

POLITECNICO DI TORINO

Facoltà di Ingegneria

Corso di Laurea in Ingegneria delle Telecomunicazioni



Tesi di Laurea

Implementazione di un Data Lake in un ambiente Aero Cargo

Relatore: Prof. Paolo GARZA

Candidato: Evariste E. FANSI NDENGOUE

INDICE

Capitolo 1.....	5
Introduzione.....	5
1.1 Descrizione del problema affrontato	5
1.1.1 Contesto generale	5
1.1.2 Problematica.....	6
1.2 Sommario	7
Capitolo 2 Inquadramento generale	9
2.1 Presentazione dell'azienda	9
2.1.1 Qualche numero di WFS	10
2.2 Presentazione della business line cargo in WFS Global	11
2.2.1 Struttura organizzativa di WFS Cargo nel mondo.....	11
2.2.2 Processi di esportazioni e messaggi Cargo IMP.....	13
Capitolo 3 Problemi da affrontare e GAP analisi	18
3.1 Definizione dei problemi da affrontare: GAP Analysis	18
3.2 Obbiettivo.....	18
3.3 Requisiti del business: KPI per una reportistica operativa	18
3.4 AS – IS: Sistema esistente in azienda	21
3.4.1 Obbiettivo.....	21
3.4.2 Descrizione del sistema	21
3.4.3 Funzionalità di base.....	21
3.4.4 Architettura	22
3.4.5 Modello di dati	24
3.4.6 Reportistica	25
3.5 Il concetto di Data Lake e lo stato futuro della soluzione.....	28
3.5.1 Data Lake	28
3.5.2 Stato futuro della piattaforma del Business Intelligence Tool in WFS	36
3.6 Gap analysis.....	41
Capitolo 4 Benchmark e soluzione proposta	44
4.1 Requisiti funzionali	44
4.2 Requisiti non funzionale tecnici	46
4.3 Benchmark delle soluzioni sul mercato.....	47

4.3.1	Benchmark delle soluzioni di Data Integration soluzioni.....	48
4.3.2	Benchmark delle soluzioni per lo stoccaggio dei dati	58
4.3.3	Concetto del Cloud datawarehouse e Benchmark delle soluzioni cloud datawarehouse.....	61
4.3.4	Benchmark delle soluzioni cloud datawarehouse.....	69
4.3.5	Benchmark della soluzione di reportistica	76
4.4	Raccomandazione della soluzione finale	84
4.5	Descrizione dell'architettura della soluzione finale	84
4.5.1	Soluzione di data integration: Informatica	84
4.5.2	Soluzione di cloud datawarehouse: Snowflake	86
4.5.3	Soluzione di Reporting: Power BI.....	93
Capitolo 5 Processo della logistica cargo specificazione dettagliata dei KPI da implementare		95
5.1	Presentazione del trasporto logistico aero cargo	95
5.2	Analisi dei requisiti	104
5.2.1	Key Performance Indicator (KPI)	104
5.2.2	Qualche linea concetti linee guide dell'attività cargo.....	105
5.2.3	Identificazioni dei KPI	107
5.3	Processo di Data Warehousing.....	110
5.3.1	Specifica dei requisiti e progettazione iniziale dei data mart	114
Capitolo 6 Progettazione iniziale del datawarehouse.....		117
6.1	Data mart Cargo operation	117
6.1.1	Specifica dei requisiti del fatto F_OPERATION_D	118
6.1.2	Specifica del fatto F_OPERATION_BUD_M	120
6.1.3	Specifica del fatto F_WORKFORCE_BUD_M.....	122
6.1.4	Specifica del fatto F_WORKFORCE_D.....	124
6.2	Data mart Finance.....	126
6.2.1	Fatto finance F_FINANCE	126
6.3	Data mart HR People.....	129
6.3.1	Fatto F_HEADCOUNTS_M.....	129
6.3.2	Fatto F_HIRES_M	131
6.3.3	Fatto F_LEAVERS_M.....	132
Capitolo 7 Progettazione concettuale finale e logica		135
7.1	Elementi in comune tra i data mart nel datawarehouse	135
7.2	Cargo Operation.....	136

7.2.1	Elementi comune tra i data mart di cargo operation.....	136
7.2.2	Progettazione concettuale finale del data mart Cargo Operation_D.....	138
7.2.3	Progettazione logica del data mart Cargo Operation_D	138
7.2.4	Progettazione concettuale finale del data mart Cargo Operations_BUD_M	140
7.2.5	Progettazione logica del data mart: OPERATION_BUD_M	141
7.2.6	Progettazione concettuale finale del data mart Workforce DAY	141
7.2.7	Progettazione logica del data mart: Workforce_Day	142
7.2.8	Progettazione concettuale finale del data mart Workforce BUD Month.....	143
7.2.9	Progettazione logica del data mart: Workforce Budget Month	143
7.3	HR People	144
7.3.1	Elementi comune tra i data mart di HR People o entità risorse umane.	144
7.3.2	Progettazione concettuale del data mart Headcounts Month.....	145
7.3.3	Progettazione logica del data mart Headcounts Month.....	146
7.3.4	Progettazione concettuale del data mart Hires	146
7.3.5	Progettazione logica Hires	147
7.3.6	Progettazione concettuale Leaves	147
7.3.7	Progettazione logica Leaves	148
7.4	Data Mart Finance	149
7.4.1	Progettazione concettuale finale FINANCE	149
7.4.2	Progettazione Logica FINANCE.....	149
7.5	Progettazione logica del Datawarehouse.....	150
Capitolo 8	Procedure di estrazione, trasformazione e caricamento (ETL).....	151
8.1	Il processo ETL	151
8.1.1	Fase 1 — Estrazione.....	152
8.1.2	Fase 2 — Trasformazione	152
8.1.3	Fase 3 — Caricamento	154
8.2	Componenti principale di Informatica Intelligent Cloud Service	154
8.3	Principio di alimentazione del Datawarehouse.....	164
8.3.1	I vari tipi di zone di stoccaggio	164
8.3.2	Regole di alimentazione	166
8.4	Processo ETL: Specificazione e design delle mappings tasks per il caricamento dei dati	168
8.5	Processo ETL per il data mart Risorse umane, operazioni e finanze.....	172
8.5.1	Implementazione in informatica delle task flow.	172
8.5.2	Taskflow del processo ETL dell'entità risorse umana	173
8.5.3	Taskflow del processo ETL dell'entità Cargo Operation: PMP.....	174

8.5.4	Taskflow del processo ETL dell'entità l'entità Finanza: SAPFC	180
8.5.5	Esecuzione delle procedure ETL.....	181
Capitolo 9 Risultati e discussioni		182
9.1	Introduzione alla reportistica.....	182
9.1.1	La reportistica offline.....	184
9.1.2	La reportistica online.....	184
9.2	BI Tool per il reporting.....	184
9.2.1	Microsoft Power BI.....	184
9.2.2	SQL Server Reporting Services (SSRS).....	191
9.3	Dashboard realisti come esito del progetto	192
9.4	Risultati e discussione	193
9.4.1	Collegamento alla sorgente dei dati	193
9.5	Presentazione dei diversi Dashboard per entità.	197
9.5.1	Dashboard Cargo Operation.....	197
9.5.2	Finance Dashboard.....	215
9.6	Problemi riscontrate e miglioramenti.....	223
9.6.1	Problemi riscontrate	223
9.6.2	Miglioramenti.....	225
Capitolo 10 Conclusione		227
10.1	Conclusione	227
10.2	Sviluppi futuri.....	228
Glossario		231
Bibliografia.....		234
Elenco figure		235
Elenco Tabelle		239

Capitolo 1

Introduzione

1.1 Descrizione del problema affrontato

1.1.1 Contesto generale

È in un ambiente altamente complesso e altamente competitivo che la maggior parte, se non tutte, le imprese operano. Questo clima altamente competitivo richiede a queste aziende di monitorare molto attentamente il mercato per non rimanere indietro rispetto alla concorrenza, rispondendo il più rapidamente possibile alle aspettative del mercato, dei loro clienti e dei loro partner.

A tal fine, i dirigenti dell'azienda, qualunque sia il loro settore di attività, devono essere in grado di svolgere i compiti sotto la loro responsabilità in questo campo. In particolare, dovranno prendere le decisioni più appropriate. Queste decisioni, che influenzeranno notevolmente la strategia dell'azienda e quindi il suo futuro, non devono essere prese con leggerezza o troppo frettolosamente, date le loro conseguenze per la sopravvivenza dell'azienda. Si tratta di prendere decisioni fondate sulla base di informazioni chiare, affidabili e pertinenti. Il problema è quindi come identificare e presentare queste informazioni alle persone giuste, tenendo presente che le aziende sono sopraffatte da una notevole quantità di dati e che i sistemi operativi "transazionali" sono limitati o addirittura incapaci di fornire tali informazioni e costituiscono quindi un supporto apprezzabile per il processo decisionale. È in questo contesto che sono nati i "sistemi decisionali". Essi forniscono ai responsabili delle decisioni informazioni di qualità su cui possono basare le loro scelte decisionali. Per fare ciò, questi sistemi utilizzano una vasta gamma di tecnologie e metodi, di cui i "datawarehouse" rappresentano l'elemento principale ed essenziale per l'implementazione di un buon sistema decisionale.

Grazie alla sua dimensione economica e alla sua posizione sul mercato della logistica aero cargo europeo e americano, l'attività quotidiana di WFS Global genera dati complessi e voluminosi. Questi dati rappresentano una preziosa fonte di informazioni che potrebbero migliorare

significativamente il processo decisionale. Tuttavia, questi dati non vengono utilizzati in modo soddisfacente, mettendo così a repentaglio il processo decisionale a tutti i livelli del Gruppo.

Il presente progetto mira a creare un sistema in grado di consolidare i dati dei sistemi transazionali e di fornire informazioni di qualità ai responsabili delle decisioni. L'obiettivo è quello di fornire ai responsabili delle decisioni dati che possano informarli e facilitare un processo decisionale rapido e informato. Un sistema di questo tipo richiede la creazione di un datawarehouse affidabile contenente le informazioni necessarie per lo svolgimento dei processi decisionali.

1.1.2 Problematica

Il problema da affrontare sarebbe la consolidazione dei dati provenienti da differenti sistemi sorgente infine di aggregarli per avere indicatori di performance significative necessario alla dirigenza/al comitato esecutivo per le prese di decisioni efficaci.

Questa tesi si svolge in un contesto dove

- I dati sorgenti sono sparsi in diversi paesi in una infrastruttura IT locale 20 Paesi nel mondo e 4 continenti.
- Le applicazioni sorgenti sono eterogenee con un impatto sulla la struttura dei modelli dei dati.
- Esistenza di un datawarehouse limitato ai fini della gestione dei dati delle operazioni nel magazzino
- Dati in silo, nessuna possibilità di incrociare i dati provenienti di dipartimenti, paesi diversi con la finalità di estrarre informazioni pertinenti per una presa di decisione efficiente.
- Questo problema da risolvere pone due maggiore problema:
- L'architettura esistente necessita di essere ridisegnata per poter estrarre, elaborare e trasmettere i dati provenienti da diversa sorgente.
- Un approccio per la risoluzione di questi problemi sarà le esplorazioni di un architettura cloud per lo stoccaggio dei dati, cloud datawarehouse per la gestione dati e in fine degli strumenti di business intelligence per la presentazione dei dati.

1.2 Sommario

WFS è una multinazionale americana la cui attività principale è la movimentazione di merci, e passeggeri a terra. Ha un'impronta globale e si trova all'interno degli aeroporti e nell'area cargo. Assiste più di un centinaio di compagnie aeree.

Nel corso della sua esistenza è stata oggetto di numerose acquisizioni seguite da trasformazioni all'interno dell'azienda. L'ultimo dato è l'acquisizione da parte di Cerberus Capital con sede a New York.

Una delle prime osservazioni fatte da Cerberus Capital, dopo l'acquisto di WFS Global, è l'inesistenza di uno strumento interno di BI per la gestione dell'azienda a livello di comitato esecutivo. WFS Global è fondamentalmente una società di servizi aeroportuali e quindi soggetta a diverse influenze esterne come Politiche (Iran Embargo), Economiche (Guerra economica tra Cina e USA) Sanitarie (virus Corona in Cina) Tecnologiche (Amazon, FedEx). Di fronte a questa constatazione, è imperativo creare uno strumento di business intelligence che consenta di restituire le informazioni e di presentarle sotto forma di KPI. I KPI definiti consentiranno ai comitati esecutivi di essere altamente reattivi alle influenze esterne che possono influire sul buon funzionamento dell'azienda. Per ovviare a questa situazione, il piano strategico prevede diversi progetti di trasformazione, uno dei quali affronta le tematiche della BI, ovvero l'implementazione di 44 KPI suddivisi nelle seguenti aree: Risorse Umane, Finanza, Relazioni con i Clienti, Operations, HSSE Health, Safety, Security and Environment.

Questa richiesta coincide con la trasformazione dei sistemi informativi in quanto influenzerà l'implementazione e l'evoluzione del datawarehouse esistente. Questa trasformazione avverrà da due punti di vista:

1. Trasformazione dell'infrastruttura che memorizzerà i dati di origine: basata sulle metodologie interne e sui criteri di selezione per le soluzioni IT di WFS Global. Sul mercato sono state individuate soluzioni di cloud storage. E la scelta del WFS si è concentrata su Amazon S3. La scelta ha due vantaggi, il primo è che WFS opta per la migrazione di tutte le sue applicazioni verso il cloud Amazon, il secondo è

l'associazione di questo spazio di archiviazione con la soluzione di cloud datawarehouse dati denominata ***Snowflake*** che approfondiremo più avanti.

2. L'implementazione di una nuova architettura data lake sotto la tecnologia a fiocco di neve su cui verrà costruito il datawarehouse globale. Ed è su questa base che si costruiranno i diversi cubi di dati, che ci permetteranno di avere i 44 KPI.

Capitolo 2

Inquadramento generale

2.1 Presentazione dell'azienda

WFS ha oltre 45 anni di esperienza nel soddisfare le aspettative di servizio dei principali clienti in tutto il mondo [1]. I nostri clienti sanno che comprendiamo la loro attività e la nostra rete globale fornisce la portata di cui hanno bisogno per fornire un servizio clienti di alta qualità in tutte le principali sedi aeroportuali del mondo. Come organizzazione di successo e in crescita, WFS ha la capacità di investire in strutture, attrezzature, sistemi e nuove opportunità di business. Lavoriamo secondo i più alti standard del settore per la sicurezza, la protezione e la qualità, e in conformità con le normative locali e internazionali. WFS offre un servizio completo che include servizi cargo, passeggeri, rampa, bagagli, servizi tecnici e premium. Il nostro team globale di persone è composto da professionisti dei servizi aeroportuali altamente addestrati, altamente qualificati e sicuri di sé, supportati dalla formazione e dallo sviluppo locale e dalla scuola di formazione interna di WFS Airport College in Francia. Siamo appassionati, capaci e sempre pronti ad esplorare nuove opportunità presentate dai clienti di tutto il mondo.

WFS, World Flight Services, è un punto di riferimento mondiale nei servizi aeroportuali e opera su piste e terminal. L'azienda è strutturata attorno a cinque attività:

- **Cargo:** assistere le compagnie aeree e i noleggiatori nello stoccaggio, nella pallettizzazione, nella preparazione dei documenti di volo e nella movimentazione delle merci pericolose.
- **Passeggero:** accompagnare i passeggeri dai controlli fino all'imbarco migliorando l'esperienza del cliente.
- **Tarmac:** eseguire operazioni a terra combinando prestazioni e sicurezza.
- **Tecnico:** Progettare e mantenere soluzioni tecniche per migliorare l'efficienza operativa.

Bagaglio: Garantire operazioni efficienti per i bagagli.

Premium: per offrire servizi personalizzati

DATA	OLTRE 45 ANNI DI CRESCITA
1971	Creazione di SFS (Société de Fret et de Services). Si afferma rapidamente come operatore aereo di successo per il trasporto merci in Europa.
1983	American Airlines (AA) costituisce AMR Services Corporation per esternalizzare i suoi servizi di rampa, passeggeri e terminal.
1993	SFS e AMR Services Corporation si fondono per creare uno dei gruppi di servizi aeroportuali leader a livello mondiale.
1996	WFS apre un ufficio regionale a Hong Kong per sviluppare il terreno, servizi di terminal merci e passeggeri in Asia.
1999	La fusione SFS/AMR viene rinominata Worldwide Flight Services (WFS) per riflettere l'espansione globale del gruppo.
2002	WFS acquisisce SEN Group e inizia un significativo sviluppo dell'assistenza a terra in Europa sotto l'ombrello di EFS European Flight Services.
2004	WFS crea Bangkok Flight Services (BFS) per gestire le licenze di assistenza a terra e di movimentazione merci nel nuovo aeroporto di Bangkok.
2006	LBO France acquisisce WFS dal Gruppo Vinci. WFS ottiene le licenze per l'assistenza a terra a Parigi CDG e Parigi ORY, ampliando ulteriormente il business del gruppo.
2008	WFS acquista una quota di 60 partecipazioni in British Airways Regional Cargo e ottiene l'accesso a una rete di 8 aeroporti nel Regno Unito.
2009	WFS India lancia i servizi di handling rampa e passeggeri a Nuova Delhi e Cochin.
2013	La joint venture WFS AFS apre nuovi magazzini cargo a Città del Capo e Johannesburg.
2014	WFS compie il primo passo nel mercato dell'assistenza a terra in America Latina acquisendo una partecipazione di controllo in Orbital Group.
2015	Platinum Equity acquisisce WFS da LBO France. WFS acquisisce una partecipazione di 51 azioni in Fraport Cargo Services.
2016	WFS acquisisce CAS in North America. Craig Smyth nominato CEO di WFS.
2018	Un'affiliata di Cerberus Capital Management, L.P. ha completato l'acquisizione di Worldwide Flight Services da Platinum Equity.

Tableau 2.1: Cronologia di WFS Global

L'esperienza di WFS si estende oltre i confini francesi, con una presenza in tutti e 5 i continenti, in 22 paesi e 198 aeroporti. WFS dispone attualmente della rete di movimentazione merci più estesa ed efficiente d'Europa.

WFS in Francia comprende 3 società, EFS on the Ground Handling che fornisce servizi a terra agli aeromobili, **FH** (*France Handling*) e **SFS** (*Société de Fret et de Services*) sul settore Cargo per garantire il trasporto aereo di merci.

2.1.1 Qualche numero di WFS

- **27 000** professionisti del servizio aeroportuale che lavorano per voi in tutto il mondo.
- **6,3 milioni di tonnellate** di merci movimentate ogni anno in tutto il mondo
- **180 sedi** globali nei cinque continenti.

- **1,3 miliardi di euro** di fatturato annuo, che ci permette di investire nel servizio di cui avete bisogno.
- Oltre **45 anni** di esperienza nel mondo dell'aviazione

2.2 Presentazione della business line cargo in WFS Global

2.2.1 Struttura organizzativa di WFS Cargo nel mondo

WFS ha un'impronta globale, quindi è presente in diversi paesi e continenti. Questa presenza globale si concretizza in due entità regionali, ciascuna guidata da un VP Executive. Avremo quindi le entità regionali seguenti:

- **EMEA:** Europa, Medio Oriente, Africa, Asia:

Questa entità è costituita per:

- Europa: Francia, Spagna, Belgio, Svezia, Danimarca, Irlanda, Regno Unito,
- ASIA: Bangkok, Hong Kong
- Medio Oriente: Qatar
- Africa: Kenya, Sudafrica
- **AMERICA:** L'America consiste principalmente di:
 - Nord America: Stati Uniti
 - Canada
- **SUD AMERICA:**
 - Brasile

Questa organizzazione è supportata da entità legali a livello nazionale e da un'entità globale che funge da holding e sede centrale con sede a Parigi. Ogni entità giuridica a livello nazionale dispone di stazioni a loro volta composte da magazzini e uffici.

Entità legale: un country manager è a capo di quest'entità. Responsabile della gestione dei clienti, delle operazioni di budget e delle previsioni.

Stazione: È associata ad un aeroporto per esempio la stazione (CDG: Charles De Gaulles, EWR: Newark ...) ogni stazione sarà codificata in modo univoco nella base secondo un quadro di riferimento predefinito (Lista degli aeroporti), ed è strutturata come segue:

Ufficio: Svolge due ruoli principali: controllo e ricevimento delle merci in seguito ai seguenti compiti:

- Elaborazione della documentazione
- Verifica del numero di omologazione
- Verifica del passaporto
- Controllo di sicurezza
- Se la merce non è assicurata, chiedere la sicurezza e restituirla alla reception.
- In caso di trasferimento sicuro della merce al negozio

Magazzino: Il magazzino è il luogo in cui le merci vengono immagazzinate e lavorate. Due processi principali supportano le attività del Magazzino:

- Esportazione
- Importazione
- Transito

L'attività operativa è supportata da uno strumento informatico chiamato Cargospot. È un software professionale che permette di registrare e gestire tutte le operazioni di trasporto aereo di merci in modo computerizzato. Viene utilizzato sia per il processo di importazione che per quello di esportazione. Ha anche come funzionalità principali:

- Comunicare inviando messaggi Cargo IMP alle compagnie aeree e poi ai clienti.
- Ricezione del carico di importazione, invio del carico di esportazione, fatturazione, tracciamento, inventario, ricezione di preallarmi, inserimento di informazioni AWB.
- **Funzione di supporto:** a supporto di tutte le sue attività operative, ci sono funzioni internazionali, HR, IT, sicurezza e protezione, Finanza, Relazione con il cliente.
- **Funzione HR:** WFS è una società di servizi che fornisce assistenza alle compagnie aeree, il cui principale asset è costituito dalle risorse umane. Ciò richiede prima di tutto una selezione adeguata delle risorse, una pianificazione equa, una gestione flessibile per adattarsi ad elementi esterni come scioperi, embargo o altri. Ci sono diverse categorie di risorse: operai, impiegati, impiegati e fornitori di servizi esterni per soddisfare alcune esigenze urgenti. L'orario di lavoro è di 7 giorni alla settimana

e 24 ore al giorno. L'organizzazione nei negozi si basa sugli orari dei voli: arrivo e partenza degli aerei.

- **Funzione IT:** Il ruolo del dipartimento IT è quello di proporre di implementare e supportare tutte le applicazioni che devono supportare i processi di business. Questi processi possono essere operativi o di supporto. Ha un'infrastruttura che varia a seconda dei diversi paesi, con server ospitati internamente ed esternamente. Su ogni sito è presente un responsabile IT che fa capo a un CIO centrale.
- **Sicurezza e protezione:** La sicurezza e la protezione è un pilastro importante nelle attività di WFS Global. Si parte dalla sicurezza delle merci e delle persone, fino alla sicurezza delle infrastrutture aeroportuali.
- **Finanza:** la finanza comprende anche la fatturazione e la contabilità. Gestisce tutti i flussi finanziari in entrata e in uscita. Responsabile del budget e delle previsioni. Consolida tutti i costi e i ricavi delle diverse entità giuridiche. Responsabile della rendicontazione finanziaria agli investitori e alle istituzioni.
- **Relazione con il cliente:** Questo dipartimento mira a gestire la relazione tra WFS Global ed i suoi clienti in previsione di potenziali non conformità osservate.

2.2.2 Processi di esportazioni e messaggi Cargo IMP

Durante l'esportazione, molti giocatori interagiscono tra loro:

- Mittenti (privati o aziende)
- Spedizionieri
- Compagnie aeree
- Operatori (rappresentanti delle compagnie aeree e/o spedizionieri)
- Guardie di sicurezza
- Autotrasportatori
- Dogana
- Agenti di rampa
- Il capitano
- I controllori del traffico aereo...

È possibile creare la seguente cronologia:

1. Un cliente di spedizione (privato/azienda) desidera esportare pacchi

2. Il cliente contatta il suo spedizioniere
3. Si stabilisce una stima in base al numero di colli, al peso, al volume, al valore commerciale e alla destinazione della merce)
4. Dopo l'accordo del cliente, lo spedizioniere programma il pre-instradamento della merce da consegnare presso il cliente.
5. Dopo il ricevimento dell'Ordine di trasporto, il vettore effettuerà il trasporto tra la sede del cliente e il magazzino doganale dello spedizioniere.
6. Il carico è assicurato (raggi X, squadra cinofila, carico libero...)
7. Lo spedizioniere effettua una prenotazione con la compagnia aerea.
8. La compagnia aerea deve trasmettere allo spedizioniere un numero di **MAN (Manifesto)**.
9. Lo spedizioniere chiede al suo **GHA (Ground Handling Agent)** di presentare il carico con l'**ETA (Estimate Time of Arrival)** alla società **GHA**
10. Il **MAN** viene inserito su Cargospot.
11. Il carico viene caricato su **ULD (Unit Load Device)** o lasciato alla rinfusa
12. Il carico deve essere messo a disposizione dell'agente di rampa per il carico a bordo dell'aeromobile.
13. Il capitano controlla la spedizione caricato a bordo e prende nota delle specifiche di carico (**DGR (Dangerous Goods Regulations)**, corpi... Usando il **NOTOC (NOTification TO Captain)**) [2].
14. I controllori del traffico aereo danno al pilota l'autorizzazione a rullare l'aereo.
15. L'aereo si sta dirigendo verso la pista per il decollo.

Standard CARGO IMP per lo scambio dei messaggi

L'attività di trasporto merci si basa essenzialmente sullo scambio di messaggi tra lo spedizioniere, la compagnia aerea e l'handler. Questa attività genera un'enorme quantità di documenti da scambiare preventivamente tra le parti coinvolte. Infine, per risolvere questo problema, è stato istituito un sistema di scambio di documenti.

L'automazione dello scambio di dati di carico tra le compagnie aeree e le altre parti funziona in modo molto più efficiente se si applicano le procedure e gli standard di

messaggistica. Cargo-IMP di IATA è lo standard ereditato per lo scambio di informazioni critiche sulle operazioni di carico.

Sviluppato congiuntamente dalla IATA, dalle compagnie aeree che ne fanno parte e da Airlines for America (A4A), Cargo-IMP è la fonte ufficiale di messaggi per le specifiche riguardanti l'assegnazione dello spazio, la lettera di trasporto aereo, il manifesto di volo, la contabilità, lo stato, la discrepanza, l'embargo, le dogane, la fatturazione CASS, le merci pericolose, le assegnazioni e il trasporto di superficie. Il Cargo-IMP include anche liste di codifica e decodifica di tutti i codici e le abbreviazioni approvate.

Il Cargo-IMP contiene le edizioni aggiornate e storiche del manuale Cargo-IMP, compresi i collegamenti di capacità di messaggistica del vettore e la richiesta di elementi di dati.

Vantaggi principali

Massimizzate i benefici delle vostre operazioni di carico attraverso una maggiore automazione della messaggistica, assicurando che le vostre procedure di scambio di messaggi di carico siano conformi alle normative e alle pratiche internazionali e locali:

- Specifiche dei messaggi, inclusi elementi di dati incorporati, codici di abbreviazione e funzionalità di stampa con funzioni di ricerca avanzate
- Tabelle interattive che includono una directory di dati elementi, codici di abbreviazione e un elenco delle capacità dell'utente
- Accesso a una fonte completa di informazioni sui messaggi di interscambio merci e sulle relative procedure

Codifica per esportazione

Un FSU (Freight Status Update) è un messaggio inviato alla compagnia aerea che le permette di conoscere e aggiornare lo stato di una spedizione nella catena di elaborazione [3].

La principale FSU per il processo di Export

FSU	Significato	Utilità
FBL	Freight Booking List	Consente al GHA su Cspot di essere preallarmato dell'arrivo di una LTA Lettera di trasporto Aero. Questo riduce la quantità di input per il GHA.
RCS	Cargo Received from Shipper	Dichiara il ricevimento del carico/merche alla compagnia
RCT	Received Cargo from another carrier	
PRE	Dichiara che il carico è stato pre-manifestato su un volo	
MAN	Dichiara che il carico è ufficialmente assegnato al volo.	
DEP	Menziona il decollo dell'aereo	
FHL	House Air Waybill List	Consente l'ingresso dell'HAWB in modo che la dogana possa controllare le merci in partenza e in arrivo in un paese.
FWB	Full Air Waybill	Permette l'inserimento completo delle informazioni sulla AWL. In questo modo, può essere consultato e stampato da tutti gli utenti Cspot.

Tableau 2.2: Standard dei messaggi Flight Status Update nel Aero Cargo Business

Importazione: Elaborazione del carico e ricezione dell'importazione, carico/Merce in uscita

D'altro canto, distinguiamo due schemi complementari:

- Un circuito Elaborazione e ricezione del carico di importazione
 - Un circuito Carico in uscita
- Troviamo in importazione diversi attori:
- Compagnie aeree
 - Trasportatori di merci dalla pista ai magazzini
 - Un team di magazzino incaricato della ricezione, della marcatura, dello stoccaggio e della fornitura di merci.
 - Un team dell'ufficio operativo incaricato della gestione amministrativa e operativa delle merci (prenotazione dei voli, consegna dei documenti, accoglienza dei clienti, ecc.)
 - Spedizionieri aerei
 - Aziende indipendenti / individui
 - Autotrasportatori
 - I doganieri

Circuito di movimentazione e ricezione della merce nel processo Import [4].

1. Il volo atterra

2. Il carico viene scaricato dall'aereo e portato al WFS dalle società di traino.

3. L'ufficio operativo recupera la busta per elaborare il volo e informa su Cspot le informazioni sul destinatario, gli addebiti...)

4. Il negozio Ecore Freight ("Ecore Company")

5. La partita è trascritta computerizzata su Cspot

6. Il destinatario viene informato dell'arrivo della sua merce (in tutto o in parte tramite un avviso di arrivo).

Nel corso del suo viaggio, un AWB passa attraverso le mani di diversi operatori, situati in diversi servizi, aeroporti e paesi. Così, le procedure IATA hanno stabilito una standardizzazione degli scambi di messaggi e delle procedure. Cspot si è ispirato a questa standardizzazione creando degli statuti per ogni AWB.

Elenco delle principali Codice d'importazioni FSU

FSU	Significato	Invio della FSU
AWD	I documenti sono stati messi a disposizione del cliente	A seguito scrittura dell'agente Rcv e del carico
RCF	Il carico/merce del volo è stato ricevuto	Al momento dell'inserimento del "ecore"
NFD	Il cliente è stato informato dell'arrivo della merce.	Al momento dell'inserimento del "ecore" (se viene menzionato l'agente Rcv)
DIS	Differenza tra il numero di pacchi ricevuti e il numero di pacchi manifestati	A seguito della variazione del numero di colli ricevuti rispetto alla quantità manifestata
	Carta AWB trovata o mancante	Quando si controlla l'AWB trovata o l'AWB mancante
RCT	È stato ricevuto il carico da un'altra compagnia aerea	Dopo aver inserito la voce AWB per l'esportazione (OK)
DLV	Il carico è stato consegnato al cliente	Dopo l'emissione della merce sulla scheda Consegna
CRC	La dogana è stata appositamente avvertita di questa spedizione	
CSN	La compagnia aerea ci informa del suo scambio di informazioni con la dogana.	Quando l'azienda viene inviata (prima è meglio è fare i nostri preparativi)
TGC	La spedizione è stata inoltrata alla Camera di Commercio.	Quando l'uscita è fatta a favore della CCI

Tableau 2.3:Elenco delle principali Codice d'importazioni FSU

Capitolo 3

Problemi da affrontare e GAP analisi

3.1 Definizione dei problemi da affrontare: GAP Analysis

Dopo aver presentato il contesto generale dell'azienda, i tipi di attività, import, Export, i processi del aero cargo e in fine i messaggi di per lo scambio di informazioni tra diversi parti. Si nota una manipolazione in grande quantità di informazioni e di dati a traverso il gruppo. Queste informazioni provengono di sorgente eterogene.

Il problema da affrontare sarebbe la consolidazione dei dati provenienti da differenti sistemi sorgente infine di aggregarli per avere indicatori di performance significative necessario alle dirigenze per le prese di decisioni efficace.

Requisiti del business, avere l'informazione in base settimanale del toneggio trattato nel magazzino. Per quanto riguarda il business ground, e i numeri di aeri/turni trattato durante la settimana.

3.2 Obbiettivo

La risoluzione del problema identificato si riassume in due maggiore obbiettivi.

- Dotare l'azienda di una infrastruttura tecnica permettendo di poter raccogliere tutte le informazioni in una singola fonte or (Repository) in modo da rompere le barriere tra diversi dipartimenti e dare la possibilità a tutti di poter usufruire i dati proveniente da ente differente.
- Creare un datawarehouse appoggiandosi sul singolo ripositorio creato come prerequisito per provvedere al top management i KPI richieste e necessari per le prese di decisioni

3.3 Requisiti del business: KPI per una reportistica operativa

Avremmo due livelli di requisiti:

- Requisiti del business: Sono requisiti ad alto livello raccolto da business che saranno può tradotto in requisiti funzionali. In questo capitolo parleremmo dei requisiti business
- Requisiti funzionali: Sono requisiti più dettagliati permettendo di poter specificare meglio il bisogno per la sua implementazione.

Dopo l'acquisizione di WFS da parte di CERBERUS, l'azienda ha definito progetti strategici e una roadmap per la trasformazione di WFS. Questa tabella di marcia fa parte del Project PIP Performance Indicator Program, che mira concretamente a creare una Balance Scorecard di 44 KPI per monitorare le prestazioni del WFS. Questi indicatori sono rappresentati per dominio come segue:

HSSE

#Kpi	HSSE
1	Occupational Safety
2	Lost time injuries /200k hours worked (WFS staff only, all hours worked)
3	Work days lost (absence) due to work related injury
4	Aircraft damage
5	Flight Safety
6	Aviation Security
7	Asset Protection
8	Equipment Damage
9	Environment
10	Number of HSSE 'self-inspections' conducted
11	Number of Near-Miss reports made
12	Regulatory Non-Compliance
13	Internal Non- Compliance

Tableau 3.1:KPI Safety and Security

FINANCE

#Kpi	Financial Performance / Shareholder
39	Yield (Revenue per Kilo / Turnaround)
40	Labour Costs per Kilo / Turnaround
41	Indirect Costs per Kilo / Turnaround
42	Total Costs per Kilo / Turnaround
43	Warehouse rental cost / sqm
44	Contribution per Kilo / Turnaround

Tableau 3.2:KPI Finance

CUSTOMER SATISFACTION

#Kpi	Customer satisfaction
16	% of DEP sent On-time
17	% of RCS sent On-time
18	% of MAN sent On-time
19	% of FAP (Flown as planned)
20	% of NFD sent On-time
21	% of RCF sent On-time
22	% Number of Flights Delay over Total Operated Flights
23	% Number of Flights Delay SEVERITY 1 over Total Operated Flights
24	% Number of Flights Delay SEVERITY 2 over Total Operated Flights
25	% Flights First Baggage Delivery On Target (15 after ATA)
26	% Flights Last Baggage Delivery On Target (35 after ATA)
27	Lost Baggage Index Lost Bags/Departed Passengers

Tableau 3.3:KPI Customer Satisfaction

OPERATION

#Kpi	Operations
28	Trend - Number of Turnarounds
29	Trend - Number of Handled Tons
30	Kilos per Hour
31	AWB per Hour
32	Paid Hours / Turns
33	% Overtime
34	% of Payroll, Third Party and Overtime over Total Paid Hours
35	Cost per Paid Hour (Payroll + Overtime + Third Party)
36	Manpower Costs per Paid Hour - WFS (Payroll + Overtime)
37	Manpower Costs per Paid Hour – Third Party
38	% Occupancy Warehouse

Tableau 3.4:KPI Cargo Operation

HR

#Kpi	People
14	Staff turnover rate
15	% Sickness over Total Payroll Paid Hours

Tableau 3.5:KPI Human resources

3.4 AS – IS: Sistema esistente in azienda

3.4.1 Obiettivo

La prima cosa da fare dopo aver raccolto i requisiti espressi dal management e operazioni e di fare una valutazione del sistema esistente con l'obiettivo di avere una percentuale di copertura dei requisiti espressi dal management. I requisiti sono i 44 KPI presentati nel capitolo precedente e copre tutti i domini dell'azienda. HR, Operazioni, Finanze, relazione clienti e Safety and security

3.4.2 Descrizione del sistema

Per il monitoraggio delle attività del cargo, WFS a messo in piedi una soluzione di reportistica, chiamato **PMP** (*Performance Management Platform*) basato su una interfaccia web. Questa soluzione è stata sviluppata e mantenuta da un software vendor esterno e specializzato delle soluzioni del contesto aero cargo. I dati integrati in questa soluzione provengono da sorgente proveniente da diversi paesi estratte principalmente da soluzione per la gestione delle operazioni, cargospot, ePIC, e Hermes. Le caratteristiche di questi dati sono, i toni, ore lavorate, ore supplementare, AWB o lettera di trasporto aereo. Frequenza è giornaliera per i toni e settimanale per le ore/ore straordinarie [5].

3.4.3 Funzionalità di base

La piattaforma di Performance Management chiamato PMP la soluzione presente da WFS Global, ha come ruolo, l'acquisizione di dati reali (storici) tramite feed automatici dai sistemi sorgente IT locali. Offre anche la possibilità di integrare i dati della previsione annuale di tonnellaggio e, del rolling forecast (Previsione a rotazione) mensile, per un uso analitico. Confrontare il tonnellaggio previsto con reale o con l'anno precedente.

Nella prima versione della soluzione installata, verranno acquisiti i volumi di tonnellaggio, “*Airway bills*” lettera di trasporto aereo, le ore e i pezzi.

Il Performance Manager fornisce una visione unificata dei dati operativi all'interno dell'organizzazione WFS. Attraverso una moderna interfaccia web i gestori delle

stazioni possono rivedere e correggere i dati delle loro stazioni. Gli amministratori delle applicazioni hanno una visione d'insieme.

Le entità organizzative si baseranno sulla struttura dell'organizzazione come definita nel **MDM Master**.

Date di chiusura e promemoria

Per l'immissione mensile è possibile definire una data di chiusura. Oltre a questi dati non è possibile l'inserimento di dati (ad eccezione di utenti specifici). Separatamente possono essere impostati dei promemoria per l'invio automatico di e-mail nel caso in cui i gestori di stazione siano in ritardo con l'inserimento dei dati.

Previsioni di fine mese

Tramite un algoritmo avanzato il volume del tonnellaggio di fine mese è stimato a qualsiasi giorno del mese. L'algoritmo tiene conto delle tendenze settimanali e mensili per magazzino/vettore. Questa stima è riportata nei Rapporti, nelle esportazioni e per il gestore della stazione per vedere come sta procedendo la sua stazione rispetto alle Rolling Forecast di quel mese.

3.4.4 Architettura

L'architettura funzionale è rappresentata dalla figura sotto. Si nota tre blocchi principale della soluzione. Da sinistra a destra:

- **Sys1...4:** I sistemi sorgente sono sistemi IT operativi locali che forniscono dati

Dati da acquisire

- **Volume del tonnellaggio:** La granularità sarà KGs per Magazzino / Cliente / Tipo di aeromobile / Giorno / Processo / Consegna / Prodotto.
- **Unità AWB (Lettera di trasporto Aero):** Granularità sarà Importo per Magazzino / Cliente / Tipo di aeromobile / Giorno / Processo / Consegna / Prodotto.

- **Pezzi:** La granularità sarà Pezzi per Magazzino / Cliente / Tipo di aeromobile / Giorno / Processo / Consegna / Prodotto.
- **Ore:** Granularità sarà Ore per Magazzino / Giorno / Tipo (Pagato, Straordinario, Malato, Terza parte ecc.) / Reparto (Magazzino, Ufficio, Rampa)

Come ulteriore livello di dettaglio, tutti i dettagli del livello di volo saranno catturati e disponibili attraverso il database.

- **MDM:** Master Data Management tool 'traduce' le diverse definizioni di codice in uso all'interno dei vari sistemi in uno standard WFS uniforme senza che ogni sistema richieda l'adeguamento del proprio standard di codifica. In può essere utilizzato per la codifica delle stazioni, tipi di aeromobili, ecc.
- **PM:** Performance Manager Module è il portale e l'interfaccia web per la presentazione e/o l'inserimento delle informazioni. I dati proveniente dal sistema sorgente può essere rivisto, corretto o approvato.
- **DW:** Datawarehouse è un'applicazione che trasforma i dati della tabella dei fatti da PM in cubi aggregati per un uso facile. Questi cubi di reporting costituiscono la base per i rapporti in linea, con possibilità di estrarre in Excel o PDF, come prima sorgente di dati per Tableau (server e desktop), e input per altre applicazioni aziendali (come

SAP)

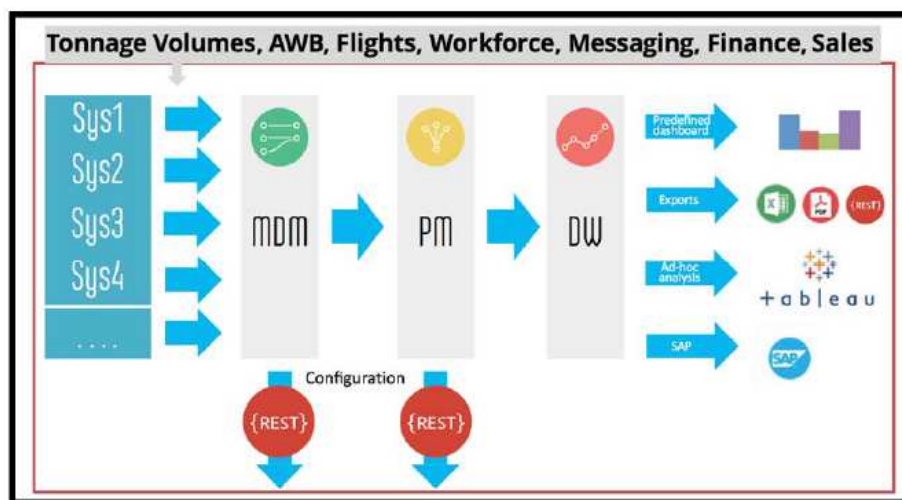


Figure 3.1:Architettura della piattaforma PMP

3.4.5 Modello di dati

Il cuore della soluzione è un datawarehouse rudimentale con 15 dimensioni come presentata nella figura 1, con 9 tabella dei fatti.

Come spiegato prima, la prima funzionalità di PMP sarà di fornire lo stato attuale delle attività nei magazzini, le misure essenziale per questa reportistica sono:

- **Pezzi:** In numero di pezzo manipolato nel magazzino
- **Tonneggi:** Toni trattati nel magazzino
- **AWBs (Air Way Bill):** Numero di Air Way Bill elaborato negli uffici di preparazione
- **Costo:** Il costo sarà calcolato dipendentemente delle dimensioni

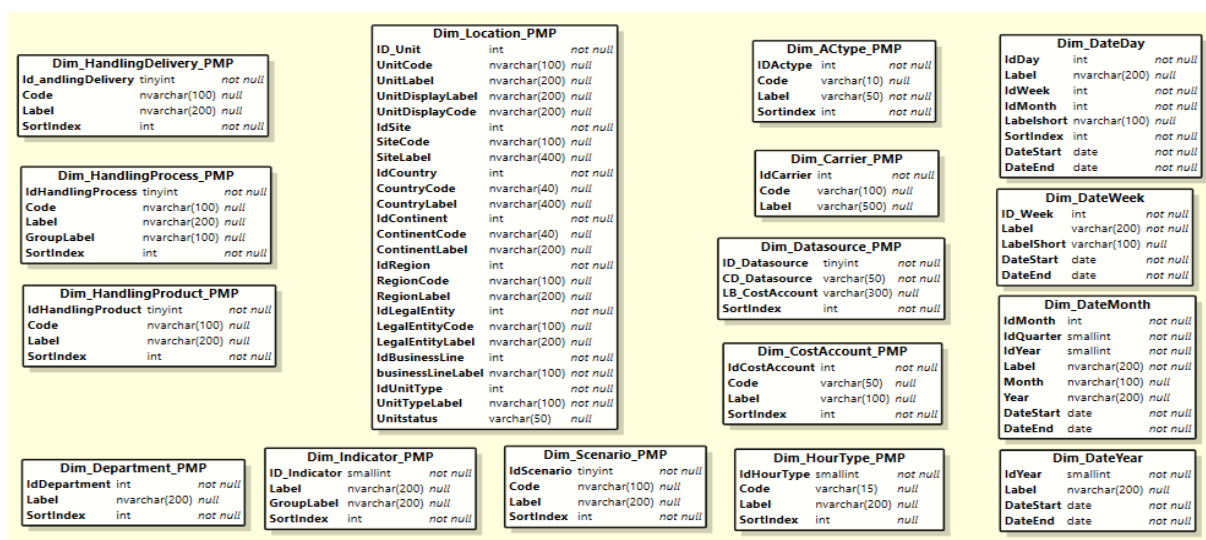


Figure 3.2: Dimensione del datawarehouse di PMP

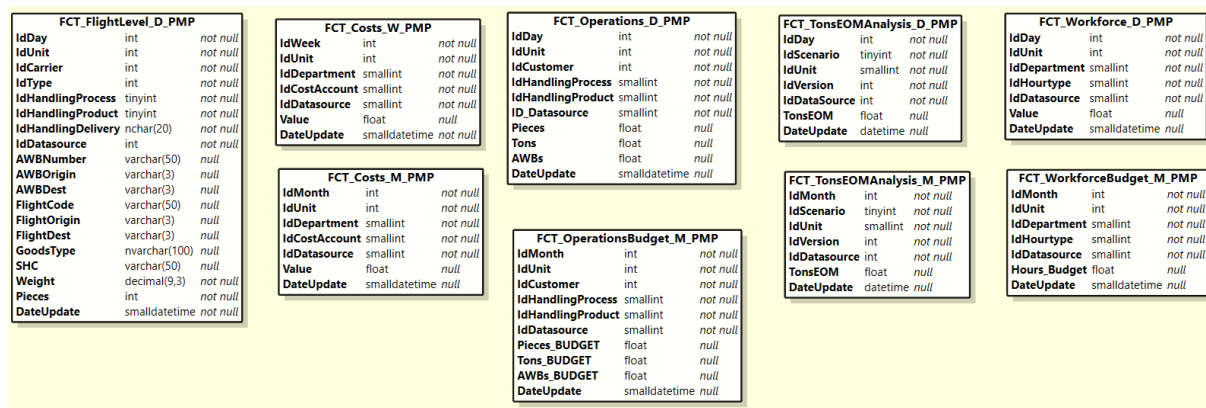


Figure 3.3: Tabelle dei Fatti del Datawarehouse PMP

3.4.6 Reportistica

Rapporti commerciali

Performance Manager contiene report standard predefiniti sulle tipologie di attività che sono disponibili per gli utenti del sistema.

Rapporti sulla qualità dei dati

Diversi rapporti di qualità dei dati creati per la divisione Cargo Operation saranno disponibili nella piattaforma Cargo Data Management. I report mostreranno la deviazione Actual-Budget al mese, con la possibilità di attivare un'e-mail per casi eccezionali.

Interfacce di importazione ed esportazione

I dati operativi possono essere importati nel sistema in cinque modi:

- Tramite alimentazione automatica (modalità preferita)
- Tramite inserti manuali / correzioni. Utile se non è disponibile un'alimentazione automatica per una stazione o se i dati non sono affidabili.
- Tramite caricamento del foglio di calcolo. Per caricare i dati può essere utilizzato un file di foglio di calcolo appositamente formattato. I dati all'interno del sistema possono essere esportati in vari modi:
- Esportazione XLS: tramite un'interfaccia intuitiva tutti i dati del sistema possono essere esportati in varie granulometrie
- File TXT appositamente formattato per l'importazione in SAP

Piattaforma di reporting tableau

Il software Tableau è leader nell'analisi della BI e negli strumenti di reporting. Questo software può essere utilizzato per interrogare in modo interattivo il Performance manager Datawarehouse (o altre fonti di dati, come banche dati, fogli di calcolo, cubi OLAP ecc.)

Applicazione desktop Tableau

I Power User possono utilizzare il desktop Tableau per interrogare in modo interattivo il datawarehouse di Performance Manager. I report e i cruscotti ad hoc possono essere facilmente creati e condivisi.

Applicazione Tableau Server

Il server Tableau consente alla piattaforma Performance Manager di estrarre i report predefiniti (creati dagli utenti desktop di Tableau) come mostra la figura 3.4 e di renderli disponibili in formato PDF agli utenti non autorizzati.

Overall

Payroll Paid Hours	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Warehouse	4,298.70	4,298.70	4,298.70	4,298.70	4,298.70	4,298.70	4,298.70	4,298.70	4,298.70	4,298.70	4,298.70	4,298.70	51,584.40
Office	2,507.57	2,507.57	2,507.57	2,507.57	2,507.57	2,507.57	2,507.57	2,507.57	2,507.57	2,507.57	2,507.57	2,507.57	30,090.90
Ramp													
Ops Support	358.23	358.23	358.23	358.23	358.23	358.23	358.23	358.23	358.23	358.23	358.23	358.23	4,298.70

Overtime Hours

Warehouse	256.43	256.43	256.43	256.43	256.43	256.43	256.43	256.43	256.43	256.43	256.43	256.43	3,077.12
Office	149.58	149.58	149.58	149.58	149.58	149.58	149.58	149.58	149.58	149.58	149.58	149.58	1,794.99
Ramp													
Ops Support	30.09	30.09	30.09	30.09	30.09	30.09	30.09	30.09	30.09	30.09	30.09	30.09	361.09

3rd Party Hours

Warehouse	708.45	472.30	551.02	747.81	1,023.32	826.53	1,062.68	983.97	708.45	708.45	708.45	708.45	9,209.91
Office	125.02	83.35	97.24	131.97	180.59	145.86	187.53	173.64	125.02	125.02	125.02	125.02	1,625.28
Ramp													
Ops Support	17.01	11.34	13.23	17.95	24.57	19.84	25.51	23.62	17.01	17.01	17.01	17.01	221.13

Figure 3.4:Esempio di rapporto prodotto da PMP



Figure 3.5:Esempio di rapporto prodotti da PMP sul toneggio mensile

Dopo elaborazione dei dati estratti da PMP e puoi importati in un file Excel si ottiene un rapporto più dettagliato come monstra la figura 3.6; Questo rapporto contiene le metriche operativa e finanziaria.

PROJECTIONS END OF THE MONTH		Generated on avril 11, 2019				
LEVEL N-1 Station 1		MTD	Month Estimate	Budget	Budget Deviation	Last Year Last Year Deviation
OPS METRICS						
Tonnage	19 539	60 118	65 587	● -5 469	65 077	● -4 959
Productive Hours WFS Employees excluding Overtime	15 605	62 420	69 030	● -6 610	62 719	● -299
Overtime WFS	705	2 821	3 119	● -299	2 834	● -14
Third Party Hours	1 985	7 939	8 779	● -841	7 977	● -38
Other Non-Productive	0	0	0	● 0	0	● 0
KG per Man/Hour	1 068	822	810	● 11	885	● -64
KG per AWB's	650	620	615	● 5	695	● -75
FINANCIAL METRICS						
	Last Month	Month Estimate	Budget	Budget Deviation	Last Year	Last Year Deviation
Revenues	1 465 425	4 508 850 (1)	5 246 953	● -738 103	5 010 893	● -502 043
Yield	75	75	80	● -5	77	● -2
Manpower Costs	879 255	3 517 020 (2)	3 641 799	● -124 779	3 382 400	● 134 620
Cost per Hour	48	48	45	● 3	46	● 2
Rent Costs	250 000	250 000 (3)	250 000	● 0	225 000	● 25 000
Other Costs	150 000	150 000 (4)	150 000	● 0	135 000	● 15 000
EBITDA	186 170	591 830	1 205 155	● -613 325	1 268 493	● -676 663
	12,70%	13,13%	22,97%	● -9,84%	25,31%	● -12,19%

Figure 3.6:Rapporto finale con misure ops e finanze

I limiti di questo sistema

- Il sistema dovrebbe essere una soluzione globale e non tiene conto dei dati proveniente da soluzioni sulla sicurezza aeroportuali, risorse umane, finanze e relazione cliente.
- Le soluzioni per la gestione del magazzino sono eterogenee, distinguiamo tre soluzioni diversi per la gestione delle operazioni Cargo all'interno del Gruppo WFS, il che rende complessa l'estrazione e l'allineamento dei dati.
- Per rimediare al problema di eterogena della soluzione sorgente si e proposto un interfaccia sviluppate per l'inserimento manualmente i dati nel sistema PMP, e anche

alternativamente i flat file messi a disposizione degli utenti per l'inserimento dei dati con tutti i rischi che ciò comporta e l'impatto che questo ha sul caricamento dei dati nelle ETL.

- Questa soluzione è una soluzione pronta all'uso sviluppata in passato per un'azienda concorrente, pertanto il modello dei dati non è adattato al contesto operativo, il che rende la soluzione rigida a qualsiasi richiesta di evoluzione o personalizzazione.
- La soluzione è oggi una delle premesse che richiede compiti di amministrazione del server e rende rigida l'evoluzione dei database e dell'infrastruttura e infine l'ambiente della soluzione PMP.

3.5 Il concetto di Data Lake e lo stato futuro della soluzione

Il concetto di data lake è un concetto relativamente recente (2011), la cui paternità del termine è comunemente attribuito a James Dixon, CTO di Pentaho, specialista in integrazione e analisi dei dati. L'idea del data lake è nata dalla consapevolezza che le aziende stanno sottoutilizzando il loro capitale informativo e che presto non saranno in grado di far fronte alla sua crescita esponenziale in termini di volume e velocità, e che gli approcci tradizionali alla cattura e all'elaborazione di questa massa di informazioni hanno raggiunto i loro limiti.

L'emergere del concetto di data lake ha subito un'accelerazione grazie alla convergenza tra la necessità di piattaforme di federazione nelle aziende e i nuovi mezzi tecnici economici portati dalle tecnologie dei big data.

Concetto legato al movimento big data, il data lake designa uno spazio di archiviazione globale per le informazioni presenti all'interno di un'organizzazione. L'idea è di farlo con una flessibilità sufficiente per interagire con i dati, siano essi grezzi o altamente raffinati. Una delle chiavi di questa flessibilità è l'assenza di uno schema rigido imposto ai flussi in entrata. Questo permette di inserire tutti i dati, indipendentemente dalla loro natura e origine. Al di là dell'archiviazione, una delle sfide del data lake è quella di poter elaborare e trasformare molto facilmente le informazioni per accelerare i cicli di innovazione, e quindi essere un supporto per diverse iniziative sui dati.

Un data lake è considerato anche un metodo di memorizzazione dei dati usato dai big data. I dati sono conservati nei loro formati originali o sono trasformati molto poco.

- Un data lake contiene quindi dati di tipo diverso, come ad esempio:
- Dati strutturati provenienti in particolare da database relazionali (righe e colonne),
- Dati semi-strutturati (CSV, log, XML, JSON...),
- Dati non strutturati (e-mail, documenti, PDF),
- File blob (immagini, audio, video, ecc.).

Questi dati vengono così conservati nel database fino a quando non sono necessari per soddisfare le esigenze dell'utente. Si tratta di una soluzione di gestione dei dati ibrida e diversificata, in grado di adattarsi ai big data e di generare livelli di analisi in tempo reale senza precedenti per le aziende.

La qualità principale del data lake sta nella sua flessibilità. Deve essere in grado di interagire con i dati, indipendentemente dal loro formato. Quando i dati sono integrati nel Data Lake, gli viene assegnato un identificatore univoco e vengono contrassegnati con una serie di tag di metadati estesi. In caso di necessità, è sufficiente interrogare il Data Lake per ottenere informazioni rilevanti. Il set di dati più piccolo che ne risulta può essere analizzato per soddisfare la necessità.

Ognuno ha così un'importante fonte di informazioni grezze che può sfruttare per effettuare le proprie operazioni, evitando così la perdita di tempo legata alla formattazione delle informazioni necessarie per la loro conservazione. Per trovare le giuste informazioni memorizzate in un Data Lake, sono possibili diversi tipi di analisi, come ad esempio le query SQL, l'analisi dei dati di grandi dimensioni, la ricerca full-text, l'analisi in tempo reale o l'apprendimento automatico. Poiché le organizzazioni con datawarehouse vedono i vantaggi dei data lakes, stanno evolvendo il loro datawarehouse per includere i data lakes.

3.5.1 Data lake & Datawarehouse

La tentazione è molto spesso forte di confrontare il data lake con un classico datawarehouse, ma le differenze tra i due sono importanti, e questo su più livelli. Il data lake è progettato per assorbire i flussi di dati grezzi e renderli utilizzabili trasformandoli per soddisfare le diverse esigenze di analisi. Alla fine, questo rimane estremamente classico e non porta nulla di nuovo

a ciò che il trio "ETL - datawarehouse - datamart" potrebbe fare. Tuttavia, questo nuovo approccio è diverso in quanto permette di caricare i dati e poi di trasformarli per renderli utilizzabili. Le iniziative sui dati sono molto spesso limitate dalle difficoltà inerenti alle fasi di raccolta e di ingestione dei sistemi. Su questo punto, essere in grado di caricare i dati su una piattaforma in uno stato quasi grezzo e poi iterare rapidamente per utilizzarli è un innegabile vantaggio. Questo è spesso indicato come approccio ELT (Extract-Load-Transform) piuttosto che ETL (Extract-Transform-Load) a cui siamo abituati. Dove un datawarehouse spinge i dati dalle sue origini ai suoi consumatori lungo un percorso relativamente fisso dove ogni datamart deve soddisfare un'esigenza, qui abbiamo una flessibilità molto maggiore. Spetta infatti a ciascun consumatore materializzare il suo bisogno ed estrarre i diversi dati di origine e poi combinarli per dargli un senso.

Un altro fattore che differenzia il data lake dal suo antenato è il lato operativo che può essere associato ad esso. La capacità di ingerire flussi in tempo reale e di reagire ai dati permette alle applicazioni di interagire direttamente su di essi. Questo va oltre l'aspetto di Business Intelligence del datawarehouse, la creazione di valore non si limita più all'utilizzo dei dati a fini di reporting.

	DATA WAREHOUSE	DATA LAKE
<i>DATI</i>	Puliti Strutturati	Grezzi Strutturati, semi-strutturati, non strutturati
<i>SCHEMA</i>	Schema-on-write	Schema-on-read
<i>MEMORIZZAZIONE & COSTO</i>	Grandi volumi di dati a costi moderati	Volumi di dati estremi a basso costo
<i>AGILITA'</i>	Configurazione fissa, pochi cambiamenti	
<i>ACCESSO</i>	SQL e strumenti di BI	Programmi ad-hoc
<i>UTENTI</i>	Professionisti del business	Data scientist
<i>VANTAGGI</i>	Maturo Facile da usare Performance, sicurezza ed integrazione garantite	Meno maturo e meno facile da usare Possibilità di eseguire analisi complesse ed analisi esplorative Altamente scalabile

Tableau 3.6:Confronto tra datawarehouse e data lake

Architettura del data lake

All'interno di un Data Lake, le zone consentono la separazione logica e/o fisica dei dati che mantiene l'ambiente sicuro, organizzato e agile. In genere, l'uso di 3 o 4 zone è incoraggiato, ma possono essere sfruttate meno o più zone. La figura 3.6 mostra un'architettura standard, possono essere semplici o complesse aggiungendo o sottraendo ulteriori zone, con la denominazione delle zone proposte da Zaloni [6].

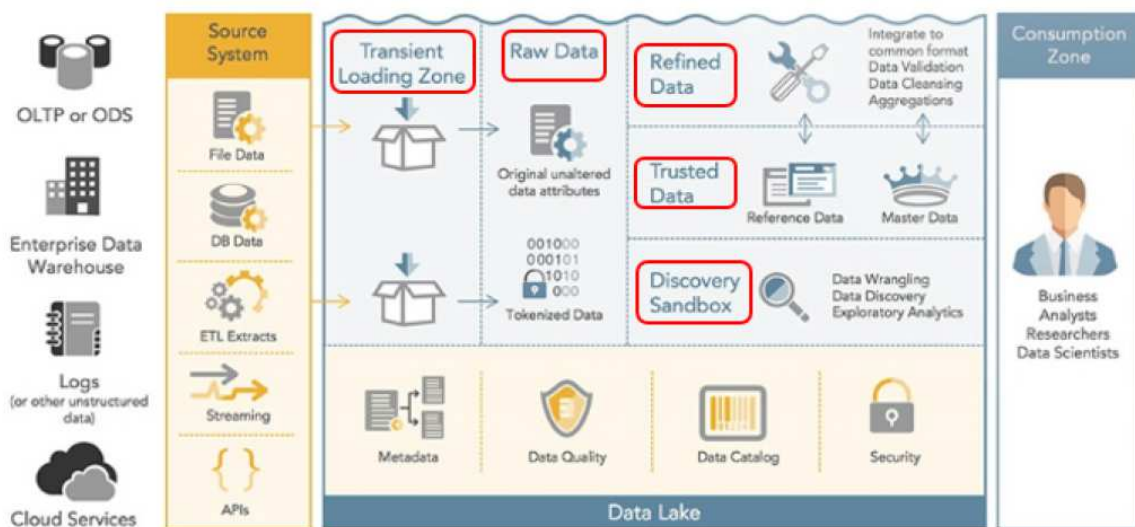


Figure 3.7: Arquitectura del Data Lake

Un sistema generico a 4 zone può includere quanto segue:

Raw Data Zone

Nella zona grezza i dati vengono spostati nel loro formato nativo, senza trasformazioni o vincolati ad alcuna regola aziendale. Spesso l'unica organizzazione o struttura aggiunta in questo livello è quella di delineare quali dati provengono da quale sistema sorgente. Anche se tutti i dati iniziano nella zona grezza, è un paesaggio troppo vasto per gli utenti meno tecnici. Gli utenti tipici sono gli sviluppatori ETL, gli amministratori dei dati, gli analisti dei dati e gli scienziati dei dati, che sono definiti dalla loro capacità di ricavare nuove conoscenze e intuizioni in mezzo a grandi quantità di dati. Questa base di utenti tende ad essere piccola e passa molto tempo a setacciare i dati, per poi spingerli in altre zone.

Trusted Data Zone

I dati di origine vengono ingeriti nell'EDW, quindi utilizzati per costruire data mart condivisi nella zona dati di fiducia. La terminologia è standardizzata a questo punto. La zona dati di fiducia contiene dati che servono come verità universale in tutta l'organizzazione. Un gruppo più ampio di persone ha applicato un'ampia governance a questi dati, che ha definizioni più complete che l'intera organizzazione può sostenere.

Refined Data Zone

Il significato viene applicato ai dati grezzi in modo che possano essere integrati in un formato comune e utilizzati da specifiche aree di attività. I dati nella zona raffinata sono raggruppati in data mart. Il data mart diventa la fonte di verità per specifici domini. Prendono sottoinsiemi di dati dal pool più grande e aggiungono un valore significativo a una finanza, alle operazioni, alla supply chain o ad altre aree amministrative.

I dati raffinati vengono utilizzati da un ampio gruppo di persone, ma non sono ancora approvati da tutti i membri dell'organizzazione. In altre parole, le persone al di fuori di specifiche aree tematiche potrebbero non essere in grado di ricavare un significato da dati raffinati. Un Data mart viene promosso alla zona di fiducia quando le definizioni applicate ai suoi dati si sono estese a un gruppo di persone molto più ampio.

La zona di esplorazione (Sandbox)

Chiunque può decidere di spostare i dati dalle zone grezze, fidate o raffinate nella zona di esplorazione. Qui, i dati di tutte queste zone possono essere trasformati per uso privato. Una volta che le informazioni sono state esaminate, vengono promosse per un uso più ampio nella zona dei dati raffinati.

Gestione dei metadati

La Metadata Management, o gestione dei metadati in italiano, consiste nell'amministrazione delle informazioni sui dati esistenti. Scoprite la definizione precisa di questo concetto, i suoi vantaggi e alcuni suggerimenti per implementare una buona gestione dei metadati nella vostra azienda.

I metadati sono le informazioni che descrivono i diversi aspetti di un insieme di informazioni (dati). I metadati sono le informazioni che descrivono i diversi aspetti di un insieme di informazioni (dati) e forniscono un contesto per la comprensione dei dati. Questo prolunga la vita dei dati permettendo alle aziende di riutilizzarli più volte.

Abilita la catalogazione e la ricerca dei metadati

Qualsiasi progetto di data lake dovrebbe incorporare una strategia di memorizzazione dei metadati per consentire agli utenti aziendali di cercare, localizzare e conoscere i set di dati disponibili nel lago. Mentre il data warehousing tradizionale memorizza un insieme fisso e statico di definizioni e caratteristiche di dati significativi all'interno del livello di memorizzazione relazionale, il data lake storage ha lo scopo di supportare in modo flessibile l'applicazione dello schema al momento della lettura. Tuttavia, ciò significa che è necessario un livello di archiviazione separato per ospitare i metadati di catalogazione che rappresentano il significato tecnico e commerciale. Mentre le organizzazioni a volte si limitano ad accumulare contenuti in un data lake senza uno strato di metadati, questa è una ricetta certa per creare una palude di dati ingestibile invece di un utile data lake. Esiste un'ampia gamma di approcci e soluzioni per garantire la creazione e il mantenimento di metadati appropriati. Ecco alcuni importanti principi e modelli da tenere a mente.

La sensata strutturazione di un data lake (Figura 3.8) non garantisce il successo nella gestione dei dati al suo interno. Per evitare che i data lakes si trasformino in data swamps (figura 3.9), la gestione dei metadati è essenziale. I metadati possono aiutare gli utenti a trovare i dati che corrispondono alle loro esigenze, ad accelerare gli accessi ai dati, a verificare l'origine dei dati e la cronologia dell'elaborazione per guadagnare fiducia e trovare dati rilevanti per arricchire le loro analisi. [12].

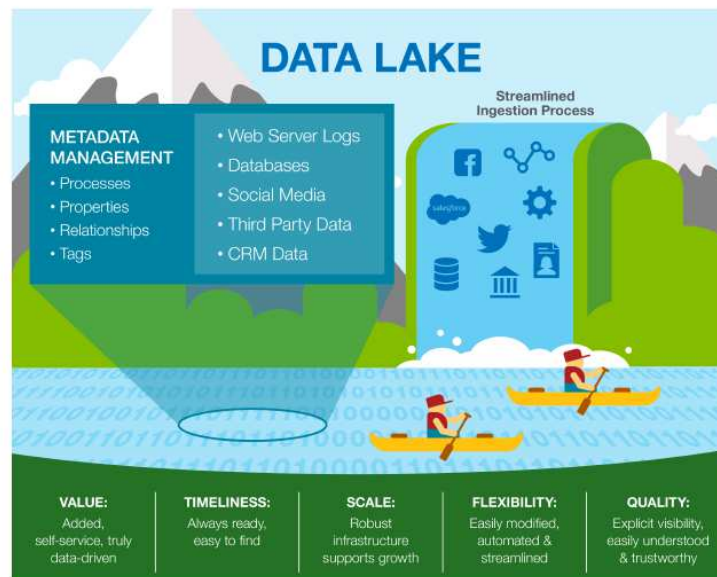


Figure 3.8:Caratteristiche del Data Lake

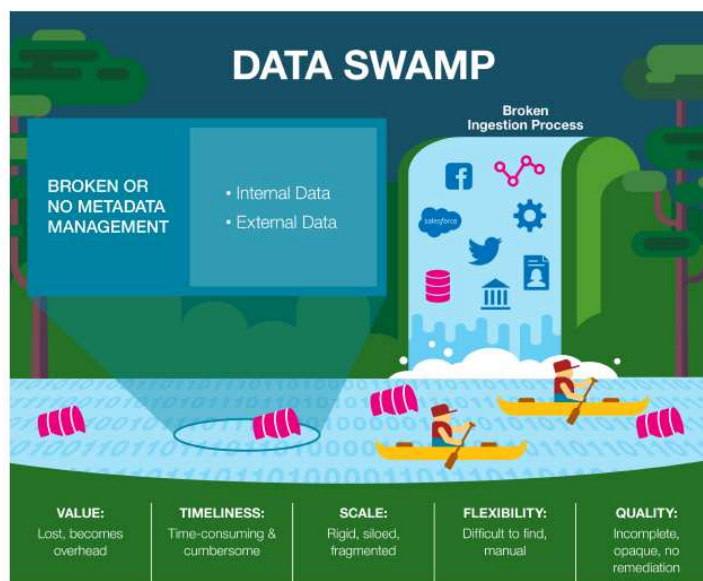


Figure 3.9:Caratteristiche del data Swamp

I vantaggi dei metadati

Come già detto, grazie ai metadati, una raffinata mappatura dei dati permette una contestualizzazione delle informazioni e porta molti vantaggi nella fase di sfruttamento: risparmio di tempo, gestione della conformità, qualità.

Le virtù dei metadati sono molteplici, tra le altre

- Accesso alle informazioni da parte di utenti senza conoscenze tecniche. I dati creano valore quando vengono sfruttati dall'azienda per scopi commerciali. Con una contestualizzazione dettagliata, è facile per tutti accedere alle informazioni richieste utilizzando un semplice motore di ricerca che indicizza la struttura, il contenuto, la qualità e la natura di ogni dato.
- Qualità dei dati. La valutazione della qualità dei dati è più facile. Una volta qualificati, i dati non hanno più bisogno di essere setacciati per vedere se sono utili, tempestivi e rilevanti. I metadati hanno fatto il lavoro.
- Risparmia tempo. Dando un profilo completo e dettagliato ad ogni dato, il tempo dell'utente è speso per operare, non per valutare, le informazioni.
- I metadati aiutano a proteggere i dati sensibili. Il regolamento della DPMR impone un nuovo quadro per l'utilizzo dei dati personali. La mappatura dei dati personali e dei dati sensibili facilita il lavoro di protezione (cifatura, gestione degli accessi...) di questi dati e l'aggiornamento del registro di trattamento, richiesto dal regolatore.
- Funzionamento più semplice e collaborazione. La tracciabilità fornita dai metadati fornisce trasparenza sulle elaborazioni effettuate. Questa conoscenza dà all'utente una maggiore fiducia nell'utilizzo dei dati, ad esempio, assicurando che l'impatto di queste azioni sui dati sia sotto controllo.
- Scoprire i dati nascosti. L'azienda ha spesso un serbatoio di dati nascosti, e quindi viziati. Possono provenire da mainframe, applicazioni complesse o altro, il che rende quasi impossibile la loro analisi e il loro sfruttamento. Una volta mappati con i metadati, questi set di informazioni vengono alla luce e possono essere sfruttati dalle linee di business dell'azienda.

A questi vantaggi, possiamo anche aggiungere la rinnovata tranquillità del dipartimento IT di fronte ai vincoli commerciali e normativi (RGPD), una volta assicurata l'implementazione operativa di questa governance dei metadati.

Data lake on premise o nell cloud?

In generale, i data lakes sono implementati in loco, con memorizzazione su Hadoop Distributed File System (HDFS) ed elaborazione tramite YARN nei cluster Hadoop. Hadoop è una soluzione scalabile e a basso costo che offre prestazioni eccellenti grazie al vantaggio

intrinseco della "localizzazione" dei dati (dati ed elaborazione risiedono nella stessa posizione).

Tuttavia, ci sono alcune sfide nella creazione di un'infrastruttura on-premise:

- Spazio - I server potenti occupano molto spazio, con conseguente aumento dei costi.
- Installazione e configurazione - Le fasi di approvvigionamento dell'hardware e di installazione/configurazione di un data center sono tipicamente complesse e possono richiedere diverse settimane o mesi per essere messe in funzione.
- Scalabilità - Se è richiesta una capacità di stoccaggio supplementare, l'operazione richiede tempo e fatica a causa dell'aumento dello spazio richiesto e della necessità di ottenere l'approvazione delle spese da parte dei responsabili delle decisioni.
- Valutazione delle esigenze - Poiché la scalabilità non è più facile in loco, è molto importante valutare correttamente il fabbisogno di materiale all'inizio del progetto. E poiché i dati aumentano ogni giorno, ma con volumi variabili, questo tipo di stima è difficile da effettuare.
- Costo - Le stime dei costi sono in genere più elevate in loco che nel cloud.

D'altra parte, un lago di dati sulle nuvole può aiutarvi ad affrontare queste sfide. I vantaggi di un lago di dati sulle nuvole sono i seguenti:

- Più facile e veloce da impostare. Piuttosto che imporre un approccio di trasformazione radicale, il cloud permette agli utenti di procedere in modo incrementale.
- Implementazione a basso costo con modello di fatturazione pay-per-use.
- Più facile da scalare man mano che le esigenze crescono, eliminando il doppio stress della valutazione dei bisogni e dell'approvazione delle spese.
- Ai vantaggi in termini di costi si aggiunge anche la riduzione del fabbisogno di spazio.

3.5.2 Stato futuro della piattaforma del Business Intelligence Tool in WFS

Dopo aver presentato in precedenza il concetto di data lake, discuteremo il design dell'architettura target al centro della piattaforma di business intelligence.

L'architettura target per il data lake sarà un'architettura strutturata su quattro livelli. Come mostrato in figura 3.10, un livello sarà associato ai sistemi sorgente, il secondo con il caricamento e la trasformazione dei dati, il terzo con l'archiviazione e l'analisi dei dati e il quarto con la presentazione delle informazioni agli utenti finali.

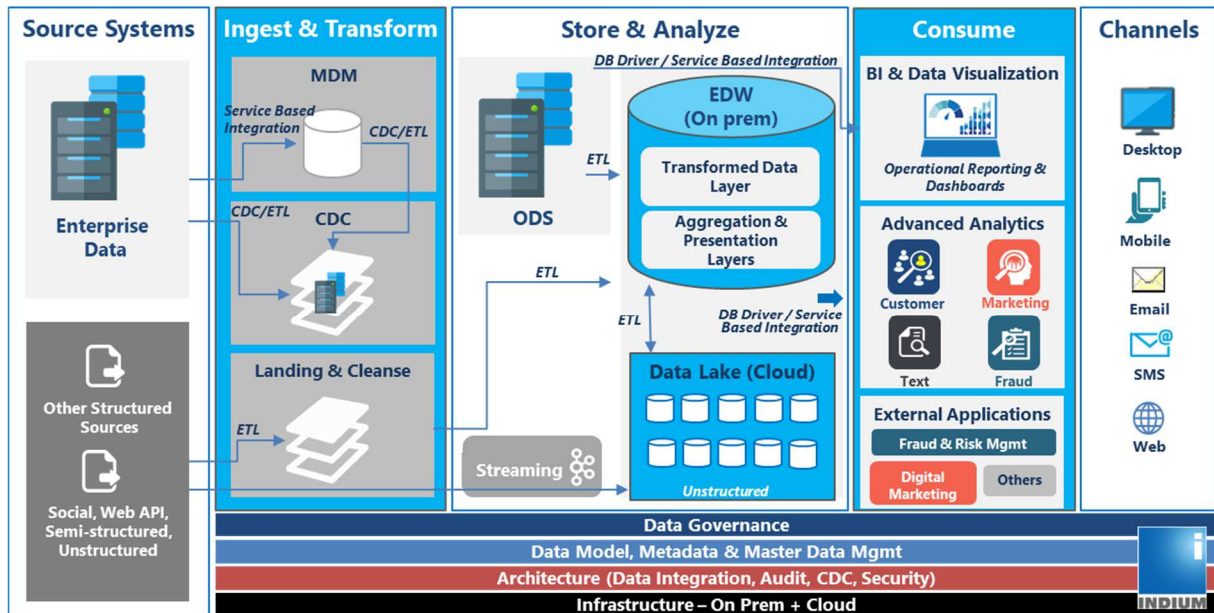


Figure 3.10:Architecture Target della piattaforma BI

Source system

I sistemi sorgente identificati nel progetto sono di tre tipi:

- Sorgenti di dati da applicazioni esistenti al WFS, incluso CargoSpot per i dati operativi. In questo specifico contesto il tool ETL non caricherà i dati direttamente da CargoSpot ma lo farà attraverso la soluzione PMP che è la soluzione di BI esistente. Questa fonte di dati consolida i dati per la gestione delle operazioni di cargo da tutti i siti WFS del mondo. Avremo anche fonti di dati provenienti dalle applicazioni CQC, la soluzione per la gestione delle relazioni con i clienti,
- La seconda tipologia di dati è costituita dai dati delle applicazioni SAAS. Il portafoglio di applicazioni WFS è costituito da un'applicazione SAAS di terze parti ospitata nel cloud. Pertanto, l'importazione dei dati dai suoi sistemi di sorgente al data lake sarà effettuata tramite API o Web Services. Inoltre, uno dei criteri di

valutazione per i sistemi di integrazione dei dati sarà questa capacità di estrarre i dati dal cloud attraverso API o Web Services. WFS ha come soluzione ospitata nel cloud la soluzione PULSE del dipartimento HSSE per la gestione di incidenti, problemi e rischi negli ambienti aeroportuali. La soluzione Tagetik per la pianificazione del budget e molto altro ancora.

- L'ultima tipologia di sorgenti di dati è costituita da flat file e dall'inserimento manuale dei dati in un sistema sorgente attraverso un'interfaccia web creata a tale scopo. Il flat file verrà fornito dalle risorse umane su un modello costruito su un file Excel. Questo file è comune a tutti i siti WFS Global. Un'altra sorgente sotto forma di flat file è costituita dai dati finanziari estratti da SAP per essere caricati nel data lake. Per inserire i dati attraverso l'interfaccia manuale sono i dati sul carico di lavoro in ore del personale operativo. Le sue informazioni vengono inserite e memorizzate manualmente nella soluzione PMP.

In conclusione, il futuro data lake deve essere in grado di ricevere dati di ogni tipo da qualsiasi sorgente e in qualsiasi formato.

Ingest and transform

Il blocco di ingestione e trasformazione è uno dei pilastri del futuro data lake, sarà responsabile dell'integrazione dei dati dalle sorgenti al data lake. E con l'obiettivo di produrre dati puliti e facilmente accessibili che possano essere sfruttati in modo efficiente dall'analitica, dalla business intelligence e/o dalle operazioni commerciali. La raccomandazione per questo strumento di integrazione dei dati è una piattaforma cloud. Dovrebbe essere strutturato come segue:

- Uno strumento per caricare i dati dalla sorgente all'area di elaborazione delle informazioni nel lago dati, chiamata zona di atterraggio. Con questa capacità di organizzare i diversi approcci al caricamento dei dati: API, servizi web, ...
- Una zona di scarico dove avverranno le estrazioni, le trasformazioni e il caricamento dei dati nel datawarehouse, nel data mart o nel cubo.
- La trasformazione dei dati sarà essenziale nella futura piattaforma perché consiste nell'applicare le regole interne dell'azienda ai dati grezzi per soddisfare i requisiti di

reporting: i dati grezzi vengono puliti e convertiti nei formati di report appropriati (se i dati non vengono puliti, diventa più difficile applicare le regole interne di reporting). La trasformazione applica le regole definite internamente. Gli standard che garantiscono la qualità e l'accessibilità dei dati devono tenere conto delle seguenti pratiche:

- **Standardizzazione:** definire i dati da elaborare, il loro formato e la modalità di memorizzazione e altre considerazioni di base che definiranno i passi successivi.
- **Deduplicazione:** segnalare i dati duplicati ai responsabili della governance dei dati; escludere e/o rimuovere i dati ridondanti.
- **Verifica:** eseguire controlli automatici per confrontare dati simili, come la durata delle transazioni o il tracciamento degli accessi. I compiti di verifica eliminano i dati inutilizzabili e segnalano anomalie del sistema, dell'applicazione o dei dati.
- **Ordinamento:** massimizzare l'efficienza del datawarehouse raggruppando e memorizzando gli oggetti per categoria. Le vostre regole di trasformazione ETL determinano la categorizzazione di ogni oggetto e la sua prossima destinazione. Il processo ETL viene talvolta utilizzato per generare tabelle di aggregazione che vengono poi proposte nei report di base; in questo caso è necessario ordinare e poi aggregare i dati.
- **Caricamento:** L'ultima fase del processo di integrazione dei dati è il caricamento dei dati estratti e trasformati nella nuova posizione. In generale, i magazzini dati supportano due modalità di caricamento dei dati: a pieno carico e a carico incrementale.

MDM (Master Data Management)

La gestione del data repository è una componente importante nell'architettura del nuovo sistema di reporting da implementare.

- I dati del deposito saranno codificati e mantenuti in un unico luogo, riducendo i costi operativi di manutenzione e codifica.

- Il sistema è il master dei dati. Lo controlla selezionando quali dati trasmette a quale sistema.
- Il sistema contiene un'unica versione attiva (possono esserci diverse versioni inattive, sia passate che future; questo è anche raccomandato per una soluzione più flessibile). È quindi garante dell'unica versione della verità e, in caso di controversia, la sua versione prende il posto della versione ufficiale.

Immagazzinare e analizzare

Questa parte dell'architettura sarà il cuore del data lake e sarà costituita da uno spazio di archiviazione dati nel cloud e da un datawarehouse nel cloud. I componenti principali saranno:

- **ODS (Operazionale Data Store):** è un database progettato per centralizzare i dati provenienti da fonti eterogenee al fine di facilitare le operazioni di analisi e di reporting. L'integrazione di questi dati comporta spesso l'eliminazione di informazioni ridondanti. Un ODS è tipicamente destinato a contenere dati di livello fine come il prezzo o l'importo delle vendite, in contrapposizione ai dati aggregati come l'importo totale delle vendite. I dati aggregati sono memorizzati in un datawarehouse.
- **Cloud datawarehouse:** Il datawarehouse del cloud sarà costituito da tabelle di fatti, dimensioni e data mart. I dati che popoleranno questo datawarehouse proverranno da diversi paesi e saranno immagazzinati nel datawarehouse per l'analisi e la produzione di KPI.

Uno degli elementi della strategia IT interna è il cloud come prima opzione nella scelta delle soluzioni, pertanto una soluzione datawarehouse cloud sarà preferita per questo e per altri motivi.

Consume

Questa parte riguarda le modalità di consumo delle informazioni memorizzate e rese disponibili dal data lake. Il reporting è l'elemento finale nella costruzione del futuro stato

dell'architettura della BI da costruire. Deve permettere agli utenti finali di visualizzare, analizzare, presentare e analizzare i dati.

3.6 Gap analysis

I paragrafi precedenti si sono concentrati sullo stato attuale e futuro dell'ambiente decisionale del WFS. La seguente tabella 3.7 fornirà una panoramica di questa situazione e identificherà le discrepanze. Seguirà un piano d'azione per la risoluzione di questa discrepanza.

Gli assi di analisi di questa gap saranno i moduli chiave che costituiscono il fondamento e la base di un ambiente decisionale. I suoi moduli saranno: fonti di dati, integrazione dei dati, archiviazione e analisi dei dati e infine il consumo di informazioni.

	AS IS	GAP	TO BE
Source systems	Une seule source de donnée existante PMP	Extraire les données provenant des sources system restant.	Donnée provenant des système opérationnels PMP
			Donnée provenant la solution CQC. Pour la gestion des relation clients
			Données provenant des solutions SAAS. Anaplan/Enablon
	Fichier envoyer par les sites sous format excell		Interface utilisateur
			Fichier plat / Excel
Ingest & Transform	Solution d'integration de données rudimentaire	Se doter d'une solution d'integration de donnée avec les fonctionnalités d'extration de données a travers les API et Webservices	Plateforme d'integration des données provenant des trois sources: Base de données, fichier plat,Extraction de données des solutions SAAS
Store & Analyse	Dase de donnée operationelle servant de base pour le reporting (PMP)	MDM	MDM: Gestion des données de reference
	Une seule table de fait dans le "datawarehouse" existant.		ODS: Gestion et historisation des données
		Data lake/Cloud datawarehouse	Staging area: Environnement pour le chargement et la transformation des doonnées
			Data Lake/Data warehouse/Data Mart
Consume	Extraction et realaboration des données pour une presentation adequate	Outil de reporting pour la presentation des KPI provenant du datawarehouse	Fournir 44 KPI pour les entités Cargo operation,Finance,HSSE, Customer relation,Human resources
	KPI uniquement sur les données operationelle		Presentation des données automatiser charger automatiquement avec la plateforme d'integration sur l'outilde reporting

Tableau 3.7:Tabella riassuntiva del Gap Analysis

GAP Analysis

Il progetto mira ad implementare una serie di KPI basati su un database decisionale. Abbiamo a disposizione i KPI forniti dall'azienda che servono come specifica

funzionale. Lo sviluppo di questi KPI richiede prima di tutto un inventario del sistema esistente e della sua capacità di fornire i KPI espressi come necessità, poi una proiezione della soluzione futura nel caso in cui il sistema attuale non sia in grado di fornire i KPI richiesti.

- Abbiamo individuato delle discrepanze nella soluzione esistente che rendono impossibile l'implementazione dei KPI richiesti.
- Dall'analisi delle fonti di dati risulta che la soluzione attuale gestisce oggi una sola fonte di dati, che rimane insufficiente per la generazione di KPI.
- Gli strumenti di trasformazione e integrazione dei dati sono inesistenti
- Datawarehouse esistente ridotto per l'utilizzo delle operazioni di cargo
- Difficili strumenti di visualizzazione sono stati inseriti nel datawarehouse. Ciò rende difficile il recupero dei KPI e costringe ad estrarre i dati per un'ulteriore manipolazione in un file Excel al fine di avere dati presentabili e qualche KPI utilizzabile ma insufficiente.

La conclusione è che per rimediare a questa discrepanza, l'intero sistema decisionale deve essere rivisto e aggiornato. Questo miglioramento o trasformazione del sistema di BI avrà un impatto sull'architettura applicativa esistente perché nuove soluzioni IT integreranno il paesaggio applicativo. Per ragioni di allineamento con la strategia IT del Gruppo, la scelta delle soluzioni sarà orientata alle applicazioni in hosting nel cloud.

D'altra parte, tutti i sistemi sorgente devono essere identificati, possono essere database, flat file o dati provenienti da piattaforme cloud attraverso API o servizi web, analizzare i formati dei dati e assicurarsi che questo formato sia supportato dalla soluzione di destinazione. Nell'ambito di WFS Global prendere in considerazione tutte le stazioni di cargo.

La soluzione di integrazione dei dati è un must in questa fase, in quanto sarà necessaria per i processi di estrazione, trasformazione e caricamento dei dati nel datawarehouse, ODS e cubi.

Il mattone MDM per la gestione del data repository è uno degli elementi importanti per allineare i dati di tutte le stazioni del mondo sotto lo stesso repository.

Il mattone del datawarehouse è il più importante, perché sarà necessario costruire un intero database decisionale che dovrà tenere conto di tutti i dati provenienti da entità diverse dall'operazione di carico che è l'unica operata oggi. Ciò richiederà la progettazione di un nuovo modello di dati, una tabella dei fatti e delle dimensioni... e un data mart.

Infine, lo strumento di reporting beneficerà dello strumento di integrazione dei dati per la presentazione dei KPI.

Conclusione

Dall'analisi ricavato dallo stato futuro si vede che siamo di fronte della realizzazione di uno strumento di business Intelligence basato su un datawarehouse. E l'obiettivo che emerge o la questione da risolvere è: Come implementare un datawarehouse per il supporto delle attività del business. Il prossimo capitolo tratterà del processo utilizzato per la scelta delle soluzioni che supporteranno l'architettura dello stato futuro discusso in questo capitolo.

Capitolo 4

Benchmark e soluzione proposta

In questo capitolo daremmo una proposizione di soluzioni per il problema discusso nel capitolo precedendone. Come implementare i KPI richiesto dal business. Il procedimento sarà il seguente:

4.1 Requisiti funzionali

I prerequisiti funzionali, come già discusso nel capitolo precedente, sono i quarantaquattro KPI da implementare nella nuova piattaforma di BI aziendale. Questi prerequisiti sono forniti dal comitato esecutivo e richiedono l'implementazione di KPI che forniscano visibilità sullo stato dell'azienda attraverso indicatori e su una piattaforma di BI condivisa dai principali attori del gruppo.

Le entità interessate da questo approccio saranno:

Cargo opération :

Entità	KPI
CARGO OPERATION	Kilos per hour
	AWB per Hour
	%Overtime
	%of Payroll, Third Party and overtime over Total Paid Hours
	Cost per paid Hour (Payroll+Overtime+Third Party)
	Manpower Cost per Paid Hour – WFS (Payroll + Overtime)
	Manpower Costs per paid Hour – Third Party
	%Occupacy warehouse

Tableau 4.1: KPI entità Cargo Operation

Main: Dashboard OPS Cargo YtY

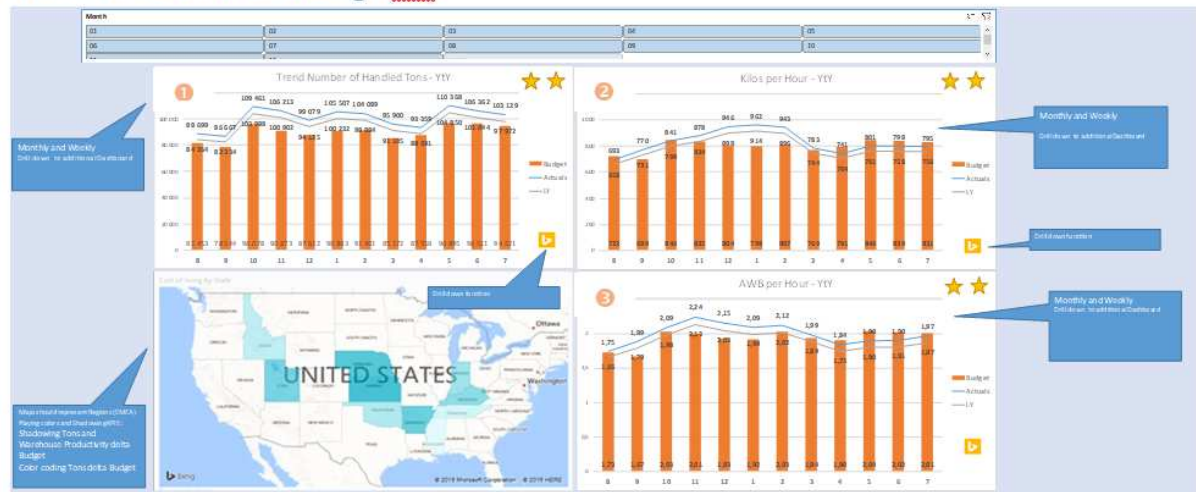


Figure 4.1: Esempio di un mockup del dashboard Cargo Operation

HR/People KPI

Entità	KPI
HR /Manpower	Staff turnover rate
	%Sickness over Total Payroll Paid Hours

Tableau 4.2: KPI entità risorse umana

Main: Dashboards Manpower Cargo



Figure 4.2: Esempio di un mockup della dashboard Manpower Cargo

Financial KPI

Entità	KPI
<i>FINANCE</i>	Yield (Revenue per kilo)
	Labour Costs per kilo
	Indirect cost per kilo
	Total cost per kilo
	Warehouse rental cost
	Contribution cost per kilo

Tableau 4.3:KPI entità Finance

Main: Dashboards Finance Cargo "YtY"



Figure 4.3:Esempio di un mockup della dashboard Finance

Questo è solo un esempio illustrativo dei KPI e del visual da produrre. Queste immagini sono state realizzate con lo strumento Excel e servono come mockup per la costruzione della reportistica finale. Descriveremo i KPI elencati in questo paragrafo in modo più dettagliato nei capitoli successivi.

4.2 Requisiti non funzionale tecnici

I prerequisiti non funzionali sono i prerequisiti nella maggior parte dei casi tecnici, che considerano tutti gli aspetti relativi all'architettura dei server, al dimensionamento, all'infrastruttura di rete e alla disponibilità di larghezza di banda, nonché alla sicurezza

sia a livello di applicazione che di rete. Nell'ambito dell'implementazione dello strumento decisionale del WFS.

La messa in atto di una nuova strategia sulla gestione dei sistemi informativi in WFS avrà un forte impatto sulla scelta dello strumento decisionale da implementare. Gli assi principali di questa strategia sono la migrazione dell'infrastruttura, comprese le applicazioni da WFS al cloud più precisamente AMAZON (AWS). Da quel momento in poi, la scelta dello strumento di BI è più precisa e orientata verso soluzioni cloud. Questo avrà altri vantaggi, come la scalabilità e l'elasticità, di cui parleremo più avanti.

I criteri da prendere in considerazione nella scelta della soluzione saranno:

- **La tecnologia** utilizzata per lo sviluppo della soluzione è una soluzione SAAS dove noi premettiamo. Verificare se la soluzione è multi-holder o istanza singola.
- **Integrazione** della soluzione nell'ecosistema applicativo esistente, assicurando che la soluzione sia in grado di integrarsi bene nell'ecosistema applicativo WFS.
- **La sicurezza** è anche un punto molto importante perché la scelta di soluzioni nel cloud impone misure di sicurezza rigorose, come i protocolli di sicurezza TSL v3 sono raccomandati.
- **La piattaforma di hosting** della soluzione è quello di prendere in considerazione la soluzione su AWS sono privilegiate nella selezione.
- **GDPR:** Si dovrebbe anche garantire che il fornitore della soluzione sia conforme agli standard di protezione dei dati GDPR.
- **Modello di supporto:** Permette di valutare la capacità dell'editor di accompagnarci quando la soluzione è in produzione. Ciò avverrà mediante un monitoraggio basato sui KPI contrattuali tra l'editore e WFS.

4.3 Benchmark delle soluzioni sul mercato

Tutte le analisi effettuate finora portano a privilegiare le soluzioni cloud per ogni strato dell'architettura della soluzione di BI.

L'architettura di destinazione si articolerà come segue:

- Soluzione di integrazione dei dati

- Storage: Azure blob / Amazon S3/Google platform
- Cloud datawarehouse
- Reportistica

Per ciascuno degli strati o livelli, saranno identificate sul mercato diverse soluzioni che saranno valutate secondo i criteri funzionali e non funzionali (tecnici) sopra descritti.

4.3.1 Benchmark delle soluzioni di Data Integration soluzioni

Definizione di data integrazioni

Una piattaforma di integrazione come servizio - iPaaS fornisce funzionalità per consentire l'integrazione di dati, applicazioni, API e processi che abbracciano soluzioni cloud e on-premises. Questo si ottiene sviluppando, implementando, gestendo e monitorando le interfacce di integrazione, ovvero le soluzioni di integrazione che collegano più endpoint in modo da poter lavorare insieme. iPaaS è tipicamente utilizzato per l'integrazione cloud, l'integrazione da applicazione ad applicazione, l'integrazione da applicazione ad applicazione, l'integrazione da business a business, l'integrazione di applicazioni mobili, la pubblicazione di API e gli scenari di integrazione Internet of Things. iPaaS fornisce varie tecnologie che consentono di costruire robuste soluzioni di integrazione. Le funzionalità di iPaaS possono essere utilizzate per costruire flussi di integrazione che collegano la combinazione di applicazioni cloud-based e on-premises, comprese quelle per i partner commerciali. Questi flussi di integrazione sono distribuiti e gestiti nel cloud, quindi gli utenti non hanno bisogno di procurarsi hardware e middleware di integrazione per eseguirli. Né devono preoccuparsi della distribuzione, dell'amministrazione, dell'aggiornamento e del monitoraggio della piattaforma di integrazione sottostante, in quanto si tratta di una piattaforma gestita.

Gardner Magic quadrant per l'enterprise Integration Platform

Il riferimento per analisi e confronto delle applicazioni sarà il magic quadrant di Gardner che adatteremmo al contesto aziendale per fare la scelta della soluzione più adatta [7].



Figure 4.4: Magic Quadrant per Enterprise Integration Platform as a Service

Informatica

Informatica, fondata nel 1993 e con sede a Redwood City, California, Stati Uniti, è una società privata e un noto fornitore di tecnologia per la gestione dei dati. È stata uno dei pionieri del mercato EiPaaS, che ha fatto il suo ingresso nel 2007. La più recente incarnazione della strategia EiPaaS della società è Informatica Intelligent Cloud Services (IICS), che è stata rilasciata all'inizio del 2018 e ha sostituito l'offerta della precedente generazione, Informatica Cloud.

IICS, rilasciato nel 2016, è la nuova generazione di Informatica, EiPaaS basata su microservizi. Fornisce un'ampia gamma di servizi, tra cui l'integrazione di applicazioni, dati e B2B, la gestione delle API, l'hub di integrazione digitale, la gestione dei dati master ("Customer 360"), la qualità dei dati e la governance. Le capacità IICS sono equivalenti, compatibili e interoperabili con quelle dei corrispondenti prodotti software di Informatica. L'IICS e le sue controparti software possono essere gestite, monitorate e

gestite con l'ausilio delle funzionalità di approfondimento operativo unificato della piattaforma, secondo un modello HIP.

Punti di forza

Comprensione del mercato e innovazione: Grazie alla sua lunga esperienza con le tecnologie EiPaaS, Informatica ha spesso anticipato le tendenze del mercato. Ne sono un esempio le tendenze per gli hub di integrazione dei dati, i cataloghi di dati, il supporto a più personaggi di integrazione, la governance, l'HIP, la convergenza dell'integrazione di applicazioni e dati, l'architettura dei microservizi e l'AI/ML per assistere gli sviluppatori di integrazione.

Presenza, crescita e soddisfazione del cliente: Informatica aveva stimato un fatturato di circa 300 milioni di dollari nel 2018, ed è il principale fornitore di iPaaS in termini di quota di mercato. Nonostante abbia oltre 11.500 clienti e debba affrontare una concorrenza sempre più intensa, Informatica ha tenuto il passo con la crescita del mercato nel 2018 senza compromettere l'alto livello di soddisfazione dei clienti per il loro rapporto con l'azienda e per le capacità dell'IICS.

Ricchezza di capacità ed evoluzione pianificata: IICS offre molte capacità e opzioni architetturali per un'ampia gamma di casi d'uso. I clienti di riferimento per Informatica apprezzano in particolare il ricco set di adattatori della piattaforma, l'integrazione dei dati, la sicurezza, la facilità d'uso tra i vari personaggi, la gestione dei metadati, il disaster recovery, la gestibilità, la conformità agli standard e il supporto tecnico. Informatica prevede di investire per rafforzare il proprio supporto alle implementazioni ibride e multicloud, la versatilità, le capacità aziendali e la facilità d'uso.

Avvertenza

Costo e complessità dei prezzi: Informatica ha recentemente introdotto la tariffazione per le organizzazioni di medie dimensioni. Ciononostante, i clienti di riferimento riferiscono che i prezzi IICS sono complessi. Nonostante l'introduzione dei prezzi basati sui casi d'uso da parte di Informatica nel 2018, la loro opinione è che l'IICS potrebbe essere più costoso delle offerte concorrenti.

La percezione del marchio: Sebbene Informatica sia un marchio noto, agli occhi di molti clienti rimane ancorata alle sue radici "dati". L'IICS è tipicamente considerata per l'integrazione dei dati e, in misura minore, per le questioni di integrazione delle applicazioni, anche nelle impostazioni da on-premise a on-premise. Sebbene l'IICS supporti il B2B, la pubblicazione di API e l'integrazione di applicazioni mobili, Informatica non ha educato il mercato su queste capacità. Pertanto, l'IICS non è considerato così spesso per questi casi di utilizzo popolare. Questo potrebbe limitare l'attrattiva di Informatica per le organizzazioni che sono alla ricerca di un EiPaaS strategico per affrontare una vasta gamma di casi d'uso.

Copertura funzionale: Nonostante abbia un'ampia gamma di funzionalità, l'IICS manca ancora di supporto per i casi d'uso emergenti come la RPA e l'elaborazione di eventi. I clienti di riferimento per Informatica considerano la gestione del ciclo di vita degli artefatti di integrazione come da migliorare. Sebbene IICS supporti diverse organizzazioni con scenari impegnativi, alcuni clienti di riferimento segnalano problemi di performance e scalabilità per casi d'uso specifici.

Dell Boomi

Dell Boomi, con sede a Chesterbrook, Pennsylvania, Stati Uniti, è una società interamente controllata da Dell Technologies. L'originaria società Boomi è stata costituita nel 2000, è entrata nel mercato iPaaS nel 2005 ed è stata acquisita da Dell nel 2010. L'azienda fornisce una suite estesa di offerte di tecnologia EiPaaS nella piattaforma Dell Boomi (AtomSphere). Queste includono l'integrazione di applicazioni e dati di base, la gestione delle API, la gestione B2B/EDI, il master data hub e l'automazione del flusso di lavoro a basso codice e lo sviluppo di applicazioni.

Ogni offerta di tecnologia disponibile sulla piattaforma Dell Boomi è disponibile in diverse edizioni, differenziate in base al mercato di destinazione, al caso d'uso, all'ampiezza delle funzionalità e al numero di endpoint collegati, alle transazioni API e ai record di dati. Tutte le edizioni includono il supporto standard, con servizi premium disponibili come opzioni. L'azienda fornisce anche un'ampia gamma di servizi gestiti e servizi professionali a complemento della sua offerta tecnologica di base.

Punti di forza

Comprensione del mercato: Come uno dei pionieri dello spazio iPaaS, Dell Boomi ha una profonda comprensione delle esigenze chiave del mercato. Ciò gli ha permesso di identificare tempestivamente le opportunità, come la possibilità di utilizzare i fornitori SaaS come canale, la convergenza dell'integrazione di applicazioni, dati, API e B2B e il supporto di più persona di integrazione e multicloud. Questa conoscenza del mercato sta ora informando le mosse strategiche di Dell Boomi verso l'IoT, il edge computing, la blockchain, lo sviluppo low-code e l'intelligenza artificiale (AI)/apprendimento della macchina (ML).

Esecuzione delle vendite e strategia: Nel 2018, Dell Boomi ha investito in modo significativo nel marketing, che ha migliorato notevolmente la sua visibilità e la sua posizione, tanto da essere ora percepita dai suoi clienti di riferimento come un partner innovativo e strategico. Anche il suo numero di clienti è cresciuto notevolmente, raggiungendo oltre 8.000 clienti, così come il suo fatturato, di oltre il 60%, pur mantenendo un basso tasso di abbandono di circa il 3%. La sua strategia di vendita consiste nel lavorare a più stretto contatto con la casa madre, nel concentrarsi sui settori, nel rafforzare il canale indiretto e la rete di partner e nell'espandersi in EMEA e in Asia/Pacifico.

Versatilità e innovazione della piattaforma: EiPaaS di Dell Boomi è potente e versatile - supporta casi d'uso come l'integrazione terra-terra, il flusso di lavoro, l'IoT e l'intermediazione di eventi. I clienti di riferimento premiano i suoi strumenti di sviluppo collaborativo multipersona, la gestione del ciclo di vita, la mappatura e le funzionalità EDI. La sua roadmap richiede una nuova esperienza utente personalizzata, un nuovo gateway API e un nuovo portale API, un uso esteso dell'AI/ML, il supporto per la blockchain e il Edge computing e un motore di esecuzione distribuito.

Avvertenza

Servizio e supporto: Alcuni clienti di riferimento esprimono preoccupazione per il servizio e l'assistenza tecnica. Ad esempio, i loro feedback per la tempestività e la completezza delle sue risposte ai problemi di prodotto sono meno buoni di quelli di

alcuni concorrenti. La soddisfazione complessiva dei clienti di Dell Boomi è inferiore a quella di altri concorrenti leader.

Funzionalità mancanti: Sebbene la piattaforma Dell Boomi disponga di ampie funzionalità, alcuni clienti di riferimento notano una mancanza di supporto per funzionalità quali MFT, integrazione con strumenti RPA, virtualizzazione dei dati/federazione e analisi dei flussi di eventi. La maggior parte dei suoi clienti di riferimento ritiene che le sue attuali funzionalità di gestione delle API non siano alla pari con le funzionalità fornite da altri fornitori EiPaaS.

Strategia geografica: Dell Boomi ha operazioni di vendita diretta in Nord America, Europa occidentale, Australia, Singapore, Hong Kong e Giappone, ma solo una presenza diretta limitata in altre regioni. Nei paesi in cui non ha una presenza diretta, le attività di vendita sono svolte da Dell Technologies o dai rivenditori. L'assistenza ai clienti è fornita dagli hub di supporto e dai partner integratori di sistemi Dell Boomi. Pertanto, Dell Boomi potrebbe non essere un fornitore interessante per le organizzazioni, soprattutto quelle di grandi dimensioni, che preferiscono che le attività di vendita e di supporto siano svolte direttamente dal fornitore EiPaaS.

Microsoft

Microsoft, fondata nel 1975 e con sede a Redmond, Washington, USA, è entrata nel mercato EiPaaS nel luglio 2016 con il lancio di Azure Logic Apps. Attualmente, la sua principale offerta EiPaaS è Azure Integration Services (AIS), che combina Azure Logic Apps con Azure API Management, Azure Service Bus (un servizio di messaggistica in coda e di pubblicazione degli abbonamenti) e Azure Event Grid (un servizio di ingestione di eventi di massa). Altri servizi relativi all'integrazione includono Azure Data Factory, uno strumento per l'integrazione dei dati e Microsoft Flow, uno strumento per l'integrazione dei cittadini costruito su Azure Logic Apps.

Nella valutazione di Microsoft, ci concentriamo su AIS. Consideriamo Azure Data Factory e Microsoft Flow come estensioni di questa offerta principale perché vengono vendute attraverso diversi canali.

Punti di forza

Adozione in rapida crescita: Nonostante sia sul mercato da meno di tre anni, Azure Logic Apps (il componente principale di AIS) è già una delle poche offerte EiPaaS con più di 10.000 clienti. Questo notevole successo deriva dallo slancio di mercato di Microsoft Azure, dalla fedeltà dei clienti Microsoft e dalla reputazione di Microsoft come fornitore IT solido, visionario e innovativo, di cui ci si può fidare come partner commerciale strategico.

Copertura globale: AIS è disponibile in tutto il mondo, con il supporto dell'organizzazione globale Microsoft e di circa 1.000 partner abilitati AIS. La maggior parte dei componenti AIS sono disponibili in 26 centri dati Azure (30 sono previsti entro la metà del 2019) distribuiti in tutte le principali regioni del mondo. Questo, insieme al supporto di Azure per una serie di standard internazionali, regionali e industriali, rende AIS una delle poche offerte EiPaaS veramente globali.

Visione tecnologica: AIS, Azure Data Factory, Microsoft Flow e la ricchezza dei servizi Azure consentono ai clienti di affrontare un'ampia gamma di casi d'uso mainstream. Gli investimenti pianificati per AIS includono l'integrazione con le piattaforme AI/ML di Microsoft, blockchain, IoT, edge computing, funzione come servizio e disponibilità su Azure Stack per le implementazioni on-premise (prevista nei prossimi 12-18 mesi). Questi miglioramenti estenderanno l'applicabilità di EiPaaS di Microsoft a diversi importanti scenari di integrazione emergenti.

Avvertenza

Maturità funzionale: Nonostante il continuo sforzo di Microsoft per migliorare l'offerta EiPaaS, alcuni clienti di riferimento ritengono che ci sia ancora spazio per migliorare le funzionalità dell'AIS. Tra le aree in cui individuano la necessità di una maggiore maturità ci sono la trasformazione e la mappatura dei dati, il supporto EDI, la gestione del ciclo di vita degli artefatti di integrazione, i test automatizzati e la scoperta dei metadati. Alcuni clienti di riferimento nutrono anche preoccupazioni circa la gestibilità, la scalabilità e il throughput di AIS.

Complessità e costo dei prezzi: Sebbene i prezzi di AIS siano altamente flessibili e trasparenti, i clienti di riferimento trovano anche abbastanza complessi da gestire perché richiedono una comprensione dettagliata dei loro scenari per capire quale sarà il costo. Microsoft sta lavorando ad un modello di pricing più semplice e prevedibile, ma i clienti di riferimento esprimono anche preoccupazioni sul costo effettivo del servizio e sulla mancanza di flessibilità di Microsoft durante le trattative.

Strategia di vendita e di marketing incentrata su Microsoft: Comprensibilmente, le vendite e il marketing di Microsoft sfruttano principalmente la popolarità della piattaforma cloud Azure e, in misura minore, il crescente successo delle sue applicazioni SaaS (Office 365 e Dynamics 365) per promuovere l'adozione dell'AIS. Pertanto, molti potenziali clienti percepiscono l'AIS come "solo per i membri" e non particolarmente attraente per le organizzazioni alla ricerca di un EiPaaS con capacità multicloud.

Oracle

Oracle è stata fondata nel 1977 e ha sede a Redwood Shores, California, USA. Offre un'ampia gamma di tecnologie PaaS che si sono evolute insieme alle sue applicazioni SaaS e all'infrastruttura sottostante che Oracle fornisce attraverso IaaS, PaaS, SaaS e tecnologie di integrazione di applicazioni e dati.

L'Oracle Integration Cloud si è evoluta a partire dall'offerta iniziale di Oracle Integration Cloud Service e Oracle Process Cloud Service nell'aprile 2015. Da allora, Oracle ha aggiunto Oracle SOA Cloud Service, Oracle Managed File Transfer Cloud Service, Oracle API Platform Cloud Service, Oracle Data Integration Platform Cloud e Oracle Self-Service Integration Cloud Service.

Punti di forza

Versatilità della piattaforma: Come ci si aspetterebbe da un'azienda con una lunga storia nel campo della tecnologia di integrazione, così come le offerte nei vicini settori tecnologici PaaS, IaaS e SaaS, Oracle fornisce una delle suite EiPaaS più coesive e versatili.

Presenza globale: Oracle ha una forte presenza globale. Anche se non è completamente focalizzata sulle tecnologie di integrazione, gli utenti finali beneficiano dell'ampia rete di Oracle, poiché significa che possono acquistare Oracle Integration Cloud a livello locale, direttamente o attraverso i canali SaaS e PaaS di Oracle, quasi ovunque nel mondo.

Prezzi: Con prezzi basati su volumi, consumi e "porta la tua licenza" per i clienti con investimenti Oracle esistenti, Oracle offre un'ampia gamma di modelli di prezzo. Le modalità di determinazione dei prezzi Oracle sono disponibili sul suo sito web per la consultazione diretta, insieme alle calcolatrici per le diverse valute e ai modelli di determinazione dei prezzi. Il packaging orientato al SaaS e al settore industriale si aggiunge ai modi in cui i clienti possono consumare i servizi EiPaaS di Oracle.

Avvertenza

Focus sul mercato: Sebbene il marchio Oracle sia ben noto, Oracle è ancora generalmente percepito nel mercato generale come un fornitore di costosi software di integrazione. La sua base clienti è costituita prevalentemente da grandi organizzazioni con investimenti già esistenti in prodotti Oracle. I clienti che cercano di integrare applicazioni non Oracle possono trovarsi a doversi affidare alla rete di partner Oracle, piuttosto che alla competenza diretta di Oracle.

L'esperienza del cliente: I punteggi dei clienti di riferimento di Oracle per quanto riguarda l'esperienza del cliente, il servizio e il supporto sono inferiori a quelli ricevuti dalla maggior parte dei suoi vicini concorrenti.

Innovazione: Oracle è più un "seguace veloce" che un "first mover" quando si tratta di aree di sviluppo di nuovi prodotti. Sebbene Oracle abbia investito nella sua tecnologia e abbia tutte le caratteristiche innovative che ci si aspetterebbe, le organizzazioni di utenti finali che cercano innovazioni "first to market" dovrebbero guardare altrove.

Criteri di selezione

I criteri di selezione si baseranno sulle forze e debolezze delle applicazioni, e certi prerequisiti funzionali e tecnici, identificati durante le riunioni con il mestiere e l'IT

Vendor	ETL Product	Strengths	Weaknesses
Informatica	Data Integration Platform	<ul style="list-style-type: none"> • Highly rated by analysts • Extensive product portfolio 	<ul style="list-style-type: none"> • Reputation for high prices • Overlapping products
Dell Boomi	Dell Boomi Platform	<ul style="list-style-type: none"> • Strong growth • Large customer base • Versatility 	<ul style="list-style-type: none"> • Poor user experience
Oracle	Data Integration Platform Cloud	<ul style="list-style-type: none"> • Autonomous capabilities • Zero downtime 	<ul style="list-style-type: none"> • Lack of documentation • New to the market
Microsoft	Azure Data Factory	<ul style="list-style-type: none"> • Low price • Familiar interface 	<ul style="list-style-type: none"> • New to the market

Tableau 4.4:Forze e debolezze delle iPaaS

Scoring

Enterprise iPassS Rating Comparison				
Feature	Dell Boomi	Informatica Intelligent Cloud	Oracle Integration Cloud	Microsoft
Prebuilt Connectors	9,0	8,2	8,6	7,4
Connector modification	8,3	7,8	7,8	7,2
Support real-time and batch integration	7,0	8,5	8,0	7,1
Data quality services	8,6	7,7	6,8	7,2
Data security features	7,6	7,8	8,2	6,7
Monitoring Console	6,4	8,0	8,4	6,4
Somma	46,9	48,0	47,8	42,0

Tableau 4.5: Scoring delle soluzioni ETL

Soluzione ritenuta e caratteristiche e motivazione

Informatica e la soluzione ritenuta come soluzione di data integrazioni

Vantaggi della piattaforma iPaaS Informatica Intelligent Cloud

- Soluzione a basso rischio, poiché iPaaS offre funzionalità pronte all'uso
- I componenti della soluzione sono altamente scalabili e configurabili
- Efficiente in termini di costi per lo sviluppo, la gestione e l'esecuzione di soluzioni software in quanto si tratta di un modello a pagamento
- iPaaS può aiutarti nel tuo viaggio verso la riduzione dei rischi e il raggiungimento della conformità al GDPR attraverso strumenti e processi.
- Agile e reattivo per soddisfare le richieste del business, con un'implementazione accelerata delle applicazioni e una continua agilità di progettazione
- Conforme ai protocolli di sicurezza comuni

- Permette e sfrutta l'architettura di integrazione, il quadro di governance e gestione, le migliori pratiche e gli standard.
- Quadro di governance che contribuirà a guidare il futuro sviluppo del business e il processo decisionale aziendale
- Sfruttamento delle soluzioni Cloud & On-premise che facilitano l'integrazione ibrida
- Costruire visivamente flussi di lavoro con strumenti facili da usare
- Iniziare più velocemente con i modelli di piattaforma IPaaS, i flussi di lavoro e i connettori out of box
- Collegare sistemi disparati in ambienti eterogenei
- Fornisce un eccellente supporto per l'integrazione aziendale e gli scenari B2B
- Scrivere una volta, riutilizzare spesso, il che significa riutilizzabilità ed estensibilità incorporate nella piattaforma

4.3.2 Benchmark delle soluzioni per lo stoccaggio dei dati

Con l'avvento del cloud è possibile archiviare i vostri dati nel cloud pubblico o privato. Ciò offre vantaggi come la rapida implementazione dell'infrastruttura di storage, il ridimensionamento dello storage se necessario. In questa parte analizzeremo due soluzioni leader nel mercato del cloud data storage: Azure blob / Amazon S3

Amazon S3 vs Azure Storage: Differenza tra queste due soluzioni

Amazon S3 servizio per l'archiviazione dei dati è descritto come "l'archiviazione e il recupero di qualsiasi quantità di dati, in qualsiasi momento, da qualsiasi punto del web". Il servizio Amazon Simple Storage rende disponibile un'infrastruttura dati in ogni punto del web in ogni tempo.

D'altra parte, **Azure Storage** è descritto in dettaglio come "Archiviazione cloud affidabile ed economica per dati grandi e piccoli". Azure Storage offre la flessibilità di memorizzare e recuperare grandi quantità di dati non strutturati, come documenti e file multimediali con Azure Blobs; dati strutturati basati su nosql con Azure Tables; messaggi affidabili con Azure Queues, e utilizza Azure Files basati su SMB per la

migrazione delle applicazioni on-premises verso il cloud. Amazon S3 e Azure Storage possono essere classificati come strumenti "Cloud Storage".

Alcune delle caratteristiche offerte da Amazon S3 sono:

- Scrivere, leggere e cancellare oggetti contenenti da 1 byte a 5 terabyte di dati ciascuno. Il numero di oggetti che si possono memorizzare è illimitato.
- Ogni oggetto viene memorizzato in un secchio e recuperato tramite una chiave unica, assegnata dallo sviluppatore.
- Un secchio può essere memorizzato in una delle diverse Regioni. È possibile scegliere una Regione per ottimizzare la latenza, ridurre al minimo i costi o soddisfare i requisiti normativi. Amazon S3 è attualmente disponibile nelle Regioni US Standard, US West (Oregon), US West (California del Nord), UE (Irlanda), Asia Pacifico (Singapore), Asia Pacifico (Tokyo), Asia Pacifico (Sydney), Sud America (San Paolo) e GovCloud (USA). La Regione Standard degli Stati Uniti instrada automaticamente le richieste alle strutture della Virginia settentrionale o del Pacifico nord-occidentale utilizzando le mappe di rete.

D'altra parte, Azure Storage fornisce le seguenti caratteristiche chiave:

- Blob, tabelle, code e file
- Altamente scalabile
- Durevole e altamente disponibile

"Affidabile" è il motivo principale per cui gli sviluppatori considerano Amazon S3 rispetto alla concorrenza, mentre "soluzione di storage all-in-one" è stato dichiarato come il fattore chiave per la scelta di Azure Storage.

AWS Vs Azure Vs Google: Storage

La memorizzazione è una delle funzioni chiave dei servizi cloud; i servizi offerti dal dominio di memorizzazione sono legati alla memorizzazione dei dati. AWS fornisce servizi di storage di lunga durata, mentre i servizi di storage forniti da Microsoft Azure e Google Cloud Platform sono anche opzioni affidabili e rispettabili.

Storage Services	AWS	Azure
Object Storage Service for Use Cases	Simple Storage Services(S3)	Storage (Block Blob)
Archive Storage	S3 Infrequent Access Glacier Data Archive	Storage (Cool) Storage (Archive)
Hybrid Storage	Storage Gateway	StorSimple
Automatic Protection and Disaster Recovery	Disaster Recovery	Site Recovery
Bulk Data Transfer Solutions	Import/Export Disk Snowball Edge SnowMobile	Import/Export Azure Data Box
Backup	Object Storage Cold Archive storage Gateway	Backup

Tableau 4.6: Confronto tra funzionalità di Azure e Amazon per lo storage

Criteri di selezione

Feature	Azure Blob	S3	WFS Criterias (M/S/C/W)	Scoring : Azure Blob	Scoring : Amazon S3
Protocol	HTTP(REST)	HTTP(REST/SOAP)	M	1	1
Tolls for large files	NO	Export/Import	M	0	1
Archiving Storage	NO	YES	M	0	1
Users Accessing/Storing data paid by customer	NO	YES	S	0	1
Server Side Encryption	NO	YES	C	0	1
Container limit	Unlimited	100	M	0	1
CDN Endoint	Unique endpointper account	Unique endpoint per container	M	1	1
Reduced Redundancy Storage Option	NO	YES	M	0	1
Private Content Support	YES	YES	M	1	1
Virtual Hosted style	NO	YES	M	0	1
Path Style	YES	YES	M	1	1
Auto delete	NO	YES	M	0	1
Deleting container without deleting entries	YES	NO	M	1	0
Mounting VHD from blob storage	YES	NO	M	1	0
Hosting from Static or dynamic content	NO	YES	M	0	1
Logging	Storage Analytics	S3 Logging	C	1	1
Valid Key name characters	Alphanumeric and "/"	Alphanumeric, "/" and "."	W	1	1
Storage Emulator	YES	NO	M	1	1
Maximun Blob Size	1TB	5TB	M	1	1
Maximum size per account	100TB	Unlimited	M	1	1
Maximum Containers	Unlimited	100	W	1	1
Scoring				12	19

Tableau 4.7:Esito della selezione tra le soluzioni Amazon e Azure

Soluzione ritenuta

In base allo scoring fornito confrontano le funzionalità la scelta sarà Amazon S3

4.3.3 Concetto del Cloud datawarehouse e Benchmark delle soluzioni cloud datawarehouse

Concetto del cloud datawarehouse

Il concetto e la soluzione di datawarehouse in-the-cloud è piuttosto nuovo e si propone come alternativa per le soluzioni di datawarehouse on-premise. In questa parte presenteremo i vantaggi del cloud datawarehouse per illustrare la nostra scelta di questa tecnologia e quindi il punto di riferimento delle soluzioni di cloud datawarehouse sul mercato.

I seguenti vantaggi si trovano nella maggior parte delle soluzioni analitiche basate sul cloud:

- **Scala infinita:** A differenza delle soluzioni on-premises based che sono limitate dallo spazio fisico del rack del data center, una soluzione cloud-based può (in teoria), scalare ad una dimensione quasi infinita. In realtà, poiché ogni soluzione ha dei limiti pratici, questi saranno indicati dove applicabile.
- **Punto di ingresso basso:** Poiché le implementazioni basate sul cloud sono pagate in abbonamento o a pagamento, non ci sono grandi spese in conto capitale da finanziare, un vantaggio significativo per le piccole e medie imprese.
- **Sempre disponibile:** Ogni soluzione esaminata include un'alta disponibilità integrata. Ciò si confronta bene con le soluzioni on-premise, che tipicamente necessitano di un centro dati di disaster recovery off-site con le corrispondenti spese in conto capitale aggiuntive.
- **Controllo dei costi:** Il modello pay-as-you-go supporta la capacità di controllare i costi su base continuativa ed evita una significativa spesa in conto capitale.
- **Time to Market:** La maggior parte dei database basati su cloud può essere implementata in pochi minuti su un nuovo hardware, il che velocizza la fornitura di nuove soluzioni.
- **Agilità:** Molte soluzioni basate sul cloud possono essere rapidamente implementate, scalate o temporaneamente sospese, il che le rende un'eccellente soluzione per la prototipazione, la prova di concetto e l'analisi di scoperta. - Costo ridotto: La

possibilità di implementare o rimuovere rapidamente un database aiuta a ridurre i costi, evitando database di test di accettazione degli utenti e di prestazioni allocati in modo permanente.

Valutazione

Il database relazionale è stato creato per quasi 40 anni ed è tipicamente valutato in base alle prestazioni e all'usabilità. Tuttavia, le recenti innovazioni sia nell'architettura del database che nelle implementazioni basate sul cloud, aprono una serie completamente nuova di criteri in base ai quali giudicare una soluzione. Questi includono:

- **Semplicità e riduzione delle competenze richieste:** Le moderne soluzioni basate sul cloud sono spesso notevolmente facili da implementare e gestire, soprattutto se paragonate alla tradizionale soluzione on-premises.
- **Elasticità on the Fly:** Che include la capacità di aumentare o diminuire le risorse di calcolo per soddisfare le richieste. Alcune soluzioni sono più veloci da rispondere e possono essere scalate automaticamente.
- **Concorrenza:** Capacità di segmentare il carico del magazzino: la capacità di evitare il tiro alla fune tra l'elaborazione batch ETL e le richieste dell'utente finale a bassa latenza. Questo supporta un alto livello di concorrenza. - Calcolo e stoccaggio in scala indipendente: Alcune soluzioni basate su cloud possono espandere lo storage indipendentemente dalle risorse di calcolo. Ciò fornisce un maggiore grado di flessibilità nella distribuzione e significa che si paga solo per le risorse di cui si ha bisogno.
- **Gestione di dati diversi:** Compresa la possibilità di caricare e interrogare dati JSON semi-strutturati spesso presenti nelle applicazioni basate sul web.
- **Archiviazione avanzata basata su colonne:** Che ha un impatto enorme sulle prestazioni delle query analitiche, massimizzando al contempo la compressione dei dati.
- **Forte supporto SQL:** Che include funzioni di finestra analitica e funzioni definite dall'utente (UDF). Il resto di questo rapporto descriverà questi criteri in dettaglio ed esaminerà come gli attuali leader di mercato si confrontano.

Semplicità e ridotta abilità richiesta

Dieci anni fa il Dr. Michael Stonebraker ha pubblicato i documenti di ricerca *One Size [No Longer] Fits All* e *The End of an Architectural Era* in cui sosteneva che le soluzioni di database ereditate da Oracle, Microsoft e IBM non erano più adatte allo scopo.

Costruite su un'architettura OLTP basata su una serie di file ideata quasi 40 anni fa, da allora sono state ampliate per includere un'ampia gamma di funzioni per l'archiviazione dei dati. Egli ha sostenuto che queste soluzioni troppo complesse e generiche sarebbero state sostituite da sistemi dedicati, ciascuno progettato per un unico scopo.

Come esempio della complessità, Oracle11g include una vasta gamma di opzioni di tuning delle prestazioni aggiunte nel tempo, tra cui quasi 500 parametri di tuning individuali e 16 tipi di indici.

Il passaggio al cloud apre una vera e propria opportunità per semplificare o addirittura eliminare l'onere amministrativo del sistema su questi sistemi legacy. Non solo la manutenzione dell'hardware e del sistema operativo, ma anche il database stesso.

A tal fine, alcune soluzioni analitiche sono state riprogettate da zero e offrono un'architettura innovativa che elimina quasi completamente la necessità di risorse tecniche altamente specializzate. Ciò significa che gli ingegneri e gli architetti di database di valore sono liberi dalle esigenze dell'amministrazione del database per concentrarsi sulla vera sfida, fornendo le intuizioni dei clienti.

Alcuni esempi sono:

- **Index Design:** Per massimizzare le prestazioni delle ricerche. Questo può essere un settore impegnativo in quanto il progettista deve bilanciare i requisiti prestazionali spesso conflittuali delle query di lettura con la necessità di caricare rapidamente grandi volumi di dati. Questo porta spesso a disabilitare gli indici durante i grandi carichi di batch, solo per trovare che le prestazioni delle query online ne risentano.
- **Acquisizione dati statistici:** Per supportare un ottimizzatore basato sui costi, la maggior parte dei database include strumenti per analizzare e catturare le statistiche dei metadati per massimizzare le prestazioni delle query. Tuttavia, con l'inserimento

di voci aggiuntive, queste statistiche possono diventare obsolete, il che porta a una strategia di gestione della casa per aggiornare regolarmente le statistiche che possono essere un compito impegnativo per la CPU. 8 Sonra Intelligence Limited Un confronto tra le piattaforme di cloud datawarehouse

- **Partizionamento orizzontale e replica:** Il che comporta la specificazione di un metodo ragionevole di distribuzione dei dati per frammentare (diffondere), i dati su più server. Anche se ancora utilizzato da alcuni sistemi Massively Parallel Computing (MPP) basati su cloud, questo è stato in alcuni casi sostituito da un'architettura completamente nuova e innovativa che elimina questo sforzo di progettazione aggiuntiva. Molti di questi compiti di progettazione sono eliminati dalle soluzioni basate sul cloud. Alcune offrono anche funzionalità aggiuntive, tra cui:
- **Clonazione a copia zero:** Una tecnica usata per replicare rapidamente un database per costruire un ambiente di test completamente popolato. Questo funziona clonando il database senza alcuna copia fisica dei dati. Questo può alleviare l'onere dei DEVOP, poiché i terabyte di dati possono essere clonati in pochi secondi con successivi inserti e aggiornamenti consentiti sul nuovo set di dati.
- **Condivisione dei dati:** Che fornisce l'accesso sia alle risorse di calcolo che a quelle di dati a partner esterni o a società affiliate in sola lettura. In questo modo si evita la necessità di costruire più pipeline ETL (Extract Transform and Load) per utenti esterni ed evita la necessità di routine di Change Data Capture (CDC) quando i dati del magazzino vengono aggiornati, in quanto gli utenti visualizzano sempre i dati più recenti.

Al volo Elasticità/Elasticità a richiesta

Le soluzioni on-premise forniscono una piattaforma hardware fissa con poche opzioni di elasticità - la possibilità di aggiungere o rimuovere la potenza di calcolo e lo storage a seconda delle necessità. Una soluzione ideale basata su cloud fornirebbe una piattaforma hardware flessibile che cresce e si riduce man mano che il carico di lavoro e il numero di utenti cambia nel tempo.

Tuttavia, non tutte le soluzioni basate su cloud hanno le stesse caratteristiche che dovrebbero includere:

- Risposta immediata: La capacità di allocare e rimuovere rapidamente le risorse di calcolo in base alle necessità senza alcuna interruzione del servizio o impatto sulle prestazioni
- Scala su o giù: la possibilità di aggiungere o rimuovere automaticamente le risorse hardware al variare della domanda
- Piccoli aumenti: La possibilità di aggiungere risorse aggiuntive in incrementi relativamente piccoli invece di enormi cambiamenti di passo
- Scala Indipendente: La capacità di aggiungere (o rimuovere) l'archiviazione e di calcolare le risorse in modo indipendente per far corrispondere le risorse alle richieste attuali.

Valuta Concorde: Capacità di segmentare il carico del magazzino

I sistemi on premise tendono ad essere di dimensioni fisse con risorse di macchina condivise da gruppi di utenti diversi, spesso con requisiti di elaborazione molto diversi tra loro.

I processi ETL in batch, ad esempio, tendono ad includere operazioni di inserimento o di aggiornamento di massa di lunga durata e hanno un profilo di calcolo molto diverso da quello degli utenti del cruscotto che richiedono un tempo di risposta rapido per le query analitiche online.

Questa caratteristica si riferisce alla capacità di massimizzare la concomitanza attraverso l'elaborazione di partizionamento o segmentazione per evitare il tiro alla fune per le risorse della macchina. Idealmente, la soluzione dovrebbe consentire a diversi gruppi di utenti o carichi di lavoro di funzionare indipendentemente e in modo isolato, condividendo gli stessi dati.

Alcune soluzioni basate sul cloud, tuttavia, adottano l'approccio tradizionale di mettere gli utenti in coda alle risorse. Tuttavia, questo porta ad un uso meno efficiente delle

risorse delle macchine, poiché la capacità viene condivisa tra più gruppi di utenti concorrenti.

Calcolo e memorizzazione in scala indipendente

Le architetture hardware standard dagli anni '80 in poi possono essere categorizzate in base alla loro capacità di scalare. Queste includono:

Memoria condivisa: Descrive un sistema in cui tutto (compresa la memoria), è condiviso. L'opzione disponibile per scalare la soluzione è l'acquisto e la migrazione su una piattaforma più grande, e queste soluzioni necessitano di una macchina standby potenzialmente costosa per supportare un'elevata disponibilità.

Disco condiviso: Descrive un sistema in cui le risorse di calcolo sono strettamente collegate in un cluster, ma condividono una singola risorsa di storage, tipicamente implementata utilizzando una Storage Area Network (SAN).

Shared Nothing: Conosciuto anche come Massively Parallel Processing (MPP), distribuisce sia l'elaborazione che i dati su un certo numero di macchine. I dati vengono allocati a ciascun nodo utilizzando un metodo round-robin o un metodo di hashing coerente, e possono essere aggiunti altri nodi nel tempo. Sebbene Teradata e Netezza abbiano prodotto con successo appliance di data warehouse su questa architettura, essa presenta degli inconvenienti in quanto l'elaborazione e lo storage sono legati, e non è possibile scalarli in modo indipendente.

Recenti ricerche su database hanno prodotto alcune architetture innovative che sfidano la saggezza percepita degli ultimi 40 anni. In particolare, fornendo la possibilità di scalare le risorse di elaborazione e di memorizzazione in modo indipendente.

Questo porta ai seguenti vantaggi

Scalabilità flessibile: È importante essere in grado di aggiungere in modo flessibile le risorse di storage e di calcolo man mano che le richieste cambiano nel tempo. Ciò consente di scalare da gigabyte a terabyte e petabyte mentre si regolano in modo indipendente le risorse di calcolo. Ciò significa che è possibile supportare query

complesse, a bassa latenza e ad alta intensità di risorse su sistemi di dimensioni di terabyte, oppure operazioni di elaborazione in batch massicce su storage di dimensioni di petabyte sulla stessa soluzione.

Pagate solo ciò di cui avete bisogno: Una soluzione ideale vi consentirebbe di scalare lo storage indipendentemente dalle risorse di calcolo. Ad esempio, i sistemi MPP sono in teoria bilanciati in quanto le macchine aggiuntive aggiungono sia lo storage che le risorse di calcolo. Tuttavia, poiché questi sono così strettamente legati, se le richieste di storage superano la necessità di capacità di elaborazione, il costo per terabyte può aumentare in modo sproporzionato.

Separare lo storage dalle risorse di calcolo significa che si può scalare in modo indipendente, il che significa che si paga solo lo storage e le risorse di calcolo appropriate di cui si ha effettivamente bisogno.

Gestione dei dati diversificati

Uno dei vantaggi offerti da Hadoop HDFS è la capacità di gestire dati diversificati, tra cui dati strutturati provenienti da fonti di database, e dati semi-strutturati, tra cui JSON e XML. Tuttavia, è sempre più possibile analizzare, memorizzare ed elaborare JSON e XML direttamente nel magazzino, e questo è particolarmente importante quando si analizzano i feed dei social media e i sensori delle macchine o i dati trasferiti dalle applicazioni web. Alcune soluzioni basate sul cloud includono estensioni SQL per supportare l'unione di dati JSON semi-strutturati a dati strutturati esistenti. Idealmente una soluzione di magazzino moderna dovrebbe catturare, memorizzare e interrogare i dati contro dati semi-strutturati senza la necessità di analizzare e tradurre i dati in una struttura relazionale predefinita. Dato l'aumento della domanda di una soluzione Data Lake, ci sono anche vantaggi nella capacità di memorizzare i dati in (ancora) formato dati indeterminato per supportare lo schema in lettura.

Conservazione avanzata delle colonne

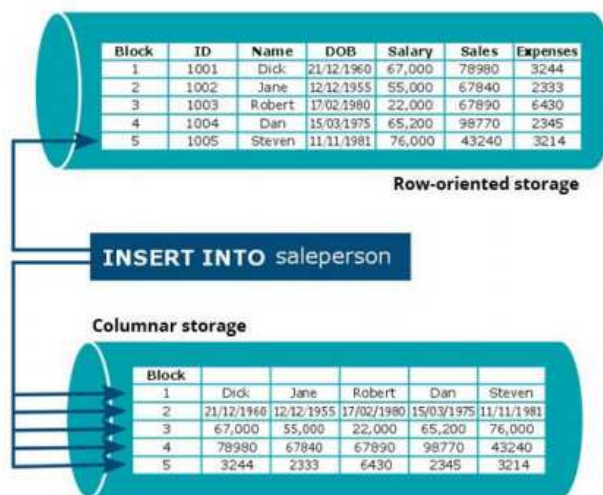


Figure 4.5:Tipologie di stoccaggio

Conservazione avanzata delle colonne

I database tradizionali di Oracle, Microsoft e IBM estendono una soluzione esistente basata su file progettata per applicazioni OLTP con funzionalità aggiuntive per un datawarehouse. Da allora, tuttavia, un'ampia ricerca ha dimostrato che i database basati su colonne sono fino a 100 volte più veloci per le query analitiche e la maggior parte (ma non tutte) le moderne soluzioni di datawarehouse sono costruite su un datawarehouse a colonne. Il diagramma qui sopra illustra la differenza tra le due soluzioni. L'archiviazione basata sulle righe di cui sopra funziona bene per i sistemi OLTP che tipicamente aggiornano tutte le colonne su una singola riga. Tuttavia, le query analitiche tendono a riassumere o raggruppare per un numero limitato di colonne su milioni di righe. Nell'esempio di cui sopra, una somma di SPESE per NOME interrogherebbe solo due colonne, e su un'ampia tabella di fatto, i guadagni di performance possono essere significativi. Memorizzare i dati in formato colonnare significa anche che ci sono maggiori possibilità di comprimere i dati eliminando i valori ripetuti, il che può portare a notevoli risparmi sui costi.

4.3.4 Benchmark delle soluzioni cloud datawarehouse



Figure 4.6: Magic Quadrant per il cloud datawarehouse

Snowflake

Snowflake, che ha sede a San Mateo, California, USA, offre un datawarehouse “as service” sulle infrastrutture AWS e Microsoft Azure. Supporta l'elaborazione relazionale ACID-compliant, oltre al supporto nativo per i formati di archiviazione dei documenti come JSON, Avro, Optimized Row Columnar (ORC), Parquet e XML. Un connettore Apache Spark nativo, l'integrazione R, il supporto per funzioni definite dall'utente, l'elasticità dinamica, il supporto temporale e le capacità di condivisione dei dati completano il nucleo dell'offerta. Le partnership recentemente annunciate con Qubole e Databricks estendono la portata di Snowflake ai casi d'uso esplorativi di data lake.

Punti di forza

- **L'esperienza del cliente:** I punteggi di riferimento dei clienti di Snowflake per l'esperienza complessiva lo collocano ai vertici dei venditori in questo Magic

Quadrant. Gli utenti hanno elogiato le capacità del prodotto, il rapporto qualità-prezzo, la flessibilità dei prezzi e dei contratti e la facilità di implementazione complessiva.

- **Capacità di elasticità dinamica:** La moderna architettura cloud di Snowflake è costruita intorno alla separazione delle risorse, che consente di scalare in modo efficace ed efficiente in risposta a carichi di lavoro misti e dinamici.
- **Efficienza operativa:** Le referenze dei clienti per Snowflake riferiscono di essere in grado di eseguire più carichi di lavoro e di supportare più casi d'uso di quanto non fossero in grado di fare con la loro precedente piattaforma, spesso con meno risorse.

Avvertenze

- **Mancano alcune caratteristiche:** Snowflake è generalmente disponibile da meno di quattro anni. Alcune referenze dei clienti hanno evidenziato la mancanza di caratteristiche come le viste materializzate e le procedure memorizzate, e un'interfaccia utente amministrativa basata sul web relativamente immaturo. Tuttavia, il fornitore sta lavorando attivamente per colmare queste lacune. Le visualizzazioni materializzate sono ora generalmente disponibili e le stored procedure sono attualmente in anteprima, anche se non ancora disponibili.
- **Ipercrescita:** Snowflake è nel bel mezzo di un periodo di ipercrescita e da quando è diventato generalmente disponibile nel 2015, la sua base di clienti è cresciuta fino a superare le 1.000 organizzazioni. Potrebbe essere difficile per l'organizzazione continuare a fornire lo stesso livello di coinvolgimento dei clienti che i suoi clienti si aspettano.
- **Risorse di formazione e supporto:** I punteggi di riferimento dei clienti di Snowflake lo collocano nella metà inferiore dei fornitori di questo Magic Quadrant per la qualità complessiva delle risorse di formazione e dell'esperienza di supporto. Queste erano le uniche due categorie in cui il punteggio di Snowflake non lo collocava nella metà superiore dei fornitori valutati, e anche per questi, i punteggi erano all'interno della deviazione standard rispetto alla media.

Amazon Web Services (Amazon Redshift)

Amazon Web Services (AWS) è una società interamente controllata da Amazon, con sede a Seattle, Washington, USA. AWS offre Amazon Redshift, un servizio di datawarehouse in the cloud. Amazon Redshift include Redshift Spectrum, un motore di interrogazione senza server e con misurazione che utilizza lo stesso ottimizzatore di Amazon Redshift, ma interroga i dati sia in Amazon Simple Storage Service (S3) che nello storage locale di Redshift. Offre anche Amazon S3, un negozio di oggetti cloud; AWS Lake Formation, un servizio sicuro di data lake; AWS Glue, un servizio di integrazione dati e catalogo metadati; e Amazon Elasticsearch, un motore di ricerca basato sulla libreria Lucene. Altre offerte includono Amazon Kinesis, un servizio di analisi dei dati in streaming, Amazon Elastic MapReduce (EMR), un servizio Hadoop gestito; Amazon Athena, un motore di ricerca senza server e metrica per i dati che risiedono in Amazon S3; e Amazon QuickSight, uno strumento di visualizzazione di business intelligence (BI). Amazon Neptune fornisce funzionalità grafiche.

Punti di forza

- **Cloud colossus:** AWS è generalmente riconosciuto come il principale fornitore di servizi cloud con un margine significativo. Questo fornisce una grande flotta di risorse che possono servire come base per l'apprendimento automatico (ML) e i dati di ottimizzazione, che possono poi essere applicati alle varie offerte DMSA. Ciò significa anche che ci sarà un pool pronto di risorse di terze parti disponibili nel mercato tecnologico e del lavoro.
- **Forza finanziaria e stabilità:** La crescita di AWS ha rappresentato oltre il 40% della crescita dell'intero mercato dei DBMS nel 2017, il che non solo gli fornisce la forza finanziaria, ma amplia la base di competenze per i suoi servizi.
- **Ampliamento della portata:** AWS ha iniziato ad offrire la possibilità di mettere insieme diversi servizi con offerte più recenti come AWS Glue, un servizio di catalogazione dei metadati e di integrazione dei dati, e Amazon QuickSight, uno strumento di visualizzazione della BI. AWS ha anche iniziato una spinta verso una presenza on-premises, rendendo disponibili su VMware alcuni dei suoi Relational

Database Services. Ha anche annunciato gli Outposts di AWS, nel novembre 2018, che includeranno Amazon EMR quando verrà lanciato nel 2019, o poco dopo.

Avvertenze

- **Complessità dell'integrazione:** AWS ha costantemente adottato un approccio ingegneristico "best fit", offrendo diversi servizi per affrontare i diversi casi d'uso. Questo approccio può portare a problemi di integrazione sempre maggiori man mano che le organizzazioni cercano di utilizzare dati e offerte che si estendono su più casi d'uso. La maggiore adozione degli archivi di oggetti cloud, come livello unificato di persistenza dei dati, e gli strumenti di metadati come Glue possono aiutare a mitigare le sfide di gestione.
- **Rapporto qualità/prezzo, prezzi e flessibilità contrattuale:** Le referenze dei clienti per AWS hanno generalmente dato punteggi elevati per la valutazione e la negoziazione dei contratti. Sorprendentemente, tuttavia, lo stesso non si è verificato quest'anno per i punteggi relativi al rapporto qualità-prezzo, ai prezzi e alla flessibilità dei contratti. AWS è stata collocata nel quartile inferiore dei venditori in questo Magic Quadrant per il rapporto qualità/prezzo percepito, e nella metà inferiore per la flessibilità dei prezzi e dei contratti. Questo potrebbe riflettere una minore consapevolezza dei miglioramenti delle prestazioni del Redshift e delle nuove funzionalità di ottimizzazione dei costi, come il Redshift Concurrency Scaling di AWS, che sono disponibili solo di recente.
- **Capacità del prodotto:** Sebbene tutti i clienti di riferimento per AWS abbiano ottenuto un punteggio medio o migliore per le capacità del prodotto, AWS ha avuto uno dei punteggi medi più bassi per le sue capacità di prodotto. Inoltre, un numero relativamente elevato di clienti di riferimento ha segnalato problemi di performance tra i problemi incontrati. Questo probabilmente riflette il fatto che Redshift è stato relativamente lento ad adottare alcune caratteristiche chiave che sono attese dai moderni ambienti DMSA cloud - come l'elasticità dinamica, la sintonizzazione automatica e la separazione delle risorse di calcolo e di memorizzazione. Il fornitore sta aggiungendo in modo aggressivo nuove funzionalità e capacità al portafoglio, per cui ci aspettiamo un miglioramento nel corso del prossimo anno.

Microsoft Azure SQL Data Warehouse

Microsoft, che ha sede a Redmond, Washington, USA, offre SQL Server come soluzione solo software con configurazioni certificate. Vende anche Azure SQL Datawarehouse (completamente gestito, MPP cloud datawarehouse), Azure HDInsight (distribuzione Hadoop basata su Hortonworks), Azure Databricks (piattaforma di analisi basata su Apache Spark) e Azure Data Lake (grande archivio dati e piattaforma di analisi) come servizi cloud.

Punti di forza

- **Capacità e visione LDW:** Con Azure Data Lake Storage, Azure Data Lake Analytics, Azure SQL Datawarehouse e Azure Databricks, Microsoft articola una forte visione per gli ambienti di elaborazione distribuiti inerenti all'LDW. Fornisce un approccio ben articolato per un percorso di produzione - dalla scoperta ed esplorazione nel data lake all'ottimizzazione nel datawarehouse.
- **Esecuzione del mercato e focus sul cloud:** Microsoft è cresciuta al doppio del tasso di crescita del mercato complessivo dei DBMS nel 2017. Se da un lato Gartner non abbatte le differenze tra gli usi operativi e analitici delle soluzioni, dall'altro la forte esecuzione di mercato di Microsoft si applica anche al DMSA. Microsoft continua a beneficiare del suo forte e precoce cloud focus. È vista come uno dei leader del cloud, ed è un fornitore leader di DMSA e un forte concorrente nel mercato dei DMSA basati sul cloud. I clienti non guardano solo alle capacità di Azure SQL Datawarehouse, ma all'insieme complessivo dei servizi di gestione e analisi dei dati offerti.
- **Fidelizzazione del cliente:** Praticamente tutti i clienti di riferimento per Microsoft hanno indicato che intendono acquistare licenze o capacità aggiuntive nei prossimi 12 mesi. Ciò dimostra ulteriormente che i clienti Microsoft apprezzano i suoi prodotti e la sua roadmap.

Avvertenze

- **Soddisfazione delle prestazioni nel cloud:** Un numero relativamente alto di clienti di riferimento, la maggior parte dei quali ancora in possesso dell'architettura Gen1

di Azure SQL Datawarehouse, ha indicato di aver affrontato problemi di performance con Azure SQL Datawarehouse, e che è necessaria un'ottimizzazione e una messa a punto specifica. L'architettura Gen2, che recentemente è diventata generalmente disponibile, fornisce prestazioni migliori, basate su benchmark e sull'utilizzo precoce da parte dei clienti, per cui ci aspettiamo che queste preoccupazioni diminuiscano man mano che un numero maggiore di clienti completa la transizione a Gen2.

- **Rapporto qualità/prezzo e flessibilità dei prezzi:** I punteggi dei clienti di riferimento per Microsoft per il rapporto qualità-prezzo lo collocano al di sotto della media complessiva di questo Magic Quadrant, e la flessibilità dei prezzi come media. Con l'aumento dell'adozione del cloud e l'elasticità della soluzione Microsoft, questi risultati dimostrano che Microsoft deve aiutare ulteriormente i suoi clienti con la prevedibilità dei costi e l'ottimizzazione dei costi. I clienti di SQL Server dovrebbero prendere in considerazione l'Azure Hybrid Benefit, che consente loro di passare le licenze on-premises nel cloud - fornendo alcuni risparmi sui costi delle licenze.
- **Funzionalità e maturità del prodotto:** I clienti di riferimento per Microsoft hanno evidenziato una serie di problemi di funzionalità e maturità del prodotto, in particolare per i prodotti cloud. Hanno anche messo in evidenza le disomogeneità delle funzionalità del prodotto tra i prodotti on-premises e il cloud.

Criteri di selezione

Criteri	Definizione
Semplicità e riduzione delle competenze richieste per l'uso della piattaforma	Le moderne soluzioni basate sul cloud sono spesso notevolmente facili da implementare e gestire, soprattutto se paragonate alla tradizionale soluzione on-premise.
Elasticità in volo:	Il che include la possibilità di aumentare o diminuire le risorse di calcolo per soddisfare le richieste. Alcune soluzioni sono più rapide da rispondere e possono essere scalate automaticamente.
Concorrenza: Capacità di segmentare il carico del Datawarehouse Loading:	La capacità di evitare il tiro alla fune tra l'elaborazione batch ETL e le richieste dell'utente finale a bassa latenza. Questo supporta un alto livello di concorrenza.
Calcolo e memorizzazione in scala indipendente:	Alcune soluzioni basate sul cloud possono espandere lo storage indipendentemente dalle risorse di calcolo, offrendo un maggior grado di flessibilità nell'implementazione, e significa che si paga solo per le risorse necessarie.
Gestione di formato di dati diversi:	Compresa la possibilità di caricare e interrogare dati JSON semi-strutturati spesso presenti in applicazioni web based
Archiviazione avanzata basata su colonne:	Il che ha un impatto enorme sulle prestazioni delle query analitiche, massimizzando al contempo la compressione dei dati.
Supporto SQL robusto:	Che include funzioni di finestra analitica e funzioni definite dall'utente.

Figure 4.7:Criteri di selezione di un cloud datawarehouse

Scoring

#	Criteria	AMAZON REDSHIT	MICROSOFT AZURE SQL DATA WAREHOUSE	SNOWFLAKE
1	Semplicità e riduzione delle competenze richieste per l'uso della piattaforma	3	3	5
2	Elasticità in volo:	4	3	4
3	Concorrenza:Capacità di segmentare il carico del Warehouse Loading:	3	3	5
4	Calcolo e memorizzazione in scala indipendente:	2	3	5
5	Gestione di formato di dati diversi:	4	4	4
6	Archiviazione avanzata basata su colonne:	3	3	4
7	Supporto SQL robusto:	4	4	4
	Total Rating 1-5	3,29	3,29	4,43

Tableau 4.8:Scoring per la soluzione finale

Soluzione ritenuta

La soluzione ritenuta è Snowflake un cloud datawarehouse

4.3.5 Benchmark della soluzione di reportistica



Figure 4.8: Magic Quadrant per la reportistica

Microsoft Power BI

Microsoft è leader in questo Magic Quadrant. Ha una roadmap di prodotti completa e visionaria e una massiccia presenza sul mercato attraverso il suo canale Microsoft Office.

Microsoft offre la preparazione dei dati, la scoperta di dati basati sulla visualizzazione, i cruscotti interattivi e l'analitica aumentata nella Power BI. È disponibile come opzione SaaS in esecuzione nel cloud Azure o come opzione on-premises in Power BI Report Server. Power BI Desktop può essere utilizzato come strumento di analisi personale autonomo e gratuito. L'installazione di Power BI Desktop è necessaria quando i power user stanno creando complessi mashup di dati che coinvolgono fonti di dati on-premise.

Microsoft rilascia un aggiornamento settimanale al suo servizio cloud, che ha aggiunto centinaia di funzionalità nel 2019. Le recenti aggiunte includono la visualizzazione dell'albero di decomposizione, la connettività dei dati LinkedIn e i miglioramenti della mappatura geografica.

Punti di forza

- **La diffusione " virale":** Anche se il prezzo di Power BI Pro, a 10 dollari al mese per utente, ha aiutato la trazione del prodotto sul mercato, questo è secondario rispetto al suo inserimento nell'Office 365 E5, che lo rende "autoseminante" in molte organizzazioni. I prompt di altri prodotti Microsoft Office, come Excel, incoraggiano gli utenti a "visualizzare in Power BI" per aumentare ulteriormente la sua esposizione - i suoi clienti di riferimento hanno rivendicato un maggior numero di implementazioni con più di 1.000 utenti rispetto a quelli di qualsiasi altro fornitore in questo Magic Quadrant.
- **Capacità del prodotto:** Per anni dopo il suo lancio nel 2013, Power BI è stato un prodotto "seguace" che doveva essere solo "abbastanza buono", dato il suo prezzo. Ora non è più così - e con le release del 2019, il servizio cloud Power BI Pro ha superato la maggior parte dei suoi concorrenti in termini di funzionalità. Ha superato molti, includendo capacità innovative per l'analisi aumentata e l'ML automatizzato. AI, come l'analisi di testi, sentimenti e immagini, sono disponibili all'interno di Power BI e attingono alle funzionalità di Azure. La stragrande maggioranza dei clienti di riferimento di Microsoft intervistati consiglierebbe Power BI senza riserve.
- **Visione completa del prodotto:** Microsoft continua a investire in un ampio set di capacità visionarie e a integrarle con Power BI. Questo si allinea bene con le tendenze di apertura, consumerizzazione e automazione identificate da Gartner come fattori chiave del mercato.

Avvertenze

- **Versione on-premise:** Rispetto al servizio cloud Power BI Pro, l'offerta on-premises di Microsoft presenta notevoli lacune funzionali, tra cui dashboard, analisi dello streaming, contenuti precostruiti, Q&A in linguaggio naturale, aumentazioni (ciò che

Microsoft chiama Quick Insights) e avvisi. Nessuna di queste funzioni è supportata in Power BI Report Server.

- **Solo azzurro:** Microsoft non offre ai clienti la flessibilità di scegliere un'offerta IaaS cloud. La sua offerta è disponibile solo in Azure.
- **Connettività:** Power BI offre una gamma molto ampia di connettori dati, ma il feedback degli utenti del servizio di richiesta clienti di Gartner indica che le prestazioni delle query dei data gateway on-premises sono variabili e richiedono uno sforzo di ottimizzazione. La connettività alle query dirette SAP BW e HANA è problematica - un problema noto su cui Microsoft sta lavorando. I clienti in genere scelgono invece di caricare i dati in Power BI, che è più performante.

Tableau

Tableau è un Leader in questo Quadrante Magico. Offre un'esperienza di esplorazione visiva che consente agli utenti aziendali di accedere, preparare, analizzare e presentare i risultati nei loro dati. Dispone di potenti capacità di marketing e di espansione delle capacità dei prodotti aziendali, ma c'è una certa incertezza sulla sua direzione come parte della forza vendite.

Nel 2019, Tableau ha ampliato in modo significativo la portata della sua offerta di prodotti, in particolare per quanto riguarda le capacità di analisi e di governance. Per l'analisi aumentata, Tableau ha introdotto sia Ask Data che Explain Data per fornire query in linguaggio naturale e approfondimenti automatici. Per la governance, Tableau ha migliorato Tableau Prep Builder (che viene fornito con Tableau Creator) e ha introdotto Tableau Prep Conductor per programmare e monitorare le attività di gestione dei dati. Tableau Prep Conductor viene fornito in bundle con il Catalogo Tableau come parte dell'Add-On per la gestione dei dati. Tableau ha anche introdotto il Server Management Add-On, che fornisce la gestione del server, la migrazione dei contenuti e l'ottimizzazione del carico di lavoro. Tableau ha anche spostato una parte significativa della sua base clienti nel cloud con Tableau Online.

Il 1° agosto 2019 Salesforce ha completato l'acquisizione di Tableau. Questa acquisizione crea opportunità e sfide per Tableau. Salesforce rafforza Tableau in tre aree

chiave emergenti del mercato della piattaforma ABI: AI, Cloud, e l'embedded analytics. Tuttavia, Tableau aveva già compiuto progressi significativi in tutte e tre le aree prima dell'acquisizione e deve ora conciliare una roadmap di prodotti complicata e sovrapposta.

Nota: Salesforce ha annunciato un piano per l'acquisizione di Tableau nel giugno 2019. L'acquisizione è stata completata il 1° agosto 2019. Tuttavia, l'autorità britannica per la concorrenza e i mercati (CMA) ha implementato un ordine "hold separate", che ha richiesto a Salesforce e Tableau di operare separatamente, in attesa di una revisione. La CMA ha revocato tale ordine alla fine di novembre 2019. Di conseguenza, i piani di integrazione dei prodotti e delle aziende non sono stati sviluppati e non sono stati disponibili per essere condivisi con Gartner in tempo per essere presi in considerazione per questo Magic Quadrant. In quanto tale, rappresentare l'offerta congiunta come un'unica entità non era giustificato, né sarebbe stato utile ai lettori a questo punto. Salesforce e Tableau sono quindi rappresentati separatamente in questo Magic Quadrant.

Punti di forza

- **L'entusiasmo dei clienti:** I clienti dimostrano un atteggiamento entusiasta nei confronti di Tableau, come dimostrano gli oltre 20.000 utenti che hanno partecipato alla sua conferenza annuale degli utenti del 2019. I clienti di riferimento hanno ottenuto un punteggio di Tableau ben al di sopra della media dell'esperienza complessiva. Questi utenti sono dei forti campioni per Tableau.
- **Facilità di esplorazione visiva e manipolazione dei dati:** Tableau consente agli utenti di ingerire rapidamente dati da un'ampia gamma di fonti di dati, di miscelarli e di visualizzare i risultati utilizzando le migliori pratiche di percezione visiva. I dati possono essere facilmente manipolati durante la visualizzazione, ad esempio durante la creazione di gruppi, bidoni e gerarchie.
- **Momento:** Tableau ha incrementato le sue entrate totali a poco più di 900 milioni di dollari nel primo trimestre del 2019, e ha raggiunto una crescita del 14% dai primi sei mesi del 2018 ai primi sei mesi del 2019. Tableau rimane una presenza costante nelle shortlist dei valutatori e continua ad espandersi all'interno della sua base

installata. I clienti di riferimento intervistati sono stati per lo più aggiornati all'ultima versione di Tableau e hanno espresso opinioni positive sull'esperienza di migrazione.

Avvertenze

- **Nuovi rischi in un mercato che cambia:** Tableau ha dominato l'era della scoperta dei dati visivi del mercato della piattaforma ABI, ma mentre il mercato si muove verso l'era aumentata, i nuovi entranti potrebbero rivelarsi dirompenti. Finora, tuttavia, Tableau ha fatto delle scelte valide in termini di equilibrio tra le priorità della roadmap di prodotto a breve e a lungo termine.
- **La governance:** Nonostante le nuove release di prodotti per la gestione dei dati e dei server che hanno aggiunto capacità di governance e amministrative nel 2019, persistono le percezioni di una governance e di un'amministrazione debole tra alcuni dei clienti di riferimento di Tableau. Queste sono evidenti anche durante alcune chiamate di Gartner.
- **Esperienza di vendita, contratti e costi:** La negoziazione con Tableau ha sempre avuto i suoi pro e i suoi contro. In generale, i clienti come Tableau Viewer come opzione a basso costo per i consumatori analitici, e stanno assecondando la spinta di Tableau verso i prezzi degli abbonamenti. Tuttavia, con il Server Management Add-on e il Data Management Add-on, i clienti di Tableau si troveranno di fronte a prezzi “à la carte”, il che significa che dovrebbero aspettarsi di pagare un extra per le nuove funzionalità.

Qlik

Qlik è un Leader in questo Quadrante Magico. La sua forte visione di prodotto per l'incremento di ML e AI è chiara, ma lo è anche il suo minore slancio di mercato, rispetto ai suoi principali concorrenti.

La soluzione ABI leader di Qlik, Qlik Sense, si basa sull'esclusivo Qlik Associative Engine, che da 20 anni alimenta i prodotti Qlik. Il motore consente agli utenti di tutti i livelli di competenza di combinare i dati ed esplorare le informazioni senza i limiti degli strumenti basati su query. Il motore cognitivo di Qlik aggiunge funzionalità AI/ML al

prodotto e lavora con l'Associative Engine per offrire suggerimenti di approfondimento consapevoli del contesto e per aumentare l'analisi.

Qlik continua a migliorare l'architettura basata sui microservizi della sua piattaforma e le capacità multicloud. È disponibile una versione SaaS completa di Qlik Sense Enterprise e costituisce la base della nuova esperienza di prova di Qlik basata sul SaaS. Qlik ha introdotto le "intuizioni associative" nel giugno 2019 come capacità di analisi aumentata che utilizza il motore cognitivo di Qlik per scoprire intuizioni altrimenti nascoste. L'acquisizione di Attunity di Qlik, il cui prodotto rimane autonomo, amplia le capacità di integrazione dei dati dell'ecosistema Qlik.

Punti di forza

- **Flessibilità di dispiegamento:** Qlik è stato uno dei primi fornitori ad offrire un'esperienza utente finale senza soluzione di continuità e capacità di gestione attraverso distribuzioni multicloud. La flessibilità di implementazione on-premise, o con qualsiasi importante fornitore di cloud, o di utilizzare una combinazione di entrambi gli approcci, o di utilizzare l'offerta SaaS completa di Qlik, rimane un punto focale della visione di Qlik.
- **Espansione delle capacità della piattaforma:** Il portafoglio di offerte di Qlik abbraccia una serie di fasi del ciclo di vita dell'analisi. Qlik Sense offre capacità di ricerca di dati visivi self-service per gli analisti o gli utenti aziendali, supportando al tempo stesso l'analitica integrata nello sviluppatore dalla stessa piattaforma. Qlik Data Catalyst è utilizzato per la catalogazione e la governance aggiuntiva. Inoltre, sebbene la piattaforma di integrazione dati Qlik Data Integration Platform (in precedenza Attunity) sia un'offerta stand-alone, aggiunge potenti capacità di integrazione e di movimento dei dati sotto l'ombrello di Qlik.
- **Aumento e alfabetizzazione dei dati:** La capacità di approfondimento associativo utilizza l'esclusiva "esperienza associativa" di Qlik per scoprire automaticamente gli approfondimenti sui dati che altrimenti potrebbero essere stati tralasciati da strumenti basati su query. Mentre gli utenti delle funzionalità incrementate possono essere persone non analisti, il progetto di alfabetizzazione dati di Qlik aiuta gli utenti di tutti i livelli, clienti Qlik o meno, a comprendere e utilizzare meglio i dati.

Avvertenze

- **Momento:** Dopo un periodo di riallineamento nel 2019, Qlik sta ora aggiungendo di nuovo personale. Nel 2019 l'azienda ha compiuto alcune mosse visionarie: acquisizioni di tecnologia, importanti rilasci di prodotti e offerte di migrazione aggiornate. Tuttavia, rispetto ad altri Leader il suo slancio rimane basso, a giudicare dai dati di ricerca e dalle richieste dei clienti di Gartner e da una serie di altri indicatori. Inoltre, meno della metà dei clienti di riferimento di Qlik ha dichiarato di aver fornito il proprio strumento ABI standard per le imprese.
- **Migrazione dei prodotti:** Nonostante l'enfasi posta da Qlik sulla fornitura di supporto e risorse dedicate ai clienti che passano da QlikView a Qlik Sense, i clienti di riferimento intervistati per questo fornitore hanno identificato l'esperienza di migrazione come una preoccupazione chiave, rispetto a quella di tutti gli altri fornitori.
- **Utilizzo self-service:** Sebbene Qlik Sense sia progettato per supportare il self-service guidato visivamente, i clienti di riferimento di Qlik hanno riferito che la maggior parte dei loro utenti sta consumando cruscotti parametrizzati. Detto questo, l'esperienza associativa di base di Qlik offre un modo alternativo per scoprire automaticamente le intuizioni, che può ridurre la necessità di alcune forme di self-service.

Criteria di selezione

Item	Power BI	Tableau	Qlikview
Versioni di base	La versione desktop di Power BI è gratuita e gli utenti non devono pagare nulla. Possono semplicemente scaricare il software e iniziare ad usarlo	Anche una versione base di Tableau è gratuita, ma ha caratteristiche molto limitate per iniziare..	Anche la versione base di Qlikview è gratuita, ma come Taleau ha caratteristiche molto limitate con cui lavorare.
Versioni avanzate	Power BI viene fornito con due versioni avanzate, ovvero Power BI Pro & Power Premium Packages. Questi pacchetti costano 10 dollari al mese per utente, sembrano meno costosi di Tableau e Qlikview	La versione avanzata di Tableau costa 100 dollari per utente, che sembra essere costosa per le industrie di media scala.	30 dollari al mese per utente è il costo per l'acquisto di una versione avanzata di Qlikview. Questo si trova a metà strada tra questi tre in termini di costo.
Analisi e interpretazione	Power BI ha un sacco di funzioni avanzate con cui lavorare così bene per l'analisi e l'interpretazione dei dati.	Il tableau è il vincitore tra questi tre. A causa di molte opzioni di drill-down e di filtraggio.	Qlikview non dispone di un'analisi comprovata, quindi deve lavorare su questi aspetti.
Acquisizione e immagazzinamento dei dati	Power BI è in grado di recuperare dati da qualsiasi luogo e in qualsiasi momento. Ha una grande varietà di fonti di dati come web, cloud SQL, Azure, ecc..	Tableau anche in grado di avere buone connessioni alla sorgente dati, probabilmente alla pari con la Power BI	La trasformazione dei dati è l'elemento chiave di Qlikview e anche questo strumento è in grado di recuperare i dati da qualsiasi luogo.
Facilità di apprendimento	Power BI sembra solo una versione avanzata di Excel, quindi la gente trova facile lavorare con questo strumento a causa della somiglianza dell'interfaccia excel, delle formule e di altre caratteristiche	Tableau non è il prodotto comune dove tutte le persone comuni possono usarlo. Solo chi è a livello di scienza dei dati è a conoscenza delle caratteristiche di questo prodotto.	Anche Qlikview richiede un background in Data Science a causa delle competenze di programmazione necessarie per lavorare con questi strumenti.
Comunità dei clienti	Poiché Power BI viene fornito con il pacchetto Office 365, coloro che acquistano altri prodotti Microsoft possono essere sottoposti anche a questo strumento	Tableau ha una storia a sé stante, così come un'enorme quantità di comunità di clienti	Qlikview può essere utilizzato solo da persone con conoscenze di programmazione, quindi non sono molte le persone con un background di programmazione in termini di comunità di clienti

Tableau 4.9: Criteri di selezioni per le soluzioni di reportistica

Scoring

Criteria	Power BI	Tableau	Qlikview
Cost	Start at \$9,99 /user/Month	Start at \$500/user/Year	Start at \$1300/User/Year
Ease of Use (10%)	4	5	3
Implementation (20%)	5	4	2
Customization (20%)	2	3	4
Integrations(20%)	5	3	3
Customer Support(15%)	5	4	3
Features & Add-ons(15%)	2	3	3
Total Rating 1-5	3,85	3,55	2,17

Tableau 4.10: Esito per le soluzioni di reportistica

Soluzione ritenuta

La soluzione ritenuta in base allo scoring è Power BI e SSRS un prodotto Microsoft

4.4 Raccomandazione della soluzione finale

Dopo aver effettuato un benchmarking e analizzato le soluzioni proposte dal mercato, abbiamo individuato diverse soluzioni in grado di soddisfare le nostre esigenze. Descriveremo prima l'architettura di destinazione e poi la presentazione dettagliata delle soluzioni selezionate.

4.5 Descrizione dell'architettura della soluzione finale

L'architettura della soluzione finale è strutturata come segue:

- La soluzione Informatica Cloud servirà come piattaforma di integrazione e sarà la piattaforma di progettazione per le ETL
- Abbiamo scelto Amazon S3 come soluzione di archiviazione dati.
- Snowflake è la soluzione datawarehouse selezionata e sarà composta dagli elementi descritti nella figura seguente (Fig 4.9)
- Infine, lo strumento di reporting scelto è Power BI di Microsoft.

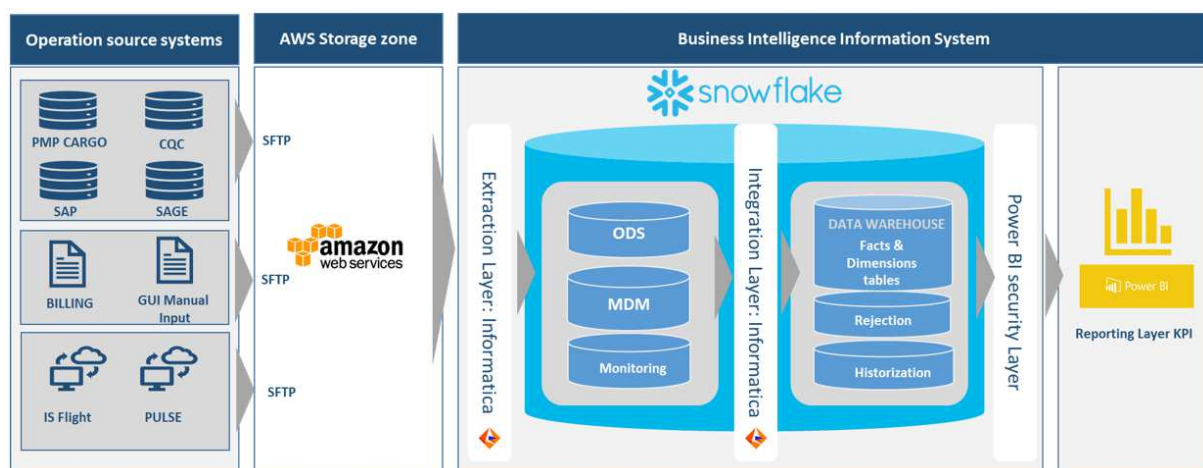


Figure 4.9: Architettura target

Nei paragrafi seguenti verranno presentate in dettaglio le soluzioni scelte.

4.5.1 Soluzione di data integration: Informatica

Informatica Intelligent Cloud Services permette di rilevare le sfide legate all'iPaaS

Informatica offre un'iPaaS di nuova generazione, che ti permette di soddisfare le esigenze della digital transformation guidata dai dati in un ambiente multicloud. Informatica Intelligent Cloud Services™

(IICS) è una soluzione end-to-end, modulare e completa per la gestione dei dati in un mondo ibrido e multicloud. È creata per scalare e fornire dati alla velocità del business, soddisfacendo al tempo stesso le esigenze dei nuovi e più complessi casi d'uso.

IICS offre una “user experience” moderna e unificata tra una vasta gamma di modelli di integrazione iPaaS ed è costruito su un'architettura di microservizi modulare che è personalizzabile e semplice da esplorare. IICS offre un ambiente produttivo per la gestione dei dati, accelerato dall'engine CLAIRE™, che fornisce i consigli guidati dall'intelligenza artificiale e dal machine learning per automatizzare e accelerare il lavoro di integrazione e gestione dei dati. La nuova iPaaS supporta l'innovazione del business con modelli di integrazione avanzati che collegano tutti i tipi di dati tra ambienti Cloud, ibridi e on-premise e si adattano agli utenti del business o dell'IT.

Funzioni, vantaggi

Le seguenti sono funzionalità fondamentali aggiuntive di un'iPaaS di nuova generazione:

- **Data Integration nel Cloud.** Automatizza e velocizza i tuoi workload nei datawarehouse Cloud con template di soluzione out-of-the-box su misura per i datawarehouse Cloud Snowflake
- **Integrazione di applicazioni nel Cloud.** Integra le fonti dati in tempo reale con processi di business intelligenti che coprano applicazioni Cloud e on-premise (ad es. Salesforce, NetSuite, Marketo, SAP).
- **Cloud Integration Hub.** Semplifica l'architettura di data integration, sostituendola con una semplice architettura publish-and-subscribe, dove le fonti dati pubblicano i dati una sola volta per tutti i subscriber. I subscriber ricevono i dati nel formato e nel timeframe necessario.
- **Data Quality nel Cloud.** Profila i dati, esegui la discovery dei problemi legati alla data quality e crea ed esegui regole di data quality per correggere i problemi

in modo continuo. Questo garantisce dati affidabili per le decisioni e i principali processi di business.

- **Master Data Management nel Cloud.** Fornisci una vista a 360 gradi dei dati sulle entità di business fondamentali (ad es. cliente, fornitore, partner) da diversi sistemi di fonti dati interne ed esterne.
- **Sicurezza dei dati nel Cloud.** Identifica rapidamente i dati sensibili e proteggili in modo da adempiere alle normative e alle policy di sicurezza dei dati.

4.5.2 Soluzione di cloud datawarehouse: Snowflake

Architettura di snowflake: Technical assessment

Snowflake è un datawarehouse di dati cloud offerto come Software-as-a-Service (SaaS) su più CLOUD (AWS e Azure) per il carico di lavoro di analisi. Questo è un servizio cloud simile a AWS Redshift, Google Big Query [8].

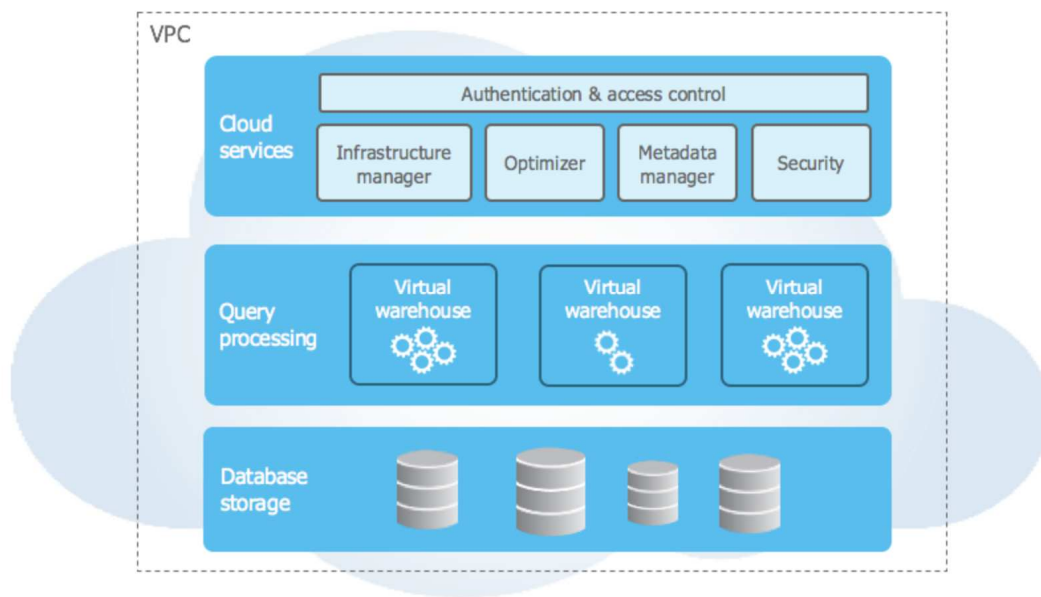


Figure 4.10: Architettura di Snowflake

Come potete vedere ci sono tre componenti principali:

1. **Servizi cloud:** Questo componente si occupa di una varietà di servizi come l'autenticazione, il controllo degli accessi, la gestione dei metadati, la gestione delle infrastrutture, l'analisi delle query e l'ottimizzazione.

2. **Elaborazione delle query:** Snowflake ha un concetto di warehouse virtuali che sono come un cluster MPP separato che può essere istanziato su richiesta. Sono disponibili in diverse dimensioni da X-small a 4X-Large e Snowflake si carica in base alle dimensioni del warehouse virtuale. Ogni magazzino virtuale è un'entità indipendente e non condivide risorse di calcolo con altri magazzini virtuali e le prestazioni di un warehouse virtuale non sono influenzate da altri.
3. **Archiviazione di database:** I dati memorizzati in snowflake sono automaticamente ottimizzati, compressi in un formato proprietario a colonne e sono memorizzati in un Cloud storage come S3 o Azure blob storage. Snowflake gestisce come viene memorizzato, dove viene memorizzato ed è accessibile solo da query SQL eseguite utilizzando snowflake.

Questo paragrafo descrive in dettaglio cinque caratteristiche del nostro multi-cluster, l'architettura dei dati condivisa, che sono una ragione sufficiente per spostare il datawarehouse on-premises e i carichi di lavoro analitici nel cloud.

Il problema con la maggior parte delle piattaforme nel cloud

Nell'ultimo decennio, probabilmente avete sentito da varie aziende tecnologiche che il modo migliore e più conveniente per creare rapidamente approfondimenti da un mix di dati, in particolare un mix di dati multi-strutturati (JSON, Avro, tabelle, ecc.), consiste nello spostare i dati analitici sui dati. La Figura 1 rappresenta l'architettura tipica di MPP o cluster. La teoria afferma che se si collegano i dati e si calcolano le risorse insieme e le si diffonde attraverso l'hardware delle materie prime attraverso l'elaborazione in parallelo massiccia (MPP) o forse Hadoop su server distribuiti, l'analisi sarebbe rapida e a basso costo.

Typical MPP/Cluster Architecture

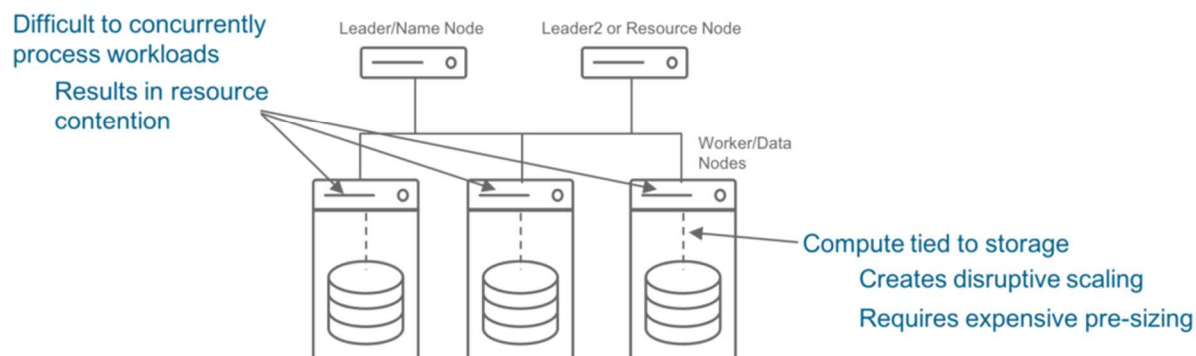


Figure 4.11: Architettura tipica di MPP / Cluster

Tuttavia, il passare del tempo e le frustrazioni ben pubblicizzate degli utenti hanno dimostrato che legare il calcolo allo storage crea una serie di conseguenze negative per l'azienda. L'accordatura, per esempio, diventa un lavoro ingrato. Il ridimensionamento dell'ambiente è dirompente, la concorrenza delle query è limitata e il supporto simultaneo di carichi di lavoro separati (ad es. Caricamento di nuovi dati e query di BI) è quasi impossibile. Inoltre, l'accesso ai dati per interrogarli frequentemente richiede di coinvolgere i professionisti dell'ingegneria dei dati della propria organizzazione con le competenze linguistiche Java, Python o Scala necessarie per estrarre i dati dai cluster.

Ma passare attraverso il team di ingegneria dei dati spesso significa ritardi. Non è il tipo di processo da attuare se si desidera essere un'organizzazione agile, rapida e basata sui dati. Inoltre, i dirigenti preferirebbero che i data engineer si concentrassero sullo sviluppo di nuovi prodotti e servizi.

Ora, tuffiamoci nell'architettura multi-cluster di dati condivisi di Snowflake e sui motivi principali.

Motivo 1: Ridimensionamento senza interruzioni

Snowflake è stato fondato sulla convinzione che il collegamento tra elaborazione e archiviazione non sia un approccio efficace per un ridimensionamento senza limiti e senza interruzioni. L'architettura multi-cluster di dati condivisi di Snowflake (vedere la Figura 2) separa il ridimensionamento delle risorse di calcolo dalle risorse di

archiviazione, consentendo in tal modo un ridimensionamento continuo e senza interruzioni.

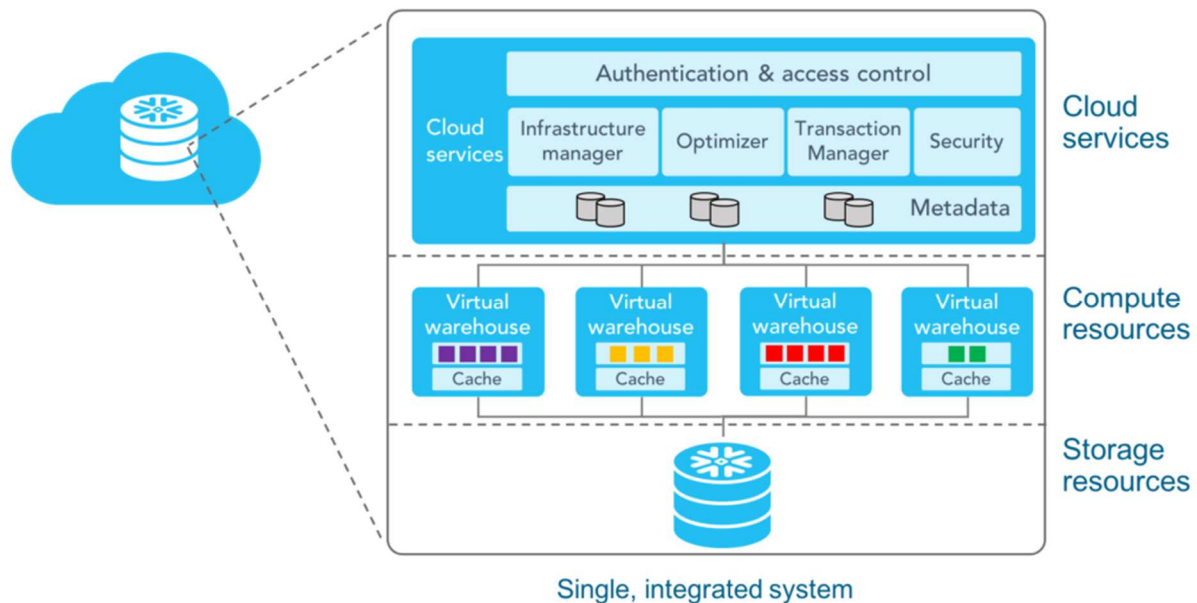


Figure 4.12: Snowflake Multi-Cluster Share Data Architecture

In Snowflake, mentre le query sono in esecuzione, le risorse di calcolo possono essere ridimensionate senza interruzioni o tempi di inattività e senza la necessità di ridistribuire / riequilibrare i dati (archiviazione). Il ridimensionamento delle risorse di calcolo può avvenire automaticamente, con il rilevamento automatico. Ciò significa che il software Snowflake è in grado di rilevare automaticamente quando è necessario il ridimensionamento e ridimensionare l'ambiente senza l'intervento dell'amministratore o dell'utente.

Per l'archiviazione, è possibile ridimensionare le risorse praticamente a qualsiasi capacità senza il costo aggiuntivo di aggiungere risorse di calcolo non necessarie. Entrambi i gruppi di risorse possono aumentare o diminuire, inclusa la sospensione nel caso di risorse di calcolo. Con Snowflake puoi personalizzare facilmente e senza soluzione di continuità le risorse in base alle esigenze specifiche della tua organizzazione.

I cluster MPP tradizionali nel cloud potrebbero richiedere di mettere il database in modalità di sola lettura o potrebbero richiedere di portare il cluster offline. Con altre

architetture, se si è in grado di ridimensionare il calcolo "automaticamente", è comunque necessario impegnare manualmente il sistema per comunicarlo.

Motivo 2: Separazione fisica con integrazione logica

Altre offerte di datawarehouse cloud stanno iniziando a separare elaborazione e archiviazione. Tuttavia, dal punto di vista del datawarehouse cloud, la separazione dei livelli di elaborazione e archiviazione non è sufficiente. Se tutto ciò che è stato realizzato fosse la separazione di elaborazione e archiviazione, si avrebbe una topologia che assomiglia al server più topologie di rete di archiviazione (LAN o SAN) tipiche dei data center locali che, in alcuni casi, risalgono al 1990. In questi ambienti, in genere è necessario integrare separatamente funzionalità quali alta disponibilità, protezione dei dati, conservazione dei dati, gestione dei metadati e sicurezza. Questo è più complesso e costoso.

Il porting di un ambiente locale sul cloud lascia gli stessi problemi. Inoltre, fare affidamento su ciò che offre intrinsecamente un provider cloud non fornirà tutte le solide funzionalità in grado di soddisfare una vasta gamma di requisiti. Le esigenze di una grande impresa saranno diverse da quelle di una PMI.

Come mostrato nella *Figura 4.12*, l'architettura di dati condivisi multi-cluster di Snowflake brilla perché le risorse di elaborazione e archiviazione sono fisicamente separate, ma fanno logicamente parte di un singolo sistema di datawarehouse integrato, moderno e integrato nel cloud. L'architettura include servizi cloud integrati come provisioning trasparente delle risorse, gestione automatica dei metadati e resilienza (ad es. Backup / conservazione dei dati protezione dai guasti dei nodi).

Un singolo sistema integrato elimina gli sforzi di costruzione dei cluster che è necessario fare per far lavorare insieme strati separati.

Motivo 3: Elaborare carichi di lavoro separati contemporaneamente su dati coerenti

Quando si tratta di supportare carichi di lavoro multipli e separati con architetture che non sono costruite su cloud, in genere accade una delle due cose. In primo luogo, non è possibile isolare i carichi di lavoro perché la maggior parte delle architetture sono cluster

singoli: i carichi di lavoro condividono risorse. Questo porta a una lotta per le risorse e prestazioni scadenti, il che significa che potrebbe essere necessario ricorrere alla pianificazione fuori orario. In alternativa, separerai i carichi di lavoro simultanei in cluster diversi ma ci sarà il rischio di incoerenza dei dati poiché i cluster separati apportano autonomamente modifiche ai dati. Ciò può comportare letture sporche e analisi imprecise.

Per evitarlo, dovrai applicare un certo livello di sovraccarico sotto forma di gestione e / o sincronizzazione di tabelle di modifica o metriche esterne.

L'architettura multi-cluster di Snowflake ti consente di supportare tutti i carichi di lavoro separati che puoi immaginare. È possibile fornire a ciascun carico di lavoro il proprio motore di calcolo e riposare facilmente. Ci occupiamo di garantire la coerenza dei dati (un'unica fonte di verità) con integrità conforme ACID.

Motivo 4: Caricare, trasformare e integrare rapidamente JSON e altri dati semi-strutturati

L'architettura di Snowflake consente un rapido consolidamento di tutti i diversi dati su un'unica piattaforma, consentendo l'analisi su set di dati più grandi e più ampi. Più è possibile eseguire analisi su dati consolidati, più approfondimenti completi è possibile sviluppare. Ciò è reso possibile dal nostro tipo di dati VARIANT brevettato. Snowflake carica i dati semi-strutturati come tipo di dati VARIANT, consentendo di interrogare rapidamente JSON in modo completamente relazionale (vedere la Figura 3.).

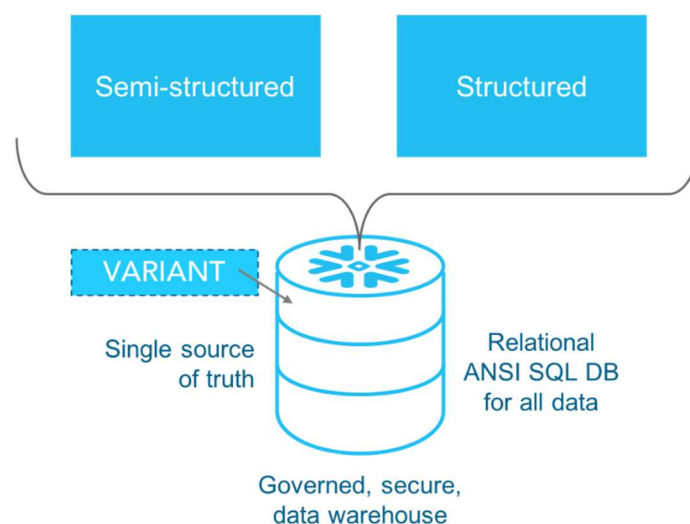


Figure 4.13: Consolidamento di diversi formati di dati

Altre piattaforme potrebbero richiedere più archivi di dati e griglie di query per gestire analisi e data warehousing su dati misti. Questo è un approccio di architettura dei dati più complesso.

Motivo 5: Condivisione dei dati in tempo reale e sicura

L'architettura dei dati condivisi multi-cluster di Snowflake consente un rapido accesso e analisi per gestori di programmi, team di marketing, dirigenti, analisti aziendali e data scientist. Puoi anche condividere i dati all'interno della tua organizzazione e al di fuori della tua azienda per includere il tuo ecosistema di partner commerciali e qualsiasi consumatore di dati esterno (vedi Figura 4.14.).

Puoi farlo con un accesso granulare e sicuro e viste sicure dei tuoi dati senza alcun movimento di dati. Abilitata dalla gestione completa dei metadati, l'architettura di Snowflake consente di specificare i dati live che si desidera condividere. Concedi semplicemente l'accesso e puoi condividere un intero database o solo una parte di esso con accesso a dati sempre aggiornati.

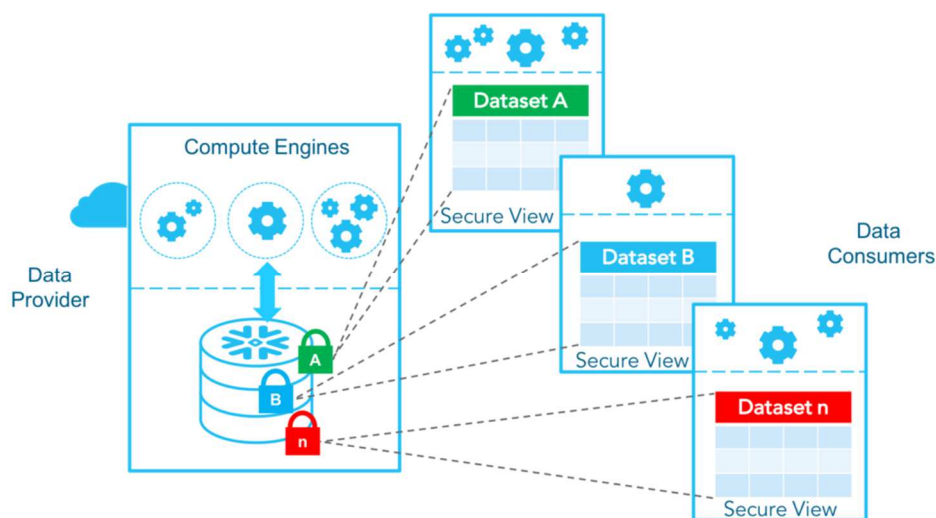


Figure 4.14: Condivisione dei dati del fiocco di neve - Nessuna copia o spostamento dei dati

Altri approcci alla condivisione dei dati ti limiteranno alla condivisione di un intero database o ti richiederanno sforzi laboriosi per crittografare e trasmettere dati, incluso l'onere di separare i dati privati e sensibili. A peggiorare le cose, i dati saranno stantii nel momento in cui un elemento cambia dalla tua parte.

4.5.3 Soluzione di Reporting: Power BI

Power BI è un insieme di servizi software, applicazioni e connettori che lavorano insieme per trasformare diverse fonti di dati in informazioni visive dettagliate e interattive. I dati possono essere sotto forma di un foglio di calcolo Excel o di una raccolta di datawarehouse ibridi locali o cloud-based. Power BI vi permette di accedere facilmente alle fonti di dati, di visualizzare e scoprire ciò che è importante e di condividere queste informazioni con chiunque desideriate.

Componenti di Power BI

La Power BI è composta dai seguenti elementi:

- Un'applicazione desktop per Windows chiamata Power BI Desktop.
- Un servizio online Software as a Service (SaaS) chiamato Power BI service.
- Applicazioni mobili Power BI per dispositivi Windows, iOS e Android.

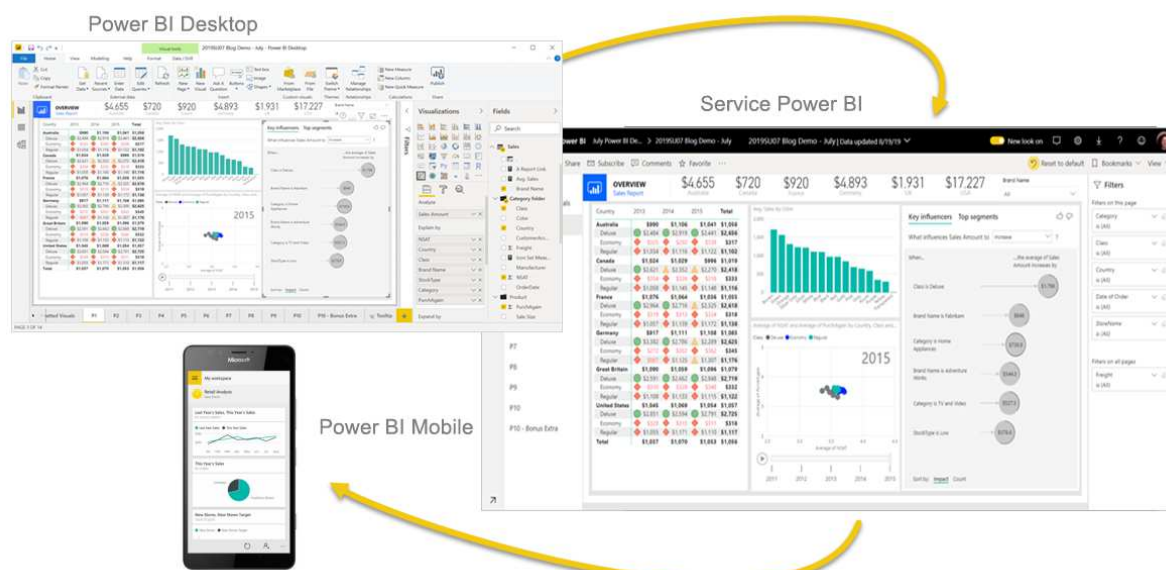


Figure 4.15:Componenti di Power BI

Questi tre elementi - Power BI Desktop, il servizio e le applicazioni mobili - sono progettati per consentire di creare, condividere e consumare intuizioni di business in modo ottimale, in base alle necessità o al proprio ruolo.

Un quarto elemento, Power BI Report Server, consente di pubblicare i report di Power BI su un server di report locale dopo averli creati in Power BI Desktop.

Flusso di lavoro nella Power BI

Un flusso di lavoro standard in Power BI inizia con la connessione alle fonti di dati e la creazione di un report Power BI Desktop. Pubblicare quindi questo rapporto da Power BI Desktop al servizio Power BI, e diffondete il rapporto in modo che gli utenti finali di Power BI e coloro che lavorano su dispositivi mobili possano visualizzarlo e interagire con esso. Questo flusso di lavoro è comune e mostra la complementarità tra i tre elementi principali di Power BI.

Reportistica locale con Power BI Report Server

È possibile creare, distribuire e gestire report di Power BI mobili e a paginazione in locale con la suite di strumenti e servizi "out-of-the-box" di Power BI Report Server.



Figure 4.16: Processo di distribuzione reportistica

Power BI Report Server è una soluzione da implementare dietro il firewall. Poi distribuite i rapporti agli utenti appropriati in vari modi per visualizzarli in un browser web, su un dispositivo mobile o come e-mail. E poiché Power BI Report Server è compatibile con Power BI nel cloud, è possibile migrare nel cloud quando è possibile.

Capitolo 5

Processo della logistica cargo specificazione dettagliata dei KPI da implementare

5.1 Presentazione del trasporto logistico aero cargo

L'aero cargo è un'area logistica speciale che consente l'interscambio terra-aria. Si tratta di un aeroporto dedicato al trattamento delle merci, spesso parte integrante di un sistema aeroportuale ampio in cui sono offerti servizi sia ai passeggeri che alle merci. In via generale le funzioni aeroportuali possono essere schematicamente rappresentate come in Fig. 5.1.

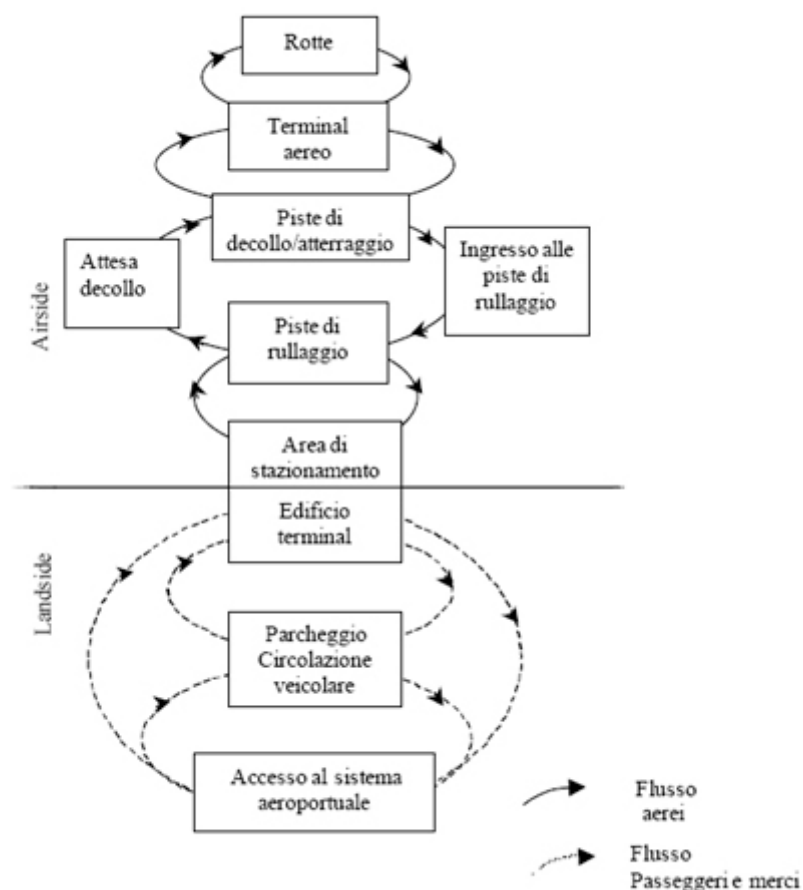


Figure 5.1: Funzioni aeroportuali

All'interno di un aeroporto è possibile distinguere due differenti aree operative: Air side e land side. Dell'air side fanno parte le piste, le aree di stazionamento dei velivoli ed il sistema di controllo del traffico aereo. Fanno invece parte del land side le strutture destinate all'amissione biglietti, servizi, banchi di accettazione, area consegna e ritiro. Si considerano appartenenti al land side anche i sistemi di accesso all'area aeroportuale e di movimento all'interno della stessa [9].

La struttura di una rete logistica aeroportuale a supporto del trasporto delle merci via aria è molto complessa e prevede tre linee di magazzini (Figura 4.2).

La prima linea, direttamente a contatto con l'aeroporto e situata nei pressi delle piste di atterraggio, è rappresentata dai magazzini di transito (Gateway) in cui vengono svolte attività di cross-docking e di attraversamento. È necessario che la merce in transito soste per il minimo tempo possibile nei magazzini di prima linea per garantire fluidità ai flussi di merce arrivo e in partenza dallo scalo aeroportuale.

La seconda linea invece è costituita dai magazzini degli integratori e degli spedizionieri, localizzati nelle vicinanze dell'aeroporto (10-15 Km), per garantire una continuità al flusso. Nella seconda linea, la merce può sostare per tempi più prolungati, allo scopo di permettere all'integratore e/o allo spedizioniere di ottimizzare i carichi e organizzare gli ordini, che proseguiranno verso la terza linea, ovvero verso i magazzini dei fornitori di servizi logistici o dei grandi clienti industriali e commerciali.

In tale stadio, la merce consegnata dalla seconda linea viene stoccata e subisce interventi strettamente concordi con le politiche di inventory management specifiche di ciascun utilizzatore. È opportuno sottolineare come la terza linea non “veda” l'aeroporto, né le attività che generano i flussi, grazie alla schermatura della seconda linea, che rappresenta di fatto un punto di disaccoppiamento, che permette di separare completamente le attività di distribuzione sul territorio dalle attività aeroportuali.

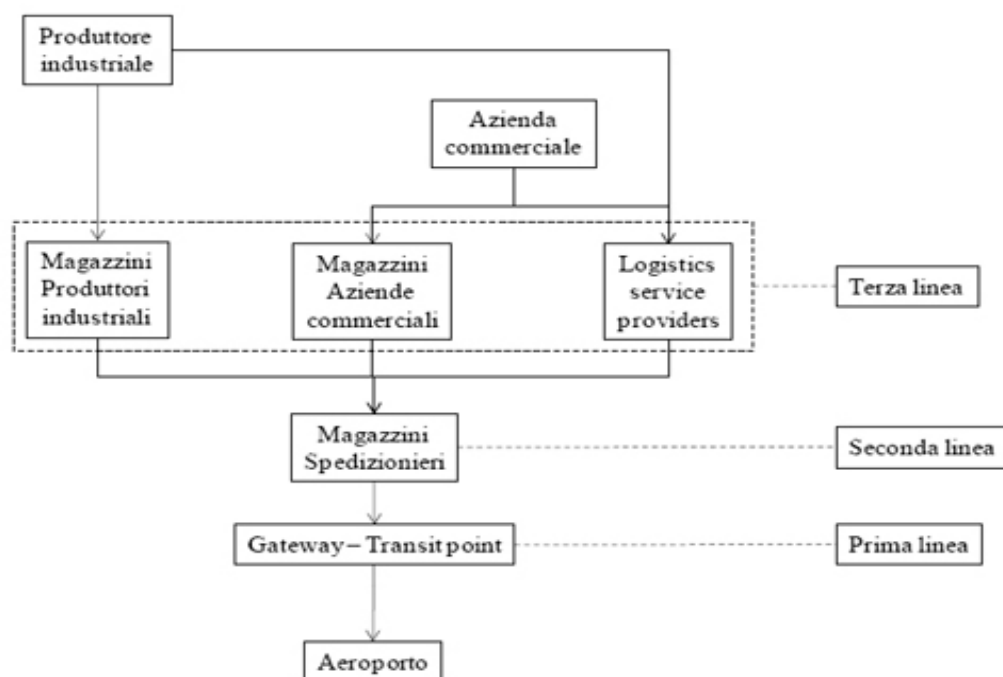


Figure 5.2:Struttura rete logistica aeroportuale

Di seguito si farà riferimento esclusivamente ai magazzini di prima linea localizzati all'interno dell'aeroporto.

Le operazioni di movimentazione merce in un aeroporto riguardano la preparazione dei carichi, il carico/scarico dall'aero e il trasferimento delle merci in magazzini opportunamente attrezzati. Nello specifico, le merci in uscita dall'aeroporto, provenienti via aia, vengono controllate e consolidate attraverso la formazione di pallet e container. Invece le merci in ingresso all'aeroporto, dopo aver attraversato i controlli doganali, vengono deconsolidate al fine di formare i carichi da spedire via aia. Infine, le merci in attraversamento, che quindi subiscono un trashingment aero-aero, vengono scaricate dall'aero di origine, eventualmente stoccate in magazzino e ricaricate sull'aero di destinazione. In relazione alle attività da svolgere, l'area cargo di un aeroporto deve essere dotata di magazzini attrezzati per accogliere mere deperibile e non, per effettuare controlli anche mediante l'ausilio di apparecchiature a raggi X, per garantire un rapido deflusso della merce tra la modalità aerea e le modalità di trasporto terrestri (strada e ferroviaria)

In definitiva, in un magazzino cargo aeroportuale si possono avere tre tipologie di flussi:

- **Flussi in export**, ovvero merce proveniente via terra e destinata al trasporto via aria;
- **Flussi in import**, ovvero merce proveniente via aria e destinata al trasporto via terra
- **Flussi di transito**, costituiti dalla merce che subisce un trashipment da aero ad aereo; questi flussi sono presenti solo nei magazzini cargo di aeroporti Hub

Le aree funzionali ed operative di un magazzino cargo sono (Fig. 5.3):

- **Area avio-camionata** di interfaccia tra l'air side ed il land side, si tratta di un'area di ricevimento/spedizione delle merci (ULD et pallet aeronautici ULD) provenienti o diritti agli aerei;
- **Area trattamento ULD** costituita da isole di pallettizzazione e spalettizzazione, per il confezionato e lo sconfezionamento delle ULD;
- **Area stoccaggio** attrezzata per accogliere sia ULD che merce sfusa; ULD in arrivo o in partenza;
- **Area di stazionamento ULD** per l'allineamento, sul lato aria, delle ULD in arrivo via terra.
- Area di interfaccia con i trasporti terrestri per l'arrivo/partenza della merce via terra.

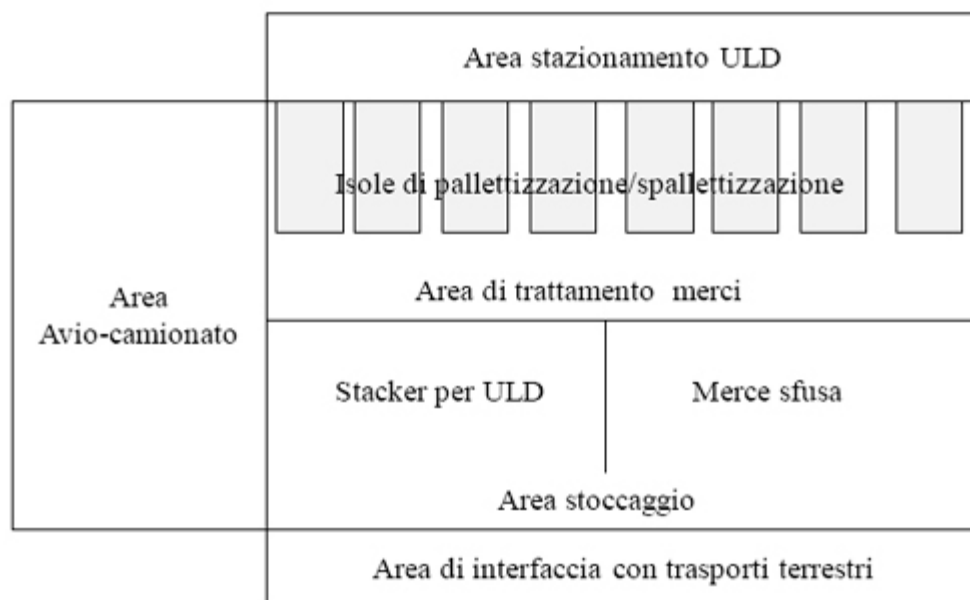


Figure 5.3: Aree funzionali di un magazzino cargo aeroportuale

L'area avio-camionato è generalmente costituita da linee di rulliere con testa una piattaforma di interfacciamento con camion (Fig) per il trattamento delle ULD in arrivo all'aeroporto già confezionate e pronte per essere caricate sui velivoli. In particolare

l'ULD, scarico dal camion, viene pieno pesato su apposito dispositivo optoelettronico multiraggio in grado di rilevarne la lunghezza, l'altezza e la larghezza. Successivamente l'ULD viene trasferito sulle rulliere di accumulo della linea e avviato all'area di interfaccia con l'airside (Fig.5.4).



Figure 5.4:Area avio-camionate lato land side (Milano Malpensa)

L'area di trattamento delle ULD è la zona dove si procede alla preparazione dei carichi provenienti via terra e destinati al trasporto via aria e viceversa. L'area è costituita da isole di confezionamento o pallettizzazione (Build Up) e da isole di confezionamento o spallettizzazione (Break down). Nelle isole di pallettizzazione le ULD in Export vengono confezionate, partendo dai singoli colli di merce sfusa, e preparate per spedizione via aria. Ogni isola è costituita da una o più piattaforme elevabili, denominate work station, comandate da un circuito oleodinamico, sulle quali sono montati dei piani a rulli motorizzati che permettono la movimentazione delle ULD. Le piattaforme sono collegate tra di loro con la linea in uscita verso l'air side da una serie di rulliere. Le piattaforme sono alloggiate in fosse in cemento armato, profonde circa 3 m, per permettere alla piattaforma stessa di scendere sotto piano pavimentato in modo che l'operatore addetto al confezionamento dell'unità di carico (ULD) possa operare in condizioni di sicurezza ed ergonomia. Prima della spedizione si procede al controllo del peso e delle dimensioni della ULD formata, mediante un sistema di pesatura a celle di carico posto sulla piattaforma ed un dispositivo optoelettronico per il controllo di sagoma. Nelle isole di spallettizzazione (Fig.5.5), invece, le ULD in import vengono

decomposte, ossia, i singoli colli vengono rimossi dalla ULD per poter procedere a la formazione di pallet e container da trasportare via terra. Queste isole sono in tutto identiche a quelle di palletizzazione.

L'area di stoccaggio è suddivisa in due parti: un magazzino per merce sfusa ed un magazzino per ULD.



Figure 5.5:Isola di spallettizzazione

Tali veicoli possono essere guidati da un uomo a bordo oppure possono essere complemento automatici.



Figure 5.6:Stacker (Milano Malpensa)

Generalmente lo stacker è suddiviso in due zone:

- **Zona import**, dove vengono stoccate le ULD, provenienti dagli aeromobili, in attesa di essere sconfezionate o consegnate agli autocarri tramite l'avio-camionato.
- **Zona export**, dove vengono stoccate le ULD, già confezionato nelle isole di build-up o provenienti dagli autocarri tramite l'avio-camionato, in attesa di essere spedite per via aerea.

La "staging area" è l'area di parcheggio delle ULD, è un termine tecnico (Fig 5.7), è un'area del deposito merci o delle ULD su banchine speciali. Queste banchine sono costituite da trasportatori a rulli in attesa di essere inseriti nel sistema di movimentazione ULD, ripartiti in volo e pronti per essere prelevati con i mezzi di rampa. Le unità di carico ULD vengono raccolte o selezionate in banchina attraverso un televisore che può essere guidato da un operatore o in modalità automatica.



Figure 5.7:Staging Area (Milano Malpensa)

In fine l'aera di interfaccia con i trasporti terrestri (Fig.5.8) in generale è costituita da ribalte a rampa telescopica per consentire ai TIR di accostarsi al magazzino e garantire le operazioni di trasferimento della merce dal magazzino cargo al veicolo utilizzato per la distribuzione terrestre mediante carrelli elevatori.



Figure 5.8:Area di interfaccia con i trasporti terrestri (Milano Malpensa)

Analisi della terminologia: AWB

La Lettera di Trasporto Aereo (AWB) è un documento necessario per l'esportazione e l'importazione di merci trasportate per via aerea. Essa è considerata come prova del contratto tra il vettore (compagnia aerea) e il mittente, la prova dell'accettazione della merce, la fattura, un certificato di assicurazione e un manuale di istruzioni per la movimentazione. Può essere rilasciata dalla compagnia aerea, dal mittente o dal destinatario.

Nel caso di una spedizione al dettaglio, viene rilasciata dall'agente della compagnia aerea. Se si tratta di una spedizione consolidata, il consolidatore emette una lettera di vettura aerea della casa. Viene poi firmata dalla compagnia aerea prima che la merce venga effettivamente caricata e consegnata al mittente. L'AWB certifica la spedizione effettiva una volta che la compagnia aerea ha inserito la data e il numero di volo sull'AWB.

Informazioni obbligatorie sull'orario di arrivo previsto

- La data e il luogo di emissione del contratto;
- I punti di partenza e di destinazione;

- Le fermate previste, fermo restando il diritto del vettore di stabilire che può modificarle, se necessario e senza tali modifiche, facendo perdere al trasporto il suo carattere internazionale;
- La data e l'ora prevista per la partenza e/o il ritorno;
- Le referenze del mittente;
- I riferimenti dello spedizioniere alla partenza e alla destinazione;
- Le referenze del vettore;
- Le referenze del destinatario;
- Riferimenti della compagnia aerea di movimentazione merci alla partenza e alla destinazione;
- La natura dei beni;
- Il numero, il metodo di imballaggio, i segni particolari o i numeri dei colli;
- Il peso, la quantità, il volume o le dimensioni della merce;
- Lo stato apparente della merce e dell'imballaggio;
- Il costo del trasporto se previsto, la data e il luogo di pagamento e la persona che deve pagare;
- Dettagli su tasse e tariffe;
- Se la spedizione è in contrassegno, il prezzo della merce e, se applicabile, l'importo delle spese;
- L'importo del valore dichiarato;
- Il numero di copie della lettera di vettura aerea;
- I documenti trasmessi al vettore per accompagnare la lettera di trasporto aereo;
- Il limite di tempo per il trasporto e una breve indicazione del percorso da seguire, se previsto;
- L'indicazione che il trasporto è soggetto al regime di responsabilità stabilito dalla Convenzione di Varsavia;
- Informazioni legali doganali e di sicurezza aerea a seconda del tipo di merce;
- Il profilo dello spedizioniere e degli eventuali spedizionieri.

5.2 Analisi dei requisiti

Dopo aver presentato nel precedente capitolo il processo del magazzino, si procede con l'analisi dei requisiti e progettazione concettuale iniziale dei datamart.

Prima di entrare nei dettagli della progettazione verrà fornita una panoramica sulle caratteristiche e sugli aspetti architetturali di un Datawarehouse.

5.2.1 Key Performance Indicator (KPI)

KPI Definition

Il primo punto è di listare le KPI da implementare nella soluzione. Questi KPI sono raggruppate per dipartimento. E ripartiti in questo modo:

- Finance
- HR
- Operazioni
- HSSE

I requisiti del progetto prevedono le implementazioni di 43 KPI; In un primo tempo daremo una definizione dei KPI da implementare. Sono 11, Prima di definire i KPI definiremmo il processo de business CARGO.

HSSE					
1	Occupational Safety	CA/GH	23	% Number of Flights Delay SEVERITY 1 over Total Operated Flights	GH
2	Lost time injuries /200k hours worked (WFS staff only, all hours worked)	CA/GH	24	% Number of Flights Delay SEVERITY 2 over Total Operated Flights	GH
3	Work days lost (absence) due to work related injury	CA/GH	25	% Flights First Baggage Delivery On Target (25 after ATA)	GH
4	Aircraft damage	GH	26	% Flights Last Baggage Delivery On Target (XX after ATA)	GH
5	Flight Safety	CA/GH	27	Lost Baggage Index Lost Bags/Departed Passengers	GH
6	Aviation Security	CA/GH	Operations		
7	Asset Protection	CA/GH	28	Trend - Number of Turnarounds	GH
8	Equipment Damage	CA/GH	29	Trend - Number of Handled Tons	CA
9	Environment	CA/GH	30	Kilos per Hour	CA
10	Number of HSSE 'self-inspections' conducted	CA/GH	31	AWB per Hour	CA
11	Number of Near-Miss reports made	CA/GH	32	Paid Hours / Turns	GH
12	Regulatory Non-Compliance	CA/GH	33	% Overtime	CA/GH
13	Internal Non- Compliance	CA/GH	34	% of Payroll, Third Party and Overtime over Total Paid Hours	CA/GH
People			35	Cost per Paid Hour (Payroll + Overtime + Third Party)	CA/GH
14	Staff turnover rate	CA/GH	36	Manpower Costs per Paid Hour - WFS (Payroll + Overtime)	CA/GH
15	% Sickness over Total Payroll Paid Hours	CA/GH	37	Manpower Costs per Paid Hour – Third Party	CA/GH
Customer satisfaction			38	% Occupancy Warehouse	CA
16	% of DEP sent On-time	CA	Financial Performance / Shareholder		
17	% of RCS sent On-time	CA	39	Yield (Revenue per Kilo / Turnaround)	CA/GH
18	% of MAN sent On-time	CA	40	Labour Costs per Kilo / Turnaround	CA/GH
19	% of FAP (Flown as planned)	CA	41	Indirect Costs per Kilo / Turnaround	CA/GH
20	% of NFD sent On-time	CA	42	Total Costs per Kilo / Turnaround	CA/GH
21	% of RCF sent On-time	CA	43	Warehouse rental cost / sqm	CA
22	% Number of Flights Delay over Total Operated Flights	GH	44	Contribution per Kilo / Turnaround	CA/GH

Figure 5.9:44 PKI per entità e per business line

5.2.2 Qualche linea concetti linee guide dell'attività cargo

In questa sessione presenteremo qualche concetto chiavi necessario per le comprensioni dell'attività cargo svolto nel magazzino. Questi concetti permettono di capire e misurare le attività svolte nel magazzino.

Activity Driver Cargo Business

Metric	Definition
Loose Tons Import	Checked-in Import Tons not pre-palletized to be delivered locally (destination same station).
Loose Tons Export	Manifested Export Tons not pre-palletized with origin same station.
Loose Tons Transit Import	Checked-in Import Tons not pre-palletized with destination different than same station.
Loose Tons Transit Export	Manifested Export Tons not pre-palletized with origin different than same station.
Nr. Of AWB Import	Number of AWB handled in Regular Import Process.
Nr. Of AWB Export	Number of AWB handled in Regular Export Process.
Nr. Of AWB Import Transit	Number of AWB handled in Transit (Import).
Nr. Of AWB Export Transit	Number of AWB handled in Transit (Export)
BUP Tons Import	Checked-in Import Tons pre-palletized and not broken down to be delivered locally (destination same station).
BUP Tons Export	Manifested Export Tons pre-palletized and not built-up with origin same station.
BUP Tons Import Transit	Checked-in Import Tons pre-palletized and not broken down with destination different than same station.
BUP Tons Export Transit	Manifested Export Tons pre-palletized and not built-up with origin different than same station.

Figure 5.10:Misure delle attività Cargo

Profilo CARGO

Profile Type	Definition
Warehouse	<p>Employees participating in the production process in the warehouse including 3rd party labour.</p> <p>Tasks included: Breaking down, building up and shifting shipment at the warehouse acceptance; cargo physical acceptance and checks, weight and volume checks, transport of goods to the storage location, taking goods out of storage location, ULD build up and loading of belly carts, ULD break-down and unloading of belly carts, Loading and unloading of trucks, BUP staging, pre-flight staging, ULD preparation, tagging and weighing ULD's, ULD control.</p> <p>Cargo screening and any other security checks like ETD checks, physical checks are also included.</p> <p>Shift-Leaders, Duty Managers, Team Leaders, Supervisors are included (assuming they are directly and actively participating in the production by e.g. coordinating work).</p> <p>Station Managers, Ops Manager are excluded (assuming they are not actively participating in the production process directly.)</p>
Office	<p>Employees in the office with main activities related with all kind of documentation handling, basically AWB data capture, Manifesting, Customs documentation, cargo documentation acceptance and check, DGR and other special cargo documents checks, Delivery notes, T1 and ADR, NOTOC issuance, ULD control, Document scanning or making photocopies related with flight documents.</p>
Ramp	<p>Ramp employees dedicated to aircraft loading/off loading and ramp transportation.</p> <p>The entities performing transportation from and to the Ramp, should record these employees under ramp employees.</p>
Support	<p>Managers of different sections at station level.</p> <p>Administrative and Payroll people also included (Station Manager, Admn and HR, Operational and Quality, Others).</p> <p>Plus employees not actively participating in the production process (no physical work nor document processing nor direct production coordination). Planning, rostering and any other supporting activities not driven by the volumes + Weight and Balance, for units with trucking they should be reported under Support employees.</p>

Figure 5.11: Profilo attività Cargo

Cost Drivers Item

Item	Metric	Definition
Productive Hours	Payroll Paid	Payroll paid hours of our own employees as per WFS contract, excluding Overtime. Includes part time contracts and also salary paid and hourly paid. Holidays worked and paid should be included in this category.
	Overtime	Paid overtime hours of our own employees with a WFS contract. This metric would include the Overtime, the Double time as well as night shifts when special rate.
	Temporary Labor	Hours worked by subcontracted 3 rd Party staff (agencies). By definition, Subcontracted does not have Absence.
Non-Productive Hours	Sickness	Paid absent hours of own staff but not attended due to sickness.
	Vacations	Paid absent hours of own staff but not attended due to vacations. Vacations are the legally authorized XX days off.
	Long term sickness	Paid absent hours of own staff but not attended due to long term sickness paid by WFS .
	Other	Paid absent hours of own staff but not attended due to other absences scheduled by the Station (training, meetings, etc.).
Calculations	Payroll Working Hours	Difference of Payroll Paid Hours and Absence (of any kind). Number of Hours with physical presence in our premises, performing ops tasks of our own employees as per WFS contract. Payroll Worked hours are effective and "productive" worked and paid hours of our employees <ul style="list-style-type: none"> - Paid hours for meal breaks : non-productive hours => are excluded - Paid hours but "non-productive" because of late flight arrival => NOT excluded
	Total Working Hours	Payroll Working Hours + Overtime + Temporary Labor
	Total Paid Hours	Payroll Paid Hours + Overtime + Temporary Labor

Tableau 5.1:Induttore di costo per l'attività cargo

5.2.3 Identificazioni dei KPI

Nella tabella seguente verremmo la definizione del KPI Operazione lo spesso processo sera svolto per gli altri dipartimenti o gruppi di KPI

Tendenza - Numero di tonnellate gestite

Codice	Nome tecnica	Elementi di calcolo	Business rule
	#KPi29	▪ Tonnellate	<p>Tonnellate gestite (misurate come Tonnellate manifestate e controllate) compresi i transiti.</p> <p>I transiti sono contati in entrambe le tratte, ovvero quando i transiti sono manifestati e a destinazione quando i transiti sono stati registrati.</p> <p>Tonnellate totali gestite per l'importazione e l'esportazione (compresi i transiti). Le Tonnellate manifestate e ricevute (Check-In) sono il riferimento, supportato da un FFM.</p>

Tableau 5.2:Tendenza - Numero di tonnellate gestite

Chilogrammi per ora

Codice	Nome tecnica	Elementi di calcolo	Business rule
	#KPi30	<ul style="list-style-type: none"> ▪ Tonnellate ▪ Ore Produttive 	Chili di carico movimentato (inclusi BUP e Loose) diviso per il totale delle ore di lavoro come paga pagata + lavoro straordinario + ore di lavoro temporaneo - Assenza Altri - Assenza Malattia - Assenza a lungo termine - Vacanze. Ore e assenze legata esclusivamente al Profilo "Magazzino" come definito nel presente documento (vedi "Def Shared e Cargo specifico")

Tableau 5.3:KPi30 Chilogrammi per ora

Air waybill per ora

Codice	Nome tecnica	Elementi di calcolo	Business rule
	#KPi31	<ul style="list-style-type: none"> ▪ AWB ▪ Ore Produttive 	Numero di AWB documentati (compresi BUP e Loose) diviso per il totale delle ore di lavoro come paga pagata + ore di lavoro straordinario + ore di lavoro temporaneo - Assenza Altri - Assenza Malattia da assenteismo - Assenza a lungo termine - Vacanze. Ore e assenze riferite esclusivamente al profilo "Ufficio" come definito nel presente documento. (vedi "Def Shared e Cargo specifico")

Tableau 5.4:KPI i30 AWB per Ora

Ore pagate / Turni

Codice	Nome tecnica	Elementi di calcolo	Business rule
	#KPi32	<ul style="list-style-type: none"> ▪ Turnaround ▪ Orari pagati (Retribuzioni e Terzi) 	Numero di turnaround diviso per il totale delle ore pagate, compresi gli stipendi pagati + gli straordinari + le ore di lavoro temporaneo.

Tableau 5.5:Ore pagate / Turni

% di paghe su totale ore pagate

Codice	Nome tecnico	Elementi di calcolo	Business rule
	#KPi34	<ul style="list-style-type: none"> ▪ Ore di lavoro straordinario ▪ Orari di lavoro ▪ Orari di terzi 	<p>Tre tariffe contemporaneamente.</p> <p>% del libro paga sul totale delle ore pagate (libro paga + lavoro straordinario + terzi)</p> <p>% di ore di lavoro straordinario sul totale delle ore pagate (libro paga + lavoro straordinario + terzi)</p> <p>% di Terzi sul totale delle ore pagate (Retribuzioni + Lavoro straordinario + Terzi)</p> <p>Fornire un riferimento del nostro livello di flessibilità</p>

Tableau 5.6: % di paghe su totale ore pagate

Costo per ora pagata (libro paga + lavoro straordinario + terzi)

Codice	Nome tecnico	Elementi di calcolo	Business rule
	#KPi35	<ul style="list-style-type: none"> ▪ Totale costi salariali ▪ Totale ore pagate 	<p>Tutti i conti di stipendio, comprese le assicurazioni e la previdenza sociale più i costi di terzi e gli straordinari, divisi per il totale delle ore pagate (salari, terzi e straordinari).</p> <p>Questo tasso dovrebbe includere sia le spese generali di campo che le spese generali di campo (per definizione, non abbiamo spese generali a livello di stazione, ma attività di supporto come definite in "Def Shared").</p>

Tableau 5.7: Costo per ora pagata (libro paga + lavoro straordinario + terzi)

% Magazzino di lavoro

Code	Nome tecnico	Elementi di calcolo	Business rule
	#KPi38	<ul style="list-style-type: none"> ▪ Tonnellate gestite ▪ Capacità totale del magazzino 	<p>Tonnellate totali gestite (importazione, esportazione e transito) divise per capacità di magazzino mensile.</p> <p>La capacità totale deve essere riportata come capacità operativa e non come standard IATA.</p>

Tableau 5.8: % Magazzino di lavoro

5.3 Processo di Data Warehousing

La prima, e sicuramente più diffusa, definizione di Datawarehouse ci viene fornita da [9] “un datawarehouse è una raccolta di dati organizzata per soggetti, integrata, non volatile, e variabile nel tempo, di supporto ai processi decisionali”.

- **Organizzata per soggetti:** nei sistemi di supporto alle decisioni i dati sono organizzati per analizzare dei soggetti di interesse. Il focus si sposta dalle transazioni, caratteristica dei sistemi operazionali, all’analisi di una specifica applicazione o funzione aziendale.
- **Integrata:** la caratteristica distintiva principale di un datawarehouse è l’integrazione di dati provenienti da fonti eterogenee. L’integrazione avviene attraverso un processo di estrazione, trasformazione e caricamento dei dati dalle sorgenti al datawarehouse.
- **Non volatile:** I dati sono caratterizzati da accessi in sola lettura e non sono soggetti a modifiche; per questo motivo mantengono la loro integrità nel tempo. Ciò comporta un notevole vantaggio anche in fase di progettazione in quanto non deve essere gestita la concorrenza tra l’aggiornamento dei dati e gli accessi in lettura.
- **Variabile nel tempo:** mentre i database operazionali conservano solo i dati più recenti, un datawarehouse è pensato per fornire supporto alle decisioni strategiche, dunque necessita di un orizzonte temporale più ampio al fine di poter analizzare i cambiamenti nel tempo.
- **Supporto alle decisioni:** lo scopo principale è di supportare il processo decisionale inteso come capacità di rispondere rapidamente alle business questione dei dirigenti aziendali.

Come già evidenziato precedentemente, un datawarehouse è una base di dati organizzata per soggetti; ciascun soggetto di interesse può essere considerato, a sua volta, come un datawarehouse di più piccole dimensioni, un sottoinsieme della base di dati di partenza.

È il concetto di *Data Mart*.

Un data mart è un sottoinsieme dei dati già presenti nel datawarehouse e viene generalmente progettato per soddisfare le esigenze di una specifica funzione di un business. Il datawarehouse rappresenta perciò l'intero sistema informativo dell'azienda, mentre le informazioni presenti all'interno di un singolo data mart sono strettamente legate all'area aziendale di interesse.

Come mostrato in Figura 5.9 e 5.10, la progettazione di un datawarehouse può avvenire attraverso due diversi approcci: top-down oppure bottom-up.

L'approccio top-down, introdotto da Inmon [9], consiste nel progettare l'intero datawarehouse e, solo successivamente, estrarre da esso i singoli data mart. Questa soluzione si basa su una visione globale dell'obiettivo e garantisce la realizzazione di un datawarehouse consistente e ben integrato. Gli svantaggi sono legati all'impossibilità di prevedere nel dettaglio tutte le esigenze che potrebbero sorgere nelle diverse aree aziendali e ai lunghi tempi di realizzazione. Tale soluzione è anche conosciuta nella letteratura come data-driven, cioè guidata dai dati.

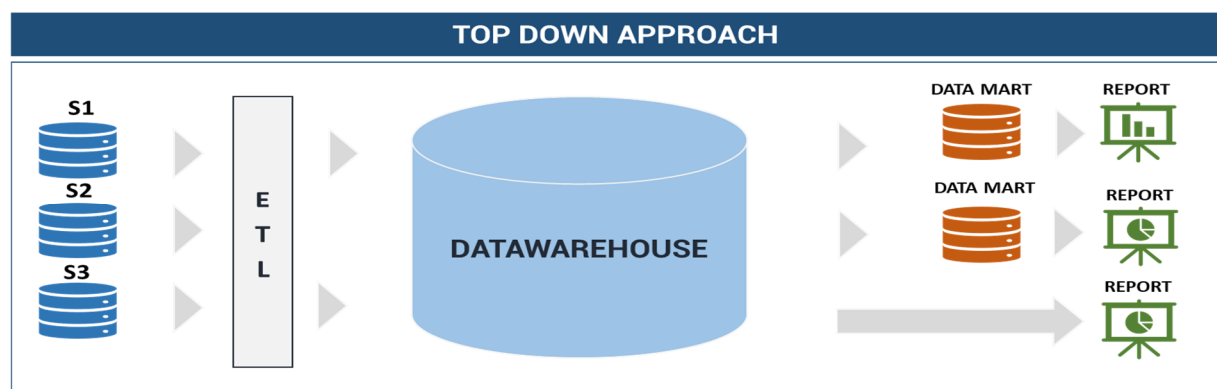


Tableau 5.9:L'approccio top-down

L'approccio bottom-up, invece, introdotto da Kimball [10], è un approccio di tipo incrementale: il datawarehouse viene costruito assemblando iterativamente i diversi data mart. Questa soluzione permette, in fase di progettazione, di spostare il focus sulla singola area di business e quindi di realizzare in tempi brevi i singoli data mart, ma presenta lo svantaggio di essere poco mantenibile. Tale approccio è anche conosciuto nella letteratura come approccio analysis-driven, cioè guidato dalle analisi.

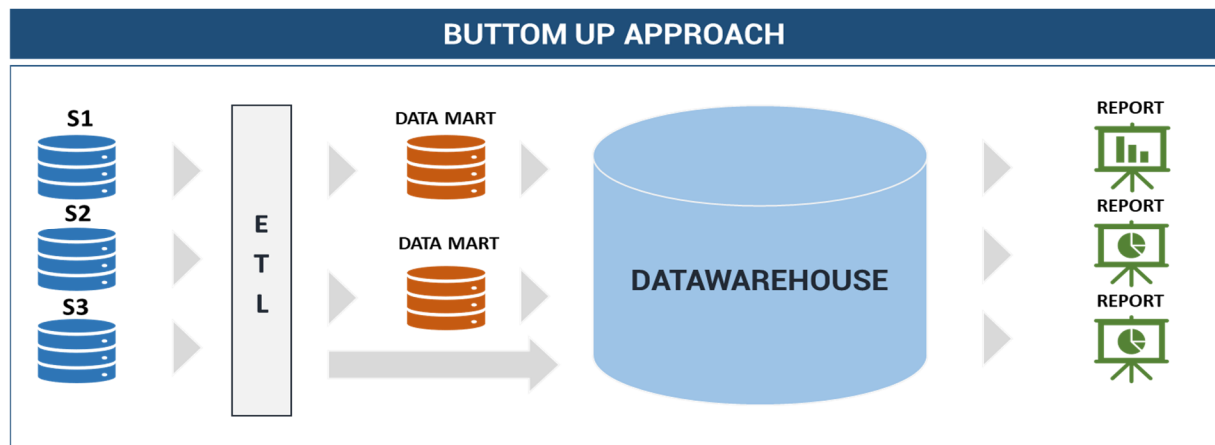


Tableau 5.10:L'approccio bottom-up

Nel caso in esame si è scelto di adottare la strategia bottom-up, partendo quindi dalla progettazione dei singoli data mart per arrivare alla struttura completa e finale del datawarehouse.

Scelta la modalità con la quale si è deciso di affrontare il problema, si procede con la progettazione vera e propria.

Seguendo il metodo riportato in Albano [11], la progettazione di un datawarehouse è un'attività complessa organizzata nelle seguenti fasi:

Analisi dei requisiti. L'obiettivo è di raccogliere e definire i requisiti di analisi; la loro individuazione avviene attraverso le interview, interviste ai committenti con lo scopo di comprendere a fondo il processo di business e interpretare le esigenze analitiche per arrivare a creare modelli di dati e strumenti di presentazione efficaci e coerenti. La fase di analisi dei requisiti si suddivide in due sottofasi principali: la raccolta e la specifica dei requisiti.

- Progettazione concettuale. Si definisce un modello concettuale dei dati da analizzare; in questa fase sono descritti i fatti, le dimensioni e gli attributi dimensionali.
- Progettazione logica. Si trasforma il modello concettuale nelle strutture logiche da adoperare per la gestione del datawarehouse attraverso un DBMS11 relazionale.

- Progettazione fisica. Si definiscono le strutture dati necessarie (indici e viste materializzate) per memorizzare le tabelle definite durante la progettazione logica e agevolare le operazioni di analisi.

Nei paragrafi successivi verranno trattate nel dettaglio la fase di analisi dei requisiti e la progettazione concettuale iniziale dei data mart. Saranno introdotti i concetti di fatto, misura e dimensione e descritte le metriche e gli indicatori chiave di prestazione dei processi aziendali.

Durante la fase di analisi dei requisiti si raccolgono e definiscono con i committenti i requisiti delle analisi di supporto alle decisioni che si desiderano eseguire. Il risultato è una serie di documenti formali che andranno a evidenziare i primi aspetti essenziali della progettazione, ovvero la base per poter individuare i fatti, le dimensioni e le misure coinvolte.

Prima di procedere con la specifica dei requisiti e quindi presentare una prima bozza di modello concettuale, si presenta il risultato delle interviste ai committenti.

Nella Tabella 4.1 sono descritti i requisiti di analisi; saranno omesse per semplicità tutte le misure di base già descritte nel Capitolo 3.

Operations		
28	Trend - Number of Turnarounds	GH
29	Trend - Number of Handled Tons	CA
30	Kilos per Hour	CA
31	AWB per Hour	CA
32	Paid Hours / Turns	GH
33	% Overtime	CA/GH
34	% of Payroll, Third Party and Overtime over Total Paid Hours	CA/GH
35	Cost per Paid Hour (Payroll + Overtime + Third Party)	CA/GH
36	Manpower Costs per Paid Hour - WFS (Payroll + Overtime)	CA/GH
37	Manpower Costs per Paid Hour – Third Party	CA/GH
38	% Occupancy Warehouse	CA

Tableau 5.11: Metriche della business line Cargo

Financial Performance / Shareholder		
39	Yield (Revenue per Kilo / Turnaround)	CA/GH
40	Labour Costs per Kilo / Turnaround	CA/GH
41	Indirect Costs per Kilo / Turnaround	CA/GH
42	Total Costs per Kilo / Turnaround	CA/GH
43	Warehouse rental cost / sqm	CA
44	Contribution per Kilo / Turnaround	CA/GH

Tableau 5.12:Metriche dell'entità Finanze

5.3.1 Specifica dei requisiti e progettazione iniziale dei data mart

In questa sezione si introducono i concetti di fatto, misura e dimensione, necessari per procedere con le fasi di progettazione. Successivamente, per ciascun data mart individuato, saranno fornite le specifiche dei requisiti del fatto, le dimensioni coinvolte e le misure interessate. Per concludere, si presenteranno i relativi schemi concettuali iniziali.

Nel processo di data warehousing, si parla di **fatto** perché il management di un'organizzazione ragiona in termini di collezioni di dati che riguardano particolari funzioni o processi aziendali. Ogni fatto è caratterizzato da un insieme di attributi numerici, le misure, che riguardano il comportamento di un fenomeno aziendale.

In base al tipo di funzione di aggregazione¹³ che è possibile associare, una misura può essere:

- **Additiva:** se ha senso effettuarne la somma per ogni dimensione.
- **Semi additiva:** se non può essere sommata per alcune dimensioni, tipicamente quella temporale.
- **Non additiva:** se non ha senso effettuarne la somma. Il valore di una misura non additiva è definito applicando altre funzioni di aggregazione o come rapporto di misure additive.

Nella terminologia del data warehousing l'intervallo di tempo in cui ogni misura è considerata prende il nome di granularità o grana. La granularità descrive il livello di dettaglio, o di sintesi, dei dati raccolti. La scelta del giusto livello da adottare è uno step critico da affrontare durante la progettazione iniziale del datawarehouse. Il numero di

record da elaborare e, di conseguenza, i requisiti tecnologici di sistema aumentano esponenzialmente al crescere del livello di dettaglio. Dietro il concetto di granularità risiede, infatti, un'esigenza estremamente pratica: l'economicità delle elaborazioni legate alle analisi.

I manager ragionano in modo multidimensionale, ovvero sono interessati ad analizzare le misure dei fatti secondo prospettive diverse di analisi, **le dimensioni**.

Le dimensioni sono quindi utilizzate per contestualizzare le misure dei fatti e per applicare filtri e raggruppamenti di dati;

L'ultima considerazione riguarda il tipo di strategia da adottare in caso di dimensioni con attributi che possono cambiare nel tempo. Come riportato in Kimbal [10], si considerano quattro strategie principali:

- **Tipo 1 (Riscrittura della storia).** È la soluzione più semplice e comporta la perdita della storia: il valore di un attributo dimensionale che cambia deve essere sostituito con il nuovo valore.
- **Tipo 2 (Aggiunta di un nuovo record).** È la soluzione più comune che permette di mantenere la storia dei valori. Si modifica la struttura della dimensione aggiungendo due campi **START_DATE** e **END_DATE** che indicano l'intervallo di validità di ciascun record; il campo **END_DATE** è inizialmente settato a null. Nel momento in cui è necessario aggiornare il valore di un attributo vengono eseguite le seguenti operazioni:
 - Il campo **END_DATE** è aggiornato con la data in cui è avvenuta la modifica.
 - Si crea un nuovo record con la stessa chiave naturale del precedente.
 - Si imposta il campo **START_DATE** del nuovo record con la data in cui è avvenuta la modifica.
 - Il campo **END_DATE** del nuovo record viene settato a null.
- **Tipo 3 (Aggiunta di uno o più campi).** È la soluzione meno utilizzata in quanto consiste nell'aggiungere tanti campi quante sono le versioni dell'attributo dimensionale che si vuole mantenere. Il vantaggio di questa soluzione risiede nell'avere accesso immediato alla storia dei valori; lo svantaggio principale è che

deve essere previsto, in fase di progettazione, il numero massimo di versioni da mantenere.

- **Tipo 4.** Si fa uso di una tabella aggiuntiva per tenere traccia della storia dei valori degli attributi dimensionali. La struttura della tabella è del tutto analoga alla soluzione di Tipo

Capitolo 6

Progettazione iniziale del datawarehouse

Il datawarehouse è stato progettato prendendo in conto le KPI espressi come bisogno. E per questo che il concepimento delle tabelle dei fatti saranno fatto sulla base dei gruppi di KPI (Operazione, Finanza, Resource Umane, HSSE e qualità). Noi ci focalizzeremo per questa tesi su:

- Cargo Operazione
- Risorse umane
- Finanza

E provvederemo per ciascuna delle entità una specifica della tabella dei fatti, le misure e in fine lo schema concettuale.

L'obiettivo di questa sessione è quello di costruire i diversi data mart in funzione delle entità Cargo Operation, finanza e delle risorse umane del Cargo business line. Questo segue l'approccio incrementale raccomandato da KIMBALL, quindi avremo per ogni entità data mart che raccolgono in alcuni casi, diverse tabelle di fatto e le relative dimensioni.

6.1 Data mart Cargo operation

Il data mart delle operazioni di cargo business sarà costituito dopo l'analisi di quattro tabelle di fatti, ovvero:

- **F_OPERATION_D:** In questa tabella dei fatti saranno memorizzate le misurazioni sull'attività operativa in tonnellaggio, con una frequenza giornaliera.
- **F_OPERATION_BUD_M:** in questa tabella di fatto saranno memorizzate le misure sull'attività operativa in tonnellaggio in previsione, con frequenza mensile.
- **F_WORFORCE_D:** in questa tabella di fatto saranno memorizzate con frequenza giornaliera le misure sulle risorse necessarie in termine di ore lavorate per l'esecuzione delle operazioni.

- **F_WORKFORCE_BUD_M**: in questa tabella di fatto saranno memorizzate le misure sulla proiezione mensile delle risorse necessarie in termine di ore lavorate per l'esecuzione dei compiti operativi.

Il data mart di Cargo Operation sarà costruito intorno a queste tabelle di fatto che specificheremo nei paragrafi seguenti.

6.1.1 Specifica dei requisiti del fatto F_OPERATION_D

Di seguito si descrive il fatto **F_OPERATION_D**, specificando, nella *Tabella 6.1*, il suo significato, le dimensioni e le misure preliminari interessate. La granularità del fatto **F_OPERATION_D** determina la dimensione del data mart e il tipo di analisi che si può effettuare sui dati.

Descrizione	Dimensioni preliminari	Misure preliminari
Una riga del fatto corrisponde all'insieme delle misure registrate durante les operazioni nel magazzino.	Cliente, Giorno, Handling Delivery, Handling Process, Handling Product, Unit,	Pieces, Tons, AWBS (Air Waybill/Lettera di trasporto)

Tableau 6.1: Specifica dei requisiti del fatto Operazione giornaliero

Dimensione

In *Tabella 6.2* si descrivono le dimensioni specificando per ognuna di esse il nome, una breve descrizione, quali sono gli attributi modificabili nel tempo e il tipo di trattamento da utilizzare per le modifiche. Per i data mart *Cargo Operation* avremmo sei dimensione associate, con granularità mensile per la dimensione **MESE**.

Nome	Descrizione	Granularità	Attributi modificabili	Trattamento modifiche
CUSTOMER	Dimensione contenente il nome et codifica di tutti i Clienti, compagnie aeree	La singola compagnie aerea	CD_UNIT	Tipo_2
MESE	Dimensione temporale che identifica il verificarsi di un fatto	Mensile	No	
HANDLING DELIVERY	Gestione de la consegna spedizione nel processo IMPORT, la spedizione puo essere in arrivo il locale o in transit	La singola destinazione della spedizione	No	
HANDLING PROCESS	Processo di movimentazione spedizione nel magazzino, Processo IMPORT or EXPORT	Il singolo tipo di processo	No	
HANDLING PRODUCT	Trattamento di tipo spedizione nelle magazzino, Mail AWB, BUP,...	Il singolo tipo di spedizione da gestire	No	
ORGANIZATION UNIT	Unita organizzativa sparsa nelle regioni	Il singolo unita organizzativa	ID_UNIT	Tipo_2

Tableau 6.2: Dimensione associato al fatto Operazione giornaliera

Misure

Si presenta la tabella riepilogativa delle misure per il fatto **F_OPERATION_D**. Tali misure sono fondamentali per il calcolo delle metriche e dei KPI presentati nel Paragrafo 5.2.3 Per ciascuna di esse si specifica il nome, il tipo di aggregazione e se il valore deriva o memo dalla combinazione di due o più misure. Avremmo tre misure fondamentale PIECES, TONS a AWS nel business line Gargo e nel l'entità Cargo operation.

Misura	Aggregabilità	Derivata
PIECES	Addivita	No
TONS	Addivita	No
AWS	Addivita	No

Tableau 6.3: Misure della tabella del fatto Operazione giornaliero

Schema concettuale

La figura mostra lo schema concettuale definito a partire dalle informazioni prodotto durante l'analisi dei requisiti.

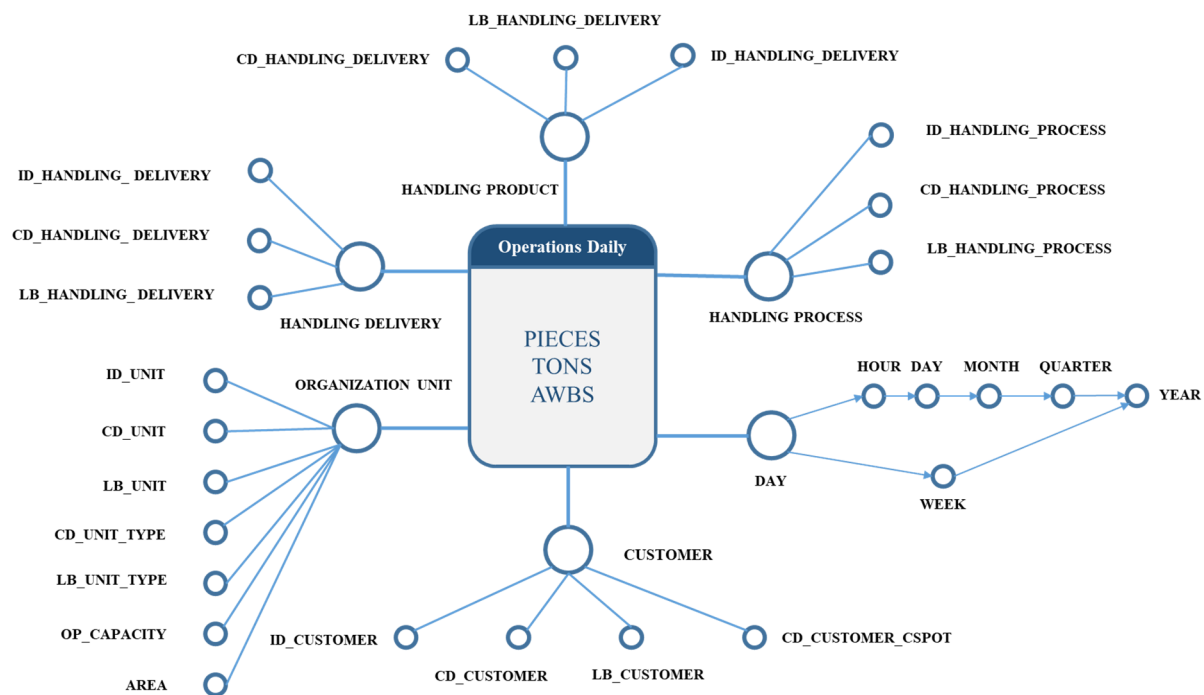


Figure 6.1: Schema concettuale del data mart operazione giornaliero

Lo schema riporta le misure all'interno delle tabelle del fatto, dal quale originano i collegamenti verso le dimensioni e i relativi attributi; si noti come gli attributi Hour, Day, Month, Quarter e Year della dimensione Day fanno parte di una gerarchia bilanciata con quattro livelli. Per definizione, In una gerarchia bilanciata tutti i rami scendono fino allo stesso livello di dettaglio e il membro padre logico di ogni membro è situato al livello immediatamente superiore rispetto al membro.

6.1.2 Specifica del fatto **F_OPERATION_BUD_M**

Raccoglie informazioni sulla proiezione mensile dei dati operativi di Cargo Operation, in opposizione al data mart **OPERATION_D** dove le informazioni operative sono raccolte su base giornaliera.

Di seguito si descrive il fatto Operazionale **F_OPERATION_BUD_M**, specificando, nella **Tabella 6.2**, il suo significato, le dimensioni e le misure preliminari interessate. La granularità del fatto determina la dimensione del data mart e il tipo di analisi che si può effettuare sui dati.

Descrizione	Dimensioni preliminari	Misure preliminari
Una riga del fatto corrisponde all'insieme delle misure registrate durante les operazioni nel magazzino.	Cliente, Mese, Handling Delivery, Handling Process, Handling Product, Unit,	Pieces, Tons, AWBS (Air Waybill/Lettera di trasporto)

Tableau 6.4: Specifica dei requisiti del fatto Operazione giornaliero

Dimensione

In **Tabella 6.5** si descrivono le dimensioni specificando per ognuna di esse il nome, una breve descrizione, quali sono gli attributi modificabili nel tempo e il tipo di trattamento da utilizzare per le modifiche.

Nome	Descrizione	Granularità	Attributi modificabili	Trattamento modifiche
CUSTOMER	Dimensione contenente il nome et codifica di tutti i Clienti, compagnie aeree	La singola compagnie aerea	CD_UNIT	Tipo_2
MESE	Dimensione temporale che identifica il verificarsi di un fatto	Mensile	No	
HANDLING DELIVERY	Gestione de la consegna spedizione nel processo IMPORT, la spedizione puo essere in arrivo il locale o in transit	La singola destinazione della spedizione	No	
HANDLING PROCESS	Processo di movimentazione spedizione nel magazzino, Processo IMPORT or EXPORT	Il singolo tipo di processo	No	
HANDLING PRODUCT	Trattamento di tipo spedizione nelle magazzino, Mail AWB, BUP,...	Il singolo tipo di spedizione da gestire	No	
ORGANIZATION UNIT	Unita organizzativa sparsa nelle regioni	Il singolo unita organizzativa	ID_UNIT	Tipo_2

Tableau 6.5: Dimension del data mart Cargo Operation Mensile

Misure

Si presenta la tabella riepilogativa delle misure per il fatto **F_OPERATION_BUD_M**. Tali misure sono fondamentali per il calcolo delle metriche e dei KPI presentati nel Paragrafo 5.2.3

Per ciascuna di esse si specifica il nome, il tipo di aggregazione e se il valore deriva o memo dalla combinazione di due o più misure. (cfr. Tabella 6.6).

Misura	Aggregabilità	Derivata
PIECES_Budget	Additiva	No
TONS_Budget	Additiva	No
AWS_Budget	Additiva	No

Tableau 6.6:Misure del data Mart Cargo Operation Mensile

Schema concettuale

La figura mostra lo schema concettuale definito a partire dalle informazioni prodotto durante l'analisi dei requisiti.

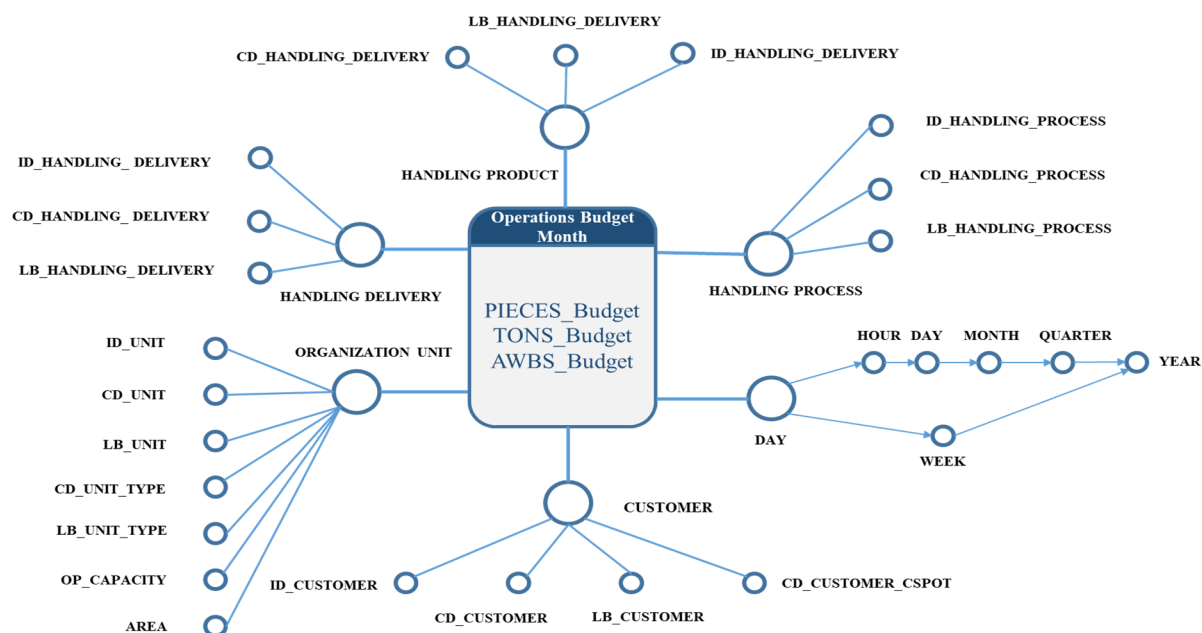


Figure 6.2:Schemas concettuale del data mart operation mensile

Lo schema riporta le misure all'interno delle tabelle del fatto, dal quale originano i collegamenti verso le dimensioni e i relativi attributi; si noti come gli attributi Hour, Day, Month, Quarter e Year della dimensione Day fanno parte di una gerarchia bilanciata con quattro livelli.

6.1.3 Specifica del fatto F_WORKFORCE_BUD_M

La tabella **F_WORKFORCE_BUD** raccoglie le misure sulle ore di lavoro previste per un periodo mensile necessario per svolgere le attività nei magazzini.

Specifica del fatto

Di seguito si descrive il fatto **F_WORKFORCE_BUD_M**, specificando, nella Tabella 6.7, il suo significato, le dimensioni e le misure preliminari interessate. La granularità del fatto determina la dimensione del data mart e il tipo di analisi che si può effettuare sui dati.

Descrizione	Dimensioni preliminari	Misure preliminari
Ore necessario previsto a bilancio per il personale o mano d'opera del magazzino e ufficio per svolgere l'attività di handling per un mese	Dipartimento, Mese, Tipologie di ore, Unità organizzativa.	Ore Previsto a bilancio

Tableau 6.7: Specifica del data Mart Workforce BUD M

Dimensione

In Tabella 6.8 si descrivono le dimensioni specificando per ognuna di esse il nome, una breve descrizione, quali sono gli attributi modificabili nel tempo e il tipo di trattamento da utilizzare per le modifiche.

Nome	Descrizione	Granularità	Attributi modificabili	Trattamento modifiche
DEPARTMENT	Nome del reparto dove si svolge l'attività	Il singolo dipartimento	No	-
MONTH	Dimensione temporale che identifica il verificarsi di un fatto	Mensile	No	-
HOURLY TYPE	Tipo di ore da pagare, straordinari, vacanze,...	Orario	No	-
ORGANIZATION UNIT	Unità organizzativa sparsa nelle regioni	Il singolo unità organizzativa	ID_UNIT	Tipo_2

Tableau 6.8: Tabella della dimensione del data mart Workforce Mensile

Misure

Si presenta la tabella riepilogativa delle misure per il fatto **F_WORKFORCE_BUD_M**. Tali misure sono fondamentali per il calcolo delle metriche e dei KPI presentati nel Paragrafo 5.2.3.

Per ciascuna di esse si specifica il nome, il tipo di aggregazione e se il valore deriva o meno dalla combinazione di due o più misure. (cfr. Tabella 6.9)

Misura	Aggregabilità	Derivata
HOURS_BUD	Additiva	No

Tableau 6.9:Tabella Misure Data Mart Workforce Mensile

Schema concettuale

La figura mostra lo schema concettuale definito a partire dalle informazioni prodotto durante l'analisi dei requisiti.

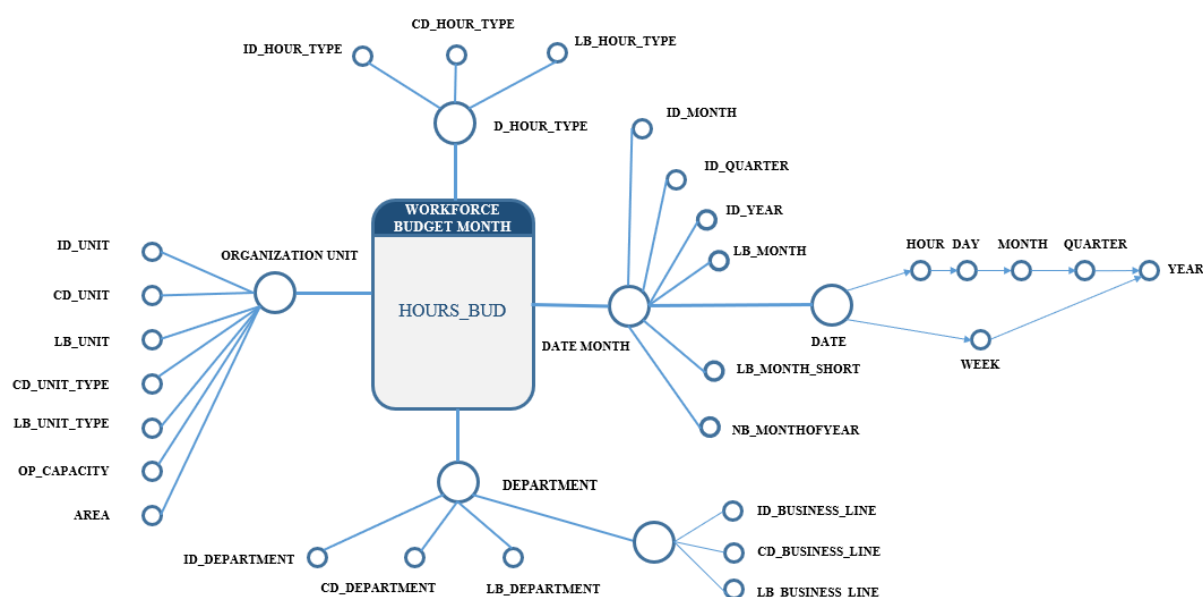


Tableau 6.10:Schemas concettuale Workforce Mensile

Lo schema riporta le misure all'interno delle tabelle del fatto, dal quale originano i collegamenti verso le dimensioni e i relativi attributi; si noti come gli attributi Hour, Day, Month, Quarter e Year della dimensione Day fanno parte di una gerarchia bilanciata con quattro livelli.

6.1.4 Specifica del fatto F_WORKFORCE_D

La tabella **F_WORKFORCE_D** raccoglie le misure sulle ore di lavoro effettuato per un dato giorno la misura e l'ora.

Specifica del fatto

Di seguito si descrive il fatto **F_WORKFORCE_D**, specificando, nella Tabella 6.11, il suo significato, le dimensioni e le misure preliminari interessate. La granularità del

fatto determina la dimensione del data mart e il tipo di analisi che si può effettuare sui dati.

Descrizione	Dimensioni preliminari	Misure preliminari
Ore necessario per il personale o mano d'opera del magazzino e ufficio per svolgere l'attività di handling per la giornata	Dipartimento, Giorno, Tipologie di ore, Unità organizzativa.	Ore

Tableau 6.11: Specifica del data Mart Workforce BUD_D

Dimensione

In Tabella 6.12 si descrivono le dimensioni specificando per ognuna di esse il nome, una breve descrizione, quali sono gli attributi modificabili nel tempo e il tipo di trattamento da utilizzare per le modifiche.

Nome	Descrizione	Granularità	Attributi modificabili	Trattamento modifiche
DEPARTMENT	Nome del reparto dove si svolge l'attività	Il singolo dipartimento	No	-
DAY	Dimensione temporale che identifica il verificarsi di un fatto	Giornaliero	No	-
HOURL TYPE	Typo di ore da pagare, straordinari, vacanze,...	Orario	No	-
ORGANIZATION UNIT	Unità organizzativa sparsa nelle regioni	Il singolo unità organizzativa	ID_UNIT	Tipo_2

Tableau 6.12: Tabella della dimensione del data mart Workforce Giornaliera

Misure

Si presenta la tabella riepilogativa delle misure per il fatto **F_WORKFORCE_D**. Tali misure sono fondamentali per il calcolo delle metriche e dei KPI presentati nel Paragrafo 5.2.3.

Per ciascuna di esse si specifica il nome, il tipo di aggregazione e se il valore deriva o meno dalla combinazione di due o più misure. (cfr. Tabella 6.13)

Misura	Aggregabilità	Derivata
HOURS	Additiva	No

Tableau 6.13: Tabella Misure Data Mart Workforce Giornaliera

Schema concettuale

La figura mostra lo schema concettuale definito a partire dalle informazioni prodotto durante l'analisi dei requisiti.

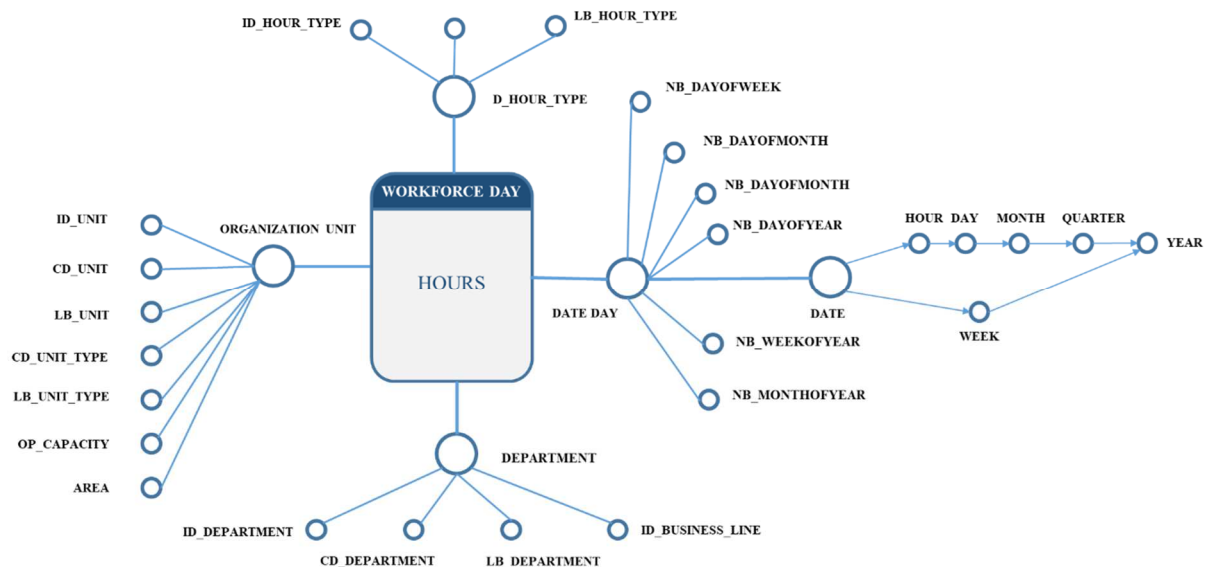


Tableau 6.14: Schemas concettuale Workforce Giornaliero

Lo schema riporta le misure all'interno delle tabelle del fatto, dal quale originano i collegamenti verso le dimensioni e i relativi attributi; si noti come gli attributi Hour, Day, Month, Quarter e Year della dimensione Day fanno parte di una gerarchia bilanciata con quattro livelli.

6.2 Data mart Finance

Il data mart dell'entità finanza sarà costituito dopo l'analisi di quattro tabelle di fatti, ovvero:

- **F_FINANCE:** In questa tabella dei fatti saranno memorizzate le misure sull'attività dell'entità finanza.

6.2.1 Fatto finance F_FINANCE

Specifica del fatto

Di seguito si descrive il fatto F_FINANCE specificando, nella Tabella 6.15, il suo significato, le dimensioni e le misure preliminari interessate. La granularità del fatto determina la dimensione del data mart e il tipo di analisi che si può effettuare sui dati.

Descrizione	Dimensioni preliminari	Misure preliminari
Ore necessario previsto a bilancio per il personale o mano d'opera del magazzino e ufficio per svolgere l'attività di Handling per un mese	Dipartimento, Mese, Tipologie di ore, Unità organizzativa.	social account amount locale, social account amount mensile, social account amount locale,

Tableau 6.15: Specifica del data Mart Finance

Dimensione

In Tabella 6.16 si descrivono le dimensioni specificando per ognuna di esse il nome, una breve descrizione, quali sono gli attributi modificabili nel tempo e il tipo di trattamento da utilizzare per le modifiche.

Nome	Descrizione	Granularità	Attributi modificabili	Trattamento modifiche
ACCOUNT	Conto contabile in SAP	Il singolo conto	CD_ACCOUNT	Tipo_2
CURRENCY	Contiene le informazioni sulla valuta	Singola valuta	-	-
CUSTOMER	Dimensione contenente il nome et codifica di tutti i Clienti, compagnie aeree	La singola compagnie aerea	CD_UNIT	Tipo_2
DATE MONTH	Data in mese	Mese	-	-
BUSINESS LINE (ORG)	Descrive le filiali per tipo di attività	Singola filiale	-	-
LEGAL ENTITY (ORG)	Entità giuridica, nell'organizzazione aziendale	Singola entità giuridica	-	-
REPORTING UNIT (ORG)	Dimensione usato per fare il reporting per entità giuridica e business line	Singolo reporting unit	CD_REPORTING_UNIT	Tipo_2
STATION (ORG)	Stazione di un'entità giuridica e business line	Singola stazione	-	-

Tableau 6.16: Tabella della dimensione del data Mart Workforce Giornaliera

Misure

Si presenta la tabella riepilogativa delle misure per il fatto **F_OPERATION_D**. Tali misure sono fondamentali per il calcolo delle metriche e dei KPI presentati nel Paragrafo 5.2.3.

Per ciascuna di esse si specifica il nome, il tipo di aggregazione e se il valore deriva o meno dalla combinazione di due o più misure. (cfr. Tabella 6.17)

Misura	Aggregabilità	Derivata
SOCAMOUNT_LOCAL	additiva	No
SOCAMOUNT_M	additiva	No
SOCAMOUNT_YTD	additiva	No
CONVAMOUNT	additiva	No
CONVAMOUNT_YTD	additiva	No
CONSAMOUNT	additiva	No
CONSAMOUNT_M	additiva	No
CONSAMOUNT_M	additiva	No
CONSAMOUNT_YTD	additiva	No

Tableau 6.17:Tabella Misure Data Mart Finance

Schema concettuale

La figura mostra lo schema concettuale definito a partire dalle informazioni prodotte durante l'analisi dei requisiti.

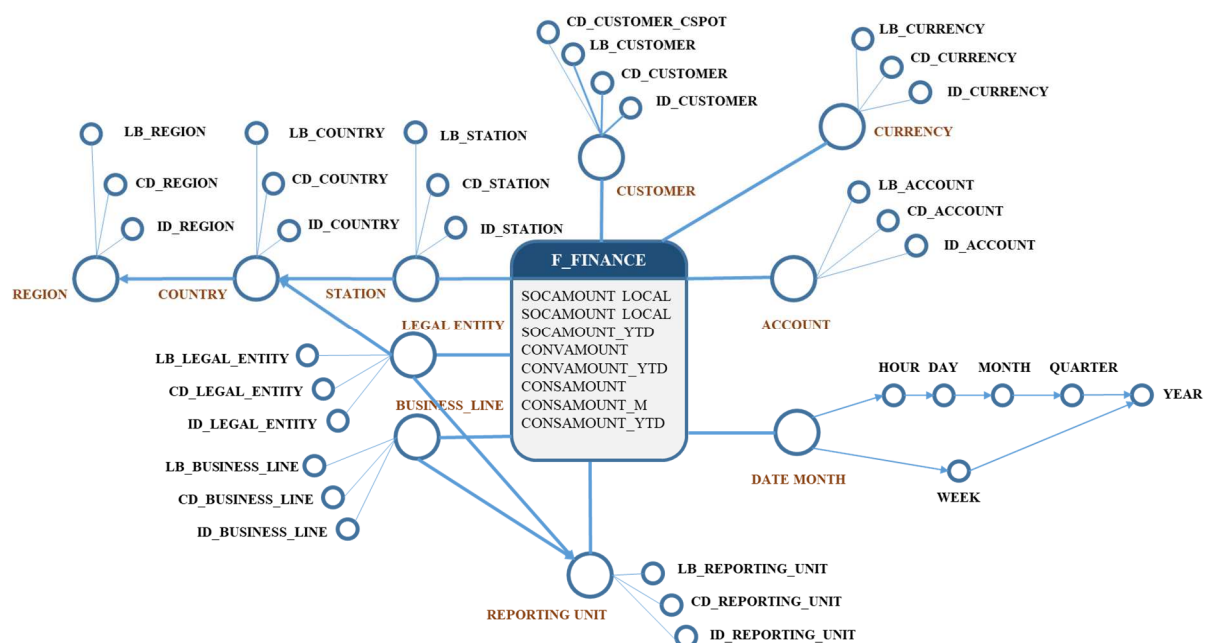


Figure 6.3:Schemas concettuale Data Mart Finance

Lo schema riporta le misure all'interno delle tabelle del fatto, dal quale originano i collegamenti verso le dimensioni e i relativi attributi; si noti come gli attributi Hour,

Day, Month, Quarter e Year della dimensione Day fanno parte di una gerarchia bilanciata con quattro livelli. Lo stesso vale per la dimensione, STATION, COUNTRY, REGION.

6.3 Data mart HR People

Il data mart HR & PEOPLE sarà costituito dopo l'analisi di quattro tabelle di fatti, ovvero:

- **F_HEADCOUNTS_M:** In questa tabella dei fatti saranno memorizzate le misure numero di impiegati attivi nel gruppo WFS
- **F_HIRES_M:** in questa tabella di fatto saranno memorizzate le misure sul numero di persone assunto nell'ultimo mese nel gruppo WFS
- **F_LEAVERS_M:** in questa tabella di fatto saranno memorizzate le misure sul numero di partenze nel gruppo WFS;

6.3.1 Fatto F_HEADCOUNTS_M

Specifica del fatto

Di seguito si descrive il fatto **F_HEADCOUNTS_M**, specificando, nella Tabella 6.18, il suo significato, le dimensioni e le misure preliminari interessate. La granularità del fatto determina la dimensione del data mart e il tipo di analisi che si può effettuare sui dati.

Descrizione	Dimensioni preliminari	Misure preliminari
Descrive l'organico necessario per lo svolgimento dell'attività nelle stazioni. Si diversi tipi d'organico. Quelli fissi e quelli flessibili, proposto agenzie di lavori collaborare	Dipartimento, Mese, socio	Organico

Tableau 6.18: Specifica del data Mart Headcounts

Dimensione

In Tabella 6.19 si descrivono le dimensioni specificando per ognuna di esse il nome, una breve descrizione, quali sono gli attributi modificabili nel tempo e il tipo di trattamento da utilizzare per le modifiche.

Nome	Descrizione	Granularità	Attributiti modificabili	Trattamento modifiche
MONTH	Dimensione temporale che identifica il verificarsi di un fatto	Mensile	No	-
PARTNER	Descrizione del socio/Partner	Singolo socio /Partner	No	-
DEPARTMENT	Nome del reparto dove si svolge l'attività	Il singolo dipartimento	No	-

Tableau 6.19:Tabella della dimensione del data mart Headcounts

Misure

Si presenta la tabella riepilogativa delle misure per il fatto **F_HEADCOUNTS_M**. Tali misure sono fondamentali per il calcolo delle metriche e dei KPI presentati nel Paragrafo 5.2.3.

Per ciascuna di esse si specifica il nome, il tipo di aggregazione e se il valore deriva o meno dalla combinazione di due o più misure. (cfr Tabella 6.20)

Misura	Aggregabilità	Derivata
HEADCOUNT	additiva	No

Tableau 6.20:Tabella della Misura per il Data Mart Headcounts

Schema concettuale

La figura mostra lo schema concettuale definito a partire dalle informazioni prodotto durante l'analisi dei requisiti.

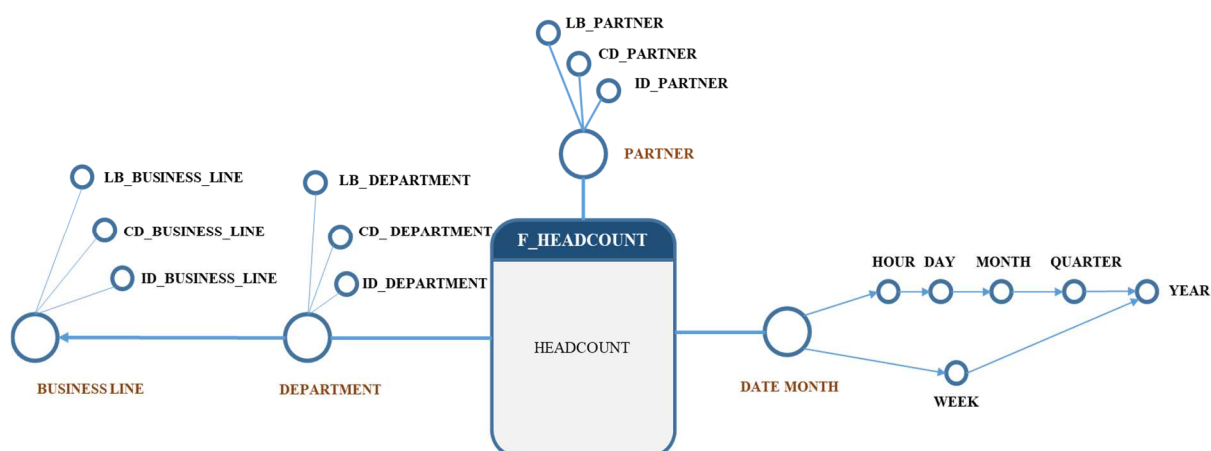


Figure 6.4:Schemas concettuale Data Mart Headcounts

Lo schema riporta le misure all'interno delle tabelle del fatto, dal quale originano i collegamenti verso le dimensioni e i relativi attributi; si noti come gli attributi Hour, Day, Month, Quarter e Year della dimensione Day fanno parte di una gerarchia bilanciata con quattro livelli. Lo stesso per DEPARTEMENT e BUSINESSLINE.

6.3.2 Fatto F_HIRES_M

Di seguito si descrive il fatto **F_HIRES_M**, specificando, nella Tabella 6.21, il suo significato, le dimensioni e le misure preliminari interessate. La granularità del fatto determina la dimensione del data mart e il tipo di analisi che si può effettuare sui dati.

Descrizione	Dimensioni preliminari	Misure preliminari
Descrive l'organico necessario per lo svolgimento dell'attività nelle stazioni. Si diversi tipi d'organico. Quelli fissi e quelli flessibili, proposto agenzie di lavori collaborare	Dipartimento, Mese, socio	Organico

Tableau 6.21: Specifica del data Mart Hires Mensile

Dimensione

In Tabella 6.22 si descrivono le dimensioni specificando per ognuna di esse il nome, una breve descrizione, quali sono gli attributi modificabili nel tempo e il tipo di trattamento da utilizzare per le modifiche.

Nome	Descrizione	Granularità	Attributi modificabili	Trattamento modifiche
MONTH	Dimensione temporale che identifica il verificarsi di un fatto	Mensile	No	-
PARTNER	Descrizione del partern	Sigolo partern	No	-
DEPARTMENT	Nome del reparto dove si svolge l'attività	Il singolo dipartimento	No	-

Tableau 6.22: Tabella della dimensione per il Data Mart Hires Mensile

Misure

Si presenta la tabella riepilogativa delle misure per il fatto **F_HIRES_M**. Tali misure sono fondamentali per il calcolo delle metriche e dei KPI presentati nel Paragrafo 5.2.3

Per ciascuna di esse si specifica il nome, il tipo di aggregazione e se il valore deriva o meno dalla combinazione di due o più misure. (cfr. Tabella 6.23)

Misura	Aggregabilità	Derivata
HIRES	additiva	No

Tableau 6.23:Tabella della Misura per il Data Mart Headcounts

Schema concettuale

La figura mostra lo schema concettuale definito a partire dalle informazioni prodotto durante l'analisi dei requisiti.

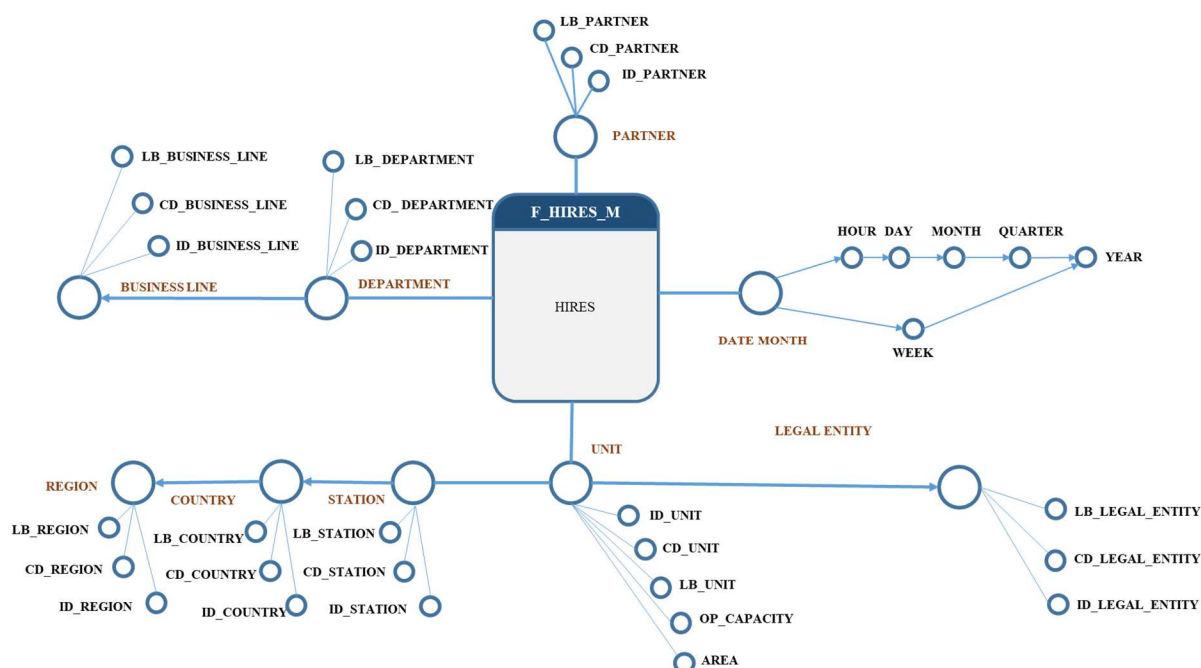


Figure 6.5:Schemas concettuale Data Mart Hires

Lo schema riporta le misure all'interno delle tabelle del fatto, dal quale originano i collegamenti verso le dimensioni e i relativi attributi; si noti come gli attributi Hour, Day, Month, Quarter e Year della dimensione Day fanno parte di una gerarchia bilanciata con quattro livelli. Lo stesso per STATION, COUNTRY, REGION.

6.3.3 Fatto F_LEAVERS_M

Di seguito si descrive il fatto **F_LEAVERS_M**, specificando, nella Tabella 6.24, il suo significato, le dimensioni e le misure preliminari interessate. La granularità del fatto determina la dimensione del data mart e il tipo di analisi che si può effettuare sui dati.

Specifica del fatto

Descrizione	Dimensioni preliminari	Misure preliminari
Ore necessario previsto a bilancio per il personale o mano d'opera del magazzino e ufficio per svolgere l'attività di handling per un mese	Dipartimento, Mese, Tipologie di ore, Unità organizzativa.	Ore Previsto a bilancio

Tableau 6.24: Specifica del data Mart Leavers Mensile

Dimensione

In Tabella 4.3 si descrivono le dimensioni specificando per ognuna di esse il nome, una breve descrizione, quali sono gli attributi modificabili nel tempo e il tipo di trattamento da utilizzare per le modifiche.

Nome	Descrizione	Granularità	Attributi modificabili	Trattamento modifiche
DEPARTMENT	Nome del reparto dove si svolge l'attività	Il singolo dipartimento	No	-
MONTH	Dimensione temporale che identifica il verificarsi di un fatto	Mensile	No	-
HOURLY TYPE	Tipo di ore da pagare, straordinari, vacanze,...	Orario	No	-
ORGANIZATION UNIT	Unità organizzativa sparsa nelle regioni	Il singolo unit0 organizzativa	ID_UNIT	Tipo_2

Tableau 6.25: Tabella della dimensione del data mart Leaves Mensile

Misure

Si presenta la tabella riepilogativa delle misure per il fatto **F_LEAVERS_M**. Tali misure sono fondamentali per il calcolo delle metriche e dei KPI presentati nel Paragrafo 5.2.3 Per ciascuna di esse si specifica il nome, il tipo di aggregazione e se il valore deriva o meno dalla combinazione di due o più misure. (Tabella)

Misura	Aggregabilità	Derivata
LIVERS	additiva	No

Tableau 6.26: Tabella della Misura per il Data Mart Leaves mensile

Schema concettuale

La figura mostra lo schema concettuale definito a partire dalle informazioni prodotte durante l'analisi dei requisiti.

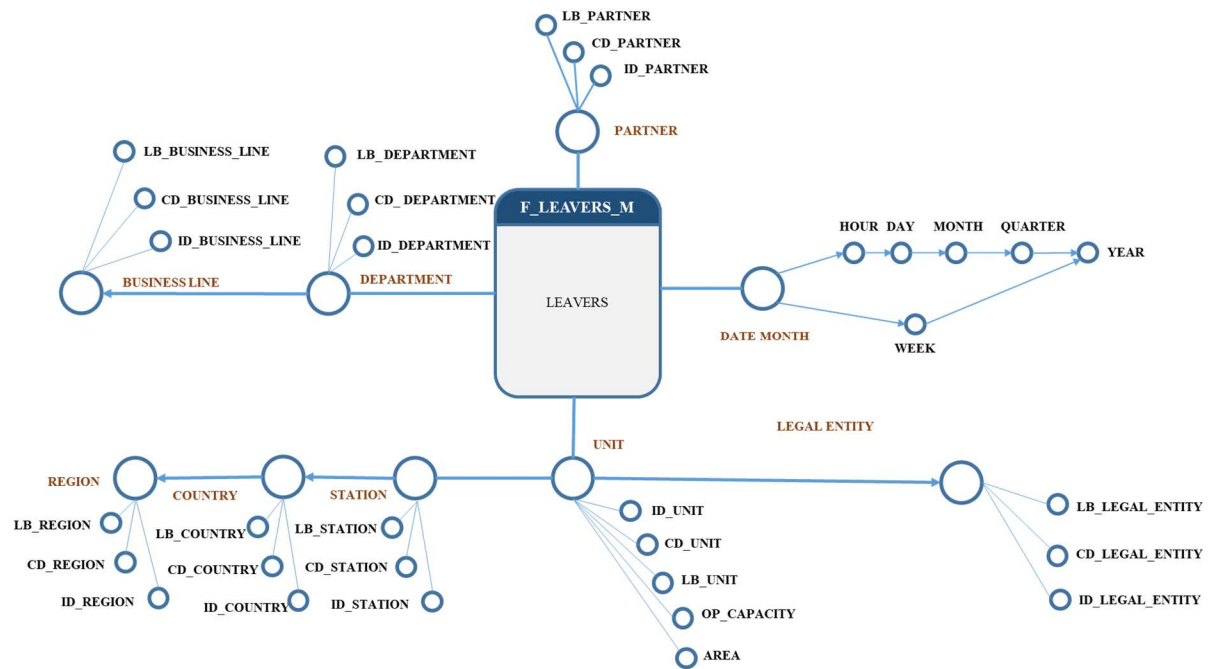


Figure 6.6: Schemas concettuale Data Mart Leaves Mensile

Lo schema riporta le misure all'interno delle tabelle del fatto, dal quale originano i collegamenti verso le dimensioni e i relativi attributi; si noti come gli attributi Hour, Day, Month, Quarter e Year della dimensione Day fanno parte di una gerarchia bilanciata con quattro livelli.

Capitolo 7

Progettazione concettuale finale e logica

Nel seguente Capitolo verranno affrontati i passi finali della progettazione concettuale con lo scopo di arricchire gli schemi presentati nel capitolo precedente.

Per ciascun data mart si passerà successivamente alla fase di progettazione logica, ovvero alla traduzione dei modelli concettuali in modelli relazionali.

Infine, nel **Paragrafo 5.5** sarà presentato lo schema relazionale dell'intero datawarehouse.

7.1 Elementi in comune tra i data mart nel datawarehouse

Prima di procedere con la modellazione concettuale finale si presenta l'elenco delle dimensioni individuate specificando in quali data mart sono impiegate.

Questo riepilogo è utile per evidenziare quali informazioni sono in comune a processi diversi e quindi dovrebbero avere interpretazione e rappresentazione unica, per essere poi condivise nel datawarehouse.

	FINANCE	HR PEOPLE			OPERATION			
DIMENSIONS	F_FINANCE_M	F_HEADCOUNTS_M	F_HIRES_M	F_LEAVERS_M	F_OPERATIONS_D	F_OPERATIONS_BUP_M	F_WORKFORCE_BUD_M	F_WORKFORCE_D
D_ACCOUNT	X							
D_CUSTOMER	X				X	X		
D_DEPARTMENT		X	X	X			X	X
D_DT_DAY					X			X
D_DT_WEEK								
D_DT_MONTH	X	X	X	X		X	X	
D_DT_YEAR								
D_HANDLING_DELIVERY					X	X		
D_HANDLING_DIRECTION								
D_HANDLING_PROCESS					X	X		
D_HANDLING_PRODUCT					X	X		
D_HOUR_TYPE							X	X
D_ORG_BUILDING								
D_ORG_BUSINESS_LINE	X							
D_ORG_COUNTRY								
D_ORG_LEGAL_ENTITY	X							
D_ORG_REGION								
D_ORG_REPORTING_UNIT	X							
D_ORG_STATION	X							
D_ORG_UNIT			X	X	X		X	X
D_ORG_PARTENER	X	X						

Tableau 7.1:Tabella riepilogativa delle dimensioni in comune tra i data mart

La Tabella 7.1 mostra come la dimensione che non ci sono dimensione comune tra le diverse aree; i data mart Cargo Operation e finance condividono D_DT_MONTH

Si presenta la tabella riepilogativa aggiornata delle dimensioni in comune tra i data mart

	FINANCE	HR PEOPLE			OPERATION			
DIMENSIONS	F_FINANCE_M	F_HEADCOUNTS_M	F_HIRES_M	F_LEAVERS_M	F_OPERATIONS_D	F_OPERATIONS_BUP_M	F_WORKFORCE_BUD_M	F_WORKFORCE_D
D_ACCOUNT	X							
D_CUSTOMER	X				X	X		
D_DEPARTMENT		X	X	X			X	X
D_DT_DAY					X			X
D_DT_WEEK								
D_DT_MONTH	X	X	X	X		X	X	
D_DT_YEAR								
D_HANDLING_DELIVERY					X	X		
D_HANDLING_DIRECTION								
D_HANDLING_PROCESS					X	X		
D_HANDLING_PRODUCT					X	X		
D_HOUR_TYPE							X	X
D_ORG_BUILDING								
D_ORG_BUSINESS_LINE	X							
D_ORG_COUNTRY								
D_ORG_LEGAL_ENTITY	X							
D_ORG_REGION								
D_ORG_REPORTING_UNIT	X							
D_ORG_STATION	X							
D_ORG_UNIT			X	X	X		X	X
D_ORG_PARTENER	X	X						

Tableau 7.2:Tabella riepilogativa finale delle dimensioni in comune tra i data mart

7.2 Cargo Operation

Per l'entità di Cargo Operation avremo quattro tabelle di fatti come segue:

- **F_OPERATION_D:** In questa tabella dei fatti saranno memorizzate le misurazioni sull'attività operativa con una frequenza giornaliera.
- **F_OPERATION_BUD_M:** in questa tabella di fatto saranno memorizzate le misure sull'attività operativa prevista a budget, con frequenza mensile.
- **F_WORFORCE_D:** in questa tabella di fatto saranno memorizzate con frequenza giornaliera le misure sulle risorse necessarie per l'esecuzione delle operazioni.
- **F_WORKFORCE_BUD_M:** in questa tabella di fatto saranno memorizzate le misure sulla proiezione mensile delle risorse necessarie per l'esecuzione dei compiti operativi.

7.2.1 Elementi comune tra i data mart di cargo operation

Prima di procedere con la modellazione concettuale di cargo operation si presenta l'elenco delle dimensioni individuate specificando in quali data mart sono impiegate.

Questo riepilogo è utile per evidenziare quali informazioni sono in comune a processi diversi e quindi dovrebbero avere interpretazione e rappresentazione unica, per essere poi condivise nel datawarehouse.

	OPERATION			
	OPERATIONS_D	OPERATIONS_BUP_M	WORKFORCE_BUD_M	WORKFORCE_D
D_CUSTOMER	X	X		
D_DT_DAY	X			X
D_HANDLING_DELIVERY	X	X		
D_HANDLING_PROCESS	X	X		
D_HANDLING_PRODUCT	X	X		
D_ORG_UNIT	X	X	X	X
D_DT_MONTH		X	X	
D_DEPARTMENT			X	X
D_HOUR_TYPE			X	X

Tableau 7.3:Tabella riepilogativa delle dimensioni in comune tra i data mart dell'entità cargo operation

La Tabella 7.3 mostra come la dimensione D_ORG_UNIT sia l'unica in comune tra le diverse aree; I data mart Operation_D, Operation_BUD_M, condividono invece anche la gerarchia composta dalle dimensioni D_CUSTOMER, D_HANDLING_DELIVERY, D_HANDLING_PRODUCT e ORG_UNIT.

Da ulteriori analisi con i committenti ed emerso che la dimensione D_CUSTOMER, Rappresenterebbe un livello aggiuntivo della gerarchia dimensionale D_ORG_UNIT. L'introduzione di tale dimensione nella gerarchia consentirebbe non solo di effettuare analisi a livello l'unità organizzativa ma anche di calcolare alcuni KPI fondamentali. Inoltre, la dimensione D_ORG_UNIT Sarà soggetta a frequenti modifiche in futuro e quindi non dovrebbe essere trattata come dimensione degenere

Si presenta la tabella riepilogativa aggiornata delle dimensioni in comune tra i data mart (cfr. Tabella 7.4)

	OPERATION			
	OPERATIONS_D	OPERATIONS_BUP_M	WORKFORCE_BUD_M	WORKFORCE_D
D_CUSTOMER	X	X	X	X
D_DT_DAY	X			X
D_HANDLING_DELIVERY	X	X		
D_HANDLING_PROCESS	X	X		
D_HANDLING_PRODUCT	X	X		
D_ORG_UNIT	X	X	X	X
D_DT_MONTH		X	X	
D_DEPARTMENT			X	X
D_HOUR_TYPE			X	X

Tableau 7.4:Tabella riepilogativa finale delle dimensioni in comune tra i data mart dell'entità cargo operation

7.2.2 Progettazione concettuale finale del data mart Cargo Operation_D

Lo schema concettuale del data mart Cargo Operation Daily, mostrato in figura 7.1, non presenta alcune differenze con lo schema introdotto nel capitolo precedente.

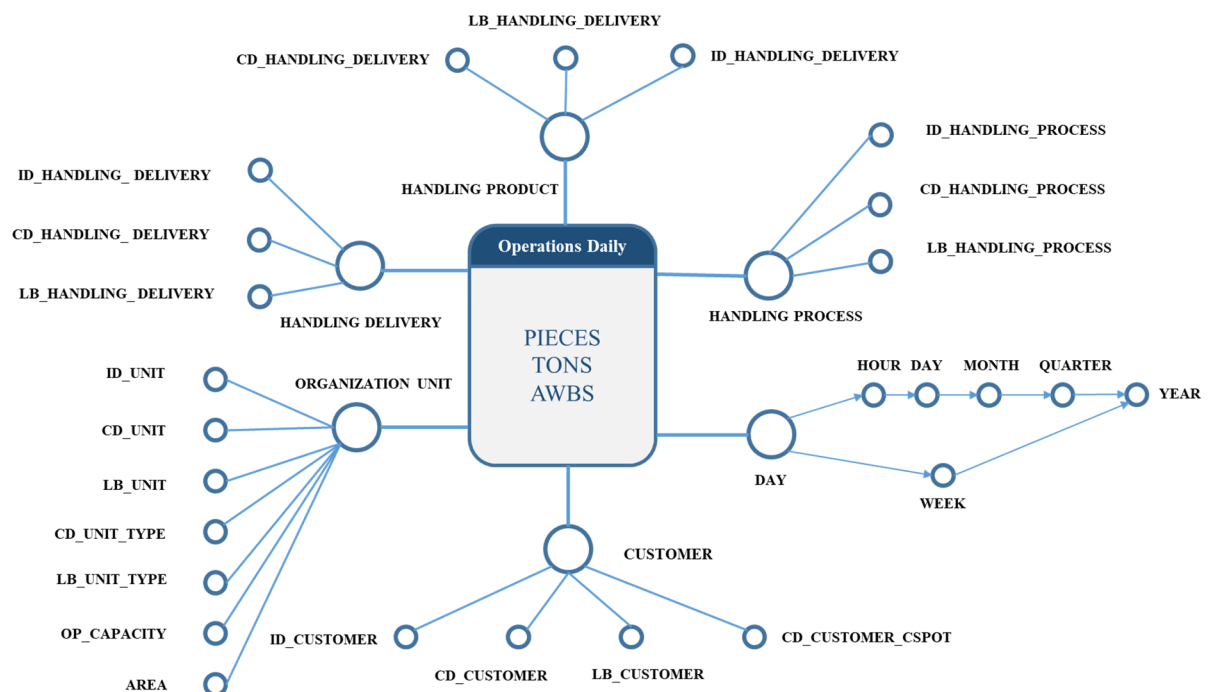


Figure 7.1:Schema concettuale del data mart Operation Daily

7.2.3 Progettazione logica del data mart Cargo Operation_D

Per la traduzione da schema concettuale a schema logico sono state adottate le seguenti convenzioni, come suggerito in [Albano 14]:

- Utilizzo di schemi a **stella**, **fioco di neve** e **costellazione** per la rappresentazione dei singoli data mart del datawarehouse finale. Lo schema a stella consiste in una tabella del fatto che referencia due o più tabelle dimensionali; lo schema a fiocco di neve è un'estensione dello schema a stella dove le tabelle dimensionali referenzino a loro volta altre tabelle; infine, lo schema a costellazione consiste in due o più tabelle del fatto che condividono alcune dimensioni.
- Introduzione di una **chiave primaria surrogata** in aggiunta alla chiave naturale per ciascuna tabella dimensionale. Negli schemi logici le chiavi primarie sono contrassegnate dalla sigla PK, primary key
- Introduzione di chiave primaria surrogate e di una o più **chiave esterne** per la tabella del fatto. Negli schemi logici le chiavi esterne sono contrassegnate dalla sigla FK, *foreign key*
- I nomi della tabella seguono la convenzione *D_nome_dimension* per quanto riguarda le tabelle dimensionali e *F_nome_fatto* per quanto riguarda le tabelle del fatto.

La Figura 7.1 mostra lo schema logica del data mart Cargo Operation. Lo schema segue la rappresentazione a fiocco di neve: La tabella del fatto F_OPERATION...e infatti collegata con la dimensione D_ORG_UNIT, che contiene elementi della gerarchia organizzativa, D_DT_MONTH per elementi temporali, D_CUSTOMER che contiene informazioni sul clienti, D_HANDLING_DELIVERY, il modo per il delivery della merche, D_HANDLY_PRODUCT, che contiene informazioni sul tipo del prodotto e in fine la dimensione D_HANDLING_PROCESS, informazioni sul tipo di processo: Import, export o transit.

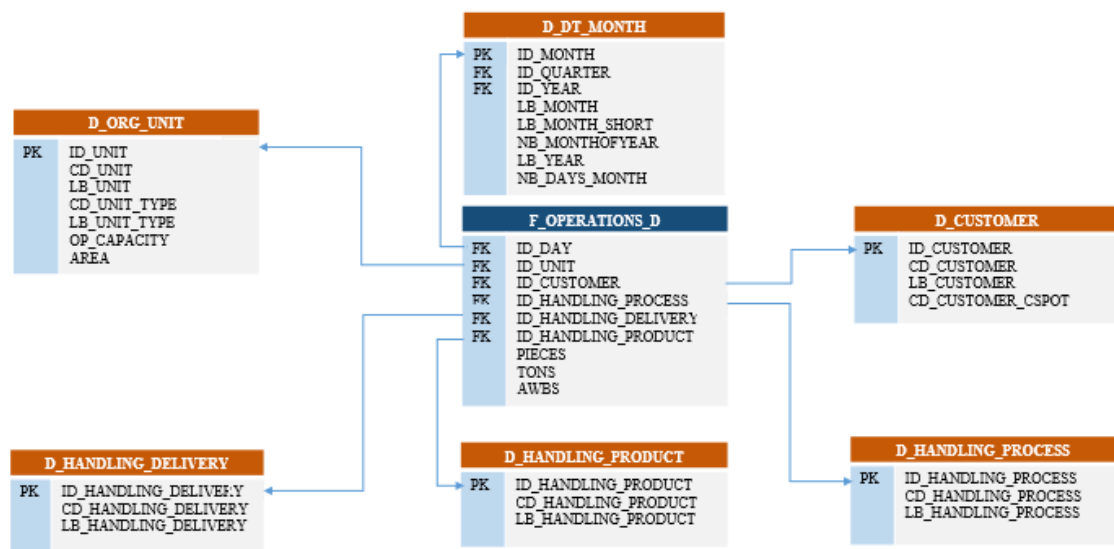


Figure 7.2:Schema logico del data mart Cargo Operation Daily

7.2.4 Progettazione concettuale finale del data mart Cargo Operations_BUD_M

Lo schema concettuale del data mart Cargo Operation Budget Month, mostrato in figura 7.3, non presenta alcune differenze con lo schema introdotto nel capitolo precedente.

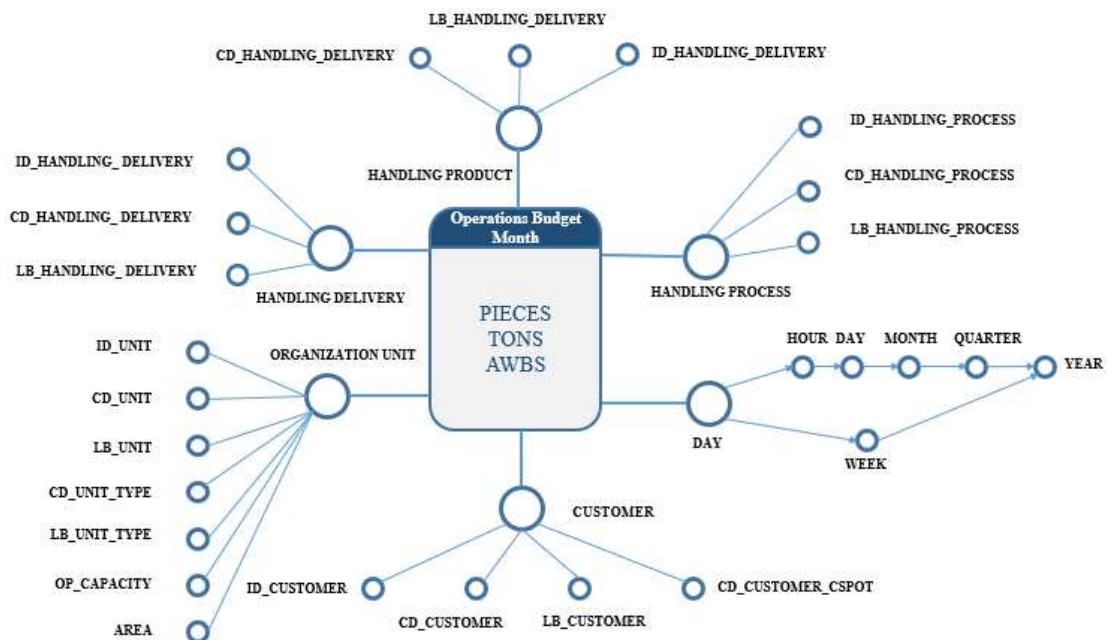


Figure 7.3:Schema concettuale del data mart Operation Budget Month

7.2.5 Progettazione logica del data mart: OPERATION_BUD_M

Lo schema utilizzato per rappresentare il data mart *Operation Budget Month* è, anche in questo caso, lo schema a fioco di neve. La tabella del fatto F_OPERATION_BUD_M, riferenzia sei dimensioni (cfr. figura 7.4)

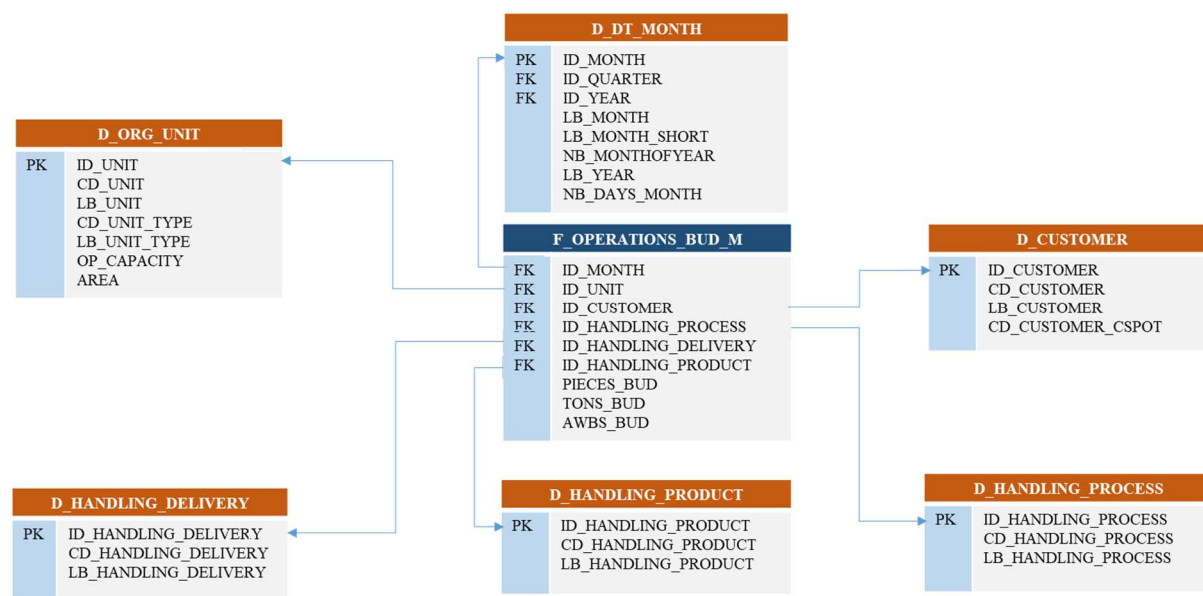


Figure 7.4:Schema logico del data mart Cargo Operation Month

7.2.6 Progettazione concettuale finale del data mart Workforce DAY

Come già accennato nel paragrafo precedente, una prima modifica allo schema concettuale riguarderà l'aggiunta della D_CUSTOMER come nuova dimensione

In figura 7.5 Viene mostrato lo schema concettuale finale del data mart Workforce Day, con l'integrazione della dimensione D_CUSTOMER.

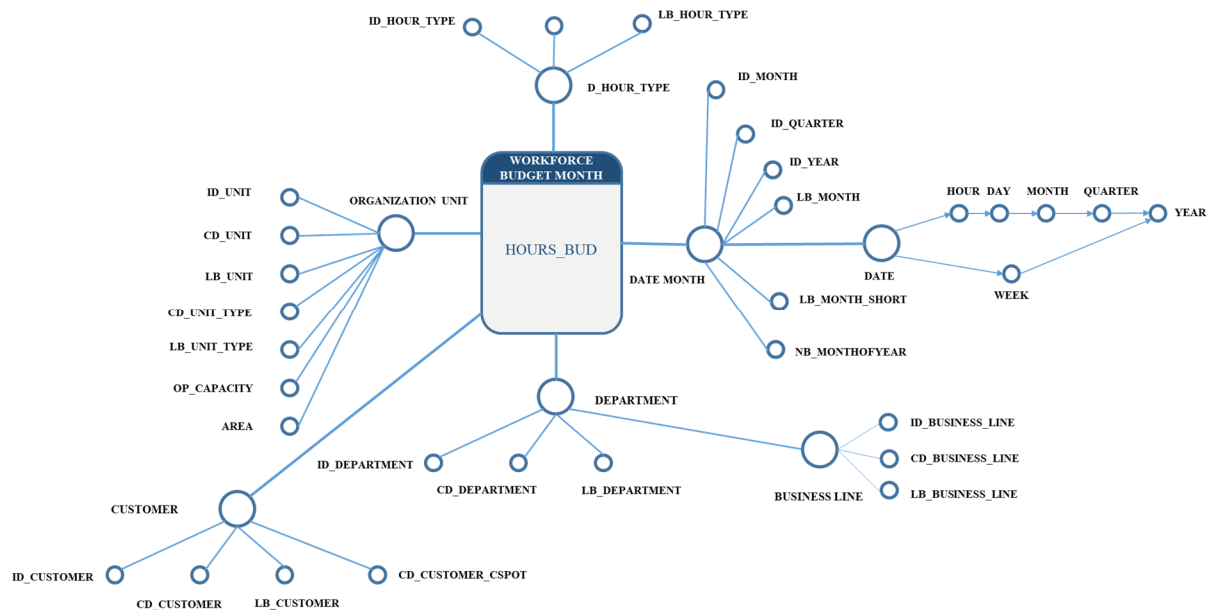


Figure 7.5: Schema concettuale finale del data mart WORKFORCE DAY

7.2.7 Progettazione logica del data mart: Workforce_Day

Lo schema utilizzato per rappresentare il data mart Workforce Day è, anche in questo caso, lo schema a fiocco di neve. La tabella del fatto, **F_WORKFORCE_D**, referenziata cinque dimensioni e le due gerarchie dimensionale **D_DT_WEEK – D_DT_MONTH** e **D_DEPARTMENT – D_ORG_BUSINESS_LINE**.

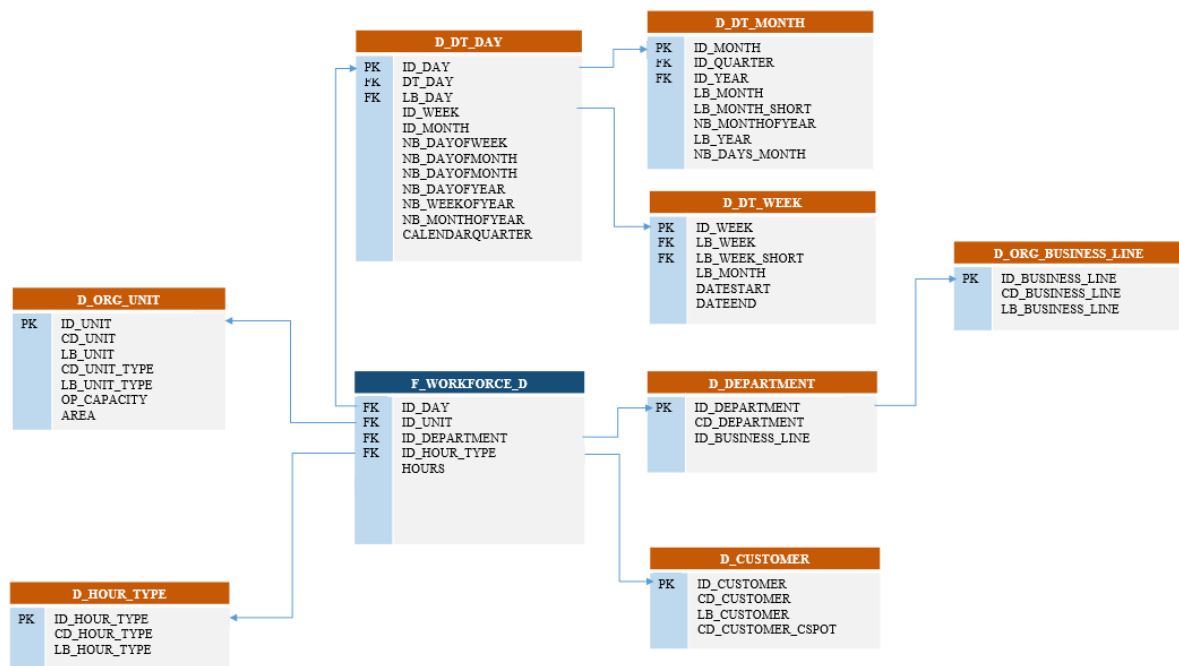


Figure 7.6: Schema logico del data mart Workforce Day

7.2.8 Progettazione concettuale finale del data mart Workforce BUD Month

Anche per questa data mart, una prima modifica allo schema concettuale riguarderà l'aggiunta della **D_CUSTOMER** come nuova dimensione

In figura 7.7 Viene mostrato lo schema concettuale finale del data mart Workforce Day, con l'integrazione della dimensione D_CUSTOMER.

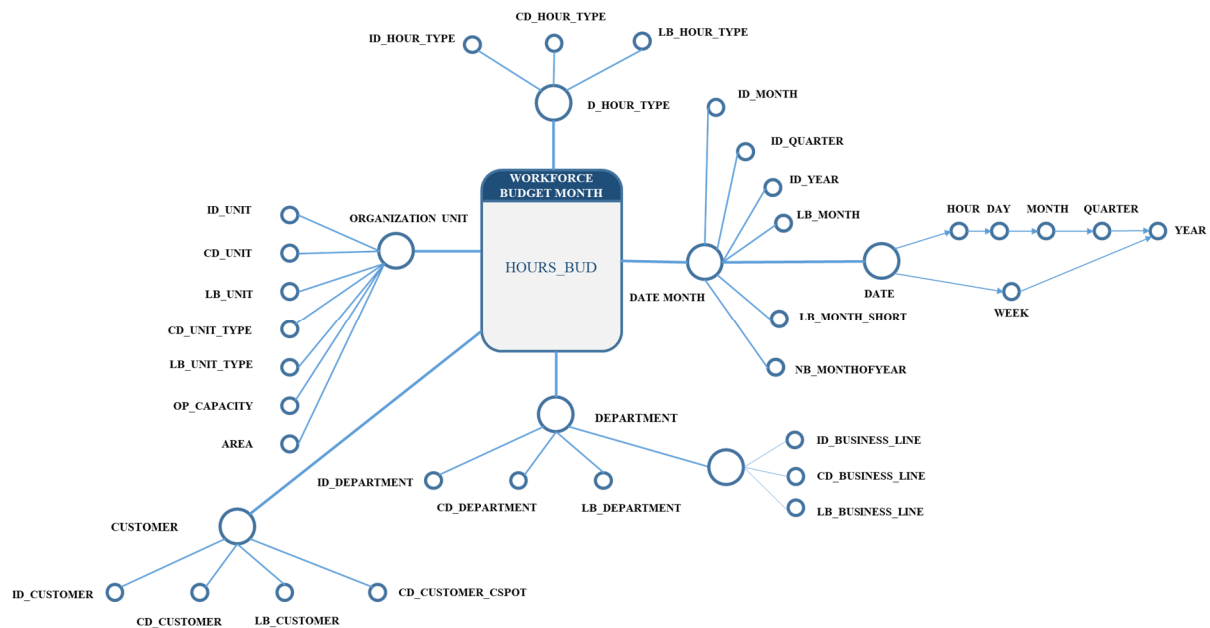


Figure 7.7: Schema concettuale del data mart Workforce Budget Month

7.2.9 Progettazione logica del data mart: Workforce Budget Month

Lo schema utilizzato per rappresentare il data mart *Workforce Budget Month* è, anche in questo caso, lo schema a fiocco di neve. La tabella del fatto, **F_WORKFORCE_D**, referenziata cinque dimensioni e le due gerarchie dimensionale **D_DT_WEEK – D_DT_MONTH** e **D_DEPARTMENT – D_ORG_BUSINESS_LINE**.

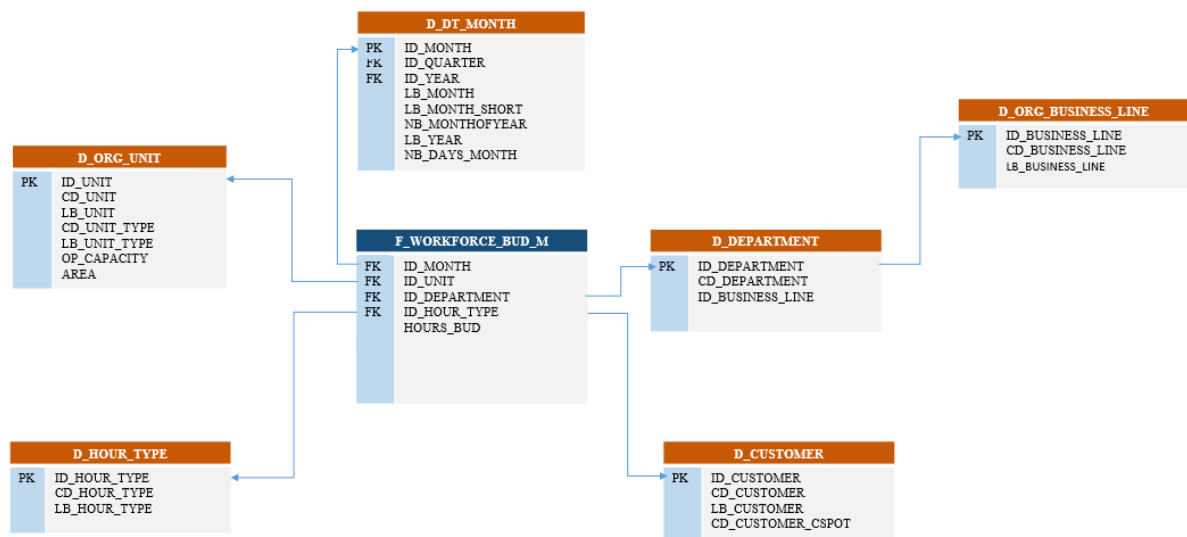


Figure 7.8:Schema logico del data mart Workforce BUD Month

7.3 HR People

7.3.1 Elementi comune tra i data mart di HR People o entità risorse umane.

Prima di procedere con la modellazione concettuale dell'entità risorse umane (HR People) si presenta l'elenco delle dimensioni individuate specificando in quali data mart sono impiegate.

Questo riepilogo è utile per evidenziare quali informazioni sono in comune a processi diversi e quindi dovrebbero avere interpretazione e rappresentazione unica, per essere poi condivise nel datawarehouse.

	HR PEOPLE		
	F_HEADCOUNTS_M	F_HIRES_M	F_LEAVERS_M
D_DEPARTMENT	X	X	X
D_DT_MONTH	X	X	X
D_ORG_UNIT		X	X
D_ORG_PARTENER	X		
D_SOURCE_SYSTEM	X	X	X

Tableau 7.5:Tabella riepilogativa delle dimensioni in comune tra i data mart dell'entità HR People

La Tabella 7.5 mostra come le dimensioni **D_DEPARTMENT**, **D_DT_MONTH** e infine **D_SOURCE_SYSTEM** sia l'unica in comune tra le diverse aree; I data mart

HIRE MONTH e **LEAVERS MONTH**, condividono invece le dimensioni **ORG_UNIT**.

Da ulteriori analisi con i committenti ed emerso che la dimensione **D_ORG_PARTNER**, Rappresenterebbe una nuova dimensione. L'introduzione di tale dimensione consentirebbe non solo di effettuare analisi a livello l'unità organizzativa ma anche di calcolare alcuni KPI fondamentali. Inoltre, la dimensione **D_ORG_PARTNER** Sara soggetta a frequenti modifiche in futuro e quindi non dovrebbe essere trattata come dimensione degenerare

Si presenta la tabella riepilogativa aggiornata delle dimensioni in comune tra i data mart (cfr. Tabella 7.6)

	HR PEOPLE		
	F_HEADCOUNTS_M	F_HIRES_M	F_LEAVERS_M
D_DEPARTMENT	X	X	X
D_DT_MONTH	X	X	X
D_ORG_UNIT		X	X
D_ORG_PARTNER	X	X	X
D_SOURCE_SYSTEM	X	X	X

Tableau 7.6:Tabella riepilogativa delle dimensioni in comune tra i data mart dell'entità HR People

7.3.2 Progettazione concettuale del data mart Headcounts Month

Lo schema concettuale del data mart *Headcounts Month*, mostrato in figura 7.9, non presenta alcune differenze con lo schema introdotto nel capitolo precedente.

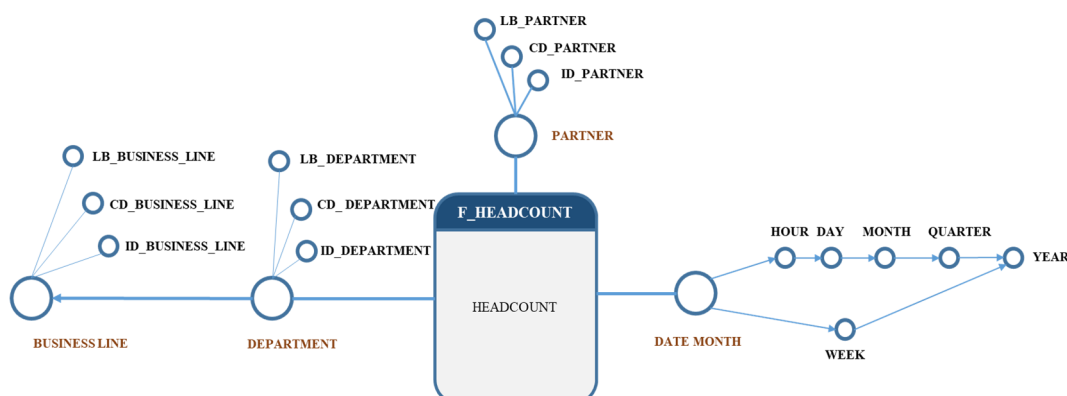


Figure 7.9:Schema concettuale del data mart *Headcounts Month*

7.3.3 Progettazione logica del data mart *Headcounts Month*

Lo schema utilizzato per rappresentare il data mart Headcounts Month è uno schema a fiocco di neve. La tabella del fatto, **F_HEADCOUNT**, referenzia tre dimensioni e la gerarchia dimensionale **D_DEPARTMENT – D_ORG_BUSINESS_LINE**.

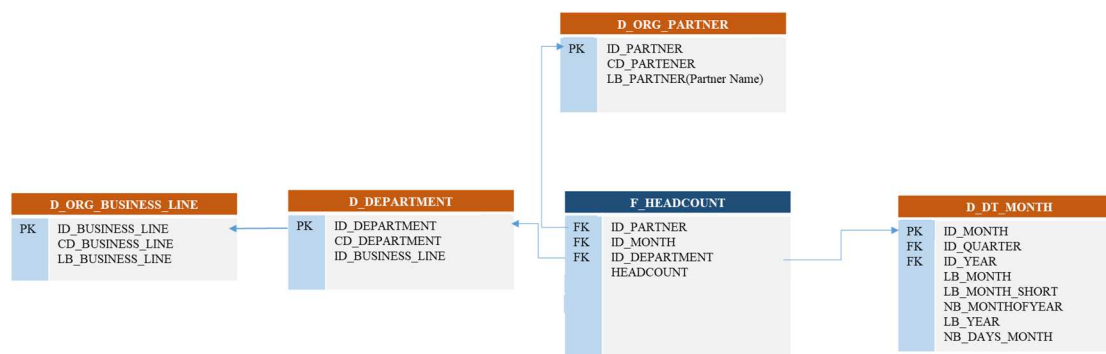


Figure 7.10: Schema logico del data mart *Headcounts Month*

7.3.4 Progettazione concettuale del data mart Hires

Una prima modifica allo schema concettuale del data mart *Hires* riguarderà l'aggiunta della **D_PARTNER** come nuova dimensione

In figura 7.11 Viene mostrato lo schema concettuale finale del data mart Hires, con l'integrazione della dimensione **D_PARTNER**.

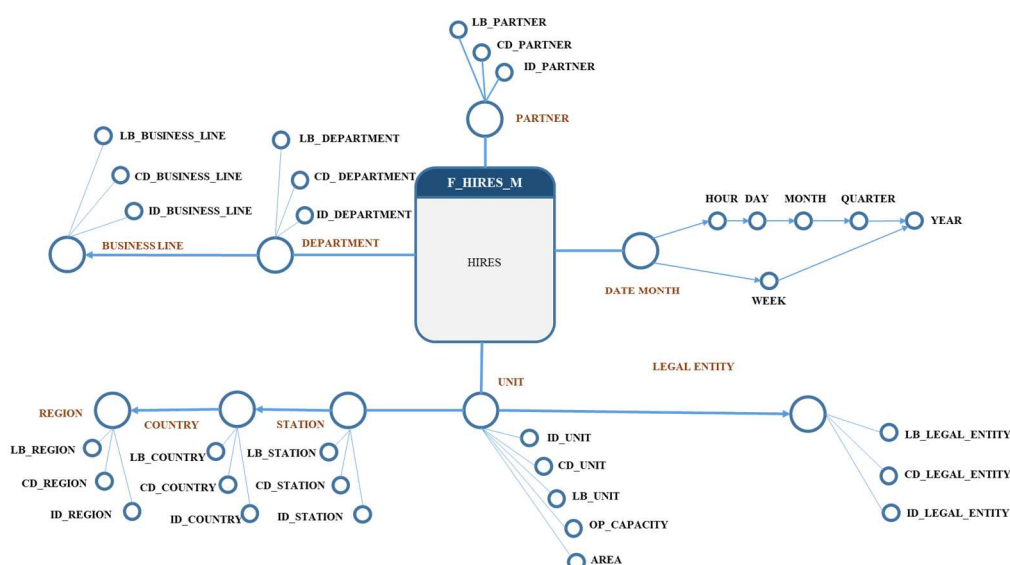


Figure 7.11: Schema concettuale del data mart Hires

7.3.5 Progettazione logica Hires

Lo schema utilizzato per rappresentare il data mart Hires è uno schema a fiocco di neve. La tabella del fatto, F_HIRES_M, referenziata quattro dimensioni e tre gerarchie dimensionale **D_DEPARTMENT – D_ORG_BUSINESS_LINE**, **D_ORG_UNIT-D_ORG_COUNTRY-D_ORG_REGION** e in fine **D_ORG_UNIT-D_ORG_LEGAL_ENTITY**;

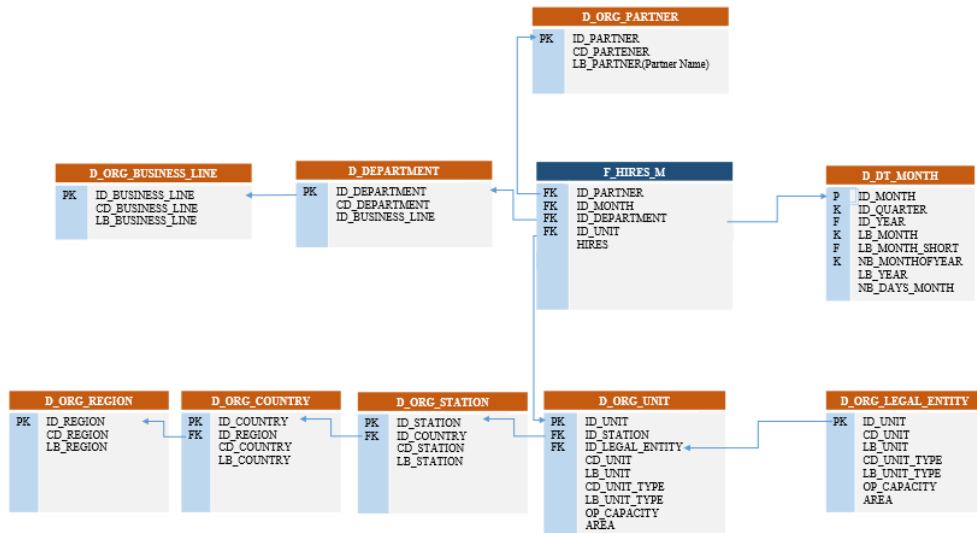


Figure 7.12:Schema logico del data mart Hires

7.3.6 Progettazione concettuale Leaves

Lo schema utilizzato per rappresentare il data mart *Leaves* è uno schema a fiocco di neve. La tabella del fatto, F_HIRES_M, referenziata quattro dimensioni e tre gerarchie dimensionale **D_DEPARTMENT – D_ORG_BUSINESS_LINE**, **D_ORG_UNIT-D_ORG_COUNTRY-D_ORG_REGION** e in fine **D_ORG_UNIT-D_ORG_LEGAL_ENTITY**;

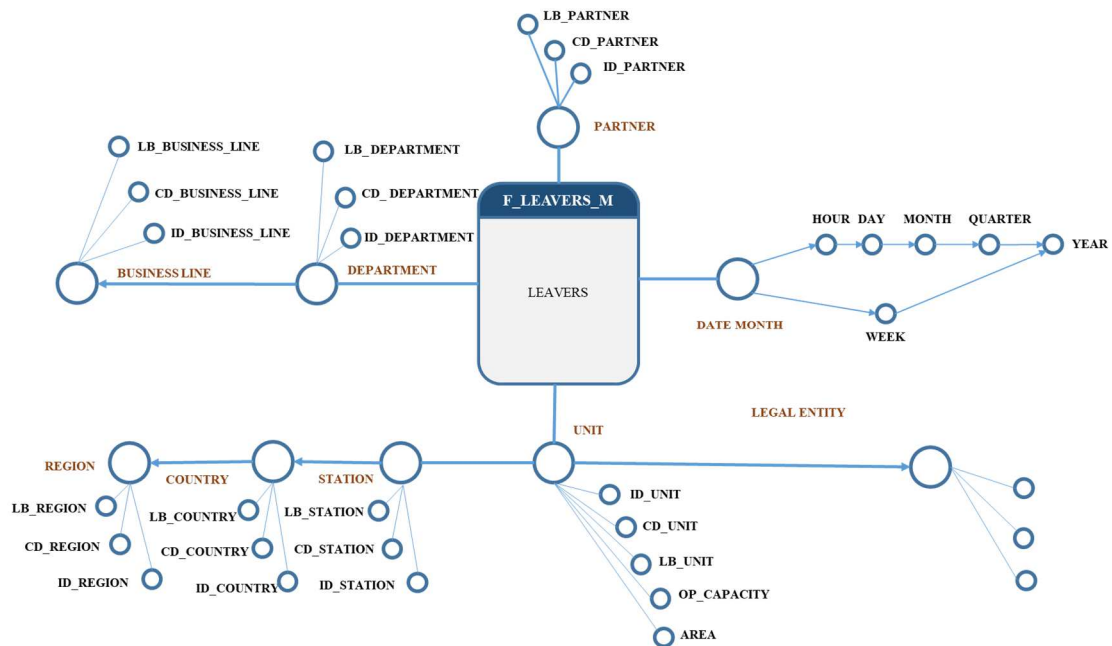


Figure 7.13:Schema concettuale del data mart Leaves

7.3.7 Progettazione logica Leaves

Lo schema utilizzato per rappresentare il data mart Headcounts Month è uno schema a fiocco di neve. La tabella del fatto, **F_HEADCOUNT**, referenziata tre dimensioni e la gerarchia dimensionale **D_DEPARTMENT – D_ORG_BUSINESS_LINE**.

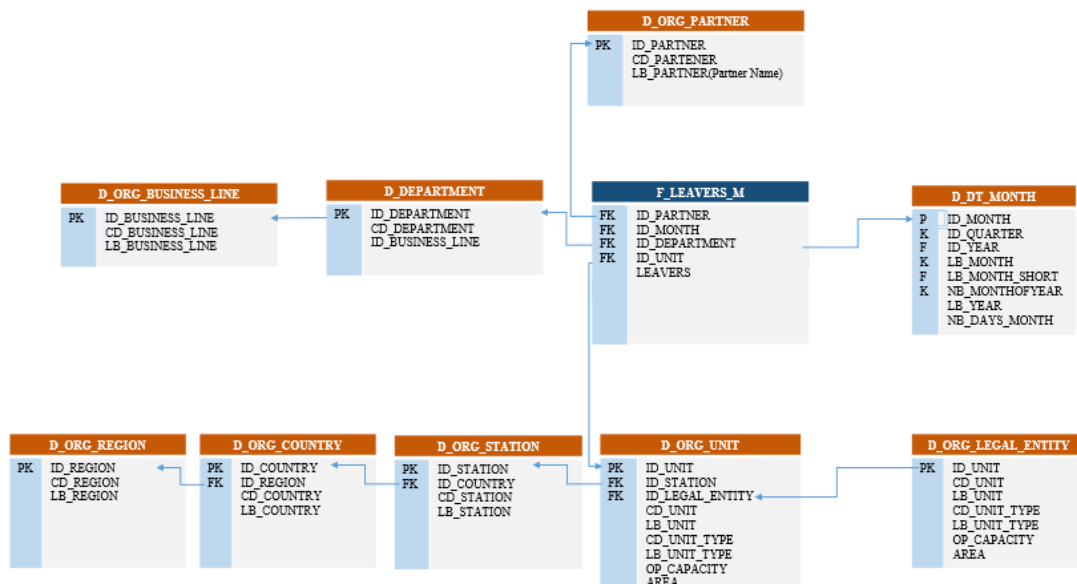


Figure 7.14:Schema logico del data mart Leaves

7.4 Data Mart Finance

7.4.1 Progettazione concettuale finale FINANCE

Lo schema concettuale del data mart *Headcounts Month*, mostrato in figura 7.9, non presenta alcune differenze con lo schema introdotto nel capitolo precedente.

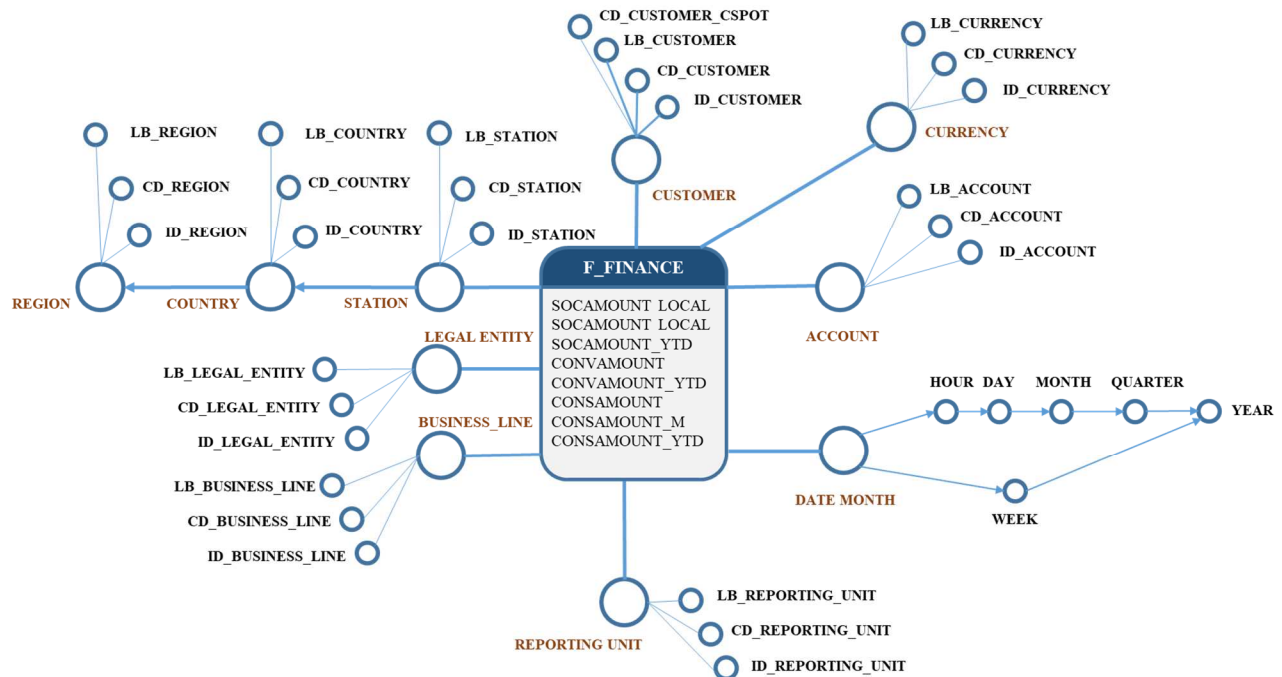


Figure 7.15: Schema concettuale del data mart Finance

7.4.2 Progettazione Logica FINANCE

Lo schema utilizzato per rappresentare il data mart Headcounts Month è uno schema a fiocco di neve. La tabella del fatto, **F_HEADCOUNT**, referenziata tre dimensioni e la gerarchia dimensionale **D_DEPARTMENT – D_ORG_BUSINESS_LINE**.

Capitolo 8

Procedure di estrazione, trasformazione e caricamento (ETL)

Terminate le fasi di analisi dei requisiti e di progettazione del cloud datawarehouse in snowflakes, si procede al design della piatta forma di dati di supporto alle decisioni. In questo capitolo, dopo una breve descrizione del processo ETL, verranno analizzate nel dettaglio le singole fasi. Sarà affrontato il problema dell'eterogeneità delle sorgenti dati e la scelta delle tecniche per garantire che il cloud datawarehouse sia costantemente aggiornato.

8.1 Il processo ETL

Il processo di estrazione e compilazione dei dati grezzi, la loro trasformazione per renderli comprensibili e il loro caricamento in un sistema di destinazione, come un database o un datawarehouse per un facile accesso e analisi, è noto come processo ETL. Abbreviazione di Extract - Transform - Load, Process of ETL è una componente importante nell'ecosistema dei dati di qualsiasi impresa moderna.

Poiché i dati provenienti da più fonti hanno un modello diverso, ogni set di dati deve essere trasformato in modo diverso prima di poter essere utilizzato per la BI e l'analisi. Ad esempio, se si compilano dati da sistemi sorgente come Amazon Redshift e Google Analytics, queste due fonti dovranno essere trattate singolarmente con l'intero processo ETL.

Implementazione del processo ETL: le tre fasi

Per quanto riguarda l'implementazione del livello di processo ETL, il flusso di lavoro può essere suddiviso nella forma completa del suo acronimo.

1. E - Estrazione
2. T - Trasformazione
3. L - carico

Scopriamo come convertire i dati grezzi in informazioni intelligibili.

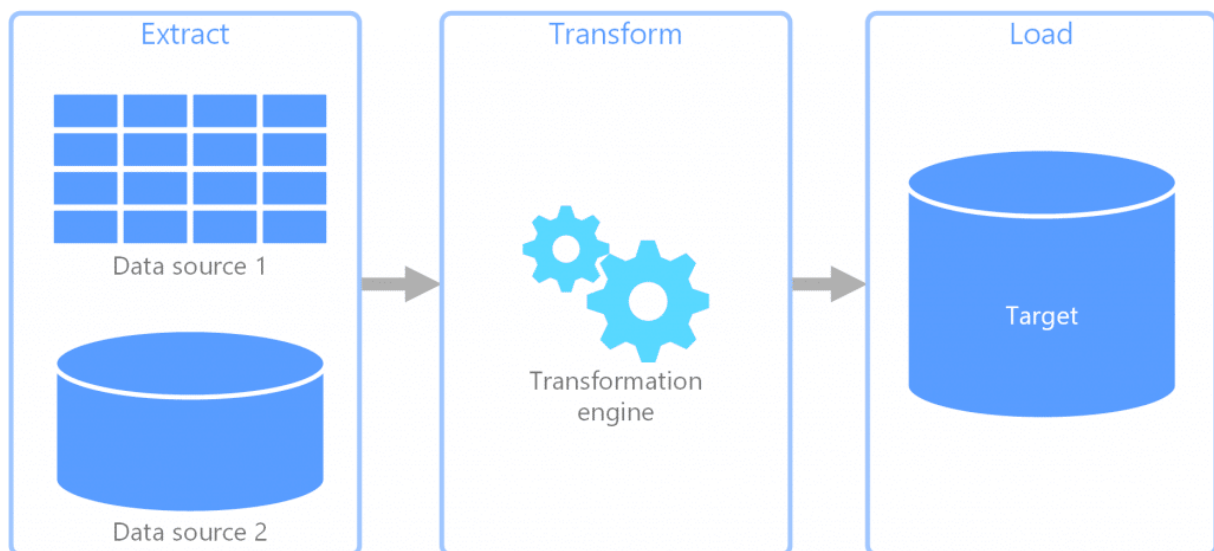


Figure 8.1: Il processo ETL: dai sistemi sorgenti al datawarehouse

Le tre fasi dei processi ETL sono le seguenti:

8.1.1 Fase 1 — Estrazione

Il primo passo prima di iniziare a organizzare i dati è quello di estrarre o estrarre i dati da tutte le fonti rilevanti e compilarli. Queste fonti possono includere database on-premise, sistemi CRM, piattaforme di automazione del marketing, file non strutturati e strutturati, applicazioni cloud e qualsiasi altra fonte di dati da cui si vuole imparare attraverso l'elaborazione analitica.

Una volta consolidati tutti i dati critici, noterete che i dati provenienti da diverse fonti sono datati e strutturati in diversi formati. In questa fase, i dati compilati devono essere organizzati in termini di data, dimensione e fonte per corrispondere al processo di trasformazione. È necessario un certo livello di coerenza in tutti i dati in modo che possano essere inseriti nel sistema e convertiti nella fase successiva. La complessità di questa fase può variare notevolmente a seconda del tipo di dati, del volume di dati e delle fonti di dati

Fasi di estrazione ETL

- Compilare i dati da fonti rilevanti.
- Organizzare i dati per renderli coerenti.

8.1.2 Fase 2 — Trasformazione

La trasformazione dei dati è la seconda fase del processo ETL nel datawarehouse. Qui i dati compilati vengono convertiti, riformattati e puliti nell'area di transito per l'inserimento nella base dati di destinazione nella fase successiva. La fase di trasformazione consiste nell'esecuzione di una serie di funzioni e nell'applicazione di un insieme di regole ai dati estratti, convertendoli in un formato standard per soddisfare i requisiti di schema del database di destinazione. Il livello di manipolazione richiesto nella trasformazione ETL dipende esclusivamente dai dati estratti e dai requisiti di business. Ciò include la convalida dei dati e il loro rifiuto se non sono accettabili.

Una buona fonte di dati non richiederà molte trasformazioni, mentre altri set di dati potrebbero averne bisogno in modo significativo. Per soddisfare i requisiti tecnici e commerciali del vostro database di destinazione, potete sottoporlo a diverse tecniche di trasformazione.

Fasi di trasformazione ETL

- Convertire i dati in base alle esigenze aziendali.
- Riformattare i dati convertiti in un formato standard per garantire la compatibilità.
- Ripulire i dati irrilevanti da insiemi di dati.
 - Ordinare e filtrare i dati.
 - Cancellare le informazioni duplicate.
 - Tradurre se necessario.

8.1.3 Fase 3 — Caricamento

Il passo finale è quello di caricare i set di dati precedentemente estratti e trasformati nel database di destinazione. Ci sono due modi per farlo. La prima è una routine di inserimento SQL che prevede l'inserimento manuale di ogni record in ogni riga della tabella del database di destinazione. L'altro approccio al caricamento utilizza un processo chiamato Bulk Data Loading, che è riservato al caricamento di dati alla rinfusa.

L'inserimento SQL può essere lento, ma esegue controlli di integrità ad ogni inserimento. Sebbene il caricamento alla rinfusa sia molto più veloce per il caricamento di enormi quantità di dati, non esegue controlli di integrità ad ogni inserimento. Il caricamento alla rinfusa è l'ideale per set di dati che siete sicuri siano privi di errori.

Fasi di carico ETL

- Caricare set di dati ben trasformati tramite caricamento alla rinfusa.
- Caricare set di dati discutibili tramite inserti SQL.

8.2 Componenti principale di Informatica Intelligent Cloud Service

Informatica Intelligent cloud service consente di creare, testare ed eseguire flussi di dati per il popolamento del datawarehouse.

Quando ci si connette all'IICS la versione cloud dell'informatica, vengono offerti diversi servizi. Il servizio a cui siamo interessati nell'ambito del progetto è il servizio data integration [12].

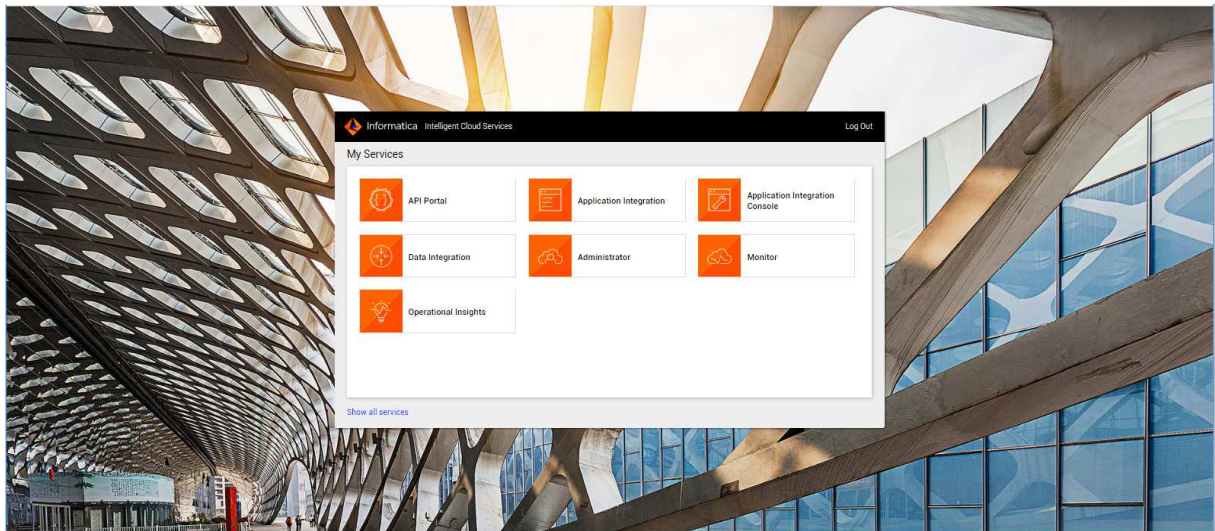


Figure 8.2: Interfaccia di informatica Cloud

Quando si seleziona la sezione Integrazione dei dati, viene presentata una dashboard che fornisce i punti principali della propria organizzazione dal punto di vista dello sviluppatore. Controllare la seguente screenshot per vedere i riferimenti

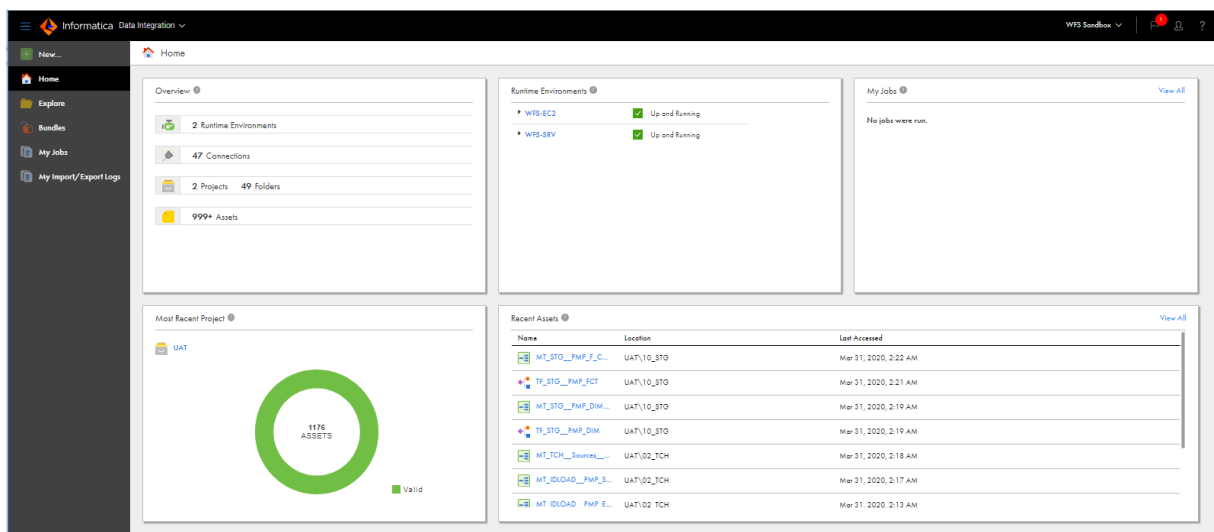


Figure 8.3: Dashboard di Informatica Cloud

Esaminiamo i componenti presenti nella dashboard.

1. La sezione **Overview** mostra i componenti attualmente attivi nella struttura, come il numero di agenti di sicurezza configurati, il numero di connessioni create. Rappresenta anche il numero di progetti, asset, ecc.
2. La sezione **Runtime Environments** elenca gli agenti sicuri attualmente disponibili per l'organizzazione o l'account. Elenca tutti gli agenti sicuri in

esecuzione e non in esecuzione. È possibile fare clic sul singolo agente di sicurezza per visualizzarne i dettagli.

3. La sezione **My Jobs** visualizza l'elenco della mappatura ETL o dei compiti già creati. Al momento non abbiamo imparato a creare i posti di lavoro. Man mano che andiamo avanti, potrete vedere l'elenco qui.
4. La sezione **Most Recents Project** elenca i dettagli dell'ultimo progetto su cui si è lavorato. **Default** è il progetto predefinito creato da Informatica Cloud. Si può aggiungere nuovi Progetti.
5. **Recent Assets** visualizza l'elenco delle attività che sono state create. Per asset in informatica cloud si intendono fondamentalmente le mappature ETL o i task.

Con questo abbiamo descritto la dashboard o la console per il design dell'ETL.

Assets- Mappings or tasks in informatica cloud

Come detto in precedenza, l'Asset non è altro che Mappings o Tasks che rappresentano i flussi ETL e l'implementazione del processo del Business. Diversi tipi di Asset possono essere creati in Informatica cloud. Discuteremo di ognuno di essi in dettaglio. Per creare un nuovo Asset che sia un nuovo lavoro o un nuovo processo ETL, cliccate su Nuovo come si vede nell'immagine qui sotto:

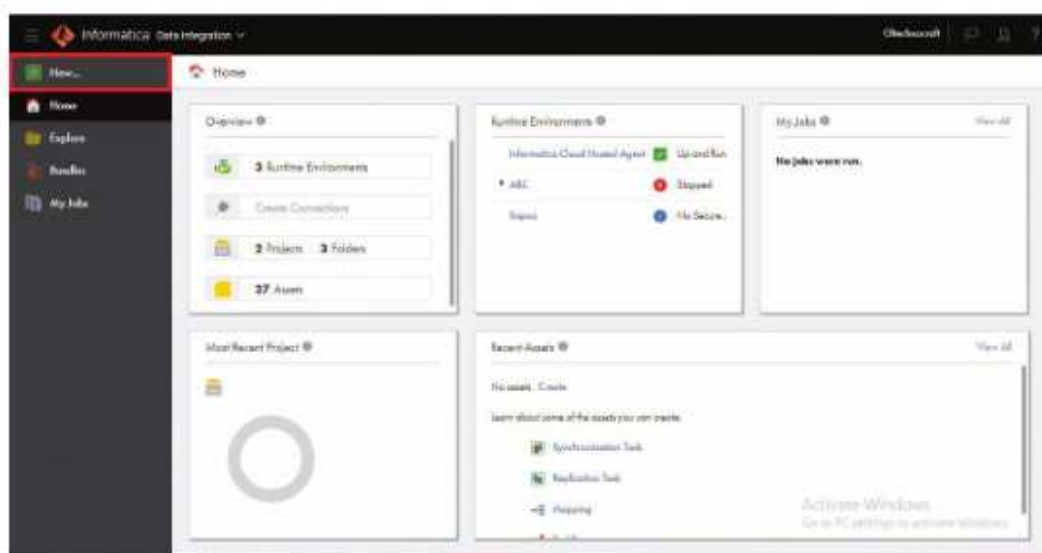


Figure 8.4:Interfaccia per la creazione di nuovi asset

Una nuova finestra pop-up consente di selezionare il tipo di asset che si desidera creare, come mostrato nell'immagine riportata di seguito:

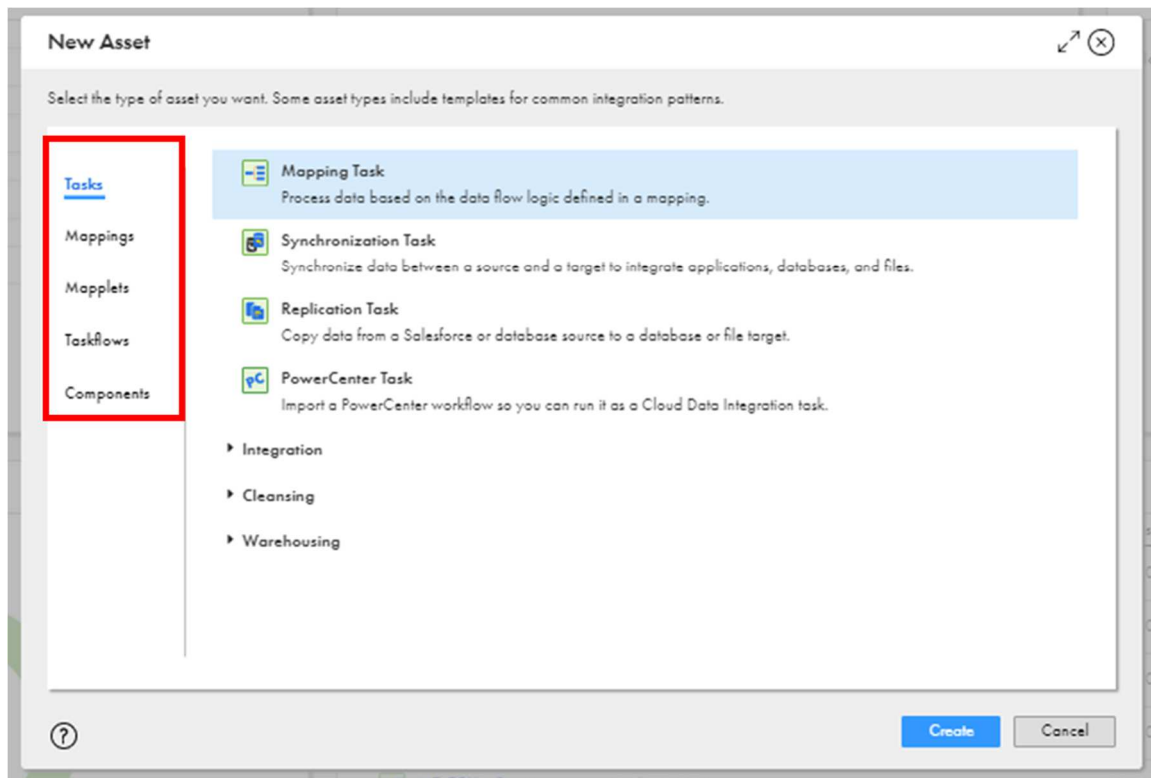


Figure 8.5:Tipologie di asset in Informatica Cloud

Come detto in precedenza, in Informatica Cloud è possibile creare diversi tipi di asset.

I tipi di asset disponibili sono di diversi tipi:

1. Tasks
2. Mappings
3. Taskflows
4. Components

Discuteremo in dettaglio di ogni tipo di asset nel libro. In informatica cloud è disponibile una funzione extra che consente di eseguire direttamente il *mappings* senza la necessità di creare una *tasks*.

Discutiamo in dettaglio i diversi tipi di tasks. Parleremo prima di tutto del Mappings degli asset.

Mappings

Questa sezione consente di creare processi ETL. Nella sezione di Mappings è possibile definire la sorgente, il target e la trasformazione. Il Mapping una volta salvato e validato possono essere eseguite per ottenere il risultato desiderato.

I vari tipi di mapping che possono essere creati sono visibili cliccando sui mapping. Una volta cliccato su Mappings, si visualizzeranno diversi tipi o modelli disponibili, come mostrato nella schermata sottostante:

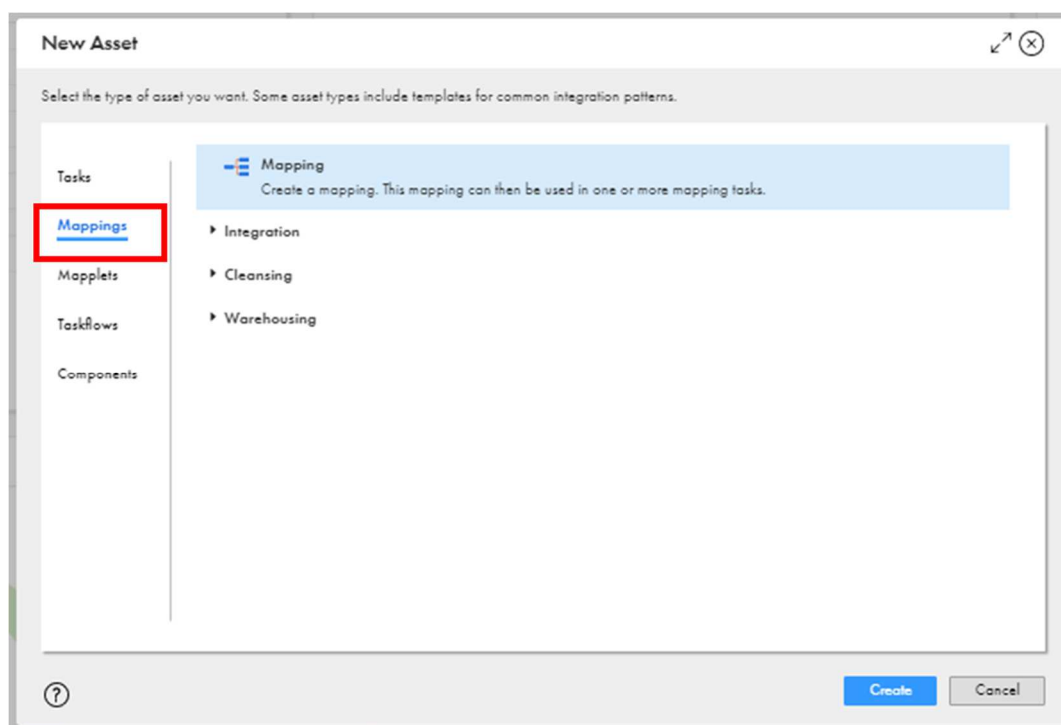


Figure 8.6: Interfaccia per la creazione di nuovi asset di tipo Mapping

Vari tipi di Mapping sono modelli predefiniti disponibili in informatica Cloud, che possono essere utilizzati come riferimento per costruire i processi ETL. Le mapping disponibili sono:

- a. **Mapping:** È possibile utilizzare questa opzione per creare un mapping da zero aggiungendo Source, Target e varie trasformazioni
- b. **Integration:** Questa opzione permette di selezionare vari modelli in base ai quali costruire i propri mapping. Come join due sorgenti con inner join.

c. **Cleansing:** Questa opzione permette di selezionare i modelli che ci aiuteranno a pulire i dati dai vostri sistemi di dati, come la rimozione di nulli o la rimozione di duplicati.

d. **Warehouse:** Questa opzione fornisce i modelli per varie mappature relative all'implementazione di concetti di Data Warehousing come la creazione di mappature SCD1, SCD2 ecc.

Tasks

I task sono una forma di processi ETL, ma con caratteristiche limitate che possono essere programmate per funzionare ad un determinato intervallo di tempo. I Task sono generalmente utilizzati per eseguire i Mappings creati sotto la sezione Mappings asset e la funzionalità di sincronizzazione o di Replica. Se si desidera utilizzare più trasformazioni o è necessario creare scenari complessi nella propria logica, è necessario creare prima i Mappings come descritto sopra in Mapping Asset.

Per creare un nuovo Task, fare clic su Tasks come mostrato nella schermata seguente:

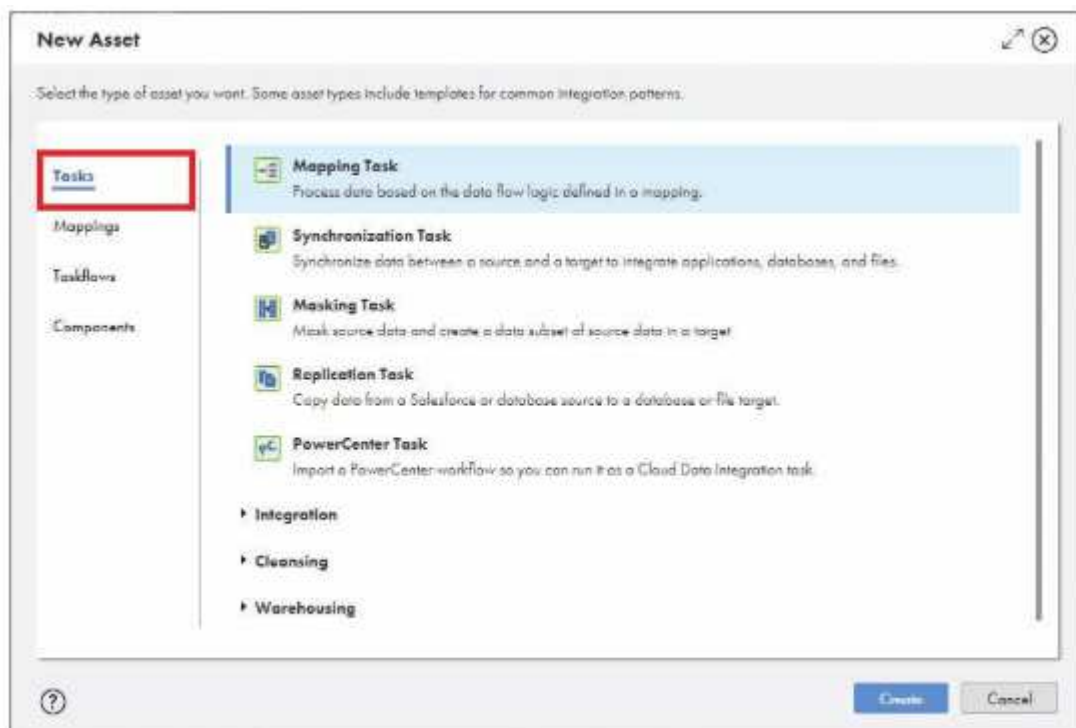


Figure 8.7:Interfaccia per la creazione di nuovi asset di tipo Tasks

Diversi task che possono essere creati sono:

a. **Mapping Task:** I processi ETL creati sotto gli asset di Mappings possono essere usati per collegarsi in Mappings Task ed essere eseguiti secondo una particolare programmazione.

b. **Synchronization Task:** Questo asset può essere utilizzato per sincronizzare i dati tra due sistemi di dati come Oracle, Salesforce, file ecc. Puoi usare questa opzione se si vuole sincronizzare i dati da Salesforce a Oracle ecc. Quando non c'è bisogno di complesse manipolazioni logiche ma solo di sincronizzare i dati dal sistema sorgente al sistema di destinazione, Synchronize Task si adatta meglio. Il compito di sincronizzazione offre funzioni di manipolazione di base.

c. **Replication Task:** Questo task può essere usato per replicare i dati da un sistema sorgente ad un altro. Simile al compito di sincronizzazione, il compito di replica non offre funzioni di trasformazione complesse.

d. **PowerCenter Task:** Usare questa sezione per eseguire un Mappings creata in Power Center. Qui si può importare il Mappings che è stata esportata da Power Center. Se si pensa di migrare da PowerCenter a Informatica Cloud, questo Task può essere usato per importare tutti i Mapping da PowerCenter a Informatica Cloud.

e. **Mass Ingestion:** Questa opzione può essere usata per trasferire file di grandi dimensioni dal sistema sorgente al sistema di destinazione. Compiti simili possono essere creati per l'integrazione, la pulizia e le esigenze del warehouse.

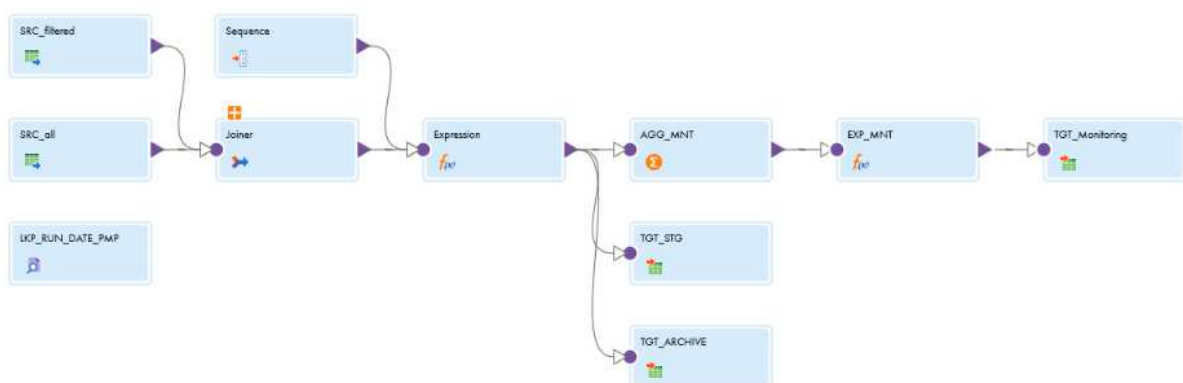


Figure 8.8:Esempio di Mapping task in Informatica Cloud

Taskflow

Come accennato in precedenza, i Taskflow sono una combinazione di più task creati nell'ambito dei Tasks assets disposti in sequenza per ottenere particolari funzionalità di business. Per creare un nuovo Taskflow cliccate su Taskflow come mostrato nella schermata seguente:

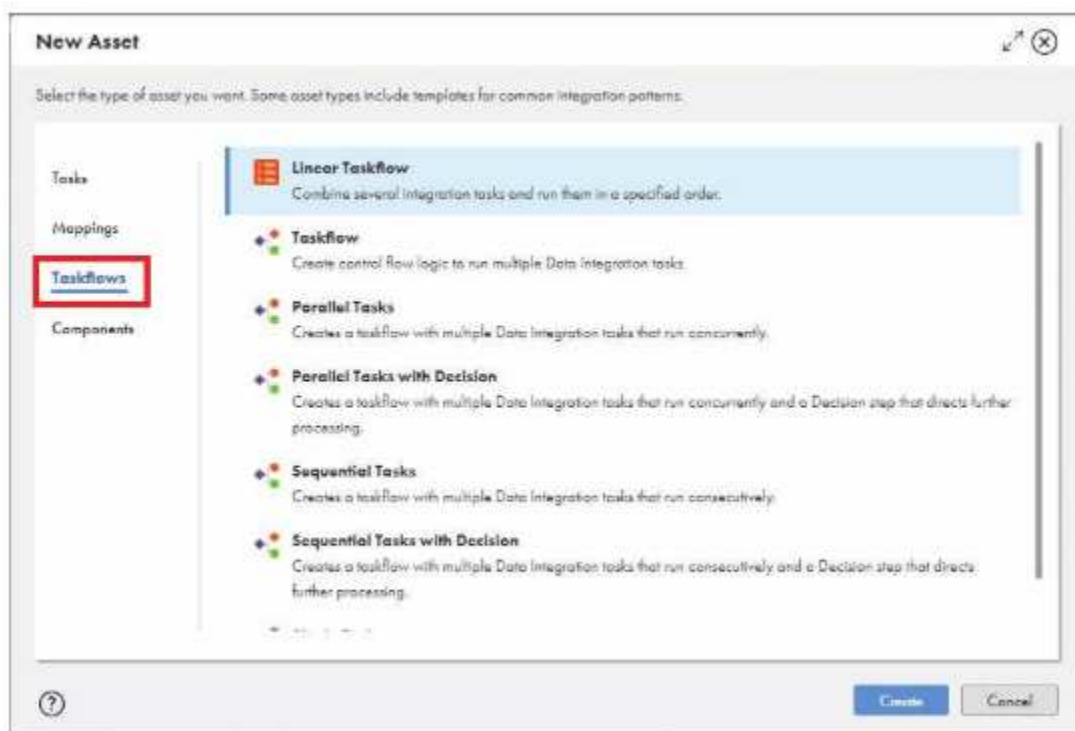


Figure 8.9:Schermata per la creazione dei Taskflow

I task flow sono principalmente divisi in due sottosezioni:

a. Linear taskflow: Questa opzione permette di associare più compiti tra loro. È possibile aggiungere più attività in un taskflow lineare da eseguire secondo la sequenza necessaria per gli scenari di business. È necessario semplicemente aggiungere un'attività e una pianificazione. Il taskflow lineare eseguirà le attività una dopo l'altra.

b. Taskflow: Questa opzione consente di creare flussi di business complessi. È possibile progettare scenari complessi come l'aggiunta di task da eseguire in parallelo, l'esecuzione di task basati su un particolare controllo delle condizioni, ecc. In parole semplici si può dire, Taskflow permette di creare un complesso Linea Taskflow. Ci sono

altri modelli disponibili come Attività parallele, Attività parallele con decisione, Attività sequenziali, ecc.

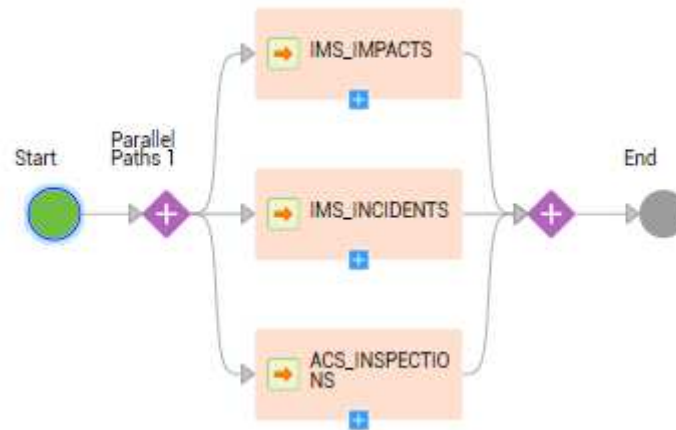


Figure 8.10:Esempio di taskflow

Components

Questa opzione permette di creare componenti come Mapplet o di salvare le query SQL che possono essere utilizzate nella sezione Tasks o Mappings. La maggior parte dei componenti di questa sezione sono di alto livello e richiedono la comprensione di diversi aspetti come Web Services, file XML, formati JSON e quindi fuori dallo scopo del libro. Questi sono componenti usati raramente

Per creare un nuovo componente, fare clic sui componenti come mostrato nella screenshot qui sotto:

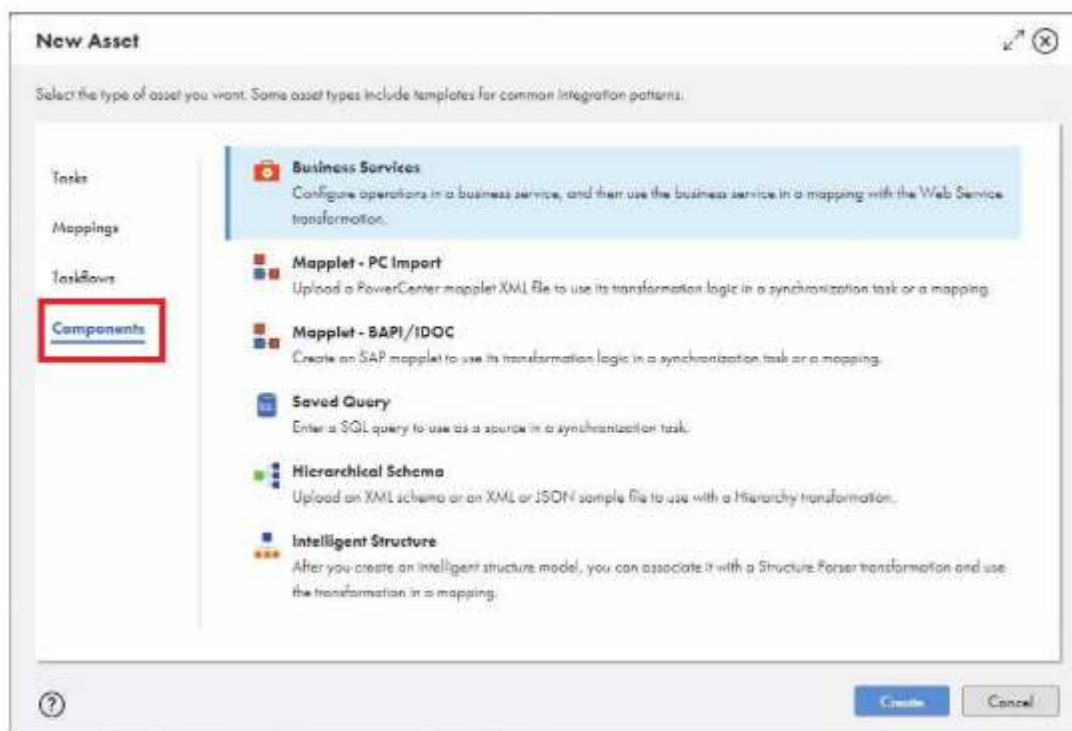


Figure 8.11: Interfaccia di creazione di asset di tipo components

Con questo abbiamo ora una comprensione di alto livello su vari Assets che possono essere creati in Informatica Cloud.

Fase di estrazione

La fase di estrazione consiste nell'acquisizione dei dati da numerose fonti tra loro eterogenee: possono infatti basarsi su tecnologie diverse e presentarsi sia come fonti relazionali, sia non relazionali. La fase di estrazione deve tener conto delle caratteristiche della base di dati sorgente e dei singoli campi da estrarre. Capita spesso, infatti, che vi siano volumi ragguardevoli di record che non è necessario processare; è consigliabile quindi filtrare tali record a monte del processo e recuperare esclusivamente i campi che saranno effettivamente utilizzati nelle fasi successive in modo da ridurre notevolmente il carico di lavoro.

I dati estratti, vengono depositati in un'area temporanea, detta staging area, costituita generalmente da una specifica insieme di tabelle all'interno del datawarehouse o di un database relazionale separato. I dati che raggiungono l'area di staging, a meno di filtri sui campi e piccole conversioni dei metadati, costituiscono una copia esatta dei sistemi

sor-gente. Introdurre un'area di staging nel processo di realizzazione di un datawarehouse porta con sé alcuni vantaggi significativi:

- Indipendenza dai sistemi sorgenti. In caso di variazioni nei sistemi sorgenti, le modifiche sono limitate alla fase di estrazione e alla staging area, lasciando inalterata la struttura del datawarehouse e minimizzando quindi il disservizio percepito dall'utente finale.
- Caricamento delta (o incrementale). Viene effettuato il caricamento nel datawarehouse dei soli record non ancora elaborati o che hanno subito variazioni rispetto all'ultimo caricamento.
- Separazione dei processi di estrazione e trasformazione. Ogni flusso dati proveniente dai sistemi sorgenti deve essere estratto e depositato nell'area di staging prima di procedere con le fasi successive; in questo modo le procedure di trasformazione sono eseguite localmente e risultano indipendenti da eventuali malfunzionamenti dei sistemi sorgenti.

Dall'area di staging inizia il vero e proprio processo di ETL che porta al popolamento del datawarehouse. In questo paragrafo saranno presentate, per ciascun sistema sorgente, le procedure di estrazione dati ritenute più interessanti.

8.3 Principio di alimentazione del Datawarehouse

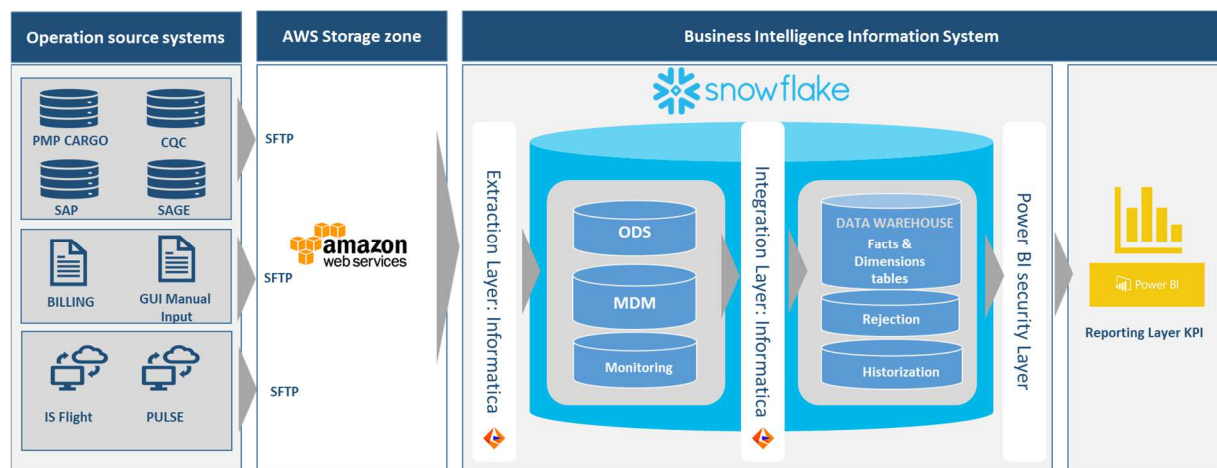


Figure 8.12: Flusso di alimentazione del datawarehouse

8.3.1 I vari tipi di zone di stoccaggio

- La zona di stoccaggio dell'AWS S3 è permanente e contiene:

- Tutti i file inviati dai paesi dai loro sistemi sorgente di operazioni o dal Web Gui, per mezzo di un SFTP
- La zona di preparazione contiene:
 - L'ODS, che contiene tabelle senza trasformazione o join. Sono tabelle che contengono tutti i file che arrivano nella S3.
 - Gli schemi MDM (Mappatura e tabelle di riferimento): per standardizzare i dati all'interno della WFS Holding.
 - La tabella di monitoraggio: utilizzata per caricare solo informazioni incrementali
- Il datawarehouse (DWH) contiene
 - Tabelle dei fatti e delle dimensioni
 - La tabella dei fallimenti
 - La tabella di storicizzazione

L'AWS S3

Questa zona di memorizzazione permette di caricare i dati in modo identico ai dati del sistema sorgente dell'operazione. Essa libera i sistemi di operazione utilizzati dagli utenti finali.

In questa fase non viene effettuata alcuna trasformazione. Tutti i dati vengono caricati senza trasformazione. Per default, gli attributi funzionali, anche se non utilizzati in una prima fase, vengono caricati per recuperare facilmente i dati senza influenzare i sistemi sorgente.

L'area di memorizzazione dell'AWS S3 è permanente e storicizzata, il che permette di fare questo:

- Conservare un archivio dei file di dati provenienti dai sistemi sorgente.
- Alimentare le tabelle ODS
- Caricare il DWH solo con i nuovi dati provenienti da flat file o dal database sorgenti.

ODS

L'ODS: una memorizzazione storicizzata dei dati di ingresso

L'ODS viene fornito dall'AWS S3. Il suo scopo è quello di conservare per un po' di tempo un'immagine dei dati provenienti dai sistemi sorgente e di recuperarli senza influenzare questi sistemi.

Datawarehouse

Il Datawarehouse (DWH): memorizzazione in un modello di dati dimensionali (tabelle di fatti e dimensioni) Il DWH è la base dell'attività di reporting. I dati vengono consolidati nella loro forma definitiva durante il trasferimento dall'ODS al DWH attraverso lo schema MDM e sono pronti per il reporting. Il feed DWH viene effettuato utilizzando tutte le regole fornite da WFS Holding.

8.3.2 Regole di alimentazione

Data flow

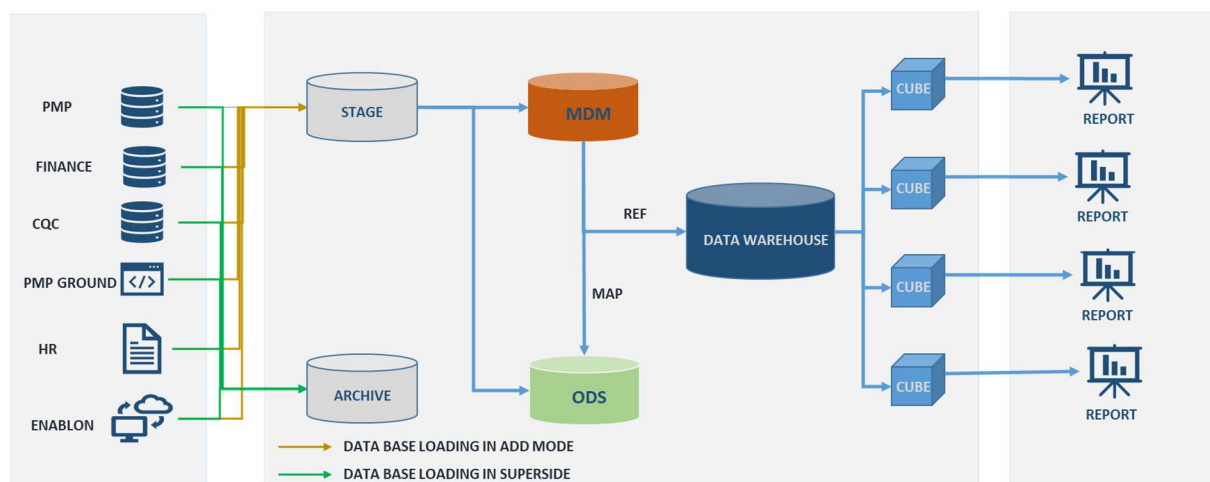


Figure 8.13:Flussi di dati nella piattaforma di datawarehouse

L'alimentazione avviene attraverso tre fasi distinte e indipendenti. Il rigetto di alcuni dati in ogni fase non impedisce l'alimentazione parziale della fase successiva. I dati respinti verrebbero caricati dopo la pulizia.

Alimentazione di archiviazione:

AWS S3

Durante questa fase, i dati provengono da varie fonti (database, flat file).

I file di sistema sorgente vengono inviati automaticamente (o manualmente nella fase di inizializzazione: estratto completo) attraverso un SFTP e memorizzati nell'AWS S3, all'interno di specifiche cartelle. Ogni paese ha una cartella individuale nel secchio S3, (idem per le cartelle PMP e SAP BFC per i file provenienti da questi database).

Non vi è alcuna richiesta diretta da parte delle banche dati.

Ogni giorno, alle 20.00 GMT, i job vengono eseguiti automaticamente per caricare i dati.

ODS

L'ODS è organizzato in tabelle e fornito da file AWS S3 in un processo incrementale. Vengono aggiunti solo nuovi dati. Non viene effettuata alcuna trasformazione tra AWS S3 e ODS. Durante questa fase c'è un processo di monitoraggio, che permette il tracciamento. MDM è fatto anche garantire l'integrità dei dati in tutta la WFS Holding.

DATA WAREHOUSE

Durante il trasferimento dei dati da ODS a DWH, i dati vengono trasformati - in base alle regole aziendali. Poi vengono storicizzati e consolidati definitivamente in DWH.

L'integrità referenziale è implementata nell'ETL per garantire la corrispondenza tra dati e riferimenti.

Quando il caricamento finale nel DWH ha successo, i file di input vengono spostati nella cartella degli archivi e conservati per 3 mesi. Ogni tre mesi, questi file vengono compressi e conservati nella memoria "glacier" S3 e i file non compressi vengono cancellati.

Dal punto di vista funzionale, vengono trattati cinque campi indipendenti:

- I dati relativi alla protezione della salute, della sicurezza, della società e dell'ambiente proverrebbero dalle seguenti fonti: Anaplan e Pulse
- I dati "Persone" proverrebbero dalle seguenti fonti: Sistema delle risorse umane (13 fonti diverse), PMP (compresi gli input manuali) e Anaplan
- I dati sulla soddisfazione del cliente provengono dalle seguenti fonti: CQC, IS Flight (3 diverse fonti), Anaplan e Web Gui (inserimento manuale)
- I dati di funzionamento dovrebbero provenire dalle seguenti fonti: IS Flight, Payrol, SAP Conso, PMP (incl. ingressi manuali) e Anaplan
- I dati relativi alla performance finanziaria dovrebbero provenire da: Anaplan, IS Flight, PMP, SAP Conso, sistemi di fatturazione (13 fonti diverse) e Web Gui (inserimento manuale)

Se c'è un problema di alimentazione su uno qualsiasi di questi cinque campi, solo questi campi relativi saranno impattati e non interferiranno con gli altri

8.4 Processo ETL: Specificazione e design delle mappings tasks per il caricamento dei dati

L'obiettivo dello strumento ETL, Informatica Intelligent Cloud Services (IICS o INFA), è quello di caricare i dati dalle 6 fonti descritte in precedenza e nel data lake su Snowflake.

In IICS gli oggetti sono localizzati e raggruppati in cartelle, ognuna delle quali rappresenta una fase dell'integrazione dei dati. Cioè ogni cartella contiene tutti gli oggetti necessari per una data fase per tutte le fonti.

<input type="checkbox"/>	Name	Type	Updated On	Location	Description
<input type="checkbox"/>	00_Global_TF	Folder	Aug 13, 2019, 4:33 PM	PREPROD	Taskflows of the whole ETL chain for each source category
<input type="checkbox"/>	02_TCH	Folder	Aug 13, 2019, 4:33 PM	PREPROD	technical parameters (ID_LOAD & SOURCES) used in 10_STG
<input type="checkbox"/>	10_STG	Folder	Aug 13, 2019, 4:33 PM	PREPROD	map source (external db or S3 files) to STG
<input type="checkbox"/>	15_STG_dim_to_MDS	Folder	Aug 13, 2019, 4:33 PM	PREPROD	map dim tables from STG to MDS
<input type="checkbox"/>	16_	Folder	Aug 28, 2019, 1:25 PM	PREPROD	run post-session scripts on EC2 for FILE & SAPFC
<input type="checkbox"/>	20_FLG	Folder	Aug 13, 2019, 4:33 PM	PREPROD	validate fact tables in STG
<input type="checkbox"/>	25_STG_fact_to_MDS	Folder	Aug 13, 2019, 4:33 PM	PREPROD	map fact tables from STG to MDS
<input type="checkbox"/>	30_STG_delete_ODS_and_DWH	Folder	Aug 13, 2019, 4:33 PM	PREPROD	use fact tables from STG to delete data in ODS and DWH
<input type="checkbox"/>	40_STG_fact_to_ODS	Folder	Aug 13, 2019, 4:33 PM	PREPROD	map fact tables from STG to ODS
<input type="checkbox"/>	42_MDM_SP	Folder	Aug 13, 2019, 4:33 PM	PREPROD	execute stored procedures to validate staging MAP tables in MDS
<input type="checkbox"/>	48_MDM_to_ODS	Folder	Aug 13, 2019, 4:33 PM	PREPROD	map validated entries in MAP tables from MDS to ODS
<input type="checkbox"/>	49_ODS_to_local	Folder	Aug 13, 2019, 4:33 PM	PREPROD	map MAP tables from ODS to local FF for lookups in 50_STG_to_DWH
<input type="checkbox"/>	50_STG_to_DWH	Folder	Aug 13, 2019, 4:33 PM	PREPROD	map fact tables from STG to DWH with lookups to FF
<input type="checkbox"/>	51_DWH_ID_UPDATE	Folder	Sep 13, 2019, 10:54 PM	PREPROD	
<input type="checkbox"/>	55_MDM_REF_to_DWH	Folder	Aug 13, 2019, 4:33 PM	PREPROD	copy REF tables from MDS to DWH

Figure 8.14:Lista dei tasks create durante il progetto

Gli oggetti sono principalmente di 3 tipi:

- **Mapping:** un mapping personalizzata da campi sorgente a campi target con possibili trasformazioni intermedie; non funziona da solo
- **Task di mapping:** per eseguire un mapping, si sceglie un ambiente di runtime, definire i parametri di mapping, se presenti. Un task di mappatura può includere i pre e post-comandi che vengono eseguiti nel prompt dei comandi.
- **Taskflow:** una sequenza di task di mappatura e/o altri taskflow con logica e ramificazione

Per convenzione:

- Mapping inizia con 'M_'.
- Mapping task iniziano con 'MT_'.
- I taskflow iniziano con 'TF_' (ad eccezione di quelli nella cartella '00_Global_TF')

I taskflow che dovrebbero essere eseguiti manualmente o pianificati si trovano in 00_Global_TF. Essi comprendono altri oggetti ETL (task di mappatura e taskflow) situati nelle altre cartelle e necessari per integrare una data sorgente nel data lake.

La Figura 8.13 mostra una panoramica di tali taskflow (chiamati LOAD_xxx dove xxx è la sorgente, ad esempio LOAD__CQC o LOAD__PMP). La maggior parte delle

sorgenti (alias categorie) differiscono da questo flusso generale in un modo o nell'altro che viene descritto più avanti nella sua stessa sezione.

In IICS i task vengono eseguiti in un ambiente runtime, che è o EC2 (un'istanza EC2) o SRV (on-premise server). Idealmente tutti i task dovrebbero essere eseguiti su EC2, ma PMP e MDS sono raggiungibili solo dietro il firewall WFS da SRV. I task che devono connettersi al database PMP e i task che leggono o scrivono su MDS devono essere eseguibili su SRV.

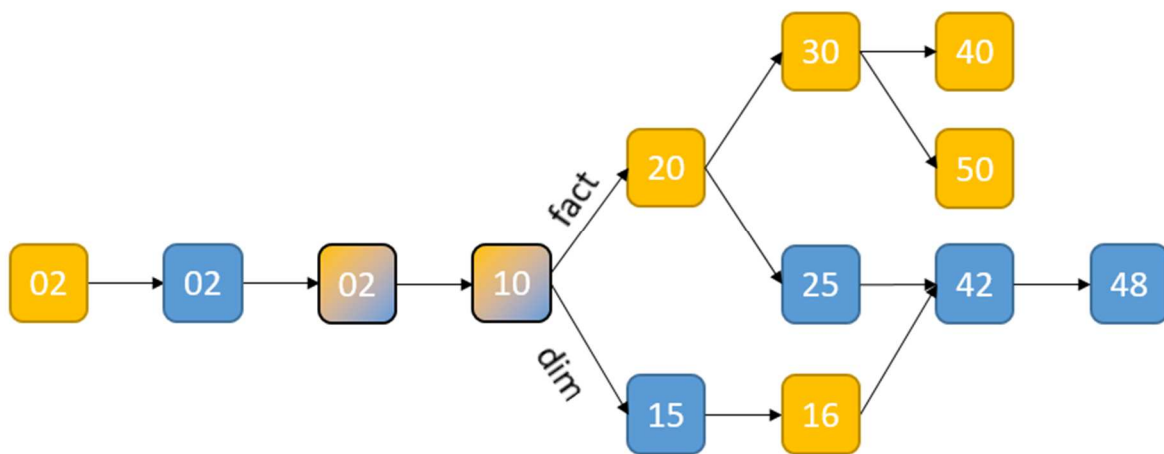


Figure 8.15:Paranoramica dei taskflow

Figura 8.1: Sequenza generale dei task per l'integrazione dei dati di una sorgente. I numeri mostrano la cartella dove si trovano gli oggetti. Il colore rappresenta l'ambiente di runtime (EC2: oro, SRV: blu). Ogni quadrato può rappresentare un singolo task o un insieme di task (taskflow o task in esecuzione parallela per tutte le tabelle).

N°	Description	SA	Category	Source	Target	Operation
02a	Generate new ID_LOAD	EC2	All	SQL query	ODS.TCH	Update
02b	Copy ID_LOAD	SRV	All	ODS.TCH	Flat file	Replace
02c	Retrieve technical parameters	SRV	PMP	ODS.TCH	Flat file	Replace
10	map a source (database table or S3 file) to a table in our STG database	EC2	FILE, CQC, SAPFC	Flat file on S3	STG	Truncate + insert
15	map a dimension table from STG to MDS	SRV	ANAPLAN	ANAPLAN	+	Insert
16	run post-session scripts to handle dimension files on S3	EC2	CQC	CQC		Insert
20	validate fact tables with business rules	SRV	PMP	PMP	ARCHIVE	
25	map fact tables from STG to MDS	SRV	All	STG	SSMDS.stg	Insert
30	use fact tables from STG to delete data in ODS and DWH	EC2	All but CQC	STG	ODS	Delete
40	map fact tables from STG to ODS	EC2	CQC	STG	DWH	Delete
42	execute stored procedures to validate staging MAP tables in MDS	EC2	All but CQC	STG	ODS	Insert
48	map validated entries in MAP tables from MDS to ODS	SRV	CQC	STG	ODS	Upsert
50	map fact tables from STG to DWH with lookups to FF	SRV	All	-	-	Stored procedure call
		SRV	All	MDS	ODS	Truncate + insert
		EC2	All but CQC	STG	DWH	Insert

Tableau 8.1: Sequenza generale dei task per l'integrazione dei dati di una sorgente

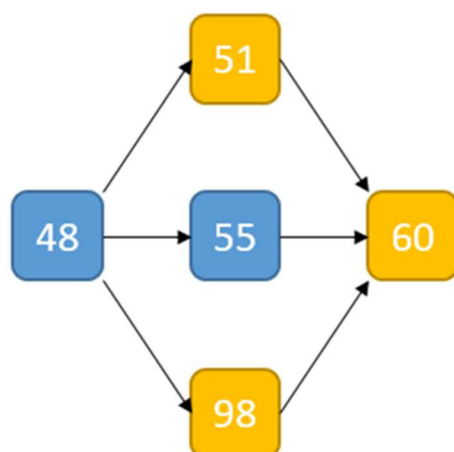


Figure 8.16: Esempio di sequenza di taskflow

N°	Description	SA	Category	Source	Target	Operation
48	map validated entries in MAP tables from MDS to ODS	SRV	All	MDS	ODS	Truncate + insert
51	update IDs in DWH fact tables	EC2	All	DWH	DWH	Update
55	copy REF tables from MDS to DWH	SRV	All	MDS	DWH	Upsert
98		EC2	FILE			
60	Call Snowflake stored procedures to generate cubes	EC2	All	-	CUBE ARCHIVE	+ Stored procedure call

Tableau 8.2: Sequenza dei task da 48 a 60

8.5 Processo ETL per il data mart Risorse umane, operazioni e finanze

8.5.1 Implementazione in informatica delle task flow.

Durante il corso del progetto è stato designato diversi task flow riusabile, presenteremmo in questa sessione in dettaglio il taskflow chiamato LOAD_XXX, XXX sta per il sistema sorgente. Questo taskflow è comune a tutti i sistemi sorgente da caricare. La figura (N° di figura) rappresenta il taskflow del sistema sorgente della base di dati operativa PMP.

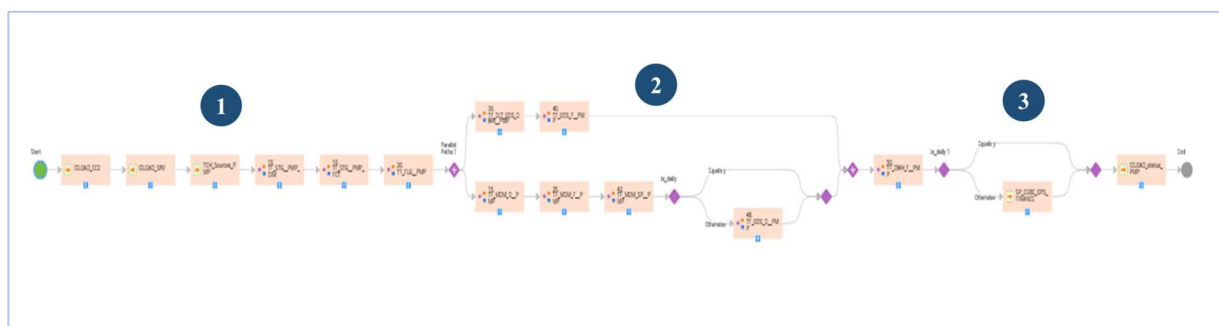


Figure 8.17: Esempio di taskflow per il caricamento dei dati di Cargo Operation

Questo taskflow a tre parte che spiegheremmo in dettaglio nel processo ETL di caricamento dei dati dell'entità cargo operation.

8.5.2 Taskflow del processo ETL dell'entità risorse umana

Molti aspetti sono simili a SAP in quanto consiste di file piatti su S3. Il dipartimento IT o HR di ogni paese invia i file sull'AWS S3 tra il 5 e il 10 di ogni mese. Da un punto di vista ETL, questa fonte ha alcune peculiarità sulla gestione dei file:

- Utilizza una tabella di configurazione (TCH.ETL_SOURCES) e script powershell per creare dinamicamente elenchi di file da recuperare da S3.
- Molti task hanno un post-command per spostare i file (da S3 a EC2, e tra le cartelle S3) nei passi 02, 16, 50.
- Il passo 20 è particolarmente importante per rilevare gli errori nei file creati manualmente. In caso di rifiuto viene inviata una mail.
- Esiste un rapporto per seguire le consegne dei paesi. I file consegnati sono contrassegnati con la lettera 'Y'. Prima del 10 del mese, i file non consegnati sono annotati 'P(fine)'. A partire dall'11, i file non consegnati sono contrassegnati con 'N'. I file con righe rifiutate sono contrassegnati con 'E'.
- I tipi di file 'Hires' e 'Leavers' possono essere vuoti. In tal caso viene inviata una mail e può essere necessario contrassegnare tale file come vuoto in MDS (tabella TCH_MONITORING_File_Uploads).

Global Taskflow: CARICARE IL FILE

Questo taskflow lancia tutte le attività di mappatura:

- Recuperare file da Amazon S3
- (genera lo script powershell usato per copiare e archiviare i file)
- Carico, dati in tabelle STG e tabelle ARCHIVIO
- Carico tabelle ODS da tabelle STG e MDM
- Carico tabelle DWH
- Carico, tabelle CUBO

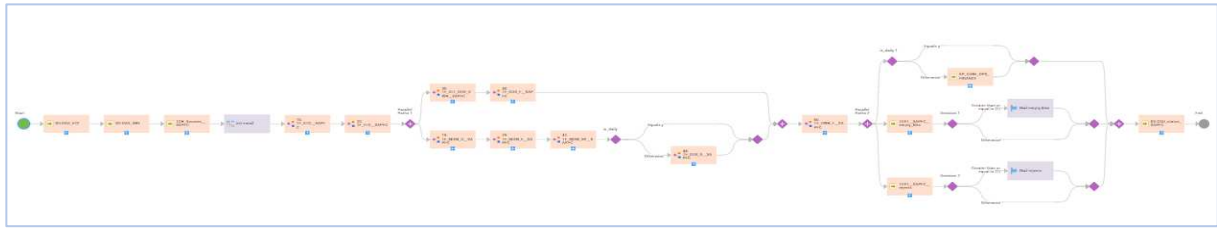


Figure 8.18: Processo ETL per l'entità risorse umane

8.5.3 Taskflow del processo ETL dell'entità Cargo Operation: PMP

Global Taskflow: LOAD__PMP

This taskflow launches all subtasks:

- Load data into STG tables
- Load ODS tables from STG and MDM tables
- Load DWH Tables
- Load Cube tables.

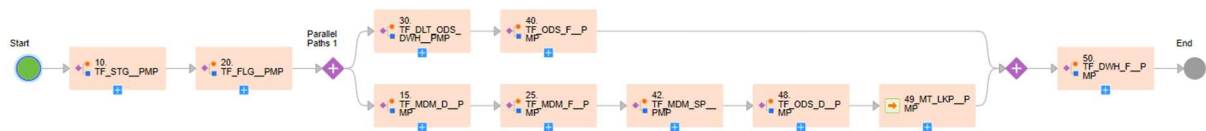


Figure 8.19: Processo ETL per l'entità Cargo Operation

Details task

Start:

IDLOAD_EC2

Generazione di un ID per la sessione in corso. Si tratta di un ID tecnico generato dal timestamp assegnato al processo all'avvio. Questo ID viene memorizzato in Snowflake -> BI_PROD -> TCH -> MONITORING (c'è un log per ogni task mapping eseguito).

IDLOAD_SRV

Copia l'ID generato dal task precedente in un flat file: RUN_DATE_PMP.txt utilizzando il collegamento CON_FF_SRV_INFA_LOOKUP (accessibile in INFORMATICA -> CONNECTION)

TCH_SOURCES_PMP

Attività di mappatura memorizzata in PROD\02_TCH

Questo oggetto elenca tutte le diverse fonti di dati nel nostro caso: SOURCE_CATEGORY= PMP.

Da qui recuperiamo il valore della colonna RETRIEVE_DEPTH_DAYS nella tabella ETL_SOURCES schema TCH Database BI_PRD

Questo valore è il numero di giorni che vogliamo estrarre. Attualmente il valore è 2

Ma questa è una data tecnica: questo significa per esempio che se il 03 aprile qualcuno carica 6 mesi di dati in PMP, questi dati saranno caricati in Snowflake dall'esecuzione automatica del processo INFORMATICA il 04 aprile. Nel caso in cui questi dati debbano essere caricati immediatamente la soluzione è quella di eseguire manualmente il Taskflow PROD -> 00_Gloabl_TF -> LOAD_PMP (cliccare con il tasto destro del mouse sui tre punti della linea & run)

Non è necessario troncare la tabella Fiocco di neve per "ricaricare" i dati. Ogni volta che il Taskflow è in esecuzione tutti i dati vengono aggiornati automaticamente.

10.TF_STG__PMP_DIM

Per copiare tutti i dati delle tabelle PMP DIM su Snowflake nello schema STG

Data Task

PMP_DIM_ACMODEL

PMP_DIM_ACTYPE

PMP_DIM_CARRIER

PMP_DIM_DATASOURCE

PMP_DIM_DEPARTMENT

PMP_DIM_HANDLINGDELIVERY

PMP_DIM_HANDLINGDIRECTION
PMP_DIM_HANDLINGPROCESS
PMP_DIM_HANDLINGPRODUCT
PMP_DIM_HOURTYPE
PMP_DIM_LOCATIONS
PMP_DIM_SCENARIO

10.TF_STG__PMP_FCT

Per copiare tutti i dati delle tabelle PMP FACT su Snowflake nello schema STG

Data Task

PMP_F_FLIGHTLEVEL_D
PMP_F_OPERATIONS_D
PMP_F_OPERATIONSBUDGET_M
PMP_F_TONSEOMANALYSIS_D
PMP_F_TONSEOMANALYSIS_M
PMP_F_WORKFORCE_D
PMP_F_WORKFORCEBUDGET_M
PMP_F_BAG_BAGSSORTED_D
PMP_F_BAG_QUALITY_D
PMP_F_GHA_ACTIVITIES_D
PMP_F_GHA_FLIGHTLEVEL_D
PMP_F_GHA_OPERATIONAL_D
PMP_F_GHA_QUALITY_D
PMP_F_GHA_TURNAROUNDS_D

20.TF_FLG__PMP

Per alcune tabelle dello schema STG, per verificare se i dati sono ben formattati, ad esempio per una data di tipo campo. Questo è più usato quando la sorgente dei dati è un file piatto come per esempio un file HR. Per esempio, controlliamo che la colonna HEADCOUNT sia un campo numerico.

Data Task

PMP_F_FLIGHTLEVEL_D

PMP_F_OPERATIONS_D
PMP_F_OPERATIONSBUDGET_M
PMP_F_TONSEOMANALYSIS_D
PMP_F_TONSEOMANALYSIS_M
PMP_F_WORKFORCE_D
PMP_F_WORKFORCEBUDGET_M
PMP_F_BAG_BAGSSORTED_D
PMP_F_BAG_QUALITY_D
PMP_F_GHA_ACTIVITIES_D
PMP_F_GHA_FLIGHTLEVEL_D
PMP_F_GHA_OPERATIONAL_D
PMP_F_GHA_QUALITY_D
PMP_F_GHA_TURNAROUNDS_D

A questo punto si deve eseguire due Subtaskflow in parallelo

Primo ramo

15.TF_MDM_D__PMP

Per copiare tutte le tabelle DIM da PMP e inserire o aggiornare le tabelle delle mappe (entità) in MDS

Data Task

DIM_ACMODEL
DIM_ACTYPE
DIM_CARRIER
DIM_DATASOURCE
DIM_DEPARTMENT

25.TF_MDM_F__PMP

Subtaskflow: Fare una mappatura delle tabelle dalla staging area STG a Microsoft master data MDS

42.TF_MDM_SP__PMP

Subtaskflow: eseguire procedure memorizzate per convalidare le tabelle MAP di staging in MDS

Secondo ramo

30.TF_DLT_ODS_DWH__PMP

Subtaskflow: Usare le tabelle dei fatti da SNOWFLAKE, DB:BI_PROD, schema: STG per cancellare i dati nello schema ODS e nello schema DWH

Data Task

PMP_F_FLIGHTLEVEL_D
PMP_F_OPERATIONS_D
PMP_F_OPERATIONS_BUDGET_M
PMP_F_TONS_EOM_ANALYSIS_D
PMP_F_EOM_EOM_ANALYSIS_M
PMP_F_WORKFORCE_D
PMP_F_WORKFORCE_BUDGET_M
PMP_F_BAG_BAGSORTED_M
PMP_F_BAG_QUALITY_D
PMP_F_GHA_ACTIVITIES_D
PMP_F_GHA_OPERATIONAL_D
PMP_F_GHA_QUALITY_D
PMP_F_GHA_TURNAROUNDS_D

40.TF_ODS_F__PMP

Subtaskflow: Fare una mappatura delle tabelle dalla staging area STG a ODS operational Data storage

Data Task

ODS__PMP_F_FLIGHTLEVEL_D
ODS__PMP_F_OPERATIONS_D
ODS__PMP_F_OPERATIONSBUDGET_M
ODS__PMP_F_TONSEOMANALYSIS_D
ODS__PMP_F_TONSEOMANALYSIS_M

ODS__PMP_F_WORKFORCE_D
ODS__PMP_F_WORKFORCEBUDGET_M
ODS__PMP_F_BAG_BAGSSORTED_D
ODS__PMP_F_QUALITY_D
ODS__PMP_F_GHA_ACTIVITIES_D
ODS__PMP_F_GHA_FLIGHTLEVEL_D
ODS__PMP_F_GHA_OPERATIONAL_D
ODS__PMP_F_GHA_QUALITY_D
ODS__PMP_F_GHA_TURNAOUNDS_D

48.TF_ODS_D__PMP

Subtaskflow: Fare una mappatura degli ingressi validi delle tabelle MAP da Microsoft master data a ODS operational Data storage

Chiusura dell'esecuzione in parallelo

50.TF_DWH_F__PMP

Subtaskflow: Fare una mappatura degli ingressi validi delle tabelle MAP da Microsoft master data a ODS operational Data storage

Task Flow	Mapping Task	Mapping	Folder	Source	Target
TF__step_by_step__PMP			00_GLOBAL_TF		
TF__step_by_step__PMP_groun			00_GLOBAL_TF		
TF__step_by_step__PMP_cargo			00_GLOBAL_TF		
	MT_STG_PMP_DIM_ACModel	M_STG_PMP_DIM_ACModel	10_STG	Dim_ACmodel	PMP_DIM_ACModel
	MT_STG_PMP_DIM_ACTYPE	M_STG_PMP_DIM_ACTYPE	10_STG	Dim_Actype	PMP_DIM_ACTYPE
	MT_STG_PMP_DIM_CARRIER	M_STG_PMP_DIM_CARRIER	10_STG	Dim_Carrier	PMP_DIM_CARRIER
	MT_STG_PMP_DIM_COSTACCOUNT	M_STG_PMP_DIM_COSTACCOUNT	10_STG	Dim_CostAccount	PMP_DIM_COSTACCOUNT
	MT_STG_PMP_DIM_DATASOURCE	M_STG_PMP_DIM_DATASOURCE	10_STG	Dim_Datasource	PMP_DIM_DATASOURCE
	MT_STG_PMP_DIM_DEPARTMENT	M_STG_PMP_DIM_DEPARTMENT	10_STG	Dim_Department	PMP_DIM_DEPARTMENT
	MT_STG_PMP_DIM_HANDLINGDELIVERY	M_STG_PMP_DIM_HANDLINGDELIVERY	10_STG	Dim_HandlingDelivery	PMP_DIM_HANDLINGDELIVERY
	MT_STG_PMP_DIM_HANDLINGDIRECTION	M_STG_PMP_DIM_HANDLINGDIRECTION	10_STG	dim_HandlingDirection	PMP_DIM_HANDLINGDIRECTION
	MT_STG_PMP_DIM_HANDLINGPROCESS	M_STG_PMP_DIM_HANDLINGPROCESS	10_STG	Dim_HandlingProcess	PMP_DIM_HANDLINGPROCESS
	MT_STG_PMP_DIM_HANDLINGPRODUCT	M_STG_PMP_DIM_HANDLINGPRODUCT	10_STG	Dim_HandlingProduct	PMP_DIM_HANDLINGPRODUCT
	MT_STG_PMP_DIM_HOURLTYPE	M_STG_PMP_DIM_HOURLTYPE	10_STG	Dim_HourType	PMP_DIM_HOURLTYPE
	MT_STG_PMP_DIM_INDICATOR	M_STG_PMP_DIM_INDICATOR	10_STG	Dim_Indicator	PMP_DIM_INDICATOR
	MT_STG_PMP_DIM_LOCATIONS	M_STG_PMP_DIM_LOCATIONS	10_STG	Dim_Locations	PMP_DIM_LOCATIONS
TF_STG_PMP_DIM	MT_STG_PMP_DIM_SCENARIO	M_STG_PMP_DIM_SCENARIO	10_STG	Dim_Scenario	PMP_DIM_SCENARIO
	MT_STG_PMP_F_COSTS_M	M_STG_PMP_F_COSTS_M	10_STG	F_Costs_M	PMP_F_COSTS_M
	MT_STG_PMP_F_COSTS_W	M_STG_PMP_F_COSTS_W	10_STG	F_Costs_W	PMP_F_COSTS_W
	MT_STG_PMP_F_FLIGHTLEVEL_D	M_STG_PMP_F_FLIGHTLEVEL_D	10_STG	F_FlightLevel_D	PMP_F_FLIGHTLEVEL_D
	MT_STG_PMP_F_OPERATIONS_D	M_STG_PMP_F_OPERATIONS_D	10_STG	F_Operations_D	PMP_F_OPERATIONS_D
	MT_STG_PMP_F_OPERATIONS_BUDGET_M	M_STG_PMP_F_OPERATIONS_BUDGET_M	10_STG	F_OperationsBudget_M	PMP_F_OPERATIONS_BUDGET_M
	MT_STG_PMP_F_TONSEOMANALYSIS_D	M_STG_PMP_F_TONSEOMANALYSIS_D	10_STG	F_ToncEOMAnalysis_D	PMP_F_TONSEOMANALYSIS_D
	MT_STG_PMP_F_TONSEOMANALYSIS_M	M_STG_PMP_F_TONSEOMANALYSIS_M	10_STG	F_ToncEOMAnalysis_M	PMP_F_TONSEOMANALYSIS_M
	MT_STG_PMP_F_WORKFORCE_D	M_STG_PMP_F_WORKFORCE_D	10_STG	F_Workforce_D	PMP_F_WORKFORCE_D
	MT_STG_PMP_F_WORKFORCE_BUDGET_M	M_STG_PMP_F_WORKFORCE_BUDGET_M	10_STG	F_WorkforceBudget_M	PMP_F_WORKFORCE_BUDGET_M
	MT_STG_PMP_F_BAG_BAGSORTED_D	M_STG_PMP_F_BAG_BAGSORTED_D	10_STG	F_BAG_BagsSorted_D	PMP_F_BAG_BAGSORTED_D
	MT_STG_PMP_F_BAG_QUALITY_D	M_STG_PMP_F_BAG_QUALITY_D	10_STG	F_BAG_Quality_D	PMP_F_BAG_QUALITY_D
	MT_STG_PMP_F_GHA_ACTIVITIES_D	M_STG_PMP_F_GHA_ACTIVITIES_D	10_STG	F_GHA_Activities_D	PMP_F_GHA_ACTIVITIES_D
	MT_STG_PMP_F_GHA_FLIGHTLEVEL_D	M_STG_PMP_F_GHA_FLIGHTLEVEL_D	10_STG	F_GHA_FlightLevel_D	PMP_F_GHA_FLIGHTLEVEL_D
	MT_STG_PMP_F_GHA_OPERATIONAL_D	M_STG_PMP_F_GHA_OPERATIONAL_D	10_STG	F_GHA_OperationsL_D	PMP_F_GHA_OPERATIONAL_D
	MT_STG_PMP_F_GHA_QUALITY_D	M_STG_PMP_F_GHA_QUALITY_D	10_STG	F_GHA_Quality_D	PMP_F_GHA_QUALITY_D
	MT_STG_PMP_F_GHA_TURNAROUNDS_D	M_STG_PMP_F_GHA_TURNAROUNDS_D	10_STG	F_GHA_Turnarounds_D	PMP_F_GHA_TURNAROUNDS_D

Tableau 8.3: Tabella dettagliata del processo ETL dell'entità Cargo Operation

8.5.4 Taskflow del processo ETL dell'entità l'entità Finanza: SAPFC

Global Taskflow: LOAD_SAPFC

This task flow launches all subtasks:

- Retrieve files from Amazon AWS S3
- (Generate powershell script used to copy and archive files)
- Load data STG tables
- Load ODS tables from STG and MDM tables
- Load DWH Tables
- Load Cube tables.

Questo flusso di attività lancia tutti i subtask:

- Recupera i file da Amazon AWS S3
- (Genera script powershell usato per copiare e archiviare i file)
- Caricamento dei dati nelle tabelle STG
- Caricamento delle tabelle dell'ODS da tabelle STG e MDM
- Caricamento delle Tabelle del DWH
- Caricamento Tabelle del Cube.

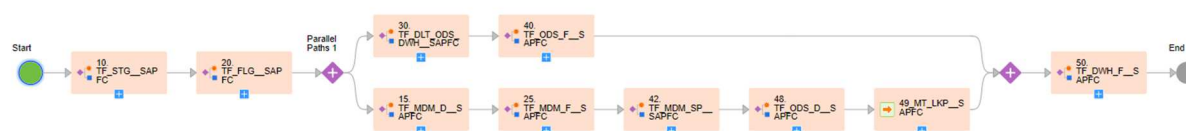


Figure 8.20:Processo ETL per l'entità finance

8.5.5 Esecuzione delle procedure ETL

Le operazioni di estrazione, trasformazione e caricamento, descritte nei paragrafi precedenti, devono essere eseguite periodicamente così da poter mantenere costantemente aggiornato il cloud datawarehouse. La frequenza di aggiornamento e le modalità di esecuzione di tali procedure sono state definite attraverso lo strumento *Informatica cloud* introdotto nel Paragrafo 8.2 e descritte di seguito. Tabella 8.4 fornisce, per ciascuna procedura, una breve descrizione, la frequenza di aggiornamento dei dati e le data e l'ora di avvio.

Procedura	Descrizione	Frequenza	Data e ora
00/LOAD_PMP	Procedura di importazione dai operativo dalla sorgente PMP	Una volta al giorno	20:05 GMT
00/BATCH_DAILY_HR	Procedura per il caricamento dei file SAPFC proveniente dalle risorse umane	Una volta al giorno	20:05 GMT
00/BATCH_DAILY_FINANCE	Procedura per il caricamento dei file SAPFC proveniente dalla finanza	Una volta al giorno	20:05 GMT
00/BATCH_DAILY_PMP_CUBE	Procedura per il caricamento dei cubi proveniente da PMP	Una volta al giorno	20:05 GMT
98/MT_USER_ACCESS	Procedura di esecuzione delle mapping task.	Ogni ora	

Tableau 8.4:Procedure automatizzate

Capitolo 9

Risultati e discussioni

Reportistica offline e online: la scelta di Power BI

Nel seguente capitolo sono descritte le logiche e le procedure utilizzate per la realizzazione dei report. Nel primo paragrafo sarà introdotto il concetto di reportistica per il supporto ai processi decisionali e mostrate le tecniche e le metodologie per la rappresentazione e distribuzione dei report. Saranno infine mostrati alcuni esempi di report e dashboard richiesti dall'azienda cliente.

9.1 Introduzione alla reportistica

Il reporting è l'attività di rilevazione e misurazione dei risultati effettuata periodicamente e la sua efficacia dipende dal tipo di strumenti impiegati e dalle metodologie adottate. Fornisce un insieme di informazioni approfondite e di dettaglio, organizzate per rispondere alle esigenze di ciascuna area di business o funzione operativa. Le sue finalità sono:

- Offrire al management la rappresentazione del presente per capire e influenzare il futuro dell'azienda.
- Assumere decisioni tattiche e strategiche con maggiore cognizione di causa.
- Confrontare i risultati effettivi con quelli attesi ed effettuare il calcolo degli scostamenti, nonché l'analisi delle cause di tali scostamenti.

Secondo [Marchi 13], i determinanti della qualità di un sistema di reporting sono: tempestività, affidabilità, chiarezza ed essenzialità. I report devono essere facilmente interpretabili e privi di ambiguità; le informazioni in essi contenute devono essere ridotte al minimo indispensabile fornite al management responsabile in tempi brevi. Per agevolare l'interpretazione dei risultati delle analisi risulta quindi fondamentale la modalità con cui i dati sono presentati. Esistono diversi strumenti per farlo, le tecniche di presentazione più comuni, secondo [11], sono:

- Report tradizionale, in cui il risultato è mostrato in maniera tabulare con eventuali livelli di dati di riepilogo parziali.
- Tabella a doppia entrata, in cui il risultato di una misura aggregata viene mostrato in corrispondenza della combinazione di due o più dimensioni lungo gli assi cartesiani. Aggiungendo e rimuovendo dimensioni si ottengono rispettivamente le operazioni di drill-down e roll-up.
- Grafici di natura diversa, come istogrammi, diagrammi a barra o grafici a torta.

Inoltre, come mostrato in Figura 8.1, si identificano tre modalità di attuazione del supporto.

alle decisioni:

- **Reportistica offline.** La reportistica offline, o standard rappresenta il livello più basso di supporto alle decisioni. I report generati possono essere statici, se presentati in un formato non manipolabile in alcun modo dall'utente finale, o dinamici, il cui contenuto può variare secondo parametri selezionabili.
- **Reportistica online.** La reportistica online, conosciuta anche con il nome di Analisi Multidimensionale, consente agli utenti finali di effettuare analisi interattive sui dati raccolti nel datawarehouse.
- **Analisi esplorativa.** Sono impiegati algoritmi di Data Mining al fine di estrarre modelli predittivi dei dati. Quest'ultimo insieme di tecniche non rientra tra le soluzioni di reportistica e, per questo motivo, non verrà approfondito nel corso di questo capitolo.

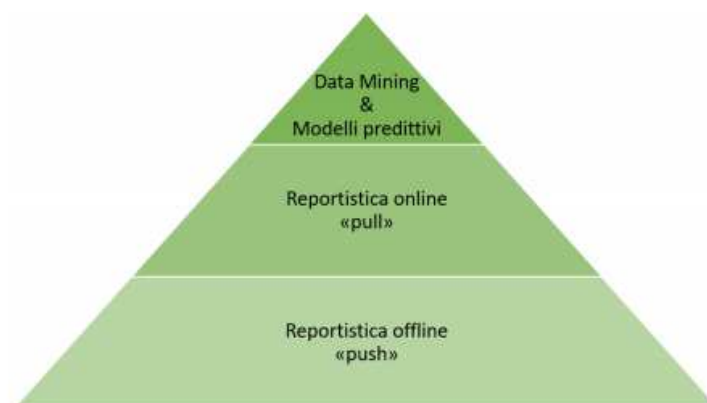


Figura 8.1 – Business Intelligence Pyramid

Un'altra classificazione riguarda le modalità con le quali sono distribuiti i report. Gli strumenti di reportistica offline offrono un sistema di sottoscrizione che permette agli utenti di ricevere periodicamente via e-mail i report desiderati, in una modalità detta push. La modalità pull, invece, è tipica della reportistica online: è l'utente finale che richiede la generazione di un report ad hoc attraverso gli strumenti OLAP a disposizione.

Nel caso in esame l'azienda cliente ha richiesto entrambe le modalità di generazione e distribuzione dei report; è stata quindi proposta una soluzione ibrida basata da un lato sulla piattaforma SQL Server Reporting Services e, dall'altro, sui software di analisi multidimensionale Microsoft Excel e Power BI.

9.1.1 La reportistica offline

Come strumento di reportistica offline è stato adottato il software SQL Server Reporting Services (SSRS). La progettazione dei report è stata realizzata per mezzo del componente Report Designer, un tool grafico che consente in pochi e semplici passaggi di generare i report desiderati. La gestione delle sottoscrizioni, nonché la pubblicazione e la schedulazione degli stessi, è avvenuta per mezzo del Report Manager. Di seguito, alcuni esempi di report generati.

9.1.2 La reportistica online

La reportistica online, a differenza di quella offline, consente agli utenti finali di creare o modificare con facilità i report desiderati, senza che siano necessarie approfondite conoscenze tecniche. È possibile inoltre utilizzare strumenti di pubblicazione per mettere a disposizione di altri utilizzatori i report creati. Come già anticipato nel Paragrafo 7.1, gli applicativi software adottati sono Power BI e SSRS.

9.2 BI Tool per il reporting

9.2.1 Microsoft Power BI

Power BI è un software di recente sviluppo, che, grazie ai grossi investimenti da parte di casa Microsoft, è riuscito a imporsi in breve tempo sui già presenti leader del settore, quali QlikSense e Tableau.

La marcia in più di Power BI sta senza dubbio, nella qualità e nell'interattività delle dashboard create. Gli utenti aziendali possono ottenere, in tempo reale, una panoramica a 360 gradi delle metriche più importanti accedendo da qualsiasi piattaforma desktop o mobile.

Esistono, infatti, tre versioni del software.

- *Power BI Desktop*: è uno strumento dotato di funzionalità avanzate per la creazione di report e mashup di dati. `E possibile combinare dati provenienti da più database, file e servizi Web con strumenti visivi utili per comprendere e correggere automaticamente i problemi di formattazione e qualità dei dati.

Power BI e Tableau si confermano, anche quest'anno, leader del settore.

- *Power BI per dispositivi mobili*: permette di monitorare e accedere ai dati aziendali creati dalla componente Desktop da qualsiasi dispositivo mobile dotato dell'app Microsoft Power BI (**Figura 9.1**).



Figure 9.1:Power BI Mobile nelle piattaforme mobili supportate

- *Power BI Online*: permette la visualizzazione e la condivisione delle dashboard aggiornate in tempo reale. Questo servizio `e fornito sia in versione free che in versione premium.

Il nostro studio sarà condotto sulla versione Desktop.

Microsoft Power BI

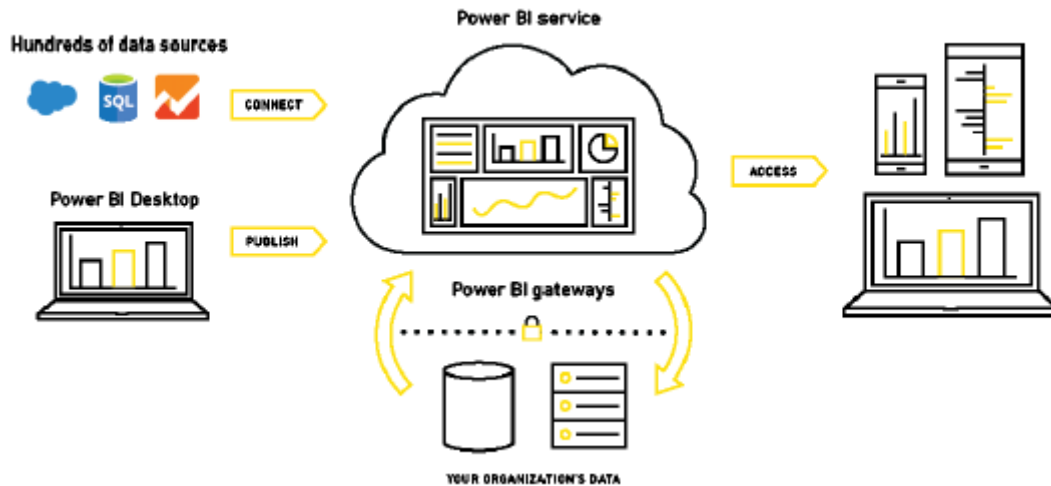


Figure 9.2:Ecosistema di Power BI

Funzionalità di Power BI

Power BI Desktop ci permette di importare un qualunque database per estrapolarne i dati di nostro interesse, elaborarli e ricavarne da essi informazioni e conoscenza.

Esso `e in grado di interfacciarsi a diversi tipi di sistemi (Figura 2.5):

- File system, quai Excel, CSV, XML, etc;
- sistemi relazionali, quali MySQL, Oracle, SQL Server, etc;
- sistemi cloud, quali Microsoft Azure, Google Analytics, Salesforce report, etc.



Panoramica di un foglio di lavoro

Una volta installata la versione Desktop di Power Bi dal sito <https://powerbi.microsoft.com/it-it/desktop/> siamo pronti ad iniziarlo e a cominciare il nostro studio.

Ci troveremo davanti un foglio con delle aree di lavoro ben definite (**Figura 2.6**). Le icone poste nell'estrema sinistra del foglio rappresentano le tre possibili “viste” (Figura 2.7); queste sono: Report, Dati e Relazioni tra tabelle.

Nella parte alta troviamo tutte le voci del menu di navigazione e tutte le varie opzioni ad esso collegate. Nella parte centrale troviamo la pagina in cui verranno elaborati e visualizzati i grafici. Sulla destra troviamo invece, il menu Visualizzazioni, con cui possiamo scegliere quale grafico utilizzare per visualizzare i dati, e il menu Filtri, che consente di definire opportunamente i valori del grafico per migliorare la visualizzazione.

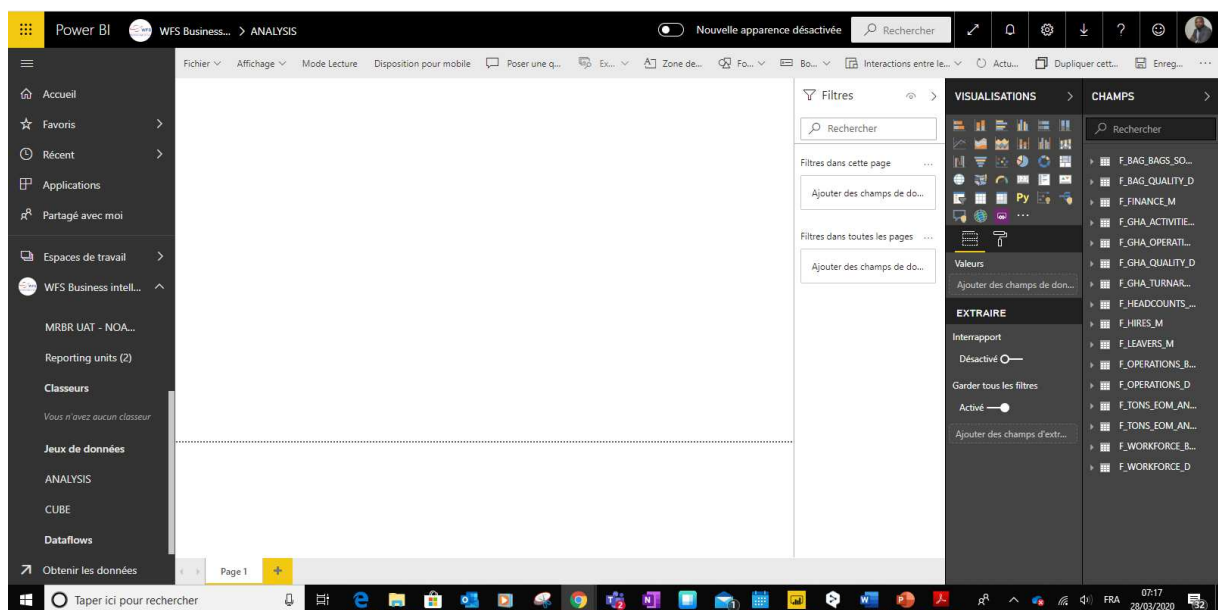


Figure 9.3: Pagina iniziale

Editor di query

Nella parte alta del foglio di lavoro è posizionato lo strumento più potente del software, ovvero il “Query editor”. Questo ci permette di effettuare operazioni di Data Shaping

sui dataset, e quindi, di modellare, filtrare ed elaborare i dati per adattarli alle nostre specifiche esigenze. Una volta cliccato sull'icona, l'editor verrà aperto in una finestra separata (**Figura 9.4**).

Nella parte destra della finestra troviamo il riquadro “Applied step”, dove verranno visualizzati, in ordine cronologico, tutte le operazioni effettuate su una data tabella (**Figura 9.5**).

Una volta completate le operazioni di querying sulle tabelle, è necessario cliccare sull'icona in alto a sinistra “Close & Apply” per salvare le modifiche apportate e chiudere l'editor.

Data	Arabi	Afghanistan	Angola	Albania	Andorra	Arab World	United Arab Emirates	Argentina	Armenia	American Samoa	Antigua and Barbuda
marcoledì 1 gennaio 1991	20				770	152281	48919	46590			700
venerdì 1 gennaio 1993			1100		780	198950	70588	112000			900
sabato 1 gennaio 1994			1824		784	205401	91488	241163			1200
domenica 1 gennaio 1996	1718		1994		2825	501551	126968	405395			1250
lunedì 1 gennaio 1996	3090		3298	2300	3458	921229	193834	667620	300		1300
martedì 1 gennaio 1997	8402		7052	3300	8618	1616551	803873	2007679	5000		1400
giovedì 1 gennaio 1998	5380		9820	5800	14117	2523542	995278	2670862	7831		1500
venerdì 1 gennaio 1999	12000		24000	11000	20600	4103008	832267	3848869	8161		1800
sabato 1 gennaio 2000	15000		25805	29781	23543	8055241	1478115	6487850	17485		1992
lunedì 1 gennaio 2001	53000		75000	392650	29429	16679594	1909303	6742791	25504		2156
martedì 1 gennaio 2002	81800	25000	140000	851000	32790	25273603	2429071	6360740	71345		2036
martedì 1 gennaio 2003	69952	200000	350000	1100000	51893	35121269	2972881	7842283	114573		2100
giovedì 1 gennaio 2004	98589	600000	740000	1255990	58966	52152873	3888117	13512385	205909		2250
sabato 1 gennaio 2005	103417	1200000	1611118	1530244	64560	84854521	4591443	22155426	318044		
domenica 1 gennaio 2006	104030	2520366	3054620	1908885	69004	125858922	5519293	31550390	1259762		
lunedì 1 gennaio 2007	113586	4668096	4961536	2321436	63503	174468139	7731508	40401771	1876411		
martedì 1 gennaio 2008	120806	7896909	6773858	1855612	84202	214082549	9357758	46508776	1443000		
giovedì 1 gennaio 2009	128000	10500000	8109421	2463741	64519	265276824	10871878	52482780	2191500		
venerdì 1 gennaio 2010	151800	10215840	9405985	2692372	65495	312247800	10928019	57082298	8853354		
sabato 1 gennaio 2011	11797879	12075218	11000000		65044	351857868	11727401	67222729	3211215		
domenica 1 gennaio 2012	115000	11340115	12785109	1500000	63855	361641859	11775252	64322647	3323817		
martedì 1 gennaio 2013	138800	16807138	13285196	3685383	63911	407704305	16063547	67361315	3346275		
martedì 1 gennaio 2014	139700	18407168	14052338	3359854	86241	418607952	16813024	61259218	3451237		
giovedì 1 gennaio 2015	141000	19709038	15884532	9400955	71338	422162187	17942360	61842011	3464490		
venerdì 1 gennaio 2016											

Figure 9.4:Query editor

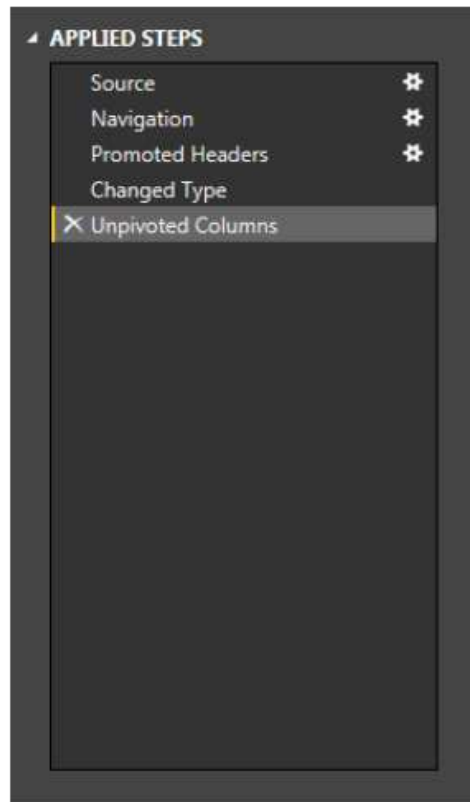


Figure 9.5: Cronologia delle query effettuate

Creazione di report e grafici supportati

Dopo aver elaborato i dati, occorre spostarsi sulla vista report tramite le icone a sinistra della pagina principale.

A questo punto, selezionando i campi di nostro interesse dal riquadro “Field”, possiamo scegliere il tipo di visualizzazione grafica. Power BI offre circa 20 visualizzazioni già preinstallate (Figura 9.6), ma molte altre sono reperibili dal Windows Store. I grafici comunemente usati e che sono più funzionali al nostro studio, sono:



Figure 9.6:Grafici preinstallati in Power BI Desktop

- *Istogrammi*: rappresentano una rappresentazione grafica di una funzione, nota solo per un certo numero di valori della variabile (Figura 9.7). Dunque, sono utilizzabili per analisi di dati che risultano essere definiti in intervalli discreti di valori.



Figure 9.7: Istogramma

- *Grafici a linee*: sono utilizzati principalmente per condurre analisi di tipo descrittivo (Figura 2.13). Ci permettono di descrivere in maniera dettagliata, l'andamento temporale di determinati valori. Attraverso la modulazione dei parametri che il grafico propone si possono effettuare studi anche di carattere predittivo.
- *Grafico a dispersione*: sono utilizzati generalmente per condurre analisi di tipo predittivo tramite l'uso di clusters. Nel grafico a dispersione almeno due variabili

di un dataset sono riportate su uno spazio cartesiano. I dati sono, poi, visualizzati tramite una collezione di punti che incrociano le coordinate cartesiane.

9.2.2 SQL Server Reporting Services (SSRS)

SQL Server Reporting Services è un software di generazione di report basato sul server Microsoft. Per lavorare con SSRS, sono disponibili diversi strumenti (componenti), a condizione che possiamo costruire rapporti, distribuire rapporti, gestire i permessi e visualizzare i rapporti.

SQL Server Data Tools (SSDT)

Business Intelligence Development Studio (BIDS) è il vecchio IDE utilizzato per sviluppare pacchetti SSIS (SQL Server Integration Services), rapporti SSRS o creare cubi SSAS (SQL Server Analysis Services). Ora si utilizzano gli SQL Server Data Tools (SSDT) per eseguire gli stessi compiti. Utilizzando SSDT è possibile creare una fonte di dati, set di dati e report condivisi. È possibile anche distribuire i rapporti al server utilizzando SSDT.

Report Builder

È lo strumento utilizzato dagli utenti per creare rapporti. Utilizzando il Report Generator, è possibile utilizzare le fonti di dati del Report Server, creare rapporti e pubblicare su Report Server se si dispone di un'autorizzazione. L'ultima versione disponibile per il Report Generator è la 3.0.

Report Server

Un'unica istanza del server di report è definita dalla raccolta completa di processori ed estensioni che forniscono un'elaborazione end-to-end dalla richiesta iniziale alla presentazione di un report finale. Attraverso i suoi sottocomponenti, il Report Server elabora le richieste di report e rende disponibili i report per l'accesso o la distribuzione come richiesto. Dal punto di vista funzionale, un Report Server consente la creazione di report, la consegna di report per una varietà di fonti di dati, nonché sistemi di autenticazione e autorizzazione estensibili.

Database report

Il server di report contiene database di report del server che memorizzano i report pubblicati, le fonti di dati condivise, i set di dati condivisi, le parti di report, i calendari e gli abbonamenti condivisi, i file di origine delle definizioni dei report, le definizioni dei modelli, i report compilati. Il server di report permette anche agli esperimenti di amministrazione di configurare il server di report per gestire le richieste di report e gestire i permessi per i report, le fonti di dati, i set di dati.

Report Manager

Il Report Manager è un'interfaccia web che può essere utilizzata per eseguire vari compiti come l'amministrazione del server di report, la visualizzazione dei report ecc.

9.3 Dashboard realisti come esito del progetto

Una volta che l'infrastruttura, il datawarehouse e la soluzione di integrazione dei dati sono stati impostati, le prossime sessioni presenterà i risultati del lavoro svolto a monte.

Ricordiamo che l'esigenza iniziale espressa dal top management è stata quella di impostare una serie di indicatori per monitorare l'evoluzione delle attività e della produttività di ogni dipartimento e di avere una visione globale delle performance aziendali.

I KPI iniziali, come mostrato nella figura sottostante, coprivano le esigenze globali del business, nell'ambito del nostro lavoro ci siamo limitati ai KPI del business Cargo e dei dipartimenti Cargo Operation, Finance e Human Resources. Il principio rimane lo stesso per gli altri dipartimenti.

HSSE					
1	Occupational Safety	CA/GH	23	% Number of Flights Delay SEVERITY 1 over Total Operated Flights	GH
2	Lost time injuries /200k hours worked (WFS staff only, all hours work)	CA/GH	24	% Number of Flights Delay SEVERITY 2 over Total Operated Flights	GH
3	Work days lost (absence) due to work related injury	CA/GH	25	% Flights First Baggage Delivery On Target (25 after ATA)	GH
4	Aircraft damage	GH	26	% Flights Last Baggage Delivery On Target (XX after ATA)	GH
5	Flight Safety	CA/GH	27	Lost Baggage Index Lost Bags/Departed Passengers	GH
6	Aviation Security	CA/GH	Operations		
7	Asset Protection	CA/GH	28	Trend - Number of Turnarounds	GH
8	Equipment Damage	CA/GH	29	Trend - Number of Handled Tons	CA
9	Environment	CA/GH	30	Kilos per Hour	CA
10	Number of HSSE 'self-inspections' conducted	CA/GH	31	AWB per Hour	CA
11	Number of Near-Miss reports made	CA/GH	32	Paid Hours / Turns	GH
12	Regulatory Non-Compliance	CA/GH	33	% Overtime	CA/GH
13	Internal Non- Compliance	CA/GH	34	% of Payroll, Third Party and Overtime over Total Paid Hours	CA/GH
People			35	Cost per Paid Hour (Payroll + Overtime + Third Party)	CA/GH
14	Staff turnover rate	CA/GH	36	Manpower Costs per Paid Hour - WFS (Payroll + Overtime)	CA/GH
15	% Sickness over Total Payroll Paid Hours	CA/GH	37	Manpower Costs per Paid Hour - Third Party	CA/GH
Customer satisfaction			38	% Occupancy Warehouse	CA
16	% of DEP sent On-time	CA	Financial Performance / Shareholder		
17	% of RCS sent On-time	CA	39	Yield (Revenue per Kilo / Turnaround)	CA/GH
18	% of MAN sent On-time	CA	40	Labour Costs per Kilo / Turnaround	CA/GH
19	% of FAP (Flown as planned)	CA	41	Indirect Costs per Kilo / Turnaround	CA/GH
20	% of NFD sent On-time	CA	42	Total Costs per Kilo / Turnaround	CA/GH
21	% of RCF sent On-time	CA	43	Warehouse rental cost / sqm	CA
22	% Number of Flights Delay over Total Operated Flights	GH	44	Contribution per Kilo / Turnaround	CA/GH

Figure 9.8:Lista degli indicatori e misure da implementare

9.4 Risultati e discussione

In questo capitolo analizzeremo i diversi dashboard creati per ogni entità: Cargo Operation, Finance e Human Resources. I Dashboard presentati sono creati da Power BI. La Power BI è impostata per il collegamento ai cubi del Datawarehouse snowflake.

9.4.1 Collegamento alla sorgente dei dati

Le tabelle e le viste utilizzate dagli strumenti di reporting (PBI & SSRS) si trovano nello schema BI_PRD.CUBE.

Le tabelle **CUBE_DATA**, **CUBE_BY_MONTH** e **CUBE_BY_WEEK** sono aggiornate rispettivamente dalle procedure memorizzate **SP_CUBE_DATA**, **SP_CUBE_BY_MONTH** e **SP_CUBE_BY_WEEK**.

Queste procedure sono chiamate dalla procedura **SP_CUBE_ALL**, che è programmata da un'attività di Informatica nel cloud. I rapporti del PBI sono aggiornati quotidianamente.

I requisiti espressi dalle entità sono implementati dal report di power BI connettendosi direttamente al cubo al fine di facilitare la creazione di report e il mantenimento dei dati aggiornati. Le viste collegate rendono facile la comprensione dei filtri grafici. Fintanto che non sono disponibili le licenze per l'accesso a Power BI dal cloud e la pubblicazione,

rimane la possibilità di distribuire i file all'interno dell'azienda in modo che chi possiede Power BI Desktop possa avere accesso alla visualizzazione dei report.

Interfaccia BI per l'accesso ai dati

Il primo contatto dell'utente con Power PI avviene attraverso l'interfaccia mostrata in figura. Dove troverà i rapporti che abbiamo sviluppato durante il nostro progetto. Ogni rapporto contiene i diversi KPI espressi in base alle necessità all'inizio del progetto.

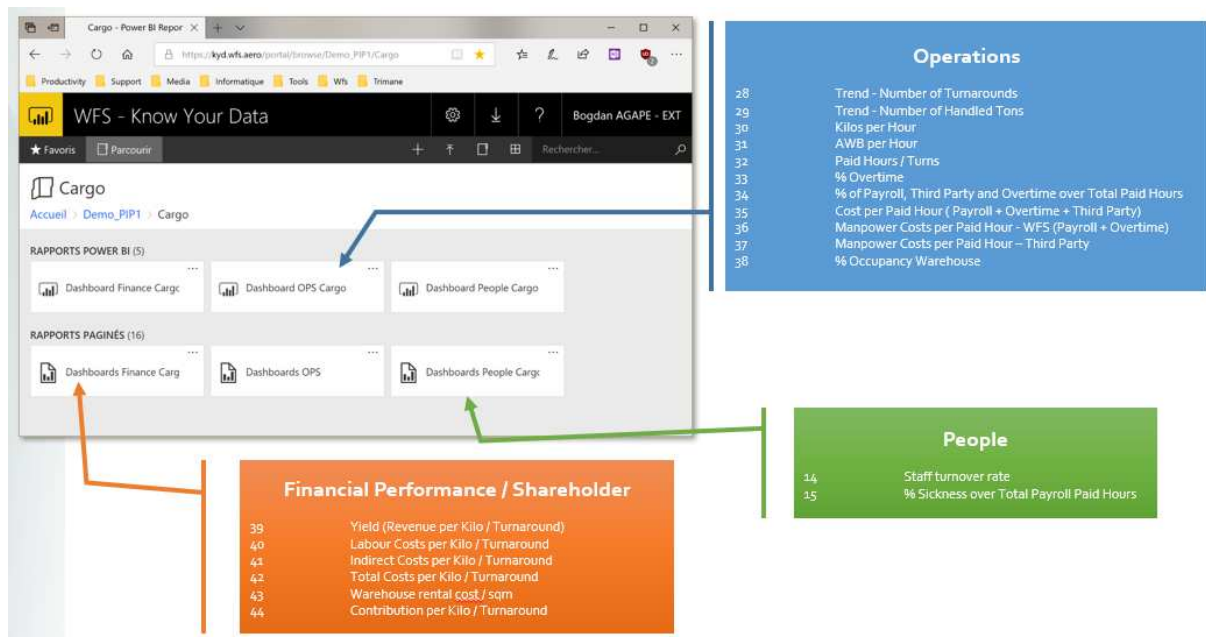


Figure 9.9: Interfaccia utente di Power BI

Analizzeremo in dettaglio i diversi rapporti a partire dal rapporto operativo, poi il rapporto della finanza e infine i rapporti delle risorse umane.

Attraverso l'interfaccia home dell'Operation Dashboard, abbiamo a sinistra i KPI sviluppati va anche notato che questa interfaccia BI sarà utilizzata in tutto il gruppo e quindi a livello internazionale.

Prima di entrare nel dettaglio, facciamo una panoramica delle funzionalità di Power PI personalizzate nell'ambito del progetto.

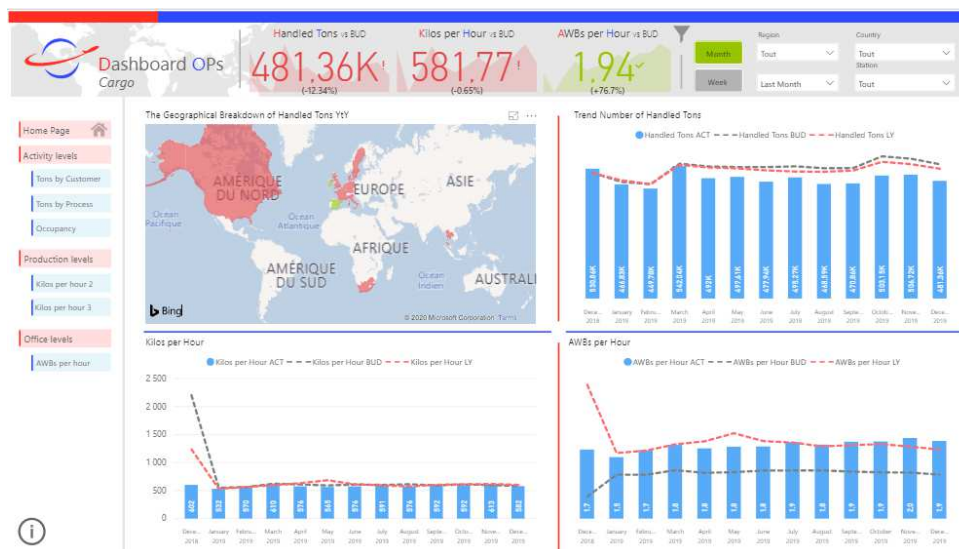


Figure 9.10: Dash board Power BI

Tab dedicata ai KPI

In questa parte e a sinistra della home page avremo l'elenco dei KPI, con ogni click sui KPI otterremo l'interfaccia e con la grafica dedicata a questo PKI. Permette di navigare su ogni KPI.

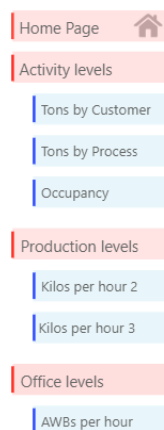


Figure 9.11:Esempio di liste di KPI Cargo Operation

Filtro per la selezione dell'area geografica e del periodo di tempo

Questa parte della tabella permette di realizzare i filtri secondo diversi parametri. La regione qui è composta da EMEEA e AMERICA, il filtro paese contiene l'elenco di tutti i paesi del gruppo WFS.

Figure 9.12:Filtro per la selezione

Indicatore principale di Cargo Operation Dashboard

La figura seguente ci mostra i KPI che mirano ad avere informazioni sul confronto tra le tonnellate elaborate e il budget, i chili all'ora rispetto ai chili all'ora previsti dal budget, gli AWB all'ora rispetto agli AWB previsti dal budget.



KPI principali al centro della dashboard

La parte centrale è composta da quattro zone, una delle quali presenta la situazione grafica e i tre altri, KPI rimanenti.

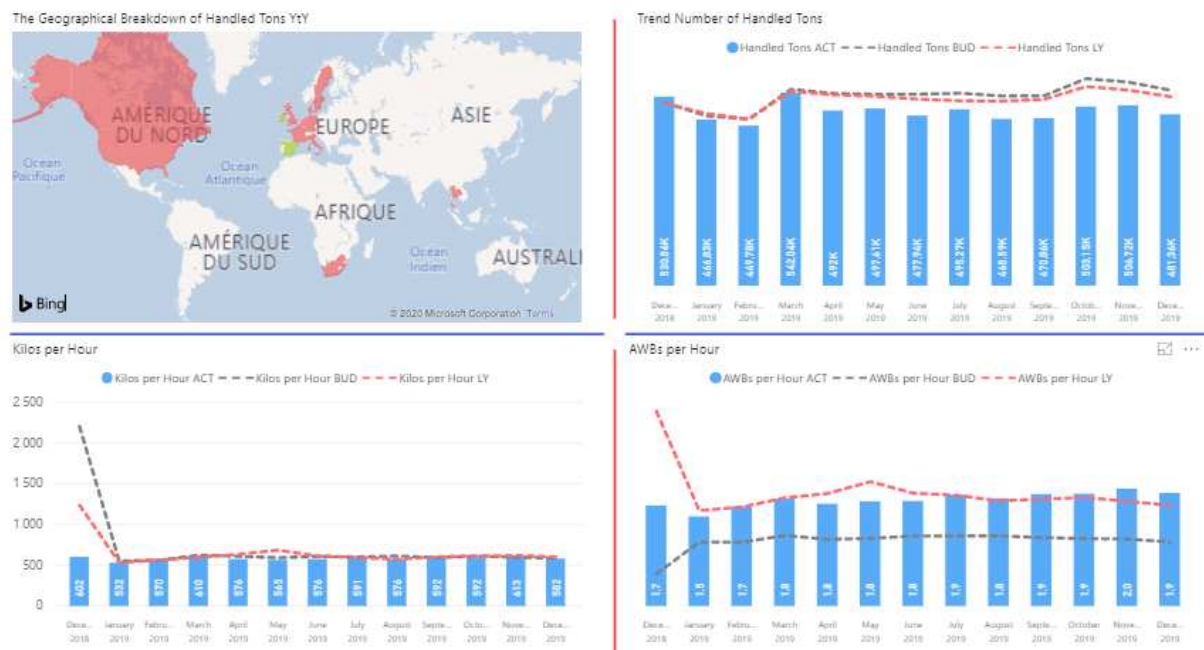


Figure 9.13:Zone di Power BI per il display dei KPI

Da sinistra a destra le descrizioni dei campi nella figura qui sopra:

1. La zona che rappresenta la geografia in quest'area appare sulla mappa la posizione geografica del sito selezionato nel filtro.
2. Nell'area del diagramma che mostra l'andamento delle tonnellate movimentate, viene mostrato un confronto tra le tonnellate attuali e le tonnellate prodotte lo scorso anno.
3. Area che mostra un confronto Chili all'ora rispetto al budget e all'anno precedente.
4. Area che mostra il numero di AWB attuali rispetto all'AWB in bilancio e all'anno precedente.

Gli elementi presentati sopra ci permetteranno di comprendere i dettagli dei KPI sviluppati per ogni entità.

9.5 Presentazione dei diversi Dashboard per entità.

Presenteremo prima le esigenze di KPI espresse dallo Staff funzionale, poi i KPI prodotti durante il Progetto e infine un'analisi dei miglioramenti e dei problemi riscontrati e una proposta di raccomandazione.

9.5.1 Dashboard Cargo Operation

Descrizione dei KPI espressi inizialmente:

Inizialmente i KPI espressi dal business (Mestiere) sono riassunti nella tabella sottostante (Tabella 9.7). Un promemoria della portata del nostro studio che si occupa solo dei KPI del settore Cargo.

#Kpi	Operations
28	Trend - Number of Turnarounds
29	Trend - Number of Handled Tons
30	Kilos per Hour
31	AWB per Hour
32	Paid Hours / Turns
33	% Overtime
34	% of Payroll, Third Party and Overtime over Total Paid Hours
35	Cost per Paid Hour (Payroll + Overtime + Third Party)
36	Manpower Costs per Paid Hour - WFS (Payroll + Overtime)
37	Manpower Costs per Paid Hour – Third Party
38	% Occupancy Warehouse

Figure 9.14:KPI da implementare per il Cargo Operation

Trend – Number of Handled Tons

Per l'analisi dei risultati abbiamo scelto come esempio la regione EMEEA, la Spagna come paese e tutte le stazioni in Spagna. Si noti che lo strumento ci dà la possibilità di andare ad una granularità più fine, cioè per stazione.



Figure 9.15:Dashboard principale dell'entità Spagna

La struttura della tabella è già stata spiegata in precedenza e possiamo direttamente all'analisi dei risultati dei KPI:

Trend – Number of Handled Tons

Misurazione del livello di attività/

Questo KPI consente di misurare il livello di attività di un determinato magazzino, sia per l'esportazione che per l'importazione, e su tutta la serie di siti del gruppo.

Formula: Numero di tonnellate gestite.

La formula utilizzata in questo caso è il numero di tonnellate movimentate in un determinato magazzino. Queste informazioni sono estratte dal database delle operazioni. Viene inoltre utilizzato per analizzare l'andamento e confrontarlo con il numero attuale di tonnellate indicato nel grafico a barre blu nel periodo dicembre 2018 - dicembre 2019, confrontarlo con le tonnellate preventivate, in linea tratteggiata grigia nella figura e infine confrontarlo con le tonnellate trattate l'anno scorso in linea tratteggiata rossa.

Misura che definisce il KPI

Tonnellate manipolate (misurate come Tonnellate manifestate e controllate) compresi i transiti. I transiti sono contati su entrambe le tratte, ovvero quando il transito si manifesta e a destinazione quando il transito è stato registrato.

Tonnellate totali gestite per l'importazione e l'esportazione (compresi i transiti). Le Tonnellate manifestate e ricevute (Check-In) sono il riferimento, e l'FFM l'elemento fisico.

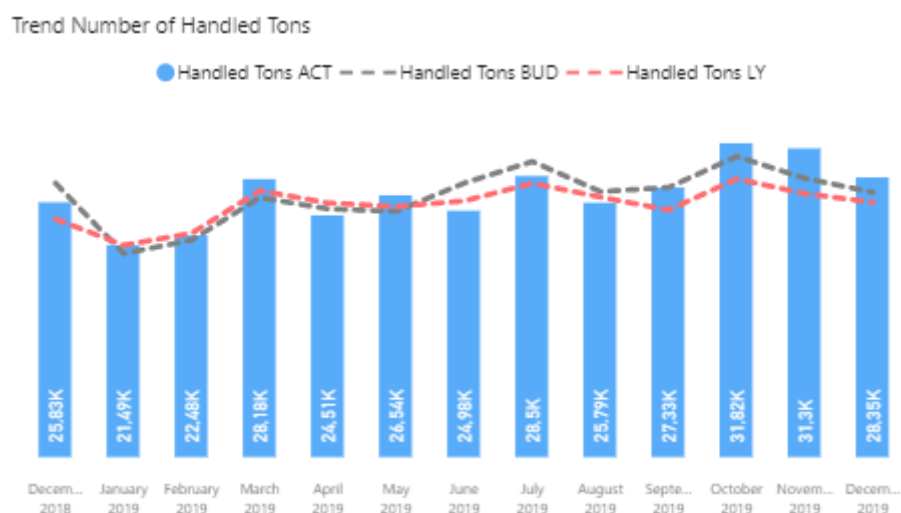


Figure 9.16:Tendenza Numero di tonnellate gestite

Misurare la produttività del magazzino

Formula: Numero di tonnellate gestite calcolate in chilogrammi diviso per il totale delle ore di lavoro del reparto "Magazzino".

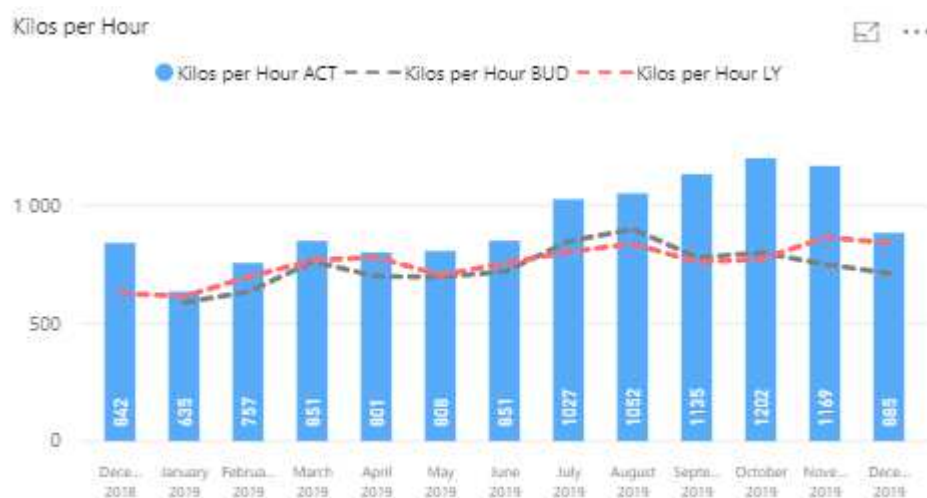


Figure 9.17:KPI presentando il numero di kili per ora rispetto all'anno precedente

Chili di Cargo trattato (inclusi BUP e Loose) divisi per il totale delle ore di lavoro retribuite (inclusi i dipendenti retribuiti in base allo stipendio e alle ore di lavoro) + Ore di lavoro straordinario + Ore di lavoro di parti terze - Assenza Altri - Assenza Malattia da assenteismo - Assenza a lungo termine - Vacanze. Esclusivamente correlato con il Profilo "Magazzino".

AWB per Hour

Misurare la produttività dell'ufficio;

Formula: Numero di AWB gestiti diviso per il totale delle ore di lavoro del reparto "Ufficio".

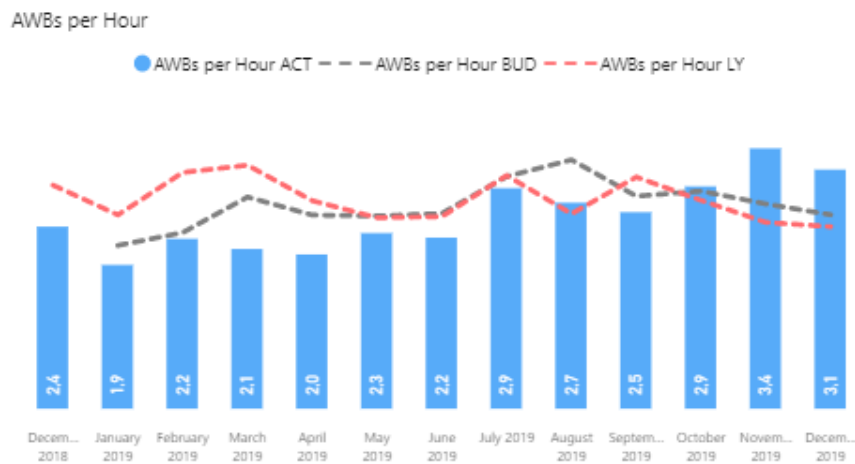


Figure 9.18:KPI presentando il numero di AWBs per ora rispetto all'anno precedente

Numero di documenti AWB (Master e Direct AWB) diviso per il totale delle ore di lavoro pagate (inclusi gli stipendi e le ore di lavoro retribuite) + Ore di lavoro straordinario + Ore di lavoro di terzi - Assenza Altri - Assenza Malattia da assenteismo - Assenza a lungo termine - Vacanze. Esclusivamente correlato con il Profilo "Ufficio".

%Overtime

Misurare l'evoluzione delle ore straordinari

Formula: Ore di lavoro straordinario divise per le ore pagate.

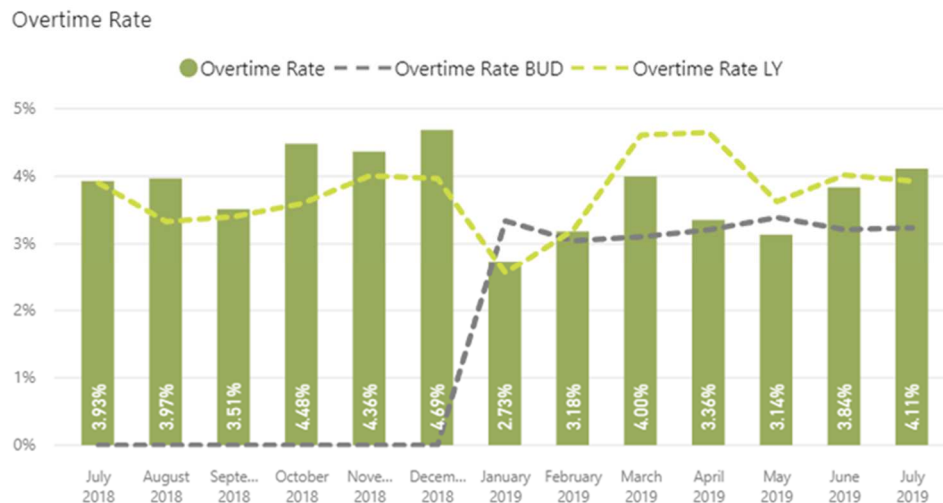


Figure 9.19: KPI per tasso di ore straordinari rispetto all'anno precedente

Metric e formula

Questo rapporto è calcolato a livello di stazione e inizialmente non includerebbe la ripartizione per profilo nonostante la disponibilità di metriche con questo livello di attributo.

La % è sempre calcolata sul totale delle ore pagate, che include i dipendenti con salario e orario, ma esclude i terzi e i subappaltatori.

%of Payroll, Third Party and Overtime Total Paid Hours

Misurare il livello di flessibilità della nostra manodopera

Formula: Totale delle ore pagate diviso per il totale delle ore pagate (%)

Totale degli straordinari diviso per il totale delle ore pagate (%)

Totale delle ore di terzi e delle ore subappaltate divise per il totale delle ore pagate (%)

Totale ore pagate = libro paga + tempo supplementare + terzi e subappalto

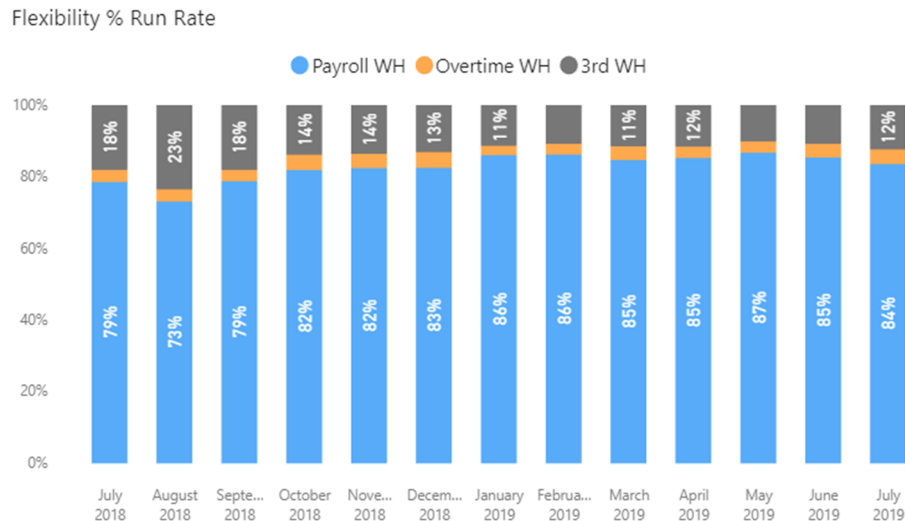


Figure 9.20:KPI presentando il tasso flessibilità

Tre tariffe contemporaneamente. Fornire un riferimento del nostro livello di flessibilità della manodopera. Basato sulle ore pagate e non sulle ore di lavoro.

Questi KPI funzionerebbero a livello di Profilo/Dipartimento per il Cargo Business. Profilo del magazzino. GH, sarebbe un totale senza discriminazione di Profilo/Dipartimento.

Cost per Paid Hour (Payrol+Overtime+Third Party)

Misurare l'evoluzione del confronto tra il costo orario e l'YTD

Formula: Costi totali delle ore pagate (inclusi dipendenti e terzi)

9R1110FD + 9R1120FD + 9R1130FD + 9R1140FD + 9R1150FD + 9R1210FD + 9R1220FD + 9R1230FD + 9R0130FD + 9R0130FD + 9R0140FD + 9R1510FD + 9R1520FD + 8R0410FD diviso per il totale delle ore pagate (compresi i dipendenti e i terzi / subappaltatori) e senza discriminazione del profilo

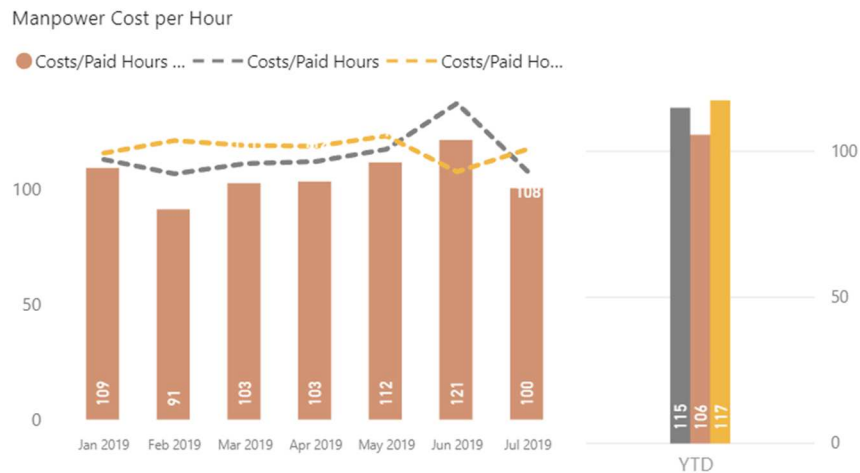


Figure 9.21:KPI sul costo della mano d'opera per ore

Tutti i conti dello Stipendio, comprese le Assicurazioni e la Previdenza Sociale più i Costi di Terzi e i Costi di Lavoro Straordinario divisi per il Totale delle Ore Pagate (Stipendio, Terzi e Lavoro Straordinario).

I costi sarebbero limitati alla sezione Field in SAP in quanto non vi è discriminazione del livello di stazione disponibile. La tariffa sarebbe quindi calcolata solo a livello di Paese.

Manpower Cost per Paid Hour – WFS (Payroll+Third Party)

Misurazione dell'evoluzione del confronto tra il costo all'ora solo WFS e YTD

Formula: Totale dei costi delle ore pagate SFP significa conti

9R1110FD + 9R1120FD + 9R1130FD + 9R1140FD + 9R1150FD + 9R1210FD + 9R1220FD + 9R1230FD + 9R0130FD + 9R0140FD + 9R1510FD + 9R1520FD + diviso per il totale delle ore pagate WFS e senza discriminazione di Profilo

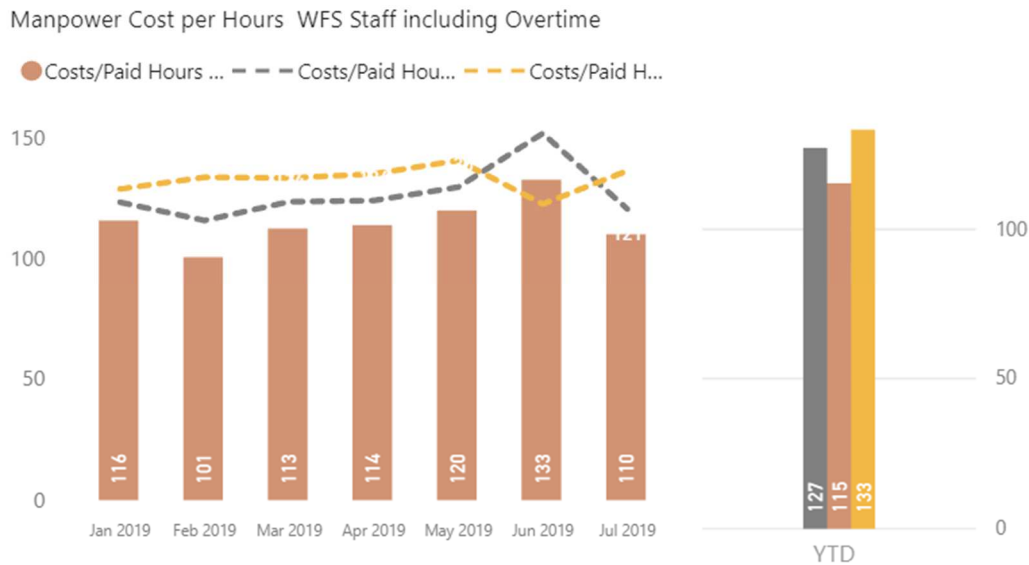


Figure 9.22:KPI sul costo della mano d'opera includendo le ore straordinarie

KPI calcolati in modo univoco a livello di Paese e senza discriminazione di Profilo come inizialmente non disponibile nella Struttura dei conti.

Questo tasso dovrebbe includere solo la sezione Field, poiché i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali.

Manpower Cost per Paid Hour – Third Party

Misurazione dell'evoluzione del Costo orario solo Terzi e Subappalto e confronto YTD

Formula: Totale dei costi delle ore pagate conti di terzi, ovvero conti di terzi

8R0410FD diviso per Totale Ore Pagate Terzi / Subappalto e senza discriminazione di Profilo.

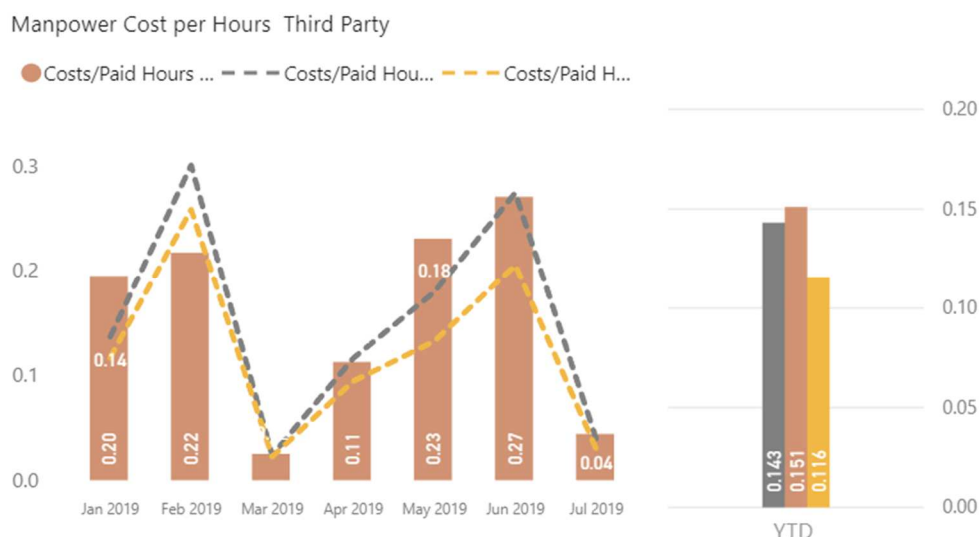


Figure 9.23:KPI sul costo della mano d'opera includendo le ore di terzi

KPI calcolati in modo univoco a livello di Paese e senza discriminazione di Profilo come inizialmente non disponibile nella Struttura dei conti.

Questo tasso dovrebbe includere solo la sezione Field, poiché i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali.

%Occupacy Warehouse

Tonnellate totali gestite diviso per Capacità Operativa Totale di un Magazzino

Definizione della metrica

Tonnellate totali gestite (importazione, esportazione e transito) divise per capacità di magazzino mensile. La capacità totale deve essere riportata come capacità operativa e non come standard IATA.

9.5.2 Manpower Cargo/People KPI

Staff Turnover: Tasso di turnover del personale %Malattia sul totale delle ore pagate

Nome tecnica: #KPi14

Misurazione del tasso di rotazione del personale

Formula: Numero di abbandoni in un mese diviso per Numero di dipendenti (Capi) l'ultimo giorno del mese.

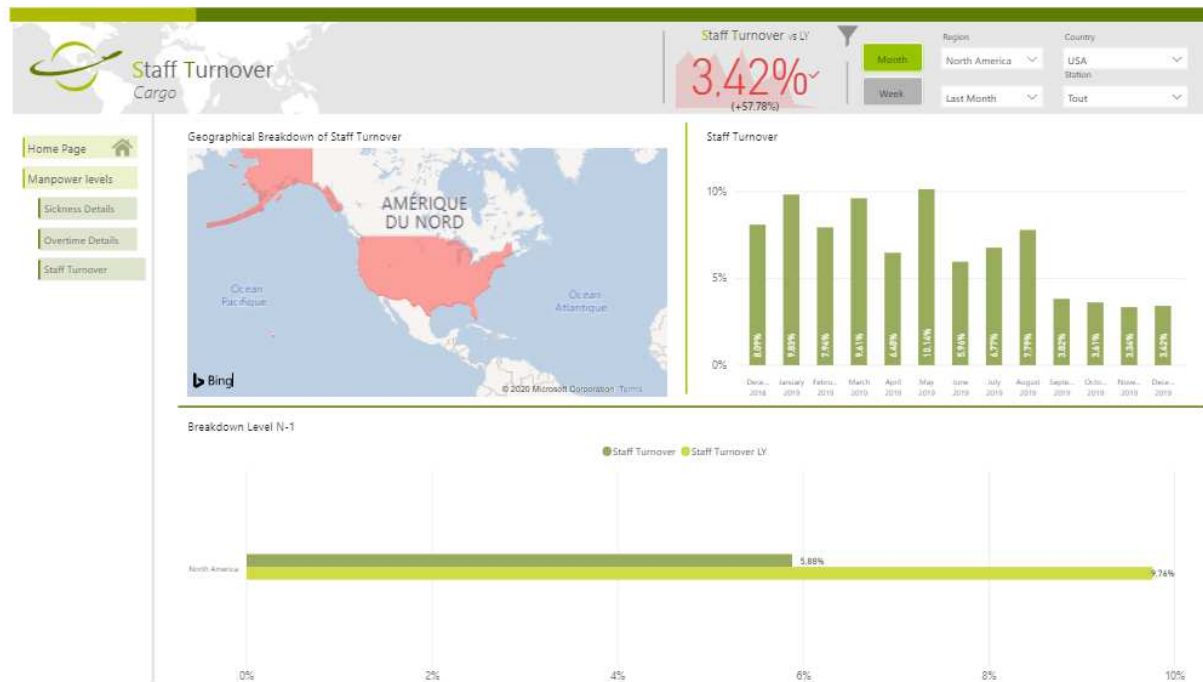


Figure 9.24: Dashboard per Manpower Cargo/People KPI

Elementi di calcolo:

- Totale delle partenze durante il mese M
- Conteggi di fine mese M-1

Business rule:

- **TASSO DI TURNOVER DEL PERSONALE = PARTENZA / NUMERO DEI DIPENDENTI**
- Totale dei dipendenti che hanno lasciato l'azienda durante il mese M: Numero di dipendenti interni che hanno lasciato l'azienda durante il mese M. Solo per quanto riguarda le retribuzioni.
- Contabilità dei dipendenti: Una persona a tempo parziale si conterebbe come un effettivo. Dipendenti fisici (solo per quanto riguarda le retribuzioni) esistenti l'ultimo giorno del mese. Una fotografia scattata l'ultimo giorno del mese. I dipendenti a tempo parziali sono conteggiati come un effettivo.

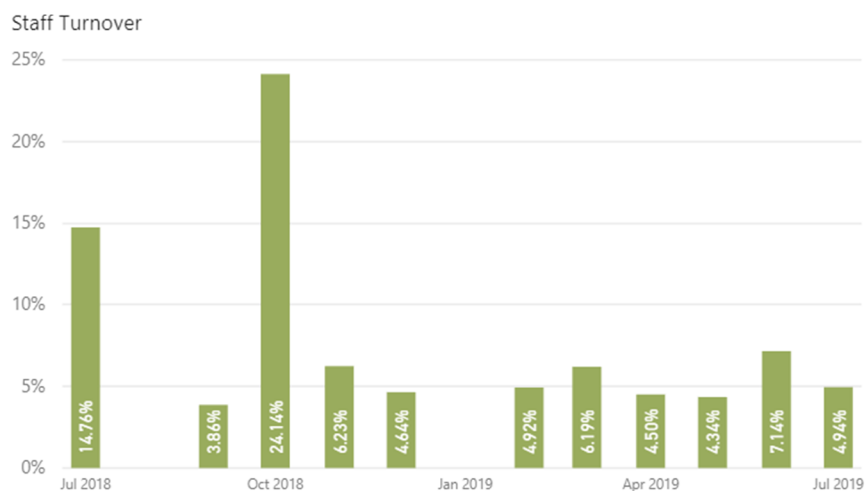


Figure 9.25: KPI sul turnover del personale

Sickness Rate

Nome tecnico: #KPi15

Misurare l'evoluzione del tasso di malattia

Formula: Ore di malattia divise per le ore di paga.



Figure 9.26: Dashboard presentando il tasso di malattie

Elementi di calcolo:

- Totale pagato Ore di busta paga pagate
- Ore totali di malattia

Business rule:

- % delle ore di malattia rispetto alle ore di paga (lavoro straordinario e lavoro temporaneo NON inclusi).
- Questo rapporto è calcolato a livello di stazione e inizialmente non includerebbe la ripartizione per profilo.
- Se le ore di malattia sono pagate o meno è irrilevante per questo rapporto.

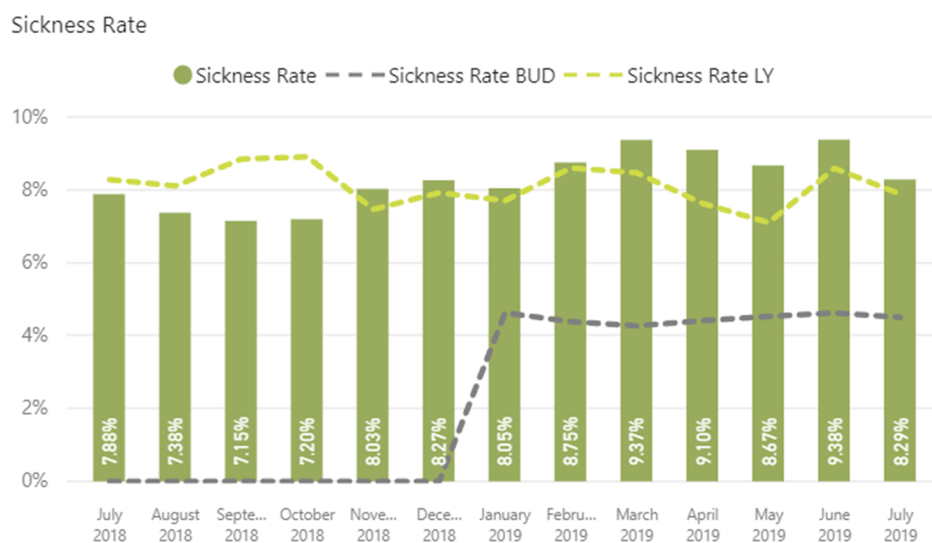


Figure 9.27: KPI sul tasso di malattie rispetto all'anno scorso

Overtime rate

Nome tecnico: #KPi33

Misurare l'evoluzione degli straordinari

Formula: Ore di lavoro straordinario divise per le ore pagate.

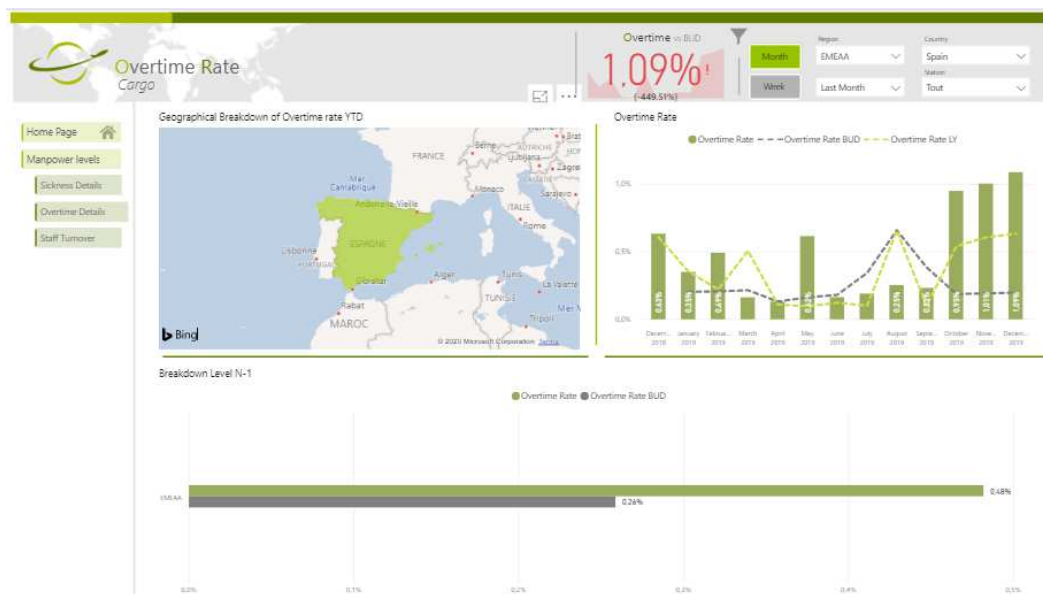


Figure 9.28: Dashboard presentando il KPI le ore straordinari

Elementi di calcolo:

- Ore di lavoro straordinario
- Totale ore di lavoro

Business rule:

- % delle ore di lavoro straordinario rispetto alle ore pagate.
- Questo rapporto è calcolato a livello di stazione e inizialmente non includerebbe la ripartizione per profilo.

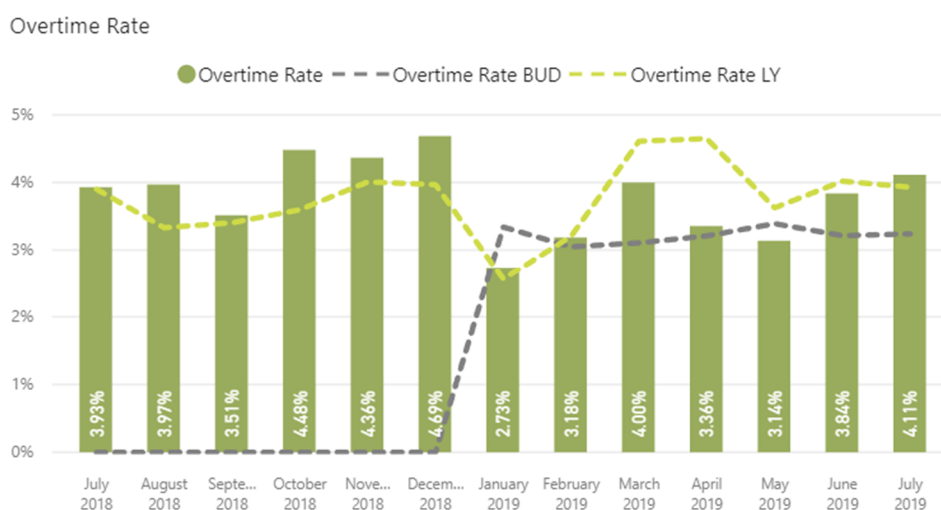


Figure 9.29:KPI sulle ore straordinari rispetto all'anno scorso

Questo rapporto è calcolato a livello di stazione e inizialmente non includerebbe la ripartizione per profilo nonostante la disponibilità di metriche con questo livello di attributo.

La % è sempre calcolata sul totale delle ore pagate, che include i dipendenti con salario e orario, ma esclude i terzi e i subappaltatori.

- Cost per Paid Hour (Payroll + Overtime + Third Party)

Nome tecnico: #KPi35

Misurazione dell'evoluzione della Formula di confronto del Costo all'ora e YTD: Costo totale delle ore pagate (inclusi dipendenti e terzi)

9R1110FD + 9R1120FD + 9R1130FD + 9R1140FD + 9R1150FD + 9R1210FD + 9R1220FD + 9R1230FD + 9R0130FD + 9R0140FD + 9R1510FD + 9R1520FD + 8R0410FD diviso per Ore Totali Pagate (inclusi i Dipendenti e Terzi / Subappaltatori) e senza discriminazione di Profilo

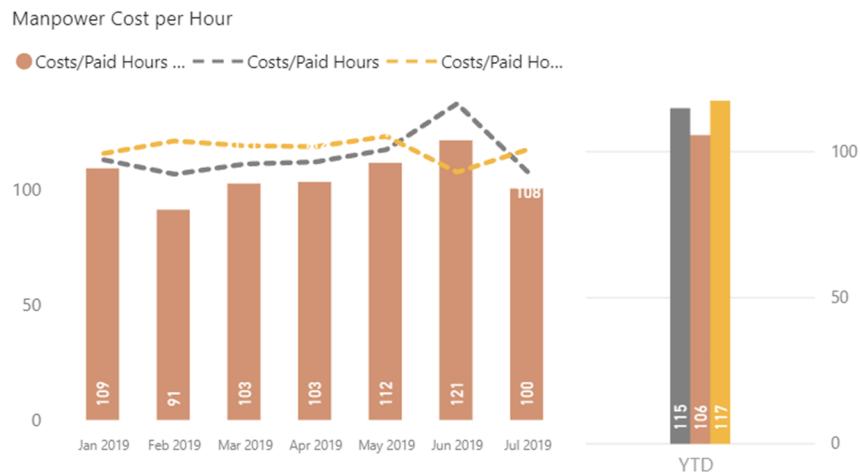


Figure 9.30:KPI sul costo della mano d'opera per ore

Tutti i conti dello Stipendio, comprese le Assicurazioni e la Previdenza Sociale più i Costi di Terzi e i Costi di Lavoro Straordinario divisi per il Totale delle Ore Pagate (Stipendio, Terzi e Lavoro Straordinario).

I costi sarebbero limitati alla sezione Field in SAP in quanto non vi è discriminazione del livello di stazione disponibile. La tariffa sarebbe quindi calcolata solo a livello di Paese.

Calculation elements:

- Total Salary Costs
- Total Paid Hours

Business rule:

- All Salary accounts including Insurances and Social Security plus Third-Party Costs and Overtime Costs divided by Total Paid Hours (Payroll, Third Party and Overtime).
- This rate should include both Field and Field Overheads (by definition, we do not have Overheads at Station Level, but Supporting Activities as defined in "Def Shared»).
- To agree with Finance account names and categories.

Elementi di calcolo:

- Totale costi salariali
- Totale ore pagate

Business rule:

- Tutti i conti dello Stipendio, comprese le Assicurazioni e la Previdenza Sociale più i Costi di Terzi e gli Straordinari, divisi per il Totale delle Ore Pagate (Stipendi, Terzi e Straordinari).
- Questo tasso dovrebbe includere sia le spese generali di campo che le spese generali di campo (per definizione, non abbiamo spese generali a livello di stazione, ma attività di supporto come definite in "Def Shared").

Manpower Costs per Paid Hour - WFS (Payroll + Overtime)

Technical name: #KPi36

Misurazione dell'evoluzione del confronto tra il costo all'ora solo WFS e YTD

Formula: Totale dei costi delle ore pagate VMR, che significa conti

9R1110FD + 9R1120FD + 9R1130FD + 9R1140FD + 9R1150FD + 9R1210FD +
9R1220FD + 9R1230FD + 9R0130FD + 9R0140FD + 9R1510FD + 9R1520FD + diviso
per Ore totali pagate WFS e senza discriminazione di profilo.

Elementi di calcolo:

- Totale costi salariali
- Totale ore pagate meno i terzi

Business rule:

- Tutti i conti di stipendio, comprese le Assicurazioni e la Previdenza Sociale più gli straordinari, divisi per le retribuzioni e le ore di lavoro straordinario.
- Questo tasso dovrebbe includere sia le spese generali di campo che le spese generali di campo (per definizione, non abbiamo spese generali a livello di stazione, ma attività di supporto come definite in "Def Shared "V).

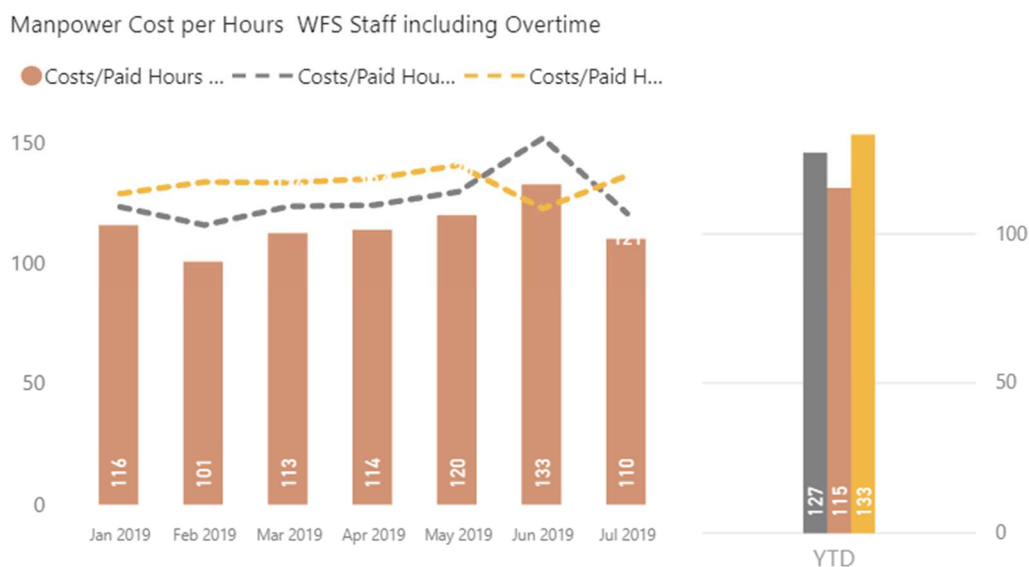


Figure 9.31:KPI sul costo della mano d'opera di WFS includendo le ore straordinarie
KPI calcolati in modo univoco a livello di Paese e senza discriminazione di Profilo come inizialmente non disponibile nella Struttura dei conti.

Questo tasso dovrebbe includere solo la sezione Field, poiché i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali.

Manpower Costs per Paid Hour – Third Party

Nome tecnico: #KPi37

Misurazione dell'evoluzione del Costo orario solo Terzi e Subappalto e confronto YTD

Formula: Totale dei costi delle ore pagate di terzi, ovvero conti 8R0410FD diviso per il totale delle ore pagate di terzi / subappalto e senza discriminazione del profilo

Elementi di calcolo:

- Totale costi di terzi
- Totale ore di lavoro di terzi

Business rule:

- Conti di Terzi divisi per Ore Terzi.
- Questo tasso dovrebbe includere sia le Spese generali di campo che le Spese generali di campo (per definizione, non abbiamo Spese generali a livello di stazione, ma Attività di supporto come definite in "Def Shared").

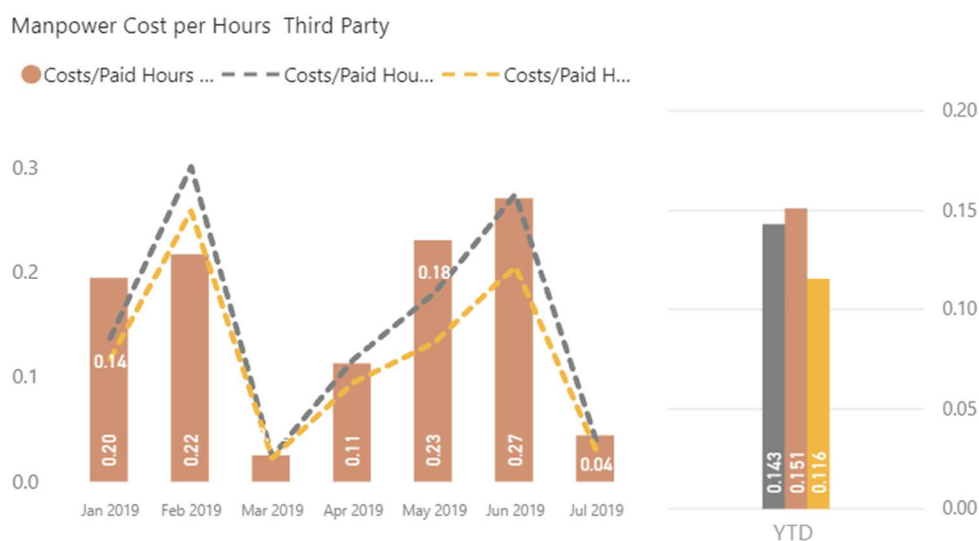


Figure 9.32:KPI sul costo della mano d'opera includendo le ore di terzi

KPI calcolati in modo univoco a livello di Paese e senza discriminazione di Profilo come inizialmente non disponibile nella Struttura dei conti.

Questo tasso dovrebbe includere solo la sezione Field, poiché i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali.

9.5.2 Finance Dashboard

Cost per Paid Hour (Payroll + Overtime + Third Party)

Nome tecnico: #KPi35

Misurazione dell'evoluzione del costo orario e confronto YTD (YearS to date)

Formula: Costi totali delle ore pagate (inclusi dipendenti e terzi)

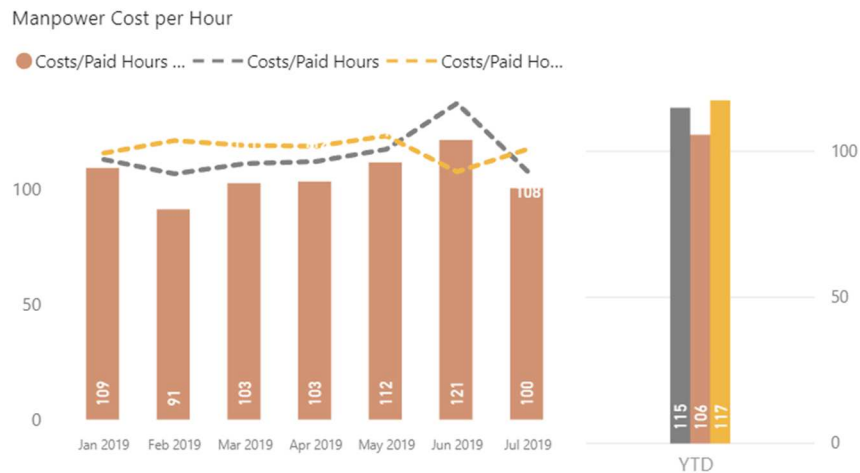
9R1110FD + 9R1120FD + 9R1130FD + 9R1140FD + 9R1150FD + 9R1210FD +
9R1220FD + 9R1230FD + 9R0130FD + 9R0140FD + 9R1510FD + 9R1520FD +
8R0410FD diviso per Ore Totali Pagate (inclusi Dipendenti e Terzi / Subappaltatori) e
senza discriminazione di Profilo.

Elementi di calcolo:

- Totale costi salariali
- Totale ore pagate

Business rule:

- "Tutti i conti dello Stipendio, comprese le Assicurazioni e la Previdenza Sociale più i costi di terza retribuzione e gli straordinari, divisi per il totale delle ore pagate (salari, terzi e straordinari).
- Questo tasso dovrebbe includere sia le spese generali di campo che le spese generali di campo (per definizione, non abbiamo spese generali a livello di stazione, ma attività di supporto come definite in ""Def Shared"").’



Tutti i conti dello Stipendio, comprese le Assicurazioni e la Previdenza Sociale più i Costi di Terzi e i Costi di Lavoro Straordinario divisi per il Totale delle Ore Pagate (Stipendio, Terzi e Lavoro Straordinario).

I costi sarebbero limitati alla sezione Field in SAP in quanto non vi è discriminazione del livello di stazione disponibile. La tariffa sarebbe quindi calcolata solo a livello di Paese.

Manpower Costs per Paid Hour - WFS (Payroll + Overtime)

Nome tecnico: #KPi36

Misurazione dell'evoluzione del confronto tra il costo all'ora solo WFS e YTD

Formula: Totale dei costi delle ore pagate VMR, che significa conti

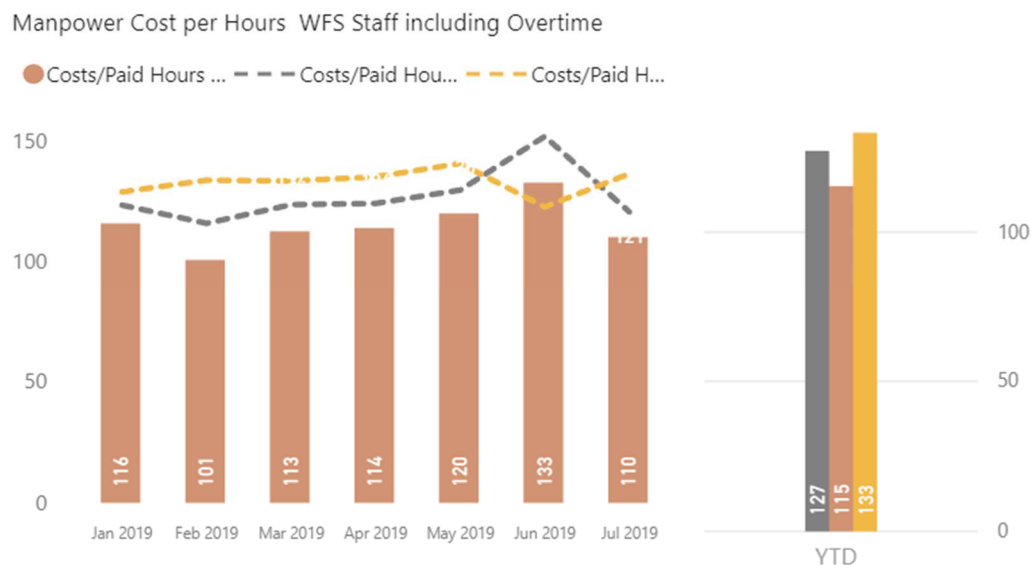
9R1110FD + 9R1120FD + 9R1130FD + 9R1140FD + 9R1150FD + 9R1210FD + 9R1220FD + 9R1230FD + 9R0130FD + 9R0140FD + 9R1510FD + 9R1520FD + diviso per Ore totali pagate WFS e senza discriminazione di profilo

Elementi di calcolo:

- Totale costi salariali
- Totale ore pagate meno i terzi

Business rule:

- "Tutti i conti dello Stipendio, comprese le Assicurazioni e la Previdenza Sociale più gli straordinari, divisi per le retribuzioni e le ore di lavoro straordinario.
- Questo tasso dovrebbe includere sia le spese generali di campo che le spese generali di campo (per definizione, non abbiamo spese generali a livello di stazione ma attività di supporto come definito in ""Def Shared"").



Manpower Costs per Paid Hour – Third Party

Nome tecnico: #KPi37

Misurazione dell'evoluzione del Costo orario solo Terzi e Subappalto e confronto YTD.

Formula: Totale dei costi delle ore pagate conti di terzi, ovvero conti di terzi.

8R0410FD diviso per Totale Ore Pagate Terzi / Subappalto e senza discriminazione di Profilo.

Elementi di calcolo:

- Totale costi di terzi.
- Totale ore di lavoro di terzi.

Business rule:

- "Conti di terzi divisi per orari di terzi.

- Questo tasso dovrebbe includere sia le spese generali di campo che le spese generali di campo (per definizione, non abbiamo spese generali a livello di stazione ma attività di supporto come definite in ""Def Shared"").

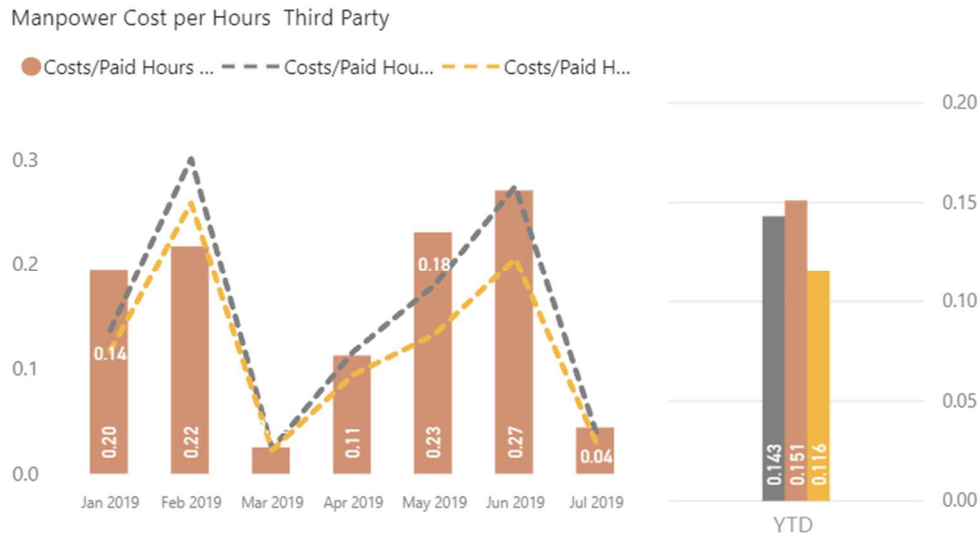


Figure 9.33:KPI sul costo della mano d'opera includendo le ore di terzi

KPI calcolati in modo univoco a livello di Paese e senza discriminazione di Profilo come inizialmente non disponibile nella Struttura dei conti.

Questo tasso dovrebbe includere solo la sezione Field, poiché i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali.

Yield (Revenue per Kilo / Turnaround)

Nome tecnico: #KPi39

Misurazione dell'evoluzione del rendimento e dell'impatto YTD rispetto a LY e Budget

Formula: Totale Ricavi riportati nel sistema di consolidamento SAP, ovvero 6R0000 GRH / CAH diviso per Totale dei chili gestiti localmente / Turnaround.

Elementi di calcolo:

- Totale ricavi
- Tonnellate fatturate
- Turnaround

Business rule:

- I ricavi totali riportati nel Consolidamento SAP divisi per il totale delle Tonnellate Gestite meno il Transito/Turnaround.
- In alternativa, le Tonnellate Fatturate verrebbero create come metrica indipendente.
- Eccezione al principio di metriche riportate "non manuali".

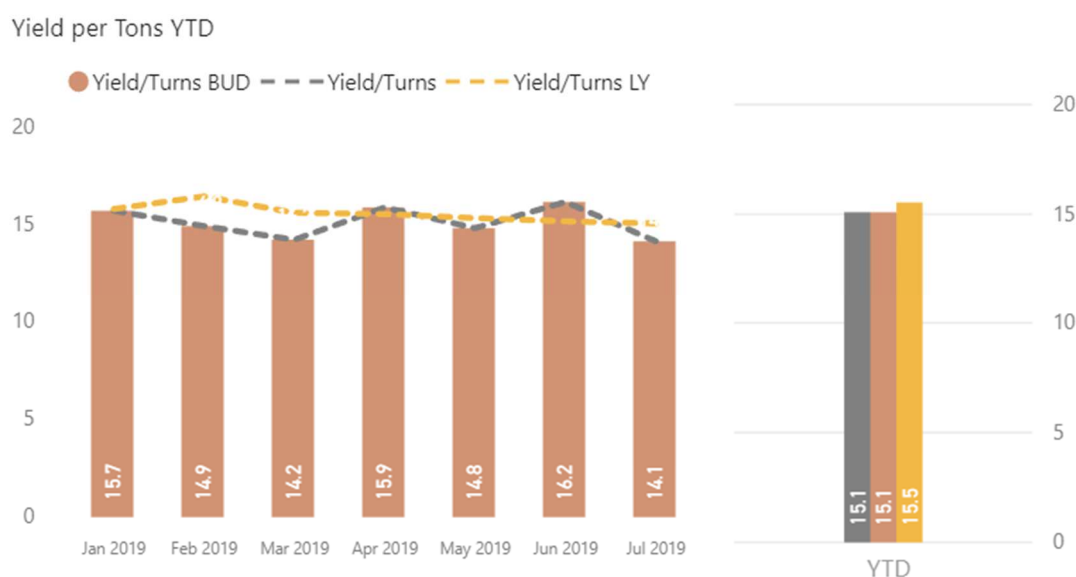


Figure 9.34: KPI sulla produzione del tonnellaggio YTD

Questo tasso verrebbe calcolato solo a livello nazionale, poiché i dati consolidati SAP non sono discriminanti a livello di stazione.

Labor Costs per Kilo / Turnaround.

Nome tecnico: #KPi40

Totale Costi del Lavoro (Stipendi inclusi nelle Assicurazioni e Previdenza Sociale e nei conti di Terzi) riportati nel Consolidamento SAP diviso per Totale dei chili gestiti/giri. Sezione di campo in SAP.

Elementi di calcolo:

- Costo totale del lavoro
- Tonnellate gestite

- Turnaround

Business rule:

- "Totale costi del lavoro (stipendi inclusi con le Assicurazioni e la Previdenza Sociale e conti di Terzi) riportati nei Consolidamenti SAP divisi per il Totale Tonnellate Gestite / Turnaround.
- Solo sezione Field in Consolidamento SAP".
- Questo tasso dovrebbe includere solo la sezione Field in quanto i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali. Di conseguenza, è disponibile solo a livello di paese.
- Informazioni non discriminate per Profilo/Settore in quanto non disponibili nell'attuale Grafico dei conti.
- A differenza dei calcoli del Rendimento, qui il riferimento del Chilo sarebbe il Totale dei Chili gestiti, compresi i Transiti.

Indirect Costs per Kilo / Turnaround

Nome tecnico: #KPi41

Costi indiretti divisi per Kilo / Turnaround Sezione di campo in SAP.

Elementi di calcolo:

- Totale costi indiretti
- Tonnellate gestite
- Turnaround

Business rule:

- Totale Altri costi diversi dal Costo del lavoro (Stipendi inclusi nelle Assicurazioni e Previdenza Sociale e nei conti di Terzi) riportati nel Consolidamento SAP diviso per il Totale Tonnellate Gestite/Ritorni.
- Solo sezione Campo nel Consolidamento SAP".

- Costi indiretti come totale degli altri costi diversi dal costo del lavoro (stipendi inclusi in Assicurazioni e Previdenza Sociale e conti Terzi) riportati nel Consolidamento SAP.
- Questo tasso dovrebbe includere solo la sezione Field in quanto i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali. Di conseguenza, è disponibile solo a livello di paese.
- Informazioni non discriminate per Profilo/Servizio in quanto non disponibili nell'attuale Grafico dei conti.

Total Costs per Kilo / Turnaround

Nome tecnico: #KPi42

Misurazione dell'evoluzione dei costi totali (livello di campo) in un periodo di 12 mesi

Formula: Costi di campo riportati nel sistema di consolidamento SAP, ovvero le spese di campo 6R0010FD diviso per Totale dei chili gestiti localmente / Turnaround.

Elementi di calcolo:

- Costi totali
- Tonnellate gestite
- Turnaround

Business rule:

- Totale Costi del Lavoro e Indiretti riportati nel Consolidamento SAP diviso per il Totale Tonnellate Manipolate / Turnaround.
- Solo sezione Field in Consolidamento SAP.

Total Cost / Turns YtY

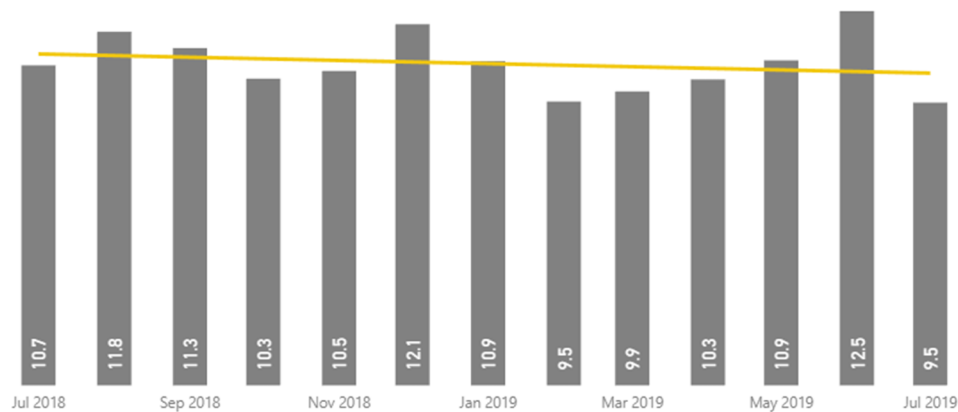


Figure 9.35:KPI sul costo totale per turno

- Totale Costi riportati nel Consolidamento SAP a livello di Field Expenses 6R0010FD.
- Questo tasso dovrebbe includere solo la sezione Field, poiché i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali. Di conseguenza, è disponibile solo a livello di paese

Contribuito per Kilo / Turnaround

Nome tecnico: #KPi44

Misurazione dell'evoluzione del contributo (livello di campo) e dell'impatto YTD rispetto a LY e Budget

Formula: Field Margin riportato nel sistema di consolidamento SAP, ovvero Field Margin 5R0000 diviso per il totale dei chili gestiti a livello locale.

Elementi di calcolo:

- Tonnellate gestite
- Turnaround

Business rule:

- Ricavi meno il totale dei costi riportati nel Consolidamento SAP diviso per il totale delle Tonnellate Gestite / Turnaround.
- Solo sezione Field in Consolidamento SAP.

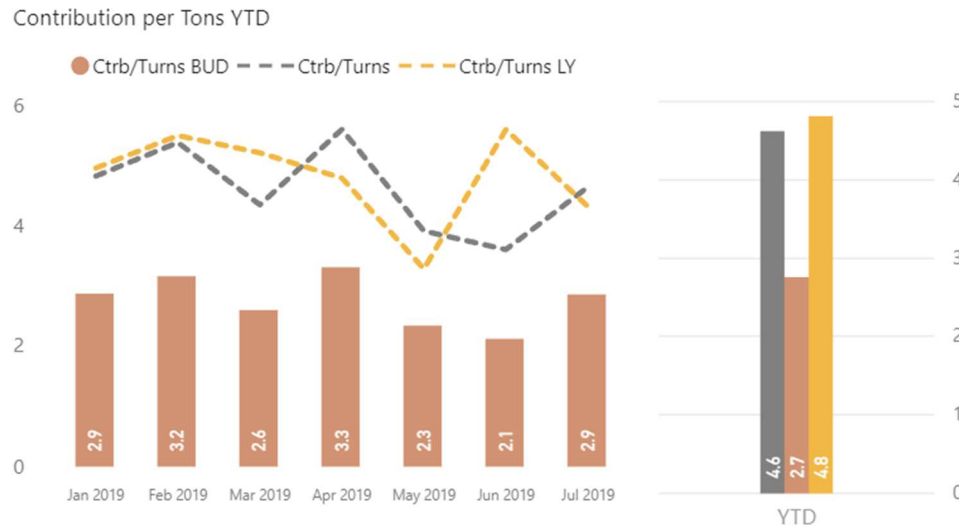


Figure 9.36:KPI sulla contribuzione del tonnage YTD

Questo tasso dovrebbe includere solo la sezione di campo, poiché i dati consolidati SAP non sono discriminanti a livello di stazione e di conseguenza questa sarebbe l'unica opzione per isolare i costi generali. Di conseguenza, è disponibile solo a livello di paese.

9.6 Problemi riscontrate e miglioramenti

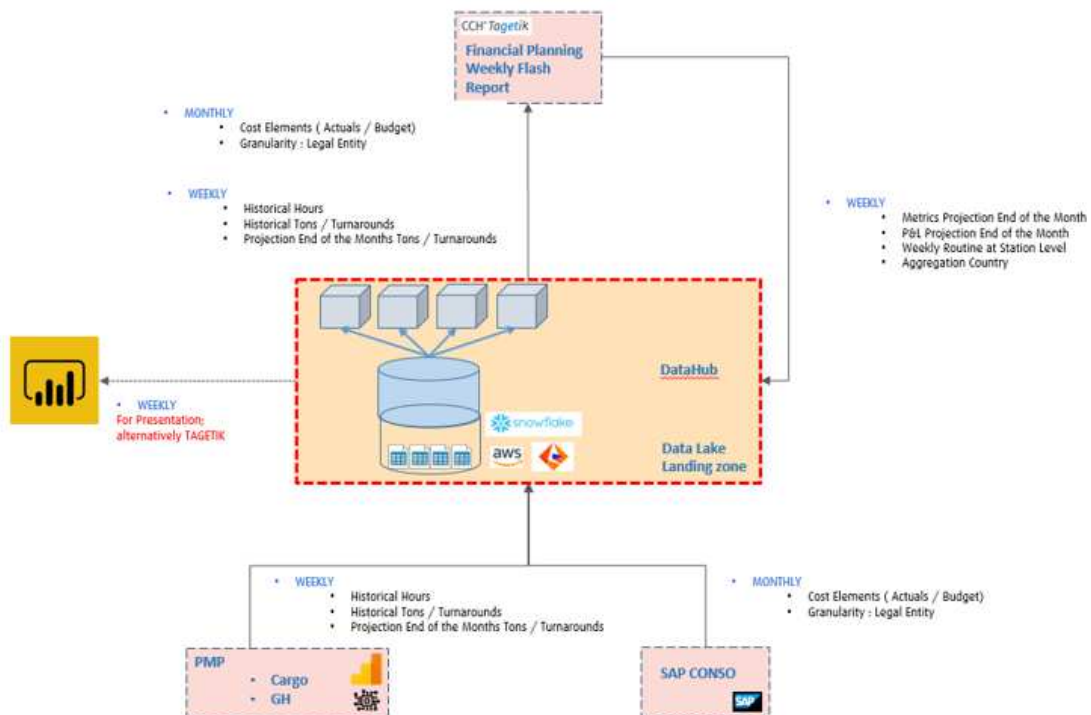
9.6.1 Problemi riscontrate

Durante tutto il progetto abbiamo affrontato il problema della qualità dei dati. Il processo di gestione di attività come la finanza, le risorse umane e le relazioni con i clienti è stato costruito senza tenere conto di ciò che veniva fatto tra di loro, per cui ogni entità ha adottato una propria codifica. Inoltre, l'unico sistema o soluzione matura è PMP - la soluzione Cargo Operations Management ha implementato un proprio MDM per la codifica e la gestione dei riferimenti.

Il data lake è stato costruito sul principio che era necessario migrare tutti i dati dalla soluzione PMP al datawarehouse della soluzione per le operazioni di carico nel data lake replicando la stessa dimensione e le stesse tabelle di fatto. Ci troviamo di fronte al

problema della coerenza dei dati. Nel processo di trasformazione dei dati abbiamo denormalizzato i dati in modo che potessero essere sfruttabili.

Il data lake rende i dati disponibili a tutte le entità dell'azienda, abbiamo anche sviluppato una soluzione Tagetik per l'implementazione di un flash report settimanale basato sui dati proveniente del entità cargo operation e di finanza stoccato nel data lake. Come mostrato nell'architettura sottostante l'architettura del data lake e la soluzione Tagetik per il weekly flash report e il flusso dei dati.



I due processi principali supportati dalla soluzione Tagetik sono: Processo di budget che si svolge annualmente e processo di flash report che viene fatto in modo settimanale.

I dati inseriti nel weekly flash report provengono dal data lake, ma i data master dimensionali sono diversi tra il data lake e la soluzione Tagetik dove il flash settimanale viene sviluppato ad esempio:

Nel data lake, un prodotto è definito da tre attributi: **IMP /LOO /LOC** per Import /Loose / Local. D'altra parte, in Tagetik un prodotto è identificato da un valore dimensionale: **CA_004 per Loose Import**, e si nota che *local* e di *transit* non è finora gestito dalla soluzione Tagetik. In termini di gestione dei master dati, per il weekly flash report, un

amministratore dovrà mantenere la coerenza dei dati e dei valori nel Data Lake e un altro amministratore manterrà i dati, i valori e la mappatura tra il codice Tagetik (CA_004) e il codice Data Lake (IMP/LOO/LOC).

Risk

Dando per scontato che un amministratore debba mantenere i dati anagrafici su base settimanale (in tutte le applicazioni), questo potrebbe essere un rischio se non c'è un processo intorno ai dati anagrafici in una prospettiva "end-to-end", e soprattutto se i valori dimensionali aumentano nel tempo

Ci potrebbero essere delle sviste, se una dimensione non viene mantenuta in questa prospettiva "end-to-end", allora alcuni valori dimensionali "inesistenti" potrebbero fluire da Data Lake a Tagetik e creare elementi di rifiuto nel processo ETL

Secondo la temporalità settimanale del processo, questo può rappresentare un rischio per gli utenti finali di non avere l'intero set di dati ogni settimana a causa dei rifiuti ETL

9.6.2 Miglioramenti

Soluzione 1: WFS definisce gli amministratori e il processo per ogni tool per la creazione/aggiornamento dei valori dimensionali

Per il processo weekly flash report, le operazioni di amministrazione devono essere effettuate prima dell'esecuzione della "chiusura settimanale" da parte degli utenti finali:

1. Creazione di nuovi valori dimensionali (ad esempio, nuovi prodotti, nuovi clienti, nuova entità giuridica/stazione, ecc....)
2. Gestione della tabella di ricerca tra codici Data Lake e codici Tagetik

Pro: Situazione attuale, nessuna "reingegnerizzazione" dei dati anagrafici di Tagetik

Contro: Possibilità di sviste se le operazioni di amministrazione non vengono effettuate su base settimanale, quindi gli articoli rifiutati non sono realmente visibili dagli utenti finali

Soluzione 2: Un'altra soluzione, che permette di ridurre le sviste e i codici "sconosciuti" da Data Lake, sarebbe quella di creare i valori dimensionali in base a quanto c'è nel data dump dei dati Data Lake ogni settimana

In questo modo, Tagetik sarà un sistema dipendente da Data Lake in termini di gestione dei master data

Pro: Riduzione delle sviste e dei rifiuti nel processo ETL (OPS e caricamento delle metriche finanziarie) fatto ogni settimana

Contro: I codici Tagetik nel sistema devono essere sostituiti per il budget

Capitolo 10

Conclusione

10.1 Conclusione

L'obiettivo principale di questa tesi ha avuto come obiettivo maggiore la progettazione e la realizzazione di una piattaforma cloud di Business Intelligence per l'analisi delle performance delle operazioni svolte in magazzini e altre entità in un'azienda di servizio per trasporto Cargo aero, con particolare attenzione al processo alla scelta di una soluzione di data warehousing. L'importanza di questa trasformazione del sistema informativo per renderlo maturo permette alle aziende di trasformare l'enorme quantitativo di dati generato, in informazioni volte al miglioramento della capacità di prese delle decisioni. L'implementazione ha seguito un processo interno per la scelta di una soluzione componente la piattaforma BI sul mercato composto dalle seguenti componenti di soluzioni: Integrazione dei dati, cloud datawarehouse e soluzioni di reporting. Per l'esecuzione del progetto, le classiche fasi di progettazione di un datawarehouse: l'analisi dei requisiti, la progettazione concettuale e logica, le fasi di estrazione, trasformazione e caricamento dei dati. L'attività di estrazione e trasformazione dei dati ha richiesto maggiore attenzione, con lo scopo di renderli aderenti alla logica di business del sistema di analisi. Durante la fase di esecuzione del progetto, sono emerse incongruenze tra i dati che hanno portato ad alcune modifiche del modello e alle logiche di calcolo degli indicatori; dove possibile le richieste sono state trasformate nel modello e alle business rule per il calcolo degli indicatori; dove possibile le richieste sono state inserite nel modello attuale, mentre quelle non immediate; dove possibile le richieste sono state rimandate agli sviluppi successivi. I benefici generate dal progetto sono molteplici. In precedenza, l'ufficio del personale dei magazzini distribuiva report che venivano generati manualmente a partire dai sistemi presenti in ciascuna paese, gli impiegati doveva quindi svolgere complesse e onerose operazioni di integrazione di basi di dati. Inoltre, tali operazioni erano limitate ai sistemi sorgenti interni alla realtà aziendale: non era possibile infatti incrociare le informazioni prodotte dai processi HR PEOPLE e CARGO OPERATION e FINANCE con le ore e il costo

delle operazioni delle risorse impiegate nel magazzino svolgendo una attività. Il sistema realizzato ha quindi complessivamente soddisfatto le aspettative, ottenendo un riscontro più che positivo da parte degli utenti finale: il top management sia durante lo sviluppo che nei periodi successivi alla messa in produzione.

10.2 Sviluppi futuri

WFS Global ha oltre manifestato interesse nell'utilizzo della piattaforma BI: Il processo HSSE per il dipartimento di sicurezza, CQC per la gestione della relazione cliente sarà esteso implementando dei ETL, per il caricamento dei dati. L'approccio bottom up adottato durante il progetto, rende la piattaforma BI più flessibile per l'aggiunto di nuovi data Mart.

Uno sviluppo ulteriore da fare è la messa in piedi di una propria governance dei dati. Applicando i *Principi di governance dei dati*. Secondo il Data Governance Institute, otto principi sono al centro di tutti i programmi di governance e di gestione dei dati di successo:

1. Tutti i partecipanti alla governance dei dati devono avere integrità nei loro rapporti reciproci. Devono essere sinceri e disponibili nel discutere i driver, i vincoli, le opzioni e l'impatto delle decisioni relative ai dati.
2. La governance dei dati e i processi di stewardship richiedono trasparenza. Deve essere chiaro a tutti i partecipanti e ai revisori come e quando le decisioni e i controlli relativi ai dati sono stati introdotti nei processi.
3. Le decisioni, i processi e i controlli relativi ai dati soggetti a governance dei dati devono essere verificabili. Essi devono essere accompagnati da una documentazione a supporto dei requisiti di revisione operativa e di conformità.
4. La governance dei dati deve definire chi è responsabile delle decisioni, dei processi e dei controlli interfunzionali relativi ai dati.
5. La governance dei dati deve definire chi è responsabile delle attività di stewardship che sono responsabilità dei singoli collaboratori e dei gruppi di data stewards.
6. Controlli e bilanci. La governance dei dati definirà le responsabilità in modo da introdurre controlli ed equilibri tra i team aziendali e tecnologici e tra coloro che creano e raccolgono le informazioni, coloro che le gestiscono, coloro che le utilizzano e coloro che introducono standard e requisiti di conformità.
7. La governance dei dati introdurrà e sosterrà la standardizzazione dei dati aziendali.

8. Gestione del cambiamento. La governance dei dati supporterà attività proattive e reattive di gestione del cambiamento per i valori dei dati di riferimento e la struttura/utilizzo dei dati anagrafici e dei metadati.

Ruoli di governance dei dati

Ogni azienda compone la propria governance dei dati in modo diverso, ma ci sono alcuni punti in comune.

Comitato direttivo (Steering committee)

Poiché la governance dei dati si estende a tutta l'impresa, generalmente inizia con un comitato direttivo composto da alti dirigenti, spesso individui di livello C o vicepresidenti responsabili di linee di business.

Proprietario dei dati (Data owner)

I proprietari dei dati sono individui responsabili di garantire che le informazioni all'interno di uno specifico dominio di dati siano governate attraverso sistemi e linee di business. In genere sono membri del comitato direttivo, anche se potrebbero non essere membri votanti. I proprietari dei dati sono responsabili per:

- Approvare i glossari dei dati e altre definizioni di dati
- Garantire l'accuratezza delle informazioni in tutta l'impresa
- Attività dirette sulla qualità dei dati
- Revisione e approvazione degli approcci, dei risultati e delle attività di gestione dei dati anagrafici
- Lavorare con altri proprietari di dati per risolvere i problemi relativi ai dati
- Revisione di secondo livello per le questioni individuate dai responsabili della gestione dei dati
- Fornire al comitato direttivo input su soluzioni software, politiche o requisiti normativi del loro dominio di dati

Responsabile dei dati (Data steward)

Gli amministratori dei dati sono responsabili della gestione quotidiana dei dati. Sono esperti in materia (PMI) che comprendono e comunicano il significato e l'uso delle informazioni, dice il Templare, e lavorano con altra data steward in tutta l'organizzazione come organo di governo per la maggior parte delle decisioni sui dati. I data steward sono responsabili per:

- Essere SMEs (Subject Matter Expert) per il loro dominio di dati
- Identificare i problemi relativi ai dati e collaborare con altri amministratori per risolverli
- In qualità di membro del data steward council
- Proporre, discutere e votare le politiche sui dati e le attività della commissione
- Segnalazione al proprietario dei dati e ad altre parti interessate all'interno di un dominio di dati
- Lavorare in modo interfunzionale tra le varie linee di business per garantire che i dati del loro dominio siano gestiti e compresi.

Glossario

<i>Acronimo</i>	<i>Descrizione</i>
AA	American Airline
ACID	Atomicity, Consistency, Isolation, Durability
API	Key Performance Indicator
AWS	Amazon Web Service
AI	Artificial Intelligence
BUP	Build Up Pallet
B2B	Business to Business
CAS	Cargo Airport Service
CARGO IMP	Cargo Interchange Message Procedures
CDG	Charles De Gaulles
CEO	Chief Executive Officer
CIO	Chief Information Officer
CQC	Cargo Quality Center
DGR	Dangerous Goods Regulations
DMSA	Data Management Solutions for Analytics
EDI	Electronic data interchange
EFS	European Flight Service
EMEA	Europe, Middle East, and Africa
EMEEA	Europe, Middle East, Africa, and Asia

EiPaaS	Enterprise Integration Platform as a Service
ETL	Extract, Transform, Load
ETA	Estimated Time for Arrival
FH	France Handling
FFM	Flight Manifest
FSU	Freight Status Update
GDPR	General Data Protection Regulation
GHA	Ground Handling Agent
HR	Human Resources
HSSE	Health, Safety, Security and Environment
HDFS	Hadoop Distributed File System
IATA	International Air Transport Association
IICS	Informatica Intelligent Cloud Services
IPAAS	Integration Platform as a Service
IT	Information Technologies
KPI	Key Performance Indicator
LAN	Local Area Network
LDW	Logical Data Warehouse
MAN	Manifested
HIP	Hybrid Integration Platform
MDM	Master Data Management

ML	Machine Learning
MPP	Massively Parallel Processing
NOTOC	Notification to Captain
OCR	Optimized Row Columnar
ODS	Operational Data Store
ORY,	Orly Airport
PMI	Piccola e media impresa
PMP	Performance Management Platform
SAAS	Software as a Service
SAN	Storage Area Network
SFS	Sociétés de Fret et de Service
SSRS	SQL Server Reporting Services
SFTP	SSH File Transfer Protocol
ULD	Unit Load Device
WFS AFS	WFS African Flight Services
WFS	World Flight Service
YTD	Years To Date
YTY	Years To Years

Bibliografia

- [1] About Us <https://www.wfs.aero/about-us/> 2020
- [2] <https://www.iata.org/> 2020
- [3] Quick Guide to Freight Status Update (FSU) Messages <http://www.ccnhub.com.au/fsu/> , 2020.
- [4] Sabeen Suharwardy *Cargospot Handling User Manual 3.0*, 10 February 2009
- [5] Performance Manager Platform – *PMP Technical Documentation and Definitions*, 2020
- [6] O'Reilly report, Castanedo F., Gidley S., *Understanding Metadata: Create the Foundation for a Scalable Data Architecture*, Febbraio 2017
- [7] Cloud Datawarehouse platform, <https://www.snowflake.com/blog/snowflake-recognized-as-a-leader-by-gartner-third-consecutive-year-positioned-in-the-magic-quadrant-report/>, 2020
- [8]
- [9] Inmon, W.H., *Building the Data Warehouse*, John Wiley & Sons, 1992.
- [10] Kimball, R. and Ross, M., *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, John Wiley & Sons, 2002.
- [11] Albano, A., *Decision Support Databases Essentials*, University of Pisa, 2014.
- [12] *Information Cloud Integration* <https://www.informatica.com/products/cloud-integration.html> ,2020.
- [13] *Microsoft Developer Network*, <https://msdn.microsoft.com>, 2014.

Elenco figure

Figure 3.1: Architectura della piattaforma PMP.....	23
Figure 3.2: Dimensionne del datawarehouse di PMP.....	24
Figure 3.3: Tabelle dei Fatti del Datawarehouse PMP	25
Figure 3.4: Esempio di rapporto prodotto da PMP.....	26
Figure 3.5: Esempio di rapporto prodotti da PMP sul toneggio mensile	26
Figure 3.6: Rapporto finale con misure ops e finanze	27
Figure 3.7: Architectura del Data Lake	31
Figure 3.8: Caratteristiche del Data Lake	34
Figure 3.9: Caratteristiche del data Swamp.....	34
Figure 3.10: Architecture Target della piattaforma BI	37
Figure 4.1: Esempio di un mockup del dashboard Cargo Operation.....	45
Figure 4.2: Esempio di un mockup della dashboard Manpower Cargo	45
Figure 4.3: Esempio di un mockup della dashboard Finance.....	46
Figure 4.4: Magic Quadrant per Enterprise Integration Platform as a Service	49
Figure 4.5: Tipologie di stoccaggio	68
Figure 4.6: Magic Quadrant per il cloud datawarehouse.....	69
Figure 4.7: Criteri di selezione di un cloud datawarehouse	75
Figure 4.8: Magic Quadrant per la reportistica.....	76
Figure 4.9: Architettura target	84
Figure 4.10: Architettura di Snowflake	86
Figure 4.11: Architettura tipica di MPP / Cluster	88
Figure 4.12: Snowflake Multi-Cluster Share Data Architecture	89
Figure 4.13: Consolidamento di diversi formati di dati.....	92
Figure 4.14: Condivisione dei dati del fiocco di neve - Nessuna copia o spostamento dei dati.....	92
Figure 5.1: Funzioni aeroportuali	95
Figure 5.2: Struttura rete logistica aeroportuale	97
Figure 5.3: Aree funzionali di un magazzino cargo aeroportuale	98

Figure 5.4: Area avio-camionate lato land side (Milano Malpensa)	99
Figure 5.5: Isola di spallettizzazione	100
Figure 5.6: Stacker (Milano Malpensa)	101
Figure 5.7: Staging Area (Milano Malpensa)	101
Figure 5.8: Area di interfaccia con i trasporti terrestri (Milano Malpensa)	102
Figure 5.9: 44 PKI per entità e per business line	104
Figure 5.10: Misure delle attività Cargo	105
Figure 5.11: Profilo attività Cargo	106
Figure 6.1: Schema concettuale del data mart operazione giornaliero	120
Figure 6.2: Schemas concettuale del data mart operation mensile.....	122
Figure 6.3: Schemas concettuale Data Mart Finance	128
Figure 6.4: Schemas concettuale Data Mart Headcounts	130
Figure 6.5: Schemas concettuale Data Mart Hires.....	132
Figure 6.6: Schemas concettuale Data Mart Leaves Mensile.....	134
Figure 7.1: Schema concettuale del data mart Operation Daily	138
Figure 7.2: Schema logico del data mart Cargo Operation Daily	140
Figure 7.3: Schema concettuale del data mart Operation Budget Month.....	140
Figure 7.4: Schema logico del data mart Cargo Operation Month.....	141
Figure 7.5: Schema concettuale finale del data mart WORKFORCE DAY	142
Figure 7.6: Schema logico del data mart Workforce Day	142
Figure 7.7: Schema concettuale del data mart Workforce Budget Month	143
Figure 7.8: Schema logico del data mart Workforce BUD Month.....	144
Figure 7.9: Schema concettuale del data mart <i>Headcounts Month</i>	145
Figure 7.10: Schema logico del data mart <i>Headcounts Month</i>	146
Figure 7.11: Schema concettuale del data mart Hires	146
Figure 7.12: Schema logico del data mart Hires.....	147
Figure 7.13: Schema concettuale del data mart Leaves.....	148
Figure 7.14: Schema logico del data mart Leaves	148
Figure 7.15: Schema concettuale del data mart Finance	149
Figure 7.16: Schema logico del data mart Finance.....	150
Figure 7.17: Schema logico del Datawarehouse.....	150

Figure 8.1: Il processo ETL: dai sistemi sorgenti al datawarehouse	152
Figure 8.2: Interfaccia di informatica Cloud	155
Figure 8.3: Dashboard di Informatica Cloud	155
Figure 8.4: Interfaccia per la creazione di nuovi asset	156
Figure 8.5: Tipologie di asset in Informatica Cloud.....	157
Figure 8.6: Interfaccia per la creazione di nuovi asset di tipo Mapping	158
Figure 8.7: Interfaccia per la creazione di nuovi asset di tipo Tasks	159
Figure 8.8: Esempio di Mapping task in Informatica Cloud	160
Figure 8.9: Schermata per la creazione dei Taskflow.....	161
Figure 8.10: Esempio di taskflow	162
Figure 8.11: Interfaccia di creazione di asset di tipo components	163
Figure 8.12: Flusso di alimentazione del datawarehouse	164
Figure 8.13: Flussi di dati nella piattaforma di datawarehouse	166
Figure 8.14: Lista dei tasks create durante il progetto.....	169
Figure 8.15: Paranoramica dei taskflow	170
Figure 8.16: Esempio di sequenza di taskflow	171
Figure 8.17: Esempio di taskflow per il caricamento dei dati di Cargo Operation	172
Figure 8.18: Processo ETL per l'entità risorse umane	174
Figure 8.19: Processo ETL per l'entità Cargo Operation.....	174
Figure 8.20: Processo ETL per l'entità finance.....	181
Figure 9.1: Power BI Mobile nelle piattaforme mobili supportate.....	185
Figure 9.2: Ecosistema di Power BI	186
Figure 9.3: Pagina iniziale	187
Figure 9.4: Query editor.....	188
Figure 9.5: Cronologia delle query effettuate	189
Figure 9.6: Grafici preinstallati in Power BI Desktop	190
Figure 9.7: Istogramma.....	190
Figure 9.8: Lista degli indicatori e misure da implementare.....	193
Figure 9.9: Interfaccia utente di Power BI.....	194
Figure 9.10: Dash board Power BI	195
Figure 9.11: Esempio di liste di KPI Cargo Operation.....	195

Figure 9.12: Filtro per la selezione	196
Figure 9.13: Zone di Power BI per il display dei KPI	196
Figure 9.14: KPI da implementare per il Cargo Operation	198
Figure 9.15: Dashboard principale dell'entità Spagna	198
Figure 9.16: Tendenza Numero di tonnellate gestite.....	200
Figure 9.17: KPI presentando il numero di kili per ora rispetto all'anno precedente	200
Figure 9.18: KPI presentando il numero di AWBs per ora rispetto all'anno precedente	201
Figure 9.19: KPI per tasso di ore straordinari rispetto all'anno precedente.....	202
Figure 9.20: KPI presentando il tasso flessibilità	203
Figure 9.21: KPI sul costo della mano d'opera per ore.....	204
Figure 9.22: KPI sul costo della mano d'opera includendo le ore straordinarie	205
Figure 9.23: KPI sul costo della mano d'opera includendo le ore di terzi.....	206
Figure 9.24: Dashboard per Manpower Cargo/People KPI.....	207
Figure 9.25: KPI sul turnover del personale	208
Figure 9.26: Dashboard presentando il tasso di malattie.....	208
Figure 9.27: KPI sul tasso di malattie rispetto all'anno scorso	209
Figure 9.28: Dashboard presentando il KPI le ore straordinari	210
Figure 9.29: KPI sulle ore straordinari rispetto all'anno scorso	210
Figure 9.30: KPI sul costo della mano d'opera per ore.....	211
Figure 9.31: KPI sul costo della mano d'opera di WFS includendo le ore straordinarie	213
Figure 9.32: KPI sul costo della mano d'opera includendo le ore di terzi.....	214
Figure 9.33: KPI sul costo della mano d'opera includendo le ore di terzi	218
Figure 9.34: KPI sulla produzione del tonnellaggio YTD.....	219
Figure 9.35: KPI sul costo totale per turno.....	222
Figure 9.36: KPI sulla contribuzione del tonnellaggio YTD.....	223

Elenco Tabelle

Tableau 2.1:	Cronologia di WFS Global	10
Tableau 2.2:	Standard dei messaggi Flight Status Update nel Aero Cargo Business ..	16
Tableau 2.3:	Elenco delle principali Codice d'importazioni FSU.....	17
Tableau 3.1:	KPI Safety and Security	19
Tableau 3.2:	KPI Finance	19
Tableau 3.3:	KPI Customer Satisfaction	20
Tableau 3.4:	KPI Cargo Operation	20
Tableau 3.5:	KPI Human resources	20
Tableau 3.6:	Confronto tra datawarehouse e data lake	30
Tableau 3.7:	Tabella riassuntiva del Gap Analysis	41
Tableau 4.1:	KPI entità Cargo Operation	44
Tableau 4.2:	KPI entità risorse umana.....	45
Tableau 4.3:	KPI entità Finance	46
Tableau 4.4:	Forze e debolezze delle IPaaS	57
Tableau 4.5:	Scoring delle soluzioni ETL	57
Tableau 4.6:	Confronto tra funzionalità di Azure e Amazon per lo storage	60
Tableau 4.7:	Esito della selezione tra le soluzioni Amazon e Azure	60
Tableau 4.8:	Scoring per la soluzione finale	75
Tableau 4.9:	Criteri di selezioni per le soluzioni di reportistica	83
Tableau 4.10:	Esito per le soluzioni di reportistica	83
Tableau 5.1:	Induttore di costo per l'attività cargo	107
Tableau 5.2:	Tendenza - Numero di tonnellate gestite.....	107
Tableau 5.3:	KPi30 Chilogrammi per ora	108
Tableau 5.4:	KPI i30 AWB per Ora	108
Tableau 5.5:	Ore pagate / Turni.....	108
Tableau 5.6:	% di paghe su totale ore pagate	109
Tableau 5.7:	Costo per ora pagata (libro paga + lavoro straordinario + terzi)	109
Tableau 5.8:	% Magazzino di lavoro.....	109
Tableau 5.9:	L'approccio top-down	111
Tableau 5.10:	L'approccio bottom-up	112

Tableau 5.11: Metriche della business line Cargo.....	113
Tableau 5.12: Metriche dell'entità Finanze	114
Tableau 6.1: Specifica dei requisiti del fatto Operazione giornaliero	118
Tableau 6.2: Dimensione associato al fatto Operazione giornaliera	119
Tableau 6.3: Misure della tabella del fatto Operazione giornaliero	119
Tableau 6.4: Specifica dei requisiti del fatto Operazione giornaliero	121
Tableau 6.5: Dimension del data mart Cargo Operation Mensile	121
Tableau 6.6: Misure del data Mart Cargo Operation Mensile.....	122
Tableau 6.7: Specifica del data Mart Workforce BUD M.....	123
Tableau 6.8: Tabella della dimensione del data mart Workforce Mensile.....	123
Tableau 6.9: Tabella Misure Data Mart Workforce Mensile	124
Tableau 6.10: Schemas concettuale Workforce Mensile	124
Tableau 6.11: Specifica del data Mart Workforce BUD_D	125
Tableau 6.12: Tabella della dimensione del data mart Workforce Giornaliera	125
Tableau 6.13: Tabella Misure Data Mart Workforce Giornaliera.....	125
Tableau 6.14: Schemas concettuale Workforce Giornaliero	126
Tableau 6.15: Specifica del data Mart Finance	127
Tableau 6.16: Tabella della dimensione del data Mart Workforce Giornaliera.....	127
Tableau 6.17: Tabella Misure Data Mart Finance.....	128
Tableau 6.18: Specifica del data Mart Headcounts	129
Tableau 6.19: Tabella della dimensione del data mart Headcounts	130
Tableau 6.20: Tabella della Misura per il Data Mart Headcounts	130
Tableau 6.21: Specifica del data Mart Hires Mensile	131
Tableau 6.22: Tabella della dimensione per il Data Mart Hires Mensile.....	131
Tableau 6.23: Tabella della Misura per il Data Mart Headcounts.....	132
Tableau 6.24: Specifica del data Mart Leavers Mensile	133
Tableau 6.25: Tabella della dimensione del data mart Leaves Mensile.....	133
Tableau 6.26: Tabella della Misura per il Data Mart Leaves mensile	133
Tableau 7.1: Tabella riepilogativa delle dimensioni in comune tra i data mart	135
Tableau 7.2: Tabella riepilogativa finale delle dimensioni in comune tra i data mart	136

Tableau 7.3: Tabella riepilogativa delle dimensioni in comune tra i data mart dell'entità cargo operation.....	137
Tableau 7.4: Tabella riepilogativa finale delle dimensioni in comune tra i data mart dell'entità cargo operation	138
Tableau 7.5: Tabella riepilogativa delle dimensioni in comune tra i data mart dell'entità HR People	144
Tableau 7.6: Tabella riepilogativa delle dimensioni in comune tra i data mart dell'entità HR People	145
Tableau 8.1: Sequenza generale dei task per l'integrazione dei dati di una sorgente	171
Tableau 8.2: Sequenza dei task da 48 a 60	172
Tableau 8.3: Tabella dettagliata del processo ETL dell'entità Cargo Operation.....	180
Tableau 8.4: Procedure automatizzate.....	181