

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

SCUOLA DI SCIENZE  
Corso di Laurea in Informatica per il Management

**Utilizzo della Business Intelligence  
per il monitoraggio delle performance  
in ambito Fashion Retail**

**Relatore:**  
Chiar.mo Prof.  
MARCO DI FELICE

**Presentata da:**  
ANGELO MUSCIA  
MASUZZO

**Correlatore:**  
Dott.ssa.  
SARA BROLLI

**I Sessione  
2020-2021**



# Introduzione

In questa tesi verrà discusso il progetto che ho effettuato durante il mio tirocinio per tesi presso l'azienda di consulenza informatica Iconsulting. Il progetto è principalmente incentrato sull'utilizzo della *Business Intelligence* nell'ambito del Fashion Retail, con lo scopo di monitorare dati relativi al Sell Out, ovvero la misura delle vendite effettuate dall'azienda cliente verso i consumatori finali, in un arco temporale ben definito. Per motivi di privacy sono stati oscurati tutti i riferimenti al cliente ed ai soggetti e alle aziende che collaborano con esso.

Il mio progetto di tesi si integra all'interno di processi aziendali già consolidati ed operativi al fine di aumentarne il grado di automatizzazione e di ottimizzazione, nonché di permettere il monitoraggio di alcuni KPI (Key Performance Indicator) d'interesse.

Il progetto è stato svolto a stretto contatto col cliente, ricevendo feedback sulle attività svolte e sulle analisi dei dati.

Per riuscire a svolgere le attività richieste, ho frequentato diversi corsi di formazione interni all'azienda, in modo da riuscire ad utilizzare i programmi necessari alla realizzazione delle attività.

Dopodiché ho affrontato vari casi d'uso per rispondere alle specifiche esigenze del cliente.

Il primo caso d'uso consiste nell'automatizzazione del processo di cancellazione di dati all'interno del Data Warehouse aziendale.

Il secondo caso d'uso consiste nell'automatizzazione del processo di inserimento di nuovi negozi all'interno dell'anagrafica del cliente. Per la realizzazione dei primi due casi d'uso, il software utilizzato è stato SAP Data Services.

Il terzo caso d'uso consiste nella realizzazione di dashboard per il monitoraggio dei dati relativi al retail. Queste sono state create utilizzando due programmi differenti (SAP Business Object e Microsoft Power BI), in modo da riuscire a confrontare i due software e permettere al cliente di valutarne il look and feel e conseguentemente di scegliere quello che meglio si adatta alle sue esigenze.

Questa tesi è suddivisa in quattro capitoli, strutturati come segue:

- *Capitolo 1:*

Questo capitolo introduce gli aspetti teorici del lavoro svolto, descrivendo il processo di Business Intelligence, con riferimenti all'evoluzione dello stesso, inoltre vengono descritti tutti gli strumenti (ad esempio i Data Warehouse) e le soluzioni necessarie per riuscire ad eseguire il processo.

- *Capitolo 2:*

In questo capitolo viene descritto il contesto progettuale, ovvero tutto ciò che circonda il progetto. Inizialmente vengono descritti l'azienda presso il quale ho svolto il tirocinio di tesi ed il cliente per il quale il progetto è destinato. In seguito viene descritto il contesto tecnologico, ovvero tutti i software e le tecnologie che sono state utilizzati per portare a termine il progetto. Inoltre in questo capitolo vengono descritti i casi d'uso, ovvero le richieste da parte del cliente. Ogni caso d'uso è stato affrontato singolarmente, ed una volta portato a termine è iniziata la risoluzione del successivo. Questo approccio metodologico ha permesso di avere un focus mirato sulle singole attività, procedendo in maniera *agile* da un caso d'uso all'altro.

- *Capitolo 3:*

Questo è il capitolo nel quale si descrive come sono state realizzate le richieste del cliente. Per ogni caso d'uso sono state descritte le fasi di analisi e di progettazione, ed in seguito come sono state effettivamente implementate.

- *Capitolo 4:*

In questo capitolo si descrivono le conclusioni relative ai casi d'uso, ovvero i risultati ottenuti. Infine per ognuno di essi si descrivono alcuni sviluppi futuri.

# Indice

<b>Introduzione</b>	<b>i</b>
<b>1 Business Intelligence e Data Warehouse</b>	<b>1</b>
1.1 Introduzione alla Business Intelligence . . . . .	1
1.2 Evoluzione Della Business Intelligence . . . . .	2
1.3 Business Intelligence come Processo Aziendale . . . . .	3
1.4 Strumenti e Soluzioni utilizzati in un processo di BI . . . . .	6
1.4.1 Fonti Dati . . . . .	6
1.4.2 ETL . . . . .	7
1.4.3 database Dimensionali e Multidimensionali . . . . .	8
1.4.4 Star Schema . . . . .	8
1.4.5 Cube . . . . .	9
1.4.6 Data Warehouse . . . . .	10
<b>2 Requisiti ed Obiettivi</b>	<b>17</b>
2.1 Iconsulting . . . . .	17
2.2 Cliente . . . . .	18
2.3 Data Provider . . . . .	18
2.4 Contesto Tecnologico . . . . .	21
2.4.1 SAP HANA . . . . .	22
2.4.2 SAP Data Services . . . . .	24
2.4.3 SAP Business Object . . . . .	24
2.4.4 Microsoft Power BI . . . . .	25
2.5 Obiettivi del Progetto . . . . .	25

---

2.5.1	Casi d'Uso . . . . .	26
<b>3</b>	<b>Realizzazione Progetto</b>	<b>27</b>
3.1	Automatizzazione cancellazione parziali di dati per consentirne il rinvio . . . . .	27
3.1.1	Analisi e Progettazione . . . . .	28
3.1.2	Fase Implementativa . . . . .	28
3.2	Automatizzazione dell'inserimento di negozi italiani in anagrafica . . . . .	31
3.2.1	Analisi e Progettazione . . . . .	31
3.2.2	Pianificazione algoritmi per la risoluzione delle problematiche . . . . .	31
3.2.3	Fase Implementativa . . . . .	33
3.3	Progettazione dashboard per la visualizzazione delle incongruenze sui dati di retail del cliente . . . . .	38
3.3.1	Analisi e Progettazione . . . . .	38
3.3.2	Implementazione dashboard . . . . .	38
3.3.3	Dashboard con SAP Business Object . . . . .	38
3.3.4	Dashboard con Microsoft Power BI . . . . .	44
<b>4</b>	<b>Conclusioni e Sviluppi Futuri</b>	<b>49</b>
4.1	Primo Caso d'Uso: Automatizzazione cancellazione parziali di dati per consentirne il rinvio . . . . .	49
4.2	Secondo Caso d'Uso: Automatizzazione dell'inserimento di negozi italiani in anagrafica . . . . .	50
4.3	Terzo Caso d'Uso: Progettazione dashboard per la visualizzazione delle incongruenze sui dati di retail del cliente . . . . .	50
	<b>Bibliografia e Sitografia</b>	<b>51</b>

# Elenco delle figure

1.1	Fasi del Processo di BI . . . . .	3
1.2	Fonti Dati . . . . .	6
1.3	Processo ETL . . . . .	7
1.4	Rappresentazione grafica di uno Star Schema . . . . .	9
1.5	Cubo Multidimensionale . . . . .	9
1.6	Approccio Top-Down . . . . .	11
1.7	Approccio Bottom-Up . . . . .	12
1.8	Modello di Immon - Corporate Information Factory . . . . .	13
1.9	Modello di Kimball - Dimensional Model . . . . .	14
1.10	Stand-Alone Data Mart . . . . .	15
2.1	Diagramma raccolta dati ed inserimento nel Data Warehouse del cliente . . . . .	19
2.2	Architettura database SAP HANA . . . . .	23
3.1	Creazione delle colonne e riempimento della colonna Customer Door Cod . . . . .	35
3.2	flusso di join tra la temple table ottenuta e la tabella delle door. . .	36
3.3	Upsert tra la template table e la tabella reale delle door. . . . .	37
3.4	Dettaglio del Retail dall'Universo dati . . . . .	39
3.5	Query creata automaticamente dal programma. . . . .	40
3.6	Filtro sulle door mancanti . . . . .	41
3.7	Dashboard generale relativa ai dati di Sell Out. . . . .	42

3.8	Dasbhhboard relativa ai dati dello Stock appartenente ad un preciso data provider. . . . .	43
3.9	Join tra due tabelle in una Calcolation View . . . . .	44
3.10	Calcolation View relativa al Sell Out . . . . .	45
3.11	Calcolation View relativa allo Stock . . . . .	46
3.12	Creazione di una colonna calcolata per il filtraggio dei Customer Code mancanti . . . . .	47
3.13	Dasbhhboard relativa ai dati dello Stock appartenente ad un preciso data provider. . . . .	48
3.14	Dasbhhboard relativa ai dati dello Stock appartenente ad un preciso data provider. . . . .	48

# Capitolo 1

## Business Intelligence e Data Warehouse

### 1.1 Introduzione alla Business Intelligence

L'evoluzione tecnologica ed il relativo adattamento ad essa da parte delle Aziende, ha portato ad un notevole incremento di dispositivi in grado di gestire l'automazione di numerose operazioni. Tali dispositivi hanno causato un enorme incremento nella quantità di dati prodotti, destinata a crescere in maniera esponenziale nel prossimo futuro. L'aumento di questi dati ha reso, in informatica, noto il termine di *Big Data*, ovvero i dati prodotti in grande quantità, con notevole quantità in formati differenti.[14] Per riuscire a stare al passo coi tempi e competitive sul mercato, le aziende moderne non possono sottrarsi all'analisi dei dati in loro possesso. Il patrimonio informativo aziendale si è arricchito in volume e dettaglio, di conseguenza per le imprese vi è l'esigenza, sempre più preponderante, di dotarsi di tecnologie e risorse che vanno oltre a quelle tradizionali, nonché di una notevole capacità di calcolo. Numerose aziende che operano nella gestione dei dati si sono mosse nella direzione di fornire soluzioni in grado di rendere utilizzabili i dati, in modo da poterne estrarre informazioni utili, tutto ciò grazie ad ingenti investimenti lato hardware o all'adozione di sistemi differenti di memorizzazione. Come i Data Warehouse, ovvero strutture nel quale memorizzare i dati (questi verranno

spiegati nel dettaglio in seguito) o le strutture basate sul cloud. [14]

La **Business Intelligence** è quindi l'insieme di tutti i processi e gli strumenti utilizzati col fine di raccogliere e analizzare dati di natura diversa ed in grandi quantità, consentendo a determinati utenti, tramite applicazioni ad accedere ad uno o più Data Warehouse, così da riuscire a svolgere operazioni strategiche ed analisi produttive. Inoltre permette alle aziende che ne fanno utilizzo di ridurre i costi dovuti alle infrastrutture IT nonché di far risparmiare tempo agli utenti ed ai fornitori di dati, questo grazie ad una distribuzione dei dati più efficiente. [15]

## 1.2 Evoluzione Della Business Intelligence

L'Evoluzione della Business Intelligence può essere classificata in tre differenti fasi: [10]

- Fase 1.0.
- Fase 2.0.
- Fase 3.0.

Ogni fase è caratterizzata dall'evoluzione delle tecnologie a disposizione e dai tipi di analisi effettuate.

La Fase 1.0 è iniziata nei primi anni del 1970, e ricopre principalmente l'analisi dei dati strutturati. Le applicazioni utilizzate sui dati si occupano di Estrarre, Trasformare e Caricare i dati, tramite un processo di ETL (del quale parleremo con dettaglio in seguito). I dati vengono immagazzinati all'interno di database Relazionali (RDBMS) o all'interno di Data Warehouse (DWH). Solo dagli anni '80, tramite l'utilizzo di tecniche di Data Mining ed applicazioni di elaborazione analitica online (OLAP) si è iniziato ad analizzare i dati tramite Report e Dashboard.

La Fase 2.0 è partita all'incirca dagli anni 2000, come conseguenza dell'evoluzione di Internet e del Web. In questa fase le aziende hanno cominciato a trarre valore

dalla raccolta e da”analisi dei dati generati dagli utenti e raccolti tramite applicazioni Web (Es. Cookies). Queste tipologie di dati, hanno permesso la nascita di nuovi modelli di business.

La Fase 3.0 è caratterizzata dall’analisi di dati provenienti da dispositivi mobili (Es. Smartphone) e dai relativi sensori appartenenti ai suddetti dispositivi (Es. Geolocalizzazione).

## 1.3 Business Intelligence come Processo Aziendale

Un obiettivo delle aziende è quello di riuscire ad ottenere dati ed informazioni di valore, in modo da riuscire ad agevolare i processi decisionali.

Le fasi del processo di Business Intelligence si possono identificare in:

- **Prima Fase:** Raccolta dei dati da analizzare.
- **Seconda Fase:** Pulizia, Validazione ed Integrazione dei Dati.
- **Terza Fase:** Elaborazione, Aggregazione ed Analisi dei Dati.
- **Quarta Fase:** Utilizzo dei dati nei processi decisionali.

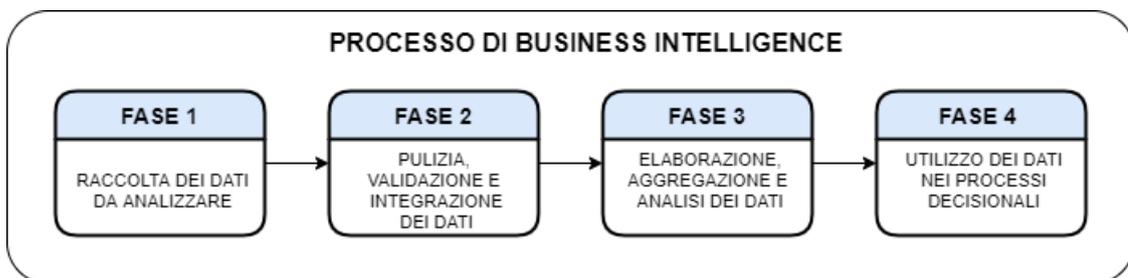


Figura 1.1: Fasi del Processo di BI

**Prima Fase:** I dati da recuperare possono provenire da *fonti esterne* o da *fonti interne* all'organizzazione aziendale. Le fonti interne sono quelle che l'azienda possiede, queste possono essere operazionali (riferite all'attività quotidiana operativa dell'azienda) o possono essere costituite dai dati strutturati per fini analitici, come quelli presenti all'interno di Data Warehouse aziendali. Quelle esterne invece, possono provenire da fonti non facenti parte dell'azienda (Es provider esterni, fonti di dominio pubblico ecc.). Dall'intero patrimonio informativo sarà possibile estrarre *informazioni qualitative* ed *informazioni quantitative*, le quantitative sono fonti basate su un'unità di misura, ed indicano una quantità, le qualitative forniscono un dettaglio o un approfondimento dell'informazione.

**Seconda Fase:** Questa fase ha lo scopo di trasformare i dati raccolti nella prima fase per permettere la loro analisi congiunta. Come conseguenza della varietà delle fonti e dei formati dei dati raccolti, è necessario effettuare operazioni di pulizia, validazione e integrazione dei dati stessi, in modo da riuscire a generare dati congrui tra loro, così da poterli elaborare insieme.

**Terza Fase:** In questa fase, i dati standardizzati vengono aggregati ed elaborati. Tramite l'analisi dei dati è possibile generare dei risultati utili per la fase finale, ovvero quella di utilizzo degli stessi nei processi decisionali.

**Quarta Fase:** In questa fase i Decision Makers, sulla base dei dati sviluppati, svolgeranno i processi decisionali.

### **Analisi dei Dati**

L'analisi dei dati è la fase finale di un processo di Business Intelligence, una volta popolato il Data Warehouse con le informazioni strutturate secondo i criteri desiderati, i *Decision Makers* sono in grado di prelevare le informazioni di loro interesse e di svolgere operazioni di analisi sui dati stessi, col supporto di determinati strumenti. I principali strumenti di analisi sono:

- **Dashboard:**

Le Dashboard sono uno strumento utile per la rappresentazione grafica di metriche e di dati importanti ai fini delle decisioni aziendali. Reagiscono in modo dinamico con l'aggiornamento dei dati presenti nel database dalle quali estraggono i dati.

- **OLAP (On-Line Analytical Processing):**

Le OLAP comprendono tutte le tecniche di analisi che permettono la navigazione di una grande quantità di dati con semplicità e rapidità ed in maniera interattiva. Le principali operazioni OLAP sono: [13]

- Roll-Up
- Drill-Down
- Slice and Dice
- Pivot

Esistono inoltre altre operazioni OLAP, ovvero le operazioni di *drilling*, queste permettono di visualizzare il dato con un maggior dettaglio. Le operazioni OLAP verranno spiegate in dettaglio in seguito.

- **Analisi What-If:**

Tramite queste analisi, attraverso operazioni simulate è possibile determinare differenti scenari futuri, aggiornandoli con la modifica di alcune variabili.

- **Data Mining:**

Nelle operazioni di Data Mining vengono utilizzate tecnologie di Intelligenza Artificiale ed analisi statistiche di vario tipo, automatizzando i processi di ricerca. Grazie a queste tecniche è possibile individuare informazioni di varia natura e non reperibili con altre tecniche di analisi.

## 1.4 Strumenti e Soluzioni utilizzati in un processo di BI

Esistono differenti piattaforme e strumenti in grado di supportare e facilitare un processo di Business Intelligence, in questo modo la comprensione dei dati stessi da parte degli utenti risulterà semplificata. Inoltre esistono anche processi in grado di semplificare la lettura e l'utilizzo dei dati.

### 1.4.1 Fonti Dati



Figura 1.2: Fonti Dati

I dati possono avere provenienze diverse, interne o esterne all'organizzazione che ne fa utilizzo. Le fonti interne possono essere costituite da dati strutturati in maniera eterogenea tra loro, e di conseguenza gli stessi dati possono essere facilmente messi in relazione gli uni con gli altri. Diverso è invece quando i dati provengono da fonti esterne all'organizzazione, questi dati sono spesso strutturati in forme diverse, di conseguenza risulta necessario un processo di conversione del dato stesso. La natura di queste fonti esterne può essere differente, i dati possono

venire da fonti gratuite (Es. ISTAT3) o da fonti private, che raccolgono dati in modi eterogenei.

### 1.4.2 ETL

L'ETL (Extraction, Transformation and Loading) è un processo di Estrazione, Trasformazione e Caricamento del dato. Lo scopo di questo processo è quello di riuscire a prendere dati grezzi da determinate fonti e riuscire a caricarli all'interno di un database in formati che permettono all'azienda un'analisi degli stessi. Il processo di ETL prevede differenti fasi, queste sono *l'Estrazione*, *La Pulizia e Trasformazione* ed il *Caricamento*. [3]

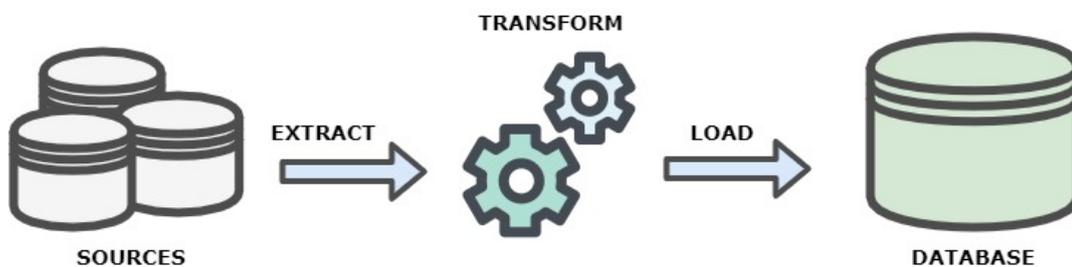


Figura 1.3: Processo ETL

- **Estrazione:** In questa fase i dati vengono presi da diverse fonti, sono strutturati in maniera differente. Questi spesso sono dati grezzi, di conseguenza necessitano, per essere resi compatibili tra loro, di pulizia e trasformazione.
- **Pulizia e Trasformazione:** Una volta estratti, i dati vengono sottoposti a processi di pulizia e trasformazione, in modo da eliminare dati incoerenti ed applicare eventuali logiche di business, costituendo una solida base per le analisi degli utenti finali. In particolare, in questa fase i dati subiscono processi di *Standardizzazione*, nella quale i dati vengono formattati secondo determinate caratteristiche; *Deduplicazione*, dove vengono eliminati i dati che si ripetono; *Verifica*, nella quale viene controllato che i dati rispettino le caratteristiche desiderate e *Ordinamento*, ovvero vengono gestiti ordinamenti e raggruppamenti che permettono di organizzare i dati in maniera efficiente.

- **Caricamento:** In questa fase i dati elaborati vengono memorizzati all'interno di una base dati, tipicamente un Data Warehouse. Le modalità di caricamento possono essere di tipo *Statico*, nel quale i dati vengono caricati all'interno del database, sostituendo di conseguenza quelli già presenti in precedenza, o di tipo *Incrementale*, dove i dati verranno periodicamente aggiornati all'interno del database, in questo modo si vanno a salvare solamente i cambiamenti avvenuti sui dati dall'ultima estrazione effettuata.

### 1.4.3 database Dimensionali e Multidimensionali

I database basati sul modello Entità-Relazione non sono adatti per l'analisi di grosse quantità di dati, di conseguenza è necessario utilizzare un **DFM (Dimensional Fact Model)**. In questo modello i DWH possono essere *Dimensionali* o *Multidimensionali*. [12]

### 1.4.4 Star Schema

Lo Star Schema (Schema a Stella) è una rappresentazione dimensionale di un Data Warehouse. Si compone di due elementi:

- **Fact (Fatti):**

I Fact rappresentano le misurazioni effettive dei dati. Sono contenuti all'interno di una *Fact Table*, ovvero la tabella centrale dello Star Schema con il riferimento a tutte le Dimension.

- **Dimension (Dimensioni):**

Le Dimension contengono il contesto nel quale i dati vengono utilizzati. Sono contenute all'interno delle *Dimension Table*, queste sono le tabelle esterne dello Star Schema, contengono tutte le colonne che rappresentano le dimensioni.

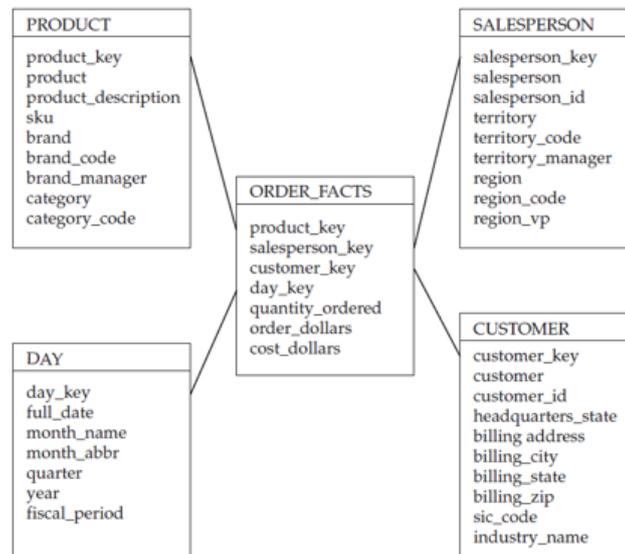


Figura 1.4: Rappresentazione grafica di uno Star Schema

### 1.4.5 Cube

Il Cube è la rappresentazione multidimensionale di un database ed è costruito sulla base dello Star Schema. Il cube viene usato principalmente all'interno di analisi OLAP.

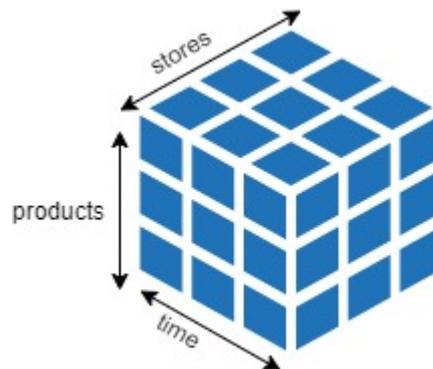


Figura 1.5: Cubo Multidimensionale

Esistono diverse operazioni possibili per l'esplorazione di un database Multidimensionale:

- **Rotation or Data Pivoting** Questa operazione permette di visualizzare i dati da differenti punti di vista.
- **Slice and Dice** Questa operazione consiste nel ridurre il numero delle dimensioni del nostro cubo.
- **Roll-Up and Drill-Down** Questa operazione consente di aumentare (Roll-Up) o diminuire (Drill-Down) il dettaglio dei dati organizzati tramite una gerarchia.

#### 1.4.6 Data Warehouse

I Data Warehouse (DWH) sono la principale struttura nel quale vengono memorizzati dati, organizzati secondo criteri ben precisi. Forniscono un modello di dati ed un punto di accesso a supporto delle analisi e delle decisioni da parte delle aziende. I DWH permettono l'interconnessione con altri sistemi applicativi e devono avere determinate caratteristiche: [13]

- **Orientato al Soggetto:** I Data Warehouse vedono i soggetti al centro dell'organizzazione dello stesso (Es. Cliente, Fornitore, Prodotto e Vendite), concentrandosi solo su ciò che fornisce una visione semplice e mirata degli argomenti d'interesse, escludendo ciò che non è strettamente necessario all'analisi.
- **Integrato:** I Data Warehouse possono contenere dati compatibili tra loro, con la stessa forma e le stesse caratteristiche. I dati provengono da fonti diverse e con caratteristiche differenti, per questo la pulizia del dato risulta fondamentale allo scopo di ottenere un'integrazione totale del dato.
- **Variabile nel tempo:** I Data Warehouse devono fornire e mantenere uno storico dei dati, tramite un elemento temporale presente all'interno dei Data Warehouse stessi.

- **Non Volatile:** In un Data Warehouse sono possibili solamente due operazioni, *Caricamento Iniziale* dei dati e *Accesso* ai dati per la lettura degli stessi.

### Data Mart

I Data Mart sono database generalmente creati sulla base del Data Warehouse, questi sono strutturati in base ad argomenti specifici, in base all'esigenza di chi ne deve fare uso. Possono esserci due approcci per l'implementazione di un Data Mart:

- *Approccio Top-Down:* In questo approccio i Data Mart vengono costruiti sulla base del Data Warehouse, risulteranno quindi sezioni dello stesso.

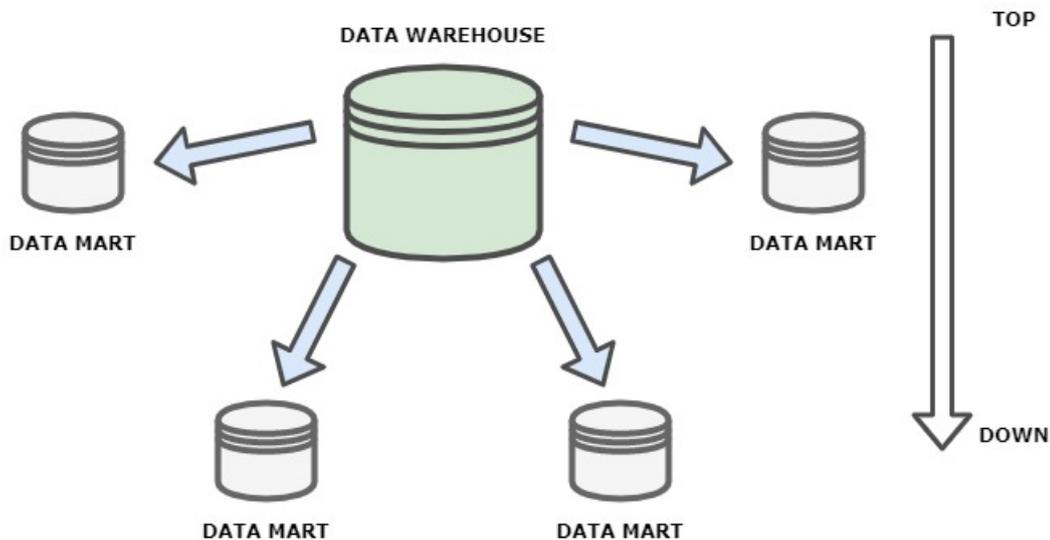


Figura 1.6: Approccio Top-Down

- *Approccio Bottom-Up:* In questo approccio i Data Mart saranno creati in base ai singoli argomenti. L'unione di questi andrà a costituire il Data Warehouse.

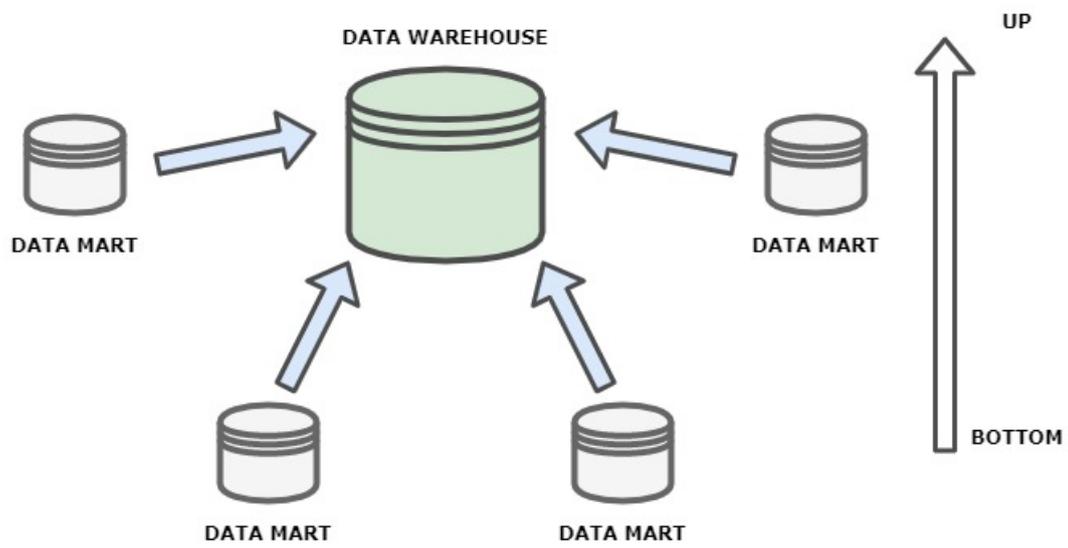


Figura 1.7: Approccio Bottom-Up

## Architettura dei DWH

Esistono diversi modi per strutturare l'architettura di un Data Warehouse, ognuna delle quali diverge dal modo in cui viene utilizzato un design dimensionale.

### Modello di Immon - Corporate Information Factory

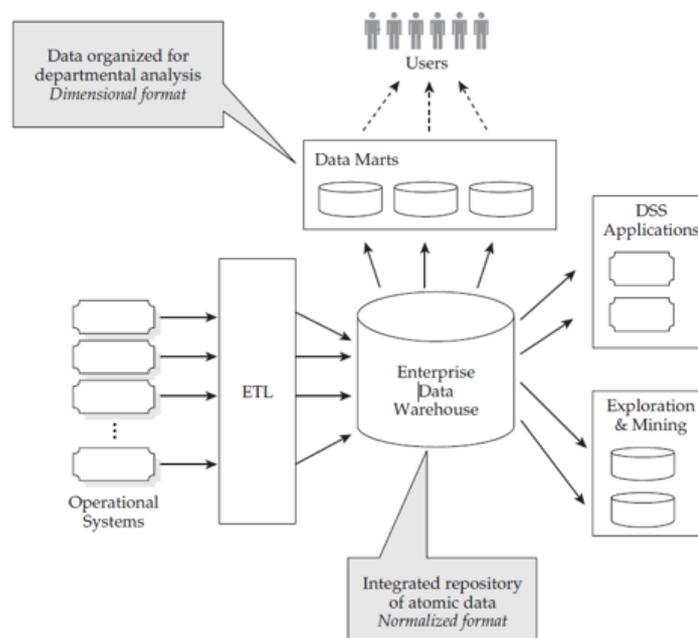


Figura 1.8: Modello di Immon - Corporate Information Factory

In questo modello dopo il processo ETL, i dati verranno caricati all'interno dell' Enterprise Data Warehouse, ovvero un archivio integrato di dati normalizzati. In seguito i dati possono essere utilizzati da altri sistemi o Data Mart, costruiti tramite un approccio Top-Down.

## Modello di Kimball - Dimensional Model

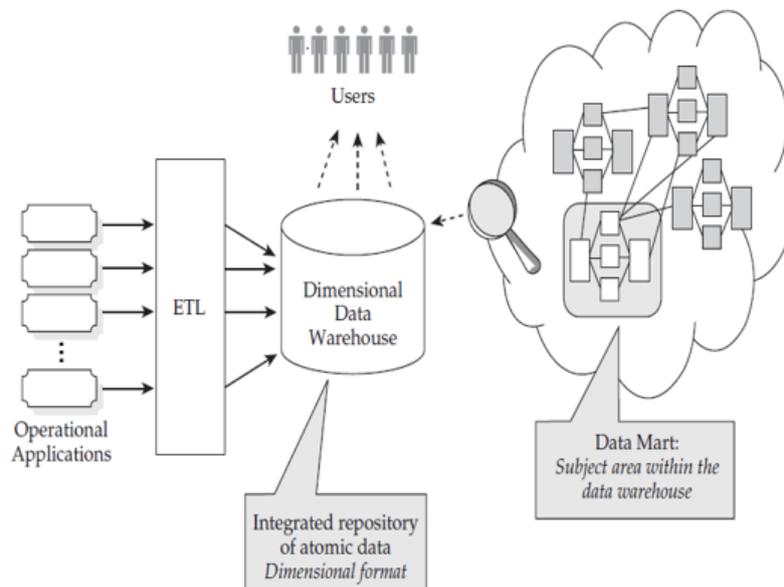


Figura 1.9: Modello di Kimball - Dimensional Model

In questo modello il Dimensional Data Warehouse è strettamente associato allo Star Schema, in quanto il Data Mart stesso è la rappresentazione di una serie di Star Schema. Il Data Warehouse verrà costruito tramite un approccio Bottom-UP.

### Stand-Alone Data Mart

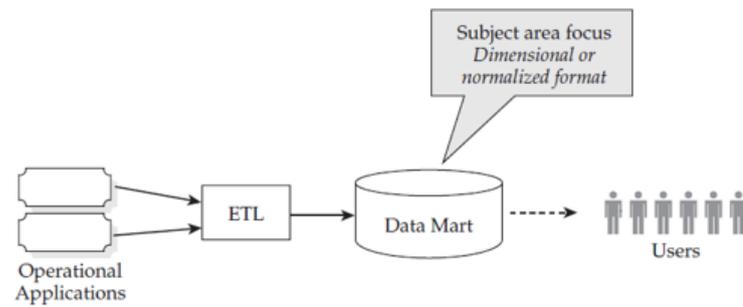


Figura 1.10: Stand-Alone Data Mart

In questa struttura i sistemi alimentano un Data Mart autonomo, che può utilizzare differenti forme di progettazione, il database risulta più rapido e semplice da implementare ed interrogare.



## Capitolo 2

# Requisiti ed Obiettivi

In questo capitolo vengono descritti tutti i requisiti necessari per la realizzazione ed il raggiungimento degli obiettivi del progetto. Inizialmente viene descritta l'organizzazione relativa ai soggetti coinvolti nel progetto in questione, successivamente viene delineato il contesto tecnologico, ovvero tutte le tecnologie utilizzate per la realizzazione del progetto. L'ultima parte di questo capitolo descrive gli obiettivi del progetto sotto forma di casi d'uso. Per la realizzazione del progetto sono intervenute all'interno di un'infrastruttura dati complessa già esistente. Il mio tirocinio è stato della durata di 4 mesi, nei quali ho imparato diverse dinamiche aziendali nonché la funzione di applicativi a me prima sconosciuti.

### 2.1 Iconsulting

The logo for Iconsulting, featuring the word "iconsulting" in a bold, lowercase, sans-serif font. The "i" is lowercase and yellow, while the rest of the letters are black.

*Iconsulting S.p.A.* è un'azienda nata nel 2001, il Core Business dell'azienda è quello di fornire consulenza informatica ad altre aziende. Iconsulting è specia-

lizzata in progetti di *Business Intelligence (BI)*, *Data Warehouse (DWH)* e di *Performance Management (PM)*. Grazie ai Know-How acquisiti tramite l'esperienza negli anni, ad oggi Iconsulting ha all'attivo più di 1000 progetti, per più di 150 clienti.

## 2.2 Cliente

Il progetto vede come cliente un'azienda che opera nell'ambito del *Fashion Retail*, precisamente in quello dell'occhialeria di lusso.

L'azienda è stata creata da una holding multinazionale nel 2014, nonostante inizialmente fosse solo una Start-Up, ha affrontato una crescita esponenziale già dai primi anni.

Il progetto nasce dall'esigenza del cliente di migliorare la gestione e il monitoraggio dei dati di Sell Out e dello Stock presente nei vari punti vendita retail sul territorio. Inizialmente il cliente si occupava solo di operazioni Business-to-Business, ma negli ultimi anni ha iniziato un processo produttivo per la realizzazione del prodotto finale, inoltre è nata la volontà di monitorare l'andamento delle vendite al consumatore finale. Il cliente utilizza una struttura *Full SAP*.

## 2.3 Data Provider

I dati relativi al Sell Out ed allo Stock provengono da differenti data provider, ognuno dei quali fa riferimento a diversi paesi ed a diversi Customer.

Ogni data provider preleva dai retailer i dati e li rilascia ad Iconsulting. Avendo più retailer, i dati che arrivano all'azienda saranno strutturati secondo un tracciato concordato.

I dati che arrivano da ogni fornitore vengono poi integrati all'interno di un Data Warehouse preesistente, e resi disponibili al cliente.

Provenendo da differenti data provider, i dati hanno richiesto un lavoro di standardizzazione ed integrazione, già effettuato al mio ingresso in azienda.

Nel corso del progetto, mi sono focalizzato su tre data provider, ognuno dei quali si concentra sulla raccolta dei dati di Sell Out da un'area geografica specifica:

- **data provider 1:** il primo data provider raccoglie dati dal Nord America.
- **data provider 2:** il secondo data provider raccoglie dati dalla zona APAC (Paesi dell'Asia Pacifica).
- **data provider 3:** il terzo data provider raccoglie dati sull'Italia.

Il cliente raccoglie ed analizza dati provenienti anche da altri data provider che non sono stati coinvolti nel progetto perché già coerenti con le attese del Business. Per questo motivo al momento il progetto di monitoring si è focalizzato sui tre casi d'uso evidenziati.

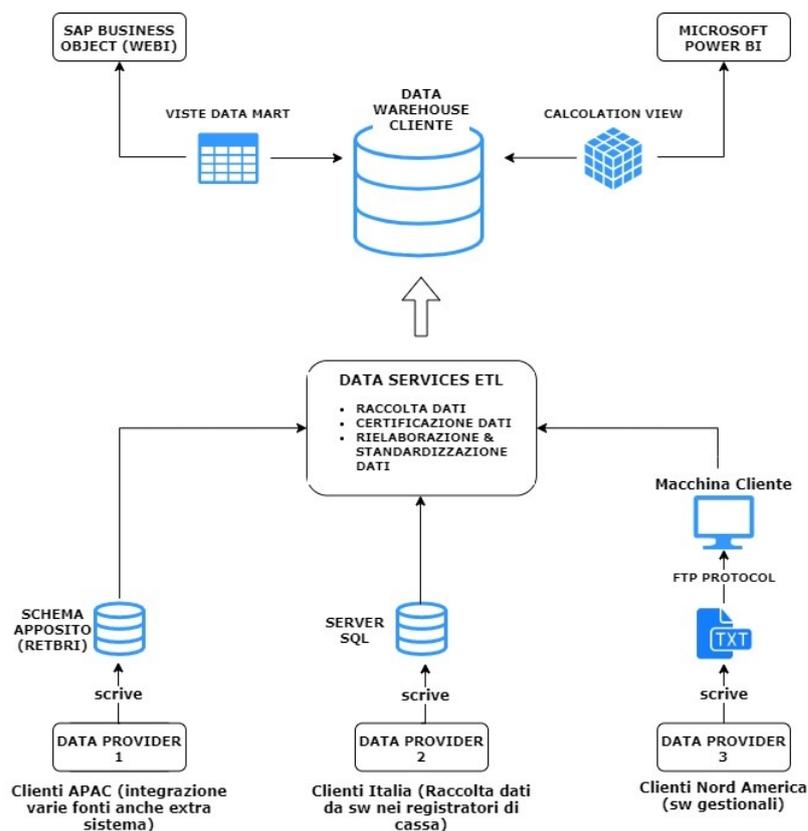


Figura 2.1: Diagramma raccolta dati ed inserimento nel Data Warehouse del cliente

In questa immagine viene descritto il processo con i quali i data provider in questione raccolgono i dati, e come questi vengono immagazzinati all'interno del Data Warehouse.

Il *data provider 1* raccoglie i dati tramite varie fonti (anche extra-sistema) e li scrive all'interno di un apposito database, in seguito tramite il processo di ETL i dati vengono scritti all'interno del Data Warehouse del cliente.

Il *data provider 2* raccoglie i dati tramite un software presente all'interno dei registratori di cassa presenti nei negozi e li scrive all'interno di un database SQL Server, successivamente tramite il processo di ETL i dati, come per il data provider precedente vengono inseriti all'interno del Data Warehouse del cliente.

Il *terzo data provider* raccoglie i dati tramite alcuni software gestionali e li scrive all'interno di una macchina del cliente in formato tramite un protocollo FTP. I dati sono inseriti all'interno di un file in formato *.txt*. In seguito come per i data provider precedenti, tramite il processo di ETL i dati verranno scritti all'interno del Data Warehouse del cliente.

## 2.4 Contesto Tecnologico

Il cliente gestisce internamente solo le operazioni di Sell In, non avendo visibilità diretta sulle vendite al consumatore finale. Negli scorsi anni ha intrapreso un percorso di recupero dei dati relativi alle vendite verso il consumatore finale ed allo Stock in possesso dei retailer.

Per recuperare questi dati, si rivolge ad enti esterni (data provider), con l'obiettivo di raccogliere in un'unica base dati il Sell Out proveniente da diverse parti del mondo.

In seguito alle prime analisi sono emerse diverse incongruenze tra i dati attesi e quelli inviati dai data provider, di conseguenza è stato creato un modello di monitoraggio per cercare di perimetrare i problemi riscontrati e di intervenire verticalmente su ognuno di essi.

La scelta implementativa si è suddivisa in due parti, quella di modellazione del dato (ETL) e la visualizzazione dei risultati ottenuti (data visualization). Per il primo ambito è stato utilizzato SAP Data Services, mentre per la parte di visualizzazione la scelta implementativa si è rivolta in direzione di due diversi strumenti con capacità di data visualization diverse:

- **SAP Business Object**
- **Microsoft Power BI**

Questi strumenti verranno descritti con maggior dettaglio successivamente.

Durante il tirocinio presso l'azienda mi sono stati presentati diversi applicativi per lo sviluppo della tesi. Gli strumenti utilizzati sono principalmente a supporto dei processi di Business Intelligence. Per la realizzazione del progetto sono stati utilizzati i seguenti strumenti:

- **SAP HANA**
- **SAP Data Services**
- **SAP Business Object**
- **Microsoft Power BI**

- Eclipse
- Microsoft Excel

### 2.4.1 SAP HANA



Il Data Warehouse sul quale l'azienda salva i dati per il seguente progetto è memorizzato all'interno di un database *SAP HANA*.

SAP HANA è un database a memoria centrale (i dati risiedono sulla memoria RAM), questa tecnologia permette di elaborare una grande mole di dati in tempo reale, in questo modo le applicazioni possono interagire direttamente con la memoria per accedere ai dati presenti al suo interno, evitando di caricarle ogni singola volta. Questo per le aziende che ne fanno utilizzo porta ad un grosso vantaggio in termini di tempistiche. Presenta le seguenti caratteristiche: [4]

- È formato da una combinazione hardware-software con lo scopo di processare grandi quantità di dati in tempo reale, usando l'elaborazione In-Memory (memorizzazione di grandi quantità di dati nella memoria principale).
- I dati sono orientati alle colonne ed alle righe. Le tabelle orientate a righe memorizzano una tabella in sequenza di righe, mentre le tabelle orientate a colonne memorizzano una tabella in sequenza di colonne.
- I dati sono allocati nella memoria RAM e non sul disco rigido (consentendo una maggiore velocità di elaborazione)

## Architettura di SAP HANA

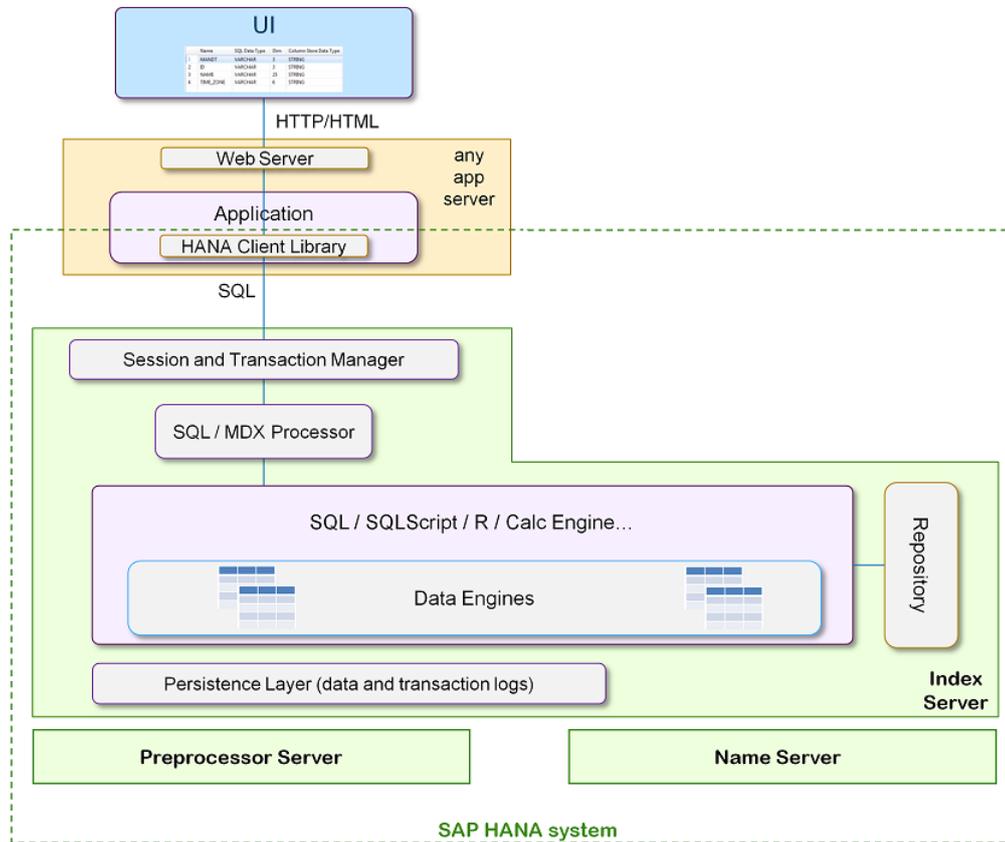


Figura 2.2: Architettura database SAP HANA

Il componente principale di SAP HANA è l'index server, questo si occupa della gestione del database e contiene gli archivi di tutti i dati ed i motori (data engines) per l'elaborazione dei dati.

Il linguaggio di scripting di SAP HANA è SQLScript, questo si basa su funzioni prive di effetti collaterali che operano sulle tabelle usando query SQL per l'elaborazione, risultando quindi parallelizzabile su più processori.

SAP HANA supporta inoltre un framework con la funzione di supportare eventuali installazioni di librerie funzionali, specializzate ed ottimizzate, queste sono integrate con i motori contenuti nell'index server. [5]

### 2.4.2 SAP Data Services



SAP Data Services è uno strumento che permette di creare ed eseguire dei flussi, composti da job, workflow e script. Grazie a questo strumento è possibile compiere un processo di ETL sui dati. Questo software si integra perfettamente con SAP HANA e con molti altri database, permettendo anche di avere come sorgente e destinazione dei flussi ETL file di varia natura (CSV, txt, xlsx etc...).

### 2.4.3 SAP Business Object



SAP Business Object è uno strumento in grado di poter permettere la creazione di report e dashboard per riuscire a rappresentare graficamente i dati, viene utilizzato dopo il processo di ETL per poter effettuare un'analisi sui dati stessi. Nel 2007 la società che ha creato il software, ovvero Business Object è entrata a far parte del gruppo SAP SE.

### 2.4.4 Microsoft Power BI



Microsoft Power BI è un programma proprietario di Microsoft, permette la realizzazione di report e dashboard al fine di supportare un processo di Business Intelligence. Dispone di due interfacce, una desktop (Power BI Desktop) ed una cloud (Services Power BI). Negli ultimi anni, grazie a questo applicativo, l'azienda Microsoft si è affermata tra i leader del settore relativo alla Business Intelligence.

## 2.5 Obiettivi del Progetto

Il progetto che ho svolto durante il mio tirocinio riguarda una porzione dei processi di elaborazione e gestione del dato che l'azienda svolge per il cliente, ovvero la gestione della parte relativa al retail.

Come descritto in precedenza, i dati vengono inviati all'azienda da diversi data provider, ed in forma differente.

L'obiettivo del progetto relativo alla mia tesi si suddivide in parti differenti:

- **Permettere una maggior consapevolezza sui dati ricevuti:** per poter lavorare su determinati dati, bisogna studiare il contesto nel quale ci si trova ed il significato che ogni dato può avere, questa è la *fase di studio* iniziale.
- **Individuazione di dati mancanti o incorretti:** tra i dati ricevuti deve essere effettuato un controllo di eventuali dati mancanti o incompleti, questo perché per poter analizzare un insieme di dati, gli stessi devono avere la stessa struttura.
- **Intraprendere azioni per correggere le situazioni di incompletezza e di errore:** l'obiettivo è quello di riuscire a normalizzare le situazioni di

errore, intraprendendo diverse azioni, questa fase si divide nella iniziale *fase di analisi* e nella successiva *fase implementativa*.

- **Presentazione dei risultati:** In questa fase finale, viene presentato al cliente il lavoro effettuato, riferendo le incongruenze e gli errori riscontrati, facendo un'analisi delle problematiche che non è attualmente possibile risolvere.

### 2.5.1 Casi d'Uso

Essendo il flusso del Retail già attivo, gli obiettivi del progetto sono in funzione del miglioramento e del controllo del flusso stesso. Gli obiettivi del progetto di tesi si possono catalogare come casi d'uso differenti:

- **I caso d'uso, data provider 2:** Automatizzazione cancellazioni parziali di dati per consentirne il rinvio.
- **II caso d'uso, data provider 3:** Automatizzazione dell'inserimento di negozi italiani in anagrafica.
- **III caso d'uso, Generale:** Progettazione dashboard per la visualizzazione delle incongruenze sui dati di retail del cliente.

# Capitolo 3

## Realizzazione Progetto

La prima fase della realizzazione del progetto si è incentrata principalmente sullo studio del cliente e delle tecnologie che sarei andato ad utilizzare per la riuscita del progetto. Per riuscire ad avere una panoramica di tutti i programmi e delle strutture dati utilizzati, le prime settimane sono state dedicate allo studio degli stessi, tramite vari corsi messi a disposizione dall'azienda. Successivamente è iniziato lo studio del cliente, mi è stato spiegato lo scopo del mio progetto e tutta l'organizzazione del cliente stesso. Mi sono stati presentati i data provider che si occupano dell'invio e della raccolta dei dati di Sell Out e Stock, ed il formato dei dati ricevuti.

### **3.1 Automatizzazione cancellazione parziali di dati per consentirne il rinvio**

Il data provider in questione (data provider 1) comunica i dati al cliente scrivendoli direttamente in uno schema apposito del database HANA in cui è presente anche il Data Warehouse aziendale. Il flusso ETL, partendo da tale schema, affina il dato e popola i vari layer del DWH, fino ad arrivare a popolare il Data Mart. Il caricamento segue una logica di *delta*, ovvero le vendite vengono caricate all'interno del Data Mart con operazioni di UPSERT. Tuttavia, i dati mandati potrebbero presentare degli errori e in questo caso non sarebbe sufficiente eseguire

delle cancellazioni sulla sorgente dell'ETL ma è necessario intervenire su tutti i layer del Data Warehouse.

Non avendo il data provider i diritti per accedere ai vari layer del Data Warehouse, ma solo alle tabelle di frontiera tra data provider e Data Warehouse, in caso di necessità di interventi o cancellazioni, lo stesso data provider manda una richiesta ad Iconulting, che interviene manualmente sul Data Warehouse. Queste cancellazioni in precedenza venivano fatte utilizzando degli script SQL, ma essendo aumentata la frequenza di queste richieste è nata la necessità di strutturare un processo standardizzato, creando un flusso ETL ad-hoc, così da automatizzare le operazioni per la rimozione dei dati errati e permettere così un rapido re-invio dei dati corretti.

### 3.1.1 Analisi e Progettazione

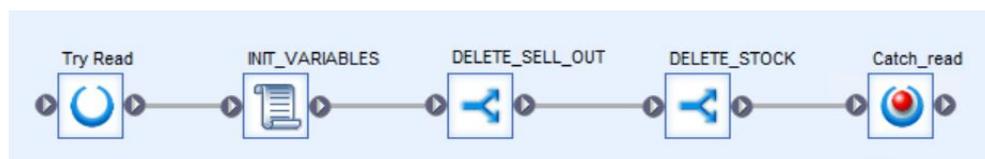
Il primo passo è stato quello di analizzare tutte le operazioni che vengono eseguite per la buona riuscita della cancellazione, questo perché andando il flusso in *delta*, era necessario un intervento su ogni singolo layer del data warehouse, identificando così tutte le tabelle interessate ed accertandosi di cancellare solamente i dati provenienti dal data provider interessato. Una volta comprese tutte le operazioni è iniziata la progettazione del flusso in tutte le sue componenti.

### 3.1.2 Fase Implementativa

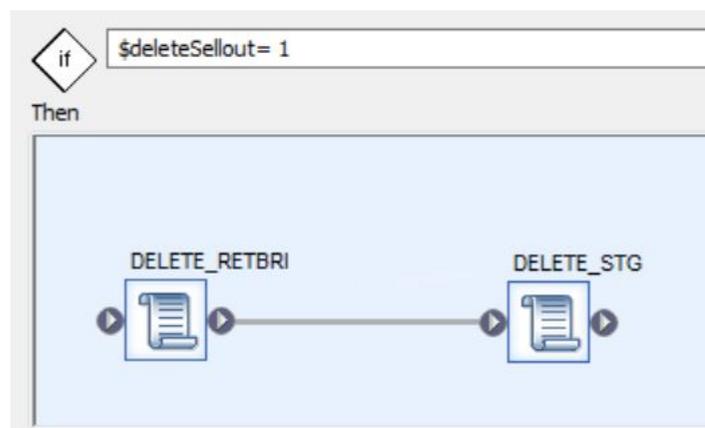
Il flusso ETL è stato sviluppato con SAP Data Services. Questo strumento permette di istanziare una serie di variabili che possono essere utilizzate per parametrizzare le query di cancellazione sul Data Warehouse. Per cui, il primo passo è stato quello di creare una serie di variabili che andranno modificate in base al customer, al periodo da eliminare e alla necessità di cancellare solo i dati di Sell Out o di Stock o di entrambi i tipi (vedi variabili *deleteStock* e *deleteSellout*).

```
Functions... ...  
  
# Variabile riferita al customer code del quale eliminare i dati  
$sold_to_customer_cod = '.....';  
  
$sold_to_customer_cod = replace_substr($sold_to_customer_cod, ',', '\\,\\');  
  
#variabile per scegliere se eliminare sell-out o stock  
$deleteSellout = 1;  
$deleteStock = 0;  
  
#Variabile relativa all'anno a cui la cancellazione fa riferimento  
$year = '2020';  
  
#Variabile relativa ai mesi a cui la cancellazione fa riferimento  
$month = '1,2,3,4,5,6,7,8,9,10,11,12';
```

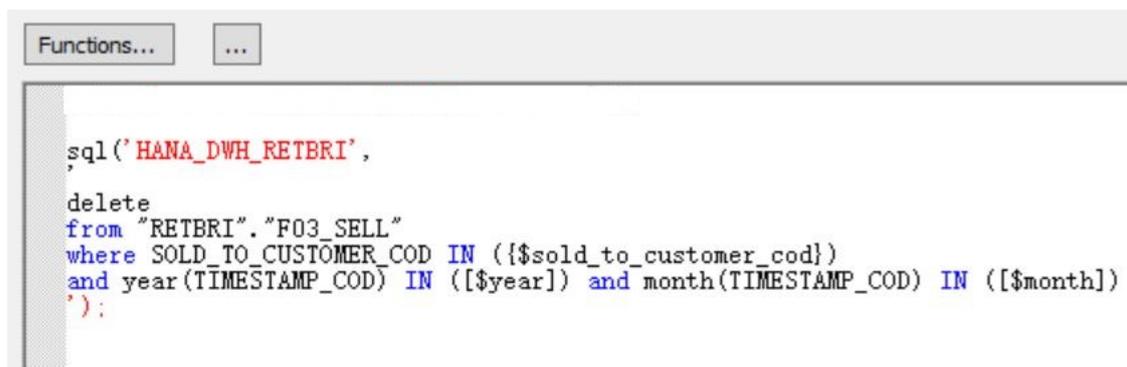
Una volta create le seguenti variabili, sono stati creati all'interno di ogni flusso gli script relativi ai database sui quali eliminare i dati.



Qui vengono creati due Workflow, uno relativo allo Stock ed uno al Sell Out, ognuno di questi contiene degli script che puntano alla tabella dalla quale eliminare i dati.

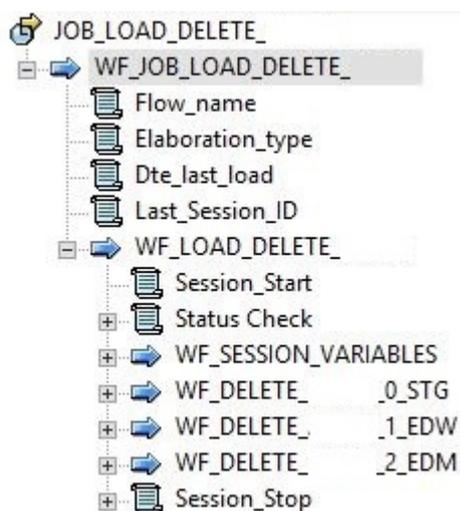


Le cancellazioni dello Stock e del Sell Out dipendono dal valore assegnato alla corrispettiva variabile, se è stata settata a 1 allora tramite un apposito controllo verranno eseguiti i relativi script.



```
sql('HANA_DWH_RETRETI',
delete
from "RETRETI"."F03_SELL"
where SOLD_TO_CUSTOMER_COD IN ({sold_to_customer_cod})
and year(TIMESTAMP_COD) IN ([$year]) and month(TIMESTAMP_COD) IN ([$month])
');
```

Le seguenti operazioni sono state eseguite per tutti i passaggi del flusso.



In questo modo il processo di cancellazione coinvolgerà tutte le tabelle contenenti i dati relativi al customer da eliminare.

## 3.2 Automatizzazione dell'inserimento di negozi italiani in anagrafica

L'obiettivo di questo caso d'uso è quello di automatizzare il processo di inserimento di uno o più negozi all'interno dell'anagrafica del cliente. In precedenza questo processo veniva eseguito in maniera manuale dagli utenti dell'azienda cliente che si occupano della gestione del mercato italiano.

### 3.2.1 Analisi e Progettazione

In questa fase è stata analizzata la richiesta del cliente, il quale chiedeva di automatizzare l'inserimento di nuovi store afferenti ai retailer presenti in anagrafica. Per riuscire a realizzare questa automatizzazione sono state effettuate diverse fasi di studio. In seguito sono state analizzate tutte le tabelle relative a quel relativo data provider e sono state messe a confronto con i campi presenti nell'anagrafica dei negozi.

### 3.2.2 Pianificazione algoritmi per la risoluzione delle problematiche

Successivamente all'analisi ed al confronto di tutte le tabelle, sono stati creati degli algoritmi per la standardizzazione della tabella relativa ai nuovi negozi. In particolare l'idea era quella di partire dalle tabelle dove il data provider comunica Sell Out e Stock per rilevare quali retailer e negozi hanno dati associati, identificando poi quali di questi non sono già presenti in anagrafica. I nostri algoritmi sono stati progettati con l'obiettivo di popolare quanti più campi possibili dell'anagrafica dei negozi partendo solamente dai dati di Sell Out e Stock e dall'anagrafica dei retailer.

#### Algoritmo II Caso d'Uso

All'interno del tracciato delle tabelle di Sell Out e Stock, era presente una colonna chiamata DOOR CUSTOMER COD, costituita dalla concatenazione di

codice cliente e codice negozio. Abbiamo analizzato il tracciato dell'anagrafica dei negozi, anche chiamati Door, per identificare delle regole con cui popolare nuove righe per i DOOR CUSTOMER CODE non già processati precedentemente. Il DOOR CUSTOMER CODE in particolare ha la seguente struttura: CodiceCliente-CodiceNegozio. Abbiamo dovuto quindi spezzare il codice nelle sue due parti (prima e dopo il trattino) per estrapolare le singole informazioni e standardizzare il formato delle stesse (es. sia il codice cliente che il codice negozio devono essere composti da 10 caratteri quindi sono stati definiti dei padding da eseguire sulle stringhe prima e dopo il trattino qualora queste non abbiano la lunghezza desiderata).

DOOR CODE	CUSTOMER_COD + (FFFF per arrivare a 20 caratteri totali) + DOOR_CUSTOMER_COD [ES. 100001 - 100002 = 0000100001FFFF100002]
VALID FROM	01/01/2016
VALID TO	31/12/9999
CUSTOMER CODE	Aggiungere "0" alla stringa prima del "-" fino ad arrivare a dieci caratteri
CUSTOMER DOOR CODE	Aggiungere "F" alla stringa dopo il "-" fino ad arrivare a dieci caratteri
DOOR TYPE COD	Se SOLD TO = CUSTOMER DOOR CODE allora "S", altrimenti "H"
DOOR NAME	SOLD TO NAME preso dalla colonna CUSTOMER_DES_ della tabella STG.CRT_D_CUS_CUSTOMER, andando in JOIN peer il SOLD TO CODE, se non va in JOIN allora "Missing"
SELL OUT TIME BUCKET	"H"
STOCK TIME BUCKET	"H"
FREQUENCY	"H"
STATUS	0

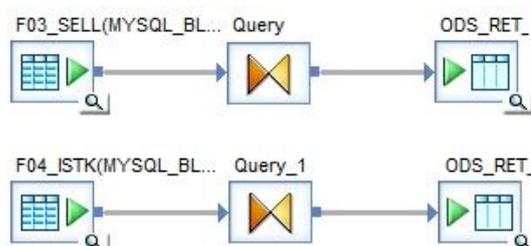
Nella pratica sono stati svolti i seguenti passaggi:

1. Fare una distinct delle combinazioni di retailer e codice negozio riscontrate nelle tabelle di Stock e di Sell Out.

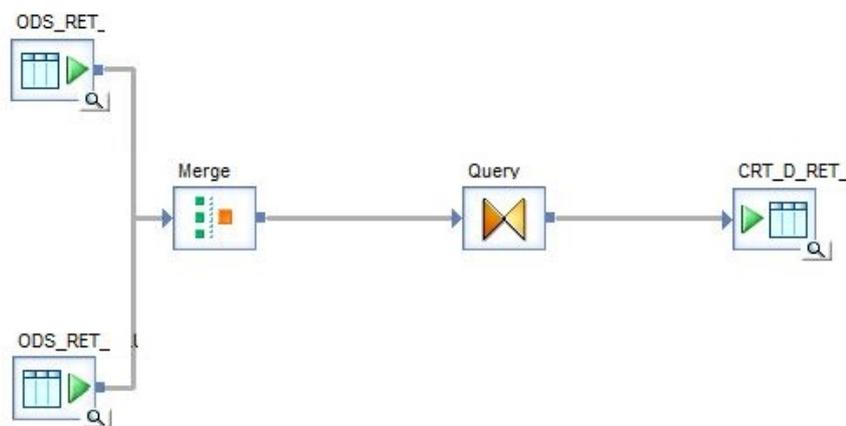
2. Fare una merge (unione) delle distinct trovate al punto 1.
3. Creare una tabella con la distinct del risultato del punto 1, in modo da aver solo chiavi univoche.
4. Eliminare tutti gli spazi dalla colonna Internal Door Code
5. Aggiungere le colonne basandosi sulla tabella D DOOR DOS.
6. *Customer Code*: Fare una substring (sottostringa) della Primary Key della tabella del punto 1, prendendo tutti i caratteri partendo da sinistra fino al carattere '-' (gestire la non presenza del carattere), inserire in seguito n. caratteri '0' fino ad arrivare ad un totale di 10 caratteri.
7. *Customer Door Code*: Fare una substring della Primary Key della tabella del punto 1, prendendo tutti i caratteri partendo dal carattere '-' (gestire la non presenza del carattere) fino alla fine della stringa.
8. *Internal Door Code*: Unire Customer Code e Customer Door Code, nella seconda aggiungere n. caratteri 'F' fino ad arrivare ad un totale di 10 caratteri.
9. Settare i valori preimpostati delle altre colonne (Es Valid From e Valid To).
10. Fare la Left Join tra la tabella e quella delle door (D DOOR DOS), così da passare quelli non presenti nella tabella delle door.
11. Inserire la tabella ricavata in quella delle door.

### 3.2.3 Fase Implementativa

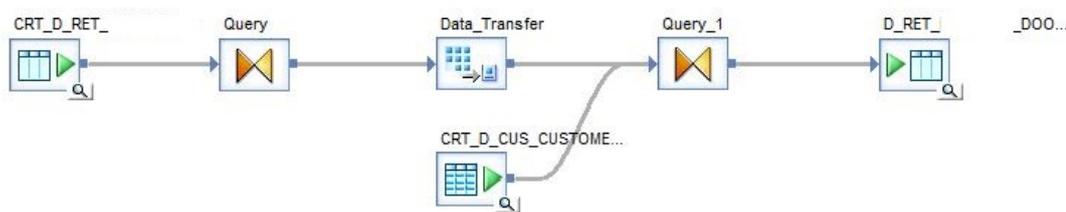
Sulla base dell'algoritmo effettuato nella fase di analisi e progettazione, tramite SAP Data Services, sono state effettuate tutte le operazioni necessarie.



In questo screen tramite una query tutti gli elementi presenti nelle tabelle di Sell Out e Stock verranno messi all'interno di una template table, ovvero una tabella in cui inserire dati in maniera rapida senza dover definire uno schema o importare metadati [9], in modo da non andare a modificare direttamente le tabelle d'origine.



In questo flusso viene effettuata prima l'operazione di merge tra le due template table di Sell Out e di Stock, in seguito tramite una query verrà fatta la distinct di tutti gli elementi, così da non avere duplicati all'interno della template che verrà creata alla fine del flusso. Le righe dovranno essere univoche in quanto questa colonna sarà la chiave primaria della mia tabella.



Nella prima query di questo flusso verranno create tutte le colonne necessarie e a sua volta riempite tramite determinati criteri.

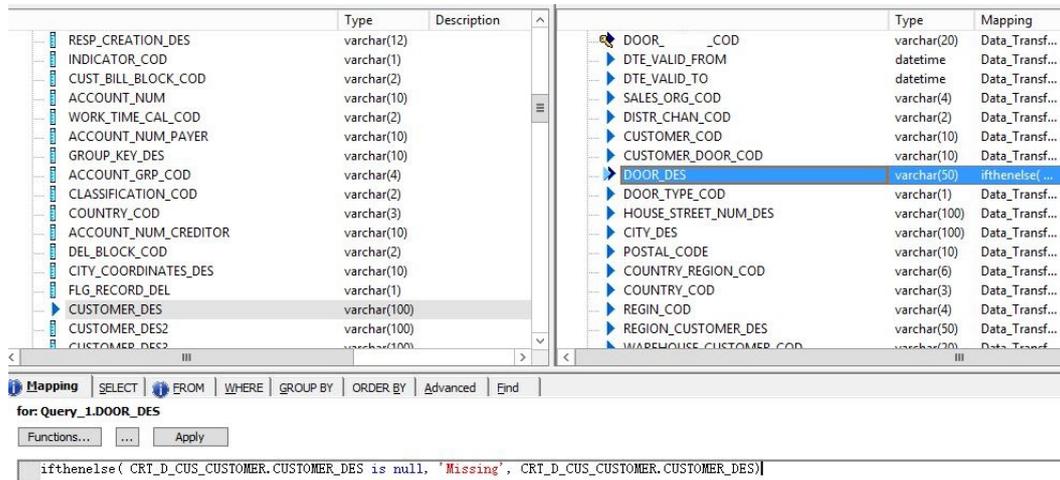
	Type	Descriptio	Type	Mapping
▶ DTE_VALID_FROM	datetime		datetime	to_date('01/...
▶ DTE_VALID_TO	datetime		datetime	to_date('31/...
▶ SALES_ORG_COD	varchar(4)		varchar(4)	'IT03'
▶ DISTR_CHAN_COD	varchar(2)		varchar(2)	'OS'
▶ CUSTOMER_COD	varchar(10)		varchar(10)	ifthenelse(i...
▶ CUSTOMER_DOOR_COD	varchar(10)		varchar(10)	ifthenelse(i...
▶ DOOR_TYPE_COD	varchar(1)		varchar(1)	ifthenelse(if...
▶ DOOR_DES	varchar(50)		varchar(50)	null
▶ HOUSE_STREET_NUM_DES	varchar(100)		varchar(100)	"
▶ CITY_DES	varchar(100)		varchar(100)	"
▶ POSTAL_CODE	varchar(10)		varchar(10)	"
▶ COUNTRY_REGION_COD	varchar(6)		varchar(6)	"
▶ COUNTRY_COD	varchar(3)		varchar(3)	"
▶ REGION_COD	varchar(4)		varchar(4)	"
▶ REGION_CUSTOMER_DES	varchar(50)		varchar(50)	"
▶ WAREHOUSE_CUSTOMER_COD	varchar(20)		varchar(20)	"
▶ F03_SELL_TIME_BUCKET_TYPE_COD	varchar(1)		varchar(1)	'H'
▶ F04_ISTK_TIME_BUCKET_TYPE_COD	varchar(1)		varchar(1)	'H'
▶ TYME_FREQUENCY_COD	varchar(1)		varchar(1)	'H'

Mapping	SELECT	FROM	WHERE	GROUP BY	ORDER BY	Advanced	Find
for: Query.CUSTOMER_DOOR_COD							
Functions...    ...    Apply							
<pre>                     ifthenelse( index(Data_Transfer_1.DOOR_..._COD, '-', 1) is not null,                     substr(Data_Transfer_1.DOOR_..._COD, index(Data_Transfer_1.DOOR_..._COD, '-', 1)+1,                     length(Data_Transfer_1.DOOR_..._COD) - index(Data_Transfer_1.DOOR_..._COD, '-', 1)), Data_Transfer_1.DOOR_..._COD)                 </pre>							

Figura 3.1: Creazione delle colonne e riempimento della colonna Customer Door Cod

Nel secondo blocco di query viene invece effettuata una left join con la tabella delle door, per prendere la descrizione delle door corrispondenti, e mettere il valore 'Missing' nel caso la join non andasse a buon fine.



In questo flusso è stato fatto il left join con la tabella D DOOR DOS, ovvero la tabella contenente l'anagrafica delle door, in modo da creare una template table contenente tutte le door non erano presenti all'interno della tabella delle door, grazie a questo join abbiamo unificato tutte le door all'interno di una singola tabella. Infine, come ultimo passaggio, il contenuto della template table è stato inserito dentro la tabella delle door, andando ad effettuare un UPSERT (inserimento dei nuovi dati ed aggiornamento di quelli già esistenti) all'interno dell'effettiva tabella di anagrafica dei negozi, utilizzata da altri flussi ETL.

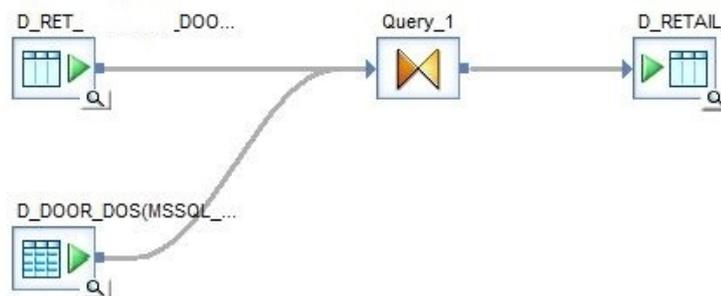


Figura 3.2: flusso di join tra la template table ottenuta e la tabella delle door.



Figura 3.3: Upsert tra la template table e la tabella reale delle door.

### **3.3 Progettazione dashboard per la visualizzazione delle incongruenze sui dati di retail del cliente**

#### **3.3.1 Analisi e Progettazione**

Una volta ricevuti i dati di Sell Out e Stock da tutti i data provider, è iniziata la fase di studio dei dati stessi. In questa fase con l'aiuto del tutor aziendale, ho analizzato le strutture e i singoli formati dei dati da noi ricevuti, individuando eventuali mancanze o errori. Una volta individuate le problematiche abbiamo strutturato una dashboard per la resa grafica:

- Dei dati ricevuti correttamente.
- Dei dati incongruenti.
- Dei dati mancanti.

#### **3.3.2 Implementazione dashboard**

In questa fase ho creato delle dashboard da proporre al cliente, per mostrare le mancanze che nonostante il processo di ETL basato sugli algoritmi da noi implementati, non è stato possibile risolvere per mancanza di informazioni. Queste dashboard sono state create in due formati diversi e tramite due strumenti differenti. Le prime sono state implementate tramite SAP Data Services, mentre le seconde sono state implementate tramite Microsoft Power BI.

#### **3.3.3 Dashboard con SAP Business Object**

Dopo aver standardizzato i dati tramite un flusso di ETL, ho creato delle dashboard da presentare al cliente. Nelle seguenti dashboard verranno raffigurati dati relativi al Sell Out ed allo Stock. Le dashboard in questione sono state realizzate tramite SAP Business Object, le fasi della creazione sono le seguenti:

### Scelta dei dati da rappresentare

In questa fase ho scelto i dati più consoni da rappresentare, in modo da semplificare l'analisi al cliente, focalizzandomi solamente sui dati strettamente necessari.

### Creazione di un sample delle dashboard

Una volta scelti i dati ho creato un sample su Microsoft Excel per organizzare graficamente la dashboard, scegliendo i grafici e le tabelle adatte alla presentazione.

### Estrazione dei dati dall'universo

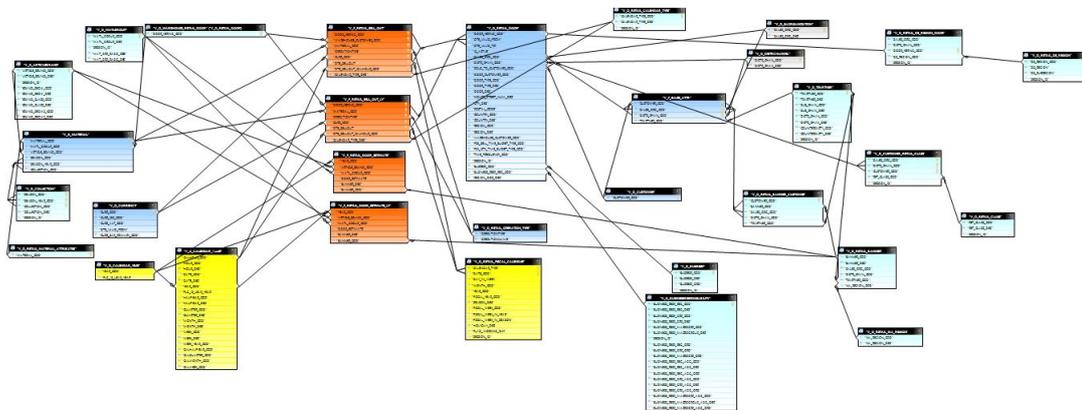
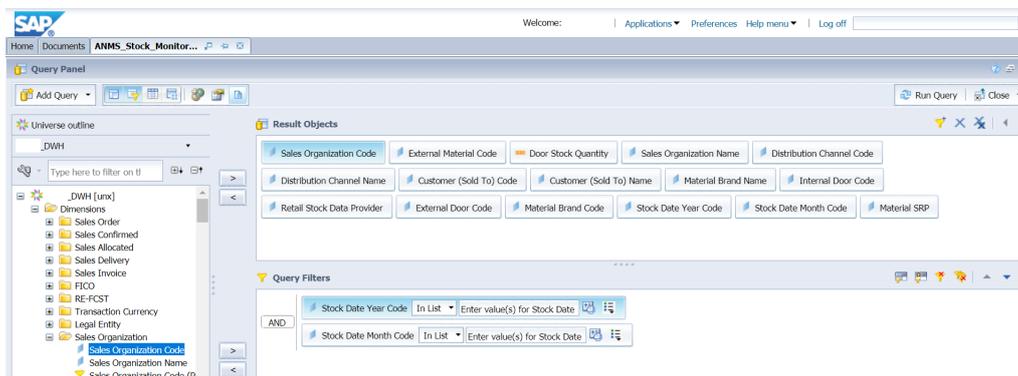


Figura 3.4: Dettaglio del Retail dall'Universo dati

In Web Intelligence, per riuscire a creare le dashboard, bisogna estrarre i dati dall'universo, formato da un insieme di metadati costruiti sulle *fact table* e *dimension table* dei Data Mart.



In questa fase ho estratto tutte le dimensioni e le misure necessarie per la creazione delle dashboard. Una volta estratte ed impostati i filtri principali, come il periodo nel quale analizzare i dati, ho eseguito la query, questa verrà automaticamente creata dal programma una volta selezionati tutti gli elementi (metadati). Una volta eseguita si potranno creare effettivamente le dashboard.

```

1 SELECT
2   Table_28."SALES_ORG_COD",
3   Table_123."RAW_MATERIAL_COD",
4   sum(Table_123."QTY_STOCK"),
5   Table_28."SALES_ORG_DES",
6   Table_27."DISTR_CHAN_COD",
7   Table_27."DISTR_CHAN_DES",
8   Table_17."CUSTOMER_COD",
9   Table_17."CUSTOMER_DES",
10  Table_49."ARTICLE_BRAND_DES",
11  Table_121."DOOR_      _COD",
12  Table_123."RET_DS_ID",
13  Table_121."DOOR_CUSTOMER_COD",
14  Table_49."ARTICLE_BRAND_COD",
15  Table_126."YEAR_COD",
16  Table_126."CALMONTH_COD",
17  Table_16."SRP"
18 FROM
19  "V_D_SAORGANIZATION" Table_28 INNER JOIN "V_D_RETAIL_DOOR" Table_121 ON (Table_121."SALES_ORG_COD"=Table_28."SALES_ORG_COD")
20  INNER JOIN "V_D_CUSTOMER" Table_17 ON (Table_17."CUSTOMER_COD"=Table_121."SOLD_TO_CUSTOMER_COD")
21  INNER JOIN "V_D_DISTRCHANNEL" Table_27 ON (Table_121."DISTR_CHAN_COD"=Table_27."DISTR_CHAN_COD")
22  INNER JOIN "V_F_RETAIL_STOCK" Table_123 ON (Table_121."DOOR_      _COD"=Table_123."DOOR_      _COD")
23  INNER JOIN "V_D_CALENDAR_HOUR" Table_124 ON (Table_123."DTE_STOCK_CALHOUR_COD"=Table_124."CALHOUR_COD")
24  INNER JOIN "V_D_CALENDAR_MONTH" Table_126 ON (Table_124."CALMONTH_COD"=Table_126."CALMONTH_COD")
25  INNER JOIN "V_D_MATERIAL" Table_16 ON (Table_16."MATERIAL_COD"=Table_123."MATERIAL_COD")
26  INNER JOIN "V_D_ARTICLEBRAND" Table_49 ON (Table_16."ARTICLE_BRAND_COD"=Table_49."ARTICLE_BRAND_COD")
27
28 WHERE
29  Table_126."YEAR_COD" IN (2020)
30 GROUP BY
31  Table_28."SALES_ORG_COD",
32  Table_123."RAW_MATERIAL_COD",
33  Table_28."SALES_ORG_DES",
34  Table_27."DISTR_CHAN_COD",
35  Table_27."DISTR_CHAN_DES",
36  Table_17."CUSTOMER_COD",
37  Table_17."CUSTOMER_DES",
38  Table_49."ARTICLE_BRAND_DES",
39  Table_121."DOOR_      _COD",
40  Table_123."RET_DS_ID",
41  Table_121."DOOR_CUSTOMER_COD",
42  Table_49."ARTICLE_BRAND_COD",
43  Table_126."YEAR_COD",
44  Table_126."CALMONTH_COD",
45  Table_16."SRP

```

Figura 3.5: Query creata automaticamente dal programma.

### Creazione delle variabili per filtrare i dati mancanti

In questa fase ho creato delle variabili ad-hoc per riuscire a filtrare i dati secondo le mie necessità. Per ogni dato ho creato una variabile che mi selezionasse i dati mancanti o inesatti.

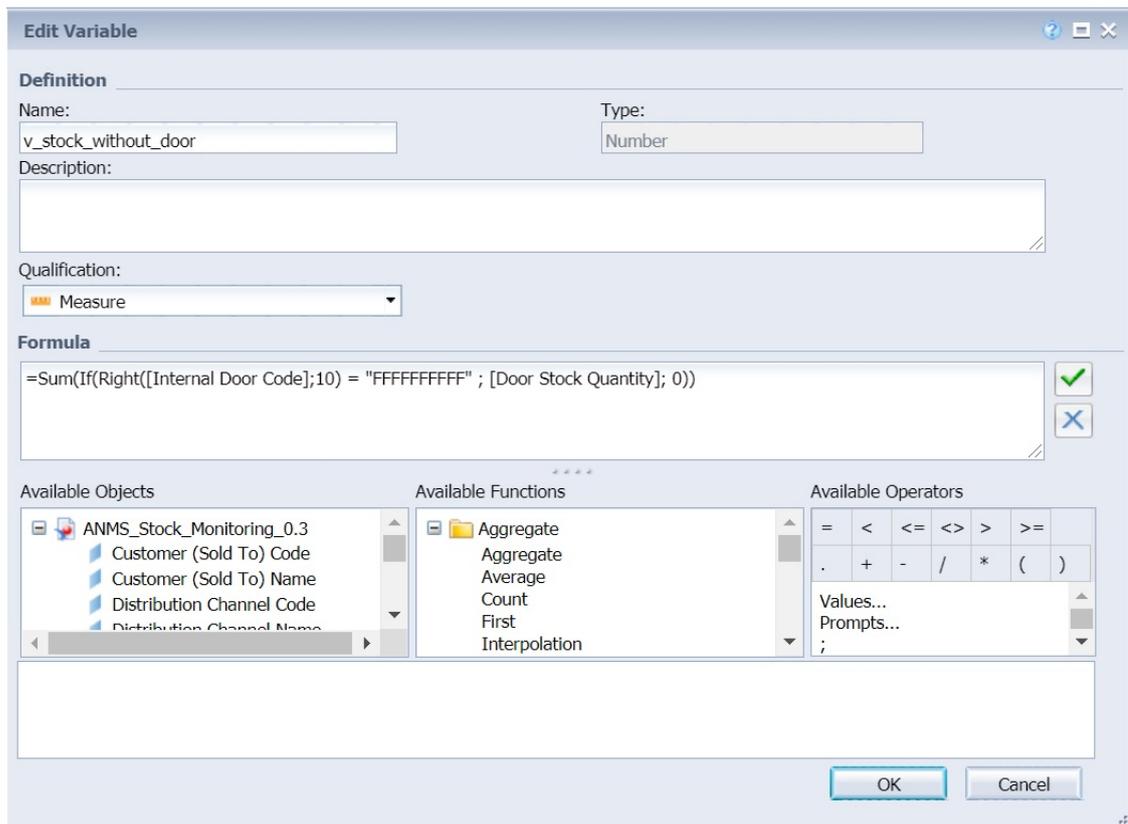


Figura 3.6: Filtro sulle door mancanti

## Dashboard

Una volta filtrati tutti i dati ed aver creato le variabili, ho iniziato a realizzare graficamente le dashboard. Le immagini successive si riferiscono a due dashboard, una riferita al Sell Out ed una allo Stock.

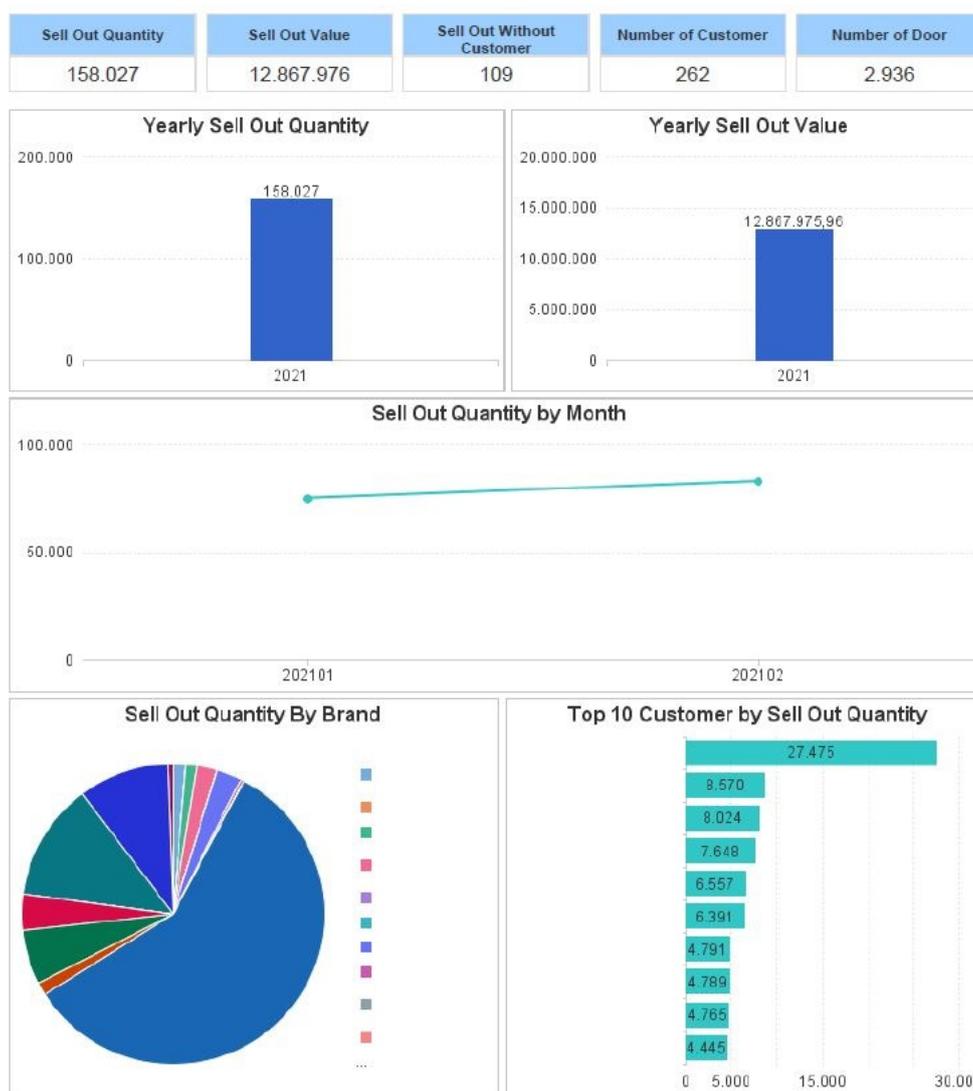


Figura 3.7: Dashboard generale relativa ai dati di Sell Out.

### 3.3 Progettazione dashboard per la visualizzazione delle incongruenze sui dati di retail del cliente



Figura 3.8: Dashboard relativa ai dati dello Stock appartenente ad un preciso data provider.

### 3.3.4 Dashboard con Microsoft Power BI

#### Calcolation View

Per riuscire a filtrare i dati da elaborare tramite Microsoft Power BI è necessario creare delle calculation view, ovvero strutture che possono essere costruite all'interno di SAP HANA e che possono essere importate all'interno di PowerBI. Gli elementi che compongono le calculation view sono i seguenti:

- **Join:** La join viene utilizzata per riuscire a fare la join tra due tabelle. Una volta selezionate le tabelle si possono modificare le caratteristiche della join stessa, come il tipo di join e la cardinalità.

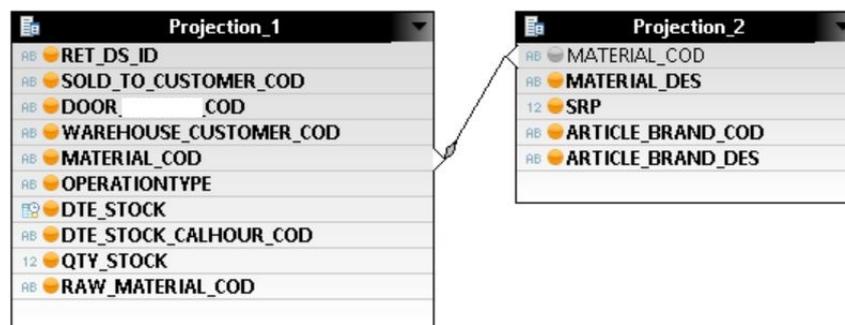


Figura 3.9: Join tra due tabelle in una Calculation View

- **Union:** Questa funzione si comporta come una *UNION ALL* in SQL, di conseguenza utilizzandola, si andranno ad concatenare i risultati di due query uno sotto l'altro, senza aggregarli. [6]
- **Projection:** Tramite la projection è possibile aggiungere, rimuovere, riordinare e rinominare le colonne, prendendo in input una tabella di origine. Una volta selezionata la tabella originaria è possibile selezionare le colonne da tenere ed impostare dei filtri sulle stesse. [7]
- **Aggregation:** L'aggregation permette di applicare aggregazioni tra i dati delle colonne, funziona come il *GROUP BY* in SQL. Le possibili aggregazioni

sono quelle di somma, media, min, max (colonne numeriche) e conteggio (sia colonne numeriche che non numeriche). [8]

### Calculation View

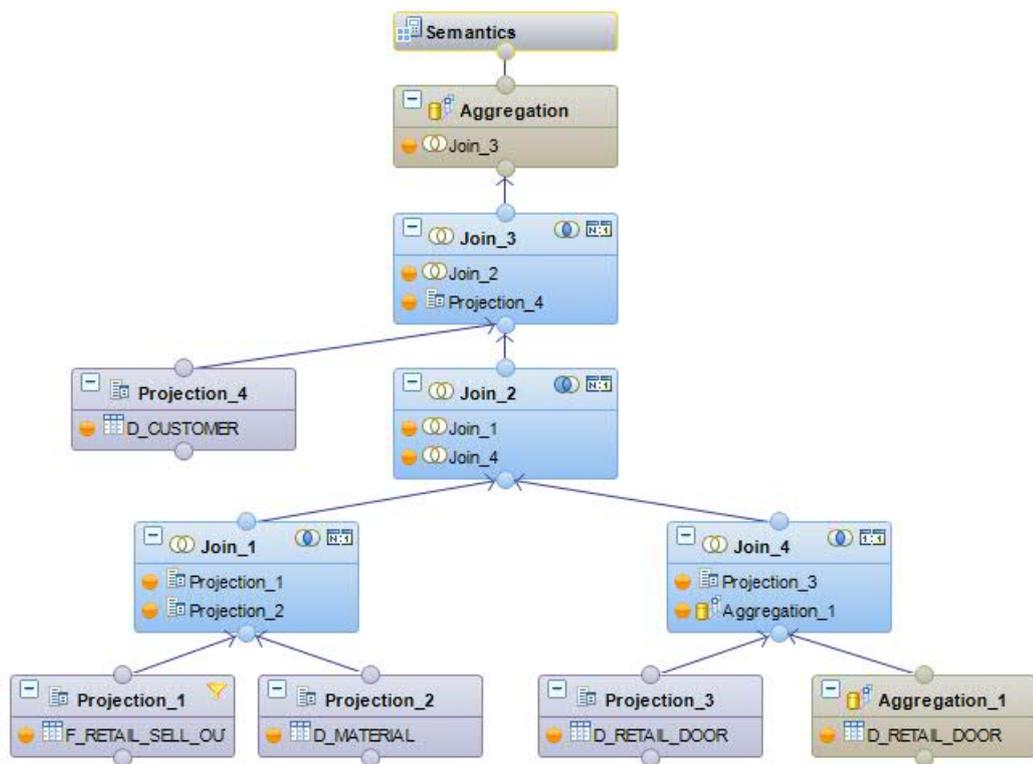


Figura 3.10: Calculation View relativa al Sell Out

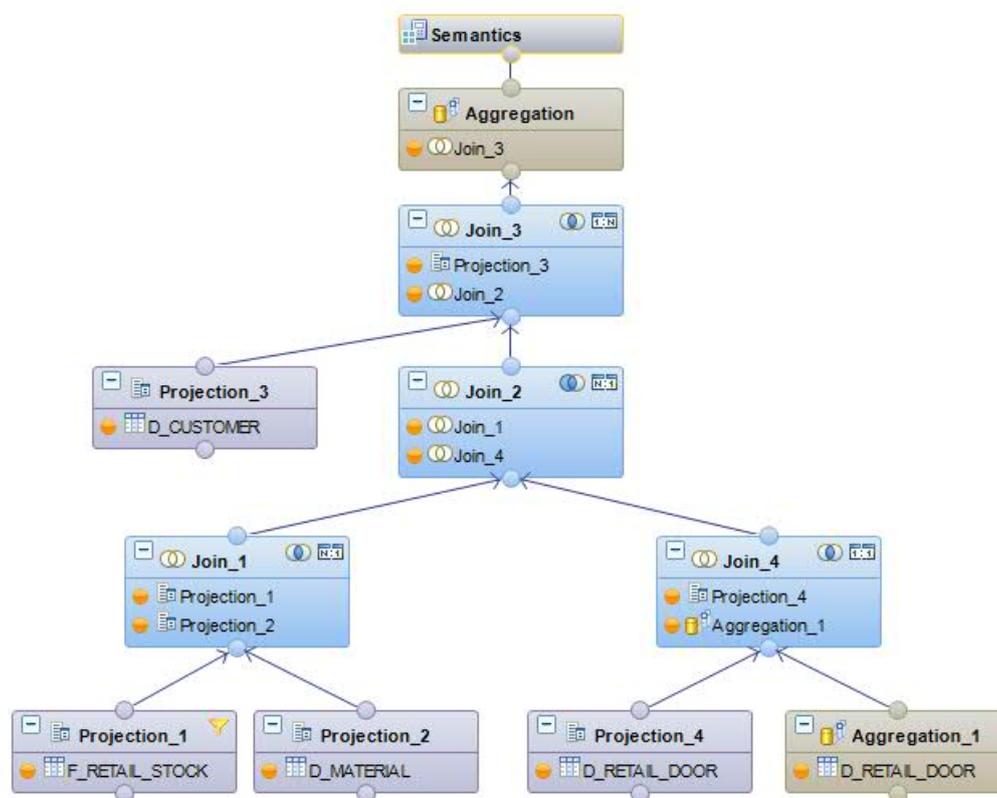


Figura 3.11: Calculation View relativa allo Stock

## Colonne Calcolate

Le colonne calcolate permettono di modificare delle colonne in base a determinate caratteristiche, inoltre è possibile aggiungere delle condizioni tramite l'apposito linguaggio (Column Engine).

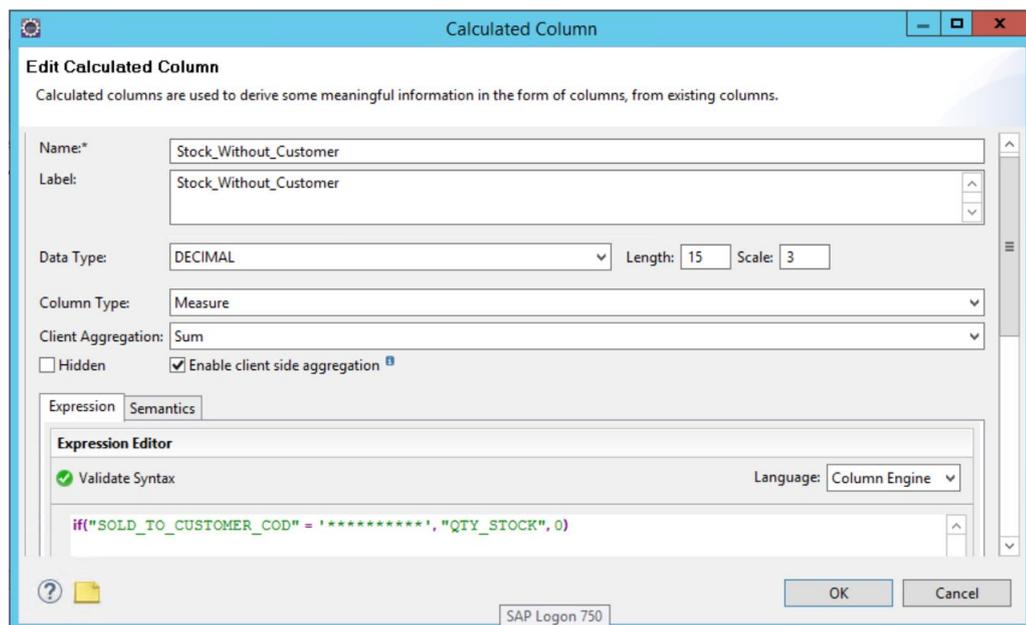


Figura 3.12: Creazione di una colonna calcolata per il filtraggio dei Customer Code mancanti

## Dashboard

Per la creazione della dashboard in Microsoft Power BI, le procedure sono simili a quelle di SAP Business Object. Le immagini seguenti sono riferite al Sell Out ed allo Stock.

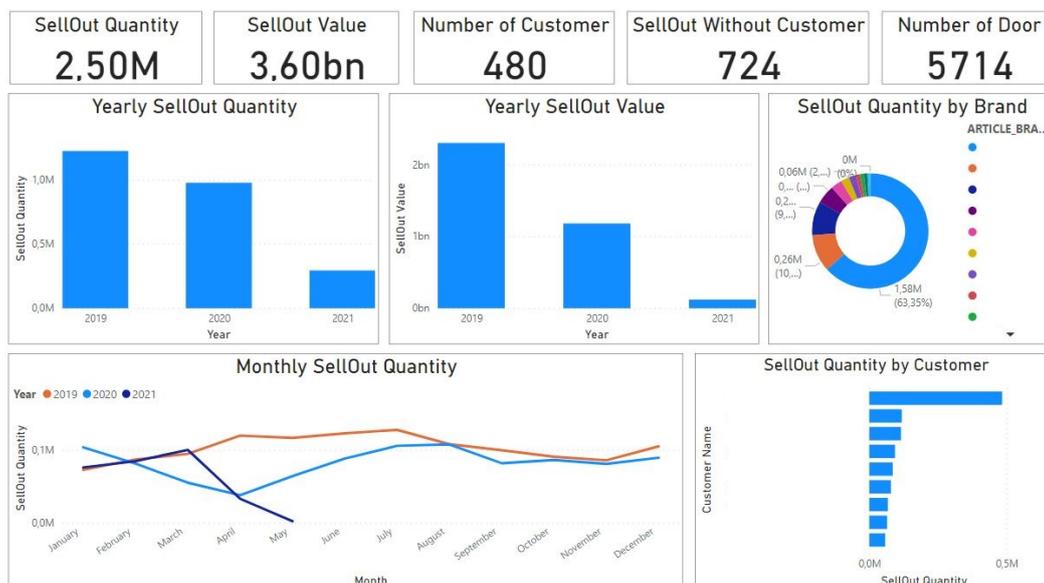


Figura 3.13: Dashboard relativa ai dati dello Stock appartenente ad un preciso data provider.

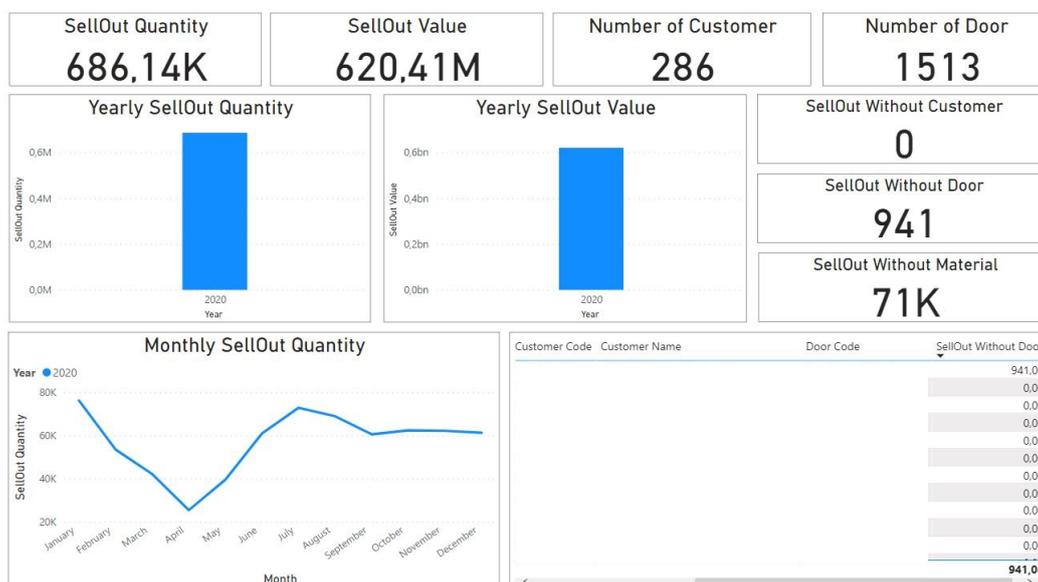


Figura 3.14: Dashboard relativa ai dati dello Stock appartenente ad un preciso data provider.

# Capitolo 4

## Conclusioni e Sviluppi Futuri

Nella realizzazione dei casi d'uso fin qui descritti, in base al controllo sui risultati ottenuti, sono state tratte determinate conclusioni, inoltre per ogni caso d'uso sono state studiati ed individuati alcuni sviluppi per il miglioramento e l'aggiunta di funzionalità aggiuntive che potranno essere implementate in seguito.

### **4.1 Primo Caso d'Uso: Automatizzazione cancellazione parziali di dati per consentirne il rinvio**

Il flusso realizzato per portare a termine questo caso d'uso è stato eseguito senza errori, ed una volta effettuati i check dell'avvenuta cancellazione in tutte le tabelle del flusso, ho constatato che gli elementi cancellati sono quelli desiderati. Il processo di cancellazione è così diventato semi-automatico, permettendo di eliminare i dati desiderati con la sola modifica di alcune variabili. In futuro lo sviluppo potrebbe evolversi, riuscendo a rendere questo processo disponibile al data provider che ne fa richiesta. Questo al momento non è implementato perché il cliente non desidera permettere al Data provider di effettuare operazioni di cancellazione e modifica sul proprio Data Warehouse.

## **4.2 Secondo Caso d'Uso: Automatizzazione dell'inserimento di negozi italiani in anagrafica**

Il flusso ha funzionato come da specifiche, popolando l'anagrafica dei negozi con quelli non ancora inseriti manualmente dagli utenti. Una volta verificato e confermato il corretto risultato dell'ETL, il caso d'uso è stato considerato come concluso. Il processo di aggiunta ora è stato automatizzato, in modo da riuscire a semplificare e velocizzare il processo di inserimento. Allo stato attuale il data provider troverà nell'apposito portale tutti i dettagli relativi ai nuovi negozi inseriti, e questi non possono essere eliminati o modificati. Gli utenti di business del cliente potranno aggiungere dettagli relativi al negozio, come la locazione ed il nome dello stesso, basandosi sui nuovi codici popolati. Un possibile sviluppo futuro potrebbe essere quello di automatizzare anche il processo di inserimento di dettagli relativi ai negozi inseriti tramite il flusso realizzato.

## **4.3 Terzo Caso d'Uso: Progettazione dashboard per la visualizzazione delle incongruenze sui dati di retail del cliente**

Grazie alla realizzazione delle dashboard, per ogni data provider visualizziamo tutti i dati che presentano errori o che non siano stati trovati, questo ci permette di poter effettuare un'analisi periodica e frequente sui dati della dashboard, in modo da riuscire ad intervenire qualora si presentassero degli errori sui dati. Abbiamo notato che gli errori si presentano principalmente per il data provider con il focus sul mercato del Nord America. Si stanno organizzando dei meeting con i business user americani per identificare i motivi per cui i dati inviati dal Data Provider non combaciano con quelli da loro attesi. Un possibile sviluppo futuro sarebbe quello di implementare dei trigger, che si attivano in caso di mancanze o incongruenze rilevate automaticamente.

# Bibliografia e Sitografia

- [1] URL: <https://www.tableau.com/it-it/learn/articles/business-intelligence>.
- [2] URL: <https://www.dataskills.it/cos-e-la-business-intelligence-applicazioni-vantaggi/#gref>.
- [3] URL: <https://www.talend.com/it/resources/what-is-etl/>.
- [4] URL: <https://saprofession.com/cosa-e-sap-hana/>.
- [5] URL: <https://help.sap.com/viewer/fc5ace7a367c434190a8047881f92ed8/2.0.03/en-US/d8d14ef53e244f6cabd10dd3f5e8c11e.html>.
- [6] URL: <http://teachmehana.com/sap-hana-union-node-calculation-view/>.
- [7] URL: <https://help.sap.com/viewer/c8a54ee704e94e15926551293243fd1d/cloud/it-IT/912f74056c3f4647821731a84680fc3a.html>.
- [8] URL: <https://help.sap.com/viewer/c8a54ee704e94e15926551293243fd1d/cloud/it-IT/328d28fa8e324f759ff87751efc3b89e.html>.
- [9] URL: <https://help.sap.com/viewer/8092b085a68941f6aaa6708685a62b0d/4.2.13/en-US/576a163d6d6d1014b3fc9283b0e91070.html>.
- [10] Alberts J Eggert M. «Frontiers of business intelligence and analytics 3.0: a taxonomy-based literature review and research agenda.» In: *Business Research* 13 (lug. 2020), pp. 685–739. DOI: 10.1007/s40685-020-00108-y.

- [11] Daniela Grigori et al. «Business process intelligence». In: *Computers in industry* 53.3 (2004), pp. 321–343.
- [12] iConsulting. *Star Schema*.
- [13] Jian Pei Jiawei Han Micheline Camber. *DATA MINING: Concepts and Techniques*. 2000. Cap. 3,4,5.
- [14] Alessandro Rezzani. *Big Data: Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*. Maggioli Editore, 2013.
- [15] Hugh Watson e Barb Wixom. «The Current State of Business Intelligence». In: *Computer* 40 (ott. 2007), pp. 96–99. DOI: 10.1109/MC.2007.331.