

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Informatica per il Management

**SOLUZIONE REAL-TIME ANALYTICS
IN AMBITO PRODUCTION BASATA
SU PIATTAFORMA SAP HANA**

Relatore:
Chiar.mo Prof.
Marco Di Felice

Presentata da:
Stefano Rotoloni

Correlatore:
Dott. Ing.
Tommaso Tesini

**Sessione III
Anno Accademico 2017/2018**

| | |
|---|-----------|
| Introduzione | iv |
| 1. Concetti Base | 1 |
| 1.1 <i>Business Intelligence</i> | 1 |
| 1.2 <i>Data Warehouse</i> | 3 |
| 1.3 <i>Data Mart</i> | 4 |
| 1.4 <i>Architettura di un Data Warehouse</i> | 4 |
| 1.5 <i>OLTP vs OLAP</i> | 6 |
| 1.6 <i>Modello Multidimensionale</i> | 8 |
| 1.6.1 <i>Concetti chiave di un DFM</i> | 9 |
| 1.6.2 <i>Star Schema</i> | 9 |
| 1.6.3 <i>Operazioni nel modello Multidimensionale</i> | 10 |
| 1.7 <i>Real-Time Analytics</i> | 11 |
| 1.8 <i>Data Visualization</i> | 13 |
| 2. Analisi e Progettazione | 14 |
| 2.1 <i>Metodologia Scrum</i> | 15 |
| 2.2 <i>Risultati di Analisi</i> | 17 |
| 2.3 <i>Mock-up dei Report</i> | 18 |
| 3. Implementazione | 21 |
| 3.1 <i>SAP HANA</i> | 22 |
| 3.2 <i>SAP FMS</i> | 29 |
| 3.3 <i>SAP Cloud</i> | 29 |
| 3.4 <i>SAP HANA Live</i> | 30 |

| | | |
|-------|---|-----------|
| 3.4.1 | <i>Calculation View</i> | 32 |
| 3.4.2 | <i>Implementazione delle Calculation View</i> | 35 |
| 3.4.3 | <i>Sperimentazione</i> | 37 |
| 3.5 | <i>SAP Business Object</i> | 38 |
| 3.5.1 | <i>Information Design Tool</i> | 40 |
| 3.5.2 | <i>Report Finali</i> | 41 |
| | Conclusioni e Sviluppi Futuri | 43 |
| | Bibliografia | 45 |
| | Sitografia | 46 |

Introduzione

Man mano che diventiamo una società digitale, la quantità di dati creati e raccolti cresce e accelera in modo significativo. L'analisi di questi dati diventa una sfida per gli strumenti analitici tradizionali che fanno sempre più fatica a stare al passo. È necessaria quindi una costante innovazione per colmare il divario tra i dati generati e i dati che possono essere analizzati in modo efficace.

I grandi strumenti e le tecnologie di dati offrono opportunità e sfide nel poterli studiare in modo proficuo per comprendere meglio le preferenze dei clienti, ottenere un vantaggio competitivo sul mercato e far crescere il loro business.

Le architetture di gestione dei dati si sono evolute dal tradizionale modello di data warehousing ad architetture più complesse che soddisfano requisiti differenti, come l'elaborazione in tempo reale e in batch, dati strutturati e non strutturati, transazioni ad alta velocità, e via dicendo.

In questa tesi andrò ad analizzare le procedure con cui un'azienda può svolgere analisi operative e strategiche in modo rapido attraverso una reportistica operativa con dati estratti in Real-Time, senza la pulizia dei dati tipica dei processi ETL in un Data Warehouse.

Iconsulting Spa, nata nel 2001, è una società di consulenza informatica specializzata in sistemi a supporto delle decisioni. In particolare, si occupa di progetti di Business Intelligence, Data Warehouse, Enterprise DW & Big data Lake e Performance management.

Grazie a tali sistemi, il cliente ha la possibilità di visualizzare le informazioni e conseguentemente di prendere una decisione più consapevole basata su fatti. Pertanto, Iconsulting supporta il cliente nella conoscenza delle proprie performance e lo aiuta ad incrementarle attraverso decisioni migliori. Il sostegno spazia su un ampio portfolio clienti in diversi ambiti di applicazione.

Per lo svolgimento del mio progetto di tesi, sono stato inserito in un team che segue un'importante azienda di occhialeria di lusso, la quale risponde alle esigenze analitiche in ambito Fashion.

Per motivi di privacy il nome dell'azienda cliente è stato omesso e mi riferirò ad essa in modo generico. Per lo stesso motivo non inserirò dati e documentazione sensibile riguardanti l'azienda.

Il cliente si occupa di gestire il settore occhialeria in campo "Fashion Retail". Si tratta di una start-up lasciata da una holding multinazionale responsabile delle vendite di prodotti da un gruppo mondiale; nonostante sia stata avviata da pochi anni, ha conosciuto una crescita esponenziale fin dai suoi inizi.

Al fine di massimizzare lo sviluppo dei propri marchi, la holding ha deciso di internalizzare la catena del valore per le attività dedicate al settore dell'occhialeria che riguardano la gestione delle vendite.

Anche attraverso questo progetto, la holding sta mettendo a punto una forma innovativa di gestione delle operazioni di analisi strategiche e operative che consente di cogliere appieno il potenziale di crescita dei suoi brand, all'interno di un mercato globale considerevole e molto competitivo, in cui il segmento di mercato sta godendo di una crescita sostanziale.

Lo scopo del progetto è stato quello di fornire una soluzione basata sul sistema SAP FMS, che verrà poi adottata dal cliente per un impianto di recente acquisizione che produce occhiali da sole di alta qualità.

La soluzione è stata progettata in base ai modelli di processo dell'attuale soluzione SAP già in uso presso il cliente, utilizzando le funzionalità standard SAP FMS.

In questo modo è stato possibile garantire le attività di analisi operative e strategiche attraverso report Real-Time e parallelamente consolidare le regole per la costruzione dei report finali. Quest'ultima fase precede la costruzione del Data Mart nell'ambito della Produzione e sarà successivamente integrato nell'Enterprise Data Warehouse già esistente, comprendente gli ambiti di *Sell-In* e *Sell-Out*.

Di seguito andrò a presentare brevemente e schematicamente i tre capitoli in cui si compone il mio elaborato. L'illustrazione dell'intero lavoro svolto è stato possibile anche grazie all'ausilio di concetti teorici fondamentali

Il primo capitolo può essere considerato un'introduzione ai concetti fondamentali della Business Intelligence e della Real-Time Analytics; in questo modo si potrà avere una panoramica generale dei concetti applicati per svolgere il lavoro.

La parte centrale del mio elaborato, che corrisponde con il secondo capitolo, sarà dedicato alla descrizione delle analisi condotte e della metodologia usata da cui sono partito per lo sviluppo del mio progetto di tesi (o lavoro). Infine, il terzo e ultimo capitolo è quello in cui andrò a delineare nel dettaglio il contesto applicativo in cui si è svolto il progetto, osservando gli strumenti utilizzati, analizzando le caratteristiche e scelte implementative. Questo capitolo risulta il più importante, in quanto verrà descritto il mio contributo al progetto, ovvero lo sviluppo di un metodo per fare reportistica su dati live, attraverso due fasi principali: una *bottom-up*, sull'analisi delle tabelle sorgenti per la creazione delle Calculation View, e una *top-down*, partendo dai mock-up dell'utente per costruire le query su SAP BO per la visualizzazione dei report finali.

1. Concetti Base

1.1 *Business Intelligence*

Con il termine Business Intelligence (BI) ci si riferisce ad una serie di processi aziendali che ruotano attorno ai dati, con operazioni di raccolta, elaborazione, analisi, cui scopo è quello di produrre informazioni al servizio del management strategico e tattico, che trova supporto analitico, storico e previsionale alle proprie decisioni. La BI è stata collocata altresì nel sottoinsieme operativo, poiché sta assumendo un ruolo sempre più importante anche nelle normali attività giornaliere delle aziende¹ [7].

Nell'ambito lavorativo moderno – in particolare nel campo della consulenza – le aziende, il cui scopo principale è fare business e diventare leader di mercato, si trovano sempre più frequentemente a confrontarsi con realtà differenti dalla propria. Ciò avviene tramite l'analisi del comportamento di aziende facente parte dello stesso settore e lo studio del mercato in cui si trovano.

L'adozione della BI da parte delle imprese permette una conoscenza più approfondita non solo di loro stesse ma anche del mercato di riferimento. Nel periodo attuale, il “cambiamento” è all'ordine del giorno, pertanto saper leggere in anticipo le tendenze dei mercati è un fattore competitivo a cui non si può e non si deve rinunciare.

Data l'elevata mole di dati generata ogni giorno, diventa necessario trovare un metodo che:

- Permetta di raccogliere e processare dati ad alta velocità (sempre più spesso si parla di processi Real-Time);

¹ Definizione più completa in [7]:

La business intelligence è un sistema di modelli, metodi, processi, persone e strumenti che rendono possibile la raccolta regolare e organizzata del patrimonio dati generato da un'azienda. Inoltre, attraverso elaborazioni, analisi o aggregazioni, ne permettono la trasformazione in informazioni, la loro conservazione, reperibilità e presentazione in una forma più semplice, flessibile ed efficace, tale da costruire un supporto alle decisioni, tattiche e operative.

- Fornisca un servizio di pulizia del dato stesso, eliminando dati sporchi, ridondanti o errati (processi ETL²);
- Definisca un sistema consolidato e stabile di memorizzazione per i dati certificati (Data Warehouse);
- Trasformi l'informazione in fonte di conoscenza attraverso analisi di business sui dati stessi, determinando nuovi KPI su cui fare nuove analisi.

Un processo con tali caratteristiche esiste ed è noto da molti anni, ma non è mai stato sfruttato come adesso che la sua applicazione permette alle aziende di ottenere risultati operativi concreti, attraverso Sistemi di Supporto alle Decisioni (DSS) sempre più efficienti.

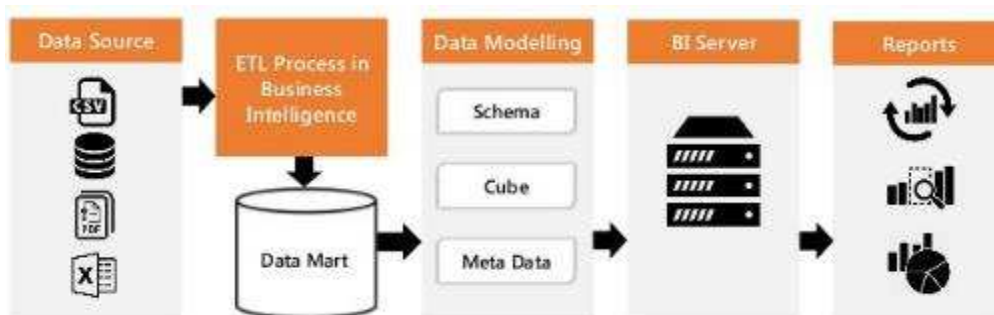


Figura 1: Attività tipiche nel processo di Business Intelligence

² “Extraction, Transformation & Loading” è il processo che preleva i dati dai sistemi alimentanti (ERP, fogli Excel etc.) e li porta nel DWH certificandoli attraverso processi di Data Quality

1.2 Data Warehouse

I Data Warehouse (DWH) sono il principale strumento a supporto della Business Intelligence. Essi permettono di collezionare dati integrati, consistenti e certificati, afferenti a tutti i processi di business dell'azienda e provenienti dalle fonti operazionali. Questi dati vengono in seguito opportunamente trasformati attraverso procedure ETL e controllati attraverso il sistema di Data Quality.

La Qualità dei dati è un requisito fondamentale per l'intero sistema informativo. In quanto, se i dati risultano sporchi, possono oltre che causare un peggioramento delle performance aziendali, portare a prendere decisioni inopportune, comportando costi aggiuntivi e perdita di opportunità.

L'obiettivo di un DWH è pertanto quello di supportare il “*knowledge Worker*” (dirigente, amministratore, gestore, analista) per aiutarlo a condurre analisi finalizzate all'attuazione di processi decisionali e al miglioramento del patrimonio informativo, e fornire un unico punto di accesso per tutti i dati dell'azienda – resi consistenti e affidabili attraverso i processi di ETL. Il DWH garantisce inoltre una profondità storica completa dei dati, poiché in esso viene persistito anche lo stato passato delle informazioni permettendo così un'analisi temporale.

Dovranno quindi essere attentamente progettati per gestire in maniera efficiente ed efficace le caratteristiche dei Big Data³ [1].

I DWH sono realizzati come principale base per i Decision Support System (DSS)⁴.

“Un sistema di supporto alle decisioni è un sistema in grado di fornire chiare informazioni agli utenti, in modo che essi possano analizzare dettagliatamente una situazione e prendere le opportune decisioni sulle azioni da intraprendere in modo facile e veloce” [8].

³ Con il termine Big Data si intende una collezione di dataset così grandi e complessi che sono difficilmente processabili con normali database relazionali.

⁴ Il DSS si appoggia su dati di uno o più database, spesso organizzati in strutture diverse con dati non omogenei.

In altre parole, un sistema di questo tipo deve supportare le attività di analisi e controllo manageriale di routine, le attività di ricerca delle cause di un problema (*focused search*) e le attività di gestione manageriale complessa (*decision making*), permettendo inoltre un facile utilizzo ad un'utenza con un tempo disponibile ridotto e riluttante verso nuove tecnologie (soprattutto nei casi in cui non riesce a percepire in breve tempo i benefici).

È possibile che un DWH sia suddiviso in diversi Data Mart, ognuno dei quali specifico per un solo processo di business fra quelli presenti all'interno dell'azienda (ordini, vendite, clienti, marketing, etc.)

1.3 Data Mart

Generalmente un Data Mart viene estratto da un DWH, ma talvolta può essere costituito anche in assenza di un sistema di dati integrato.

Nello specifico, un Data Mart è un database analitico progettato per incontrarsi con le esigenze specifiche di un'impresa. Essendo sottoinsieme logico o fisico di un Data Warehouse di dimensioni maggiori, segue le stesse regole di progettazione con dati aggregati a vari livelli di dettaglio [7].

L'implementazione può essere di due tipi:

- *Top-Down*: costruzione del DWH, e conseguente aggregazione ed esportazione nei vari Data Mart.
- *Bottom-Up*: concentrandosi su aree specifiche del business si costruiranno i vari Data Mart per poi giungere alla costruzione del DWH. In questo modo si avrà un approccio scalabile.

1.4 Architettura di un Data Warehouse

In fase di progettazione risulta fondamentale stabilire quali tipologie di architettura adottare. Chiaramente, da quando sono stati idealizzati, i modelli (descritti successivamente) si sono evoluti e, di conseguenza, un DWH deve essere costruito secondo i principi moderni [1]. I pattern descritti in questo paragrafo rimangono comunque delle basi da cui partire.

Modello di Inmon - Corporate Information Factory: I DWH si costruiscono nella loro totalità fin dal principio come un unico blocco monolitico; non è possibile vederli come la composizione dei DM. Viene adottata una visione Top-Down (fig. 2).

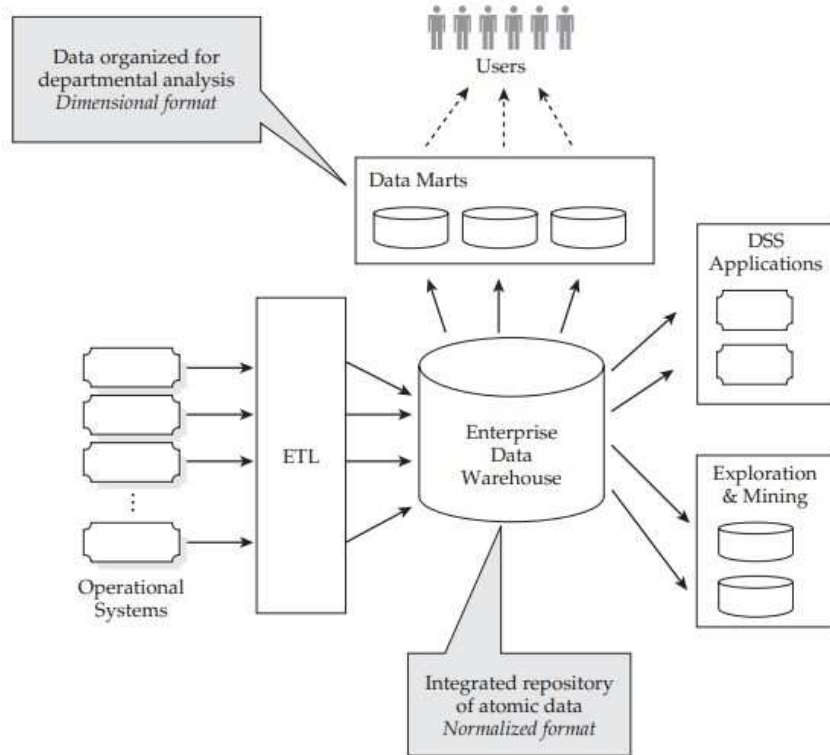


Figura 2: Modello di Inmon

Modello di Kimball - Dimensional Model: adotta un approccio Bottom-up in cui il DWH nasce dall'unione dei vari Data Mart che riferiscono ognuno ad una specifica area di business (fig. 3).

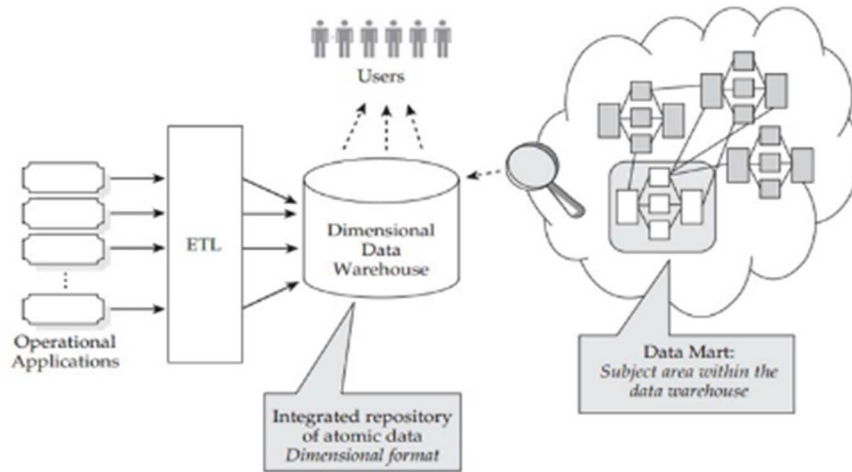


Figura 3: Modello di Kimball

1.5 OLTP vs OLAP

Online Transaction Processing (OLTP)

A livello di database, gli Online Transaction Processing si basano su query multi-access veloci ed efficaci. Le principali operazioni svolte sono INSERT, DELETE e UPDATE in quanto modificano direttamente i dati. Questi ultimi vengono quindi costantemente aggiornati e, di conseguenza, richiedono un efficiente supporto alle operazioni di riscrittura. Una caratteristica fondamentale di questi sistemi è la normalizzazione, la quale fornisce un modo rapido ed efficace per effettuare scrittura nel database.

Online Analytical Processing (OLAP)

L'Online Analytical Processing è un insieme di tecniche software per l'analisi celere e interattiva di grandi moli di dati, con la possibilità di farlo da punti di vista differenti. Questi sistemi si riveleranno molto utili per l'ottenimento di informazioni di sintesi. Queste ultime avranno lo scopo di supportare e migliorare i processi decisionali aziendale. Esempi di strumenti OLAP sono i Data Warehouse, i Cubi Multidimensionali.

Le maggiori differenze fra i due sistemi sono riportati in tabella [1]:

| | OLTP | OLAP |
|--|--|--|
| Finalità | Supporto all'operatività | Supporto al processo decisionale |
| Modalità di utilizzo | Guidata, per processi e stati successivi | Interrogazione ad hoc |
| Quantità di dati per operazione elementare | Bassa: centinaia di record per ogni transazione | Alta: milioni di record per ogni query |
| Qualità | In termini di integrità | In termini di consistenza |
| Orientamento | Per processo/applicazione | Per Soggetto |
| Frequenza di aggiornamento | Continua, tramite azioni | Sporadica, tramite funzioni esplicite |
| Copertura temporale | Dati correnti | Storica |
| Ottimizzazione | Per accessi in lettura e scrittura su una porzione di dati | Per accessi in sola lettura su tutta la base di dati |

In base alla memorizzazione dei dati, si avranno diverse architetture OLAP, ognuna delle quali con i propri pro e contro [1]:

- **Relational OLAP (ROLAP):** i dati vengono memorizzati in un database relazionale come supporto al motore OLAP. Le analisi multidimensionali vengono tradotte in query, restituendo risultati in forma multidimensionale.
- **Multidimensional OLAP (MOLAP):** si ha sia il database che il motore multidimensionale. Ci saranno difficoltà per operazioni di Drill-Down.
- **Hybrid OLAP (HOLAP):** unisce i vantaggi dei due sistemi precedenti. In particolare, pre-aggrega i dati in sistemi multidimensionali per un'analisi efficiente e veloce, mentre vengono ricercate in un data base relazionale in caso di Drill-Down.
- **Desktop OLAP (DOLAP):** i dati vengono caricati in un sistema client e vengono calcolati dal motore in locale.

1.6 Modello Multidimensionale

Il modello E-R⁵, diffuso per progettare sistemi informativi relazionali, non è adatto per esprimere e analizzare in modo dettagliato grandi moli di dati. Sarà necessaria l'adozione di un nuovo modello concettuale: il **modello multidimensionale** o **DFM (Data Fact Model)** [1].

Attraverso questa modellazione è possibile rappresentare i dati all'interno di ipercubi⁶ che forniscono istantaneamente le dimensioni di analisi e i fatti di interesse legati al business.

Gli scopi principali del sistema sono:

- Fornire supporto al design concettuale;
- Creare un ambiente dove gli utenti possano fare query in maniera intuitiva e formale;
- Favorire la comunicazione tra designer e utenti al fine di formalizzare i requisiti di progetto;
- Costruire una stabile piattaforma di design logico;
- Fornire una documentazione chiara e efficace.



Figura 4: esempio di DFM estratto da Indyco (software sviluppato in Iconulting)

⁵ Modello Entità - Relazione

⁶ Vengono definiti ipercubi proprio perché sono multidimensionali.

1.6.1 Concetti chiave di un DFM

- **Fatto:** concetto rilevante per il processo di Decision-Making. Tipicamente modella una specifica area di business (Vendite, Ordini, Produzione, etc.), ed è caratterizzato da una a più misure;
- **Misura:** rappresenta l'aspetto quantitativo del fatto che risulta di elevata importanza per l'analisi. Proprio dalle *Misure* vengono estratti dei *KPI (Key Performance Indicator)* che guideranno le imprese nelle proprie strategie di business. Alcuni esempi possono essere la Quantità prodotta, il Profitto, e il Prezzo;
- **Dimensione:** rappresenta le coordinate di analisi del *Fatto*. Tra queste possiamo trovare Data, Prodotto, Negozio;
- **Attributo Dimensionale:** è un raggruppamento logico di alcuni elementi di una stessa dimensione. Classi di elementi che consentono all'utente di selezionare i dati per specifiche caratteristiche.

1.6.2 Star Schema

Una volta costruito il DFM, viene implementato lo schema logico. Esso viene rappresentato secondo uno Star Schema (fig. 5), il cui centro è costituito da una tabella dei fatti; le punte della stella rappresentano invece le tabelle delle dimensioni che si diramano dal centro.

Solitamente, le tabelle dei fatti in uno Star Schema sono in terza forma normale (3NF), mentre le tabelle dimensionali sono de-normalizzate [1].

Le caratteristiche principali di uno Star Schema sono le seguenti:

- Struttura semplice di facile comprensione;
- Query molto performanti, perché riducono i join da effettuare tra tabelle;

- Tempo di caricamento dei dati relativamente lungo, perché la ridondanza dei dati dovuta alla de-normalizzazione, provoca l'aumento delle dimensioni della tabella;
- Ampiamente supportato da un gran numero di strumenti di business intelligence

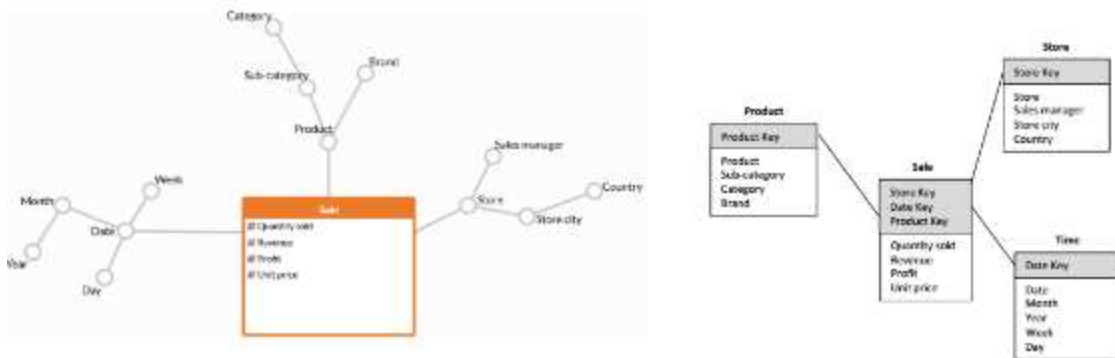


Figura 5: Trasformazione da modello Logico DFM a Tabellare

1.6.3 Operazioni nel modello Multidimensionale

Per navigare all'interno del cubo multidimensionale esistono differenti operazioni che permettono di organizzare i dati al suo interno, attraverso diverse prospettive [1].

- **Pivoting:** permette di modificare rapidamente la visualizzazione dei dati girando gli assi del cubo. Questo consentirà per l'appunto di cambiare il punto di vista da cui si analizza i dati del cubo.
- **Slice & Dice:** seleziona e proietta i dati del cubo. Nello specifico si estrarranno sotto-cubi filtrando su una (Slice) o più (Dice) dimensioni.
- **Roll-Up & Drill-Down:** Queste operazioni consentono di spostarsi all'interno di una gerarchia, scegliendo il livello di aggregazione secondo il quale l'utente desidera analizzare i dati. Nello specifico si salirà di un livello gerarchico con il roll-up, viceversa si userà il drill-down.

1.7 *Real-Time Analytics*

L'analisi in tempo reale si basa spesso su tecniche di processamento automatico dei dati ed estrazione di conoscenza attraverso modelli statistici e derivati dal machine learning, che permettono di fornire informazioni utili sfruttate successivamente per prendere decisioni opportune in modo celere. Per alcuni casi d'uso, con l'espressione "in tempo reale" si intende che l'analisi sarà completata entro pochi secondi o minuti dall'arrivo di nuovi dati [9/11].

Esiste una differenza tra *Real-Time On-demand* e *Continuous*. La prima necessita di una query elaborata dal sistema o dagli utenti per fornire un risultato analitico. La seconda, invece, è più proattiva in quanto avvisa gli utenti o innesca risposte man mano che gli eventi accadono.

Nelle realtà aziendali avere una risposta continua è costoso e spesso innecessario.

Inoltre, bisogna tenere conto che esiste una differenza tra *acquisire* i dati in tempo reale e *utilizzarli* in tempo reale.

I dati catturati dai *data streaming*, vengono trasmessi in un luogo dove possono essere immagazzinati, per far sì che non vengano persi.

Essi quindi, non vengono memorizzati, ma i campioni prelevati a intervalli regolari vengono archiviati per un ulteriore utilizzo.

Questo intervallo può variare da secondi ad anni, oltre che essere guidato da eventi.

La capacità di acquisire ed elaborare dati in tempo reale e usarli insieme a dati che sono stati pre-elaborati non è facile, non tanto a livello tecnico quanto a livello di qualità dei dati stessi.

"Nel 98% dei casi i dati di cui hanno bisogno le aziende sono storici" [10]

Essendo molto difficile determinare la qualità e l'affidabilità in tempo reale, l'elaborazione di molti dati in un DWH permette di avere il tempo di capire qual è la qualità, grazie al confronto con altri dati, e prendere la decisione più adatta in base all'informazione ottenuta.

Il costo di analisi dei dati rispetto al valore che ne deriva, definisce il confine economico che viene raggiunto molto prima del limite tecnologico.

Pertanto, le aziende devono porsi delle domande preliminari:

- Bisogno effettivo di **acquisire** i dati in tempo reale?
- Con che **intervallo** memorizzare i dati? memorizzarli o generare eventi?
- Necessita di **elaborare** i dati in tempo reale?
- Serve anche combinarli con **dati storici**?
- Bisogna **rispondere** a un evento in tempo reale?

(Molto spesso nelle situazioni aziendali, si desidera rispondere a eventi sconosciuti)

Dunque, la quantità di casi d'uso in cui è necessario elaborare o utilizzare dati in tempo reale è limitata. La maggior parte dei requisiti sono espressi da un senso di urgenza causato dalla paura di cambiare le condizioni di mercato, incomprensioni di architetture di dati presso altre aziende o fastidio al time-to-market dell'IT in progetti correlati ai dati.

Iconulting, per conto del cliente, ha svolto queste analisi determinando che una realtà Real-Time permette di mantenere le attività di analisi operative anche senza avere il consolidamento dei dati che si avrebbe da un DWH. Ciò è permesso anche dalle potenzialità dello strumento, già utilizzato dall'azienda, SAP HANA – e, andando più nello specifico HANA LIVE (di cui parlerò più avanti).

Ciò consente all'utente finale l'accesso al prodotto in evoluzione mentre è in fase di sviluppo, ottenendo così feedback in modo incrementale per evolvere e modellare in corso d'opera. Allo stesso tempo vengono consolidate le regole per la costruzione dei report finali (Fast Prototyping⁷).

È indispensabile però tenere in considerazione che all'aumentare della mole di dati questa soluzione non sarà più sostenibile. Proprio per

⁷ In simbiosi con la metodologia Agile, l'idea è quella di generare un prototipo il più rapidamente possibile in parallelo con la capacità del business partner di articolare i requisiti. Una volta che vengono perfezionati si passa ad altri requisiti e così via.

questo, parallelamente, viene definito nello specifico come dovrà essere costruita la parte di Data Mart per il consolidamento e storicizzazione dei dati.

1.8 Data Visualization

Gli strumenti di reportistica rappresentano la parte visibile della BI. Dal momento in cui esistono svariati modi di rappresentare i KPI, diventa importante capire quale sia quello più adatto a rappresentare l'informazione in base al tipo di dato fornito. I fattori che dominano la progettazione devono essere semplicità e chiarezza, oltre che attendibilità e tempestività, per far sì che un report possa essere davvero efficace, permettendogli di svolgere il proprio ruolo quale presentatore della BI[2].

Alcuni esempi di rappresentazione possono essere:

Report: presenta sia tabelle con dati disposti su righe e colonne che grafici con dati su assi cartesiani o su diagrammi a torta. La rappresentazione dei KPI avviene in maniera *statica*.

Dashboard (o cruscotto): è un contenitore di report ed altri elementi di analisi (istogrammi, mappe, diagrammi a torta) che permettere una lettura intuitiva, pertinente e di facile comprensione. La rappresentazione dei KPI avviene in maniera *dinamica*.

2. Analisi e Progettazione

In questo capitolo andrò a descrivere le analisi condotte e la metodologia usata che hanno portato alla conclusione della prima fase di sviluppo, denominata Wave 1, e che è stata da me implementata secondo i requisiti raccolti durante le fasi iniziali.

Iconsulting, attraverso le prime analisi, ha definito il piano complessivo e la metodologia necessaria per collezionare i requisiti e le soluzioni analitiche da implementare per supportare gli utenti del business attraverso report giornalieri e strategy decision-making relative alla Production Industry.

L'obiettivo, come già annunciato, era quello di creare una piattaforma di analisi integrata che potesse rispondere in modo tempestivo alle necessità del business in un Enterprise-wide context (dati comuni del business model, business glossary, business rules, etc.).

Il piano di progetto è stato strutturato in quattro fasi principali. La prima (Design Phase), consiste in interviste agli utenti chiave del business per definire il data model, KPI, requisiti dei report e project roadmap per lo sviluppo di una piattaforma Enterprise Production DW/BI.

Le altre tre fasi, ancora in via di sviluppo, avranno lo scopo di sviluppare agilmente la costruzione del DWH per le analisi della Produzione e nello specifico:

- **Fase 1 (Core Needs):** questa fase seguirà l'implementazione ERP⁸ e durerà circa 2 mesi, durante la sua stabilizzazione;
- **Fase 2 (Foundation):** questa fase sarà implementata dal terzo mese in cui l'ERP sarà operativo (Go Live). Durante questa fase il primo set di report SAP BO sarà rilasciata;

⁸ Enterprise Resource Planning (ERP), sono applicazioni gestionali integrate che creano un unico sistema informativo che coinvolge tutti i livelli e le funzioni dell'azienda.

- **Fase 3 (Advance Reporting):** questa fase verrà implementata dopo 7 mesi di ERP Go Live. Inoltre, ci sarà un'estensione della BI con set aggiuntivo di BI Reporting.

2.1 Metodologia Scrum

Scrum è il metodo Agile più diffuso ed è particolarmente indicato per progetti complessi ed innovativi. Si tratta di un framework, cioè un particolare insieme di *practice*, che divide il processo di gestione di un progetto in *sprint* per coordinare il processo di sviluppo del prodotto secondo le esigenze del cliente.

La teoria alla base di questo metodo è quella del controllo empirico dei processi, attraverso il quale prevede un processo iterativo con un approccio incrementale che ottimizza, passo dopo passo (sprint dopo sprint), la prevedibilità ed il controllo del rischio [4].

Le componenti principali di Scrum si dividono in: ruoli, artefatti ed eventi.

Sono tre i ruoli definiti all'interno dello Team Scrum, che lavorano in stretta connessione per assicurare un continuo e veloce flusso di informazioni.

- **Scrum Master:** il responsabile del processo;
- **Product Owner:** colui che conosce tutti i requisiti del prodotto e porta avanti gli interessi di tutti gli stakeholder;
- **Team Di Sviluppo:** il gruppo di professionisti cross-funzionali ed auto-organizzati. Si occupa dello sviluppo del prodotto e del testing delle funzionalità, e ha la responsabilità di organizzare le priorità trasformandole in task da completare per portare a termine quel determinato sprint.

Gli Artefatti sono progettati per massimizzare la trasparenza delle informazioni chiave (sia per il Team Scrum che per tutti gli stakeholder) e l'opportunità di ispezione e adattamento.

- **Product Backlog:** il documento che contiene la lista di tutti i requisiti necessari per la realizzazione del progetto;

- ***Sprint Backlog***: il documento che definisce tutti i task da completare nei singoli sprint;
- ***Incremento***: la somma di tutti gli elementi del Product Backlog completati durante uno sprint e durante gli sprint precedenti.

Esistono quattro tipologie di evento formali utilizzati in Scrum (con durata prefissata) per creare regolarità, sincronizzare le attività e ridurre al minimo la necessità di incontri non definiti. L'obiettivo di questi eventi è di consentire trasparenza di sviluppo e possibilità di feedback da parte degli interessati.

- ***Sprint Planning***: la riunione in cui il Product Owner ha stilato il Product Backlog e, in presenza del Team di Sviluppo e dello Scrum Master, descrive gli item più importanti e l'obiettivo da raggiungere nello sprint seguente. Al termine della riunione lo Scrum Master può compilare lo Sprint Backlog.
- ***Daily Scrum***: un breve confronto giornaliero fra Team di Sviluppo e Scrum Master;
- ***Sprint Review***: una revisione alla fine di ogni sprint per valutare se l'obiettivo prefissato è stato raggiunto e con quali risultati. Partecipa, oltre allo Scrum Team, anche il committente del prodotto.
- ***Sprint Retrospective***: un'ulteriore analisi retrospettiva effettuata con la partecipazione di tutto lo Scrum Team per valutare se sono necessarie correzioni sugli item dello sprint successivo.

Concludendo, il metodo Agile e Scrum all'interno di un'azienda è vantaggioso perché garantisce reattività e prontezza di risposta al cambiamento, così da ridurre il superfluo e far emergere le priorità.

2.2 Risultati di Analisi

Secondo quanto appreso nel precedente paragrafo, durante la Design Phase sono stati definiti gli obiettivi delle fasi successive attraverso Artefatti che saranno cruciali per lo sviluppo del progetto.

Il mio progetto si colloca all'interno della Wave 1, ovvero la macrofase di raccoglimento di più sprint che ha l'obiettivo di portare in produzione una reportistica basata sulle esigenze del business.

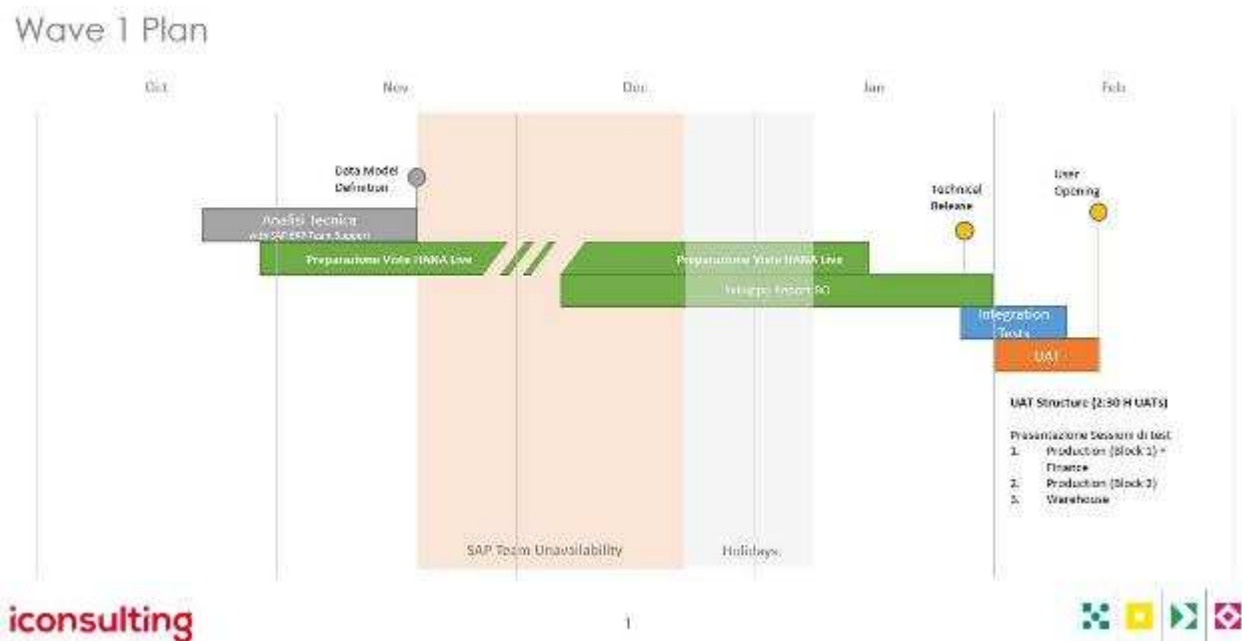


Figura 6: Dettaglio della pianificazione temporale della Wave 1

Durante questa fase sono state fatte delle assunzioni in base alle richieste del business, alla priorità di ciascuna di esse e in linea con lo stack tecnologico del cliente stesso.

È stato quindi delineato che nella Design Phase i report saranno sviluppati utilizzando il componente SAP Hana Live e basandosi solo su SAP FMS come sistema sorgente; non saranno implementati estrattori su SAP ECC⁹ o processi ETL. Tali tecnologie verranno descritte nel dettaglio nel capitolo successivo.

⁹ SAP ERP Central Component (ECC) è una delle attività più riconosciute di proprietà di SAP. Si tratta di un software di pianificazione delle risorse aziendali che consiste in diversi moduli che forniscono alle organizzazioni un grande controllo sui loro processi aziendali chiave.

Finite le analisi condotte precedentemente, inizia il progetto da me sviluppato. Attraverso una strategia bottom-up, il mio scopo è stato quello di modellare i requisiti dell'utente secondo i dati che risiedono sul database, effettuando, quindi, script SQL a basso livello. Quest'ultimi mi sono servite per definire il futuro comportamento che avrebbero dovuto avere le Calculation View una volta implementate.

Le query risultanti mi sono servite inoltre, nella fase successiva, per condurre Unit testing di quadratura sia riguardo il comportamento delle Calculation View che dei report finali su SAP BO.

2.3 Mock-up dei Report

Durante lo *sprint planning* relativo alla Wave 1, è stato concordato con il cliente la creazione di 8 report (5 + 3 di dettaglio).

I report da creare sono:

- **Production output:** permette di osservare l'andamento della produzione analizzando ogni workshop nel dettaglio e fornire metriche basate sull'intera settimana o mese corrente;



Company Name

Production Output

Report Filter Summary
 Operation: 1 Raw Operations Unit
 Execution Date: 13/09/2018

Key Operation Type of Material Total Internal Sub Mon Tues Wed Thu Fri Week Daily Avg Month Daily Avg Total MTD

| WIP | | | | | Production Quantity | | | | | | | |
|---------------|------------------|-------|----------|-----|---------------------|------|-----|-----|-----|----------------|-----------------|-----------|
| Key Operation | Type of Material | Total | Internal | Sub | Mon | Tues | Wed | Thu | Fri | Week Daily Avg | Month Daily Avg | Total MTD |
| Key Op 1 | Front | 10 | 5 | 5 | 25 | 35 | 30 | 35 | 35 | 31.6 | 30 | 305 |
| Key Op 2 | Front | 5 | 5 | 0 | 20 | 20 | 15 | 20 | 30 | 21 | 15 | 300 |
| Key Op 3 | Front | 5 | 0 | 0 | 10 | 15 | 15 | 10 | 15 | 13 | 15 | 300 |

- **WIP (Work in progress):** esso fornisce un quadro del WIP e della sua ripartizione per ciascun livello di produzione al fine di mantenere in equilibrio tutte le aree di produzione. Attraverso un *Drill-Down* verrà osservato il dettaglio di ordini di produzione;

- Inventory MM e WM:** relativi a due tipi di inventario differenti, servirà a valutare le attività di gestione del magazzino in termini di inventario e quantità di gap, se presente. Attraverso un *Drill-Down* sarà possibile osservare con maggior livello di dettaglio il materiale allocato.

Prompt

Inventory Date - From

Inventory Date - To

Stock Date

Inventory Type

Inventory Document

Storage Location

Storage Type

Inventory

CompanyName

Drill-Down to be discarded

Report Filter Summary

Inventory Date - From: 10/01/2018

Inventory Date - To: 10/09/2018

Stock Date: 10/09/2018

Inventory Type: All

Inventory Document: All

Storage Location: All

Storage Type: All

| Site | Storage Location | Stored Value Value at 10/09/2018 (€) | Gap Absolute Value (€) | Gap Net Value (€) | % of Absolute Gap on Storage Location value | % of Net Gap on Storage Location value | % of Absolute Gap on Total Warehouse Value | % of references inventoried |
|------------------------|------------------|--------------------------------------|------------------------|-------------------|---|--|--|-----------------------------|
| TOTAL INVENTORY | | 511.006,70 | 2.950,20 | -1.000,20 | 0,58% | -0,20% | 0,58% | 10% |
| Site 2 | Location 5 | 500.004,20 | 1500,20 | -1500,20 | 0,4% | -0,3% | 0,35% | 9% |
| Site 1 | Location 4 | 600,00 | 350,00 | 350,00 | 58,3% | 58,3% | 0,07% | 5% |
| Site 1 | Location 3 | 400,00 | 350,00 | 150,00 | 87,5% | 37,5% | 0,07% | 8% |
| Site 1 | Location 2 | 3.000,00 | 250,00 | 0,00 | 2,8% | 0,0% | 0,05% | 6% |
| Site 1 | Location 1 | 1.002,50 | 200,00 | 0,00 | 20,0% | 0,0% | 0,04% | 1% |

Storage Location 1

| Storage Type | Stored Value Value at 10/09/2018 (€) | Gap Absolute Value (€) | Gap Net Value (€) | % of Absolute Gap on Storage Location value | % of Net Gap on Storage Location value | % of Absolute Gap on Total Warehouse Value | % of references inventoried |
|---------------------------------|--------------------------------------|------------------------|-------------------|---|--|--|-----------------------------|
| TOTAL STORAGE LOCATION 1 | 1002,5 | 200 | 0 | 0 | 20% | 0% | 4% |
| Type 1 | 501,25 | 100 | -20 | 20,0% | -4% | 0,2% | 3% |
| Type 2 | 200,50 | 0 | 0 | 0,0% | 0% | 0,0% | 1% |
| Type 3 | 400,75 | 100 | 0 | 25,0% | 0% | 0,2% | 60% |

3. Implementazione

Per sua natura, la BI necessita di strumenti che permettano di interrogare e manipolare le informazioni. Tra i più importanti troviamo i sistemi per la gestione di basi di dati (DBMS), ovvero strumenti tecnologici in grado di gestire efficientemente grandi collezioni di dati, persistenti e condivisi, che offrono dei meccanismi che garantiscono l'affidabilità dei dati, il controllo degli accessi e la concorrenza. In questa sezione, verranno introdotti i DBMS utilizzati all'interno di questo progetto di tesi, e spiegate le motivazioni che ne hanno portato all'adozione. In particolare, approfondiremo:

- SAP HANA: caratteristiche, architettura, gestione della memoria;
- SAP HANA Cloud: caratteristiche;
- SAP HANA LIVE: caratteristiche, proprietà e utilizzo;
- SAP Business Object: caratteristiche, architettura, tool

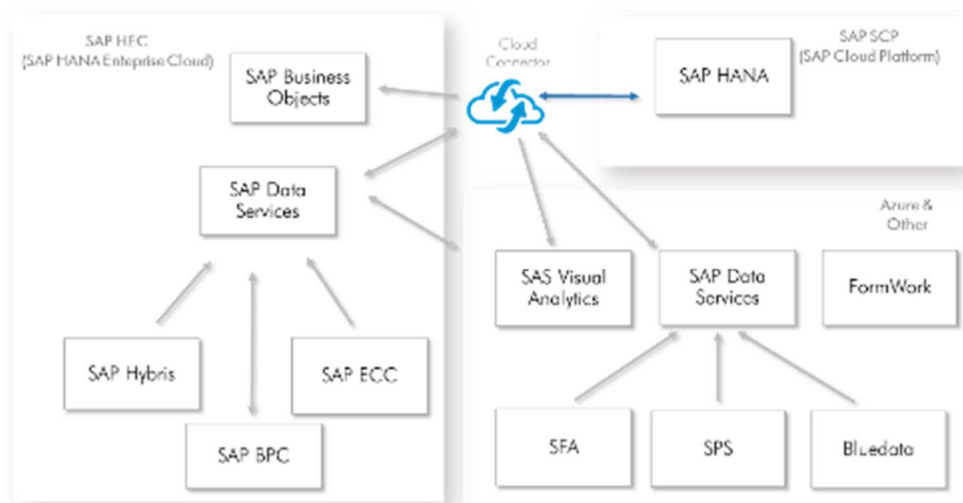


Figura 7 Architettura SAP Enterprise Cloud in essere nel DWH già esistente.

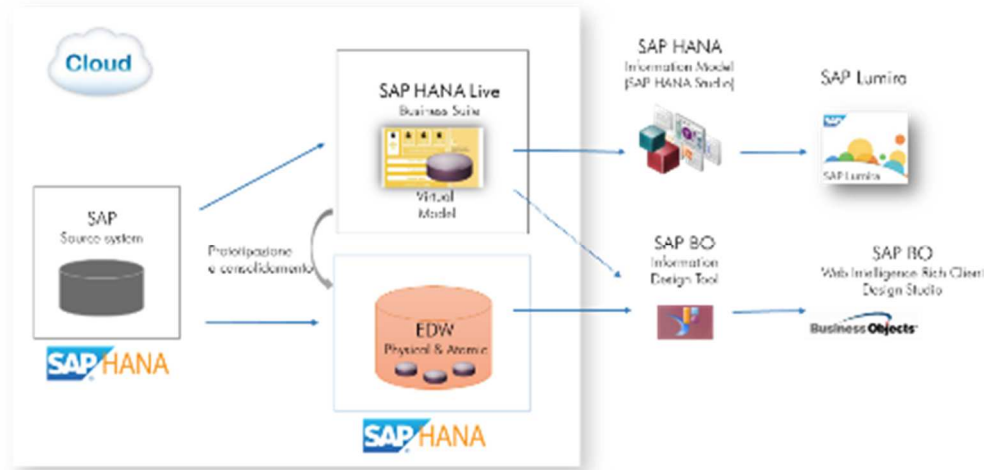


Figura 8: Organizzazione delle Tecnologie

3.1 SAP HANA

Il mercato attuale è fortemente influenzato dalle esigenze della società. Il fatto che essa si evolva in maniera costante e con ritmi veloci, obbliga le aziende a rivedere e ripensare costantemente il proprio business. In questo scenario gioca un ruolo fondamentale la scelta tecnologica.

La soluzione adottata è stata pensata e realizzata in relazione al tipo di dati di business con cui è necessario interagire.

Il Cliente ha deciso di utilizzare SAP per le sue caratteristiche e performance in ambito Enterprise.

La strategia di SAP è quella di riuscire ad abbattere le latenze e i colli di bottiglia legati all'accesso del patrimonio informativo aziendale, creando una soluzione in grado di analizzare in maniera tempestiva il proprio business permettendo di prendere decisioni consapevoli sul futuro prossimo [12]. La soluzione sviluppata da SAP è la prima appliance analitica basata sull'in-memory computing. Da qui il nome di High Performance Analytic Appliance, sul mercato conosciuta con il suo acronimo HANA:

- **Appliance:** la piattaforma è costituita da un insieme di hardware e software progettati per riuscire a ad eseguire particolari

e complesse funzioni applicative, così da garantire sempre ottime prestazioni.

- **Analytic:** è una piattaforma orientata all'analisi di enormi quantità di dati che descrivono il business supportando in maniera eccellente tutte le operazioni OLAP, mettendo a disposizione anche tecniche di data mining per fare analisi predittiva come algoritmi di clustering, di classificazione e di regressione, reti neurali e funzionalità statistiche [16].
- **High Performance:** garantita dalle ottime prestazioni grazie alla politica del In-Memory computing, che abbatte il collo di bottiglia legato all'elevata latenza su operazioni di I/O dei dischi sui quali risiedono informazioni aziendali (la totalità dei dati risiede direttamente sulla memoria centrale permettendo di avere tempi bassissimi di elaborazione dei dati). Inoltre, grazie alla memorizzazione dei dati in colonna e all'architettura multi-core, alte prestazioni sono dovute anche alla parallelizzazione dei processi.



Figura 9: Tecnologie SAP

Il sistema SAP HANA è stato completamente sviluppato in C++ ed è progettato per eseguire un sistema operativo Linux Enterprise Server. La figura sottostante mostra i componenti principali che costituiscono l'architettura di SAP HANA (fig. 10) [5].

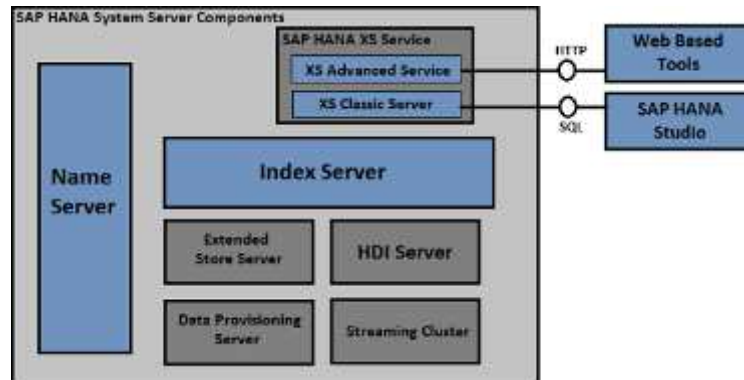


Figura 10: Architettura SAP

Occorre innanzitutto analizzare i singoli elementi e le loro funzionalità:

- ***Index Server***: anche detto SAP HANA Database, è il componente principale. Esso permette di archiviare i dati ed elaborarli. Il server richiede di gestire richieste tramite query SQL e memorizza le sessioni precedenti.
- ***Name Server***: è il componente che contiene informazioni riguardanti la tecnologia del sistema SAP HANA. In un sistema distribuito, con più istanze di HANA su server differenti, il Name Server tiene traccia per ogni server quali siano le componenti in esecuzione e su quale porzione di dati sta avvenendo l'elaborazione.
- ***XS Engine***: XS è l'acronimo di eXtended Service ed è un'estensione del database che consente alle applicazioni esterne di comunicare con la piattaforma mediante richieste HTTP. XS Classic Server è un vero e proprio server HTTP utilizzato per l'esecuzione di operazioni web che svincola gli sviluppatori dall'appoggiarsi ad un server esterno.
- ***Extended Store Server***: è il componente predisposto per fornire supporto e garantire ottime prestazioni anche nel momento in cui si lavora con grandi quantità di dati.

- ***Data Provisioning Server***: è il componente predisposto alla preparazione dei dati da fornire all'utente o a qualche altra risorsa in remote, in real-time o in modalità batch¹⁰.
- ***Streaming Cluster***: conosciuto anche come SAP Event Stream Processor è il componente predisposto alle computazioni di dati stream e alla gestione di eventi complessi.

In base a questa caratterizzazione va data importanza alla gestione delle tabelle in SAP HANA.

Tipicamente, una tabella all'interno di un database è una struttura bidimensionale organizzata in righe e colonne. La memoria del computer va in contrasto con questa gestione dal momento in cui essa viene organizzata secondo una struttura lineare. Per salvare una tabella in una memoria lineare sono possibili due soluzioni:

- ***Memorizzazione orientata per righe (row oriented)***: le tabelle vengono gestite come una sequenza di record, ognuno dei quali contiene tutti gli attributi che descrivono la tabella. Questa tecnica è adottata dalla maggior parte dei database
- ***Memorizzazione orientata per colonne (column oriented)***: le tabelle vengono gestite come sequenze di colonne, ognuna delle quali immagazzina valori di un attributo per tutti quanti i record.

¹⁰ Esecuzione non interattiva di operazioni che riguardano grandi quantità di dati



Figura 11: Tipi di Memorizzazione

La figura 11 mostra, attraverso un esempio sulla tabella delle vendite, le due tecniche di memorizzazione. I vantaggi apportati dalla tecnica colonnare sono notevoli, soprattutto in termini di performance.

Questa tecnica garantisce:

- Veloce accesso a dati, in quanto è sufficiente leggere solo la colonna interessata alla query.
- Qualsiasi colonna può essere utilizzata come indice. Salvare i dati in colonne è, dal punto di vista funzionale, molto simile a costruire un indice per ogni colonna.
- Maggior compressione dei dati in quanto la gran parte dei dati contenuti all'interno di una colonna è composta da pochi valori distinti.

- Procedure parallelizzabili, se più di una colonna è coinvolta in operazioni di ricerca, aggregazione o altro, ognuna di queste operazioni può essere eseguita da un differente processore.

SAP HANA Database supporta entrambe le tecniche, ma è particolarmente ottimizzato per la memorizzazione colonnare dei dati, riuscendo a gestire in maniera efficiente la memoria.

Un'applicazione analitica si compone principalmente di operazioni di lettura e di aggregazione, le quali richiedono velocità nel recupero e nell'elaborazione del dato.

Attraverso la memorizzazione in colonna si ottengono prestazioni notevoli per quanto riguarda operazioni di questo tipo ma non per operazioni di scrittura. La gestione dei dati in memoria centrale, per essere performante su qualsiasi tipo di operazione, viene gestita attraverso due strutture di memoria differenti, ottimizzate per singole funzioni, che collaborano tra loro: Main Storage (MS) e Delta Storage (DS).

Come evidenziato dallo schema sottostante (fig. 11), le operazioni di lettura vengono eseguite da entrambe le strutture ma sono ottimizzate solo per il MS; quelle di scrittura invece, sono eseguite e ottimizzate solo per il DS.

Il MS è quella parte di memoria centrale ottimizzata per operazioni come la compressione dei dati (effettuata prima di salvare il dato in memoria), la lettura, la ricerca e le operazioni di aggregazione sui dati salvati in *column oriented*. Per l'esecuzione di operazioni di questo tipo è necessario avere all'interno del MS la totalità dei dati che, per essere gestiti in modo snello, vengono salvati in maniera compressa.

L'operazione di scrittura in MS è un'operazione onerosa e dispendiosa in termini di tempo e risorse. Proprio per questo motivo, tale operazione non va a modificare direttamente i dati compressi, ma tutti i cambiamenti vanno a scrivere su DS [5].

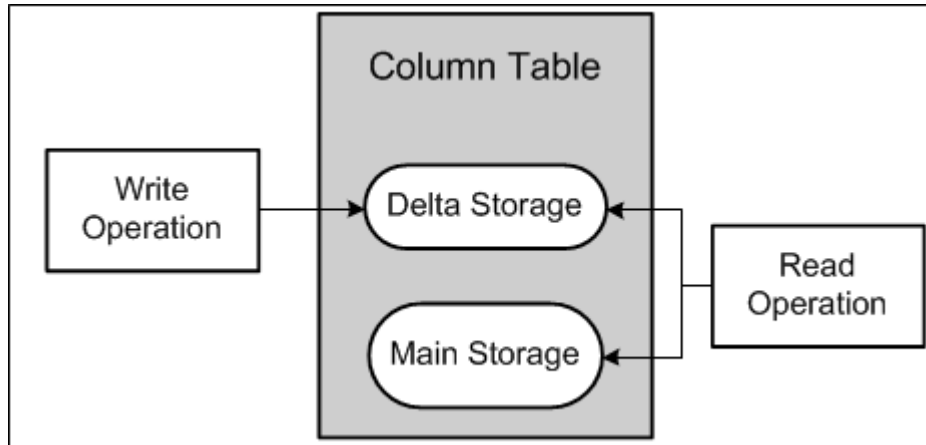


Figura 12: Operazioni Read e Write

Questa struttura ausiliaria è incaricata e ottimizzata per la scrittura dei nuovi dati in memoria centrale, e successivamente in memoria di massa, al fine di allineare i dati per operazioni di analisi con quelli di backup. Quindi il DS rappresenta la chiave di ottimizzazione dei tempi per la gestione della memoria. La procedura viene definita Delta Merge ed è rappresentata in figura:

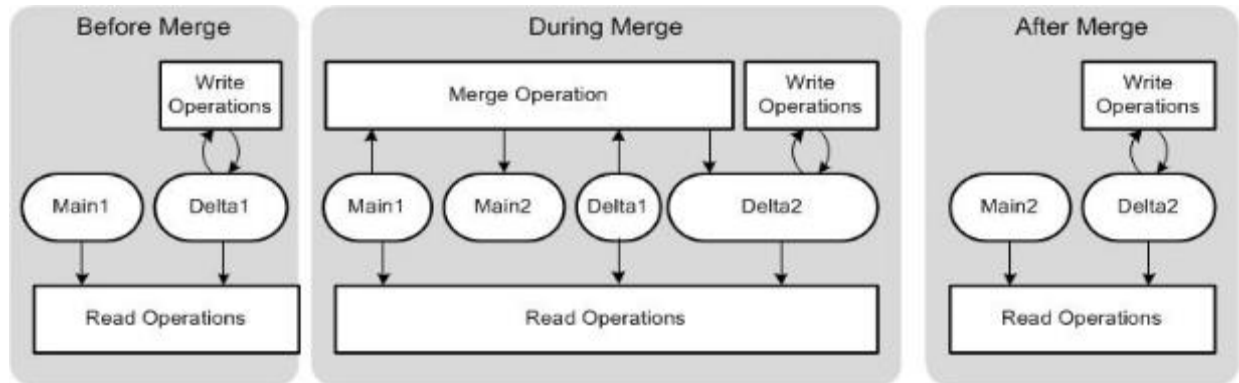


Figura 13: Before-During-After Merge

La procedura è rappresentata in tre step:

- **Before Merge:** tutte le operazioni di scrittura sono gestite dal Delta Storage mentre le operazioni di lettura sono gestite sia dal Delta Storage che dal Main Storage;
- **During Merge:** si accavallano le seguenti attività:

1. Tutte le operazioni di scrittura sono gestite dal Delta Storage 2;
 2. La lettura viene fatta da entrambi i Main Storage e dal Delta Storage 1;
 3. I nuovi dati presenti solo nel DS1 (dati “uncommitted”) vengono trascritti da DS1 a DS2;
 4. I dati presenti all’interno di MS1 e i dati “committed” all’interno del DS1, vengono uniti e scritti all’interno del MS2;
- **After Merge:** si verificano le seguenti attività:
 1. MS1 E DS1 vengono eliminate;
 2. Viene ottimizzata e rivalutata la memorizzazione dei dati all’interno dello spazio di memoria MS2. Queste operazioni consistono nel riordinamento delle righe e la compressione dei dati. Nel caso in cui le operazioni generano un cambiamento dei dati, tali dati verranno ricaricati nuovamente all’interno del MS2;
 3. Il contenuto del MS2 viene scritto in maniera permanente su disco.

3.2 SAP FMS

La soluzione più recente ed innovativa adottata dalle aziende che operano nel settore della moda è SAP Fashion Management Solution (SAP FMS). Essa offre l’opportunità di successive riorganizzazioni di alcuni processi grazie alle innovazioni tecnologiche della piattaforma SAP HANA in modalità cloud.

3.3 SAP Cloud

L’analisi di grandi set di dati richiede una capacità di calcolo significativa che può variare in base alla quantità di dati di input e al tipo di analisi. Questa caratteristica dei carichi di lavoro dei Big Data è ideale per il modello di cloud computing *pay-as-you-go*¹¹, in cui le applicazioni possono facilmente ridimensionare il proprio ambiente (in orizzontale o

¹¹ Ad esempio, AWS (Amazon Web Service)

verticale) per soddisfare le proprie esigenze, senza dover attendere l'hardware aggiuntivo o essere obbligato a investire troppo per fornire una capacità sufficiente.

SAP HANA Enterprise Cloud (SAP HEC), a suo modo, è una soluzione *infrastructure-as-a-service*, basata sulla tecnologia *in-memory*, che consente alle aziende, essendo completamente modulare, di sviluppare, integrare ed estendere nuove o esistenti applicazioni in cloud, traendo beneficio da cicli di sviluppo molto rapidi e garantendo sicurezza in quanto è gestita privatamente da SAP.

All'interno del nuovo scenario dell'economia digitale, le aziende infatti richiedono di poter innovare e differenziarsi rapidamente, di estendere sia il proprio raggio d'azione – essere presenti sul maggior numero di dispositivi, di raggiungere il maggior numero di clienti e geografie, ecc. – e di scopo, offrendo nuovi servizi, prodotti e customer experience. HEC consente alle aziende di qualsiasi dimensione, dalla start-up alla multinazionale, di soddisfare queste nuove esigenze agevolando lo sviluppo di percorsi di digitalizzazione e di business transformation.

Grazie a HEC, ma anche a SAP HCP (SAP Hana Cloud Platform), il Cliente ha potuto implementare in maniera veloce ed efficace tutti i processi core (come finance, gestione degli ordini e delivery), costruendo il proprio data warehouse direttamente in cloud e conferendo maggiori velocità e scalabilità ai propri processi [13].

3.4 SAP HANA Live

Riuscire a prevedere cosa potrebbero fare i dati, e ancor di più le informazioni ottenute attraverso un flusso continuo di dati, generati da sensori, macchine, veicoli, telefoni cellulari, social network e altre fonti in tempo reale, sta diventando sempre più un bisogno impellente per le aziende.

Man mano che il volume di dati aumenta, l'analisi richiede strumenti sempre più scalabili, flessibili e ad alte prestazioni per fornire informazioni in modo tempestivo. Tuttavia, le aziende si trovano ad affrontare un crescente ecosistema di strumenti che nascono e muoiono molto rapidamente.

Gli strumenti e le tecnologie Big Data offrono opportunità e sfide per poter analizzare i dati in modo efficiente, dando la possibilità di comprendere nella maniera migliore le preferenze dei clienti, ottenere un vantaggio competitivo sul mercato e far crescere il proprio business [6].

SAP HANA, come già accennato, rompe le tradizionali barriere di database per semplificare i paesaggi IT, eliminando la preparazione dei dati, la preaggregazione e l'ottimizzazione. È possibile accedere istantaneamente a enormi volumi di dati strutturati e non strutturati, compresi i dati di testo, da fonti diverse.

SAP HANA è una piattaforma completamente rivoluzionata per il business in tempo reale. Trasforma il business razionalizzando transazioni, analisi, pianificazione, previsione e elaborazione dei dati su un singolo database in memoria, in modo che le aziende possano operare in tempo reale.

Eseguendo il componente SAP NetWeaver Business Warehouse (SAP NetWeaver BW) distribuito su SAP HANA, si stabilisce anche un nuovo paradigma di prestazioni, agilità e semplificazione. Le prestazioni delle query e il caricamento dei dati sono esponenzialmente più celeri, quindi report e analisi complessi che normalmente richiederebbero ore o giorni possono ora essere disponibili in pochi minuti o addirittura in secondi, offrendo agli utenti business la possibilità di una vera business intelligence (BI) self-service. In questo modo è possibile consolidare livelli analitici e di pianificazione, di gestione e di archiviazione dei dati separati nel database SAP HANA, riducendo i costi generali di gestione.

Utilizzando la soluzione di rapida implementazione progettata per la migrazione del database di SAP NetWeaver BW in SAP HANA, è possibile completare agilmente la migrazione del database, senza interrompere il data warehouse esistente e proteggere completamente l'investimento BI. Cosicché è possibile lavorare sulle strutture dati esistenti per sfruttare SAP HANA a pieno ritmo.

SAP HANA Live fornisce contenuto distribuito da SAP nello stesso modo in cui viene consegnato il contenuto da SAP NetWeaver BW; tuttavia, il primo fornisce rapporti sui dati transazionali, mentre il secondo fornisce funzionalità di reporting storico. Le visualizzazioni SAP HANA

Live forniscono report operativi in tempo reale forniti nel modello di dati virtuali (VDM), utilizzando il VDM basato su tabelle di dati master e transazionali all'interno di SAP Business Suite.

Gli utenti accedono ai dati in tempo reale utilizzando *Calculation View*¹² e attraverso la piattaforma SAP BusinessObjects. L'utilizzo di HANA Live e BW in tandem concede funzionalità avanzate di analisi aziendale che creano un vantaggio reale e strategico per qualsiasi organizzazione.

Tuttavia, è naturale che un'analisi previsionale effettuata con solo dati transazionali in tempo reale è incompleta. Per questo motivo è necessario successivamente eseguire delle schedulazioni, così da garantire un piano analitico completo e strategico per ogni funzione aziendale.

Quando SAP HANA Live si connette al sistema ECC per accedere ai dati operativi in tempo reale, include informazioni più aggiornate nelle sue analisi rispetto a quanto esiste in SAP BW al momento dell'accesso tramite BI Content. Ciò è dovuto principalmente al fatto che i dati BW sono suscettibili alle trasformazioni dei dati e ai carichi già esistenti in BW e verranno eseguiti più lentamente e/o con ridondanze [14].

Concludendo, i dati transazionali in tempo reale sono diventati il nuovo indicatore del valore competitivo nel mercato. Tuttavia, i dati storici mantengono il loro valore, risolvendo molti problemi aziendali utilizzando l'integrazione di applicazioni di terze parti e richiedendo dati multi-sourcing da database transazionali precedenti.

Quindi, utilizzare BW e SAP HANA Live insieme consente non solo di conservare le lezioni del passato, ma anche di ottenere un vantaggio sul futuro.

3.4.1 *Calculation View*

Come accennato, HANA Live permette l'accesso all'utente di dati in tempo reale attraverso Calculation View, che altro non sono se non viste di calcolo basate sull'estrazione di dati direttamente dalle tabelle del DB operativo.

¹² SAP fornisce delle Calculation View pre-sviluppate ma all'occorrenza possono essere reimplementate.

Le Calculation View applicano a run-time logiche di business e di aggregazione a partire sia da tabelle fisiche del DB HANA che direttamente sulle strutture dei sistemi transazionali (SAP HANA Live). È, come già accennato, lo strumento utilizzato da HANA per modellare e fare analisi multidimensionale. Nella pratica, sono delle viste che quindi non materializzano dati ma li processano in Real-Time utilizzando il motore *in-memory*.

Possono essere analizzate tramite SQL, report Webi e dashboard, o altri sistemi di data visualisation.

Esistono due tipi di Calculation View:

- **Grafiche:** caratterizzate da una miglior ottimizzazione e utilizzo del Calculation-Engine, gestite tramite interfaccia grafica (Drag&Drop) che mette a disposizione un sottoinsieme di funzioni possibili (join, union, aggregazioni, proiezioni, rank). Esistono funzioni limitate di aggregazione, a scapito di una manutenzione più complessa (ad esempio la modifica di un attributo che si propaga in tutta la vista di calcolo);
- **Scriptate (Table Function):** permettono una maggiore libertà, consentendo di inserire codice SQL con AFL¹³. Essendo meno efficienti, vengono usate solo quando le Calculation View grafiche non hanno sufficiente potere espressivo.

In generale, per logiche comuni sono molto semplici e veloci da sviluppare, ma non banali quando le logiche diventano più complesse (per esempio effettuare Count Distinct o calcolare percentuali).

Le Calculation View grafiche sono composte mettono a disposizione quattro blocchi principali per essere modellate:

- **Projection:** viene utilizzata per selezionare colonne particolari dalla tabella di origine, filtrare i dati e aggiungere colonne calcolate. Una proiezione migliora le prestazioni perché riduce il set di risultati. L'utente può definire campi aggiuntivi nella

¹³ Application Function Library, sono componenti forniti dall'applicazione.

proiezione (ad esempio è possibile preformare i calcoli prima dell'aggregazione).

- **Aggregation:** viene utilizzato per aggregare set di dati. Ad esempio, utilizzata per ottenere la quantità totale prodotta di un materiale.
- **Join:** viene utilizzato per collegare due set di dati riga a riga utilizzando chiavi e condizioni. Nel blocco Join è necessario specificare i campi di connessione.
- **Union:** viene utilizzato per combinare i campi provenienti da due set di dati diversi. All'interno deve essere specificato per ogni campo sorgente la sua destinazione.
- **Rank:** viene utilizzato per generare classifiche, ordinamenti e aggregazioni su specifici attributi. Utilizzata per tener traccia di stati che possono cambiare nel tempo, con questo nodo è stato possibile estrarre lo stato più aggiornato di un ordine di produzione.

Inoltre, le Calculation View non sono semplici da debuggare: può essere fatta un'analisi step-by-step sui piani di accesso ma rimangono di difficile lettura.

A livello Front-End la questione si complica ulteriormente, in quanto non è possibile trovare quale query venga generata. Questo problema rende difficilmente approcciabile la costruzione di micro-cubi mirati per aumentare ulteriormente le performance.

Alcune funzionalità fornite dallo strumento per rendere ancora più performante la vista generata:

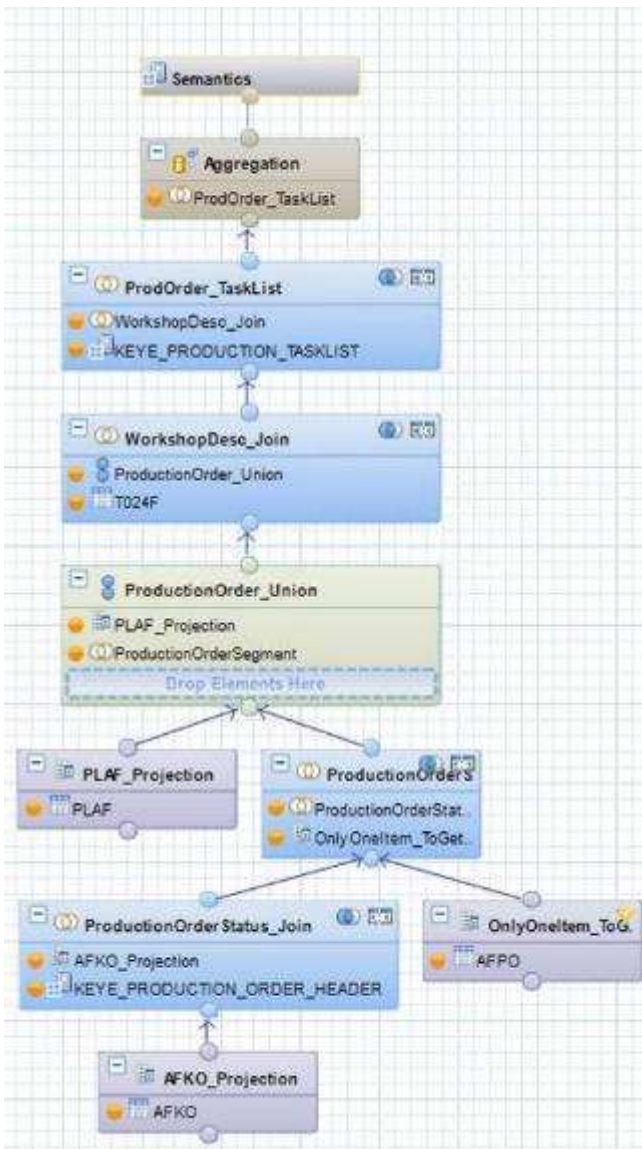
- **Parametri:** permettono di limitare alla sorgente il dataset in analisi, così da liberare il DB HANA ed effettuare un calcolo più snello;
- **Dynamic Join:** è un'impostazione settabile nelle caratteristiche dei nodi di tipo Join all'interno della Calculation View, molto utili nel caso di mescolanza di modelli a differenti granularità;

- **Transparent Filter:** permette di gestire correttamente l'attributo nel caso sia usato solo come filtro e non come output dell'aggregazione;

3.4.2 Implementazione delle Calculation View

Sono state realizzate 8 Calculation View rappresentanti i fatti utilizzati per costruire i report; esse sono basate su altre 22 che effettuano elaborazioni più a basso livello. In questo modo è stato possibile definire una chiara separazione tra i cubi finali e le modellazioni intermedie.

Ora andrò ad analizzarne una più nel dettaglio.



La figura rappresenta una Calculation View di elaborazione intermedia, relativa agli Ordini di Produzione.

Procedendo dal basso verso l'alto, incontriamo i primi nodi di tipo *projection* che estraggono dalle tabelle del gestionale solamente gli attributi considerati utili. Dal momento in cui queste tabelle hanno in media 200 colonne, per ottenere il massimo della performance è stato necessario agire nel modo più preciso possibile.

Come in un processo ETL, il problema principale non è l'estrazione ma la certificazione del dato.

Attraverso la selezione degli attributi vengono successivamente effettuati i primi join su altre tabelle del gestionale oppure,

come in questo caso, con altre Calculation View.

Raggiungendo il nodo Union viene effettuata l'unione tra gli ordini estratti dalla tabella degli ordini di produzione con gli ordini pianificati. In questo modo è possibile tener traccia contemporaneamente di due tipologie di ordini. Attraverso altri join vengono raccolte descrizioni e informazioni utili che il gestionale colloca su ulteriori tabelle.

Con l'aggregazione finale si riesce a controllare di non aver *double counting*¹⁴ e all'ultimo livello (quello semantico) si possono impostare le chiavi primarie.

Parliamo ora di un fatto con clausola star join.

Per semplificare il processo di progettazione, è stata inserita la clausola star join nella costruzione di alcuni tipi di Calculation View. È stato così possibile creare dei cubi secondo il modello già discusso nel primo capitolo, che a scapito di un tempo più lungo di caricamento dovuto alla ridondanza dei dati le query effettuate risultano più performanti, perché come già detto, riducono i join da effettuare dinamicamente.

Queste Calculation View saranno utilizzate poi per costruire i report in SAP BOBI.

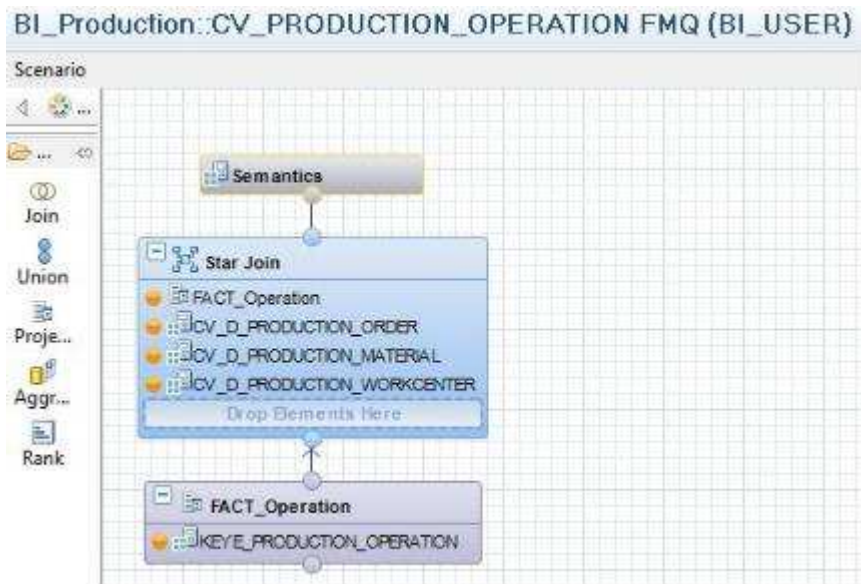


Figura 14: Calculation View con clausola Star Join

¹⁴ S'intende il conteggio di uno stesso valore più di una volta.

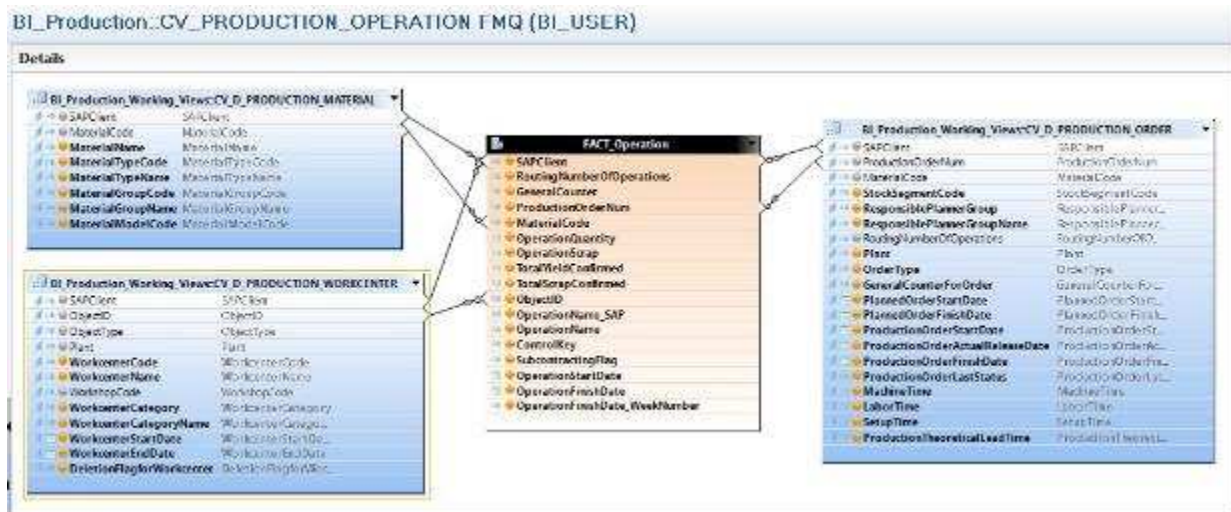


Figura 15: Dettaglio di rappresentazione di una Calculation View con clausola Star join

Possiamo ritrovare lo schema a stella riguardante il fatto discusso nella figura 16 di dettaglio.

In essa viene evidenziata la tabella dei fatti (Operazione), che contiene le misurazioni di uno specifico evento e le tabelle delle dimensioni (relative ai Materiali, agli Ordini e ai Workcenter) che si diramano da esso. Questo schema è intuitivo e di facile comprensione anche per una persona di business che non possiede conoscenze tecniche.

Le relazioni sono tutte uno a molti verso il fatto, in quanto una dimensione può avere più fatti di analisi, cioè uno stesso materiale può essere analizzato per più operazioni.

Questa Calculation View verrà usata per la costruzione del report Production Output.

3.4.3 Sperimentazione

All'inizio del progetto mi sono posto una domanda: quale convenienza c'è nel modellare le Calculation View anziché costruirle direttamente in SQL?

Le viste con script SQL vengono create con vari scopi come permettere agli utenti di avere una visione personalizzata del DB e che in parte

astragga dalla struttura logica del DB stesso. Tutto ciò permette di semplificare la scrittura di query complesse [3].

All'inizio avevo costruito due universi, uno dedicato ai fatti modellati attraverso le Calculation View e uno con delle viste in SQL Hana quanto più possibile analoghe alle prime.

Sviluppando parallelamente entrambi i sistemi ho potuto notare grandi differenze tra i due.

Attraverso la modellazione grafica delle Calculation View certi passaggi erano ottimizzati e facili da riprodurre. Infatti, come spiegato precedentemente, sono ottimizzate al momento dell'esecuzione, potendo ad esempio, eliminare join non necessari permettendo di avere prestazioni nettamente migliori rispetto al classico Script SQL.

Invece, le query scritte per generare le viste SLQ risultavano complesse e non di facile lettura, pertanto il tempo utilizzato per scriverle era di molto superiore a quello speso per implementare le Calculation View, senza per altro portare ad ulteriori benefici.

Per questo motivo, in accordo con il mio Team ho abbandonato questa strada continuando la modellazione solamente attraverso Calculation View.

3.5 SAP Business Object

Generalmente, una volta che i dati sono stati integrati e caricati all'interno del DWH, inizia la fase di studio e analisi degli stessi, garantita da SAP BusinessObject. Nel mio caso invece, il caricamento dei dati avviene, come abbiamo visto, attraverso Calculation View che permettono allo strumento di funzionare allo stesso modo.

SAP Business Objects è la soluzione di business intelligence scelta dalle imprese che vogliono migliorare i processi aziendali, individuare nuove opportunità e ottenere un vantaggio sulla concorrenza [2]. Offre soluzioni in grado di soddisfare tutte le esigenze di BI, da strumenti flessibili di reporting, query e analisi ad hoc a cruscotti e visualizzazioni avanzate. SAP Business Objects è un portafoglio di strumenti e applicazioni perfettamente integrate con i software gestionali SAP, ampiamente

scalabili per ogni esigenza di business che si presenti. Non solo semplice reportistica, ma analisi strategica dei dati, qualità delle informazioni, pianificazione, budget e consolidato.

SAP Business Object presenta una piattaforma per la BI flessibile, scalabile e potente.

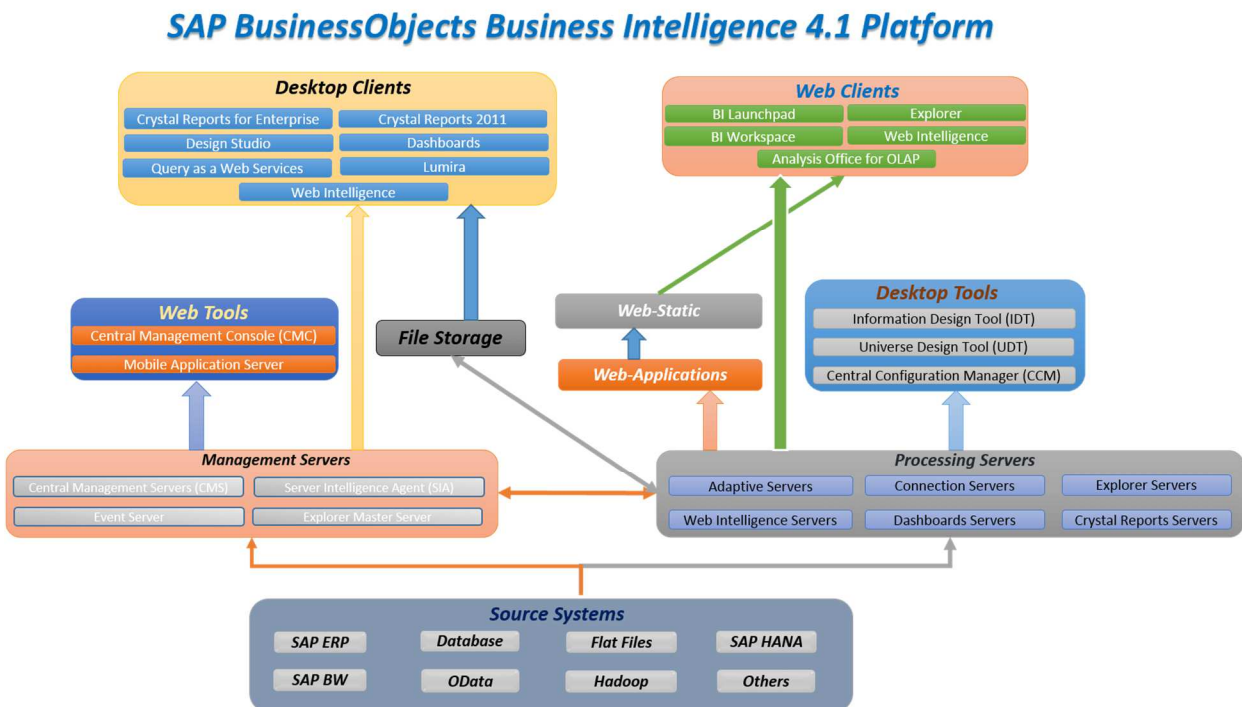


Figura 16: Piattaforma SAP Business Object 4.1

Grazie alla figura 17 è possibile osservare i principali componenti della piattaforma SAP BOBI:

- **Central Management Console:** è l'interfaccia web principale per eseguire attività amministrative nella piattaforma SAP BOBI (sicurezza dell'utente, organizzazione del contenuto e gestione del server)
- **Server Intelligence Agent:** gestisce e controlla tutti i server su un nodo assicurandone il corretto funzionamento. Un nodo è un gruppo di server SAP BOBI Platform che vengono

eseguiti con lo stesso account utente. Una macchina può contenere più nodi premettendo di eseguire processi paralleli.

- ***Central Configuration Managment***: rappresenta uno strumento utile alla risoluzione dei problemi del server e di gestione dei nodi.
- ***Semantic Layer Tool***: Alcuni degli strumenti di reportistica che abbiamo in SAP BOBI possono essere collegati direttamente ai sistemi di origine per creare report, mentre alcuni di essi necessitano di un livello semantico / aziendale intermedio. Quest'ultimo è noto anche come *Universo* in ambiente SAP BOBI, può essere creato utilizzando **Information Design Tool (IDT)** o **Universe Design Tool (UDT)**.

3.5.1 *Information Design Tool*

Information Design Tool (IDT) è un ambiente di progettazione di metadati SAP Business Objects che consente a un progettista di estrarre, definire e manipolare i metadati dalle origini relazionali e OLAP per creare e distribuire universi UNX. Gli *Universi* sono uno strato semantico che nasconde la complessità del database agli utenti finali. Estrae la complessità dei dati utilizzando il linguaggio business al posto di quello tecnico per accedere, manipolare e organizzare i dati.

La figura 18 mostra un piccolo estratto dell'universo da me creato, in cui si osservano le varie dimensioni e metriche per ogni fatto.



Figura 17: estratto dell'universo

3.5.2 Report Finali

In questo paragrafo andrò a descrivere quattro diversi tipi di report, i primi tre sono riferiti a lato production. L'ultimo invece riguarda il magazzino.

Generalmente, tutti i report creati hanno previsto dei prompt, attraverso i quali gli utenti possono personalizzare manualmente l'analisi.

Production output, come già accennato, ha lo scopo di osservare l'andamento della produzione, così da avere una misura sulla velocità di produzione di un ordine ricevuto. Per farlo, vengono effettuate quattro query su diverse Calculation View.

La prima effettua un'interrogazione per estrarre i giorni effettivi in cui la produzione è stata attiva. Il risultato viene utilizzato come filtro per determinare su una query (la seconda) la quantità "work in progress". Attraverso la terza query, grazie al risultato della prima, è possibile trasformare il livello dimensionale di analisi da giorno a mese, in questo modo si può analizzare la quantità media prodotta nel mese corrente.

Infine, l'ultima query si basa su una estrazione della Calculation View (fig. 15 e 16), che permette di ottenere un dato di confronto sulla quantità totale prodotta fino al mese corrente.

Production plan. Anche attraverso questo report si può valutare l'andamento della produzione incrociando i dati della produzione effettiva con quanto pianificato.

A livello tecnico, vengono effettuate due query sullo stesso fatto dell'ordine di produzione, basate sulla Calculation View mostrata al paragrafo 3.4.2, in cui ho evidenziato l'unione degli ordini pianificati con quelli di produzione. Le due estrazioni vengono effettuate per poter calcolare la quantità prodotta in ritardo rispetto a quanto pianificato. Grazie allo star schema è stato possibile analizzare questa metrica da una prospettiva diversa (*pivoting*).

Wip. Effettua una query direttamente sul fatto omonimo. L'aver modellato una Calculation View apposita ha semplificato la creazione del report, che fornisce un quadro sugli ordini in lavorazione per mantenere in equilibrio le varie aree di produzione, evitando di sovraccaricare alcune di esse. In questo modo si rende efficiente il processo di produzione.

L'utente può spostarsi tra vari livelli di dettaglio, grazie a un *Drill-Down*. Quest'ultimo è formato da un report di dettaglio che effettua una query che andrà a raccogliere maggiori informazioni ed effettuerà dei filtri in funzione a cosa preme l'utente (in questo modo alla query vengono passati dei parametri per mostrare il dettaglio voluto).

A livello di inventario sono stati fatti due diversi report: Inventory MM (management) e **Inventory WM** (warehouse management).

Essendo molto simili, mi concentrerò solamente sul secondo in quanto più completo.

Per la costruzione del report vengono effettuate due query, una sul fatto dell'Inventory WM e una sullo Stock WM.

In questo modo si vanno a incrociare i dati relativi all'inventario con la merce immagazzinata, così da poter valutare le attività di gestione del magazzino in termini di inventario e quantità di gap, se presente.

Attraverso un *Drill-Down* sarà possibile osservare con maggior livello di dettaglio il materiale immagazzinato.

Conclusioni e Sviluppi Futuri

Dopo aver rilasciato i report in produzione, è stato possibile iniziare la fase di UAT (User Acceptance Test).

Quest'ultima consiste in sessioni di due giorni in cui gli utenti interessati hanno potuto utilizzare i report allo scopo di fornire dei feedback su eventuali problemi da correggere o validare quanto già fatto in corrispondenza ai requisiti inizialmente espressi.

Questa tesi, come già accennato, è stato il primo step per un progetto più ampio, che precede la costruzione del Data Mart dell'ambito della Produzione.

Una volta rilasciati i report in produzione, è stato possibile giungere all'obiettivo prefissato, cioè garantire la possibilità di effettuare analisi operative e strategiche attraverso report in Real-Time, così da consolidare i KPI utili per le fasi successive. La modellazione dei fatti tratti in questa tesi, permette come sta già avvenendo, di costruire infiniti altri report secondo le necessità dell'utente.

Come già osservato, il futuro e le sorti di una azienda dipendono sempre di più dai cosiddetti decisori aziendali, ovvero quelle figure il cui compito è definire una linea strategica e quindi azioni concrete, al fine di ottenere vantaggio e valore per l'azienda. Per riuscire in questo è necessario che siano supportati da strumenti e tecnologie in grado di analizzare i dati del business per poter fare le scelte più opportune.

I risultati ottenuti al termine di questo progetto mi rendono molto soddisfatto poiché gli obiettivi posti all'inizio sono stati tutti raggiunti e gli utenti sono rimasti molto soddisfatti circa l'organizzazione, l'approccio e la qualità della delivery in tutte le fasi progettuali.

In conclusione, questo lavoro di tesi è stato molto utile per approfondire e avere riscontro pratico nella maggior parte dei topic nel campo della BI. Al contempo, mi ha permesso di acquisire nuove conoscenze dal punto di vista teorico - infrastrutturale - pratico.

Sviluppi futuri

A breve verrà fatta una revisione ad alto livello per le Wave 2 e Wave 3 con lo scopo di esaminare i requisiti e pianificare i prossimi passi, già predefiniti durante la fase di analisi preliminare.

Successivamente, seguirà l'organizzazione di un nuovo ciclo di incontri al fine di esaminare i requisiti di dettaglio per la Wave 2, uniti ai feedback raccolti durante gli UAT e i primi mesi di utilizzo dei report.

Inoltre, verranno definite con maggior dettaglio nuovi KPI in funzione delle esigenze del cliente.

Per concludere, questa tesi corrisponde ad una milestone di un processo molto più grande, che porterà alla costruzione del Data Mart dell'ambito della Produzione e che a sua volta sarà integrato nell'Enterprise Data Warehouse già esistente, comprendente gli ambiti di *Sell-In* e *Sell-Out*.

Bibliografia

- [1] Adamson C. (2010), *Star Schema: The Complete Reference*, New York, McGraw-Hill.
- [2] Administrator Guide: SAP BusinessObjects Design Studio based on SAP BusinessObjects BI Platform, 2016.
- [3] Atzeni P., Ceri S., Fraternali P., Paraboschi S., Torlone R. (2013), *Basi di dati: Modelli e linguaggi di interrogazione*, 4a Edizione, Milano, McGraw-Hill.
- [4] Craig Larman (2005), *Applying UML and patterns: an introduction to object-oriented and iterative development*, Pearson Education.
- [5] Färber F., Primsch J., Bornhövd C., Sigg S. (2011), *SAP hana Database - Data Management for modern business application*, white paper, SIGMOD Record.
- [6] Jayanthi D., Sumathi G. (2016), *A Framework for Real-Time Streaming Analytics Using Machine Learning Approach*, white paper, Department of Information Technology, Sriperumbudur.
- [7] Rezzani A. (2012), *Business Intelligence – Processi, metodi, utilizzo in azienda*, Milano, Feltrinelli Editore.
- [8] Vidette P., Patricia K. e Stephen B. (1998), “Building a Data Warehouse for Decision Support”, Prentice-Hall.
SAP HANA Tutorial, *SAP HANA Architecture Overview*, 2012

Sitografia

[9] El Meleegy A., Butsmann J. (2015), Real-Time Insights for the Digital Business, <https://sapinsider.wispubs.com/>

[10] Napel Ten M. (2019), www.linkedin.com/pulse/real-time-data-enter-confusion-zone-martijnten-napel

[11] Real-time analytics, www.gartner.com/it-glossary/real-time-analytics

[12] SAP HANA Administration Guide, [//help.sap.com/](http://help.sap.com/)

[13] SAP HANA Cloud, [//help.sap.com/viewer/product/CP/Cloud/](http://help.sap.com/viewer/product/CP/Cloud/)

[14] SAP HANA Live for sap business suite, [//help.sap.com/viewer/product/SAP_HANA_LIVE/2.0/](http://help.sap.com/viewer/product/SAP_HANA_LIVE/2.0/)

[15] SAP HANA LIVE: How to install & view its content, [//blogs.sap.com/2015/09/30/sap-hana-line-how-to-install-view-its-content/](http://blogs.sap.com/2015/09/30/sap-hana-line-how-to-install-view-its-content/)

[16] SAP HANA Predictive Analysis Library (PAL). [//help.sap.com/](http://help.sap.com/)

[17] SAP PP tutorial, www.tutorialspoint.com/sap_pp/sap_pp_tutorial.pdf