

POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Ingegneria Gestionale

Tesi di Laurea Magistrale

Strumenti di Data Visualization per la
Business Intelligence



Relatrice

Prof.ssa Tania Cerquitelli

Candidato

Fabio Tassarollo

Anno Accademico 2020/2021

Abstract

Tutti i più moderni sistemi di gestione delle organizzazioni e delle imprese sono associati a degli applicativi in grado di fornire, se opportunamente utilizzati, rapporti della performance del business basati sui dati. Questi rapporti, finalizzati a supportare i processi decisionali di medio e lungo periodo, consistono in diversi cruscotti (o dashboard) categorizzati per area aziendale nei quali la rappresentazione grafica dei dati assume un ruolo fondamentale. Nel presente documento verranno affrontate sul lato pratico le tecniche di progettazione e le funzionalità applicative di questi strumenti di analisi, con riferimenti all'attività svolta nell'esperienza di tirocinio intrapresa presso Cluster Reply, società italiana che si occupa di consulenza informatica aziendale su prodotti Microsoft. Particolare attenzione verrà dedicata a Power BI, software della Microsoft adibito a queste funzionalità. A scopo dimostrativo verranno portati alla luce alcuni casi reali di applicazione del programma nella creazione di cruscotti per conto di aziende clienti. Infine, verrà analizzato il mercato degli strumenti di Business Intelligence, le sue tendenze e gli sviluppi attesi di questo settore in ascesa.

Indice

1	Introduzione	5
1.1	Definizione di Business Intelligence	6
1.2	Definizione di Data Visualization	6
	Contesto.....	7
2	Il valore dell'analisi dati	7
2.1	Digital Transformation.....	7
2.2	Aziende informatiche.....	9
3	Cluster Reply	10
4	Dynamics 365 Finance and Operations.....	11
4.1	Mercato degli ERP.....	12
4.2	Transizione a Dynamics 365.....	13
5	Clienti	14
5.1	Azienda A	14
5.2	Azienda B	15
	Progettazione della BI.....	16
6	Architettura database	16
6.1	Sistemi OLTP	16
6.1.1	Progettazione fisica	17
6.2	Sistemi OLAP	18
6.2.1	Progettazione fisica	19
7	Cubi OLAP	20
7.1	Identificazione dei requisiti e progettazione DW.....	20
7.2	Studio e localizzazione dei dati a disposizione	21
7.3	Generazione del cubo	22
7.4	Il futuro dei cubi OLAP	24
7.5	Esempio di generazione di un cubo.....	25
8	Dashboard.....	26
8.1	Generazione della dashboard.....	26
8.2	Interazione con la dashboard.....	27

8.3	Esempi di dashboard.....	29
8.3.1	Ritardo consegne Azienda A	29
8.3.2	Analisi OTIF Azienda A	30
8.3.3	Conto economico Azienda B.....	31
9	Power BI.....	33
9.1	Importazione dei dati	33
9.2	Trasformazione della struttura dei dati	35
9.3	Esportazione delle dashboard	36
	Sviluppi futuri.....	37
10	Augmented Analytics	37
11	Il mercato della BI.....	38
	Conclusioni.....	41
	Riferimenti	42

1 Introduzione

È ormai noto a tutti l'inarrestabile impatto della tecnologia sull'umanità, i dati prodotti e la velocità di calcolo non hanno mai smesso di crescere, e a ritmi sempre più serrati. L'incremento della produzione di dati negli anni è riscontrabile nel grafico 1, diagramma a barre di una stima di International Data Corporation (IDC) sulla *global datasphere*, il volume di dati generati nel mondo in zettabytes (ZB, 10^{21} bytes). Per escludere l'effetto dell'aumento di dati dovuto alla crescita della popolazione, è stata calcolata la produzione annuale di dati per capita (grafico 2) e ne risulta che il carattere esponenziale viene mantenuto. Dividendo il volume di dati per la popolazione, infatti, il tasso di crescita medio annuo per il periodo considerato si abbassa di un solo punto percentuale, risultando del 27% circa. Di norma, avendo a disposizione più dati, i risultati ricercati sono ottenuti con margini di errore inferiori, ma, talvolta, proprio a causa della mole di informazioni a disposizione, la capacità del cervello umano di discernere la verità può essere messa a dura prova. Ogni giorno, nella vita privata, siamo bombardati da enormi quantità di informazioni e, analogamente nelle aziende, i sistemi informativi supportando a tutto tondo anche i processi più complessi, generano molti dati altrettanto complessi. È qui che interviene la rappresentazione grafica dei dati, o data visualization, uno strumento di analisi che se preceduto da un corretto processamento, è in grado di far trasparire correttamente la verità, con il potere di farlo in modo veloce, chiaro e impattante. Questi strumenti attualmente si ritrovano per lo più in applicazioni di Business Intelligence (BI), un ramo dell'informatica aziendale il cui scopo è trasformare i dati in conoscenza attraverso un processo di sintesi. Si tratta di individuare ciò che è più importante, riconoscere pattern, trovare correlazioni e valutare parametri per comparazione. La presente tesi è basata su progetti di BI svolti in un tirocinio presso l'azienda Cluster Reply.

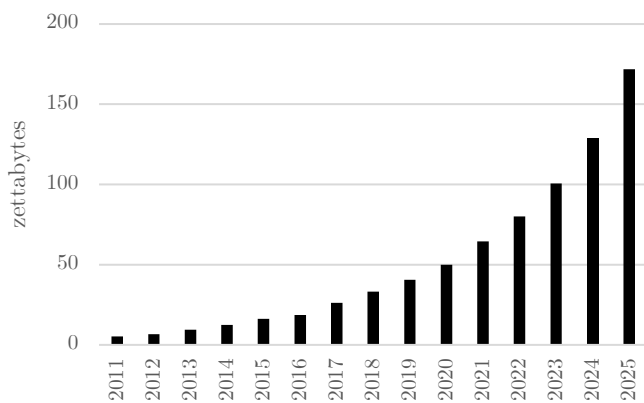


Grafico 1, Volume di dati creati, estratti e replicati globalmente dal 2011 al 2025, ZB [IDC, 2018].

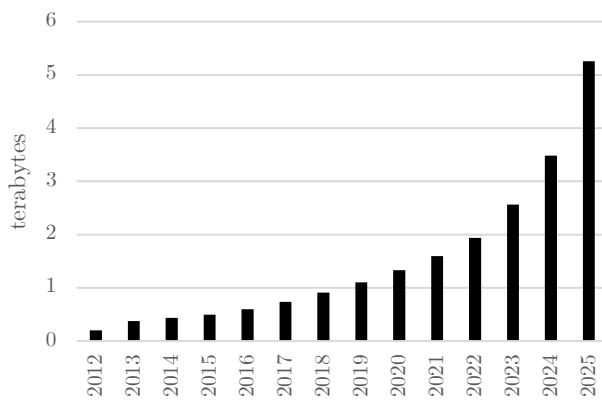


Grafico 2, Dati prodotti annualmente per capita, TB. I dati sulla popolazione provengono da Our World in Data.

1.1 Definizione di Business Intelligence

Vengono qui di seguito riportate due definizioni:

“La Business Intelligence comprende tutte le strategie, i processi, le applicazioni, i dati, i prodotti, le tecnologie e le architetture usate per permettere la raccolta, l'analisi, la presentazione e la distribuzione delle informazioni del business” [Nedim Dedić, Clare Stanier, 2016].

“La Business Intelligence (BI) è conoscenza guadagnata attraverso l'analisi delle informazioni del business. I mezzi della BI sono query, reporting, sistemi OLAP, data mining, analytics, sistemi end-user per analisi ad hoc e le dashboard per il monitoraggio della performace” [Howard Dresner, 2007].

Le dashboard per il monitoraggio sono uno strumento di Data Visualization.

1.2 Definizione di Data Visualization

La Data Visualization è un campo interdisciplinare che riguarda la rappresentazione grafica dei dati. Da un punto di vista accademico, la rappresentazione può essere considerata come una mappatura dei dati a degli oggetti grafici. La mappatura determina come gli attributi di questi elementi varino secondo i dati. Tra gli strumenti di analisi la Data Visualization è sicuramente il più accessibile come livello di complessità di comprensione.

“L’obiettivo principale della Data Visualization è permettere all’utente una comprensione qualitativa e semplice dei contenuti. Essa consiste nella trasformazione di oggetti, numeri e concetti in una forma che possa essere facilmente interpretata dall’occhio umano” [Kumar SM, Meena Belwal, 2017].

Questa pratica trova i suoi albori alla fine del XII secolo, guadagna forte interesse nel mondo accademico con la nascita dei computer e successivamente inizia ad essere sfruttata nei business per scopi di analisi dopo gli anni Settanta [Daniel Filonik, 2012]. Un'altra applicazione degna di nota si ha nelle campagne informative, in quanto i giusti oggetti grafici possono essere molto convincenti oltre a trasmettere velocemente e in modo chiaro un messaggio.

Contesto

In questa prima parte della tesi verrà analizzato il contesto nel quale si è operato: il campo dell’analisi dati, l’azienda del tirocinio, il software e le aziende da cui provengono i dati usati per i progetti di BI.

2 Il valore dell’analisi dati

«Data is the new oil» tradotto in italiano “i dati sono il nuovo petrolio” è una frase di forte impatto coniata nel 2006 da Clive Humby, matematico e imprenditore britannico [The Guardian, 2013]. La metafora è calzante perché come il petrolio, i dati estratti possono avere un alto valore potenziale, la conoscenza, ma per guadagnarla è necessario un processo di “raffinatura”, l’analisi. Inoltre, si usa questo paragone anche per descrivere un possibile scenario futuro, ormai sempre più verosimile, nel quale i dati giocheranno un ruolo analogo al petrolio, come bene driver dell’economia, fonte di potere e di competizione tra gli stati. [OECD Competition Division, 2016; Bhaskar Chakravorti, Ajay Bhalla, and Ravi Shankar Chaturvedi, 2019]. Il vantaggio competitivo sarà basato sull’accessibilità ai dati e sulle tecniche di analisi.

In questa sezione vengono analizzati i fenomeni economici e sociali che hanno causato l’incremento della produzione dei dati argomentato nell’introduzione. Il valore aggiunto apportato dagli strumenti di analisi statistiche dei dati è strettamente legato a questo incremento, in particolare si ha una maggiore accuratezza descrittiva delle entità analizzate, anche se un maggiore livello di complessità nel processo di analisi.

2.1 Digital Transformation

La Digital Transformation è un fenomeno che riguarda tutti i cambiamenti verso la creazione di sistemi tecnologici di supporto ad aspetti della società umana, o aggiornamenti verso sistemi informatici più avanzati. Sempre più organizzazioni sono orientate verso la dematerializzazione dei documenti cartacei e l’installazione di tecnologie che possano supportare, se non rinnovare, i processi che le caratterizzano. Componente fondamentale di questa tendenza è il diffondersi dei sistemi informativi nei business aziendali. Attualmente infatti, tutte le grandi aziende, che siano di produzione o di servizi, fanno uso di software gestionale. Chiaramente maggiore è il livello di integrazione delle

parti operative (HR, contabilità, logistica, etc.) maggiore è il valore aggiunto apportato da questi servizi. Allo stesso modo, maggiore è la copertura del sistema rispetto alle diverse divisioni geografiche dell'impresa migliori saranno i risultati ottenuti. Questo perché usare lo stesso sistema standardizza i processi con una conseguente maggiore efficienza, esprimibile in minori tempi e meno frequenti errori, soprattutto nelle industrie manifatturiere. Inoltre, in questa prospettiva i dati sono raccolti con lo stesso criterio ovunque e ciò permette di effettuare analisi globali che confrontino correttamente qualsivoglia divisione, macchina di linea, reparto o prodotto all'interno della realtà aziendale. Difatti, le funzionalità di Business Intelligence sono particolarmente giovate da un sistema unico che si estende in tutti gli angoli dell'organizzazione. Ciò è permesso in particolar modo dall'Internet of Things (IoT), ovvero “la crescente gamma di dispositivi connessi a Internet, misuratori e gadget incorporati in vari oggetti per consentire loro di inviare e ricevere dati” [OECD Competition Committee, 2016]. Questi strumenti permettono il controllo delle funzionalità e automatizzano la reportistica delle macchine industriali a cui sono integrati, o rendono “smart” i prodotti rivolti ai consumatori.

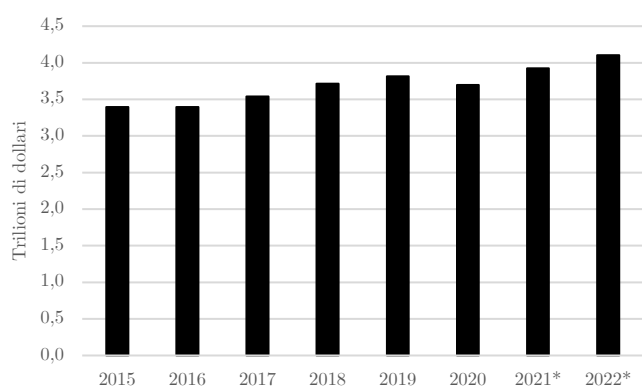


Grafico 3, Investimenti globali in Information Technology (IT) dal 2015 al 2022 (*previsione) [Gartner, 2021]

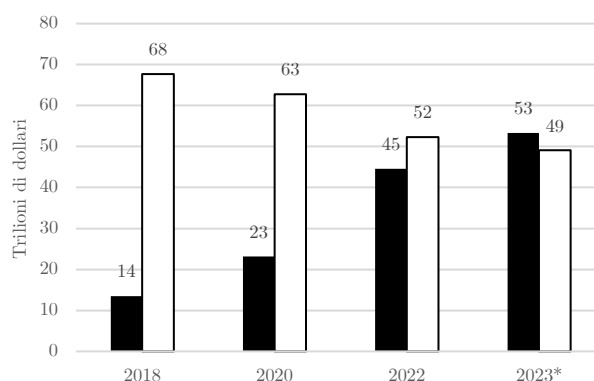


Grafico 4, PIL nominale prodotto globalmente da aziende trasformate digitalmente, in nero, e le restanti, in bianco (*previsione) [IDC, 2020]

Questi e altri fattori hanno motivato gli investimenti nel settore (grafico 3). Secondo una ricerca di IDC, le aziende che sfrutteranno il loro massimo potenziale tecnologico con i sistemi ad oggi più avanzati, produrranno più della metà del PIL nel 2023 (grafico 4). All'interno delle organizzazioni questi sistemi generano big data, grandi quantità di dati strutturate in modo complesso, formate da dati transazionali, frutto dell'interazione degli operatori con il sistema, e da dati provenienti da processi autonomi delle macchine, grazie all'IoT. Chiaramente sistemi con una maggiore copertura delle operazioni saranno anche soggetti a maggiori costi di installazione, manutenzione e switching cost, per tanto bisogna trovare il giusto compromesso tenendo in considerazione le diverse opzioni che il mercato

dell'informatica aziendale ha da offrire rispetto alle reali necessità del proprio business. Non sempre un maggiore livello di digitalizzazione comporta un miglioramento delle performance dei processi, ma ad ogni modo il trend è chiaro.

2.2 Aziende informatiche

Oltre alle organizzazioni i cui processi sono stati digitalizzati, anche le aziende con un business fortemente orientato verso il software e i dati, da considerarsi “digitali” fin dal principio, ampliano ulteriormente l'insieme delle società che concorrono all'aumento esponenziale dei dati prodotti nel mondo. E, anch'esse, usufruendo dell'analisi dei dati, ottengono importanti vantaggi. In particolare, tra queste troviamo le aziende che hanno portato alla diffusione della tecnologia nelle masse, un'altra enorme fonte di produzione di dati. Considerando le molteplici accuse di violazione delle normative sulla privacy in vigore da parte di diverse organizzazioni governative verso alcuni di questi business (si tratta di colossi tecnologici come Google e Facebook), si può ipotizzare che anche la porzione di dati privati, pur se non completamente comprovato, sono processate e analizzate per scopi economici a discapito della privacy individuale. A tal proposito si cita il documento rilasciato dall'Organizzazione per la cooperazione e lo sviluppo economico (OECD) “Big Data: bringing competition policy to the digital era” nel quale si portano alla luce le nuove problematiche derivanti dall'entrata nell'era digitale, tema tutt'oggi oggetto di dibattiti.

In conclusione, i fenomeni che hanno generato e continuano a generare i dati possono essere così divisi: trasformazione digitale delle imprese; origine e crescita di business basati su tecnologie informatiche; diffusione dell'accesso a Internet e ai servizi social nelle masse. Cambiando prospettiva, un altro modo di identificare le entità dietro la generazione dei dati è distinguere l'hardware dove tali dati vengono prodotti, o duplicati, come in un rapporto di IDC dove si distinguono: *the core* (i database tradizionali e su cloud), *the edge* (infrastrutture aziendali come sedi secondarie e gateways) e gli *endpoints* (smartphone, computer, IoT, etc.).

“Parallelamente all'espansione del volume di dati, la trasformazione di questi in conoscenza per migliorare i processi decisionali è un settore che presenta nuove sfide. Si è evoluta una filiera completamente nuova di aziende che forniscono servizi di acquisizione di dati (nuove fonti), data storage, warehousing, modellamento e analisi dei dati, e data visualization” [United Nations, 2019].

3 Cluster Reply

Cluster Reply è una società a responsabilità limitata facente parte del gruppo Reply, organizzazione fondata nel 1996 e costituita da un modello a rete di aziende altamente specializzate, che affianca i principali gruppi industriali europei appartenenti ai settori Telco & Media, Industria e Servizi, Banche e Assicurazioni e Pubblica Amministrazione nella definizione e nello sviluppo di modelli di business abilitati dai nuovi paradigmi del Big Data, Cloud Computing, Digital Media e Internet of Things. I servizi di Reply includono: Consulenza, System Integration e Digital Services [Reply, Profilo aziendale 2020]. Cluster Reply è una delle più grandi del gruppo ed è specializzata in servizi di integrazione di sistemi su tecnologie Microsoft. Cluster Reply è da più di 20 anni partner di Microsoft e quest'ultima ha riconosciuto all'azienda nell'anno 2019 un premio come migliore partner italiano. L'azienda è focalizzata sull'innovazione e offre soluzioni on-premise ed in cloud, nelle aree di Modern Workplace, Business Applications, Applications ed Infrastructure, Data ed AI [Reply, sito ufficiale Cluster Reply]. Il candidato ha intrapreso un'esperienza di tirocinio presso questa società, durante la quale è stata presa in carico la progettazione di servizi di Business Intelligence da integrare con il software ERP per il quale Cluster Reply offre la propria consulenza, Dynamics 365 Finance & Operations.



Figura 1, logo di Cluster Reply

Le aziende che compongono il gruppo Reply collaborano fra di loro su interi progetti condividendo i clienti, ma talvolta sono anche in competizione. Si tratta di un ambiente mediamente giovanile e molto dinamico, incentrato sulla ricerca e rivolto alla crescita. Questo modello è allineato con le esigenze del settore tecnologico ed è uno dei punti di forza dell'azienda. Si crea così un ambiente che per quanto nel suo complesso sia paragonabile a colossi tecnologici ben avviati, nel piccolo ha alcune delle caratteristiche vantaggiose tipiche di aziende più piccole e appena avviate, garantendo così maggiore flessibilità e allo stesso tempo bassi rischi.

Come riportato sul sito della Borsa Italiana, il Gruppo Reply, con i suoi 8157 dipendenti, opera a livello mondiale e ha chiuso l'esercizio 2019 con un fatturato consolidato di 1.182,5 milioni di euro, in crescita del 14,2% rispetto all'esercizio 2018 e con un punto percentuale

in più rispetto alla crescita annua media degli ultimi 15 anni. Il margine operativo ammonta invece a 191,3 milioni di euro. La società è quotata dal 2000 presso la Borsa valori di Milano dove è presente negli indici FTSE Italia Mid Cap e FTSE Italia STAR, segmento della borsa italiana nato per valorizzare agli occhi degli investitori le PMI con caratteristiche di eccellenza. Di seguito è riportato un sunto del bilancio dell'azienda nell'anno 2019.

Attivo	Passivo
Non liquido: 1.066.970	Patrimonio netto: 587.061
Liquido: 240.943	Altre passività: 720.853
1.307.913	1.307.913

[Reply relazione finanziaria annuale - valori in migliaia di euro, 2019]

4 Dynamics 365 Finance and Operations

Il contesto nel quale si è operato nella progettazione di servizi BI è quello della consulenza per la Digital Transformation di aziende attraverso l'installazione di sistemi software per la programmazione delle operazioni.

Enterprise Resource Planning (tradotto "pianificazione delle risorse d'impresa"), abbreviato in ERP, è un software di gestione che integra i processi di business rilevanti di un'azienda e le funzioni aziendali. Customer Relationship Management, abbreviato CRM, invece è un sistema di supporto delle operazioni di interazione con i clienti. La divisione nel quale si è operato nel tirocinio è quella dedicata al principale software gestionale di Microsoft, Dynamics 365 Finance & Operations (abbreviato D365FO). Designato alla programmazione e al controllo delle operazioni aziendali, D365FO consiste in un ERP integrato con alcune funzionalità tipiche dei CRM, una cloud aziendale che riunisce dati, operazioni e clienti studiata per imprese di medie o grandi dimensioni. Le funzionalità possono essere divise come si può dedurre dal nome in finanziarie, quindi budgeting, project management e contabilità, e in operazionali, quindi supply chain management, gestione del magazzino, programmazione della produzione e logistica.



Figura 2, logo di Microsoft Dynamics 365

Il programma viene fatto girare su degli ambienti remoti a cui si può accedere tramite browser, immettendo l'URL specifica e le credenziali dell'utente. Questa è la formula più diffusa di erogazione del servizio nel mondo dei software ERP ed è denominata SaaS, Software as a Service. Spesso, data la numerosità delle funzioni di questi sistemi e la loro complessità, la maggior parte dei dipendenti dell'azienda, come user di sistema, conoscono e sanno usare le funzionalità solo della propria area di lavoro. Come in tutte le piattaforme di gestione, inoltre, ogni utente ha limitati permessi di uso del sistema, a eccezione dell'admin che invece può accedere a tutte le proprietà. Gli ambienti si dividono in due tipologie: di testing e di produzione. Gli ambienti di testing sono usati dagli sviluppatori per testare le modifiche al software e dai consulenti per mostrare al cliente le metodologie di utilizzo. Si tratta di ambienti studiati per agevolare la transizione al nuovo sistema, adoperati per lo più da aziende di consulenza come Reply. Negli ambienti di produzione invece la piattaforma è usata dall'azienda che ha comprato il sistema dopo il completamento dell'installazione, allo scopo per cui questi sistemi sono stati progettati: uno storage dei dati raccolti in fase di produzione e una piattaforma alla quale accedere per il controllo e per la gestione delle operazioni aziendali.

4.1 Mercato degli ERP

È difficile calcolare delle quote di mercato per questa categoria di software perché ognuna delle case produttrici può offrire o non offrire determinate funzionalità, rendendo difficile la classificazione di quali, di tutto il panorama dei software aziendali, rientrino nell'insieme degli ERP. Secondo uno studio di Apps Run The World, nel 2019, i 10 più grandi fornitori di software ERP a gestione finanziaria o operativa possiedono complessivamente circa 30% delle quote di mercato. In particolare sembra che questi leader di mercato siano fornitori delle aziende più grandi, mentre la restante parte del mercato, molto frammentata, è provider di business di dimensioni medie o piccole. Prima in classifica per quote di mercato è l'azienda tedesca SAP, pioniere nel mondo di questi sistemi. Un altro nome molto noto del panorama degli ERP è Oracle, seconda solo a Microsoft per capitalizzazione di mercato tra le società di software e famosa per la sua tecnologia per i

database. Secondo quanto rilevato dalla ricerca citata, Microsoft invece si classifica in sesta posizione nel mercato degli ERP. Il primato di SAP è dimostrato anche dai risultati di una ricerca di Gartner sul fatturato delle società proveniente da software di supply chain management (grafico 5).

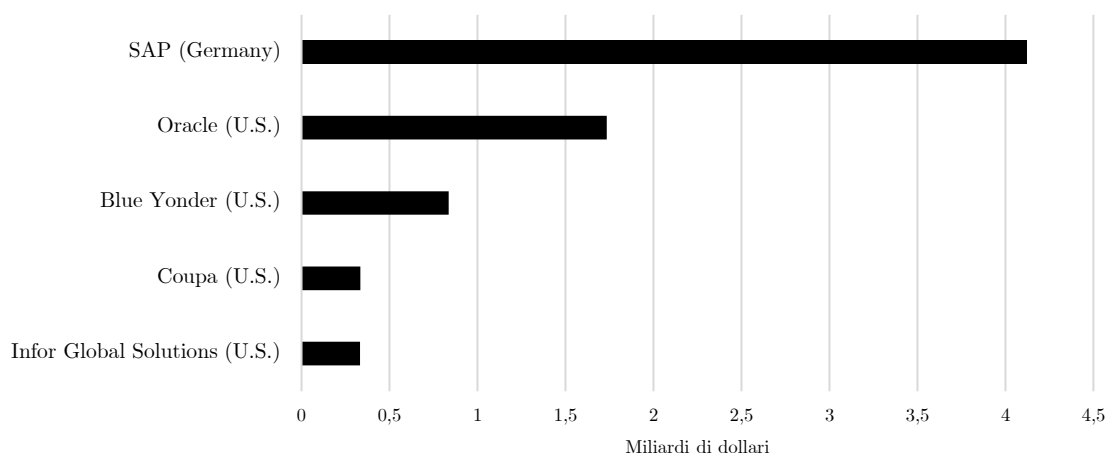


Grafico 5, Maggiori aziende al mondo di supply chain management software per fatturato derivante dal relativo segmento in miliardi di dollari nell'anno 2019 [Gartner, 2020]

4.2 Transizione a Dynamics 365

Il processo usato da Cluster Reply per portare a termine l'installazione e integrazione di Dynamics 365 nelle aziende clienti è articolato in diverse fasi. Il tempo necessario per la transizione al nuovo programma dipende dalle dimensioni e dalla complessità dell'azienda. Nel caso della divisione europea di una delle aziende per cui si ha lavorato, che ha una sede strategica e 6 filiali, sono stati stimati 20 mesi per il completamento del progetto.

La prima delle fasi è quella di analisi e diagnosi. In questa fase Cluster viene a conoscenza del business del nuovo cliente attraverso un continuo dialogo con i suoi dipendenti. Comprendere il funzionamento dei principali processi dell'azienda è di fondamentale importanza per un'integrazione efficiente con il nuovo software. Il risultato è la generazione di un documento di fogli di calcolo in cui si identificano i requisiti, suddivisi in gerarchie, e la capacità del nuovo software di soddisfarli. Si tratta dell'analisi dei fit&gap, in cui ogni sottoprocesso viene classificato come fit, se il nuovo supporto informatico può rispondere alle esigenze degli operatori, o gap, se la casistica in questione non è contemplata dal nuovo sistema. A ognuna di queste istanze viene associato un punteggio che ne esprime la priorità e in caso dei gap, una possibile soluzione, che può consistere nel rimodellare i processi dell'azienda in modo tale da adattarli al nuovo sistema,

oppure adattare il sistema ai processi dell'azienda attraverso un'operazione di customizzazione del software. Partecipando a questa fase è stato possibile farsi una prima idea dei requisiti e della struttura dei dati per la progettazione degli strumenti di analisi.

Durante le videoconferenze preparatorie, inizialmente è il cliente che condivide lo schermo al fine di mostrare le operazioni tipiche di ogni processo aziendale dall'interfaccia del sistema di gestione che verrà sostituito. Successivamente è il turno di Cluster, che facendo uso dell'ambiente di test replica gli stessi processi su D365FO nel modo più simile possibile e mostra nuove funzionalità che potrebbero interessare all'azienda in base a quanto visto. Nell'ambiente di test risiedono dati generati da Microsoft a scopo esemplificativo. Tutte le maschere sono quindi popolate da righe fatte di dati non reali, predisposti appositamente per simulare nel modo più verosimile possibile gli usi del programma.

Una criticità della fase di diagnosi è determinare quali saranno i reali costi e tempi necessari per il passaggio a D365FO rispetto a quanto stimato e promesso al cliente. In fase embrionale non è facile, per esempio, prevedere il tempo necessario agli sviluppatori per adattare il software alle necessità dell'azienda cliente e solo in fasi successive potrebbero presentarsi tematiche prima sconosciute che implicano costi non previsti. Dopo analisi e diagnosi, a cascata seguono poi le fasi di design, sviluppo e distribuzione, in cui avviene la progettazione e installazione del nuovo software.

5 Clienti

Nello svolgimento del tirocinio il candidato ha collaborato nelle attività di consulenza per due clienti che verranno brevemente analizzati in questa sezione. Per questioni di riservatezza non si potranno utilizzare riferimenti diretti alle aziende, i cui reali nomi sono stati sostituiti con Azienda A e Azienda B.

5.1 Azienda A

Azienda A è un'azienda manifatturiera giapponese specializzata nella produzione di pistole a spruzzo per la verniciatura che opera a livello globale. L'azienda ha scelto la consulenza di Cluster Reply per cambiare software ERP nella zona europea, passando da Solgenia SE7EN a Dynamics 365 Finance and Operations. Si tratta, quindi, di un nuovo cliente e la buona riuscita della commessa è importante per Cluster. Il CEO della società ha mostrato particolare interesse per le funzionalità di BI offerte da Microsoft e per questo è

stato importante evidenziare queste funzioni nella presentazione del programma. La sede di produzione dell'azienda è situata in Italia. In Germania, Polonia, Francia, Svezia e Spagna risiedono invece le filiali commerciali, che ordinano i prodotti facendo richiesta alla sede italiana per poi rivenderli nei loro territori di pertinenza, operazione di supply chain determinante e da eseguire con il software gestionale. Data l'ampiezza del catalogo di prodotti offerti e la complessità manifatturiera del prodotto, l'intensità informativa che deriva dai processi dell'azienda è relativamente alta.

Come detto, nelle prime fasi di consulenza hanno luogo una serie di incontri organizzati affinché l'azienda cliente possa esporre i suoi processi e metodi di lavoro. Nel caso di Azienda A si è trattato di più di 200 ore di meeting divisi tra Accounting and Finance e Supply Chain. Una figura chiave per la buona riuscita del progetto è stata quella dell'IT manager dell'azienda, che, come responsabile dei supporti informatici, ha partecipato a tutte le conferenze ed è stato il più coinvolto nella transizione al nuovo sistema.

In base a quanto visto dai report mostrati dal cliente, tra i diversi parametri utili ad analizzare la performance del business, è sembrato che l'azienda fosse interessata a osservare i ritardi nei tempi di produzione e consegna. Per questa ragione il candidato si è occupato della creazione di cruscotti di controllo di esempio per monitorare questi fattori che verranno analizzati in seguito.

5.2 Azienda B

Azienda B è una società italiana specializzata nel lavaggio della lana tosata, detta sucido. Aziende produttrici di capi di abbigliamento si rivolgono ad Azienda B per il servizio di lavaggio, procurando all'azienda gli ammassi di lana detti "balle" per avere indietro i lotti di lana pronti per la tessitura. Le linee di lavaggio dell'azienda sono caratterizzate da macchinari di vecchia fattura, che non dispongono di un computer di bordo. Al fine di ottenere maggiore controllo sull'impianto, su queste macchine sono stati installati dispositivi IoT per rilevare il mix di composti chimici usati nel processo, detto formula, e alcuni parametri che indicano la qualità della lana a fine lavaggio. Inoltre, l'azienda tiene traccia degli ordini nuovi, aperti e conclusi, con tutte le informazioni annesse alle balle di lana. Tutti questi dati, insieme alla contabilità, sono raccolti nel sistema informativo.

L'azienda nel periodo di svolgimento del tirocinio aveva già ricevuto la consulenza di Cluster per passare a un nuovo software aziendale. Le attività svolte hanno riguardato la consulenza per la Business Intelligence: la creazione di cubi di dati per predisporre degli strumenti di analisi sul processo di lavaggio e la conversione di alcuni report Excel della contabilità di proprietà dall'azienda in dashboard di Power BI.

Progettazione della BI

In questa seconda parte vengono analizzate le tecniche per la progettazione degli strumenti di Data Visualization per la BI. L'obiettivo principale è creare degli strumenti in grado di fornire, in modo efficace e in tempi brevi, le informazioni ricercate dagli utenti. A tale scopo è necessario studiare la pipeline di analisi a partire dall'architettura dal quale i dati provengono, primo fattore a determinare i tempi di estrazione, fino ad arrivare alle logiche di funzionamento delle dashboard, l'interfaccia dalla quale l'utente potrà condurre le proprie analisi.

6 Architettura database

I sistemi informativi aziendali per funzionare devono appoggiarsi a un database, sede di ogni dato immesso nel sistema e nel quale sono raccolte tutte le informazioni processate. Per l'analisi di tali informazioni, viene progettato un secondo spazio nel quale i dati vengono trasferiti e aggiornati periodicamente. Si profila quindi un'architettura a due livelli per consentire da una parte le elaborazioni di tipo operativo e dall'altra le elaborazioni di tipo analitico, analizzate in quest'ordine nelle sezioni seguenti.

6.1 Sistemi OLTP

I sistemi OLTP (Online Transaction Processing) rappresentano un supporto informatico delle attività core del business. Come riportato nella documentazione di Microsoft sull'argomento, essi registrano le interazioni aziendali man mano che si verificano nell'ambito delle operazioni giornaliere eseguite all'interno dell'organizzazione e supportano la creazione di query su questi dati per l'inserimento di inferenze. Queste interazioni generano dei dati transazionali, che riguardano in genere transazioni aziendali, come pagamenti ricevuti dai clienti, movimentazioni di prodotti dell'inventario, elaborazioni di ordini o fornitura di servizi. Se per esempio il sistema di gestione fa uso di una tecnologia IoT che dopo aver misurato un certo parametro registra l'informazione, avverrà un aggiornamento di una specifica tabella grazie ai sistemi OLTP. Lo stesso avviene per le informazioni di un nuovo dipendente registrato a sistema da un responsabile delle risorse umane, o per la programmazione della produzione di un lotto di prodotti. Vengono così aggiunte o modificate le istanze del database transazionale. L'esecuzione di queste interrogazioni di tipo operativo implica l'uso di query di lettura e modifica (insert, update and delete) con un basso livello di complessità e un tempo di esecuzione

relativamente basso. Tuttavia, il carico di interazioni con il database complessivamente è molto alto e i tempi di risposta necessari sono bassi. In aggiunta, le applicazioni OLTP mirano ad una garanzia di integrità e sicurezza delle transazioni. Per questa ragione, i più moderni sistemi di gestione impediscono aggiornamenti del database non provenienti dal sistema stesso. Una transazione registrata, quindi, non deve poter essere modificata tramite un accesso diretto al database da un programma terzo. Per andare più nel dettaglio, sono state identificate quattro proprietà logiche raggruppate nell'acronimo "ACID" [Hoboken, Wiley, 2003] che le transazioni devono avere affinché i meccanismi che le implementano operino in modo corretto:

- atomicità: il processo deve essere suddivisibile in un numero finito di unità indivisibili, chiamate *transazioni*. L'esecuzione di una transazione, perciò, deve essere per definizione o totale o nulla, e non sono ammesse esecuzioni parziali.
- coerenza: il database rispetta i vincoli di integrità, sia a inizio che a fine transazione. Non devono verificarsi contraddizioni tra i dati archiviati nel DB.
- isolamento: ogni transazione deve essere eseguita in modo isolato e indipendente dalle altre transazioni, l'eventuale fallimento di una transazione non deve interferire con le altre transazioni in esecuzione.
- durabilità: detta anche persistenza, si riferisce al fatto che una volta che una transazione abbia richiesto un *commit work*, i cambiamenti apportati non dovranno essere più persi. Per evitare che nel lasso di tempo fra il momento in cui la base di dati si impegna a scrivere le modifiche e quello in cui li scrive effettivamente si verifichino perdite di dati dovuti a malfunzionamenti, vengono tenuti dei registri di log dove sono annotate tutte le operazioni sul DB.

Nell'esperienza di tirocinio i sistemi OLTP sono stati osservati nell'uso dell'ERP D365FO, che si appoggia a un database transazionale denominato "AxDB".

6.1.1 Progettazione fisica

L'archiviazione fisica dei dati più adatta ai sistemi OLTP è quella tradizionale, usata già nelle prime versioni di SQL Server, chiamata rowstore-index. Con questo tipo di archiviazione il recupero dei dati richiesti è velocizzato perché la localizzazione avviene per indice di riga, evitando la scansione di ogni record all'interno delle tabelle. I dati, quindi, sono organizzati logicamente in righe e colonne ma sono immagazzinati fisicamente in righe. Questa architettura velocizza le query di ricerca di un valore particolare o di porzioni ridotte di dati [Logical Read, Choosing Between Rowstore Indexes and Columnstore Indexes].

Il classico metodo di localizzazione dei file nei sistemi OLTP è il B-Tree o una sua variante, una struttura di indicizzazione non-clustered dei dati ad albero bilanciato che permette di compiere operazioni di inserimento, cancellazione e ricerca in tempi ammortizzati logaritmicamente. Il bilanciamento sta nella lunghezza dei rami, che l'algoritmo mantiene il più uniforme possibile all'aggiungersi di nuovi record. Nelle operazioni di lettura, l'albero è percorso partendo dalla radice, passando per i nodi intermedi e giungendo in fine ai nodi foglia, che contengono il puntatore per accedere all'informazione. Il percorso è determinato dalla chiave di ricerca, input per la localizzazione del dato [Wikipedia, B-Tree].

6.2 Sistemi OLAP

I sistemi OLAP (Online Analytical Processing) invece riguardano operazioni che non hanno a che vedere con processi aziendali propriamente legati all'attività del business, bensì con operazioni rivolte alla valutazione dell'andamento del business e dei suoi risultati: l'analisi multidimensionale dei dati e il reporting mirati a supportare le decisioni aziendali.

I dati che si intende analizzare sono i dati transazionali precedentemente nominati. Come spiegato però, questi dati sono salvati in un database studiato per i sistemi OLTP, le cui elaborazioni differiscono notevolmente da quelle dei sistemi OLAP. Questa volta le interrogazioni sono di sola lettura, sono meno frequenti e riguardano enormi quantità di record. L'obiettivo è fornire delle risposte nei minori tempi possibili e l'esecuzione delle analisi direttamente nella base dati operativa non può che disattendere queste aspettative. Per questo la procedura prevede come primo passo la selezione e la copiatura dei dati di interesse dal database del sistema informativo a un secondo database progettato per le analisi multidimensionali e il data mining, chiamato data warehouse, che agevolerà le query tipiche dei sistemi OLAP e permetterà di esplorare facilmente la base dati nei diversi livelli gerarchici. Infatti, in esso i dati sono storicizzati e strutturati in aggregazioni. I dati analizzati saranno quindi dei dati derivati, mentre i dati transazionali rimangono intoccati, fattore ulteriormente positivo data la sensibilità del sistema in questo ambito. Inoltre, i dati transazionali possono essere inaffidabili o incompleti e nel passaggio di importazione al data warehouse attraverso un opportuno filtraggio si ovvia a questo problema migliorando la qualità dei dati da analizzare. Questo passaggio è detto ETL, Extract Transform Load: i dati, dopo essere stati estratti alla fonte (Extract), sono trasformati per rimuovere il rumore migliorandone la qualità (Transform), e successivamente caricati in una sorgente singola che possa a sua volta alimentare il data warehouse (Load).

Allo scopo di compiere delle analisi incrociate e complete, è bene introdurre dati aggiuntivi esterni all'organizzazione. Per tale ragione, nel data warehouse si importano anche dati

provenienti da terze parti, come ad esempio i risultati dei competitors o gli indici economici del settore. Questi dati possono essere Open Data, una categoria di informazioni pubblica e a disposizione di qualsiasi ente, o dati esterni acquisiti da agenzie specializzate nel reperimento di informazioni di mercato. Un'altra fonte esterna rispetto al database operativo potrebbe essere rappresentata da file privati creati dagli utenti in locale, come ad esempio i fogli Excel.

Siccome il database transazionale, data l'attività giornaliera dell'azienda, è aggiornato ogni secondo con nuove informazioni, affinché anche le analisi risultino altrettanto aggiornate, è necessario che dopo essere stato popolato una prima volta, il DW sia aggiornato periodicamente con una determinata frequenza. Talvolta è necessario avere i dati più recenti, in altri casi è sufficiente una frequenza giornaliera o settimanale di refresh degli applicativi di analisi. La velocità di aggiornamento è motivo di competizione fra i fornitori di servizi di analisi e Business Intelligence qualora siano necessarie delle analisi “quasi in tempo reale”.

6.2.1 Progettazione fisica

Per ottenere le migliori prestazioni in termini di velocità di risposta delle query di analisi serve studiare anche la più efficiente tecnica di gestione delle memorie. L'archiviazione dei dati nei sistemi OLAP più moderni è di tipo columnstore index, i dati in questo caso sono indicizzati e immagazzinati in gruppi che corrispondono alle colonne delle tabelle. Questa struttura agevola le query di lettura di grandi quantità di dati e con una complessità di calcolo elevata. Inoltre, siccome nelle stesse colonne i dati hanno lo stesso formato, rispetto al raggruppamento per riga si ha un miglioramento in termini di uso dello spazio di memoria, perché le colonne possono essere più efficacemente compresse. In aggiunta, lo spazio risparmiato grazie alla compressione per colonne può essere utilizzato per agevolare i processi di ordinamento dei dati. Microsoft vanta queste caratteristiche nella documentazione a riguardo [Microsoft, Columnstore indexes: Overview].

In questo caso il metodo di localizzazione può essere diverso dal classico B-Tree. Anche se generalmente considerati meno efficienti del B-Tree se non proprio per le query di lettura, gli indici bitmap possono essere una valida alternativa [Morteza Zaker, Somnuk Phon-Amnuaisuk, Su-Cheng Haw, 2008].

7 Cubi OLAP

“Un cubo multidimensionale è incentrato su un fatto di interesse per il processo decisionale. Esso rappresenta un insieme di eventi, descritti quantitativamente da misure numeriche. Ogni asse del cubo rappresenta una possibile dimensione di analisi; ciascuna dimensione può essere vista a più livelli di dettaglio individuati da attributi strutturati in gerarchie” [Golfarelli, Rizzi, 2006]. I cubi costituiscono la base di partenza per la progettazione delle dashboard e sono protagonisti nell’allestimento del data warehouse. L’addetto alla creazione dei cubi dovrà interfacciarsi con due figure, gli utenti che useranno le dashboard e i gestori del sistema informativo operativo, fonte dei dati transazionali. Superata questa fase, si è a conoscenza di quali sono i dati di interesse e come accedervi.

7.1 Identificazione dei requisiti e progettazione DW

Il primo step nella progettazione concettuale dei cubi è la comprensione dei requisiti, che per un’azienda di consulenza può avvenire solo tramite un dialogo con i clienti, gli utenti che useranno gli strumenti analitici da progettare. A seguito di questa fase si conosceranno i KPI e le metriche più importanti per il business, che implicheranno determinati carichi di lavoro sul database. Alcuni dati in media saranno più richiesti di altri, per questo conviene implementare le viste materializzate, oggetti del database pensati per conservare il risultato delle query più frequenti in modo tale da doverlo ricalcolare solo per aggiornarne i risultati, e non ogni volta che i dati sono richiesti per scopi di analisi. Creando le viste giuste, quindi, si avrà un miglioramento delle performance del database. Questi carichi di lavoro, però, non saranno costanti nell’arco dell’attività del business. Di conseguenza la progettazione del data warehouse deve essere rivista di continuo, in quanto deve adattarsi agli scostamenti del focus dell’organizzazione o a una variazione della frequenza delle interazioni, dovuta per esempio a un’espansione dell’organizzazione o alla chiusura di una divisione. Questa operazione, denominata tuning, consiste nella continua applicazione di tecniche di ottimizzazione volte ad aumentare le prestazioni del sistema in funzione delle circostanze in continuo mutamento.

Sempre a causa di questa incertezza nel trovare lo schema più corretto per soddisfare i requisiti aziendali, l’approccio preferito nella progettazione dei data warehouse è quello bottom-up, che prevede un popolamento incrementale degli elementi del data warehouse, suddiviso secondo settori aziendali specifici. All’inizio quindi si progetterà la Business Intelligence per una singola divisione dell’azienda, probabilmente la più importante, e successivamente si procederà una per volta con le altre fino ad avere il controllo di tutte. Solo a questo punto si inizierà a programmare analisi più generiche, che tengano in considerazione l’andamento complessivo del business. Talvolta, le singole aree aziendali che implicano rispettivi componenti logici del data warehouse, sono anche divise

fisicamente tra loro, ovvero si ha una scomposizione a livello hardware del data warehouse in quelli che vengono definiti data mart.

7.2 Studio e localizzazione dei dati a disposizione

Dopo aver individuato i requisiti, nella progettazione dei cubi bisogna tenere conto del volume dei dati, indicatore dell'onerosità dei processi di calcolo ottenuto come prodotto della granularità delle dimensioni. Conseguentemente, all'aumentare del livello di dettaglio e del numero delle dimensioni si avranno carichi di lavoro più gravosi. Come detto, l'obiettivo è creare strutture che possano rispondere alle richieste degli utenti in tempi rapidi, per questo bisogna trovare il minimo livello di dettaglio indispensabile ed escludere dimensioni che si ritiene possano non giovare molto alle potenzialità delle analisi, in modo tale da creare un data warehouse più efficiente. Per stimare opportunamente il volume dei dati occorre tenere in considerazione la sparsità. Un cubo con troppa sparsità e un livello di granularità elevato tenderà ad avere un volume dei dati sovrastimato. Inoltre, un alto grado di sparsità non porta a un buon funzionamento degli oggetti grafici in una dashboard, poiché questi risulterebbero incompleti per via delle vacanze di dati. La soluzione può essere aggregare, rimuovendo gli ultimi livelli gerarchici. Il massimo grado di dettaglio possibile sarà più basso ma risulteranno meno valori mancanti.

Prima di poter procedere con l'effettiva creazione dei cubi, bisogna conoscere la struttura del database transazionale, riassunto nello schema riconciliato, in quanto bisognerà recuperare le giuste informazioni per configurare il cubo. A questo scopo lo sviluppatore normalmente potrà anche accedere al database, usare delle query di lettura per comprendere dove sono situati i dati di interesse e giudicarne in prima istanza la qualità. Di norma, a tale scopo, viene fornito allo sviluppatore della BI lo schema delle fonti operazionali, cosicché possa conoscere quali sono i dati che ha a disposizione per creare i report e come reperirli. Nell'esperienza di tirocinio invece si è dovuto fare a meno di questo mezzo. Alcuni dei membri del team, esperti del sistema informativo e della struttura del database, possono fornire utili indicazioni su come adempiere a questo scopo in poco tempo, ma il più delle volte si è dovuto ricercare i dati autonomamente. L'approccio più conveniente in questo caso è accedere all'ERP, navigare fino all'area di interesse e da qui, grazie a una funzionalità di Dynamics 365, si può risalire alla provenienza dei dati riportati dopo averli selezionati, venendo quindi a conoscenza dei nomi delle tabelle nei quali sono salvati. In alternativa si possono lanciare delle query molto generiche nel database, osservarne i risultati e deducendo se si tratta di quanto cercato. In questa fase conviene valutare anche la qualità dei dati individuati. Infatti, per via di alcuni processi automatizzati, errori degli utenti o progetti sperimentali nel sistema informativo, si possono riscontrare delle anomalie che comprometterebbero le analisi. Quando capita ciò è opportuno filtrare i dati rimuovendo le anomalie, purché i dati rimanenti siano quantitativamente sufficienti per eseguire delle analisi significative.

```

SELECT      c.name AS 'NomeColonna', t.name AS 'NomeTabella'
FROM        sys.columns c
JOIN        sys.tables t    ON c.object_id = t.object_id
WHERE       c.name LIKE '%prodotto%'

```

La query SQL riportata è un esempio di interrogazione generica che permette di ricercare nel database colonne con un nome che contenga al suo interno una stringa (in questo caso “prodotto”), permettendo di comprendere in quali tabelle si potrebbero reperire le informazioni necessarie.

7.3 Generazione del cubo

Una volta determinati i dati di interesse, si può procedere a una prima progettazione del modello. A tale scopo si può fare uso del *Dimensional Fact Model* (DFM), uno schema proposto da Matteo Golfarelli, professore universitario italiano, per rappresentare intuitivamente quali saranno le analisi che il cubo permetterà. Ogni cubo è composto da un certo numero di dimensioni che fungono da coordinate di ricerca dei dati, disposte nei rami dello schema, e da delle misure, criteri di aggregazione dei dati da riportare nel box centrale (un esempio di DFM è in figura 5). Le dimensioni sono organizzate in gerarchie: diverse logiche con cui usare le dimensioni che implicano maggiore o minore granularità. Nell’ambiente Microsoft gli sviluppatori sono tenuti a modellare i cubi, definiti *aggregated measurements*, in Visual Studio.

Dopo aver individuato la localizzazione dei dati di interesse, lo step successivo consiste nello sviluppo tecnico del cubo, che consiste nella creazione di una vista, una tabella contenente tutti i dati necessari per una certa dashboard. Questo passaggio può essere eseguito usando un comando SQL (**CREATE VIEW**), ma Visual Studio facilita questa operazione: con un drag&drop degli elementi dal menù laterale, dal quale si può trovare qualsiasi oggetto del sistema, comprese le tabelle del database transazionale, è possibile unificare le tabelle in una vista generata come oggetto entità. Sarà necessario specificare attraverso quale chiave primaria ogni tabella aggiunta si lega alle altre e quali di tutti gli attributi a disposizione si vuole mantenere nella vista.

Ogni *aggregated measurement* è composto da diversi gruppi di misura, ognuno dei quali verrà collegato a una delle viste create. A queste entità si possono aggiungere delle misure, indicando il relativo campo della vista e un criterio di aggregazione tra la somma, il conteggio, la media, il massimo e il minimo. Analogamente si possono aggiungere le dimensioni, derivate da campi già disponibili nella vista del gruppo, o dimensioni esterne, provenienti da altre fonti. In questo ultimo caso sarà necessario specificare le join delle chiavi tra la tabella nel quale sono disponibili i dati della dimensione e la tabella del

gruppo di misure. Siccome alcune misure, all'interno di un'organizzazione, sono le stesse per diverse analisi, è possibile salvare degli oggetti dimensione in modo tale da non dover ripetere il settaggio per ogni cubo. Altre dimensioni inoltre valgono per qualsiasi organizzazione, in questo caso Microsoft predispone una tabella apposita contenente tutti i dati utili a eseguire le diverse aggregazioni. Un esempio è dato dalla dimensione temporale, la cui tabella, "BIDateDimensionValue", permette di eseguire tutti i passaggi gerarchici tipici di analisi temporali, vale a dire aggregazioni per mese, trimestre, anno e altre. La chiave primaria di queste tabelle equivale al minimo livello di granularità, in questo caso la data del giorno. Discorso analogo vale per la dimensione geografica.

Si riporta in figura 3 una schermata di Visual Studio, nel quale si vede un gruppo di misura. Tra le categorie si distinguono gli attributi, che sono dimensioni peculiari derivati dalla vista del gruppo di misura, ed esplose nell'immagine, si distinguono poi misure e dimensioni. Aprendo una dimensione, come si può vedere nella dimensione *Customer*, si trova la join, modificabile, che permette di legare la vista del gruppo di misura con la tabella della dimensione.

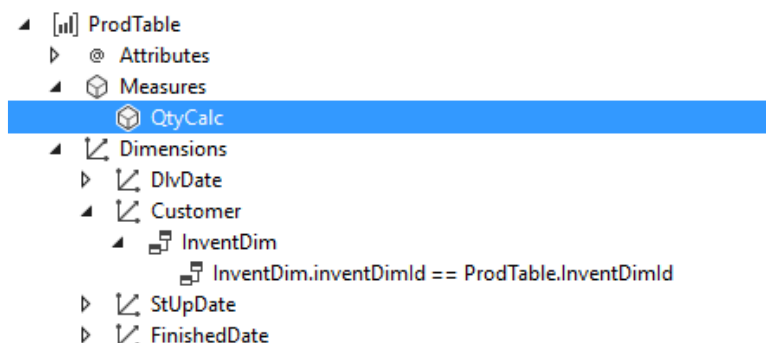


Figura 3, interfaccia del cubo di Visual Studio

Una volta creati gli oggetti, si procede con la *build* e la *sync* del progetto in cui si ha lavorato (chiamato soluzione in Visual Studio), che consiste nel rilasciare gli artefatti modellati in Visual Studio nel sistema informativo di riferimento. Se il processo va a buon fine, le *aggregated measurements* saranno visibili come oggetti entità nell'*entity store* (figura 4), un archivio che fa da intermezzo fra il database transazionale, AxDB, e il data warehouse, AxDW. Queste entità dettano al sistema come riorganizzare i dati nell'importazione da un database all'altro e da un'apposita interfaccia di Dynamics 365 è possibile impostare la frequenza di aggiornamento di ognuna di esse. Questi aggiornamenti sono riscontrabili nella maschera di Dynamics 365 *batch job*, nel quale sono tracciati tutti gli eventi di processo e il loro stato di completamento attuale. I dati passano così dalle tabelle del sistema di gestione alle tabelle dei cubi per le analisi multidimensionali.

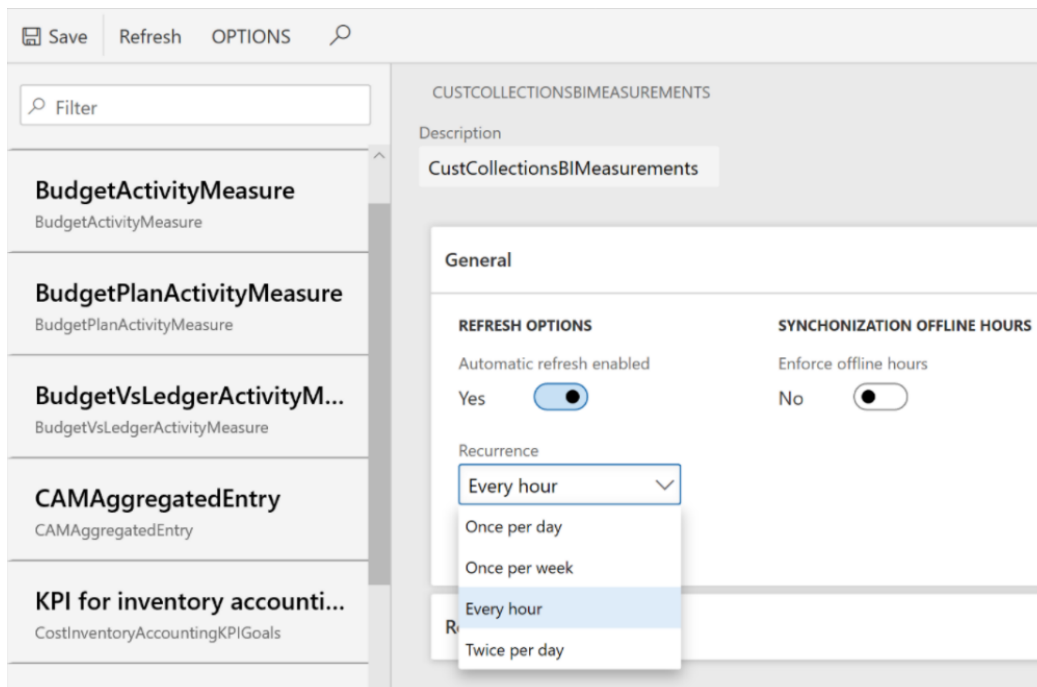


Figura 4, Schermata di Dynamics 365: Entity Store. A sinistra si può vedere l'elenco di cubi. Usando il menù a tendina si può scegliere la frequenza di aggiornamento.

7.4 Il futuro dei cubi OLAP

Purtroppo, l'ideazione di queste strutture risale a tempi in cui la velocità di calcolo e il costo della memoria erano ben altri. Al giorno d'oggi, la legge di Moore ha portato ai suoi effetti: i vincoli che rendevano onerosa la BI si sono rilassati e in alcuni casi il passaggio di creazione del cubo non porta a benefici tali da giustificarlo. In assenza di cubi la pipeline di analisi è più snella, rendendo più leggere e veloci anche le operazioni di modifica della stessa. Le grandi aziende tech, che hanno a disposizione le migliori infrastrutture tecnologiche, quando non hanno a che fare con dataset particolarmente grandi non fanno più uso dei cubi di dati perché attingendo dalle viste e dalle tabelle del data warehouse ottengono i risultati voluti in tempi ampiamente accettabili. Ciò è permesso in particolare dai sistemi a elevato parallelismo (MPP, massively parallel computing in inglese), nei quali centinaia o anche migliaia di processori sono usati in modo coordinato per singole operazioni. Infine, bisogna considerare l'efficienza dell'archiviazione indicizzata per colonne (Columnstore Index) discussa in precedenza, che costituisce un considerevole miglioramento nei tempi di risposta delle interrogazioni OLAP [The Holistics Blog, The Rise and Fall of the OLAP Cube]. Nonostante ciò, la progettazione concettuale del cubo tornerà comunque utile per rappresentare le aggregazioni che si potranno effettuare nelle analisi.

7.5 Esempio di generazione di un cubo

Di seguito è riportato il processo di creazione di uno dei cubi generati per conto dell'Azienda B. In questo caso i requisiti sono stati forniti da chi ha commissionato la progettazione del cubo e sono stati inseriti in un file condiviso anche con il cliente. In figura viene riportato il paragrafo del file relativo al cubo in questione.

2. CONSUMI (Inventrans e InventTransOrigin dei consumi) degli articoli

Articolo, Quantità, Unità, Data Effettiva, Data Finanziaria, Importo costi, valuta

Sucido Consumato:

Seriale:

- Dati del seriale: *numero di serie*
- Attributi seriale: *Container, Data_carico, Data_pesa, Fattoria, N_Balla, Nave, P_Lordo, P_Lordo_Pesa, P_Netto_Pesa, P_Tara, Piombo..*

Contromarca:

- Dati del batch: *numero batch, Paese di origine, Numero batch Fornitore*
- Attributi batch: *Lotto cliente*

Il fatto di interesse del cubo sono i consumi dei prodotti chimici nel processo di lavaggio. Viene riportato in figura 5 il DFM modellato.

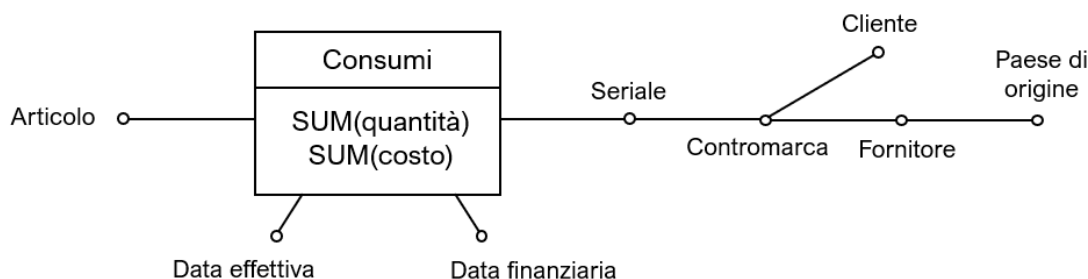


Figura 5, DFM dei consumi di Azienda B

Le misure per descrivere il fatto sono la quantità usata e il costo. La dimensione relativa all'articolo identifica il prodotto chimico; data effettiva e finanziaria indicano rispettivamente la data del reale avvenimento del consumo e la data di registrazione dell'avvenuto consumo nella contabilità; la dimensione del seriale identifica una singola balla di lana. Salendo di livello gerarchico si passa alla contromarca, assegnata a ogni lotto di balle. Ogni lotto è riferito a un solo cliente ed è proveniente da un solo fornitore, di cui si considera anche il paese di origine.

Ogni balla di lana è associata a degli attributi che la caratterizzano, informazioni ottenute attraverso l'analisi della lana indicate nel file con il nome "attributi seriale". Questi attributi non fanno parte delle dimensioni perché le aggregazioni di questi non sono utili ai fini delle analisi, se non chiaramente inappropriati per le aggregazioni, come nel caso degli attributi sul peso. Lo stesso discorso vale per unità di misura e valuta, attributi strettamente legati all'articolo utili a quantificare le misure. Per quanto non compaiano nel DFM, questi attributi sono mappati nella creazione del cubo perché sono delle informazioni di supporto la cui visualizzazione nelle dashboard può essere utile.

8 Dashboard

Le dashboard sono dei cruscotti di controllo composti da elementi grafici, quali diagrammi, istogrammi, tabelle, cartine geografiche o semplici indicatori numerici, utili a rappresentare visivamente i dati processati. Esse sono assimilabili ai cruscotti delle automobili, che aggiornano chi è al volante sullo stato del veicolo per guidare in tutta sicurezza. Allo stesso modo, tenendo d'occhio queste interfacce, utenti come manager o componenti dello *strategic apex* aziendale possono trarre conclusioni sulla performance dell'operato a livello storico, corrente e predittivo al fine della creazione di nuove opportunità di business per la crescita nel lungo termine dell'organizzazione. Ogni dashboard è pensata per utenti specifici ed è composta da insights dell'area di interesse. Si tratta di uno strumento attraverso cui utenti anche non esperti possono esplorare i dati strutturati dei cubi OLAP. L'obiettivo è permettere a qualsiasi utente di compiere delle analisi ad alto valore aggiunto pur senza disporre di competenze analitiche avanzate. La progettazione delle dashboard è a carico del responsabile della BI dell'azienda, a meno che non si decida di affidarsi a una società di consulenza. Altre volte, lo sviluppatore della dashboard coincide con l'utente che ne farà uso per le analisi.

8.1 Generazione della dashboard

Completato il processo di predisposizione delle strutture di dati, si può procedere con la composizione della dashboard con l'ausilio degli oggetti grafici disponibili su Power BI. Oltre ai diagrammi presenti nel programma di default, è possibile aggiungerne di nuovi da un portale nel quale gli sviluppatori sono abilitati ad aggiungere i propri. In questa fase bisogna conoscere le priorità delle informazioni per costruire la dashboard in modo tale da far risaltare ciò che più è rilevante per l'organizzazione. È fondamentale creare una

struttura che abbia una logica di disposizione degli strumenti di visualizzazione dei dati, che devono essere facilmente navigabili.

Molte aree di lavoro sono comuni in business diversi, si pensi al reparto risorse umane o al reparto vendite. Allo stesso modo esistono degli indicatori validi per business diversi, come quelli relativi al conto economico. Per questa ragione Power BI, come i suoi competitors, offre una libreria di dashboard standard personalizzabili (*templates* in inglese) dalla quale lo user può attingere per risparmiarsi parte del lavoro, dovendo soltanto mappare i dati nelle misure e negli oggetti grafici già predisposti.

8.2 Interazione con la dashboard

Le dashboard di Power BI, come quelle di gran parte dei servizi di BI, sono interattive. Oltre a visualizzare i risultati, con alcuni semplici passaggi è possibile esplorare a proprio piacimento i dati nelle dimensioni di analisi dei cubi, operazioni altrimenti eseguite attraverso interrogazioni SQL. Infatti, interagendo con la dashboard, l'utente fa eseguire al software delle operazioni tipiche di elaborazioni OLAP: roll-up, drill-down e slicing.

Con l'operazione di roll-up avviene una riduzione del livello di dettaglio tramite aggregazione dei dati e un aumento di livello di una gerarchia della dimensione. L'aggregazione è un processo che comporta una riduzione della quantità di dati, attraverso una sintesi che cambia la scala e ne aumenta la stabilità. Si pensi per esempio al numero di vendite giornaliere di un certo prodotto. Aggregando per mese, si riduce di un fattore $12/365$ la quantità di dati e si rimuove così la variabilità giornaliera. Si passa quindi a un livello di dettaglio inferiore risalendo lungo la gerarchia del tempo, aumentando la stabilità del dato. Il drill-down è l'operazione contraria, è come annullare il roll-up per poter analizzare il dettaglio di un insieme di dati. Per fare un esempio pratico di uso della dashboard, si pensi a un grafico a torta che riporta la redditività di un certo prodotto nei diversi stati europei, attivando il drill-down e selezionando uno stato, si potrà visualizzare la redditività di ognuno dei punti vendita dello stato (se il livello gerarchico successivo distingue ogni punto vendita dello stato). Notare che nell'uso della dashboard il drill down non è una disaggregazione di tutti i dati, ma solo dell'insieme selezionato.

Attraverso lo slicing invece si seleziona un range di una certa dimensione di analisi e si filtra il cubo di conseguenza, ottenendo una "fetta". Ciò avviene quando selezionando un qualsiasi elemento della dashboard, se questa è opportunamente progettata, tutti gli oggetti visivi vengono filtrati per mostrare i numeri di quell'elemento. In particolare, gli indicatori numerici vengono ricalcolati e nei diagrammi le porzioni di interesse vengono evidenziate. Altri filtri possono essere effettuati dalla prima colonna sul lato sinistro della dashboard, dove trascinando determinati campi dall'elenco dei dati importati si

possono selezionare precisi insiemi di dati secondo le logiche disponibili (range, selezione singola, etc.). Nelle immagini sottostanti è riportato un esempio di slicing in Power BI. In figura 6 è stato selezionato dal diagramma a barre il “prodotto a” e nel grafico a torta a fianco il programma di conseguenza evidenzia le quantità vendute del prodotto nei vari punti vendita. Sembrerebbe che il “punto vendita x” sia il principale rivenditore del prodotto a, il più venduto dei quattro, ma è secondo al “punto vendita y” in quantità vendute totali. Nella figura 7 invece viene selezionato il “punto vendita z” del grafico a torta per visualizzare le rispettive quantità vendute dei vari prodotti nel grafico a barre. I grafici sono stati realizzati con dati generati casualmente.

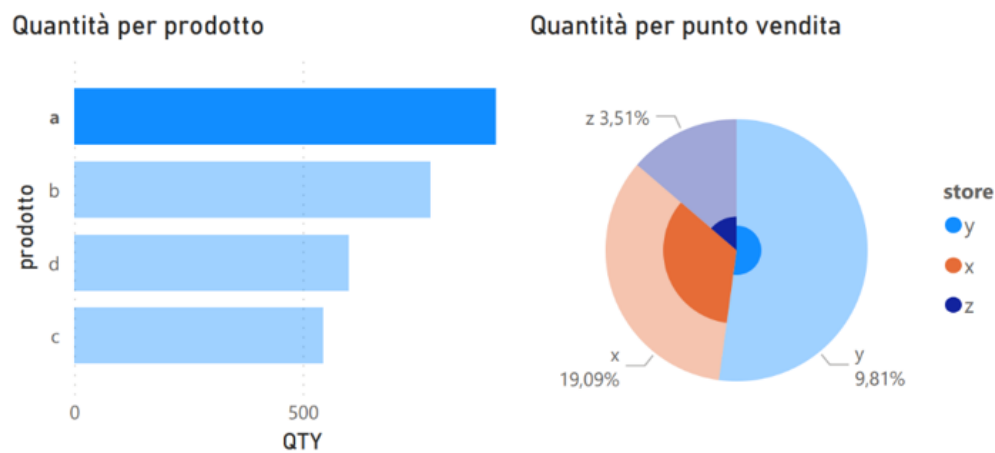


Figura 6, Slicing (filtraggio) degli elementi grafici in base al “prodotto a”

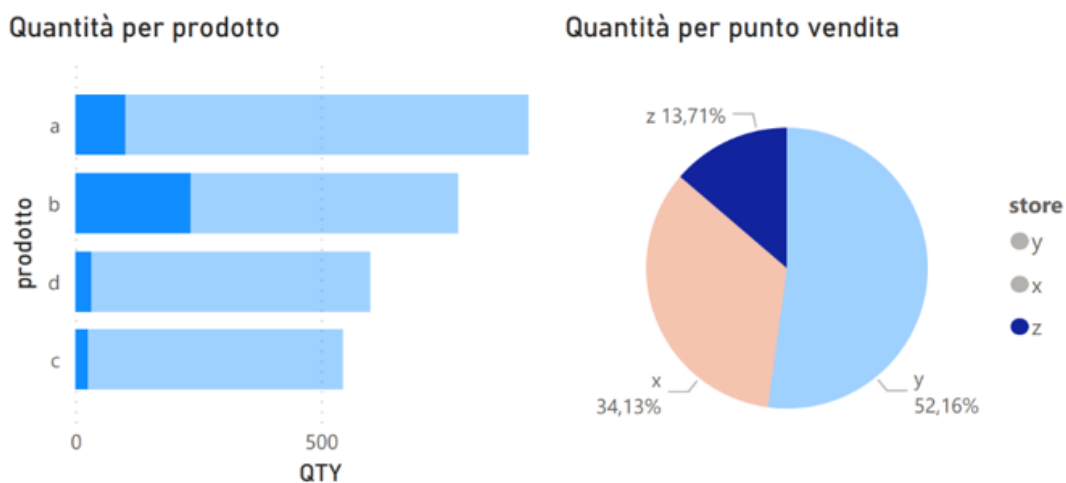


Figura 7, Slicing degli elementi grafici in base al “punto vendita z”

8.3 Esempi di dashboard

In questa sezione verranno analizzate alcune delle dashboard realizzate per conto dei clienti di Cluster Reply. Le dashboard di Azienda A sono a scopo esemplificativo di uso del software, tant'è che i dati non sono reali e provengono dalle dimostrazioni di uso di Dynamics 365 fatte durante la consulenza o dai database di esempio di Microsoft. Invece le dashboard di Azienda B sono state create sotto richiesta del cliente per il controllo di dati relativi alla contabilità aziendale.

8.3.1 Ritardo consegne Azienda A



Figura 8, Dashboard dei ritardi nelle consegne

Il cruscotto in figura 8 è molto semplice, consiste in due istogrammi che riportano la puntualità delle consegne per cliente e per prodotto di un certo anno, il primo in termini assoluti e il secondo in termini relativi, per meglio confrontare i diversi risultati della variabile delle ascisse. Comunemente, ogni ordine è caratterizzato dalla data richiesta di consegna, la data nel quale il cliente richiede di avere una certa merce e dalla data promessa, entro la quale il fornitore promette di effettuare la consegna. Rapportando il tempo impiegato a soddisfare l'ordine a partire dalla data di creazione dell'ordine, e il tempo previsto di consegna sempre rispetto alla data di creazione ordine, si ottiene una

frazione maggiore di uno quando l'ordine è in ritardo, permettendo di confrontare tempi di consegna appartenenti a scale di grandezza diverse.

$$KPI_{\text{puntualità}} = \frac{\text{data reale consegna} - \text{data creazione ordine}}{\text{data promessa consegna} - \text{data creazione ordine}}$$

Grazie a questo indicatore si può creare il report in questione e gli utenti potranno fare ipotesi riguardo la correlazione dei ritardi con un prodotto particolare, o scoprire che per un certo cliente si sta performando peggio rispetto agli altri. Nell'esempio, è stato assegnato un colore in base alla gravità del ritardo (o se la consegna è puntuale o in anticipo), basata sull'indicatore menzionato. Se il calcolo avviene in fase di trasformazione dei dati su Power BI, data la sua complessità, non è possibile impostare la funzionalità Direct Query. Per avere la funzionalità, bisognerebbe implementare nell'applicativo dell'ERP una funzione di calcolo dell'indicatore che si attiva ogni volta che c'è un nuovo ordine e che ne salvi il risultato, in modo tale che possa essere immediatamente disponibile nel database.

8.3.2 Analisi OTIF Azienda A

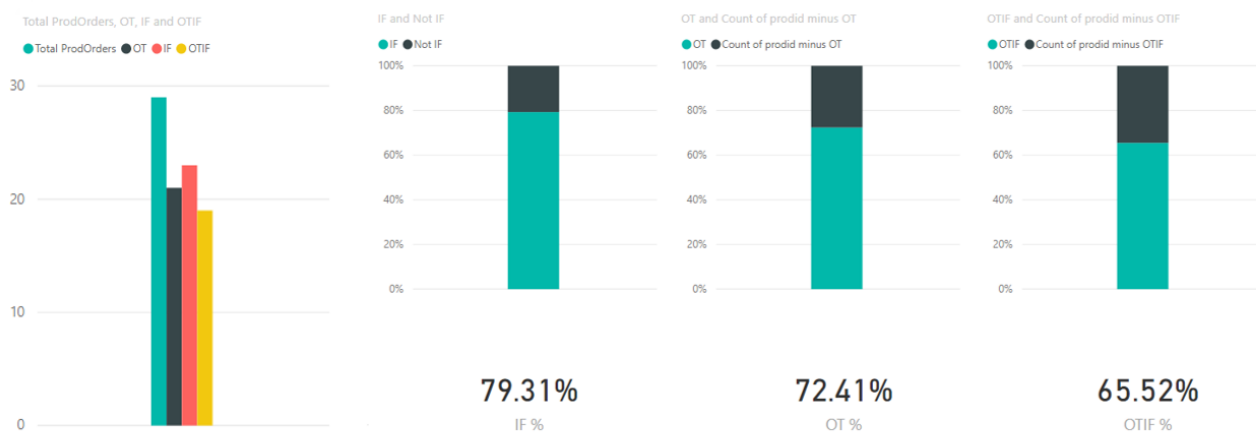


Figura 9, Dashboard OTIF

OTIF sta per On-Time In-Full, che tradotto dall'inglese significa ordine puntuale e completo. Questa analisi (figura 9) consiste nel calcolare le percentuali di ordini che sono puntuali, la cui data di consegna è uguale o antecedente a quella promessa, la percentuale di ordini completi, il cui numero di unità consegnate equivale al numero di unità promesse, e l'intersezione di questi due insiemi. Chiaramente si ambisce ad avere un'alta frazione di ordini puntuali e completi e osservando questi istogrammi l'utente può conoscere come sta performando l'azienda nell'adempimento di questo obiettivo.

8.3.3 Conto economico Azienda B

La serie di dashboard seguenti è stata creata per l'Azienda B e ha lo scopo di sostituire un report Excel fatto di molteplici fogli, creato dai dipendenti dell'azienda al fine di tenere traccia dell'andamento di importanti indicatori economici. Si tratta, infatti, di un insieme di cruscotti adibito a riassumere i dati della contabilità dell'azienda. Il vantaggio rispetto a Excel consiste nell'automatizzazione dell'aggiornamento dei dati e nella più facile comprensione delle informazioni visualizzate. Nel lavoro di creazione della dashboard il processo più critico è stato l'individuazione delle voci di entrate e uscite del bilancio. L'azienda tiene traccia di ogni voce attraverso un codice identificativo. Gruppi di voci di costo, sono identificati da altri codici, costituendo così una gerarchia a più livelli. La struttura gerarchica e la corrispondenza dei codici alle rispettive voci erano informazioni non disponibili nel database ma solo nell'Excel. Per questa ragione è stato necessario creare un nuovo file Excel contenente la gerarchia e i nomi dei conti, che è stato importato staticamente su Power BI. Mappando il file con le tabelle del cubo OLAP generato, si è potuto ricreare l'intera struttura contabile e ricalcolare le voci aggregate, che dopo la correzione di alcune imprecisioni sono risultati equivalenti a quelle riportate nell'Excel dell'azienda.

La prima dashboard, in figura 10, è la più importante: nella fascia superiore si può leggere l'ammontare dei principali componenti del conto economico: fatturato lordo, valore aggiunto, margine operativo lordo, margine operativo netto, risultato ante imposte. Questi valori, fatta eccezione del fatturato, non erano immediatamente disponibili nel database, per questo è stato necessario creare le rispettive misure nell'ambiente di Power BI. Al centro è stata creata una tabella pivot che dà allo user la possibilità di esplorare più nel dettaglio gli stessi indicatori nelle varie componenti. Il grafico in basso riassume visivamente l'ammontare degli stessi indicatori. In basso a sinistra invece, grazie al menu a tendina e allo slider inseriti, lo user potrà filtrare a piacimento tutti gli oggetti della dashboard nella dimensione temporale. Nell'immagine si può osservare l'ammontare dei valori economici raggiunti nell'anno 2019 fino al mese di ottobre. La seconda dashboard, in figura 11, invece serve a confrontare i risultati cumulati dell'anno corrente con quelli dell'anno passato in un certo mese. In basso a sinistra sono stati inseriti dei filtri analoghi a quelli della prima dashboard. Nella fascia in alto ci sono gli stessi indicatori della dashboard precedente ma come differenza rispetto all'anno passato (YOY sta per Year Over Year, dicitura inglese per indicare questo genere di analisi). Il grafico riporta il confronto di uno solo degli indicatori, il fatturato lordo. I numeri in basso a destra sono l'ammontare di balle di sughero ricevute nell'anno corrente e nel precedente. Quest'ultimo dato non è pertinente alla contabilità ma, a giudicare dell'Excel di riferimento, sembra che serva all'azienda come punto di riferimento per avere un'idea dell'attività dell'azienda.

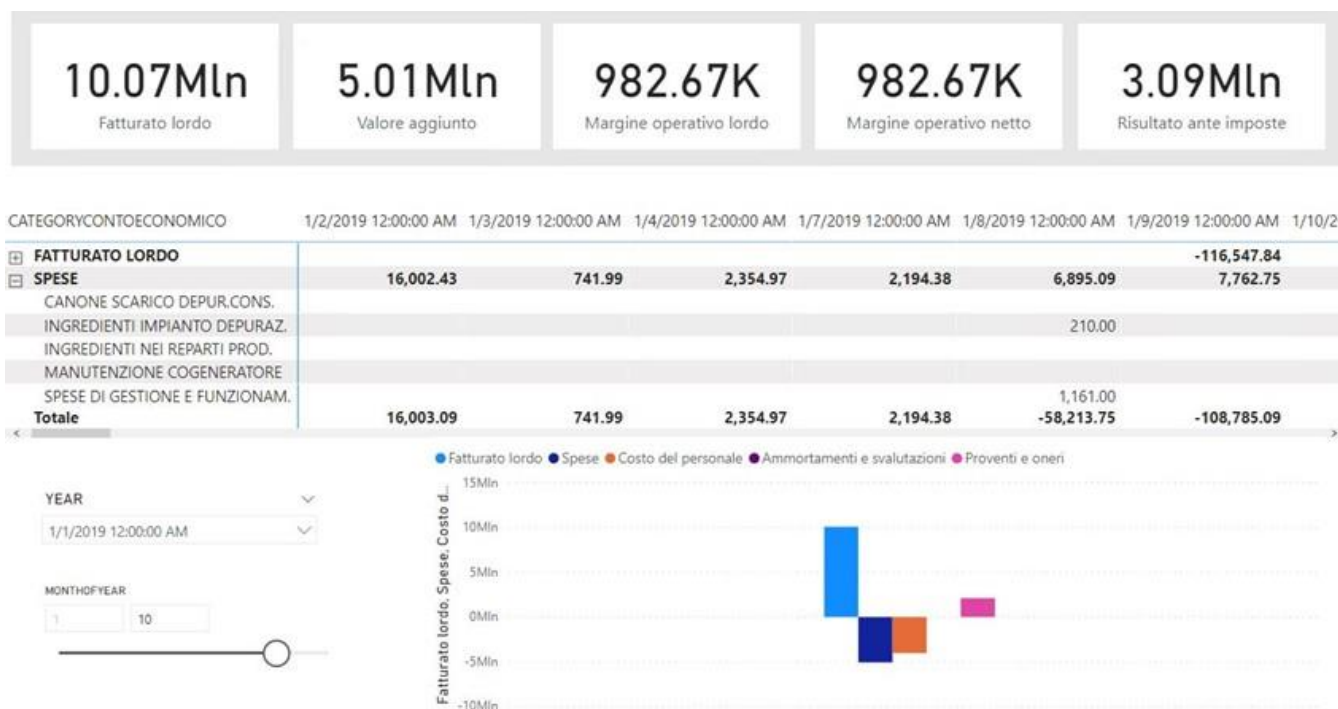


Figura 10, Dashboard conto economico

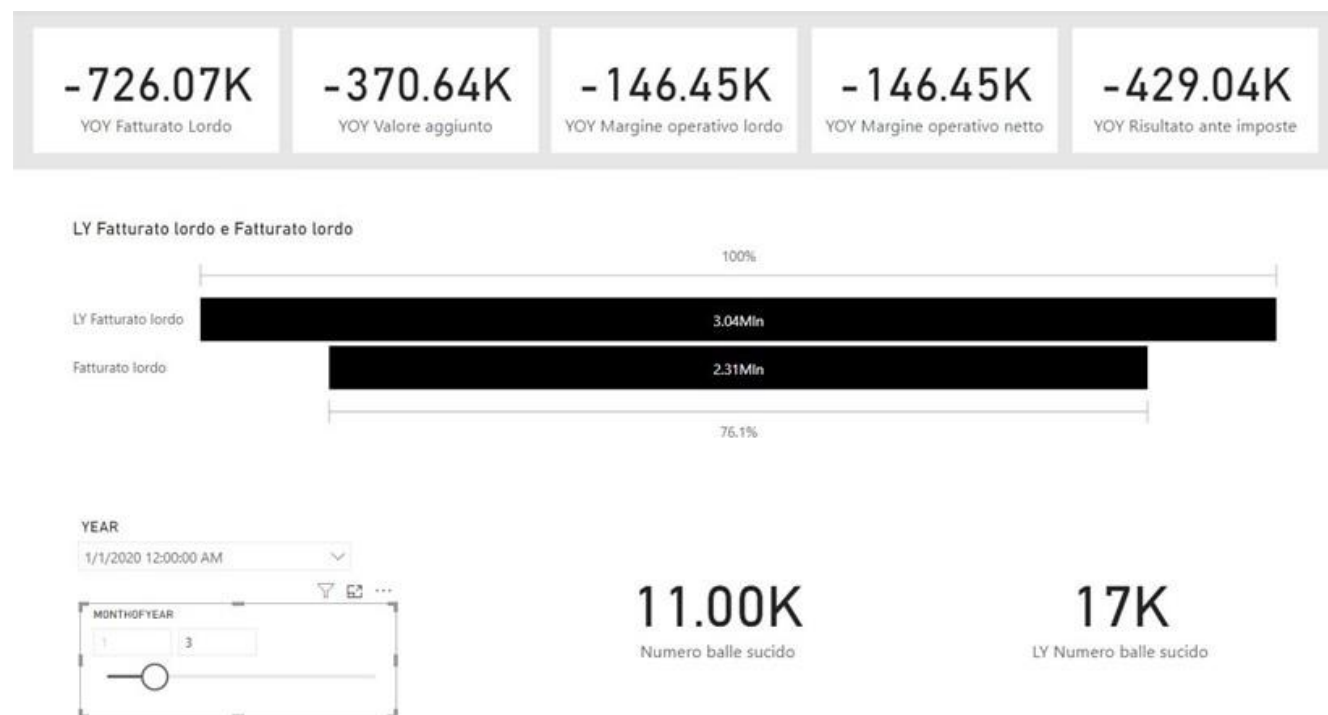


Figura 11, Dashboard conto economico YOY

9 Power BI

Nell'ambiente di produzione Dynamics 365 è possibile installare Microsoft Power BI al fine di generare delle dashboard per visualizzare i dati raccolti tramite grafici e KPI. Grazie alle dashboard create i manager dell'azienda possono valutare a colpo d'occhio la performance del business e individuare minacce e opportunità. Si tratta di un applicativo molto flessibile, intuitivo e sempre aggiornato in tutte le sue parti.

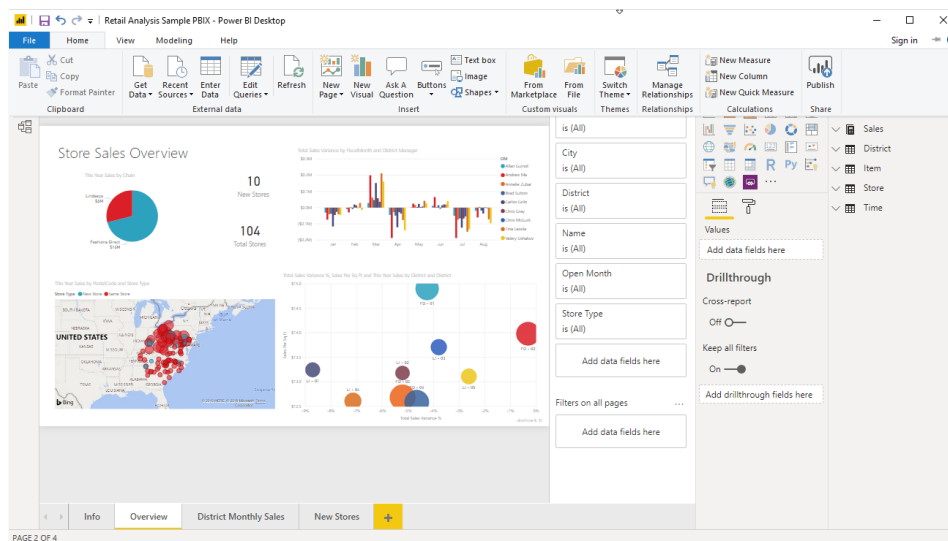


Figura 12, Schermata di Power BI [sito Microsoft]

9.1 Importazione dei dati

Tra i fattori determinanti nella creazione di una dashboard efficace, oltre alla scelta dei giusti strumenti di visualizzazione e una corretta selezione delle informazioni necessarie, è importante studiare la migliore modalità di importazione dei dati. L'obiettivo è garantire i minori tempi di aggiornamento possibili e chiaramente, evitare quando possibile l'importazione di dati statici. Con Power BI è possibile usare come origine dei dati molteplici fonti, anche nella stessa dashboard: servizi online (tra i quali PowerBI.com, ricco di dataset su ogni argomento), file in locale o i database privati ai quali si ha accesso. Si potrebbe anche usare il database transazionale, ma questo tipo di progettazione implicherebbe un'onerosità di calcolo non ottimale, poiché come spiegato in precedenza il database transazionale non è funzionale a scopi di analisi. La procedura prevede quindi di usare come fonte il data warehouse, AxDW, e richiamare i cubi OLAP generati in precedenza. In fase di importazione lo sviluppatore è tenuto a scegliere tra le modalità Import Mode e Direct Query. Con la prima i dati sono trasferiti nella cache di Power BI, una memoria temporanea che si aggiorna periodicamente, con la seconda invece ogni volta

che c'è un'interazione con la dashboard o un suo refresh, viene eseguita una nuova query direttamente nella fonte primaria. Quest'ultimo settaggio è più conveniente nei seguenti scenari [Microsoft Documentation, About using DirectQuery in Power BI]:

- Il tasso di produzione dei dati è alto e i report devono essere quasi in tempo reale: con Direct Query si crea una connessione diretta con l'origine e i nuovi dati saranno visualizzati con un ritardo di decine di minuti al massimo per via dei tempi di aggiornamento del data warehouse.
- Elevata quantità di dati: la cache di PowerBI è limitata, quindi in alcuni casi potrebbe non essere sufficiente a contenere tutti i dati utili alla corretta visualizzazione della dashboard. Con Direct Query invece i dati vengono direttamente processati e il problema non si presenta, la cache viene usata ma solo per memorizzare i risultati utili per la visualizzazione.
- Sicurezza: condividendo le dashboard con l'import mode si trasferiscono le relative porzioni di dati, quindi si deve fare attenzione a chi ha accesso alla dashboard. Con Direct Query l'accesso alla dashboard è legato all'accesso del database, quindi solo chi ha le credenziali per le suddette porzioni di dati potrà visualizzare la dashboard. Il controllo è quindi automatizzato.
- Restrizioni sulla sovranità dei dati: alcune organizzazioni impediscono la fuoriuscita dei dati dalla loro sede principale e solo usando Direct Query i dati non vengono salvati in memorie secondarie.

I vantaggi di usare questa funzionalità come spiegato sono molti, ma ci sono delle problematiche che in altre situazioni la rendono svantaggiosa:

- Le numerose query eseguite direttamente nella fonte primaria sono un carico di lavoro in più per il database principale dell'intero sistema da non sottovalutare.
- I tempi di risposta della dashboard dipenderanno dalle prestazioni del database di origine, che se non ha buone prestazioni implicherà dei tempi di aggiornamento inaccettabili o ben superiori a quelli che si avrebbero con l'import mode. Se si verifica un timeout la dashboard non viene aggiornata.
- Limitazione nelle trasformazioni dei dati e nella modellazione di misure.
- Alcuni oggetti visivi potrebbero essere incoerenti con gli altri, perché l'utente ha interagito con essi. Quando ciò si verifica, è sufficiente lanciare un aggiornamento dell'intera dashboard.

Dopo aver lanciato l'importazione, i dati selezionati saranno disponibili in una colonna sulla destra. Da qui, sarà possibile trascinarli negli oggetti visivi per comporre la dashboard.

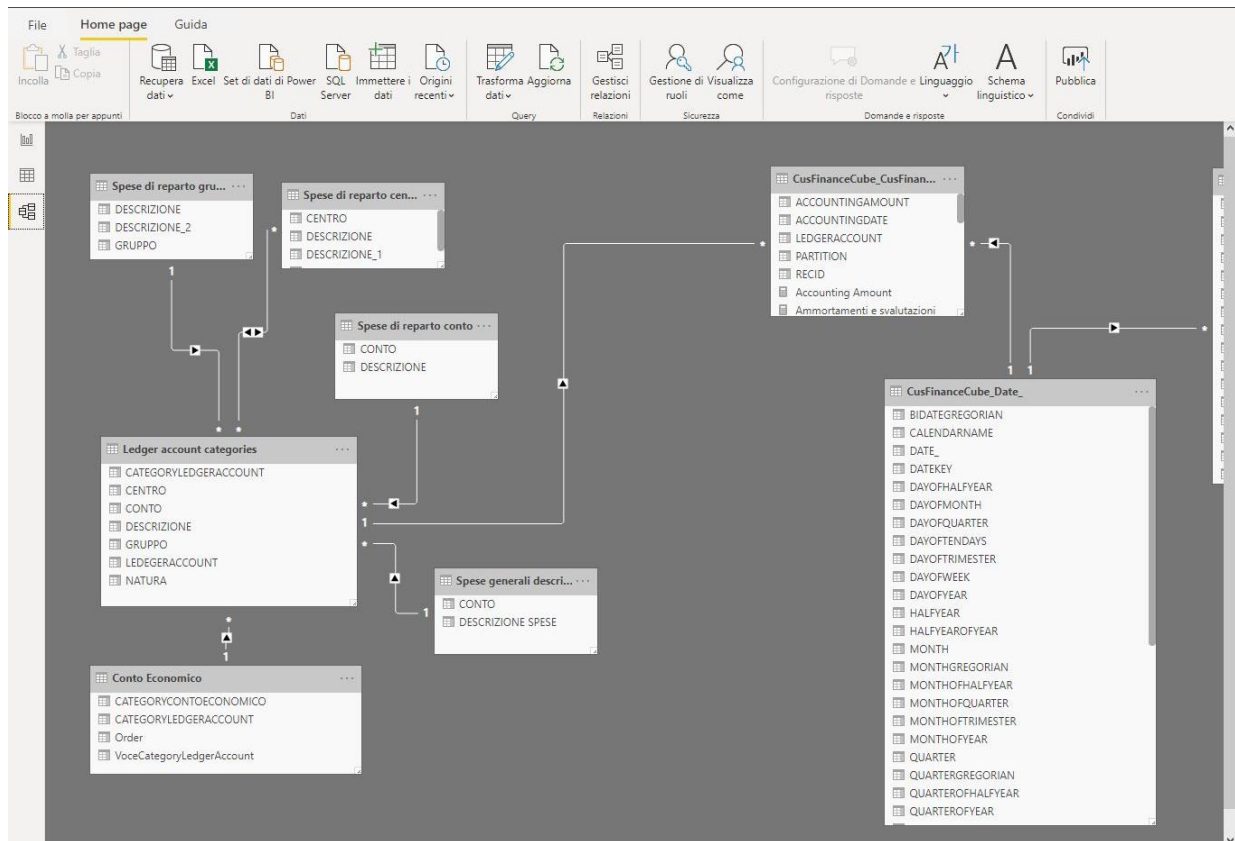


Figura 13, Schema delle relazioni di Power BI

Dopo l'importazione è necessario specificare i legami fra le tabelle, allo stesso modo in cui sono stati impostati nella creazione dei cubi. Nell'immagine in figura 13 si può vedere una maschera di Power BI pensata per impostare le join in modo intuitivo. Trascinando un attributo di una tabella sopra un secondo attributo di un'altra tabella, si va a indicare il legame delle chiavi delle due entità. È necessario specificare la cardinalità della relazione, ma il software generalmente la riconosce automaticamente. Da questa interfaccia si potranno mappare le relazioni con dati provenienti da altre fonti.

9.2 Trasformazione della struttura dei dati

Dopo l'importazione, i dati possono essere filtrati e trasformati modificando o aggiungendo colonne derivate da quelle native del data warehouse. Ogni passaggio di trasformazione viene registrato in una pipeline, cosicché si possa revisionare la catena di trasformazioni che verrà effettuata ogni volta che la dashboard viene riaggiornata con i nuovi dati.

A questo punto le tabelle e i relativi campi saranno disponibili nella sezione destra del programma, dalla quale si potranno trascinare negli oggetti grafici per il loro popolamento. Lo sviluppatore può aggiungere delle misure ulteriori oltre a quelle predisposte nel cubo,

anche con formule complesse e in funzione di altre misure, per esempio per ottenere KPI caratteristici nelle analisi da parte dell'organizzazione. La formulazione di queste misure avviene attraverso il linguaggio di interrogazione di Microsoft, DAX (Data Analysis eXpressions), che vanta centinaia di formule precostituite per le analisi.

Tutte queste operazioni, affinché la funzionalità Direct Query non si disattivi, devono implicare calcoli non diversi da aggregazioni. Il che significa che non si potranno eseguire operazioni che restituiscano un risultato singolo per ogni riga, ma solo operazioni di aggregazione. La ragione di tale limitazione è che le trasformazioni devono sempre poter essere convertite in query dal software. Diversamente, quando si usa l'import mode non ci sarebbero limitazioni alle operazioni di trasformazione, talvolta molto utili per pulire il dataset o per fare analisi più complesse. Nel momento in cui viene eseguita un'operazione che impedisce l'uso di Direct Query il sistema notifica l'utente, proponendo il passaggio a Import Mode o l'annullamento dell'ultima trasformazione eseguita. Al contrario, il passaggio di un cruscotto da Import Mode a Direct Query non può essere effettuato.

9.3 Esportazione delle dashboard

Le dashboard ultimate possono essere salvate come file di tipo pbx in locale, ma sono pensate per essere visualizzabili su altri sistemi. Per integrarle con altre piattaforme, come il sistema informativo aziendale, è necessario pubblicarle nella propria area privata di PowerBI.com. Dopo questo passaggio, su Dynamics 365 si può fare uso di una sezione apposita per la gestione delle dashboard. In altri ERP l'integrazione potrebbe risultare più difficile. Per incorporare le dashboard al proprio sito web o in piattaforme private invece sono disponibili delle API.

Sviluppi futuri

In questa ultima parte vengono analizzate le novità tecnologiche più recenti della BI, l'attuale stato del mercato di questi servizi e le prospettive future.

10 Augmented Analytics

La principale innovazione nel mondo della Business Intelligence è l'applicazione di algoritmi di intelligenza artificiale negli strumenti di analisi dei dati. Si possono distinguere due forme di utilizzo per le analisi di queste tecniche dirompenti nel mondo dei software: l'applicazione mirata alle analisi stesse, per esempio nella generazione di modelli predittivi basati su dataset, e l'applicazione pensata ai fini della creazione di strumenti di analisi più autonomi. In quest'ultimo caso i poteri predittivi di questi algoritmi sono sfruttati per anticipare alcune delle operazioni di progettazione della BI, in una serie di funzionalità riunite sotto l'espressione "Augmented Analytics" di cui si riporta qui di seguito la definizione tradotta dall'inglese proveniente dalla ricerca di Gartner "Augmented Analytics Is the Future of Data and Analytics" del 2017 nel quale il termine è stato coniato.

"Augmented Analytics è un approccio di analisi statistiche dei dati basato sull'uso di algoritmi di Machine Learning e Natural Language Processing finalizzato ad automatizzare i processi di analisi normalmente eseguiti da specialisti o data scientist".

Machine Learning (ML) è una branca dell'intelligenza artificiale che consiste nello sviluppo di algoritmi in grado di generarsi autonomamente (imparare) attraverso l'esperienza e i dati. Grazie a queste tecniche, tra le tante applicazioni, i più moderni servizi di BI offrono funzionalità di "Automated Insights" che dati pochi input dell'utente generano un quadro di analisi dettagliato selezionando variabili e strumenti grafici in automatico. Per imparare dall'esperienza, questi algoritmi tengono in considerazione i risultati che ottengono: nel caso di Power BI, dopo aver usato queste funzioni l'utente può inviare un feedback direttamente dal programma su quanto ha apprezzato l'output dell'algoritmo. Il feedback è strutturato in un semplice apprezzamento o non apprezzamento, o in una selezione multipla di possibili commenti. Queste informazioni sono raccolte a livello globale dall'algoritmo di Microsoft che impara di conseguenza.

Natural Language Processing (NLP) è un'applicazione del ML il cui obiettivo è permettere alla macchina di comprendere o generare espressioni del naturale linguaggio umano per facilitare le interazioni con il software. L'applicazione più comune del NLP è

la commutazione di interrogazioni al database formulate in linguaggio comune in query SQL, detta Natural Language Query. In Power BI, immessa un'espressione analoga a un'interrogazione nell'apposito campo, in risposta il software sceglie anche l'oggetto visivo più adatto a rappresentare i dati richiesti.

Si riporta un esempio pratico in cui sono usati entrambi i tipi di algoritmi, ML e NLP: con la funzionalità “spiega aumento” di Power BI riferita a un determinato rialzo di un grafico a linee con asse temporale, il software identifica le variabili che l'hanno causato e “spiega” allo user in linguaggio naturale quanto trovato, abbinando il tutto con un grafico dimostrativo.

11 Il mercato della BI

Gli applicativi che formano questo mercato sono pensati per aiutare le aziende a valutare i risultati ottenuti e prendere decisioni strategiche oculate. Tutte le moderne piattaforme di analisi e Business Intelligence (ABI) sono progettate per essere intuitive, facili da utilizzare e in grado di supportare l'intero flusso di analisi, dalla preparazione dei dati alla visualizzazione e generazione degli insight. Secondo quanto trovato in una ricerca di mercato di Gartner del 2021, nel 2019 il fatturato derivato dalle piattaforme BI è cresciuto del 7%, di tre punti percentuali in meno rispetto all'anno precedente, aggirandosi intorno ai 15 miliardi di dollari. La diminuzione del valore economico del settore è dovuta a una forte competitività sui prezzi che ne ha portato un calo, ma, nonostante ciò, il numero di utenti continua a incrementare enormemente.

I fornitori di questi servizi spaziano da affermati colossi tecnologici a start-up finanziate da venture capital. Le più grandi solitamente integrano in offerta anche i servizi di gestione dei dati cloud, come nel caso di Microsoft Azure.

Attualmente il mercato è indirizzato verso l'implementazione di cloud più efficienti. La maggior parte degli investimenti, riguarda proprio la ricerca verso la creazione di un sistema il più coeso possibile, per quanto complesso ed esteso.

Le piattaforme ABI non sono più differenziate dalle funzionalità di visualizzazione, ormai date per scontate. Gli oggetti grafici più efficaci sono stati adottati da tutti i fornitori, quindi cruscotti di case diverse sono simili fra loro. Ciò che ad oggi determina principalmente il vantaggio competitivo sono le proprietà di Augmented Analytics, ovvero, come spiegato in precedenza, l'uso dell'intelligenza artificiale per supportare lo user nelle

analisi. Una tendenza importante infatti è data proprio dall'implementazione di applicativi in grado di facilitare il più possibile il processo di analisi, allo scopo di estendere la gamma di utenti in grado di trarre vantaggio dalla Business Intelligence.

Oltre a queste caratteristiche, si elencano qui altri fattori che sono critici per competere nel mercato:

- Sicurezza: requisito fondamentale, le aziende sono interessate a preservare la segretezza dei dati, per tanto ogni accesso alla piattaforma deve essere sotto controllo.
- Cloud Analytics: l'abilità di supportare la costruzione, distribuzione, e gestione degli strumenti di analisi nelle piattaforme cloud, che siano basati su dati allocati su cloud o in locale.
- Preparazione dei dati: possibilità di combinare informazioni provenienti da diverse fonti e di personalizzare i modelli analitici in termini di misure, gruppi e gerarchie.
- Dashboard interattive: dashboard con cui l'utente può interagire per condurre le proprie analisi.

I modelli di prezzi offerti dai leader di settore sono incrementali, pensati in una maniera tale da motivare i clienti a comprare servizi più completi dopo aver osservato i primi risultati positivi di versioni limitate del software.



Grafico 6, fatturato derivato da servizi BI in milioni di dollari dal 2014 al 2018 di alcune importanti aziende del settore [SAS Institute, 2018]

Secondo la ricerca di mercato di Gartner, Power BI è leader indiscusso del mercato. Tra i vantaggi del programma sicuramente non si può non considerare la casa produttrice, Microsoft infatti tramite Office, diffusissimo nelle organizzazioni, ha un ottimo canale di distribuzione. Inoltre, la BI di Microsoft primeggia nelle funzionalità di Augmented Analytics. L'unica debolezza di questo strumento è il basso livello di compatibilità con altri sistemi, tant'è che Power BI da cloud è disponibile solo se come IaaS si fa uso di Microsoft Azure. Seppur in seconda posizione rispetto a Microsoft, altri importanti esponenti nei servizi BI identificati da Gartner come leader visionari sono Tableau, acquistata nel 2019 da Salesforce, importante azienda di servizi CRM, e Qlik, azienda svedese incentrata in particolare sui servizi di DataVisualization.

Conclusioni

L'aumento della produzione di dati in combinazione al progresso tecnologico renderà l'utilizzo degli strumenti di analisi un fattore chiave nella competizione tra i business e un potente mezzo di supporto per le scelte strategiche dei governi. Il prerequisito per accedere a questi mezzi è la digitalizzazione della propria organizzazione. Effettuata questa transizione, un primo strumento che non può mancare per monitorare le performance dei processi è la Data Visualization.

Nella presente tesi l'attenzione è stata focalizzata su questa pratica, ad oggi fondamentale nella Business Intelligence, che consente un'analisi di statistica descrittiva per l'esplorazione di dataset che si distingue dagli altri strumenti in quanto accessibile anche da profili non specializzati. La Data Visualization permette direttamente agli utenti che necessitano di risposte di effettuare le analisi, pur non disponendo di competenze statistiche o informatiche avanzate, interagendo con le dashboard.

Sulla base delle ricerche effettuate e da quanto appreso durante l'esperienza di tirocinio, si può dedurre che l'intento dei provider di servizi di data visualization sia quello di potenziare il valore fornito da questi strumenti mantenendo invariata, se non aumentando, la facilità di utilizzo attraverso l'aggiunta di nuove funzionalità. In secondo luogo, si cerca di permettere a questi servizi di fornire risposte in tempo reale o ad ogni modo di minimizzare i tempi di aggiornamento. Vista però la diminuzione del costo delle memorie e l'aumento della velocità di calcolo derivanti dal progresso delle infrastrutture tecnologiche questo obiettivo sarà raggiunto senza particolari sforzi da parte del settore dei software. Oltre alla semplicità di utilizzo e rapidità negli aggiornamenti, tra gli altri fattori che determinano la buona progettazione di un cruscotto per il monitoraggio bisogna considerare infine il contenuto, in termini di organizzazione dei dati, scelta dei KPI e chiarezza di presentazione, questa volta a cura delle aziende di consulenza che predispongono i sistemi informatici per i propri clienti.

Nel documento si è sempre fatto riferimento a un contesto aziendale, ma non è difficile trasporre l'applicazione di questi strumenti ad aspetti della vita sociale e individuale. Per esempio, esistono già dashboard interattive sui propri parametri vitali misurati da *wearable technologies*, o dashboard di domotica.

Per concludere, si prospetta un futuro nel quale dato il livello di dettaglio dei dati a disposizione e i poteri degli algoritmi di intelligenza artificiale, i cruscotti che consulteremo ci forniranno delle risposte sempre più precise su quale sia la scelta giusta da intraprendere.

Riferimenti

- United Nations, “Digital Economy Report 2019, United Nations conference on trade and development, value creation and capture: implications for developing countries”, 2019, https://unctad.org/system/files/official-document/der2019_en.pdf
- IDC, “The Digitization of the World From Edge to Core”, 2018, <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>
- Our World in Data, Population, consultato il 25/2/2021, <https://ourworldindata.org/grapher/projected-population-by-country>
- Nedim Dedić, Clare Stanier, “Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. 10th International Conference on Research and Practical Issues of Enterprise Information Systems”, 2016
- Howard Dresner, “The Performance Management Revolution: Business Results Through Insight and Action”, 2007
- Kumar SM, Meena Belwal, “Performance Dashboard - Cutting-Edge Business Intelligence and Data Visualization”, 2017
- Daniel Filonik, “Developing a Dashboard for Real-Time Data Stream Composition and Visualization”, 2012
- OECD Competition Committee, “Big Data: bringing competition policy to the digital era”, 2016, [https://one.oecd.org/document/DAF/COMP\(2016\)14/en/pdf](https://one.oecd.org/document/DAF/COMP(2016)14/en/pdf)
- Harvard Business Review, Bhaskar Chakravorti, Ajay Bhalla, and Ravi Shankar Chaturvedi, “Which Countries Are Leading the Data Economy?”, 2019, <https://hbr.org/2019/01/which-countries-are-leading-the-data-economy>
- Gartner, Inc., “Information technology (IT) worldwide spending from 2005 to 2022 (in billion U.S. dollars)”, 2021, <https://www-statista-com.ezproxy.biblio.polito.it/statistics/203935/overall-it-spending-worldwide/>
- IDC, “Nominal GDP driven by digitally transformed and other enterprises worldwide from 2018 to 2023 (in trillion U.S. dollars)”, 2019, <https://www->

statista-com.ezproxy.biblio.polito.it/statistics/1134766/nominal-gdp-driven-by-digitally-transformed-enterprises/

- Reply, Profilo aziendale Reply, 2020,
https://www.reply.com/InvestorsDocuments/it/Company_Profile_ita.pdf
- Reply, sito ufficiale di Cluster Reply, consultato il 10/01/2021,
<https://www.reply.com/cluster-reply/it/>
- Borsa Italiana, Scheda Reply, consultato il 10/01/2021,
<https://www.borsaitaliana.it/borsa/azioni/scheda/IT0005282865.html?lang=it>
- Reply, Reply annual financial report, 2019,
https://www.reply.com/InvestorsReports/en/2019_REPLY_Annual_Report_EN.pdf
- Apps Run The World, “Top 10 ERP Software Vendors, Market Size and Market Forecast 2019-2024, 2020”, <https://www.appsruntheworld.com/top-10-erp-software-vendors-and-market-forecast/>
- Gartner, Inc., “Main supply chain management software suppliers worldwide in 2019, based on related segment revenue (in million U.S. dollars)”, 2020
- Microsoft Documentation, OLTP (Online Transaction Processing), consultato il 12/12/2020, <https://docs.microsoft.com/it-it/azure/architecture/data-guide/relational-data/online-transaction-processing>
- Hoboken, Wiley, “Dictionary of E-Business”, 2003
- Logical Read, “Choosing Between Rowstore Indexes and Columnstore Indexes”, consultato il 7/01/2021, <https://logicalread.com/choosing-rowstore-indexes-columnstore-indexes/#.YBHI4uhKiUk>
- (EN) Wikipedia, B-Tree, consultato il 27/12/2020,
<https://en.wikipedia.org/wiki/B-tree>
- Microsoft Documentation, Columnstore indexes: Overview, consultato il 19/11/2020, <https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-overview?view=sql-server-ver15>
- Morteza Zaker, Somnuk Phon-Amnuaisuk, Su-Cheng Haw, “An Adequate Design for Large Data Warehouse Systems: Bitmap index versus B-tree index”, 2008
- Golfarelli, Rizzi, “Data warehouse, teoria e pratica della progettazione”, 2006

- The Holistics Blog, The Rise and Fall of the OLAP Cube, consultato il 27/11/2021, <https://www.holistics.io/blog/the-rise-and-fall-of-the-olap-cube/>
- Microsoft Documentation, About using DirectQuery in Power BI, consultato il 3/03/2021, <https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-directquery-about>
- Gartner, Inc., “Augmented Analytics Is the Future of Data and Analytics”, 2017
- Gartner, Inc., “2021 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms”, 2021
- SAS Institute, “Business Intelligence and analytics tools software revenue worldwide by vendor from 2014 to 2018 (in million U.S. dollars)”, 2018, <https://www-statista-com.ezproxy.biblio.polito.it/statistics/473009/business-intelligence-and-analytics-software-revenue-worldwide-by-vendor/>