



**Politecnico  
di Torino**

# Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale

Anno Accademico 2023/2024

## **Integrazione e armonizzazione dei dati in un datawarehouse aziendale**

Un progetto di business intelligence applicato

Relatrice:

prof.ssa Tania Cerquitelli

Candidato:

Luca Perilli

Tutor aziendale:

dott. Andrea Bolatto



# Indice

<b>Indice</b> .....	3
<b>Introduzione all’elaborato</b> .....	6
<b>1. Cenni di Business Intelligence</b> .....	8
<b>2. Gli applicativi software utilizzati</b> .....	10
<b>2.1 La piattaforma di front-end - Qlik Sense</b> .....	10
<b>2.1.1 Panoramica generale</b> .....	10
<b>2.1.2 Interfaccia grafica ed elementi chiave</b> .....	12
<b>2.2 La gestione dati in backend</b> .....	15
<b>2.2.1 Greenplum</b> .....	15
<b>2.2.2 DBeaver</b> .....	17
<b>3. Il progetto aziendale</b> .....	20
<b>3.1 Panoramica del mercato e del cliente</b> .....	21
<b>3.1.1 Il contesto</b> .....	21
<b>3.1.2 L’azienda cliente - CAAB</b> .....	22
<b>3.1.3 I partner aziendali coinvolti</b> .....	23
<b>3.2 Architettura del datawarehouse aziendale</b> .....	25
<b>3.2.1 Introduzione</b> .....	25
<b>3.2.2 Il Customer Relationship Management</b> .....	27
<b>3.2.3 Processo di caricamento dati (ELT) in Greenplum</b> .....	29
<b>3.2.4 Struttura del modello dati</b> .....	35
<b>3.2.5 Caricamento e integrazione dei dati su Qlik Sense</b> .....	39
<b>3.3 Lo sviluppo del progetto</b> .....	43
<b>3.3.1 Lo scopo</b> .....	43
<b>3.3.2 Analisi preliminari</b> .....	45
<b>3.3.3 Unificazione dati e reportistica</b> .....	53

<b>3.3.4 Modifica struttura del datawarehouse</b> .....	59
<b>3.3.5 Test sulle modifiche</b> .....	62
<b>3.3.6 Recupero storico e rilascio in produzione</b> .....	64
<b>3.3.7 Cronoprogramma</b> .....	66
<b>4. Conclusioni e sviluppi futuri</b> .....	67
<b>Sitografia</b> .....	68
<b>Acronimi e significati</b> .....	69
<b>Figure</b> .....	70



## Introduzione all'elaborato

L'obiettivo del seguente elaborato di tesi è quello di presentare un caso applicato relativo ad un progetto di *data integration* realizzato all'interno di una nota azienda di consulenza internazionale.

Questo è stato svolto nell'ambito della consulenza ad un noto istituto di credito italiano ed europeo, e specificatamente alla divisione della banca dedicata ai servizi finanziari legati all'acquisto di veicoli.

Nello specifico, il progetto ha riguardato l'analisi e lo sviluppo di un nuovo processo di integrazione dei dati, volto ad includere un nuovo flusso all'interno del sistema informativo esistente. Tale iniziativa mirava a standardizzare e rendere organica l'analisi dei dati provenienti dai molti fornitori IT dell'azienda cliente.

L'integrazione dei dati appartenenti a questo nuovo flusso nel datawarehouse (DWH) si è infatti resa imprescindibile per l'azienda committente, in quanto la sua divisione di business richiedeva l'armonizzazione e la sistematizzazione dell'analisi dei dati provenienti da varie nazioni europee, e non più esclusivamente dall'Italia, a causa dell'accelerata espansione aziendale in territori internazionali.

Precedentemente all'implementazione di questo progetto, l'analisi dati era caratterizzata da una notevole inefficienza e da una mancanza di integrazione con le preesistenti strutture informative, comportando ritardi e maggiori costi dovuti alla disfunzionalità del processo nel suo complesso. La realizzazione di questo progetto ha quindi permesso all'azienda di affrontare con successo le sfide legate alla gestione di una mole crescente di dati provenienti da diversi contesti nazionali.

Attraverso l'implementazione dei dati in maniera più efficiente e sistematica, l'azienda è ora in grado di gestire in modo ottimale l'analisi dei dati provenienti da una vasta gamma di fonti, consentendo una visione più completa e accurata delle proprie attività commerciali. In tal modo, il progetto non solo ha contribuito a ridurre i ritardi e i costi associati alla gestione dei dati, ma ha anche fornito una base solida per migliorare la qualità delle decisioni aziendali e sostenere la crescita e l'espansione internazionale dell'azienda nel lungo periodo.

Il percorso di sviluppo del progetto ha richiesto quindi un'analisi approfondita delle fasi tecniche, delle strategie di implementazione e delle funzionalità finali del sistema, ponendo particolare attenzione alle esigenze e alle aspettative dell'azienda cliente.

Per fare ciò, si andrà ad esaminare il percorso di sviluppo che ha portato alla definizione della soluzione finale implementata nel sistema

aziendale del cliente, concentrando l'attenzione sui passaggi tecnici chiave, le strategie di implementazione e le funzionalità dell'applicazione risultante.

L'analisi non si limita però alla fase di sviluppo, ma include anche aspetti accessori relativi al progetto, ma fondamentali per la sua realizzazione e per comprendere a pieno il contesto di lavoro. È infatti essenziale considerare il contesto del progetto e le motivazioni funzionali che hanno guidato le scelte di progettazione e sviluppo, in risposta alle esigenze del cliente, con il quale è stato stabilito un dialogo costante al fine di migliorare efficacemente il risultato finale basandosi sui feedback e le verifiche effettuate durante l'intero periodo di sviluppo.

Inoltre, la tesi fornirà una rapida panoramica degli aspetti organizzativi del progetto, analizzando i tempi medi dedicati all'analisi e allo sviluppo dello stesso, e di alcune delle motivazioni alla base di alcune scelte strategiche aziendali e di progetto.

Saranno anche esposte informazioni dettagliate sulla tecnologia e il sistema implementati, con particolare focus sul software Qlik Sense, i principali elementi e processi legati alla gestione del datawarehouse e sui principali tool di frontend e backend utilizzati.

# 1. Cenni di Business Intelligence

Per poter illustrare al meglio il progetto svolto, in questo capitolo si analizza brevemente la nozione e l'applicazione della business intelligence in ambito aziendale, in modo da esporne lo stato dell'arte attuale.

Il concetto di business intelligence ha le sue radici negli anni '60 e '70 del secolo scorso, quando le prime aziende iniziarono a riconoscere il valore dei dati per la gestione aziendale. Tuttavia, nel corso degli anni la BI si è evoluta molto rapidamente, incorporando tecnologie sempre più avanzate per l'analisi e la visualizzazione dei dati, di pari passo con l'avvento del World Wide Web e l'esplosione dei dati digitali (Big Data) negli anni 2000.

Con il termine *Business Intelligence (BI)* ci si riferisce infatti al ramo dell'analisi dati che si occupa di raccogliere, analizzare e trasformare dati aziendali in informazioni significative e utili per prendere decisioni informate. Questo è un processo complesso che coinvolge l'uso di strumenti software e tecniche analitiche per estrarre, elaborare e presentare i dati al fine di supportare le decisioni aziendali. Tale processo richiede diverse fasi, tra cui:

- raccolta dati da sistemi IT interni ed esterni (clienti).
- data pre-processing, per operazioni di pulizia, integrazione e verifica della qualità dei dati raccolti.
- interrogazione dei dati, tipicamente tramite linguaggio SQL.
- creazione di visualizzazioni, dashboard e report, solitamente tramite programmi di BI dedicati (es: MS PowerBI o Qlik Sense).

Le caratteristiche fondamentali di un sistema di Business Intelligence includono:

- facilità d'uso e velocità, poiché il sistema deve essere in grado di gestire grandi quantità di dati in un formato facilmente comprensibile e garantire tempi di risposta rapidi.
- integrazione e storicizzazione, infatti i dati provenienti da varie fonti, sia interne che esterne all'azienda, devono essere integrati e bisogna mantenere un registro storico delle informazioni.
- identificazione di trend e anomalie, tramite strumenti interattivi che devono consentire agli utenti di individuare facilmente le tendenze e le deviazioni significative nei dati.
- adattabilità nel tempo, ovvero il sistema deve potersi adattare alle evoluzioni tecnologiche in modo da capitalizzare le nuove opportunità di miglioramento aziendale.

- indipendenza dall'IT, infatti gli utenti devono poter creare autonomamente i propri report senza dover dipendere dall'area IT, abbracciando il concetto di *self-service BI*; ciò vale specialmente per visualizzazioni non troppo complesse.
- sicurezza, dovendo garantire un controllo rigoroso e selettivo (rispetto al ruolo e al progetto assegnato) all'accesso ai dati, essendo questi solitamente dei dati sensibili.

Quindi, il concetto di BI è strettamente correlato a quello di *Big Data*, con cui ci si riferisce a insiemi di dati estremamente vasti e complessi che superano la capacità delle tradizionali tecniche di gestione e analisi dei dati; essi sono caratterizzati da volumi ingenti, alta velocità di generazione e accumulo all'interno di database, ma anche dalla diversità dei formati dei dati. Al giorno d'oggi, questi ultimi provengono infatti da una notevole molteplicità di fonti diverse, ognuna con uno specifico formato e frequenza di aggiornamento.

È dunque evidente che l'unico metodo per l'analisi di ingenti quantità di dati di questo tipo è proprio la business intelligence che, per questa ragione, è diventata fondamentale per le aziende moderne, indipendentemente dalla dimensione e dal settore in cui esse operano, permettendo:

- supporto alle decisioni basate sui dati, consentendo l'identificazione di tendenze, opportunità e sfide, supportando decisioni più informate.
- competitività, in quanto si riescono così a comprendere meglio i bisogni dei clienti, ottimizzare le operazioni e rispondere rapidamente ai cambiamenti del mercato.
- personalizzazione ed esperienza cliente, migliorando la soddisfazione e la fedeltà del cliente sul lungo periodo.
- riduzione dei costi, poiché le aziende riescono a identificare inefficienze e aree di spreco, consentendo di prendere misure correttive per ridurre i costi operativi.
- scalabilità, in quanto essa è l'unico modo per trattare ed analizzare enormi quantità di dati provenienti da fonti diverse.

In sintesi, la business intelligence è diventata sempre più importante per le aziende moderne a causa del maggior uso dei dati e della necessità di prendere decisioni informate in un ambiente aziendale altamente competitivo. Le aziende che investono nella BI hanno infatti come obiettivo quello di ottenere un vantaggio strategico e migliorare la loro capacità di adattamento ai cambiamenti del mercato, aumentando i ricavi, migliorando l'efficienza operativa e raggiungendo una maggiore competitività commerciale.

## **2. Gli applicativi software utilizzati**

Per ragioni di chiarezza espositiva, prima di passare alla descrizione di dettaglio del contesto aziendale e di progetto, in questo capitolo si analizzano brevemente gli applicativi software utilizzati in ambito aziendale per lo sviluppo del progetto, in modo da chiarirne funzionalità, logiche e vantaggi.

### **2.1 La piattaforma di front-end - Qlik Sense**

La piattaforma scelta dall'azienda come strumento di front-end per l'analisi dei dati è Qlik Sense, una potente applicazione di business intelligence (BI) sviluppata da QlikTech International, azienda leader nel settore dell'analisi dati, che ha l'obiettivo di sviluppare soluzioni end-to-end all'avanguardia che si adattino alle esigenze specifiche.

Si procede dunque ora a presentare il software e ad esplorarlo in dettaglio, concentrandosi sulle funzioni principali ed effettuando un rapido confronto con altri software simili e i metodi tradizionali di analisi dati.

#### **2.1.1 Panoramica generale**

Qlik Sense è stato lanciato nel 2014 da Qlik come successore di QlikView, un altro noto strumento di BI. La nascita di Qlik Sense segna una svolta nell'approccio fornito dall'azienda all'analisi dati, poiché si concentra sulla creazione di dashboard interattive e visualizzazioni analitiche, rendendo l'analisi più accessibile agli utenti finali. Nel corso degli anni, il software ha subito numerose iterazioni ed è diventato uno dei principali strumenti di BI utilizzati in tutto il mondo.

Qlik Sense offre una vasta gamma di funzioni e caratteristiche, tra cui le principali sono:

- creazione di dashboard interattive, in cui gli utenti possono facilmente creare dashboard personalizzate trascinando e rilasciando elementi visivi come grafici, tabelle e mappe; questo tipo di approccio, grafico ed intuitivo, semplifica quindi notevolmente l'approccio al programma e all'analisi nel complesso.

- associatività dei dati, che rappresenta una delle caratteristiche distintive di Qlik Sense, consentendo agli utenti di esplorare in modo dinamico i dati e di scoprire relazioni nascoste. Infatti, mentre i sistemi tradizionali effettuano la ricerca dei dati in maniera sequenziale, Qlik Sense consente l'accesso non sequenziale, dunque indipendentemente dalla posizione in memoria del dato, in maniera molto simile a come avviene nei processi cognitivi umani.
- analisi self-service, infatti gli utenti possono eseguire analisi senza la necessità di competenze tecniche avanzate, in qualunque momento, consentendo a chiunque di prendere decisioni basate sui dati; il software suggerisce infatti anche tipologie di analisi e grafici in base al tipo di dato utilizzato. È quindi evidente come questa soluzione permetta così di ottenere visualizzazioni rapide, flessibili e facili da utilizzare, al contrario di quanto avviene nei tradizionali cubi OLAP.
- analisi *in-memory*, quest'innovativa tecnologia di analisi dei dati fa sì che tutti i dati vengano caricati in maniera dinamica nella memoria e le metriche siano definite poco alla volta, una volta che l'utente esegue le selezioni. Inoltre, durante il caricamento dei dati da diverse fonti, QlikSense crea un modello dati *in-memory* che garantisce un accesso rapido a milioni di dati con un minor consumo di spazio, ottimizzando le informazioni, eliminando i valori ridondanti e stabilendo automaticamente le relazioni tra le tabelle. Il risultato di questa struttura *in-memory* è che il software esegue calcoli ad alta velocità e offre funzionalità di analisi multidimensionale a tutti gli utenti, permettendo di ottenere risposte immediate alle query, di elaborare grandi dataset e di rispondere alle esigenze di numerosi utenti in modo efficiente.
- gestione dei permessi e controllo della visibilità, questa funzionalità è fondamentale per garantire che solo gli utenti autorizzati possano accedere e visualizzare determinati dati sensibili o dashboard personalizzate. Qlik Sense offre infatti un robusto sistema di gestione dei permessi basato su ruoli e regole, impostate dagli amministratori del sistema, in modo che ognuno, in base alla sua mansione, responsabilità e mercato di riferimento possa avere accesso a dashboard, informazioni o insiemi di dati specifici; ciò aumenta quindi notevolmente la sicurezza dei dati e l'efficienza operativa aziendale.
- integrazione dei dati, infatti Qlik Sense può essere facilmente integrato con diverse fonti dati, consentendo agli utenti di accedere a dati provenienti da diverse piattaforme, offrendo una visione unificata e coerente dei dati provenienti da database diversi. Inoltre, esso si distingue per la rapidità nell'ottenere risultati concreti poiché non richiede una pre-definizione delle

dimensioni; ogni dato è facilmente accessibile sia come dimensione che come misura, adattandosi alle specifiche esigenze.

Tuttavia, ci sono anche alcune limitazioni da considerare, come il costo rilevante, specialmente per le piccole aziende o le organizzazioni con budget limitato, e la complessità delle implementazioni personalizzate che richiedono una conoscenza tecnica approfondita.

Rispetto però ad altri software di BI, come Tableau e Power BI, Qlik Sense offre un approccio unico con l'associatività dei dati; ciò può essere un vantaggio significativo per le organizzazioni che cercano di scoprire relazioni complesse tra i dati. Inoltre, rispetto ai metodi tradizionali di analisi dati, Qlik Sense offre un approccio più veloce ed efficiente, consentendo agli utenti di lavorare direttamente con i dati senza dipendere da dipartimenti IT o analisti.

In conclusione, Qlik Sense è una potente piattaforma di business intelligence che offre numerosi vantaggi, tra cui un'interfaccia intuitiva accessibile anche agli utenti non tecnici, analisi in tempo reale per decisioni immediate e la scoperta naturale delle relazioni tra i dati.

Inoltre, la flessibilità di Qlik Sense consente la creazione di applicazioni personalizzate, il collegamento dei dati in modo relazionale attraverso uno schema di dati in-memory, la creazione di presentazioni basate sui dati tramite storytelling, analisi statistiche, previsionali e storiche da eseguire, e l'abilità di analizzare dati da molteplici fonti; ciò fornisce supporto ad una vasta gamma di aree aziendali, come nel caso dell'azienda cliente.

### **2.1.2 Interfaccia grafica ed elementi chiave**

L'interfaccia di Qlik Sense è progettata per offrire agli utenti un ambiente potente e flessibile per l'analisi dei dati e la creazione di applicazioni interattive. Al centro di questa interfaccia troviamo due componenti fondamentali: il *Qlik Sense Hub* e la *Qlik Sense Management Console*, le quali svolgono entrambe un ruolo cruciale nella gestione e nell'uso efficace della piattaforma.

La prima rappresenta il punto di ingresso principale per gli utenti finali, essendo infatti qui che questi interagiscono con le applicazioni create in Qlik Sense. Nel Qlik Sense Hub, gli utenti possono infatti accedere alle diverse applicazioni disponibili, esplorare dashboard interattive e visualizzazioni dei dati. Questo ambiente intuitivo consente agli utenti

di eseguire analisi, effettuare selezioni e ottenere informazioni rilevanti in modo rapido e intuitivo.

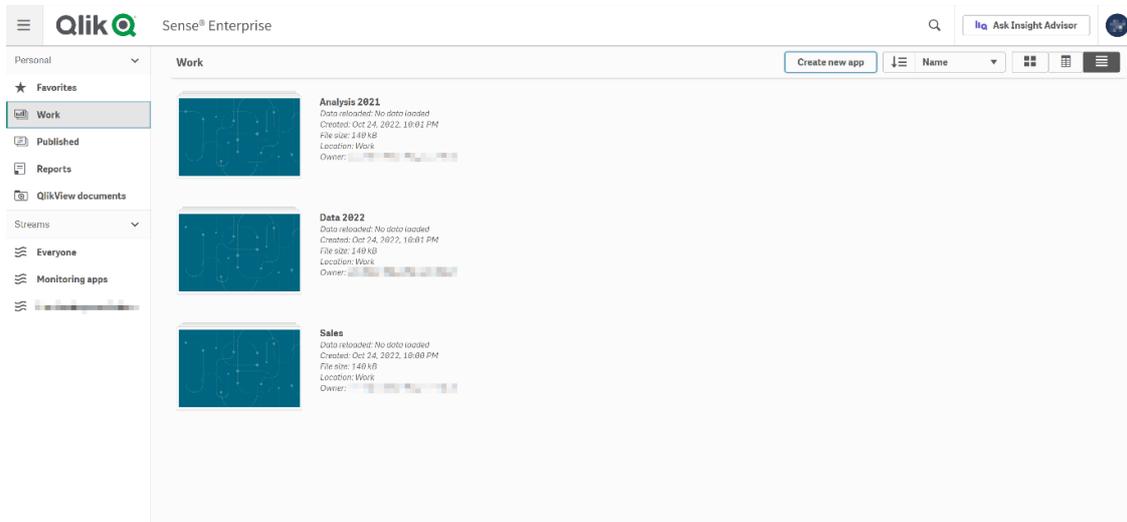


Figura 1 - Interfaccia del Qlik Sense Hub.

Fonte immagine: <https://www.qlik.com/us/products/qlik-sense>

È importante sottolineare che il Qlik Sense Hub e tutte le applicazioni presenti sono suddivise in diversi *stream*, presenti nella barra laterale sinistra del Qlik Sense Hub; ciò implica che ciascuna delle app è accessibile solo da un gruppo di utenti, in base alle specifiche definite dall'amministratore del sistema (ad esempio il ruolo e il mercato di riferimento a cui si è assegnati).

Inoltre, le applicazioni in ogni stream si suddividono in tre diverse categorie:

- *pubblico*, app revisionata ed approvata dall'amministratore del sistema o esperto (solitamente, un membro del reparto IT).
- *community*, app di un singolo utente che viene "promossa" dall'amministratore in quanto (potenzialmente) utile anche al resto degli utenti.
- *privato*, app creata da un utente per una visualizzazione o test interno.

Ovviamente, ogni utente avrà una certa visibilità e possibilità di modifica sulle applicazioni create (in base alle sue credenziali), mentre l'amministratore del sistema avrà una visibilità totale (anche sulle app private), la possibilità di modifica, e la possibilità di spostare le singole app da una categoria all'altra.

La seconda, la Qlik Sense Management Console (QMC), è invece il cuore dell'amministrazione e della gestione della piattaforma. Gli amministratori utilizzano infatti questa console per supervisionare e

controllare tutti gli aspetti di Qlik Sense, avendo la possibilità di gestire i singoli utenti, ruoli e autorizzazioni, ma potendo anche configurare le impostazioni di sicurezza, monitorare l'utilizzo delle risorse di sistema e gestire le distribuzioni di app e aggiornamenti. La QMC assicura che la piattaforma sia scalabile, sicura e conforme alle esigenze aziendali.

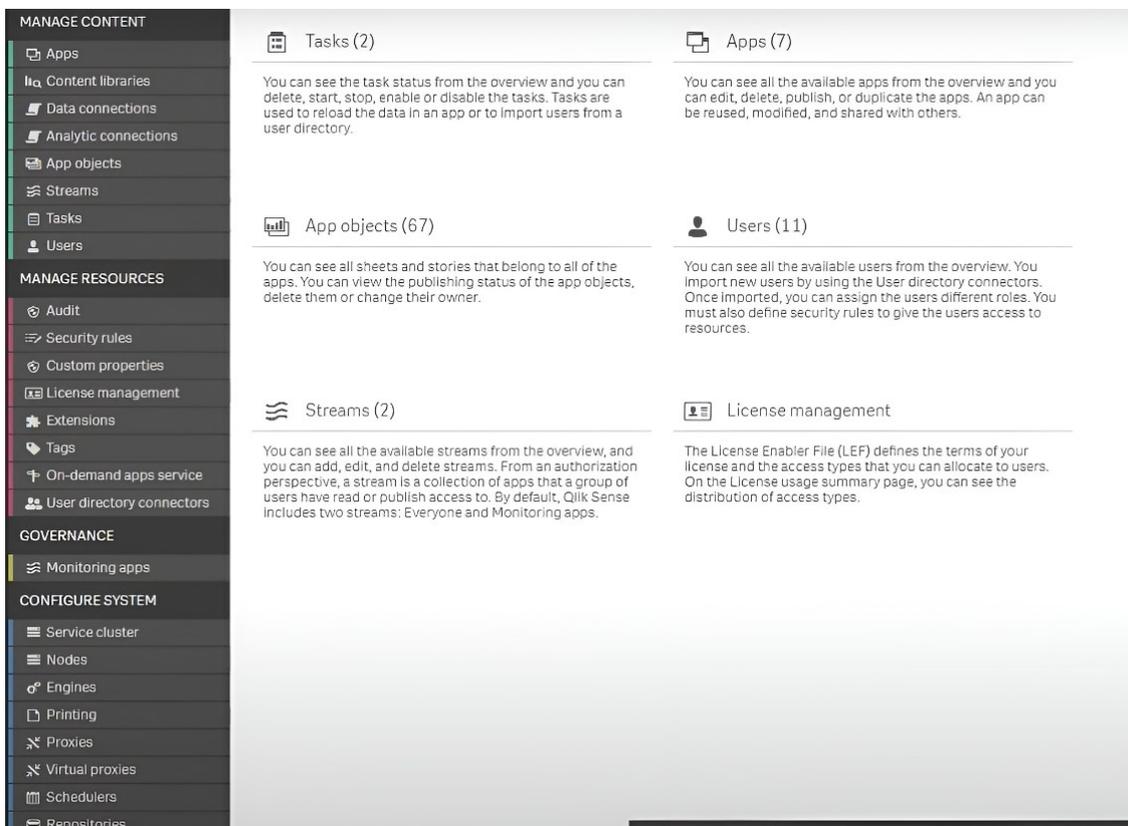


Figura 2 - Interfaccia della Qlik Sense Management Console.

Fonte immagine: <https://community.qlik.com/t5/Deployment-Management/How-to-Access-the-QMC/td-p/2048079>

Come detto sopra, ci sono poi le applicazioni, che sono gli strumenti principali di Qlik Sense per analizzare i dati; queste possono infatti essere considerate come singoli file sul quale vengono svolte le singole analisi. Gli utenti possono creare, modificare e condividere le proprie app per analizzare i dati aziendali in modo personalizzato. Ogni app può poi contenere uno o più fogli, ognuno dei quali include visualizzazioni interattive come grafici, KPI e tabelle. La creazione di applicazioni è un processo flessibile che consente agli utenti di combinare dati da diverse fonti, creare dashboard personalizzate e condurre analisi dettagliate.

Le app possono essere condivise con altri utenti all'interno dell'organizzazione, garantendo che le informazioni cruciali siano

accessibili a tutti i membri del gruppo, promuovendo la collaborazione e la diffusione delle conoscenze aziendali.

Ciò che distingue questa piattaforma è l'interconnessione tra tutti i fogli, ovvero le operazioni effettuate su uno di essi influenzano automaticamente tutti gli altri oggetti presenti nei fogli collegati.

## **2.2 La gestione dati in backend**

In accoppiata alla piattaforma Qlik Sense, utilizzata, come detto, per ciò che riguarda la parte di sistema front-end, l'azienda cliente si è affidata all'architettura Greenplum per la gestione dei dati all'interno del datawarehouse.

Si procede dunque ora alla sua presentazione, illustrandone le principali caratteristiche e punti di forza, ed effettuando un rapido confronto con il principale competitor, Oracle.

### **2.2.1 Greenplum**

Greenplum è un sistema di gestione di database relazionali altamente scalabile e open-source, sviluppato da *Pivotal Software*.

Le sue principali caratteristiche e punti di forza sono:

- architettura parallela, infatti Greenplum è progettato per eseguire query in parallelo, dunque su grandi quantità di dati distribuiti su nodi multipli. Questa architettura consente di scalare in maniera orizzontale per gestire crescenti volumi di dati e carichi di lavoro complessi. Inoltre, anche la fase di caricamento dei dati (*data ingesting*) viene svolta in parallelo, sempre per ragioni di performance totali del sistema, supportando diverse tipologie di fonti di dati, come database relazionali, file di testo, file di log e dati da fonti di terze parti.

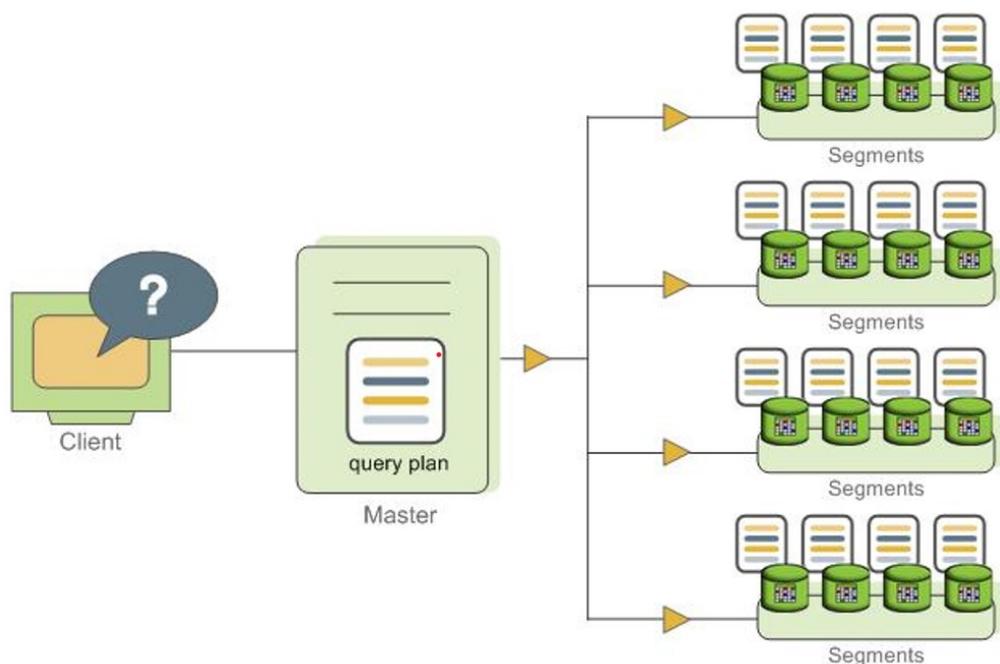


Figura 3 - Rappresentazione dell'esecuzione parallela di Greenplum.

Fonte immagine: [https://docs.vmware.com/en/VMware-Greenplum/7/greenplum-database/admin\\_guide-query-topics-parallel-proc.html](https://docs.vmware.com/en/VMware-Greenplum/7/greenplum-database/admin_guide-query-topics-parallel-proc.html)

- supporto per il data warehousing, tanto da essere ampiamente utilizzato in tutto il mondo come sistema di data warehousing per l'analisi di dati storici e la creazione di report aziendali in molte società di grandi dimensioni.
- open source, infatti il software è basato su PostgreSQL, un popolare sistema di gestione di database open source che eredita molte delle funzionalità e della compatibilità con SQL. Grazie a ciò, Greenplum ha una community di sviluppatori e utenti molto attiva che contribuiscono al suo sviluppo e forniscono supporto. Inoltre, tramite l'azienda proprietaria del software, Pivotal, sono offerti anche pacchetti specifici per il supporto a clienti aziendali di grandi dimensioni, i quali richiedono tipicamente funzionalità, integrazioni e assistenza maggiore rispetto a privati o piccole aziende.
- compressione dei dati, in quanto vengono utilizzati algoritmi di compressione dei dati per ridurre lo spazio di archiviazione richiesto, migliorando le prestazioni delle query.
- integrazione con strumenti BI e analisi avanzata, infatti è compatibile con una grande varietà di strumenti di business intelligence come Tableau e QlikSense, come nel caso del progetto in questione. Inoltre, Greenplum offre supporto a strumenti di analisi avanzata, inclusi algoritmi di machine learning, come Apache MADlib.

- sicurezza e gestione accessi, in quanto Greenplum include funzionalità avanzate di sicurezza dei dati, tra cui autenticazione, autorizzazione/gestione degli accessi in base alle credenziali e crittografia dei dati.

Nel caso aziendale in esame, le logiche e i dati attualmente presenti in Greenplum sono stati recentemente migrati dal precedente sistema al quale si affidava il cliente, ovvero Oracle, che è il leader di mercato per quanto riguarda la gestione dei database relazionali. L'abbandono del sistema sviluppato da Oracle in favore di quello di Greenplum è stato deciso principalmente per due fattori, ovvero per la sua caratteristica di essere open-source, e per ragioni economiche. Infatti, proprio per il fatto che Oracle è il sistema di datawarehouse più utilizzato al mondo, i suoi costi di assistenza e licenza sono significativamente più elevati se comparati con quelli di Greenplum; inoltre, il software di Oracle è basato su un sistema proprietario, meno flessibile rispetto a quello open source di Greenplum, rendendo la sua manutenibilità ed integrazione con altri sistemi più complessa.

Oltre alle differenze principali già elencate, Greenplum è anche noto per la sua scalabilità orizzontale, che può essere più conveniente per gestire grandi volumi di dati rispetto alla scalabilità verticale tipica di Oracle.

L'architettura del datawarehouse basata sul sistema Greenplum è poi interrogata attraverso un software di backend, DBeaver, specializzato nella gestione delle query dati. In questo contesto, DBeaver svolge un ruolo cruciale, poiché è utilizzato per eseguire interrogazioni SQL e ottenere i dati desiderati.

### **2.2.2 DBeaver**

DBeaver è un software client, open source e multiplatforma per l'interrogazione dei database, siano essi relazionali o non relazionali. Il suo principale punto di forza è il fatto di essere multi-database, ovvero potersi interfacciare e interrogare una vasta gamma di DB, consentendo agli utenti di potersi connettere simultaneamente a diverse istanze di database eterogenee; alcuni esempi di database supportati sono MySQL, PostgreSQL, Oracle, SQL Server, DB2, MongoDB. Proprio in virtù di questa sua caratteristica, DBeaver consente anche agli utenti di importare ed esportare dati da e verso il database in vari formati, tra cui CSV, Excel, JSON e molti altri, semplificando notevolmente il trasferimento di dati tra diverse fonti.

In tal senso, il software offre un'interfaccia ricca di elementi grafici per agevolare l'interazione con i database, la gestione dei dati e delle interrogazioni SQL.

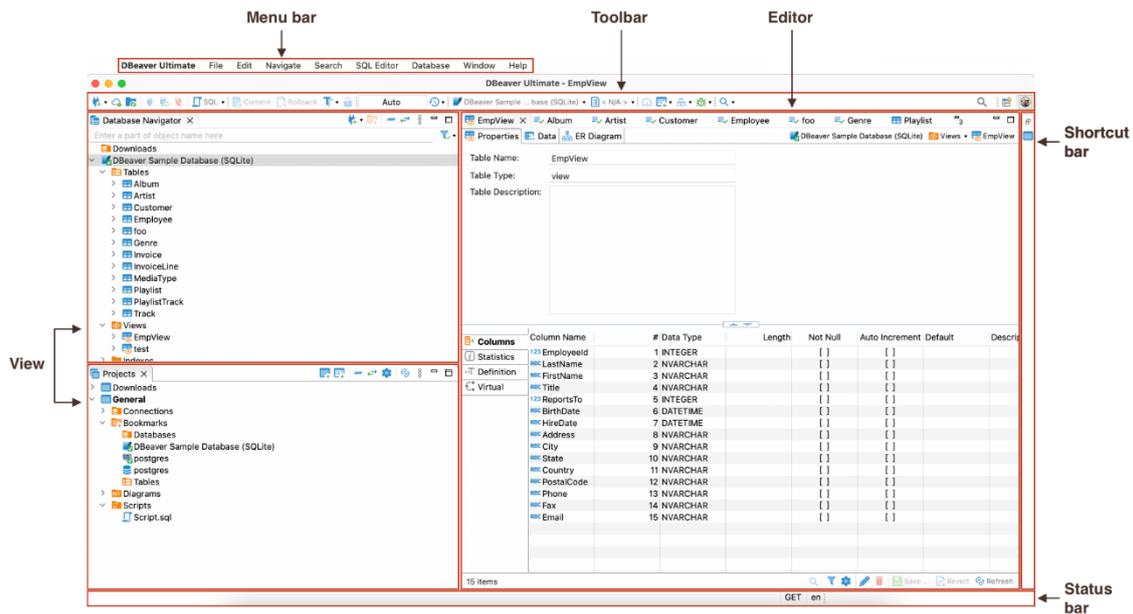


Figura 4 - Interfaccia della schermata principale di DBBeaver.

Fonte immagine: <https://dbeaver.com/docs/dbeaver/Application-Window-Overview/>

Nella parte superiore dell'applicazione si trova la barra dei menu, che permette di accedere a tutte le funzionalità e le opzioni disponibili, tra cui il collegamento ai database, l'esecuzione di query SQL e la gestione degli oggetti del database. Accanto a questa, c'è la barra degli strumenti che contiene icone e scorciatoie per le azioni comuni come il salvataggio delle query o il collegamento ai database.

L'area laterale a sinistra dell'interfaccia è occupata dall'albero di navigazione, che mostra la struttura ad albero dei database, dei server e degli oggetti connessi. Questo strumento permette quindi di navigare tra i database ed i loro oggetti in modo intuitivo.

L'elemento centrale di DBBeaver è l'editor di query SQL, che fornisce uno spazio per scrivere, modificare ed eseguire le query. Questo editor offre funzionalità avanzate, tra cui la formattazione del codice e il completamento automatico e, una volta eseguita una query, il risultato viene visualizzato nell'area sottostante.

Nell'area laterale a sinistra dell'interfaccia, si trova il pannello delle proprietà, che mostra dettagli sulle proprietà dell'oggetto selezionato nel database, come tabelle, viste o procedure. Il pannello delle connessioni, invece, elenca tutte le connessioni attive con i database, consentendo una facile gestione di questi ultimi.

È dunque evidente come DBeaver sia un software altamente personalizzabile, permettendo di adattare l'interfaccia alle esigenze specifiche del business, aggiungendo o rimuovendo pannelli, modificando i layout o configurando scorciatoie personalizzate; è, ad esempio, possibile mostrare ulteriori informazioni, come lo schema del database, la cronologia delle query o i diagrammi ER. Tutto ciò rende quindi DBeaver uno strumento flessibile e potente per la gestione dei database collegati e delle query SQL, tenendo anche in considerazione la sua struttura open-source, che gli permette di essere usato su qualunque sistema operativo (Windows, MacOS e Linux) e con la possibilità di sviluppare o installare plugin per aggiungere funzionalità personalizzate.

### **3. Il progetto aziendale**

In questo capitolo, si vuole descrivere il progetto aziendale nella sua interezza. Si partirà dall'analizzare il contesto aziendale specifico, approfondendo sia gli aspetti più generali del mercato di riferimento sia quelli più tecnici relativi alle attività svolte dall'azienda cliente. Successivamente, si descriveranno le fasi di sviluppo del progetto, partendo dall'analisi della situazione AS-IS fino ad arrivare all'effettiva modifica per arrivare alla soluzione finale implementata.

Per ragioni di riservatezza, si farà riferimento all'azienda di consulenza presso cui si è sviluppato il progetto con l'acronimo di KM mentre, per le stesse ragioni, CAAB sarà l'acronimo utilizzato per l'azienda cliente.

L'attività progettuale presso quest'ultima si è concentrata sull'integrazione di un nuovo flusso di dati relativo a nuovi mercati europei in cui CAAB si è recentemente espansa. Questo ha reso necessaria l'implementazione coerente dei nuovi dati nel sistema esistente, garantendo una transizione fluida e garantendo l'affidabilità e l'efficienza complessiva del modulo CRM.

Con quest'ultima acronimo si fa riferimento ad un modulo del DWH, ovvero il Customer Relationship Management (CRM), che verrà dettagliato in seguito. Questo è un componente del DWH dedicato all'archiviazione e all'analisi dei dati relativi ai clienti, ed è un sistema multi-sorgente in quanto i dati provengono da diverse fonti, come il CRM operativo, il sito web aziendale, i sistemi di marketing e di vendita.

Questo modulo consente quindi di raccogliere, organizzare e analizzare informazioni dettagliate sui clienti, come contatti, transazioni, interazioni e preferenze, andando a migliorare la gestione delle relazioni con i clienti, prendere decisioni basate sui dati e offrire un servizio personalizzato e mirato.

Appare quindi evidente come sia fondamentale una gestione efficace e strutturata dei dati all'interno del CRM, per tutte le fonti dati presenti, ed è proprio da questa premessa che ha origine il progetto in esame.

## **3.1 Panoramica del mercato e del cliente**

Si procede ora ad illustrare l'azienda cliente di KM con la quale è stato svolto il progetto oggetto dell'elaborato. In questo capitolo si analizzerà dunque la storia di CAAB e si farà una panoramica del mercato di appartenenza, approfondendo sia gli aspetti più generali del mercato di riferimento sia quelli più tecnici relativi alle attività svolte dall'azienda cliente.

### **3.1.1 Il contesto**

CAAB è una delle principali istituzioni finanziarie a livello globale; essa è una realtà solida, con una presenza significativa in Europa e in altre parti del mondo. Fondata agli inizi del secolo scorso, ha una lunga storia nel settore bancario e finanziario, e le sue origini risalgono alla creazione delle prime casse rurali e artigiane, cooperative di credito rurali, nel tardo XIX secolo.

Nel corso dei decenni, essa si è espansa attraverso acquisizioni e partnership strategiche; in tal modo, CAAB è riuscita a rafforzare ulteriormente la sua posizione nel settore bancario fino a diventare, ad oggi, una delle banche più grandi d'Europa.

CAAB ha una presenza significativa in diversi mercati chiave, tra cui:

- Spazio Economico Europeo (SEE), infatti questa ha una forte presenza in Italia, Spagna, Francia e altri paesi europei.
- mercati emergenti, in quanto è attiva in vari mercati emergenti, come l'Europa dell'Est, dove ha una presenza notevole.
- servizi globali, offrendo servizi bancari e di investimento attraverso la sua divisione di banca d'investimento, con una presenza globale in importanti centri finanziari.

Nello specifico, CAAB è coinvolta in diverse linee di business, tra cui:

- retail banking, offrendo servizi bancari al dettaglio, inclusi conti correnti, prestiti, carte di credito e servizi di risparmio attraverso la sua rete di filiali e canali online.
- banca d'investimento, infatti CAAB è attiva nella finanza d'investimento, nelle operazioni di mercato dei capitali, nel trading e nei servizi di consulenza per le imprese.
- gestione patrimoniale e assicurazioni, in quanto fornisce servizi di gestione patrimoniale e assicurativi per individui e aziende.

In particolare, l'inserimento aziendale in KM è avvenuto all'interno dell'area di analisi dati e gestione del datawarehouse di proprietà di CAAB, nella divisione aziendale specializzata nei servizi finanziari legati all'acquisto di veicoli, in particolare di automobili.

### **3.1.2 L'azienda cliente - CAAB**

Come detto, CAAB gestisce un reparto dedicato all'interno della sua istituzione bancaria, focalizzato sui servizi finanziari inerenti all'acquisto di veicoli. Questa si concentra sulla collaborazione con i produttori automobilistici e le reti di concessionarie per offrire soluzioni di finanziamento e servizi integrati ai clienti, che sono sia privati che aziende, di diversa dimensione. L'obiettivo della banca è offrire soluzioni finanziarie personalizzate e innovative per soddisfare le esigenze di ogni cliente.

Il principale servizio offerto dall'azienda ai propri clienti è il finanziamento per l'acquisto di automobili nuove o usate, fornendo opzioni flessibili con pagamenti rateali adattati alle esigenze finanziarie del cliente stesso. CAAB si occupa anche di offrire polizze assicurative per i veicoli, grazie alla collaborazione con diverse compagnie di assicurazione, coprendo diverse casistiche, come il danno al veicolo, la responsabilità civile, il furto e molto altro. Sempre per quanto riguarda il mercato delle automobili, vengono anche offerti servizi di assistenza, supporto tecnico e riparazioni in caso di necessità post-acquisto. Essa ha anche stabilito negli anni delle solide partnership con reti di concessionarie e produttori automobilistici per ampliare l'offerta di pacchetti finanziari integrati, permettendo ai clienti di ottenere il finanziamento richiesto direttamente presso la concessionaria al momento dell'acquisto del veicolo.

Inoltre, l'azienda offre anche servizi di leasing a breve termine e car sharing, tramite una nota azienda sua controllata, specializzata appunto in tale settore. In questo elaborato, si fa riferimento a tale azienda con l'acronimo DA, per ragioni di riservatezza.

Queste due business unit sono state cedute a quest'ultima da LY, acronimo con il quale si fa riferimento ad una nota azienda specializzata nel leasing e lungo-medio termine, i cui dati sono presenti all'interno dello stesso DW di CAAB. Questa operazione è avvenuta a seguito della dissoluzione di una joint venture precedentemente esistente tra CAAB e LY. Infatti, nonostante la separazione, le due entità hanno concordato di mantenere i dati nel medesimo DW per ridurre i costi associati alla divisione. Tuttavia, per garantire la sicurezza delle informazioni, sono

state implementate misure atte a garantire che ciascuna entità possa accedere solo ai propri dati e non a quelli dell'altra, come ulteriormente specificato in seguito.

In conclusione, CAAB ha dimostrato di voler diversificare e rafforzare la sua presenza nel settore automobilistico, sia attraverso l'espansione delle attività di finanziamento che attraverso lo sviluppo di servizi di mobilità avanzati. Ciò riflette anche la crescente importanza della mobilità sostenibile e dei servizi di condivisione nel panorama automobilistico globale.

È evidente come l'operato di CAAB sia perfettamente in linea con le recenti tendenze del mercato automobilistico. Essa è infatti impegnata nel favorire la transizione energetica del settore automotive, con l'obiettivo di offrire ai clienti soluzioni finanziarie sostenibili per l'acquisto di veicoli elettrici e ibridi, investendo allo stesso tempo nell'innovazione digitale per migliorare l'esperienza dei clienti, offrendo ai clienti servizi bancari e assicurativi semplici e convenienti.

### 3.1.3 I partner aziendali coinvolti

Nell'ambito della complessa gestione dell'infrastruttura dati, l'azienda cliente interagisce con diversi soggetti, tra i quali rientra evidentemente anche KM. All'interno di questo sistema, che comprende non solo CAAB come cliente, ma anche altre società specializzate in consulenza e supporto informatico e ingegneristico, ciascun attore svolge una specifica funzione che sarà illustrata di seguito. Per ragioni di riservatezza, queste altre aziende partner saranno chiamate con gli acronimi RP e CG.

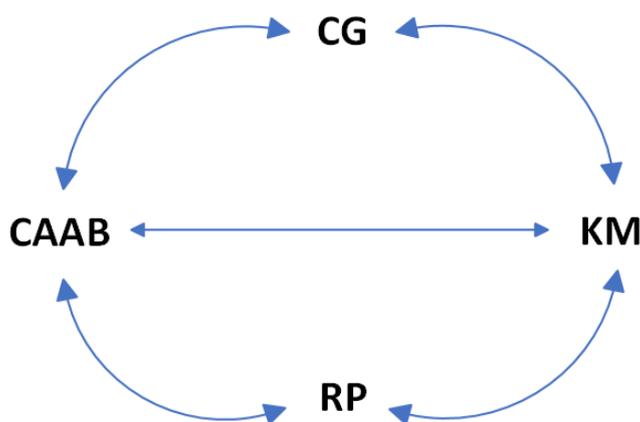


Figura 5 - Rappresentazione dei rapporti fra le aziende coinvolte nel progetto

RP, nota società di consulenza e servizi IT di respiro internazionale, è specializzata in soluzioni digitali, consulenza strategica e servizi di sviluppo software. Nel caso in esame, essa si occupa della gestione e manutenzione del database transazionale alla base del servizio che permette ai singoli concessionari, ad esempio, di registrare un cliente in ambito di vendita retail. Il DB in questo caso si occupa quindi di registrare non solamente l'anagrafica del cliente, ma anche tutte le proposte, i preventivi ed i contratti stipulati. Inoltre, sempre RP, si occupa della gestione del software frontend collegato alle informazioni contenute nel DB sopra descritto.

C'è poi CG, multinazionale di consulenza e servizi tecnologici che è una delle principali aziende nel settore dell'IT, offrendo una vasta gamma di servizi, tra cui consulenza aziendale, sviluppo di software, servizi cloud e gestione dell'infrastruttura IT. Nel progetto descritto, CG gestisce, tramite il noto software ERP SAP, i contratti validi in capo ad ogni cliente.

Infine KM, l'azienda presso cui si è svolta l'attività progettuale, è considerata una fra le più prestigiose in ambito di consulenza ed è presente in 143 paesi nel mondo, fornendo supporto e servizi multidisciplinari alle aziende, grazie alle diverse *Lines of Service* presenti, fra cui le principali sono:

- *Audit & Assurance*, che offre gli strumenti necessari per soddisfare le richieste informative del mercato e degli investitori.
- *Advisory*, per il supporto al cambiamento e al miglioramento delle performance tramite creazione di valore sostenibile sul lungo periodo.
- *Tax & Legal*, per il supporto amministrativo e burocratico.
- *Accounting*, per il supporto operativo ai processi contabili e fiscali.

Come spiegato anche in precedenza, KM gestisce il datawarehouse a valle delle operazionali transazionali sopra descritte, gestendo la parte di analisi dati e reportistica al cliente (CAAB) anche tramite l'utilizzo di Qlik Sense, noto software di business intelligence descritto nel capitolo dedicato. KM si interfaccia quindi, da un lato con RP e CG (che si occupano della valorizzazione dei campi all'interno del DW), e dall'altro direttamente con CAAB.

Oltre a quelle citate, ci sono poi diversi altri partner di supporto, fra cui il più noto ed importante è IBM, azienda leader a livello globale nella fornitura e implementazione di soluzioni hardware e software. Relativamente all'accesso ai database fisici, è infatti IBM a fornire i datacenter al cliente. In tal senso, essa si assume la responsabilità della gestione, manutenzione e scalabilità delle risorse offerte al cliente.

KM svolge quindi direttamente attività per conto di CAAB, andando ad interrogare il DW per assisterli, ad esempio, in eventuali problematiche di frontend o backend al momento della creazione di visualizzazioni dati. In tale situazione, sarà poi eventualmente compito della stessa KM dover interagire con altre aziende al fine di risolvere le segnalazioni riportate dal cliente. Infatti, la comunicazione con i clienti, ovvero gli utenti all'interno di CAAB, avviene quasi sempre in modalità non tecnica, ed è dunque compito di KM dover trasporre la richiesta del cliente in linguaggio tecnico, individuare la problematica all'interno del DW, valutare i vari passaggi da effettuare al fine di risolvere la problematica o, se necessario, comunicarla alle altre aziende partner, se il problema necessita di un loro diretto intervento.

KM non svolge però esclusivamente questo tipo di attività, ma anche quelle di *Enhancement*, ovvero l'ampliamento e la modifica della struttura dati del datawarehouse in modo da soddisfare le nuove specifiche richieste da CAAB. Quest'ultima è, tra l'altro, la fattispecie sulla quale si è svolto il progetto oggetto di questo elaborato.

## **3.2 Architettura del datawarehouse aziendale**

Si procede ora a descrivere la struttura dati del datawarehouse specifico del caso aziendale, con l'obiettivo di evidenziare le principali componenti e le fasi di popolamento e manipolazione dei dati, sia all'interno del datawarehouse che all'interno di Qlik Sense. Le specifiche implementate tramite il progetto descritto da questo elaborato (TO-BE) saranno invece descritte nel capitolo successivo, evidenziando le differenze con la struttura attuale (AS-IS). Si approfondisce inoltre il modulo specifico oggetto dell'elaborato, ovvero il modulo del Customer Relationship Management (CRM).

### **3.2.1 Introduzione**

All'interno del DW sono presenti diversi moduli, o sistemi gestionali, che si occupano di specifiche funzioni a supporto dell'operatività aziendale del cliente. Infatti, il datawarehouse è suddiviso logicamente secondo diverse aree applicative, ognuna gestita secondo uno *schema* diverso, ma tutte facenti parte dello stesso DW.

Dal punto di vista del cliente, queste aree funzionali sono totalmente diverse e scollegate fra di loro, quindi gli utenti che sono stati assegnati esclusivamente ad un'area possono lavorare solo su quest'ultima, senza possibilità di accesso alle altre (in base alle credenziali d'accesso fornite dall'amministratore del sistema).

Con *schema* si intende una struttura logica che viene utilizzata per organizzare e raggruppare oggetti del database, come tabelle, viste, procedure memorizzate e altri elementi. Essi consentono di separare e gestire in modo ordinato gli oggetti del database all'interno di un database specifico. Questi, infatti, fungono da contenitori logici per gli oggetti del database rendendo possibile raggruppare tabelle correlate, viste o altre entità all'interno dello stesso schema; questi oggetti, all'interno di uno stesso schema, non sono inoltre direttamente accessibili da altri schemi a meno che non vengano esplicitamente concessi permessi di accesso, tipicamente dall'utente superuser, ovvero il proprietario dell'intero DW. Ciò significa che è possibile controllare l'accesso e la visibilità degli oggetti tramite autorizzazioni su ogni diverso schema presente, ma anche le impostazioni di sicurezza.

Infatti, ognuno di essi ha tre tipologie di utenti: *account owner*, che è l'amministratore dello specifico schema, il quale ha diritti di lettura, scrittura e modifica di tutte le autorizzazioni degli account presenti; questa tipologia di account ha quindi accesso completo a tutti gli oggetti e funzionalità presenti, ma solo relativamente allo specifico schema, fra i vari presenti all'interno del datawarehouse. C'è poi l'*account utente* che, avendo autorizzazioni ristrette, solitamente ha possibilità di scrittura, quindi di caricamento di nuovi dati, ma non di gestione ad alto livello del DW. Infine, c'è l'*account di servizio*, che può solamente leggere il contenuto dei dati, solitamente per effettuare operazioni di analisi e di reporting.

C'è poi un'ulteriore tipologia di account, come accennato in precedenza, che fa però riferimento non più al singolo schema, ma all'intero datawarehouse, ovvero il *superuser*, il proprietario/responsabile di quest'ultimo. Quest'ultimo ha le autorizzazioni di più alto livello all'interno del DW e può regolarne ogni aspetto, compreso ciò che riguarda i singoli schemi al suo interno.

Si procede ora a presentare il modulo che è oggetto di questa trattazione, il Customer Relationship Management, sul quale è stato sviluppato l'intero progetto.

### 3.2.2 Il Customer Relationship Management

Il Customer Relationship Management (CRM) è uno dei moduli presenti all'interno del datawarehouse di CAAB, e ha la funzione di collezionare i dati anagrafici e contrattuali provenienti dal sistema transazionale dell'azienda, per entrambe le linee di business, ovvero quella *retail* (vendita) e quella *rental* (noleggio).

Nello specifico, i dati acquisiti riguardano le *chiamate*, che rappresentano tutte le diverse forme di interazione tra il servizio clienti e i clienti stessi. Queste interazioni includono chiamate telefoniche, e-mail, comunicazioni tramite PEC e interazioni online, e sono parte integrante dell'assistenza ai clienti fornita dal servizio clienti.

Tali interrogazioni generano poi dei *casi*, ognuno dei quali è aperto in seguito al contatto iniziale; ad esempio, per ogni interazione tra il servizio clienti e l'utente, l'operatore ha la possibilità di avviare più casi, ognuno dei quali affronta un aspetto specifico.

Infine, ci sono le *attività*, che sono collegate ad uno specifico caso con una relazione molti a molti, il che significa che un caso può essere collegato a diverse attività contemporaneamente.

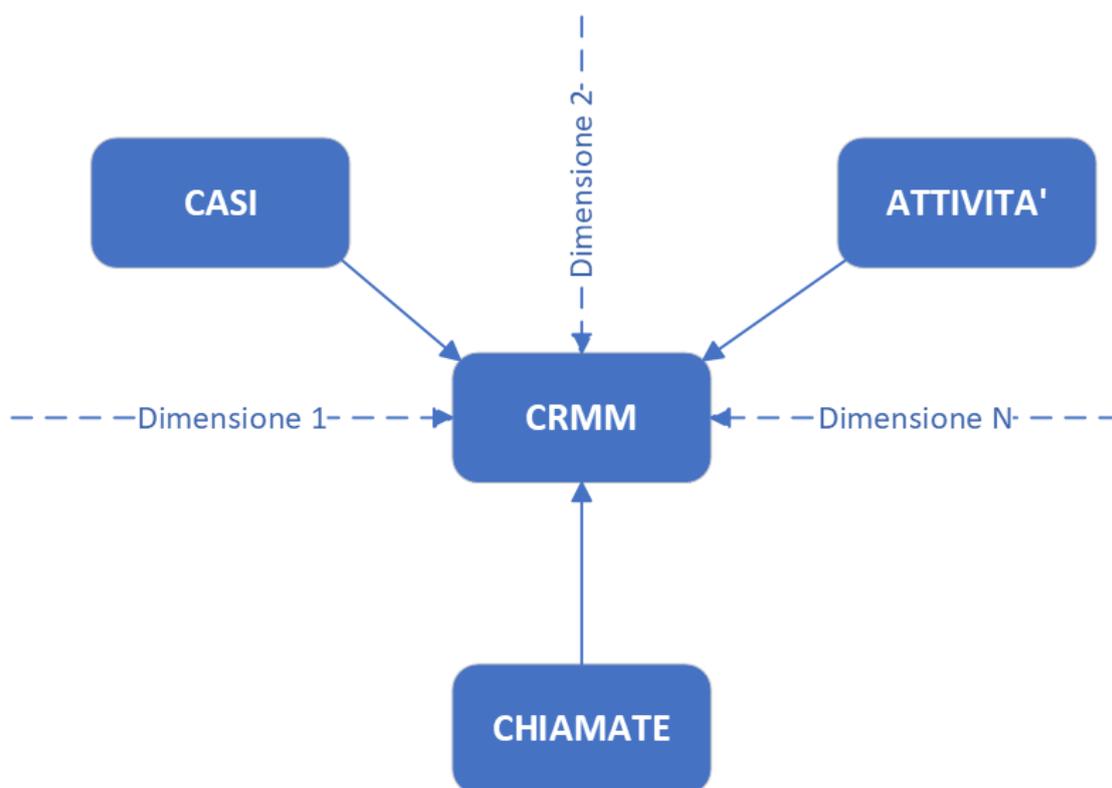


Figura 6 - Rappresentazione estremamente semplificata dello schema dati

Quindi, per quanto riguarda questo modulo del DW, il cliente che richiede il supporto di KM è principalmente il *customer care* di CAAB, che gestisce il ramo dei reclami relativi ai contratti stipulati dai clienti; questo non è però l'unico cliente interessato a questo tipo di dati. Infatti, diversi stakeholders aziendali utilizzano questi ultimi, soprattutto quelli relativi alle performance dei singoli operatori, per valutare diversi KPI. Ad esempio, uno fra i più importanti in tal senso è quello che fa riferimento ai tempi di risposta per clienti considerati 'premium', ovvero il tempo necessario alla chiusura della pratica per quest'ultima tipologia di clienti.

È importante sottolineare che quanto descritto avviene allo stesso modo sia per i dati appartenenti a CAAB sia per quelli LY, essendo tutti con la stessa struttura e sullo stesso DW, come approfondito nei capitoli successivi. Quanto detto rimane vero nonostante i cambi di proprietà delle società coinvolte che si sono succeduti negli ultimi anni, come spiegato in precedenza, per via di specifici accordi tra le aziende al momento della scissione, al fine di ridurre i costi e minimizzare l'impatto ed i tempi necessari alla stessa.

Le applicazioni create a frontend su Qlik Sense per il cliente sono:

- CRM Customer Care
- CRM Report Reclami
- CRM Customer Care – LTR Analysis Area 7

Le prime due app di reportistica fanno riferimento a CAAB mentre l'ultima a LY, e verranno approfondite nei capitoli successivi.

### 3.2.3 Processo di caricamento dati (ELT) in Greenplum

Prima di specificare nel dettaglio il processo di caricamento dei dati provenienti da database esterni all'interno del modulo CRM nel DW Greenplum del cliente (fase di *data ingestion*), si procede ad illustrare le diverse fonti dati dai quali questi vengono estratti.

Le fonti dati presenti includono database provenienti da diversi fornitori; in questo contesto, ogni attore si occupa della gestione di uno specifico aspetto del business. Nello specifico, i fornitori sono:

- Amazon Web Services (AWS), implementata in seguito allo sviluppo iniziale del DW, ma prima dello sviluppo di questo progetto.
- Genesys
- Comdata
- Microsoft Dynamics

In particolare, tutti i fornitori si occupano della gestione dati relativa alle chiamate, ma il sistema Dynamics si occupa anche della gestione dei casi e delle attività. Tuttavia, la maggior parte dei dati proviene dalla fonte dati AWS e MS DYN, e solo in parte minoritaria da Genesys e Comdata.

I dati provenienti dai flussi sopra descritti sono importati all'interno del DW tramite processi ELT. Con processo ELT, acronimo di "Extract, Load, Transform", si fa riferimento ad un componente fondamentale nell'ambito della gestione e dell'integrazione dei dati all'interno di un sistema informativo. Esso consente infatti di raccogliere, caricare e trasformare dati da varie fonti eterogenee in un formato compatibile e accessibile all'interno di un datawarehouse o di un sistema di archiviazione dati. Questo processo svolge quindi un ruolo cruciale nell'assicurare che i dati siano accurati, coerenti e pronti per l'analisi, e, nel corso della sua esecuzione, il processo ELT coinvolge diverse fasi, tra cui l'estrazione dei dati dalle sorgenti, il loro caricamento nell'archivio dati di destinazione e la loro trasformazione per adeguarli alle esigenze specifiche.

Nel caso specifico in esame, il processo si articola in quattro fasi sequenziali principali, le quali verranno esaminate dettagliatamente in seguito:

1. Avvio di una nuova schedulazione ELT da parte del pianificatore del carico di lavoro (workload scheduler), che in questo caso è *IBM Tivoli*.

2. Esecuzione dello script ELT presente all'interno del server ELT, che definisce il flusso di lavoro da eseguire richiamando le shell all'interno del DW Greenplum.
3. Svolgimento delle shell ELT, scritte in linguaggio Unix, e formate da catene, sottoprocessi e attività; definiscono le attività specifiche da eseguire per poter completare l'attività ELT richiesta.
4. Caricamento dei dati sul DW Greenplum, tramite il tool *GPLOAD*, che sfrutta i file YAML per i dettagli di configurazione.

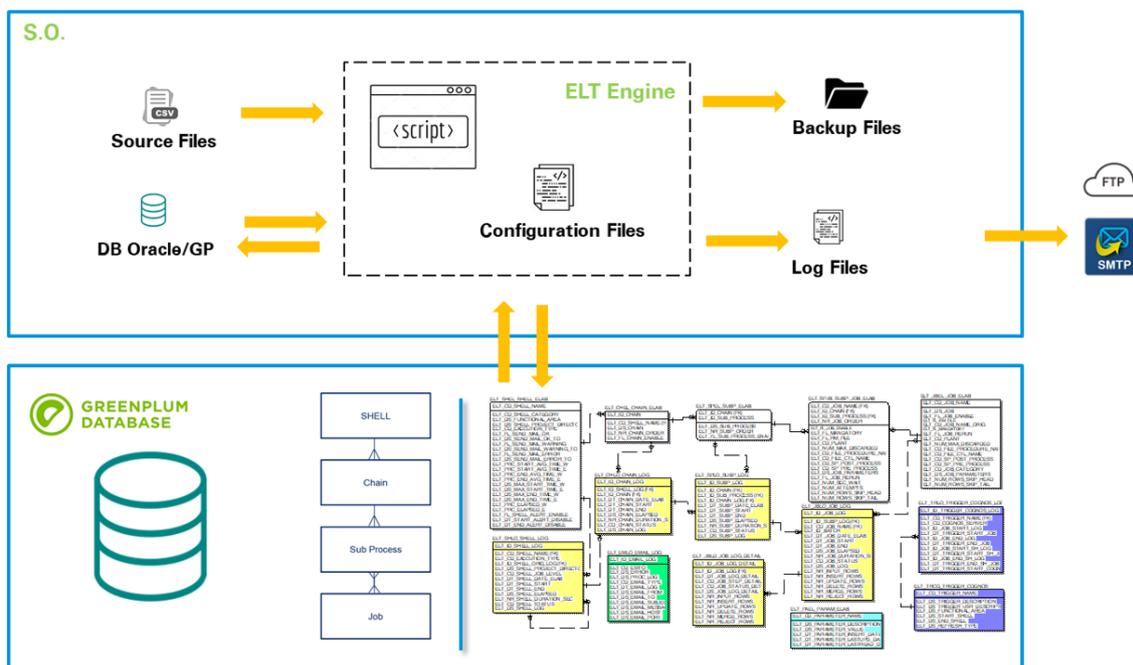


Figura 7 - Rappresentazione delle interazioni tra il server e il DW

Il processo si articola quindi a partire dalla schedulazione iniziale, che si ha a seguito dell'attivazione delle istanze di *IBM Tivoli*. *IBM Tivoli Workload Scheduler (TWS)* è infatti un sistema di pianificazione e automazione delle attività, noto anche come schedatore, ed è progettato per automatizzare le operazioni IT, tra cui l'esecuzione di processi, script, lavori batch e altre attività programmate in un ambiente IT complesso. Questa componente consente quindi agli utenti di definire, pianificare, monitorare e gestire il flusso di lavoro delle attività IT in base a una serie di regole e dipendenze e può essere utilizzato per automatizzare una vasta gamma di operazioni, tra cui backup, monitoraggio, elaborazione dei dati, esecuzione di script e molto altro. Inoltre, TWS è anche noto per essere un software altamente configurabile e flessibile.

Quindi, mentre *IBM Tivoli* offre una suite di prodotti per la gestione dell'infrastruttura IT, *IBM Tivoli Workload Scheduler* è una componente

specifica di Tivoli utilizzata dai fornitori del cliente per la pianificazione e l'automazione delle attività IT.

Il tool IBM, per poter far interagire il server ELT con il framework di Greenplum, avvia gli script ELT.

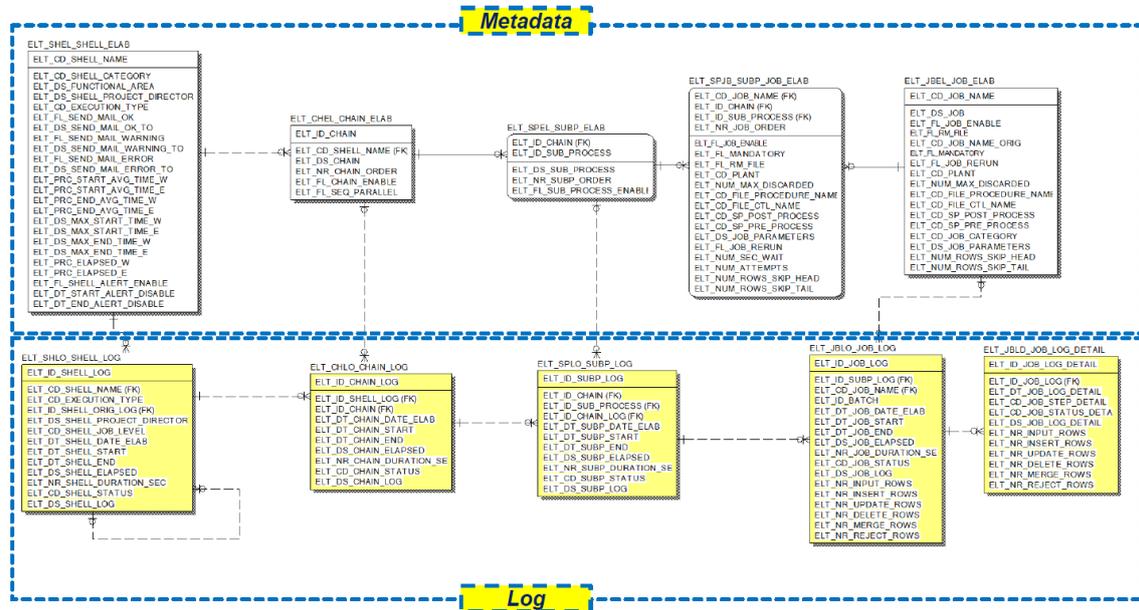
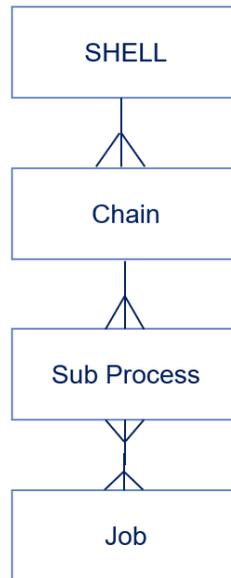


Figura 8 - Modello relazionale del framework ELT

Questi sono programmi specifici, solitamente scritti in linguaggi di programmazione come Python o Unix (o altri, a seconda delle esigenze), che si connettono al database Greenplum per eseguire le operazioni di estrazione, caricamento e trasformazione dei dati; ciò avviene utilizzando librerie o driver per connettersi a Greenplum ed eseguire le operazioni desiderate. Per poter compiere queste operazioni, è però necessario che lo script ELT abbia alcuni dati in input, relativi al dettaglio delle attività che si vuole svolgere, i loro parametri, le dipendenze tra le attività e la frequenza con cui devono essere eseguite.

Successivamente, gli script ELT comunicano con il DW Greenplum, attivando le shell relative alle operazioni da svolgere. Ogni shell è costituita da tre sotto-elementi gerarchici, che sono usati per gestire meglio il lavoro, dividendolo in unità più piccole.



*Figura 9 – Componenti di una shell*

Questi sotto-elementi sono:

1. Catena (chain), che è una sequenza di attività o passaggi ELT configurati in modo specifico; queste catene possono includere attività come l'estrazione dei dati, la pulizia dei dati, la trasformazione dei dati e il caricamento dei dati in Greenplum.
2. Sottoprocesso (subprocess), che è una sequenza di passaggi ELT associata ad una specifica catena. Ad esempio, si potrebbe avere un processo principale di una catena ELT che coinvolge l'importazione di dati grezzi in una tabella temporanea e poi sottoprocessi separati per la pulizia e la trasformazione dei dati prima del caricamento finale nel datawarehouse.
3. Attività (job), le quali sono utilizzate per definire e pianificare attività specifiche, relative ad uno o più sottoprocessi. Queste possono includere query SQL, script personalizzati o comandi per l'elaborazione dei dati e possono essere configurati per interagire direttamente con gli oggetti del database, come le tabelle, consentendo di eseguire operazioni di ELT all'interno del database stesso.

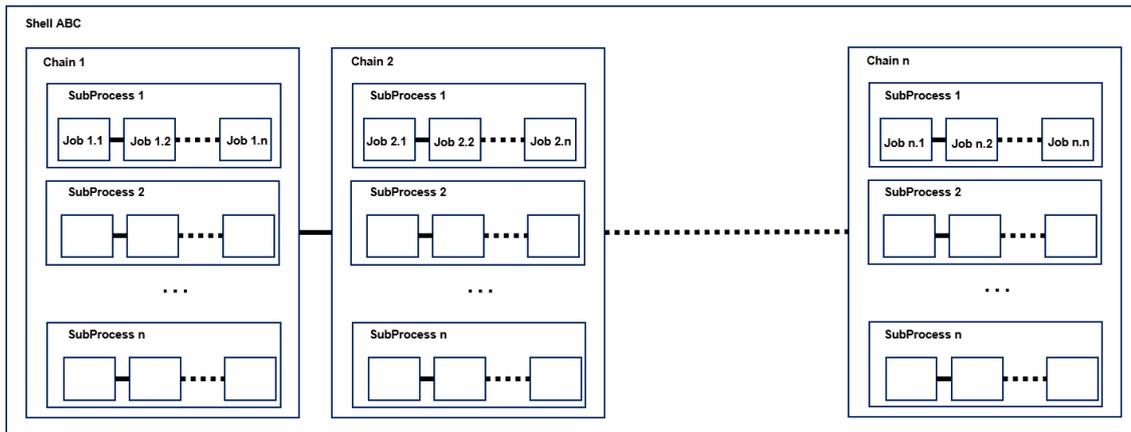


Figura 10 - Rappresentazione del processo multi-chain interno ad una shell

Per poter far partire le shell, l'azienda impiega il software open-source Putty, il quale è una utility versatile e ampiamente utilizzata per l'accesso remoto a sistemi e dispositivi attraverso connessioni sicure SSH (Secure Shell), che è un protocollo crittografato per la comunicazione sicura tra il computer locale e il server remoto. È un software molto popolare tra gli amministratori di sistema e gli utenti in quanto offre un'interfaccia a riga di comando in cui gli utenti possono inserire comandi per stabilire connessioni, autenticarsi e interagire con i sistemi remoti.

Infine, per poter effettivamente procedere al caricamento dei dati sul DW, Greenplum offre uno strumento chiamato *GPLOAD*, che è uno strumento di caricamento dati progettato per semplificare tale processo. Esso consente infatti di caricare dati da una varietà di fonti, tra cui file di testo, CSV, JSON e altri formati comuni e il suo funzionamento è basato su file di configurazione YAML; questi sono file di facile lettura e scrittura e consentono di specificare i dettagli del caricamento, come la posizione dei file sorgente, il mapping dei campi, le opzioni di trattamento degli errori, i dettagli per la connessione al DW Greenplum, il formato dei dati e delle tabelle, e altro ancora. GPLOAD identifica quindi le specifiche elencate all'interno di questo file di configurazione e procede al caricamento dei dati nel DW.

Inoltre, esso è anche in grado di sfruttare la potenza di elaborazione parallela di Greenplum per eseguire il caricamento dei dati in modo efficiente, cosa particolarmente utile quando si devono gestire grandi quantità di dati, e permette la trasformazione dei dati durante il caricamento, come la manipolazione dei valori dei campi o il riconoscimento di dati complessi. Inoltre, il software offre anche la possibilità di gestire gli errori in maniera personalizzata e monitorare l'avanzamento del caricamento dei dati, registrando informazioni dettagliate sul processo (file di log).

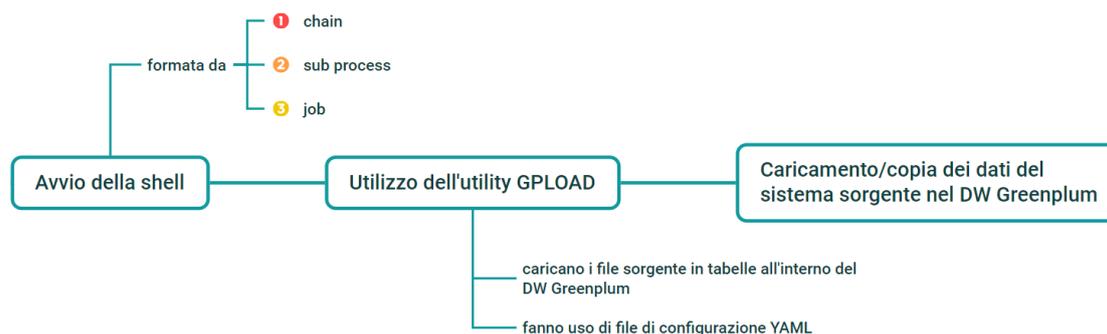


Figura 11 - Rappresentazione semplificata del processo ELT in un DW Greenplum

Nel caso specifico in esame, le shell relative al modulo CRM all'interno del DW (prima dello sviluppo del progetto) erano 3, avviate in maniera sequenziale al termine della precedente:

1. AWS, è dedicata al caricamento dei dati delle chiamate del servizio clienti, comprese e-mail, comunicazioni tramite PEC e interazioni online. Essa si occupa del trasferimento dei dati dalla sorgente esterna all'interno dell'area di staging del DW Greenplum. È però importante sottolineare che questo processo di caricamento avviene in modalità delta, ovvero vengono trasferiti solo i dati che non sono stati precedentemente caricati, garantendo così un utilizzo efficiente delle risorse e mantenendo l'integrità del sistema.
2. Inbound, si occupa del caricamento nell'area di staging del DW dei dati relativi alle chiamate Inbound provenienti dai sistemi MS Dynamics, Genesys e Comdata. Con chiamate *Inbound* ci si riferisce alle chiamate effettuate da nuovi e vecchi clienti al servizio clienti. Inoltre, questa shell si occupa anche della creazione della tabella a stella CRMM, precedentemente accennata e di cui si parlerà più in dettaglio in seguito, e della creazione del file di controllo che avvia il processo ELT di Qlik.
3. Outbound – Dynamics, in cui la shell carica nell'area di staging solo i dati delle chiamate MS Dynamics relative all'Outbound. Con chiamate *Outbound* ci si riferisce a quelle effettuate dal servizio clienti verso gli attuali clienti, al fine di rinnovare o cambiare il contratto esistente, o verso nuovi clienti per proporre nuovi contratti.

La shell AWS è la prima ad essere avviata, essendo questo il primo fornitore che procede all'invio dei dati al DW, ed è quindi indipendente dalle altre due, che invece sono dipendenti dalla rispettiva shell che precede. Nello specifico, la shell Inbound dipende dalla shell AWS, mentre la shell Outbound – Dynamics ha come dipendenza la shell Inbound.

Al termine della shell Inbound viene inoltre avviato il processo ETL relativo a Qlik Sense tramite un file di controllo apposito, come dettagliato nel capitolo successivo.

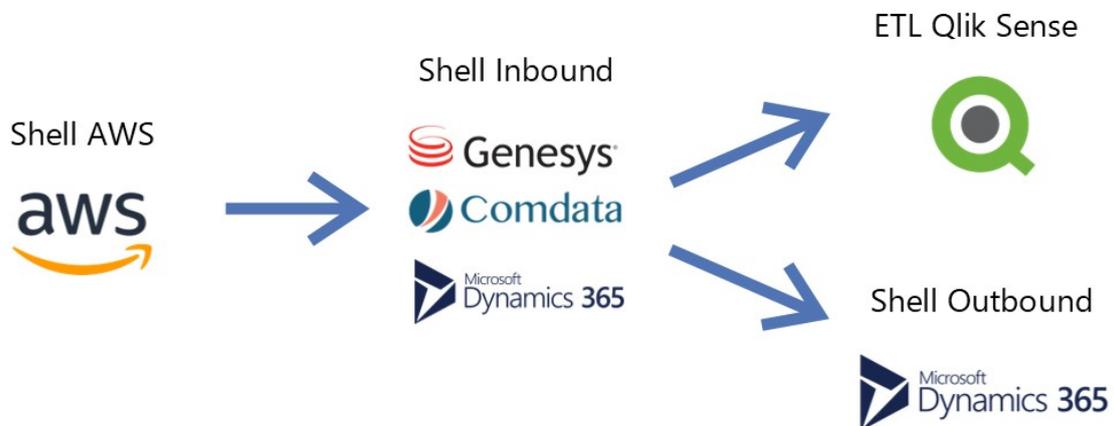


Figura 12 - Sequenza di avvio delle shell ELT prima del progetto

### 3.2.4 Struttura del modello dati

Una volta terminato il processo ELT di caricamento dei dati, dai singoli DB fino ad arrivare all'interno del DW Greenplum del cliente, i processi di manipolazione dati all'interno di quest'ultimo sono ascrivibili a tre fasi sequenziali principali:

1. Acquisizione (Data ingestion), costituita dalla fase di *Staging*.
2. Integrazione e riconciliazione (Data integration and reconciliation), dove è presente il livello *Operational Data Storage*
3. Processamento (Data Processing), di cui fanno parte il *Work layer* e il *Data mart layer*.

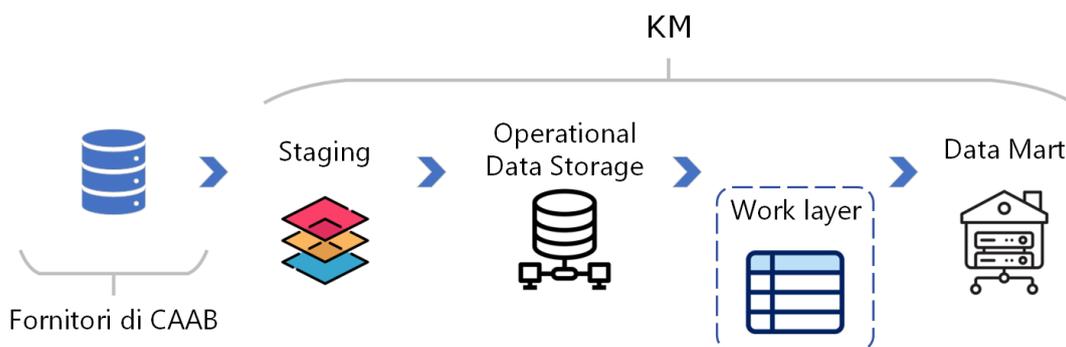


Figura 13 - Architettura del datawarehouse prima del progetto

Nel contesto del *Livello di preparazione (Staging Area - SA)*, viene creata una copia virtualmente identica (in rapporto 1:1) del sistema di origine dei dati, con poche o nessuna modifica o operazioni di pulizia preliminari prima dell'effettivo processo di elaborazione dei dati. Questa fase iniziale ha lo scopo di creare un ambiente di lavoro in cui i dati sorgente possono essere trattati in modo coerente prima di essere utilizzati per analisi o elaborazioni successive. Quindi, la copia dei dati mantiene l'integrità originale dei dati sorgente, ma coinvolge eventualmente alcune attività di pulizia per rimuovere dati duplicati, eliminare valori non validi o standardizzare i formati dei dati. L'obiettivo principale in tal senso è garantire che i dati siano pronti per essere utilizzati in modo affidabile nei processi successivi senza comprometterne l'integrità o la coerenza. Si sottolinea che il caricamento dei dati dal sistema sorgente all'interno del DW, che avviene quasi esclusivamente con cadenza giornaliera, interessa solamente i dati effettivamente modificati rispetto al precedente invio; in questo contesto, sono quindi solamente una frazione dell'intera mole di dati esistente ad essere inviati e processati giornalmente, per evidenti questioni di riduzione della capacità computazionale e di banda internet necessaria, con una significativa ottimizzazione dei processi di analisi svolti. Essendo il livello di staging alla base degli step di processamento dati successivi, quanto appena detto interessa a cascata anche i livelli sequenziali.

C'è poi il livello denominato *Archivio operativo dei dati (Operational Data Storage - ODS)* che è ottimizzato per l'accesso e l'elaborazione dei dati in tempo reale o quasi. In questo livello sono contenuti tutti i dati storici, aggiornati regolarmente dalle varie fonti sopra descritte, inseriti all'interno delle tabelle dei fatti corrispondenti alle chiamate, ai casi e alle attività. All'interno di questo livello vengono effettuate tre operazioni particolarmente importanti, per tutte le fonti dati presenti nel DW:

- storicizzazione dei dati, in cui viene definita la validità del dato andando a definire per ogni record l'inizio e la fine della validità dello stesso all'interno della tabella dei fatti relativa, tramite campi appositi.
- operazioni di pulizia e trasformazione, per rendere tra loro coerenti i dati.
- implementazione delle logiche richieste dal business, che sono specifiche per trattare coerentemente il dato con l'obiettivo del management
- categorizzazione dei dati tramite un campo specifico chiamato *Legal Entity*, ovvero il riferimento all'azienda di appartenenza per il singolo record, dato che i dati inviati fanno riferimento sia a LY che CAAB. Tramite questo campo, gli utenti a frontend possono

vedere solamente i dati di loro interesse e a cui gli è permesso di accedere, senza compromettere la sicurezza dei dati delle due aziende.

Successivamente, si trova un livello intermedio noto come *Livello di lavoro (Work Layer - WL)*, che non rappresenta uno stadio autonomo, ma piuttosto costituisce una fase necessaria all'interno del processo complessivo di elaborazione dati. Infatti, in questa fase, vengono eseguite una serie di operazioni sequenziali sui dati che convergono nel generare le tabelle dei fatti finali, insieme alle relative tabelle dimensionali. Questi elementi costituiscono la base del sistema dati d'interrogazione che è alla radice del sistema frontend di Qlik Sense; in altre parole, questo livello è il cuore del processo di preparazione dei dati che alimenta l'interfaccia utente di Qlik Sense. Si ha quindi l'applicazione di diverse logiche e procedure, come operazioni di pulizia per garantire che i dati siano accurati e privi di errori, l'integrazione per unificare fonti di dati eterogenee, la verifica della qualità dei dati per assicurare l'integrità e l'affidabilità e, infine, l'applicazione di logiche di business specifiche per conformare i dati alle esigenze aziendali.

L'ultimo passo del processo consiste nella creazione del *Magazzino dei dati (Data Mart - DM)*. In questa fase, a valle delle operazioni precedenti, si ottengono le tabelle dei fatti e le tabelle delle dimensioni associate. Il Data Mart è un repository di dati specializzato, focalizzato su un'area di business specifica, dove i dati vengono estratti, trasformati e caricati da diverse fonti dati al fine di supportare le esigenze di analisi e reporting aziendali.

Durante questa fase, il sistema procede a creare uno schema logico relazionale tra le tabelle del CRM. Questo schema è costituito da una struttura a stella, con una tabella centrale di grandi dimensioni, e le tabelle dei fatti e delle dimensioni ad essa collegate. Questo modello è chiamato CRMM (Customer Relationship Management Model) ed al suo interno è presente anche un'ulteriore fase necessaria alla storicizzazione dei dati, in cui vengono salvate le tipologie di relazioni presenti tra le tabelle presenti all'interno del modello a stella. Il contenuto delle singole tabelle, come spiegato in precedenza, è invece storicizzato all'interno dell'ODS.

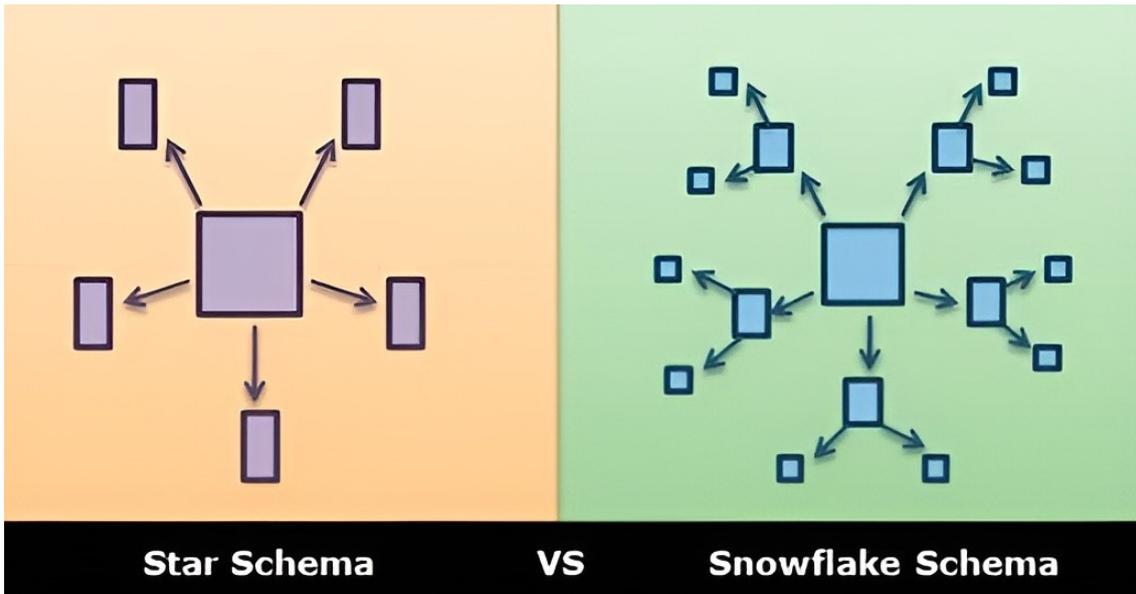


Figura 14 – Differenza grafica tra i due principali modelli dati possibili.

Fonte immagine: <https://nidhig631.medium.com/star-schema-vs-snowflake-schema-78dc9424a8a2>

In tal modo, si semplificano notevolmente le query e le analisi fatte sui dati, consentendo agli utenti di ottenere informazioni significative e dettagliate.

Nello specifico, la costruzione dello schema dati inizia con l'inserimento delle tabelle dei fatti, che rappresentano chiamate, casi e attività, e prosegue con l'aggiunta delle tabelle delle dimensioni, in modo da organizzare i dati in modo efficiente, evitando la duplicazione delle metriche e dei KPI calcolati in fase di reportistica tramite Qlik Sense.

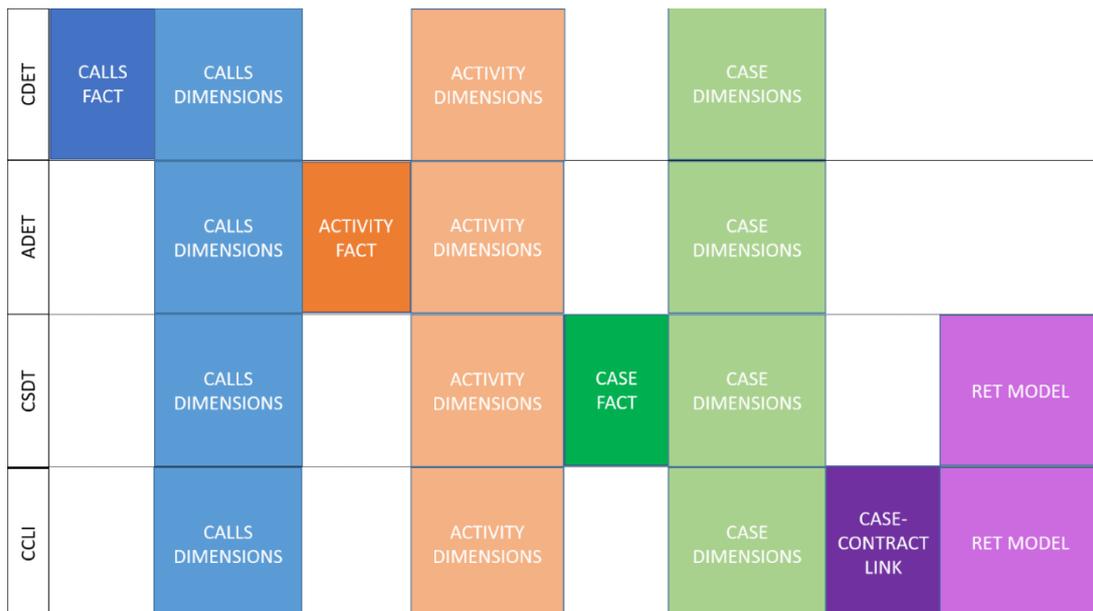


Figura 15 - Rappresentazione a blocchi della CRMM

### 3.2.5 Caricamento e integrazione dei dati su Qlik Sense

Iniziando dalle tabelle dei fatti e delle dimensioni, si dà il via al processo di caricamento dei dati all'interno della piattaforma di Qlik Sense.

Questo processo è appunto noto come ETL (Extract, Transform, Load) ed è simile al processo ELT illustrato in precedenza per il caricamento dei dati dalle fonti sorgenti al datawarehouse Greenplum. Tuttavia, Al contrario di quanto accadeva in precedenza, i dati vengono prima trasformati e poi caricati all'interno di Qlik Sense.

Inoltre, essendo i dati in input a Qlik Sense già processati tramite diverse elaborazioni e inseriti all'interno di schemi relazionali, come precedentemente specificato, il processo ETL all'interno di questa piattaforma risulta essere semplificato rispetto al precedente, ed è articolato in tre fasi principali:

1. estrazione (extract), anche denominato ETL 01, dove i dati vengono estratti dalle tabelle dei fatti e delle dimensioni, al livello del Data Mart nel datawarehouse Greenplum. Tale procedura crea così delle copie (in rapporto 1:1) identiche alla fonte dati, replicando le tabelle all'interno del database Qlik Sense, in un file con formato .QVD specifico della piattaforma stessa.
2. trasformazione (transform), chiamato ETL 02, in cui i dati estratti vengono quindi sottoposti a diverse operazioni di trasformazione; queste operazioni includono principalmente fasi di normalizzazione e aggregazione dei dati per definire il modello dei dati, mettendo in associazione le singole tabelle create. Inoltre, questo livello prevede la creazione di alias per le tabelle, quando necessario, in modo da trasformare i campi con nome tecnico in altri con una terminologia aziendale di più facile comprensione e utilizzo al momento della creazione delle applicazioni di business.
3. caricamento (load), denominato *Presentation Layer*, che include lo sviluppo di visualizzazioni e funzionalità di scoperta dei dati. A questo punto del processo, i dati sono strutturati in diverse tabelle Qlik Sense e i campi sono stati rinominati durante i passaggi precedenti; in tal modo, si consente l'aggiornamento dei dati all'interno delle diverse applicazioni create dagli utenti, permettendo così l'elaborazione, l'analisi e la creazione di report interattivi.

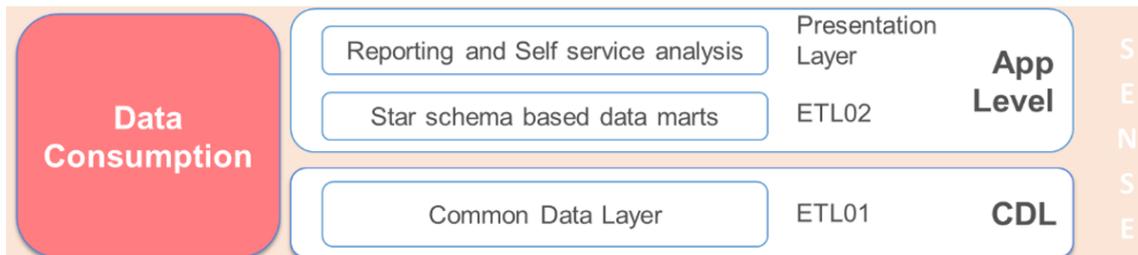


Figura 16 - Rappresentazione riassuntiva delle tre fasi ETL di Qlik Sense

Questo processo ETL è fondamentale per preparare i dati e renderli pronti per l'uso in Qlik Sense, garantendo che siano accurati, coerenti e facilmente interrogabili all'interno dell'applicazione. Infatti, una volta terminato questo processo, le applicazioni costruite sulla piattaforma possono interrogare i dati, creare dashboard (con metriche e KPI specifici) e svolgere tutte le analisi richieste dal business.

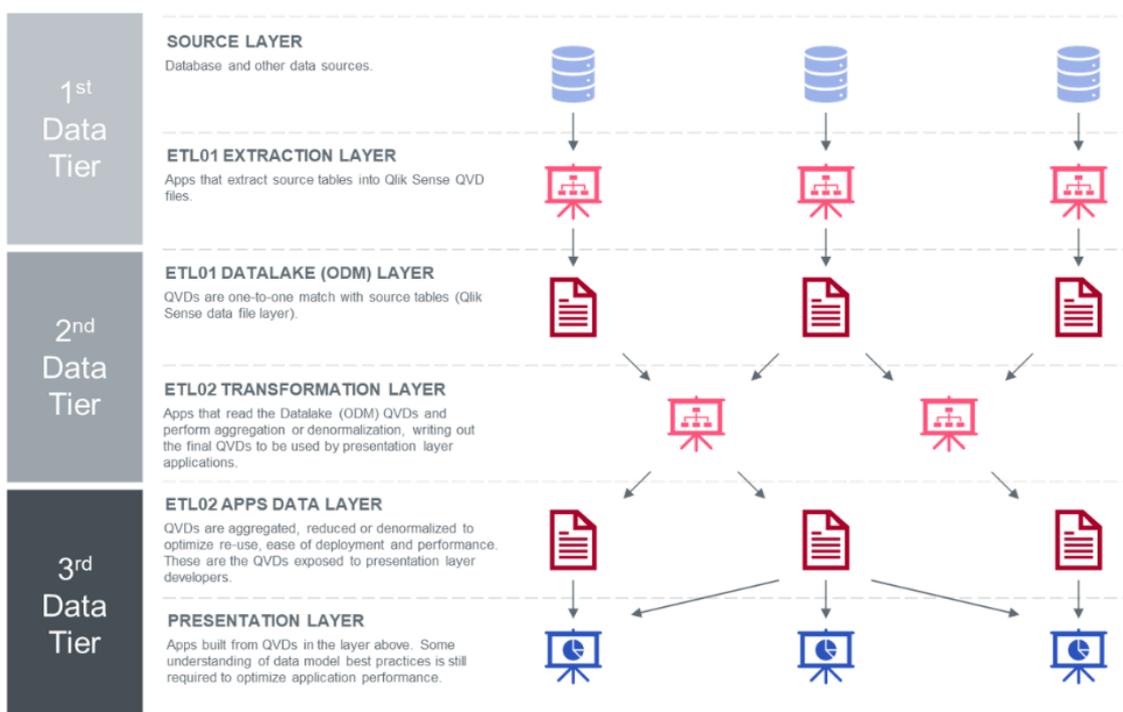


Figura 17 - Architettura dati di Qlik Sense

Tuttavia, mentre le altre shell presentate vengono richiamate da IBM TWS all'arrivo dei file sorgente ed una volta che le shell precedenti sono completate, quella relativa al processo ETL di Qlik Sense viene attivata dal programma stesso. Infatti, la sua schedulazione è determinata dall'utilizzo di un file apposito, ovvero un file di controllo che è utilizzato per gestire e coordinare l'avvio dei processi. Questo è un file con estensione .bat, ovvero un file batch che contiene una serie di comandi

eseguibili in sequenza. Questi tipi di file sono scritti in un linguaggio di scripting semplice e sono utilizzati per automatizzare processi ed attività di vario tipo, in questo caso l'intero processo ETL Qlik Sense.

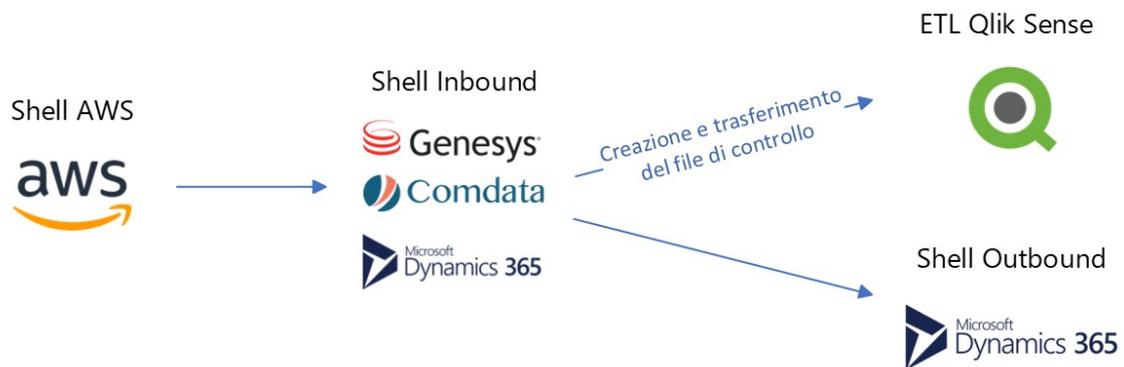


Figura 18 - Sequenza di avvio delle shell con dettaglio sul file di controllo

Il processo consiste nel creare questo file di controllo all'interno del server UNIX, al termine dello script presente nella shell Inbound. Il file viene poi trasferito dalla parte finale dello script, tramite IBM TWS, all'interno del server Qlik Sense. È poi presente un ulteriore script, questa volta interno a Qlik Sense, che si occupa di far partire l'intera schedulazione ETL sotto riportata. Tuttavia, il file di controllo è in realtà vuoto e non contiene al suo interno nessuno script, serve solo a fare da trigger per la shell ETL interna a Qlik Sense, segnalandone la partenza. Inoltre, questa shell si occupa anche di eliminare il file di controllo al termine del processo ETL, altrimenti il giorno successivo il processo si bloccherebbe.

# ETL Qlik Sense

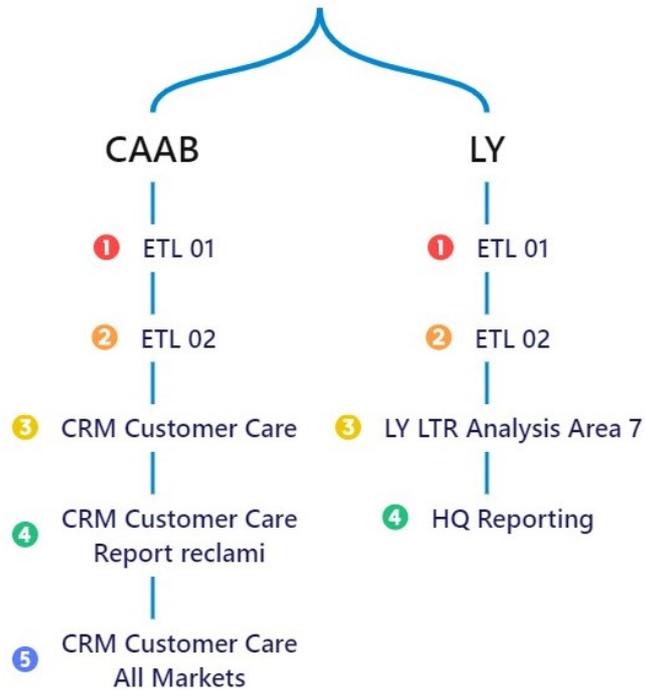


Figura 19 - Dettaglio ETL QLIK Sense per le due aree coinvolte

Nello specifico, come spiegato in precedenza, i dati presenti nel CRM sono suddivisi in quelli relativi al ramo bancario di CAAB e in quelli facenti riferimento a LY. Questa suddivisione impatta però solamente i dati e i report visibili a frontend in Qlik Sense, in quanto le strutture dati presenti nel DW e il DW stesso sono i medesimi. Tale discriminazione viene quindi fatta nel momento della creazione delle dashboard da parte degli utenti, in quanto questi ultimi avranno visibilità solamente dei dati di loro competenza; questa suddivisione è fatta a backend grazie al campo apposito, *Legal Entity*, che funge da discriminante sul dominio di appartenenza corretto del singolo dato.

## **3.3 Lo sviluppo del progetto**

Si illustra ora il contenuto del progetto in merito alle modifiche implementate, rispetto alla situazione AS-IS descritta nei capitoli precedenti. Si partirà dal riaffermare l'obiettivo alla base del progetto, fino ad arrivare alla descrizione delle motivazioni e delle varie fasi intermedie prima di arrivare all'implementazione effettiva della soluzione finale.

### **3.3.1 Lo scopo**

L'obiettivo del progetto è stato quello di integrare un nuovo flusso di dati relativo ai mercati europei all'interno del modulo CRM sopra descritto. Nello specifico, l'integrazione ha riguardato un flusso cruciale di dati riguardanti i nuovi contratti stipulati nel mercato europeo da CAAB, e provenienti da un altro provider sorgente, ovvero Salesforce (SFDC).

Salesforce è un'azienda leader nel settore del software per la gestione dei contatti dei clienti (CRM) ed offre una vasta gamma di soluzioni per aiutare le aziende a gestire e migliorare le relazioni con i clienti; infatti, essa offre un ecosistema completo e integrato per gestire l'intero ciclo di vita del cliente. Questo flusso dati viene poi integrato nel DW e in Qlik Sense, in modo tale da poter poi procedere all'implementazione nelle dashboard presenti e alla fase di visualizzazione e reporting da parte degli analisti di CAAB e dell'HQ.

Il processo di integrazione dei dati è stato però reso più complesso dal fatto che la struttura del DW è basata sulla struttura dati esistente per il flusso relativo alla sorgente Microsoft Dynamics, e dunque non sul nuovo flusso proveniente da Salesforce. Infatti, quest'ultimo disponeva di una formattazione e di una struttura diversa dai flussi già esistenti ed integrati all'interno del DW.

Lo scopo del progetto è stato quindi quello di uniformare la struttura dei flussi provenienti da tutte le diverse fonti dati, precedentemente riportate, con l'obiettivo di ottenere un processo di analisi e visualizzazione/reporting dei dati omnicomprensivo e flessibile rispetto alle esigenze specifiche richieste, cercando di mantenere inalterati il maggior numero di dimensioni e KPI precedentemente calcolati nelle dashboard. L'HQ di CAAB aveva infatti la necessità di incorporare i dati SFDC relativi al mercato europeo all'interno delle dashboard utilizzate,

in quanto in precedenza tali strumenti di analisi facevano affidamento esclusivamente su informazioni provenienti dal mercato italiano.

Prima dell'intervento oggetto di questa trattazione, i dati provenienti da SFDC venivano integrati manualmente, dove possibile, tramite i singoli analisti dati di CAAB; questo era infatti l'unico modo a disposizione di questi ultimi per poter integrare i dati da loro richiesti. Tuttavia, come facile immaginarsi, questa pratica aveva però un impatto significativo sui processi di importazione e analisi dati dell'intera azienda, causando ritardi frequenti e complicazioni nel valutare ed integrare campi diversi, essendo di fatto i due flussi dati differenti e non sovrapponibili.

In sintesi, l'attività si è quindi concentrata sull'uniformare le dashboard per l'analisi dati presenti all'interno di Qlik Sense, e, al tempo stesso, ampliarle con i dati provenienti dal mercato europeo, standardizzando così l'intero processo di analisi dati.

In questo contesto, le fasi principali del progetto sono state:

1. analisi preliminari, in cui si sono analizzate le dashboard ed i rispettivi KPI, al fine di capire quali di essi fossero quelli più impattati dalle modifiche al modello dati precedentemente implementato. In questa fase, sono anche state definite le modifiche principali da apportare in relazione alle specifiche di business dettagliate dal cliente.
2. unificazione dati e reportistica, che ha interessato l'analisi specifica dei campi dati correlati alle metriche dei KPI.
3. modifica struttura del datawarehouse di test, in cui sono state effettivamente apportate le modifiche tecniche sul DW di prova.
4. test sulle modifiche, dove si è passati a testare ogni singola modifica fino a quel momento effettuata in ambiente di sviluppo.
5. recupero dello storico e rilascio in produzione, infatti i test sono stati effettuati solo su una frazione dei dati totali, per ragioni di efficienza prestazionale e di tempistiche. Per questa ragione, in contemporanea al rilascio in produzione, è stato recuperato tutto lo storico passato insieme anche ai dati del giorno stesso di rilascio. Si è quindi poi passati alle modifiche del DW di produzione, avendo come riferimento quanto fatto in ambiente di sviluppo.

Si procede ora ad illustrare nel dettaglio ogni singolo passaggio sopra elencato.

### 3.3.2 Analisi preliminari

La prima fase del progetto ha avuto come obiettivo l'analisi del modello dati insieme ai referenti del reparto ICT dell'azienda cliente, per comprendere sommariamente quali fossero le specifiche del nuovo flusso in input (da integrare tramite file sorgente in formato CSV) e quali le principali modifiche da effettuare, incluse quelle alle dashboard presenti in Qlik Sense impattate dalla modifica.

Sono state quindi condotte verifiche e sviluppati report, prevalentemente sotto forma di fogli di calcolo e di testo, per fare in modo di capire come andare ad unificare i dati esistenti con quelli provenienti dal sistema SFDC, cercando allo stesso tempo di mantenere inalterate il più possibile le dashboard precedentemente create dagli analisti di CAAB sulla piattaforma Qlik Sense. Si è cercato quindi di coprire il maggior numero possibile di dimensioni e KPI esistenti, ampliandoli però con i flussi SFDC relativi al mercato europeo.

Nello specifico, si è partiti dall'analisi del modello AS IS sotto riportato, che rappresenta la struttura dati del DW di CAAB che è basato sul flusso dati di Microsoft Dynamics; l'immagine rappresenta solo una parte dell'intero modulo del CRM, ovvero quella impattata dalle modifiche di progetto.

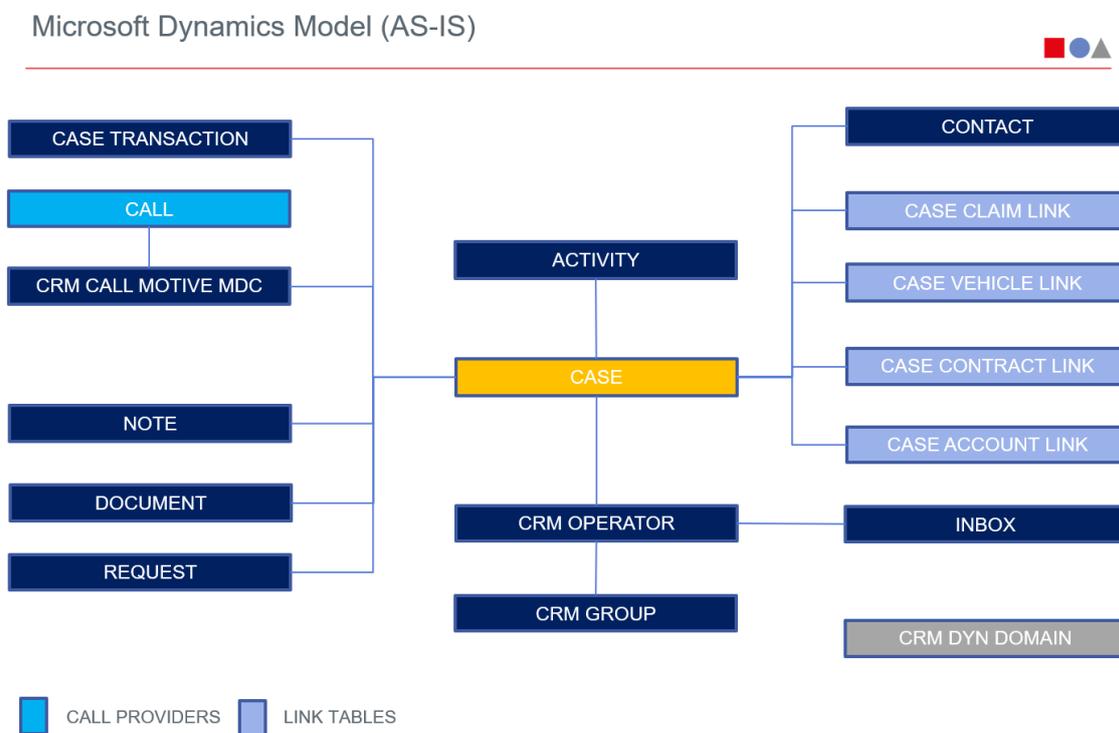


Figura 20 - Dettaglio del modulo CRM prima delle modifiche

Si procede ora ad illustrare il significato delle singole tabelle riportate:

- **Caso (case)**: rappresenta la registrazione di una richiesta, un problema o un reclamo da parte di un cliente, e il registro con tutto lo storico dei processi e delle attività collegate alla sua gestione; contiene le informazioni principali su un caso, come l'identificativo del caso, la data di apertura, la data di chiusura, lo stato, la priorità e il motivo di chiamata.
- **Transazione (transaction)**: contiene lo storico di tutti i cambiamenti che hanno interessato un singolo caso. Ad esempio, una transazione potrebbe registrare l'assegnazione di un caso a un operatore, l'aggiunta di una nota o la chiusura di un caso.
- **MDC (motivo di chiamata)**: è la dimensione principale che caratterizza un caso, in termini di tipologie di richieste e di regole per gestirle; contiene i motivi di chiamata utilizzati per classificare i casi, che possono poi essere utilizzati per automatizzare il processo di gestione dei casi, ad esempio assegnando automaticamente i casi a gruppi di operatori specifici.
- **Note (notes)**: contiene le note relative a un caso. Le note possono essere utilizzate per registrare informazioni aggiuntive sul caso, come le azioni intraprese dall'operatore o le risposte del cliente.
- **Documenti (documents)**: contiene i documenti relativi a un caso; questi possono essere utilizzati per fornire supporto alle informazioni contenute nel caso, come fatture, contratti o report di test.
- **Richiesta (request)**: contiene le informazioni su una richiesta specifica associata a un caso.
- **Attività (activity)**: rappresenta l'azione di registrazione di un'interazione tra un soggetto, tramite chiamata, email o altre forme di comunicazione, e un'attività assegnata ad uno specifico operatore o a un gruppo di operatori. Contiene quindi le informazioni su un'attività specifica associata ad un caso. Ad esempio, un'attività potrebbe riguardare una chiamata al cliente, un'email o un incontro.
- **Collegamento caso-veicolo (case vehicle link)**: contiene le informazioni sul collegamento tra un caso e un veicolo; ad esempio, un collegamento potrebbe essere utilizzato per registrare un reclamo relativo a un veicolo specifico.
- **Collegamento caso-reclamo (case claim link)**: contiene le informazioni sul collegamento tra un caso e un reclamo; ad esempio, un collegamento di questo tipo potrebbe essere utilizzato per registrare un caso relativo a un reclamo specifico.

- **Collegamento caso-profilo utente (case account link):** contiene il collegamento tra un caso e il profilo dell'utente coinvolto dal caso stesso.
- **Contatto (contact):** è la controparte che contatta il CRM, indipendentemente dal canale di contatto utilizzato. Contiene le informazioni sul contatto associato a un caso, inclusi ad esempio il nome, il cognome, l'indirizzo email e il numero di telefono.
- **Operatore (operator):** che si riferisce all'operatore o al gruppo di operatori che prendono in carico l'attività. Contiene informazioni sull'operatore, tra cui l'ID univoco di riconoscimento, il nome, il cognome e il gruppo di appartenenza.
- **Collegamento operatore-gruppo (operator\_group):** contiene le informazioni sul collegamento tra un singolo operatore ed il suo gruppo di appartenenza; ad esempio, un collegamento potrebbe essere utilizzato per registrare l'assegnazione di un operatore a un gruppo specifico.
- **Gruppo (group):** contiene le informazioni sul gruppo associato a un operatore, come il nome del gruppo e una sua descrizione.
- **Posta in entrata (inbox):** contiene informazioni per indirizzare le attività al gruppo di assistenza di riferimento.
- **Dominio del CRM (CRM domain):** contiene informazioni specifiche rispetto al DB Dynamics, come i metadati.

Per quanto concerne la tabella relativa alle Chiamate (Calls), come accennato nei capitoli precedenti, il suo flusso dati originario era limitato ai fornitori Dynamics, Comdata e Genesys. Ciascun fornitore gestiva una specifica barra telefonica, ad esempio dedicata ai finanziamenti o alle modifiche contrattuali. Successivamente, a questo flusso dati è stato integrato un nuovo fornitore, Amazon AWS, anch'esso associato a specifiche barre telefoniche. Per caricare i dati provenienti dal suo flusso dati, è stata inoltre introdotta una shell specifica in aggiunta a quelle esistenti, come dettagliato nel capitolo dedicato alla struttura del DW.

In questo contesto aziendale, una *barra telefonica* si riferisce a una divisione o segmento specifico di attività telefoniche all'interno di un sistema più ampio. Ogni barra telefonica può essere dedicata a un determinato servizio, come finanziamenti, modifiche contrattuali o altri aspetti specifici dell'attività aziendale.

I fornitori, come Dynamics, Comdata, Genesys e Amazon AWS, gestiscono ciascuno una propria barra telefonica, contribuendo così a fornire servizi telefonici specializzati e ad ampliare le capacità complessive del sistema aziendale.

In tal senso, è stato evidente come fosse necessario modificare il modello dati esistente sopra riportato. Infatti, tale modello dati (AS IS)

è stato confrontato con il flusso dati proveniente da SFDC, arrivando a definire il flusso finale da ottenere (TO BE), sotto riportato, ovvero con le modifiche da apportare al nuovo flusso dati da integrare per fare in modo di adattarlo e integrarlo a quello Microsoft Dynamics già implementato. Si sottolinea però che, nonostante i due flussi non avessero campi perfettamente sovrapponibili, essi avevano la stessa granularità dei dati.

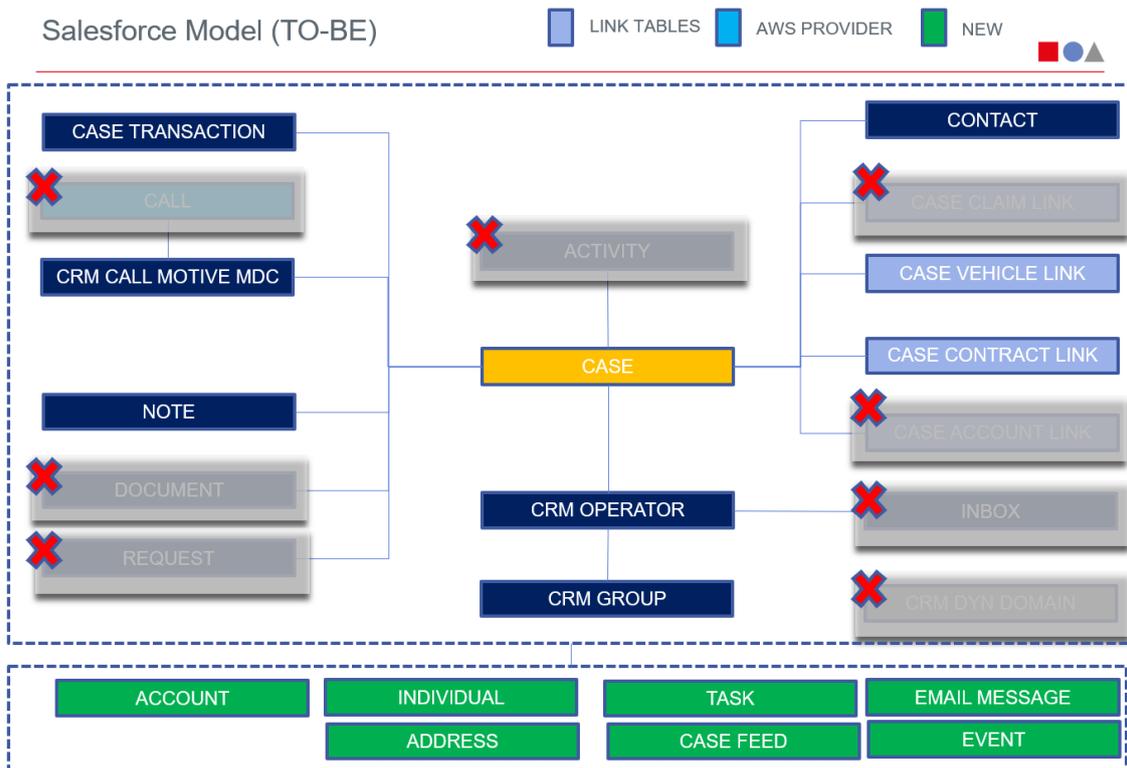


Figura 21 - Differenze tra i flussi dati DYN e SFDC

Come evidente dall'immagine, i due flussi dati non erano perfettamente sovrapponibili, ma presentavano diverse differenze. Infatti, il flusso SFDC non presentava alcune tabelle all'interno del suo modello dati, ovvero la tabella:

- Chiamata (call)
- Documento (document)
- Richiesta (request)
- Attività (activity)
- Collegamento caso-reclamo (case claim link)
- Collegamento caso-profilo utente (case account link)
- Posta in entrata (inbox)
- Dominio Dynamics del CRM (CRM Dynamics domain)

Fields					
Table	Total Fields	Spec. Italia	Missing Fields	Of which Used	Note
ACTIVITY	83	0	42	42	Salesforce has not this entity
CALL	23				<b>Out of Scope: Salesforce has not this entity. Call data will be integrated</b>
CASE	141	38	31	2	
CASE_ACCOUNT_LINK	14		10	0	Salesforce has not this entity
CASE_CLAIM_LINK	16	0	12	1	Salesforce has not this entity
CASE_CONTRACT_LINK	17	5	0	0	
CASE_TRANSACTION	74	3	10	7	
CASE_VEHICLE_LINK	21	7	1	0	
CONTACT	35		6	6	
CRM_CALL_MOTIVE_MDC	26	0	7	3	
CRM_DYN_DOMAIN	7		4	0	Salesforce has not this entity
CRM_GROUP	7	0	0	0	
CRM_OPERATOR	12	1	1	0	
DOCUMENT	14		10	1	Salesforce has not this entity
INBOX	9	2	4	0	Salesforce has not this entity
NOTE	21		1	0	
REQUEST	18		13	7	Salesforce has not this entity
<b>Grand Total</b>	<b>538</b>	<b>56</b>	<b>152</b>	<b>69</b>	

Figura 22 - Dettaglio delle tabelle coinvolte

Tuttavia, il flusso SFDC disponeva di nuove tabelle, di seguito descritte:

- Profilo utente (account): contiene le informazioni anagrafiche e di contatto di un cliente.
- Soggetto (individual): altre informazioni relative alla persona fisica.
- Indirizzo (address): contiene le informazioni sull'indirizzo dell'utente, come la via, il numero civico, il cap, la città, la provincia, e il paese.
- Attività (task): contiene le informazioni su un'attività pianificata, come la data di scadenza, la descrizione ed il suo stato.
- Messaggio Email (email message): contiene le informazioni sui messaggi di posta elettronica, come la data di invio dell'email, l'oggetto e il corpo della stessa e a quale gruppo di operatori deve essere indirizzata.
- Registro del caso (case feed): contiene aggiornamenti sullo stato di un caso, come note, transazioni e attività, ma anche comunicazioni tra gli operatori ed il business e documenti relativi al caso stesso, come fatture e contratti.
- Evento (event): contiene informazioni su eventi collegati ad uno specifico record.

Queste informazioni sono state inserite all'interno del modello dati in quanto possono essere utili per migliorare la gestione dei casi, fornendo agli operatori informazioni più complete e aggiornate sui casi da gestire; in tal modo viene assicurata al cliente una migliore e più efficiente gestione delle richieste.

L'introduzione di queste nuove tabelle è stata necessaria per colmare il divario tra i due flussi dati. Come precedentemente menzionato, l'obiettivo principale è stato quello di apportare modifiche al flusso SFDC in modo da integrarlo in modo coerente al resto del DW già presente, cercando al contempo di minimizzare le alterazioni alle dashboard. Tuttavia, due tabelle all'interno del modello dati di Dynamics, ovvero CRM Dynamics Domain e Activity, non hanno potuto essere sostituite con equivalenti SFDC, poiché assenti in quest'ultimo flusso dati. Di conseguenza, si è optato per una leggera modifica delle connessioni tra le tabelle. In particolare, la tabella Task ha acquisito un ruolo di rilievo, in quanto è stata direttamente collegata alla tabella Case. Questa modifica ha permesso di ristabilire il collegamento con la tabella Call (che altrimenti sarebbe andato perso) e con la tabella Case stessa.

Dopo aver analizzato nel complesso le tabelle coinvolte dalla modifica, si è poi passati all'analisi delle formule dei KPI all'interno delle dashboard di Qlik Sense. L'obiettivo è stato quello di andare ad individuare le metriche coinvolte e capire come poterle modificare efficacemente in modo da mantenerne una corretta visualizzazione all'interno delle dashboard stesse. Ciò è stato infatti necessario in quanto non era chiaro a priori il numero totale di KPI coinvolti dalle modifiche al modello dati, e quanti di essi non erano più in grado di essere calcolati tramite il nuovo flusso dati proveniente da SFDC, data la non perfetta compatibilità dei due modelli dati sopra descritti.

ID	App QlikSense	Sheet	KPI name	Formula	Missing/Out of scope Fields
KPI2	CRM Customer Care	Monitoraggio KPI-SLA per JLR - Riepilogo	Misura SLA	sum( distinct If( [In perimetro JLR] = 'SI' and ([[Channel Descr] <-> 'TELEFONO' and wildmatch([Mandate Descr], 'JAGUAR', 'LAND ROVER')]) or ([[Channel Descr] = 'TELEFONO' and wildmatch([CRMM Last VDN Case], '25020', '25029', '25040', '25050'))])	CRMM Last VDN Case; Mandate Descr
KPI10	CRM Customer Care	Monitoraggio KPI-SLA per JLR - AffidamentiCACs	Proposal	count({<[In perimetro JLR]=, [Filter JLR Monitoring excluded MDC]=, [Filter Requested or Created in Current Year]=, [Case Status Description-Case Status]=, [Direction Descr]=, [Mandate Descr]=	Mandate Descr
KPI26	CRM Customer Care	Monitoraggio KPI-SLA per JLR - Lista Completa	HH Lavorativi ACTV: Tempo medio per	Avg([Resolution Total Time (working) - Activity]/3600)	Resolution Total Time (working) - Activity
KPI44	CRM Customer Care	Monitoraggio KPI-SLA per JLR - Perimetro_SLAKPI	Numero Case	sum( if( (not WildMatch([Channel Descr], 'TELEFONO', 'EMAIL') and wildmatch([Mandate Descr], 'JAGUAR', 'LAND ROVER')) or ([[Channel Descr] = 'TELEFONO' and wildmatch([CRMM Last VDN Case], '25020', '25029', '25040', '25050')) or ([Channel Descr] = 'EMAIL' and	CRMM Last VDN Case; Mandate Descr

Figura 23 - Analisi delle formule di ogni KPI

Si è anche analizzata la necessità di mantenere attivo il calcolo di alcuni KPI, tramite il supporto del business, con l'obiettivo di eliminare le ridondanze e quelli ormai obsoleti e non più utilizzati.

Inoltre, per un numero esiguo di KPI, non è stato possibile integrare il flusso dati SFDC modificando le metriche, essendo queste specifiche

del mercato italiano o del DB Dynamics, e quindi si è deciso insieme al business di mantenerli attivi solo per i dati riguardanti il mercato Italia.

App QlikSense	Sheet	KPI name	Formula	Is Repeated	Computable
CRM Customer Care	Analysis Calls Inbound Daily - Sintesi Ripetizioni	Chiamate Gestite	Sum([Total Calls answered, managed by the CRM operator])	x	Out of scope Salesforce
CRM Customer Care	Analysis Calls Inbound Daily - Sintesi Ripetizioni	Chiamate Ricevute	Sum([Total Calls manageable])	x	Out of scope Salesforce
CRM Customer Care	Andamentale	Numero di Casi	sum([if([Market Descr-Market Id] = 'ITALY' and [Legal Entity Descr-Fcab Legal		Si
LTL Analysis Area 7 - CRM Customer Care	CASI - % OCS (CON ASSOCIAZIONE MDC-DIREZIONI AZIENDALI)	Numero di cases	Sum([Number of cases])	x	Si
LTL Analysis Area 7 - CRM Customer Care	CASI - % OCS (CON ASSOCIAZIONE MDC-DIREZIONI AZIENDALI)	% OCS	sum([if([Case Status Description-Case Status]='CHIUSO' And([Processing Level -		No
LTL Analysis Area 7 - CRM Customer Care	CALL CENTER - CALL_SLA_CSC - DETAIL CALL	Answer Time (Sec)	Sum([Answer Time Max (Sec)])	x	Out of scope Salesforce
LTL Analysis Area 7 - CRM Customer Care	CALL CENTER - CALL_SLA_CSC - DETAIL CALL	Conversation Time (Sec)	Sum([Conversation Time (Sec)])	x	Out of scope Salesforce

Figura 24 - Analisi della provenienza del dato

Su un totale di 374 KPI distribuiti su tre diversi applicativi Qlik Sense, 149 KPI erano misure uniche e, di questi ultimi, solamente per 14 è stata verificata la non compatibilità con il nuovo flusso dati proveniente dal sistema SFDC.

KPI Recap - CRM Customer Care		
Sheet	KPI name	Count of Formula
Monitoraggio KPI-SLA per JLR - Riepilogo	Misura SLA	1
Monitoraggio KPI-SLA per JLR - AffidamentiCACS	Proposal	1
Monitoraggio KPI-SLA per JLR - Lista Completa	HH Lavorativi ACTV	1
Monitoraggio KPI-SLA per JLR - Perimetro_SLAKPI	Numero Activity	1
	Numero Case	1
	HH Lavorativi ACTV	1
Monitoraggio Team: Lavoro Impegni Anno	N.ACTV Chiuse	1
<b>Grand Total</b>		<b>7</b>
KPI Recap - CRM Report Reclami		
Sheet	KPI name	Count of Formula
Mezzo Contatto - per report Reclami: Contatti_CC (Aperti)	Somma Tempo	1
	Number of Activity Open	1
<b>Grand Total</b>		<b>2</b>
KPI Recap - LTL Analysis Area 7 - CRM Customer Care		
Sheet	KPI name	Count of Formula
CALL CENTER - ACTIVITY_DETAIL	% OCS	1
CASI - % OCS (CON ASSOCIAZIONE MDC-DIREZIONI AZIENDALI) - BASE	Giorni Chiusura Solari	1
Email_CustomerCare - Base Dati	Resolution Working Day (Working)	1
Email_CustomerCare - Tabelle di Sintesi	Sum([Total Activities])	1
	sum([Total Activities])/sum([Resolution working day (working) - Activity])	1
<b>Grand Total</b>		<b>5</b>

Figura 25 - KPI non compatibili con SFDC

A questi 14 KPI corrispondono 96 campi appartenenti alle tabelle precedentemente illustrate. Di questi 96 campi, essendo alcuni di essi mancanti all'interno del flusso dati SFDC, dopo un'attenta analisi si è accertato che questi fossero ridondanti o non utilizzati all'interno delle dashboard presenti su Qlik Sense e quindi, previo accordo con il cliente, si è deciso di eliminarli e di escluderli dal sistema.

Row Labels	Count of field
ACTIVITY	7
CASE	32
CASE_ACCOUNT_LINK	5
CASE_CLAIM_LINK	4
CASE_TRANSACTION	1
CASE_VEHICLE_LINK	1
CONTACT	26
CRM_CALL_MOTIVE_MDC	5
CRM_GROUP	1
CRM_OPERATOR	2
DOCUMENT	4
NOTE	1
REQUEST	7
<b>Grand Total</b>	<b>96</b>

Figura 26 – Dettaglio dei campi appartenenti ai KPI non compatibili

Inoltre, 46 KPI sono stati identificati come *Out of scope* per SFDC, ovvero non impattati dalle modifiche per SFDC poiché facenti riferimento ai dati di alcune barre telefoniche delle chiamate prese in carico dal sistema AWS.

Row Labels	Si	No	Out of scope SF	Total
CRM Customer Care	58	5	29	92
CRM Report Reclami	12			12
LTR Analysis Area 7 - CRM Customer Care	19	9	17	45
<b>Grand Total</b>	<b>89</b>	<b>14</b>	<b>46</b>	<b>149</b>

Figura 27 - KPI non impattati dalle modifiche

Nello specifico, le applicazioni Qlik Sense interessate dalle modifiche sono state:

- CRM Customer Care
- CRM Customer Care – LTR Analysis Area 7
- CRM Report Reclami

Le prime due applicazioni, *CRM Customer Care* e *CRM Customer Care – LTR Analysis Area 7*, sono sostanzialmente equivalenti poiché utilizzano le stesse metriche e KPI. La differenza principale risiede nel fatto che la prima è relativa a CAAB, mentre la seconda si concentra sui contratti relativi ai veicoli LY. Entrambe includono calcoli e visualizzazioni basati sulle tabelle *call*, *activity* e *case*, con riferimento ai contratti specifici di interesse; quindi, da un lato si analizzano i contratti legati ai casi nel settore bancario e, dall'altro, quelli associati ai veicoli LY. Inoltre, vengono condotte analisi approfondite sui contratti di clienti specifici, prendendo in considerazione il loro storico.

L'ultima applicazione fra quelle citate, *CRM Report Reclami*, si concentra invece sull'analisi di casi particolari relativi a CAAB, ovvero i reclami, utilizzati per analizzare gli SLA (Service Level Agreement). Questi sono accordi contrattuali che stabiliscono i livelli di servizio attesi tra fornitori e clienti e in cui sono definiti parametri chiave come tempi di risposta, tempi di risoluzione e altri indicatori di performance, assicurando un quadro chiaro delle aspettative e degli impegni reciproci nell'erogazione dei servizi. L'applicazione CRM Report Reclami esamina quindi casi e note, valutando metriche come il numero totale di reclami, quelli accolti o respinti e lo SLA nella loro gestione; si tratta di un numero limitato di report rispetto alle altre applicazioni, poiché è focalizzata su una categoria molto specifica di dati.

### 3.3.3 Unificazione dati e reportistica

L'analisi dei KPI è stata legata strettamente a quella delle metriche usate per calcolarli, ottenute dai singoli campi delle tabelle presenti nei diversi flussi dati. Il passo successivo nello sviluppo del progetto è stato quindi quello di analizzare nel dettaglio tutti i campi coinvolti dalle modifiche ai KPI evidenziate nel passo precedente. L'obiettivo in questa fase è stato quello di capire quali di questi campi avessero una mappatura equivalente tra Dynamics e SFDC, e per quali campi invece non si trovasse alcuna relazione.

Con mappatura di un campo ci si riferisce alla corrispondenza tra i campi presenti all'interno del DW e quelli provenienti dal flusso dati SFDC. In altre parole, la mappatura definisce come i diversi elementi dei dati SFDC sono associati alle strutture di archiviazione già presenti, consentendo un'organizzazione efficace e coerente delle informazioni.

Table	field	MACROCLUS	NOTE	ENTITA' SFDC	CAMPO SFDC
ACTIVITY	OWNER_LEVEL	Dato Presente	Dati ricavabili consultando l'entità dedicata	EMAIL MESSAGE	Profile.name
ACTIVITY	OWNER_ROLE	Dato Presente	Dati ricavabili consultando l'entità dedicata	EMAIL MESSAGE	UserRole.name
ACTIVITY	ACT_OPERATOR	Dato Presente	Mapping 1:1	EMAIL MESSAGE	CreatedbyId
ACTIVITY	ACTIVITY_ID	Dato Presente	Mapping 1:1	EMAIL MESSAGE	Id
ACTIVITY	CASE_ID	Dato Presente	Mapping 1:1	EMAIL MESSAGE	ParentId
ACTIVITY	DIRECTION	Dato Presente	Mapping 1:1	EMAIL MESSAGE	Incoming
ACTIVITY	EMAIL_SUBJECT	Dato Presente	Mapping 1:1	EMAIL MESSAGE	Subject
CASE	OPEN_OPERATOR	Dato Presente	Mapping 1:1	CASE	CreatedbyId

Figura 28 - Esempio di mappatura

In questa fase di maggiore approfondimento tecnico, l'analisi ha interessato tutti i campi di tutte le tabelle impattate dalle modifiche, essendo poi questi utilizzati come metriche all'interno dei KPI. Sono

quindi stati prodotti innumerevoli report, uno per ogni singola tabella interessata, verificando, ad esempio, che campi con descrizioni uguali o molto simili avessero effettivamente il tipo di dati coerente con il campo di appartenenza e dello stesso formato di quello già presente nel DW.

Inoltre, è stato messo in rilievo quale fosse la nomenclatura utilizzata dai campi con una corrispondenza esatta tra SFDC e DYN, e allo stesso tempo anche quella relativa ai nuovi campi da inserire, in modo da creare un tracciato ben definito da utilizzare in fase di implementazione nel DW. Sono quindi state effettuate diverse verifiche tecniche sui dati nell'ambito della *data quality*.

Contact ODS	Staging DYN	Staging SFDC	Staging SFRC	Note	Datatype
cntr_cd_contact_id	contact_id	contact_sfdc_il	id		
cntr_cd_note	note	contact_sfdc_il	note_c		
cntr_fl_sap_privacy_flag	sap_privacy_flag	\	\	Popolato solo per DYN	
cntr_cd_privacy_ca	privacy_ca	\	\	Popolato solo per DYN	
cntr_ds_addr_address	addr_address	address_sfdc_il	street_c		
cntr_ds_addr_town	addr_town	address_sfdc_il	city_c		
cntr_cd_consent_marketin\	\	individual_sfdc_il	consent_marketing_phc	Nuovo campo popolato solo per SFRC	varchar
cntr_cd_consent_marketin\	\	individual_sfdc_il	consent_marketing_sms	Nuovo campo popolato solo per SFRC	varchar
cntr_cd_fcab_consent_pro\	\	individual_sfdc_il	fcab_consent_profilatio	Nuovo campo popolato solo per SFRC	varchar

Figura 29 - Analisi delle relazioni tra i campi DYN e SFDC

Per effettuare correttamente quest'analisi, sono stati anche analizzati i campi di frontend presenti nelle dashboard di Qlik Sense, andando ad evidenziare la corrispondenza con il relativo campo di backend nel DW; per ognuno di essi, sono poi state segnalate eventuali specificità rispetto al mercato di riferimento e alla presenza o meno nel flusso dati SFDC.

Note	Front end field	Back end field	Info
In Scope	External Contract ID-Case Contract Link	CCOL_CD_EXTERNAL_CONTRACT_ID	probabile specificità italia
In Scope	Filter JLR Monitoring excluded MDC	CCMM_CD_MDC_L2 CCMM_CD_MDC_L3	Presente in Salesforce
In Scope	Filter Requested or Created in Current Year	DIPE_CD_MARKET_ID DIPE_CD_DAY_YYYYMMDD	LTR
In Scope	Filter Socio Demo Included MDC- Cases	CASE_CD_MDC_MOTIV_ID	Presente in Salesforce
In Scope	fine_lavorazione	CASE_DT_CASE_CREATION_DATE CASE_CD_CURRENT_CASE_STATUS CASE_DT_CASE_CLOSING_DATE	Presente in Salesforce
In Scope	Flag ultimo contatto (by contract) - Analisi Richieste EA		Presente in Salesforce /Retail Data
In Scope	flag_conteggio_ea		Retail Data
In Scope	In perimetro JLR		Retail data

Figura 30 - Collegamento tra i campi di backend e frontend

Inoltre, per svolgere questa associazione e capire come dover valorizzare correttamente i diversi campi sono stati utilizzati i metadati,

ovvero attributi che forniscono contesto e informazioni aggiuntive sulla tipologia di dato da usare per valorizzare il campo di riferimento.

Questi sono essenziali per comprendere, organizzare e gestire efficacemente i dati, e descrivono informazioni come la tipologia di dato (testo, numero, data/ora, ...), la lunghezza massima che può avere il dato da inserire e la tipologia di campo. Ogni campo gestisce infatti una diversa tipologia di dato utile alla descrizione di ogni tabella nel flusso dati, ed è quindi fondamentale conoscere il tipo di formato corretto da utilizzare per valorizzarlo, al fine di non compromettere il sistema e avere errori durante le shell di caricamento dati.

Conoscere il formato dati corretto del campo è poi fondamentale al momento dell'integrazione delle nuove tabelle SFDC all'interno del DW, essendo necessario definire tramite codice SQL la tipologia di dato che è possibile inserire ed eventuali controlli da effettuare su di esso al momento dell'inserimento da parte dell'utente.

DYN name	SFDC name	Data Type	Length	Field Type
Address Name	Name	Auto Number	80	Standard
Account ID	Account__c	Lookup (Account)		Custom
Address Number	FCAB_AddressNumber__c	Text	10	Custom
Address Type	AddressType__c	Text	255	Custom
Address Valid From	FCAB_Valid_From__c	Date		Custom
Address Valid To	FCAB_BankValid_To__c	Date		Custom
AddressNumber	AddressNumber__c	Text	10	Custom
Alternative Fax	Alternative_Fax__c	Text	255	Custom
Alternative Mobile	Alternative_Mobile__c	Phone	40	Custom
Certified Email	FCAB_Certified_Email__c	Email		Custom
City	City__c	Text	255	Custom
Company	Company__c	Picklist		Custom
Contact ID	Contact__c	Lookup (Contact)		Custom
Created By ID	CreatedById	Lookup (User)		Standard
Created Date	CreatedDate	Date/Time		Standard
Domicile City	Domicile_City__c	Text	255	Custom
Domicile House Number	Domicile_House_Number__c	Text	10	Custom
Domicile Postal Code	Domicile_Postal_Code__c	Text	255	Custom
Domicile Province	Domicile_Province__c	Text	255	Custom
Domicile Street	Domicile_Street__c	Text	255	Custom
Domicile_Nation	Domicile_Nation__c	Text	255	Custom
Fax Number	Fax__c	Phone	40	Custom
FCAB Address	FCABAddress__c	Text	255	Custom

Figura 31 - Specifica della tipologia del campo

Sempre all'interno della fase di riconciliazione dei dati dei due sistemi, si è poi passati all'analisi della codifica dello stato dei casi, per un totale di 7 possibili stati, e all'analisi delle tipologie possibili di transazioni, ovvero 18.

Case status SFDC	Case status DYN	Case status description DYN
Open	1	APERTO
Assigned to operator	2	ASSEGNATO A OPERATORE
Assigned to group	3	ASSEGNATO A GRUPPO
Under Processing I Leve	4	IN LAVORAZIONE
Under Processing	4	IN LAVORAZIONE
Suspended	5	SOSPESO
Canceled	6	ANNULLATO
Closed	7	CHIUSO

*Figura 32 - Tabella di transcodifica utilizzata tra DYN e SFDC per lo stato del caso*

Ciò è stato necessario in quanto i due flussi dati dispongono di due codifiche diverse, ovvero campi completamente differenti per formato e logica.

Infatti, le logiche implementate all'interno delle procedure ELT del DW, essendo basate sulla struttura dati DYN, andavano a recuperare l'informazione sulla tipologia/stato del caso o della transazione tramite un campo contenente un numero intero al quale era poi associata la relativa descrizione in formato stringa. Ciò non era però il caso dei dati appartenenti al flusso SFDC, in quanto questi andavano ad inserire all'interno delle tabelle dei fatti direttamente la descrizione associata ad un caso, senza utilizzare alcun tipo di codifica intermedia.

Questo ha determinato la transcodifica di questi ultimi dati direttamente all'interno del DW, poiché il fornitore Salesforce è stato impossibilitato a modificare il formato di invio dei dati, come originariamente richiesto da CAAB.

In particolare, per la transcodifica sono state implementate tabelle di decodifica all'interno del DW che sono state progettate per essere richiamate durante le fasi di ELT nei livelli di Staging e Integration Layer. È importante notare che questo approccio è stato adottato esclusivamente per i dati relativi a Salesforce (SFDC), in quanto i dati provenienti da DYN sono già dotati di un campo di codifica numerico e non richiedono questo processo aggiuntivo.

Questo ha quindi permesso di uniformare le codifiche presenti e di avere un impatto minimo all'interno delle logiche/procedure già presenti in DW.

Transaction type SFDC	Transaction type DYN	Transaction type description DYN
Opening case	1	APERTURA CASO
Automatic opening of the case	2	APERTURA AUTOMATICA CASO
Case assignment	3	ASSEGNAZIONE CASO
Automatic case assignment	4	ASSEGNAZIONE AUTOMATICA DEL CASO
Change case	5	MODIFICA CASO
MdC reclassification	6	RICLASSIFICAZIONE MDC
Suspension case	7	SOSPENSIONE CASO
Case closure	8	CHIUSURA CASO
Sending document	9	INVIO DOCUMENTO
Reminder Added	10	INSERIMENTO SOLLECITO
Note entry	11	INSERIMENTO NOTA
Case cancellation	12	ANNULLAMENTO CASO
Reopening case	13	RIAPERTURA CASO
Change additional data	14	MODIFICA DATI AGGIUNTIVI
Change Flag Urgent case - I level	15	MODIFICA FLAG CASO URGENTE PRIMO LIVELLO
Change Flag Urgent case - II level	16	MODIFICA FLAG CASO URGENTE SECONDO LIVELLO
Rescheduling of the expiration date	17	RIPIANIFICAZIONE DELLA DATA DI SCADENZA
Case Processing	18	LAVORAZIONE CASO

*Figura 33 - Tabella di transcodifica utilizzata tra DYN e SFDC per il tipo di transazione*

In questo contesto, è stata dedicata particolare attenzione alla ridefinizione delle logiche sottostanti alle caselle di posta in arrivo (inbox) di SFDC.

Le caselle di posta degli operatori, responsabili della gestione delle richieste legate al CRM, sono infatti suddivise in gruppi attraverso logiche personalizzate definite dal business di CAAB; questa suddivisione si basa sulle specifiche aree di competenza dei gruppi di operatori. Anche in questo caso però i file inviati da SFDC avevano una gestione delle Inbox differente rispetto a quella già implementata nel DW DYN, dunque è stata necessaria una collaborazione particolare con gli utenti per capire come andare a modificare i filtri alla base delle caselle di posta, andando ad ottenere all'incirca le stesse logiche presenti in DYN.

L'obiettivo è infatti sempre stato quello di ottenere logiche simili a quelle già presenti in DYN, assicurando una coerenza e continuità nel processo di gestione delle richieste.

Inbox AS IS	DIREZIONE NEW	Responsabile Direzione	MERCATO
INBOX DI ANTONELLO PAOLA	FCA BANK	FCA BANK	Italia
INBOX DI ANTONIO TROCONIS	N.C.	TBD	Italia
INBOX DI ANTONUZZI PAMELA MARIA	SALES	NARDI PIETRO	Italia
INBOX DI APPETICCHIA NORMANNO	REMARKETING	MERLO MATTEO	Italia
INBOX DI ARCA DANIELA	CSC	D'ARCO ROLANDO	Italia
INBOX DI ARCANGELI MARTINA	CSC	D'ARCO ROLANDO	Italia
INBOX DI ARCELLAZZI VALENTINA MARIAPIA	SALES	NARDI PIETRO	Italia
INBOX DI ARIEMMA STEFANIA	FCA BANK	FCA BANK	Italia
INBOX DI BOTTICELLI ORIANA	SUPPLY CHAIN	ALLASIA RAFFAELE	Italia
INBOX DI ARMENTI GIANLUCA	FCA BANK	FCA BANK	Italia
INBOX DI ARNABOLDI NICOLETTA MARIA	CSC	D'ARCO ROLANDO	Italia
INBOX DI ARNESE DONATELLA	FCA BANK	FCA BANK	Italia
INBOX DI ASSENZA ANNA	CREDIT	PETRELLI DARIO	Italia
INBOX DI ASUNI DEBORA	CSC	D'ARCO ROLANDO	Italia
INBOX DI AULENTA ROBERTO	CREDIT	PETRELLI DARIO	Italia
INBOX DI AVELLINA NUNZIATINA	FCA BANK	FCA BANK	Italia
INBOX DI AVOLIO ARMANDO	SALES	NARDI PIETRO	Italia
INBOX DI BABANDO VERONICA	N.C.	TBD	Italia
INBOX DI BAGNATO GAIA	CSC	D'ARCO ROLANDO	Italia
INBOX DI BAGNO PAOLA ANNA	SALES	NARDI PIETRO	Italia
INBOX DI BAGNULO ALESSIA	FINANCE	HAMODRAKAS NICOLAS	Italia
INBOX DI CARGANICO AMEDEO	SUPPLY CHAIN	ALLASIA RAFFAELE	Italia

Figura 34 - Analisi delle Inbox

In conclusione, a partire dalle nomenclature dei campi presenti nei KPI delle dashboard, si sono trovati i campi DYN corrispondenti in backend ed i corrispettivi relativi a SFDC, anche facendo riferimento ai metadati.

Dove non presenti dei corrispettivi campi SFDC, si è proceduto a crearne di nuovi coerenti con la nomenclatura e le caratteristiche del dato richieste dal campo stesso.

Tuttavia, come detto in precedenza, per un numero esiguo di essi si è scelto insieme al cliente di eliminarli dal sistema poiché ridondanti o non più utilizzati all'interno delle dashboard.

### 3.3.4 Modifica struttura del datawarehouse

Le modifiche apportate alla configurazione del datawarehouse, la cui struttura originaria di partenza è stata illustrata nel capitolo dedicato, hanno inciso sui livelli di staging e ODS.

Tuttavia, tali modifiche hanno anche previsto l'introduzione di un livello intermedio tra questi, noto come Integration Layer (IL), specificamente concepito per gestire l'integrazione dei dati derivanti dal flusso proveniente da SFDC.

Nonostante anche i dati SFDC siano inizialmente copiati e gestiti all'interno del livello di staging, si è resa necessaria l'introduzione di questo livello apposito. Questo perché, essendo il formato dei dati SFDC in ingresso diverso da quello delle varie fonti dati, non sarebbe stato possibile effettuare la storicizzazione direttamente in ODS, come accade invece per gli altri dati. Ciò a causa del fatto che il formato dei dati "grezzi" di SFDC è differente rispetto a quello Dynamics presente sul DW, quindi sarebbe stato impossibile integrarli e storicizzarli correttamente.

Questo livello di integrazione assolve infatti a diversi compiti:

- effettua la storicizzazione dei dati SFDC, in quanto in staging arrivano giornalmente solo le modifiche ai dati rispetto al giorno precedente, quindi esclusivamente la variazione rispetto ai dati già presenti nel DW.
- implementazione di logiche specifiche per il flusso dati SFDC in modo tale da suddividerlo e recuperare solamente i dati necessari per ricondursi al modello DYN del DW, procedendo all'effettiva integrazione dei dati. Così facendo, questi dati possono essere utilizzati in modo efficace all'interno del sistema dati esistente, consentendo l'analisi e la generazione di report senza problemi di compatibilità o integrazione.

L'obiettivo dell'IL è dunque quello di garantire la coerenza e l'omogeneità dei dati provenienti da Salesforce rispetto agli altri dati presenti nel datawarehouse, ma anche occuparsi della loro storicizzazione.

Quindi, i dati SFDC hanno una doppia storicizzazione:

- funzionale alla riconciliazione dei dati, in IL, con l'obiettivo di riconciliare le logiche ed i formati, essendo diversi tra SFDC e DYN.
- per tenere traccia dei cambiamenti nelle tabelle dei fatti, in ODS, come avviene per tutti gli altri dati.

Ad ogni modo, per entrambe le tipologie di dati l'implementazione delle logiche di business avviene nell'ODS.

L'ultima modifica ha riguardato quest'ultimo livello, in quanto si è reso necessario cambiare le logiche di business già implementate, in modo tale da integrare i nuovi dati SFDC. Sono quindi stati modificati la mappatura dei campi, integrando anche quelli SFDC, e le formule per il calcolo di campi personalizzati.

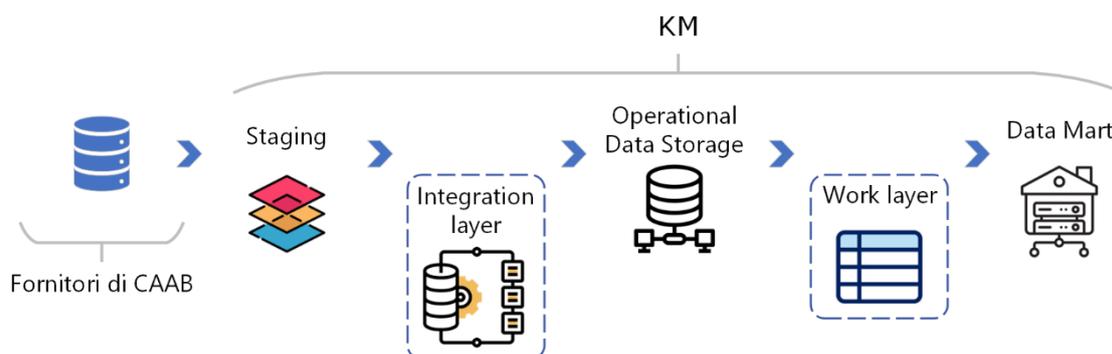


Figura 35 - Architettura del datawarehouse dopo il progetto

Ovviamente, oltre alla modifica dei campi presenti nelle tabelle coinvolte nei vari livelli del DW, si sono anche dovute creare le relative procedure per il caricamento dei dati, così come i file YAML richiesti dallo strumento GPLOAD, interno al DW Greenplum, per poter svolgere correttamente le operazioni di inserimento dati ELT.

Infine, sono state effettuate le modifiche alla programmazione delle shell tramite IBM TWS, in modo tale da creare anche i processi ELT per il caricamento di questo nuovo flusso dati all'interno del DW.

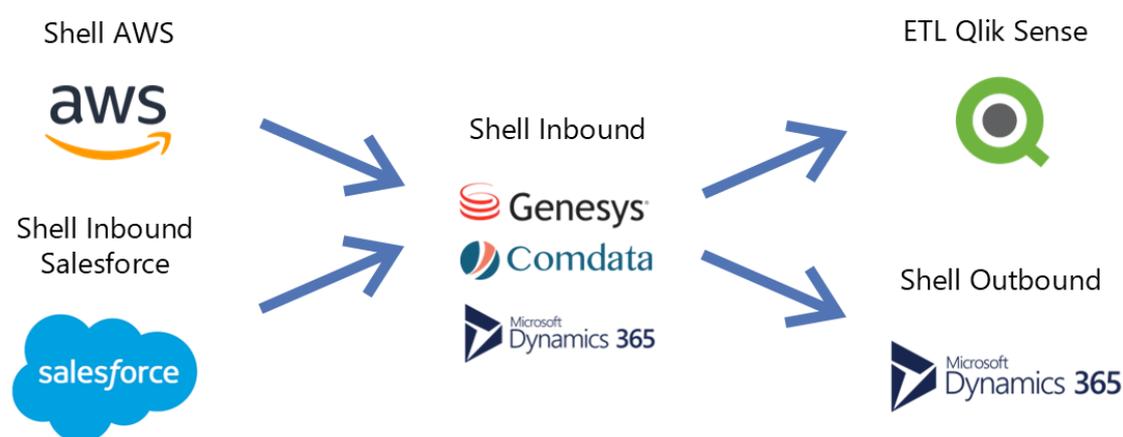


Figura 36 - Sequenza di avvio delle shell dopo il progetto

La nuova shell, facendo riferimento ai dati SFDC delle chiamate di alcune barre telefoniche, è stata programmata per essere avviata in

parallelo alla shell delle chiamate AWS. Come quest'ultima, anche la shell SFDC deve però soddisfare un requisito necessario all'avvio dello script, ovvero che sia disponibile il file sorgente con i dati inviati dal fornitore Salesforce. In particolare, l'orario di invio dei dati da parte dei due fornitori è differente (avviene comunque per entrambi durante la notte), ma le due shell, non essendo correlate tra loro, possono partire in autonomia.

In seguito, una volta che entrambe queste due shell hanno completato con successo il processo ELT dei dati, può partire la shell Inbound DYN e, successivamente, tutte le altre.

Ci sono quindi due modi in cui sono gestite le schedulazioni nel processo ELT di caricamento dei dati nel DW:

- partenza della shell all'arrivo dei file, come nel caso delle shell AWS e SFDC.
- partenza della shell all'arrivo dei file e al completamento delle shell precedenti, come nel caso della shell Inbound, Outbound e ETL Qlik Sense.

Di tutto questo processo se ne occupa lo schedulatore, che verifica che tutte le condizioni poste alla partenza della shell siano rispettate e, nel caso lo siano, lancia la shell richiesta.

L'ultima fase delle modifiche ha riguardato il modello dati presente in Qlik Sense. Si è infatti andati ad eseguire le connessioni logiche tra le varie tabelle dati, come previsto dall'analisi preliminare svolta in precedenza, insieme alle piccole correzioni di frontend ai KPI e alle visualizzazioni delle dashboard.

### 3.3.5 Test sulle modifiche

Una volta completata la modifica alla struttura del DW e del modello dati, si è passati alla fase di test sulle modifiche effettuate. I test sono stati molteplici e hanno riguardato l'analisi dei report, precedentemente identificati come "in scope SFDC" in fase di analisi preliminari, tramite fogli di calcolo creati appositamente.

In un primo momento, sono stati analizzati i report nella loro interezza, con una visione d'insieme sulle modifiche da dover testare in ogni singola pagina delle dashboard.

App QlikSense	Sheet	Analysis Area	In scope for current UATs	Notes
LTR Analysis Area 7 - CRM Customer Care	CALL CENTER CASI_BO_SLA_CSC	Calls, Case	In scope	- Test connections between Cases and Calls - Manage new MDC coding in filters /fields - Verify KPIs and aggregations
LTR Analysis Area 7 - CRM Customer Care	CALL CENTER - CASI_EMAIL_SLA_CSC_V.3	Case	In scope	- Manage missing Activity entity - Manage new MDC coding in filters /fields - Manage worklist field - Verify KPIs and aggregations
LTR Analysis Area 7 - CRM Customer Care	CASI - OVERVIEW_CASI_GENERATI	Case, Contract	In scope	- Test connections between Cases and contract - Manage new MDC coding in filters - Verify KPIs and aggregations
LTR Analysis Area 7 - CRM Customer Care	CASI - OVERVIEW_CASI_GENERATI CON TERRITORY	Case, Contract	In scope	- Test connections between Cases and contract - Manage new MDC coding in filters - Verify KPIs and aggregations
LTR Analysis Area 7 - CRM Customer Care	CASI - OVERVIEW_CASI_GENERATI CON TERRITORY_REPORT	Case, Contract	In scope	- Test connections between Cases and contract - Manage new MDC coding in filters - Manage inbox mapping in external file - Verify KPIs and aggregations
LTR Analysis Area 7 - CRM Customer Care	CASI - MONITOR BACKLOG FORNITORI - RIEPILOGO	Case, Vehicle, Contract	In scope	- Test connections between Cases, Contract and Vehicle - Manage new MDC coding in filters - Manage inbox mapping in external file - Verify KPIs and aggregations

Figura 37 - Analisi dei singoli fogli delle dashboard

Successivamente, si sono analizzati più nel dettaglio i singoli oggetti, e relativi campi, metriche e KPI, per ogni pagina all'interno delle visualizzazioni Qlik Sense create dal business.

App	Sheet	Object	Field	Notes
LTR Analysis Area 7 - CRM Customer Care	KPI_RECLAMI - Base Dati	1st field in table	ID Caso	To understand if MDC L3 = RECLAMO in DYN is equal to MDC L3 = COMPLAINT/COMPLAINTS in SFRC
LTR Analysis Area 7 - CRM Customer Care	KPI_RECLAMI - Tabelle	1st field in table (x4 tables)	Tipologia Esito	- To understand if MDC L3 = RECLAMO in DYN is equal to MDC L3 = COMPLAINT/COMPLAINTS in SFRC - to understand how to manage missing Request Result/Closure Motive
LTR Analysis Area 7 - CRM Customer Care	KPI_RECLAMI - Tabelle	2nd field in last table	Tipologia Esito	to understand how to manage missing Request Result/Closure Motive
LTR Analysis Area 7 - CRM Customer Care	KPI_RECLAMI - Tabelle	1st field in 5th and 6th table	Macro Cluster	-to identify with business users which MDC correspond to the ones needed
LTR Analysis Area 7 - CRM Customer Care	KPI_RECLAMI - Tabelle_privati	1st field in table (x5 tables)	Tipologia Esito	- To understand if MDC L3 = RECLAMO in DYN is equal to MDC L3 = COMPLAINT/COMPLAINTS in SFRC - to understand how to manage missing Request Result/Closure Motive
LTR Analysis Area 7 - CRM Customer Care	KPI_RECLAMI - Tabelle_privati	2nd field in last table	Tipologia Esito	to understand how to manage missing Request Result/Closure Motive
LTR Analysis Area 7 - CRM Customer Care	KPI_RECLAMI - Tabelle_privati	1st field in 5th and 6th table	Macro Cluster	-to identify with business users which MDC correspond to the ones needed
LTR Analysis Area 7 - CRM Customer Care	KPI_Spagna - Tabelle	Filters (x3 Bookmarks)	MERCATO	to verify the external file "TUTTE le inbox del CRM - codificate" and to eventually add the inbox received from SFRC

Figura 38 - Analisi degli oggetti all'interno di ogni dashboard

Questa fase, come evidenziato successivamente all'interno del cronoprogramma, è stata centrale all'interno del progetto e ha richiesto una notevole quantità di tempo, circa 30 giorni, pari a circa un terzo della durata complessiva del progetto.

Ciò è stato indispensabile per garantire la coerenza e l'affidabilità delle modifiche apportate, assicurando il loro pieno allineamento alle esigenze del business.

In particolare, sono stati testati i cambiamenti a:

- procedure ELT, per validare il passaggio dei dati attraverso tutti i livelli del DW.
- tabelle dati, per integrare i nuovi dati e il loro nuovo formato.
- modello dati, verificando le connessioni logiche tra le tabelle.
- contenuto dei dati, insieme al business e ai fornitori hardware e software si è verificato che i nuovi dati fossero effettivamente quelli corretti, ovvero quelli validati dal cliente.
- dashboard, andando a controllare che i nuovi dati si integrassero e fossero visualizzati correttamente dal business. Nello specifico, sono stati testati i dati, i collegamenti e i pulsanti presenti in ogni foglio di ogni dashboard esistente.
- funzionamento generale.

Sono quindi stati effettuati i *test di non regressione e di integrazione (non-regression and integration test)*, che sono due tipologie di test software fondamentali nell'ambito della *data quality* per garantire la qualità dei dati e la stabilità dell'intero processo di manipolazione durante lo sviluppo e l'evoluzione del datawarehouse.

Nello specifico, i test di non regressione sono progettati per assicurare che le modifiche apportate al codice non introducano nuovi difetti o errori nelle funzionalità esistenti, e dunque che tutto continui a funzionare correttamente, mantenendo la coerenza e l'integrità delle funzionalità e dei dati preesistenti.

Allo stesso tempo, i test di integrazione sono svolti per assicurare che i vari moduli del DW, sviluppati separatamente, funzionino correttamente quando sono integrati insieme.

Questi test verificano quindi la corretta interoperabilità tra le diverse componenti del software e si concentrano sulla rilevazione di eventuali errori o problemi che potrebbero emergere durante il processo di integrazione.

Entrambe queste tipologie di test sono state cruciali per mantenere la robustezza e l'affidabilità del sistema durante lo sviluppo e hanno permesso una verifica rapida ed efficiente in risposta alle modifiche apportate al codice precedentemente implementato.

### 3.3.6 Recupero storico e rilascio in produzione

Al momento del rilascio in produzione (*Go Live*) del nuovo flusso dati Salesforce all'interno del datawarehouse aziendale, si è proceduto ad effettuare il recupero di tutto lo storico dei dati. Ciò a causa del fatto che tutte le fasi precedenti del progetto sono state completate esclusivamente con un sottoinsieme di dati, per ragioni di efficienza. In particolare, per i test sono stati utilizzati i dati presenti nel DW relativi agli ultimi tre mesi precedenti all'inizio dello sviluppo del progetto, partito a settembre 2023.

Nel giorno di rilascio in produzione del progetto, essendo il primo invio, i dati di tutti i flussi dei diversi fornitori sono stati inviati in modalità FULL, ovvero è stato inviato all'interno del DW aziendale tutto lo storico a loro disposizione.

In questa fase, si sono quindi utilizzate le procedure ELT necessarie al corretto svolgimento di questo processo, come spiegato nei capitoli precedenti, per tutti i diversi livelli del DW. A tal proposito, sono state modificate le procedure ELT facendole partire in modalità FULL, e non DELTA, poiché essendo la prima importazione si aveva necessità di avere tutti i dati presenti e non solamente una parte di essi.

Infatti, come già sottolineato, le procedure definite durante il progetto sono state pensate per importare solamente i dati modificati nel giorno precedente dal cliente, e non l'intero dataset, per ovvi motivi prestazionali.

Ci si è inoltre assicurati che le schedulazioni automatiche definite fossero disattivate, dato che al momento del rilascio si è appunto proceduto ai caricamenti manuali dei dati, in modo da assicurarsi che tutti i dati fossero importati correttamente. Le schedulazioni automatiche definite durante il progetto sono state poi attivate correttamente al termine del rilascio in ambiente di produzione, ovvero nei giorni successivi.

Le problematiche riscontrate in quest'ultima fase di progetto hanno fatto sì che non sia stato possibile portare a termine il rilascio in produzione in una sola giornata, come inizialmente preventivato, ma siano stati necessari circa tre giorni.

I principali problemi riscontrati sono stati:

- trasmissione dei file lenta, a causa dell'enorme mole di record contenuta in ognuno di essi.

- tempi più lunghi necessari alla verifica dei dati inviati, sia da parte dei fornitori dati che da parte di KM e degli utenti business di CAAB.
- recupero dei dati relativi al giorno del rilascio, in quanto ci si è accorti solo in seguito che SFDC non lo includeva all'interno dell'invio in modalità FULL.

Questo ha comportato inoltre la necessità di recuperare i dati generati durante i giorni in cui si è effettuato il rilascio, essendosi questo protratto sfortunatamente su più giornate lavorative. Ciò è stato fatto sempre manualmente, andando ad utilizzare giorno per giorno le procedure in modalità DELTA, in modo da caricare solo i dati di interesse.

### 3.3.7 Cronoprogramma

Il cronoprogramma per l'implementazione del nuovo flusso dati SFDC all'interno del modulo CRM del DW aziendale è stato suddiviso come segue:

- progettazione del backend
  - analisi preliminari.
  - definizione della nuova struttura dati.
  - scrittura e test delle nuove procedure ELT.
- modifiche di frontend
  - aggiunta dei nuovi script in Qlik Sense, sia per la fase ETL che per l'implementazione delle nuove logiche.
  - modifica alle dashboard esistenti.
- test di accettazione dell'utente (UAT), con un innumerevole serie di test per verificare il corretto funzionamento di quanto implementato, secondo le specifiche originariamente richieste dal cliente.
- consegna in produzione (Go Live), implementazione di quanto fatto in ambiente di produzione e messa a disposizione agli utenti del nuovo sistema.

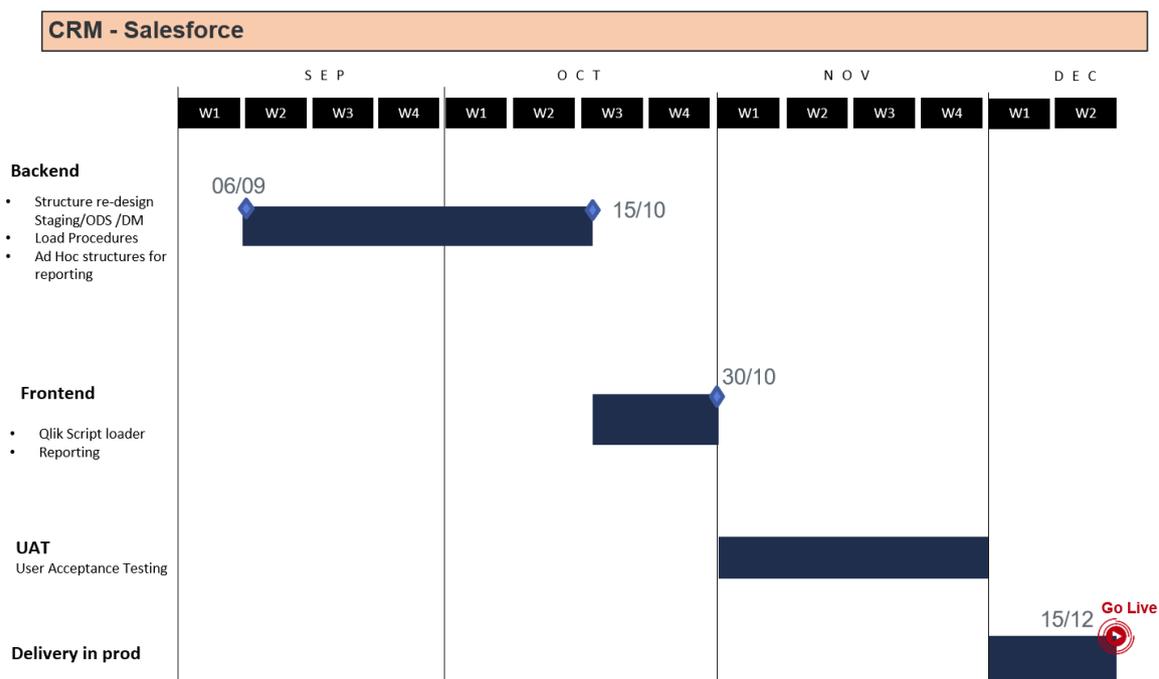


Figura 39 - Cronoprogramma del progetto

## 4. Conclusioni e sviluppi futuri

La gestione e l'ottimizzazione dell'utilizzo dei dati per l'estrazione di valore rappresentano un pilastro fondamentale per le aziende dinamiche e internazionali, come quella esaminata nel presente elaborato. Le imprese orientate ai dati (data-driven) si basano infatti sulle informazioni estratte dai dati transazionali-operativi provenienti da diverse aree di business, evidenziando la crescente necessità di definire processi e sistemi automatizzati efficienti per il corretto immagazzinamento e integrazione dei dati stessi.

In questo contesto, il progetto si è rivelato essenziale poiché ha introdotto un nuovo flusso di dati nel sistema informativo esistente, riducendo al minimo le modifiche alla struttura dati preesistente e gli impatti sulla reportistica già creata. Questo ha permesso di standardizzare ed ottimizzare l'analisi dei dati provenienti da diversi fornitori IT, svolgendo un ruolo cruciale nell'armonizzazione dell'analisi dei dati provenienti da varie nazioni europee.

Tale iniziativa ha quindi avuto un grande impatto sull'azienda in quanto ne ha supportato efficacemente l'espansione internazionale, mitigando i ritardi e i costi associati alla gestione inefficiente dei dati.

Attraverso l'implementazione di un processo di integrazione più efficiente, l'azienda è ora in grado di gestire in modo ottimale una vasta gamma di fonti di dati, consentendo una visione più completa e accurata delle sue attività commerciali.

Inoltre, la realizzazione di questo progetto ha fornito una solida base per migliorare la qualità delle decisioni aziendali e sostenere la crescita nel lungo periodo, fornendo agli stakeholder aziendali uno strumento versatile e affidabile.

Per quanto concerne i possibili sviluppi futuri, questi fanno riferimento principalmente a due aspetti:

- sviluppo di dashboard e report avanzati, per fornire agli utenti finali una visualizzazione più approfondita e interattiva dei dati, potenziando l'interfaccia grafica della dashboard sviluppata andando ad aumentare il grado di interattività concesso all'utente. In tal modo si riuscirebbe a sfruttare a pieno il contenuto informativo apportato dal nuovo flusso dati.
- ottimizzazione delle performance, per cui potrebbe rivelarsi opportuno esaminare attentamente il modello dati, le procedure di caricamento ETL e l'utilizzo effettivo dei campi e delle tabelle, eventualmente implementando opportune strategie di indicizzazione e de-normalizzazione, al fine di migliorare significativamente le prestazioni del sistema.

## Sitografia

- **IBM**, <https://www.ibm.com/topics/business-intelligence>, ultimo accesso Ottobre 2023.
- **Tableau**, <https://www.tableau.com/learn/articles/business-intelligence>, ultimo accesso Ottobre 2023.
- **Qlik**, [https://help.qlik.com/en-US/qlikview/May2023/Content/QV\\_HelpSites/Difference-qlikview-qliksense.htm](https://help.qlik.com/en-US/qlikview/May2023/Content/QV_HelpSites/Difference-qlikview-qliksense.htm), ultimo accesso Ottobre 2023.
- **Qlik**, <https://www.qlik.com/us/products/qlik-sense>, ultimo accesso Ottobre 2023.
- **Greenplum**, <https://greenplum.org/introduction-to-greenplum-architecture>, ultimo accesso Ottobre 2023.
- **VMware**, [https://docs.vmware.com/en/VMware-Greenplum/7/greenplum-database/admin\\_guide-intro-arch\\_overview.html](https://docs.vmware.com/en/VMware-Greenplum/7/greenplum-database/admin_guide-intro-arch_overview.html), ultimo accesso Ottobre 2023.
- **VMware**, [https://docs.vmware.com/en/VMware-Greenplum/7/greenplum-database/admin\\_guide-admin\\_guide.html](https://docs.vmware.com/en/VMware-Greenplum/7/greenplum-database/admin_guide-admin_guide.html), ultimo accesso Ottobre 2023.
- **VMware**, <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/docs/vmw-datasheet-greenplum.pdf>, ultimo accesso Ottobre 2023.
- **DBeaver**, <https://dbeaver.com/docs/dbeaver>, ultimo accesso Ottobre 2023.

## Acronimi e significati

- App: Applicazione
- BI: Business Intelligence
- Dashboard: interfaccia di reportistica
- ETL: Extraction, Transformation, Load
- ELT: Extraction, Load, Transformation
- IT: Information Technology
- QMC: Qlik Management Console
- SQL: Structured Query Language
- MS: Microsoft
- CAAB: azienda cliente
- OLAP: On-Line Analytical Processing
- CSV: Comma-separated values
- JSON: JavaScript Object Notation
- DB: Database
- DW: Datawarehouse
- LoB: Line of business
- SEE: Spazio Economico Europeo
- ERP SAP: sistema Enterprise Resource Planning prodotto dall'azienda SAP AG
- CRM: Customer Relationship Management
- HQ: Headquarter
- TWS: IBM Tivoli Workload Scheduler
- YAML: Ain't Markup Language
- SSH: Secure shell
- IL: Integration Layer
- ODS: Operational Data Storage
- WL: Work layer
- DM: Data Mart
- QVD: QlikView Data
- KPI: Key performance indicator
- DYN: Microsoft Dynamics

## Figure

Figura 1 - Interfaccia del Qlik Sense Hub.....	13
Figura 2 - Interfaccia della Qlik Sense Management Console.....	14
Figura 3 - Rappresentazione dell'esecuzione parallela di Greenplum.....	16
Figura 4 - Interfaccia della schermata principale di DBeaver.....	18
Figura 5 - Rappresentazione dei rapporti fra le aziende coinvolte nel progetto .....	23
Figura 6 - Rappresentazione estremamente semplificata dello schema dati.....	27
Figura 7 - Rappresentazione delle interazioni tra il server e il DW.....	30
Figura 8 - Modello relazionale del framework ELT.....	31
Figura 9 – Componenti di una shell.....	32
Figura 10 - Rappresentazione del processo multi-chain interno ad una shell.....	33
Figura 11 - Rappresentazione semplificata del processo ELT in un DW Greenplum .....	34
Figura 12 - Sequenza di avvio delle shell ELT prima del progetto.....	35
Figura 13 - Architettura del datawarehouse prima del progetto.....	35
Figura 14 – Differenza grafica tra i due principali modelli dati possibili.....	38
Figura 15 - Rappresentazione a blocchi della CRMM.....	38
Figura 16 - Rappresentazione riassuntiva delle tre fasi ETL di Qlik Sense .....	40
Figura 17 - Architettura dati di Qlik Sense .....	40
Figura 18 - Sequenza di avvio delle shell con dettaglio sul file di controllo.....	41
Figura 19 - Dettaglio ETL QLIK Sense per le due aree coinvolte .....	42
Figura 20 - Dettaglio del modulo CRM prima delle modifiche.....	45
Figura 21 - Differenze tra i flussi dati DYN e SFDC .....	48
Figura 22 - Dettaglio delle tabelle coinvolte .....	49
Figura 23 - Analisi delle formule di ogni KPI.....	50
Figura 24 - Analisi della provenienza del dato .....	51
Figura 25 - KPI non compatibili con SFDC .....	51
Figura 26 – Dettaglio dei campi appartenenti ai KPI non compatibili.....	52
Figura 27 - KPI non impattati dalle modifiche.....	52
Figura 28 - Esempio di mappatura .....	53
Figura 29 - Analisi delle relazioni tra i campi DYN e SFDC.....	54
Figura 30 - Collegamento tra i campi di backend e frontend.....	54
Figura 31 - Specifica della tipologia del campo.....	55
Figura 32 - Tabella di transcodifica utilizzata tra DYN e SFDC per lo stato del caso .....	56

Figura 33 - Tabella di transcodifica utilizzata tra DYN e SFDC per il tipo di transazione .....	57
Figura 34 - Analisi delle Inbox .....	58
Figura 35 - Architettura del datawarehouse dopo il progetto .....	60
Figura 36 - Sequenza di avvio delle shell dopo il progetto.....	60
Figura 37 - Analisi dei singoli fogli delle dashboard .....	62
Figura 38 - Analisi degli oggetti all'interno di ogni dashboard.....	62
Figura 39 - Cronoprogramma del progetto .....	66