

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

CAMPUS DI CESENA

SCUOLA DI INGEGNERIA E ARCHITETTURA

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA E SCIENZE INFORMATICHE

**REALIZZAZIONE DI UN SISTEMA DI WHAT-IF
ANALYSIS PER ANALISI RELATIVE ALLA VENDITA
E ALLA GESTIONE DEI PREMI A CLIENTE**

Tesi in

Sistemi Informativi

Relatore

Chiar.mo Prof. Matteo Golfarelli

Presentata da

Mattia Castrichino

Sessione III

Anno Accademico 2013/2014

*Alla mia famiglia ai miei amici e compagni
per avermi sostenuto e accompagnato
durante il mio percorso di studi.*

Sommario

Introduzione	1
CAPITOLO 1 Analisi What-If nei sistemi orientati all'analisi del dato.....	4
1.1 Data Warehouse.....	4
1.1 Interrogazioni nei Data Warehouse	5
1.2 ETL (Extract-Transform-Load).....	6
1.3 Modello multidimensionale.....	7
1.4 Business Intelligence	8
1.5 Analisi what-if	10
1.6 Possibilità d'uso degli scenari d'analisi.....	13
1.6.1 Benefici dell'analisi what-if	14
1.6.2 Tecniche di analisi per sistemi what-If.....	16
1.6.3 Metodologia per l'Analisi what-if	17
1.6.4 Metodologia per la rappresentazione di applicazioni di analisi What-If	19
1.6.5 Analisi pratica di un'applicazione what-if.....	21
1.6.6 Inconvenienti dell'analisi What-If.....	23
1.7 BI a supporto del budgeting.....	25
CAPITOLO 2 Analisi del dominio applicativo	27
2.1 Analisi sorgenti dati.....	27
2.1.1 Soggetti.....	27
2.1.2 Prodotti	29
2.1.3 Premi.....	30
2.1.4 Fattura.....	32
2.1.5 Vendita	32
2.2 Modellazione concettuale: Dimensional Fact Model	32
2.2.1 Identificazione del fatto	33
2.2.2 Identificazione delle misure.....	34
2.2.3 Identificazione delle dimensioni.....	34
2.2.4 Costruzione delle gerarchie	35
2.2.5 Schema di fatto	36
CAPITOLO 3 MECCANISMO DI SIMULAZIONE	39
3.1 Strumenti utilizzati	40

3.1.1	Pentaho	40
3.1.2	Toad for Oracle.....	41
3.1.3	QlikView	42
3.2	Sviluppo dell'applicazione	43
3.2.1	Proiezione premi.....	44
3.3	Applicazione QlikView	47
3.3.1	Caricamento dimensioni	48
3.3.2	Fact table	51
3.3.3	Creazione tabella simulazione	53
3.3.4	Salvataggio simulazione su data base.....	55
3.3.5	Creazione grafici.....	56
3.3.6	Scelte scartate	57
CAPITOLO 4 Interfaccia grafica dell'applicativo.....		60
4.1	Introduzione.....	60
4.2	Interfaccia grafica.....	60
4.2.1	Gestione premi.....	60
4.2.2	Simulazione premi	63
4.2.3	Mappe selezioni.....	66
4.3	Rappresentazione dei fogli di lavoro	69
Conclusioni e sviluppi futuri.....		72
Ringraziamenti		74
Bibliografia		75

Introduzione

Oggi giorno l'analisi del budgeting e del fatturato è un'operazione estremamente importante poiché consente ad un'azienda di adattarsi alle variazioni di mercato e di pianificare eventuali investimenti futuri. L'obiettivo della tesi, sviluppata presso l'azienda *Onit Group s.r.l.*, è stato quello di realizzare un sistema d'analisi *what-if* che consenta di effettuare valutazioni economiche in maniera rapida, precisa, ed in totale autonomia. L'applicativo sviluppato, richiesto dalla direzione commerciale dall'azienda *Orogel*, ha il compito di assegnare percentuali di premio agli acquisti effettuati dai clienti su determinate famiglie di vendita. Il programma è il primo progetto di tipo *data entry* sviluppato nel reparto di *Business Unit Data Warehouse e Business Intelligence* di *Onit* e offre una duplice utilità. Da un lato semplifica la gestione dell'assegnamento dei premi annuali che ogni anno sono rinegoziati, su cui l'utente della direzione commerciale può fare delle stime sulla base dei premi definiti l'anno precedente. D'altra parte rendere la direzione commerciale di *Orogel* più autonoma offrendo all'utenza un unico ambiente su cui muoversi. In passato la gestione dei premi annuali era gestita attraverso procedure Pentaho, che col tempo si sono fatte sempre più ingenti e quindi di difficile gestione. I premi potevano essere direttamente forniti al reparto *Onit* attraverso pagine Excel strutturate in maniera concordata, o depositate in cartelle condivise da cui le procedure Pentaho potevano prelevare i dati. L'applicativo sviluppato ha quindi eliminato le due fasi di lettura e scrittura, offrendo al cliente finale uno strumento che presenta una versione unificata di questi passaggi.

La struttura della tesi è suddivisa in quattro capitoli. Nel primo capitolo è riportato un breve approfondimento delle tecniche di business intelligence, motivando la diffusione di questa tipologia di sistemi, utili per condurre analisi di rilievo sui dati. Il cuore della sezione è incentrato sull'analisi *what-if*, dove sono presentati i vari aspetti positivi che possono dare supporto ad un'azienda nel gestire scelte critiche e le principali tecniche e metodologie di applicazione. Il secondo capitolo è incentrato sull'analisi del dominio, dove risiedono i principali dati che hanno costituito la fonte primaria di informazioni utili per lo svolgimento della tesi. Lo studio condotto ha quindi analizzato a fondo questi aspetti, come ad esempio i clienti e i prodotti, al fine di ottenere tutte le informazioni necessarie per la creazione della modellazione concettuale. Nel terzo capitolo invece vi è la descrizione di come l'elaborato è stato svolto, presentando brevemente anche i principali strumenti

utilizzati. La documentazione delle azioni compiute consiste nel riportare le principali procedure ideate, passando dallo strumento di estrazione e trasformazione dei dati Pentaho, alle *query* di generazioni dei dati del principale software utilizzato per realizzazione della tesi, Qlikview. Il quarto capitolo è incentrato sulla descrizione delle interfacce per la gestione dell'applicativo. Una serie di immagini con relative spiegazioni riporteranno le principali indicazioni su come eseguire l'analisi *what-if* sviluppata.

CAPITOLO 1

Analisi What-If nei sistemi orientati all'analisi del dato

Nel corso del seguente capitolo sono presentate le ragioni che hanno portato alla diffusione della *Business Intelligence* nel campo dell'analisi dei dati. Per introdurre in maniera ottimale l'argomento sono stati introdotti, all'inizio della sezione, dei brevi paragrafi che riguardano le motivazioni che hanno portato all'utilizzo dei sistemi di *data warehouse*, elencandone le funzionalità e le principali operazioni svolte. Particolare attenzione sarà rivolta invece all'analisi *what-if*, spiegando nel dettaglio cos'è, le tecniche e le metodologie di applicazione, elencandone pregi e difetti.

1.1 Data Warehouse

Un Data Warehouse è una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche:

- E' orientata ai soggetti di interesse;
- E' integrata e consistente;
- E' rappresentativa dell'evoluzione temporale;
- Non volatile.

La costruzione di un sistema di *data warehousing* non comporta l'inserimento di nuove informazioni ma piuttosto si basa sulla riorganizzazione di dati già esistenti. Un DW quindi estrae i dati dall'ambiente di produzione ovvero tutte quelle informazioni originariamente archiviate in basi di dati o provenienti da sistemi informativi esterni all'azienda. Inoltre un sistema di questo tipo permette analisi che spaziano sulla prospettiva di alcuni anni, è come se esistesse una fotografia del dato in un certo istante di tempo che non può essere aggiornata. La progressione degli scatti genera

un film che documenta la situazione aziendale da un istante zero fino al tempo attuale. Differentemente, invece, nei comuni DB operazionali abbiamo un contenuto storico limitato dato che la maggior parte delle transazioni coinvolgono i dati più recenti, dati che sono soggetti ad aggiornamenti e che perdono lo storico dell'informazione. I vari aspetti positivi del DW fanno sì che sistemi di questo tipo sono aggiornati ad intervalli regolari e ciò contribuisce alla crescita ed espansione del sistema.

In definitiva è possibile considerare il DW come un database di sola lettura in quanto i dati non vengono mai eliminati e gli aggiornamenti sono svolti solamente quando il sistema è off-line. Tale particolarità legata all'esigenza degli utenti di contenere i tempi di risposta alle interrogazioni di analisi, comporta varie conseguenze nei sistemi di *data warehouse*, non sono più necessarie sofisticate tecniche di gestione delle transazioni adottate dai DMBS. Inoltre viene abbandonata la pratica della normalizzazione delle tabelle a favore di una parziale denormalizzazione, mirata invece al miglioramento delle prestazioni.

1.1 Interrogazioni nei Data Warehouse

Le interrogazioni che vengono svolte all'interno dei sistemi di Data Warehouse sono chiamate *OLAP* acronimo che sta per *On-Line Analytical Processing*. Le funzionalità di questo tipo di elaborazione riguardano analisi dinamiche e multidimensionali che richiedono la scansione di un elevato numero di record con un metro di misura pari all'ordine dei milioni per calcolare un insieme di dati numerici di sintesi che quantificano le prestazioni dell'azienda. Tali caratteristiche fanno sì che i dati nel DW siano rappresentati in forma multi-dimensionale in modo da rappresentarli come se fossero dei punti in uno spazio dove le dimensioni corrispondono ad altre possibili dimensioni di analisi. Lo strumento *OLAP* può essere un mezzo utile per le aziende ogni qual volta si voglia analizzare i risultati delle vendite o l'andamento dei costi di acquisto delle merci, migliorando le performance nella ricerca ed effettuando un numero elevato di interrogazioni con un carico di lavoro che varia continuamente nel tempo. Le funzioni dello strumento *OLAP* sono svariate e si dividono in: *roll-up*, *drill-down*, *drill-across*, *drill-through*, *slicing*, *dicing*, *pivoting*.

1.2 ETL (Extract-Transform-Load)

Con la terminologia ETL si intende un processo di Data Warehousing svolto nella fase di costruzione del DW che ha lo scopo di estrarre le informazioni dalla sorgente originaria e immetterle nel Data Warehouse che si sta creando. In questo modo è assicurata la presenza di dati qualitativamente migliori rispetto alla fonte nativa. La definizione del processo di ETL prende il nome dalle tipologie di operazioni che sono effettuate sui dati, nello specifico la sequenza di operazioni effettuate è: **estrazione, pulizia, trasformazione e caricamento**.

Il passo di **estrazione** riveste il compito di estrarre i dati dal sistema di origine rendendoli accessibili per ulteriori elaborazioni. Per eseguire l'estrazione esistono diverse metodologie:

- *Estrazione completa*: è utilizzata in quei sistemi che non sono in grado di identificare quali dati sono stati modificati del tutto;
- *Estrazione incrementale*: è utilizzata per l'aggiornamento periodico del DW risultando molto utile in tutti quei sistemi che non sono in grado di fornire la verifica di un aggiornamento ma che possono identificare solamente i record modificati;

La fase di **pulizia** ha un ruolo molto importante in quanto garantisce di migliorare la qualità dei dati provenienti dalle sorgenti. Ad esempio controlla che non ci siano dati duplicati, dati errati e verifica che non esistano valori impossibili, errati o inconsistenti.

Lo step relativo alla **trasformazione** applica una serie di regole per modificare i dati dalla sorgente al DW di destinazione. Tale passaggio include la conversione di tutti i dati di misura alla stessa dimensione (dimensioni conformi) utilizzando le stesse unità in modo che poi possano essere uniti più avanti.

Nell'ultima fase di **caricamento** avviene l'immissione dei dati nel DW attraverso le operazioni di:

- *Refresh*: i dati del DW vengono riscritti integralmente, sostituendo quelli precedenti (tecnica utilizzata per popolare inizialmente il DW);
- *Update*: i soli cambiamenti occorsi nei dati sorgente vengono aggiunti nel DW (tecnica utilizzata per l'aggiornamento periodico del DW).

Una volta completati questi passi l'utente avrà piena libertà di eseguire le sue interrogazioni ed analisi sul sistema.

1.3 Modello multidimensionale

La progettazione di data warehouse e data mart segue il paradigma di rappresentazione multidimensionale dei dati, che apporta un vantaggio sia sotto il profilo funzionale, capace di garantire tempi di risposta rapidi a fronte di interrogazioni complesse, sia sul piano logico, dove le dimensioni corrispondono in modo naturale ai criteri di analisi utilizzati dai knowledge worker. Il modello multidimensionale è incentrato sul fatto che gli oggetti che influenzano il processo decisionale sono i *fatti* del mondo aziendale, ovvero fattori come la vendita o le spedizioni. Le occorrenze relative ad un fatto sono dette eventi, ogni singola vendita o spedizione effettuata è considerata un evento. Ciascun fatto è caratterizzato da un insieme di misure utili a descrivere quantitativamente gli eventi. In azienda la quantità degli eventi è troppo elevata per consentire di svolgere singoli analisi sugli eventi, per questo motivo si immagina di collocarli in un spazio n-dimensionale al fine di poterli agevolmente selezionarli e raggrupparli. Gli assi del cubo prendono quindi il nome di dimensioni di analisi. In uno scenario relativo al mondo vendita per esempio, le dimensioni di analisi potrebbero essere i prodotti, i negozi e le date.

Il concetto di dimensione genera la metafora del cubo che per definizione. Un cubo multidimensionale per definizione è detto incentrato su un fatto di interesse per il processo decisionale e rappresenta un insieme di eventi descritti quantitativamente da misure numeriche, dove ogni asse rappresenta una possibile dimensione di analisi. Normalmente ogni dimensione è associata ad una gerarchia di livelli di aggregazione che ne raggruppa i valori in diversi modi, tali livelli prendono il nome di attributi dimensionali. Nel caso di modelli dimensionali più complessi è possibile che le dimensioni d'analisi siano presenti in quantità elevate, facendo così prendere al modello l'appellativo di ipercubo[10].

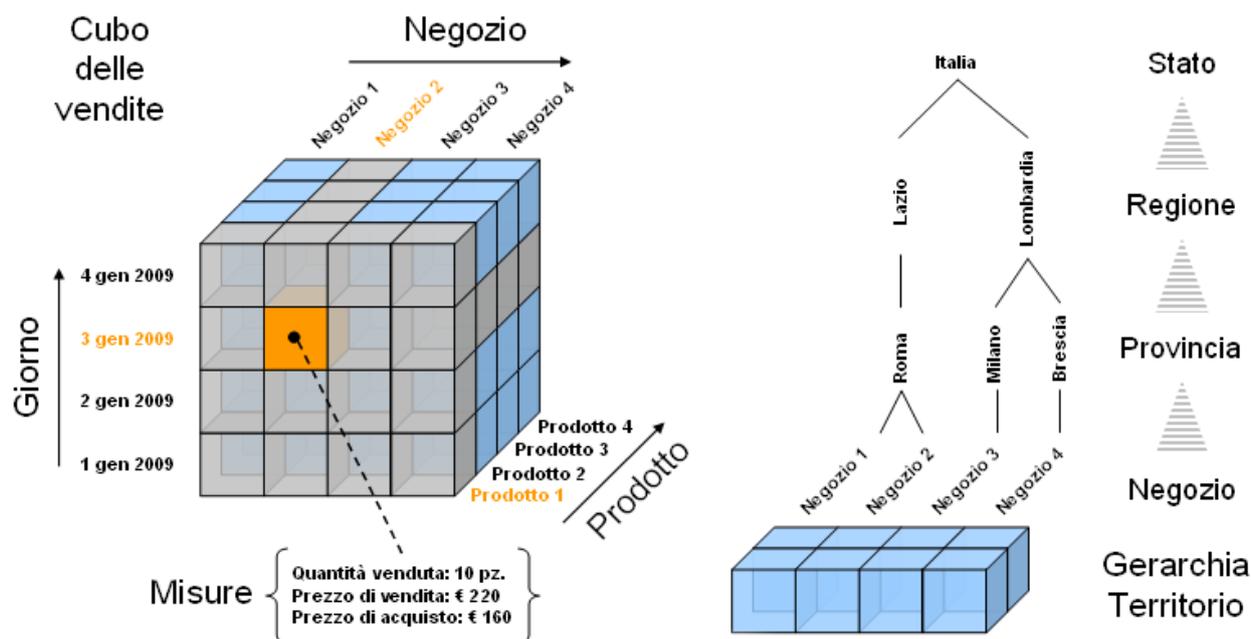


FIGURA 1. A SX ESEMPIO DI CUBO MULTIDIMENSIONALE CHE MODELLA LE VENDITE DI UNA CATENA DI NEGOZI. A DX UNA POSSIBILE GERARCHIA PER LA DIMENSIONE NEGOZI.

1.4 Business Intelligence

Il termine *Business Intelligence (BI)* si riferisce a diverse applicazioni software utilizzate per effettuare analisi sui dati grezzi di un'azienda o di un'organizzazione, composte da diverse attività correlate come: *data mining*, *online analytical processing (OLAP)*, *querying* e *reporting*. Le aziende fanno largo uso della *BI* per migliorare il processo decisionale, tagliare i costi e identificare nuove opportunità di business o persino per individuare i processi di business inefficienti che sono maturi per operazioni di *re-engineering*. Con gli strumenti disponibili oggi è possibile iniziare da subito le analisi sui dati, senza aspettare che vengano eseguiti e portati a termine complessi report. In questo modo viene favorito un incremento della produttività, in passato le attività di raccolta delle informazioni erano svolte con metodi tradizionali costringendo agli utenti di effettuare analisi e compilazioni sui dati e di scrivere di conseguenza le dovute relazioni. In uno scenario relativo a un'impresa di piccole dimensioni con pochi dipendenti questo procedimento si rivela altamente svantaggioso, attraverso un programma di *BI* invece è possibile estrarre dati e creare report con un semplice clic. Così facendo si risparmiano tempo e risorse, consentendo allo stesso modo una migliore produttività dei dipendenti. Anche l'accesso alle informazioni può trarre beneficio dall'uso

della *BI*, evitando lunghe procedure di catalogazione e valutazioni dei dati che sono continuamente in aumento e che contribuiscono ad accrescere la complessità del sistema. Attraverso un appropriato software di *BI* è possibile accumulare e monitorare le informazioni rilevanti rese disponibili solamente quando vuole l'utente. Affinché un progetto di *Business Intelligence* abbia successo è essenziale che venga seguito un processo decisionale competente basandosi sulla migliore informazione obiettiva disponibile. Il processo aiuta l'utente nella monitoraggio delle performance passate e presenti di quasi tutte le operazioni, offrendo allo stesso tempo una base solida per le previsioni future. La *BI* permette di essere più vicini agli obiettivi stabiliti consentendo di tener traccia delle informazioni, accessibili da vari soggetti ogni qual volta ne abbiano bisogno. Attraverso una migliore consapevolezza strategica, reporting più veloce, diminuzione dei costi, overhead più bassi e l'accesso alle informazioni qualitativamente più accurate, la *Business Intelligence* può influenzare positivamente anche l'utile sul capitale investito di una società.

Applicazioni reali della *BI* possono essere relative alle catene di ristoranti che sfruttano le potenzialità dello strumento per decisioni strategiche, come ad esempio l'aggiunta di nuovi prodotti al menù, piatti da rimuovere o quali stabili poco efficienti chiudere. Oltre alla gestione delle portate vengono analizzate anche questioni tattiche come la rinegoziazione dei contratti con i fornitori di prodotti alimentari e l'individuazione dei processi inefficienti da migliorare. Le catene di ristoranti sono fortemente dipendenti da queste operazioni e trovano nella *Business Intelligence* un forte alleato nel gestire le loro attività, sono il gruppo di aziende che più sta traendo beneficio dall'utilizzo di questa materia. In America la *BI* trova utilizzo anche in alcune squadre sportive di baseball e di football, aiutando le società con dati e modelli analitici utilizzati sia dentro che fuori dal campo. Analisi approfondite aiutano la squadra a selezionare i giocatori per rimanere sotto il tetto salariale mentre fuori dal campo il team effettua studi per migliorare l'esperienza dei tifosi. Ad ogni partita in casa un gruppo di 20/25 persone è incaricato di fare rilevamenti specifici che riguardano misurazioni quantitative del cibo, del parcheggio, del personale, pulizia dei bagni e altri fattori. Siti come Amazon e Yahoo non sono semplici siti di e-commerce, sono estremamente analitici e seguono un approccio "testa e impara" per i cambiamenti di business.

Nonostante i numerosi aspetti positivi della *BI* non mancano i problemi, le implementazioni possono essere messe a rischio da sfide tecniche e culturali. Il rifiuto degli utenti ad abbandonare i comuni mezzi lavorativi per affacciarsi a nuove soluzioni costituisce un grosso ostacolo alla diffusione della *Business Intelligence*. Oltre a questi fattori è considerato un ostacolo anche dover vagliare attraverso voluminose quantità di dati irrilevanti e di scarsa qualità. La chiave per ottenere

analisi precise nei sistemi *BI* sono i dati standard, che e rappresentano i blocchi di costruzione del sistema e costituiscono la componente fondamentale e per poter prendere le decisioni. Le aziende devono preoccuparsi di mantenere i loro dati archiviati e che i sistemi di *data warehouse* funzionino ottimamente prima di effettuare l'estrazione delle informazioni essenziali per gli studi di analisi, in caso contrario si rischia di svolgere scelte su informazioni errate che comportano una perdita di tempo e risorse per l'azienda. Altro fattore negativo è rappresentato dai sistemi di *BI* stessi. Nonostante questi siano più scalabili e facili da usare rispetto al passato, il nucleo della *Business Intelligence* consiste in operazioni di *reporting* piuttosto che la gestione dei processi, problema però che col tempo sta iniziando a scomparire. L'ultimo impedimento riguarda le aziende stesse, che non capiscono i loro processi di business abbastanza bene per poter determinare come migliorarli. Questa è una scelta che deve essere valutata con la massima attenzione perché se il processo di business non ha un impatto diretto sulle entrate o se non è stato standardizzato per l'azienda, tutto lo sforzo nella realizzazione del sistema potrebbe disintegrarsi. Le imprese hanno bisogno di capire tutte le attività che compongono un particolare processo di business, come l'informazione e il flusso di dati attraverso i vari processi, come i dati vengono trasmessi tra gli utenti e come le persone usano queste informazioni per eseguire la loro parte del processo. Affinché si possa migliorare il lavoro svolto dai dipendenti è essenziale che tutto questo sia compreso a fondo prima di iniziare un progetto di *Business Intelligence*. Infine è carico dei dirigenti garantire che i dati di alimentazione delle applicazioni di *BI* siano puliti e coerenti, in modo che gli utenti possano fidarsi dello strumento.

1.5 Analisi what-if

Un'analisi *what-if*, da non confondere con l'analisi di sensitività che mira a valutare la sensibilità del comportamento di un sistema a fronte di un piccolo cambiamento di uno o più parametri [3], è una tecnica di "brainstorming" utilizzata per determinare come la proiezione delle performance è influenzata dalle variazioni delle ipotesi su cui queste proiezioni sono basate. Si tratta quindi di un metodo strutturato per determinare quali previsioni relative ai cambiamenti di business possano andare male, giudicando la probabilità e le conseguenze degli studi effettuati prima che accadano. Attraverso l'analisi dei dati storici è possibile la realizzazione di tali sistemi predittivi in grado di stimare i risultati futuri (variabili dipendenti) a seguito delle ipotesi fatte su un gruppo di variabili di ingresso indipendenti, consentendo quindi di formulare alcuni scenari di previsione con lo scopo di

valutare il comportamento di un sistema reale. L'analisi dello scenario serve a determinare i valori attesi, relativi ad un investimento o ad un'attività di business. Gli scenari d'analisi possono essere applicati in diversi modi, il modello tipico è quello di eseguire analisi *multi-factor*, ovvero modelli contenenti più variabili, nelle seguenti forme [4]:

- Realizzazione di un numero fisso di scenari determinando differenza massima e minima e creando scenari intermedi tramite un'analisi del rischio. L'analisi del rischio ha lo scopo di determinare la probabilità che un risultato futuro sia diverso da quello mediamente atteso. Per mostrare questa possibile variazione si esegue una stima dei risultati positivi e negativi meno probabili. Il metodo più semplice per predire i potenziali risultati di un investimento è di produrre un caso positivo e uno negativo per ogni esito e di supporre la probabilità che essi accadano. Prendendo ad esempio una semplice analisi a due fattori, indichiamo con V il valore: $V = \text{variabile } A + \text{variabile } B$, dove A e B non sono quantità limitate, ovvero possono assumere qualsiasi valore. Assegnando i due valori, positivi e negativi, alle due variabili allora è possibile ottenere i valori dei tre scenari. Vengono assunti in modo arbitrario le probabilità di occorrenza:
 - 50% per il valore atteso, X_{exp} ;
 - 25% per il valore negativo, X_{min} ;
 - 25% per il valore positivo, X_{max} .

Una volta trovati i vari dati è possibile dedurre una distribuzione di probabilità piuttosto grezza. Attraverso la creazione di questi casi estremi si inizia ad avere una conoscenza di altri possibili risultati contenuti all'interno dell'insieme dei casi, limitato dal quello positivo e quello negativo precedentemente stimati. Assumendo che ciascuna variabile agisca indipendentemente, ossia che il suo valore non dipenda da quello delle altre variabili, è possibile costruire un caso positivo, negativo e medio per ciascuna variabile. In un modello semplicistico a due fattori questo tipo di analisi risulterebbe in un totale di nove risultati. In generale l'equazione per determinare il numero totale di risultati, tramite l'uso di questo metodo, è uguale a Y^x , dove "Y" è il numero dei possibili scenari per ogni fattore e "x" il numero di variabili nel modello.

- Analisi fattoriale casuale, attuabile attraverso l'uso di metodi Monte Carlo: un metodo Monte Carlo è un qualsiasi algoritmo che risolve un problema generando opportuni numeri casuali e osservando quella frazione dei numeri che obbedisce a unna o più particolari

proprietà. Questi metodi sono utili per ottenere soluzioni numeriche per problemi che sono troppo complicati da risolvere analiticamente [2].

Attraverso la formulazione di diverse possibili ipotesi, combinate tra loro con la probabilità che tali predizioni si verifichino, un'analista può determinare al meglio il valore di un'impresa o di un investimento e la probabilità che il dato calcolato e atteso possa effettivamente verificarsi. Determinare la distribuzione di probabilità di un investimento è pari a stabilire il rischio inerente a quell'investimento. Confrontando il rendimento atteso al rischio previsto e valutando tali fattori con la tolleranza al rischio degli investitori, si può essere in grado di prendere decisioni migliori su specifici investimenti. Ad esempio un investimento che ha fornito lo stesso rendimento ogni anno è considerato meno rischioso di uno che ha fornito rendimenti annuali con valori oscillanti tra il negativo e il positivo. Sebbene entrambi gli investimenti possono fornire lo stesso profitto complessivo per un determinato orizzonte di investimenti, i rendimenti periodici ottenuti mostrano le differenze di rischio fra di loro. Norme rigorose sul calcolo e la presentazione dei rendimenti passati garantiscono la comparabilità delle informazioni ottenute attraverso titoli, gestori di investimenti e fondi. Tuttavia, le performance passate non forniscono alcuna garanzia sul rischio di un investimento futuro o su un possibile guadagno futuro. L'analisi degli scenari cerca di comprendere il potenziale rischio/rendimento di un profilo; effettuando studi di molteplici stime pro-forma per un determinato rischio e indicando la probabilità di guadagno per ogni scenario, si comincia quindi a creare una distribuzione di probabilità (profilo di rischio) per quella particolare impresa. Le risposte a queste domande costituiscono le basi per la formulazione di giudizi di ammissibilità per tali rischi, determinando una serie di azioni per gestire quelle valutazioni ritenute inaccettabili. Un gruppo di verifica esperto può discernere in modo efficace e produttivo le principali questioni che riguardano un processo o un sistema, valutando ciò che può andare storto in base alle loro esperienze passate e alla conoscenza di situazioni simili. A tal fine gli analisti creeranno un modello multivariato, ovvero un modello con più variabili, per cercare di ottenere il valore di ogni variabile ed avere così il dato previsto. La media di ogni distribuzione di probabilità è quella che ha la più alta probabilità di occorrenza. Utilizzando un valore per ogni variabile, che si aspetta essere il più probabile, l'analista calcola il valore medio della potenziale distribuzione dei valori possibili, sapendo però che la media ha un valore strettamente informativo che non mostra alcuna variazione potenziale nei risultati.

Riassumendo, la costruzione di un modello di analisi *what-if* segue le seguenti fasi:

- Raccolta di dati storici al fine di individuare il modello matematico che rappresenta al meglio i risultati. La ricerca del modello matematico consiste nel determinare le variabili che, da un punto di vista probabilistico, hanno la maggiore influenza sulle variabili dipendenti. Oltre ad individuare le variabili con valore di probabilità più alto bisogna anche determinare la formula che meglio spiega le variazioni rilevate nelle serie storiche. Una volta determinata la relazione, è possibile costruire il modello di analisi che, grazie ad una serie di parametri, permette di prevedere i risultati futuri a fronte di una serie di modifiche operate sulle variabili in ingresso
- Analisi statistica multivariata: uno strumento statistico che usa variabili multiple per prevedere l'esito di possibili investimenti. Questa analisi prevede l'esito di situazioni influenzate da più di una variabile, e sono largamente usati nel mondo finanziario. Gli analisti lo usano per stimare flussi di liquidi e valutare nuovi prodotti. I portfolio manager e coach finanziari li usano per predire l'impatto di ogni investimento sul ritorno economico totale. Nel campo assicurativo l'analisi statistica multivariata aiuta gli assicuratori a stimare il rischio. Gli analisti includono questo tipo di modello per valutare le *stock options* (opzioni su azioni)
- Ed infine modelli e tecniche di simulazione, spesso utilizzati per i sistemi informatici di prova e scenari IT

1.6 Possibilità d'uso degli scenari d'analisi.

Considerando ad esempio il lancio di un nuovo prodotto sul mercato da parte di un'azienda, questa deve valutare i possibili scenari in merito alla commercializzazione dell'articolo. Il processo di base per la creazione di uno scenario è suddiviso in quattro parti^[5]:

1. Scegliere l'obiettivo dell'analisi inserito in una domanda *what-if*, ad esempio cosa può comportare il lancio di un prodotto in un determinato periodo dell'anno;

2. Determinare e definire i fattori più importanti che abbiano influenza sulla decisione, come ad esempio la concorrenza, la fluttuazione del prezzo delle materie prime e la maturità del mercato in merito all'accoglienza del prodotto;
3. Ora è possibile selezionare alcuni possibili scenari che siano fattibili e basati sui fattori selezionati, come ad esempio:
 - *Status quo*: descrive uno scenario in cui le variabili in gioco rimangono statiche;
 - *Revolution*: descrive uno scenario in cui le variabili in gioco subiscono uno sviluppo inaspettato;
 - *Major trends continue*: descrive uno scenario in cui le variabili in gioco hanno un trend analogo a quello attuale di mercato, senza soluzione di continuità;
 - *Trends continue and then change*: descrive uno scenario in cui le variabili in gioco hanno un trend analogo a quello attuale di mercato, che in seguito subiscono variazioni;
4. L'ultimo passo consiste nell'arricchire gli scenari con analisi dettagliate;

Come è possibile immaginare, se ogni singolo passaggio viene eseguito in maniera corretta, il team di sviluppo potrà anticipare i gusti del mercato e le mosse della concorrenza. Da ciò si evince che:

- Gli scenari possono aiutare a prevenire i rischi: se sono ben sviluppati potrebbero portare alla luce potenziali problemi prima non considerati che possono essere risolti o evitati;
- Gli scenari, se ben sviluppati, possono aumentare l'efficienza dei dipendenti spronandoli al miglioramento;
- Gli scenari aiutano a capire quali fattori sono realmente importanti per i propri scopi;
- Gli scenari possono aiutare a capire quali settori, prodotti e slogan siano più utili per il successo nell'immediato futuro.

1.6.1 Benefici dell'analisi what-if

Come già spiegato nel paragrafo precedente, lo sviluppo di un'analisi *what-if* è uno strumento molto utile per prendere le decisioni migliori, modificando le ipotesi e osservando la stima dei risultati, prevedendo così il risultato delle proprie scelte. Ad esempio, se si effettua un'analisi prima di decidere di aumentare i prezzi della propria attività, il processo sarà meno rischioso perché si avrà già determinato come l'aumento dei prezzi interesserà il proprio business, oppure un manager può

stimare quanto il margine di un prodotto può variare se vengono modificati i valori relativi al prezzo di vendita o al volume di vendita. Diversamente in uno scenario incentrato sull'economia o in un qualsiasi processo di budgeting, sono sempre presenti variabili incerte come: aliquote fiscali future, tassi di interesse, tassi di inflazione, organico, le spese di funzionamento e altri fattori che non possono essere valutati con grande precisione. L'analisi *what-if* si preoccupa di rispondere alla domanda, "se queste variabili si discostano dalle aspettative, quale sarà l'effetto sul business, o sul modello e sistema di analisi? E quali sono le variabili che stanno causando le maggiori deviazioni?". E' evidente come questa tipologia d'approccio supera i limiti delle comuni operazioni di reportistica e di analisi multidimensionali sui *data warehouse*, che oltre a fornire uno sguardo verso il futuro consente di prendere le decisioni migliori in tempi più rapidi. Lo studio di un sistema di analisi di questo tipo si assicura di rispondere a tutti gli interrogativi posti offrendo una notevole serie di vantaggi, fra cui^[6]:

- Diminuzione del rischio: l'analisi *what-if* permette ai dirigenti di prevedere al meglio l'esito delle loro decisioni, abbassando i rischi normalmente associati al processo decisionale, dando la possibilità di capire se un eventuale cambiamento porterà vantaggi o svantaggi;
- Riduzione dei tempi di decisione: l'analisi *what-if* permette agli utenti di testare rapidamente una grande varietà di scenari con dati ottenuti in tempo reale direttamente dal database. Non c'è bisogno quindi di raccogliere le informazioni o di eseguire diversi report per i diversi scenari. Consente agli utenti di prendere decisioni rapide basate su dati il più aggiornati possibili;
- Miglioramento del processo decisionale: l'analisi *what-if* prevede come le decisioni aiuteranno o danneggeranno le imprese. Inoltre, può portare l'attenzione su cambiamenti trascurati, effettuati in precedenza che potrebbero beneficiare l'azienda;
- Ottimizzazione dell'allocazione di risorse e linee guida per le future collezioni di dati;
- Testare la robustezza dei risultati di un modello o di un sistema in presenza di incertezza;
- Maggiore comprensione delle relazioni tra variabili di ingresso e di uscita in un sistema o in un modello;
- Riduzione dell'incertezza: identificare gli input del modello che causano incertezza significativa in output e che dovrebbero quindi essere al centro dell'attenzione in caso la robustezza debba essere aumentata;
- Ricerca di errori nel modello, tipicamente incontrando relazioni inaspettate tra valori in ingresso e in uscita;

- La semplificazione del modello: ovvero gestire i valori in input al modello che non hanno effetti in uscita, o identificare e rimuovere parti ridondanti della struttura del modello;
- Miglioramento della comunicazione tra modellisti e i decisori, per esempio offrendo informazioni più credibili, comprensibili, convincenti o persuasive;
- Ricerca nelle regioni nello spazio dei fattori di input per cui l'output del modello è o massimo o minimo o soddisfa un criterio ottimale (per esempio il l'algoritmo di Monte Carlo);
- In caso di taratura di modelli con un elevato numero di parametri, un primo test di analisi può facilitare la fase di calibratura concentrandosi sui parametri sensibili. Conoscere la sensibilità dei parametri è fondamentale, si correrebbe altrimenti il rischio che venga speso del tempo su fattori non sensibili e quindi di poco interesse per lo studio.

E' evidente come l'analisi *what-if* sia indispensabile per qualsiasi decisore, è estremamente importante prevedere i risultati prima che si verifichino soprattutto in un clima economico come quello attuale.

1.6.2 Tecniche di analisi per sistemi what-If

Il primo compito del progettista consiste nel riprodurre un modello, il più curato possibile, in grado di simulare il sistema in modo che i risultati ottenuti siano attendibili con quanto desiderato. La costruzione del modello avviene mediante un processo iterativo, verificando ad ogni passo i risultati ottenuti con un insieme di dati di test. Le tecniche di analisi *what-if* possono essere classificate in base al metodo utilizzato per la creazione del modello^[1]:

- *Tecniche induttive*: tipo di tecnica incentrata ad osservare gli effetti del comportamento di un sistema ignorando le cause. Il modello segue la linea di pensiero “*se fino ad ora è andata così andrà così anche dopo*” e si basano sul comportamento che ha avuto il sistema durante un certo periodo. Queste tecniche rappresentano le soluzioni più semplici e vengono anche dette estensionali;
- *Tecniche deduttive*: il modello che verrà generato è caratterizzato da rapporti di tipo causa effetto e da un'approfondita conoscenza delle regole che governano il sistema, superando

quindi i problemi dell'approccio induttivo. Lo svantaggio della tecnica deduttiva è determinato nel caso i rapporti causa effetto diano origine a dei cicli di retroazione.

A volte può capitare che i due tipi di tecniche vengano combinate al fine di massimizzare l'affidabilità del modello.

Indipendentemente dalla tecnica scelta e dal dominio applicativo, la modellazione viene sempre effettuata sui dati del data warehouse che rappresenta il principale mezzo per l'archiviazione delle serie storiche degli eventi aziendali. Nonostante ciò bisogna sempre soddisfare alcuni requisiti generali nello sviluppo di applicazioni *what-if*:

- Deve essere disponibile un modello affidabile e consistente del sistema, e i decisori devono essere in grado di formulare scenari ipotetici su di esso;
- I decisori devono essere in grado di capire e verificare le interrelazioni tra i principali fattori che influenzano il loro business;
- I decisori devono poter osservare lo stato del sistema durante e dopo la simulazione e, possibilmente, intervenire su di esso;
- I decisori devono poter aggregare e disaggregare le predizioni lungo le gerarchie di business e vedere l'impatto delle modifiche a ciascun livello.

1.6.3 Metodologia per l'Analisi *what-if*

Un applicazione *what-if* è incentrata su un modello di simulazione che rappresenta un insieme di relazioni tra i vari aspetti che caratterizzano un'azienda, come per esempio i ricavi, i clienti e il fatturato, ovvero un insieme di fattori che corrispondono ad attività di rilievo per il dominio aziendale. Tale modello costituisce il fulcro principale per permettere agli utenti di eseguire simulazioni con lo scopo di ottenere una o più predizioni, dove ogni predizione assume la forma di un cubo multidimensionale. Il cubo è composto da un insieme di celle di un dato tipo in cui le dimensioni e le misure corrispondono a variabili aziendali, che possono essere utilizzate dall'utente per effettuare analisi tramite applicazioni di front-end OLAP. Per una corretta rappresentazione del modello è necessario un formalismo che copra e integri aspetti statici, funzionali e dinamici. Tra questi risultano, di notevole importanza, gli aspetti statici che descrivono come i dati sono trasformati e derivati nel corso della simulazione. La modellazione dinamica è usata per la descrizione dei domini applicativi mentre per gli aspetti statici il modello di riferimento è quello

multidimensionale, usato per la descrizione di dati storici e dei risultati della simulazione. Nel contesto della *Business Intelligence* è preferibile scegliere come modello di riferimento il modello multidimensionale in quanto è lo schema che meglio si presta ad eseguire analisi sui dati permettendo all'utente di usufruire di più informazioni a diversi livelli di astrazione. L'analisi *what-if*, nella maggior parte dei casi è svolta utilizzando come base di partenza un sistema di *data warehousing*, che contiene al suo interno dati multidimensionali contribuendo in questo modo ad ottenere come risultato un cubo multidimensionale. Attraverso questa rappresentazione viene semplificato l'utilizzo del modello agli utenti che saranno in grado sia di valutare differenti scenari alternativi, per confrontare in un secondo momento gli esiti, sia per effettuare analisi utilizzando come punto di partenza il risultato di una simulazione precedente.

Per semplificare la specifica del modello di simulazione e favorirne la comprensione agli utenti si è adottata una scomposizione a scenari, ognuno dei quali descrive uno o più modi alternativi per costruire una previsione di interesse. Uno scenario è caratterizzato da un sottoinsieme di variabili aziendali, dette variabili sorgente e da ulteriori parametri, detti parametri dello scenario, che l'utente deve valutare per eseguire il modello e ottenere così la previsione. Mentre le variabili aziendali sono legati al dominio aziendale, i parametri dello scenario trasmettono informazioni tecnicamente correlate alla simulazione, come il tipo di regressione adottato per la previsione e il numero di anni passati da considerare per la regressione. Distinguere variabili sorgente tra le variabili aziendali è importante perché permette agli utenti di capire quali sono le "leve" su cui interagire per regolare in modo indipendente la simulazione. Differentemente le variabili aziendali sono utilizzate per memorizzare i risultati della simulazione. Ogni scenario può dar luogo a diverse simulazioni, una per ogni assegnazione ai parametri e ai valori delle variabili sorgente. Infine è importante sottolineare che le variabili numeriche nel contesto della BI sono di tipo multidimensionale, corrispondenti quindi a misure del cubo di input o della predizione.

All'interno di un modello di simulazione è importante anche descrivere la natura delle dipendenze che ci possono essere fra le variabili, determinandone il valore una volta valorizzate le variabili da cui esse dipendono. Un esempio di dipendenza potrebbe essere caratterizzato tra la quantità di due prodotti A e B venduti per mese, filiale e cliente che presentano un andamento di vendita costante. Introducendo però un nuovo prodotto sul mercato potrebbe crearsi il rischio di influenzare negativamente la vendita dei prodotti A e B. Questo fenomeno prende il nome di cannibalizzazione. Esistono due classificazioni per le dipendenze tra variabili^[1]:

- *Vincoli*: variabili che definiscono gli stati legati del sistema simulato rispettate ad ogni istante, come per esempio il ricavo che è determinato da quantità per prezzo;
- *Dipendenze temporali*: affermano che la valore de una variabile in un certo istante può influenzare il valore delle altre variabili in istanti successivi. Per esempio un maggiore vendita di un prodotto dovuta ad una determinata promozione in un certe mese, può avere ripercussioni negative sulla vendita dello stesso prodotto nel mese successivo.

Una situazione ideale prevede che le dipendenze temporali non devono violare i vincoli e che gli stessi vincoli non siano in conflitto tra loro. Nonostante tutto non è semplice raggiungere questo risultato perché le dipendenze potrebbero coinvolgere variabili a differenti granularità e una variabile potrebbe a sua volta essere coinvolta in più dipendenze.

L'ultimo aspetto non trascurabile nello sviluppo di un applicazione *what-if* è relativo alla scelta di precisione e complessità che si decide adottare. Realizzare un modello molto preciso con una granularità fine può risultare estremamente costoso, d'altro canto invece con una precisione minore si rischia di ottenere una simulazione poco affidabile. Occorre quindi prestare molta attenzione nell'effettuare una giusta scelta, nonostante non sia affatto semplice.

1.6.4 Metodologia per la rappresentazione di applicazioni di analisi What-If

L'analisi svolta nel paragrafo precedente si può riassumere nel seguente schema come riportato nel documento [1].

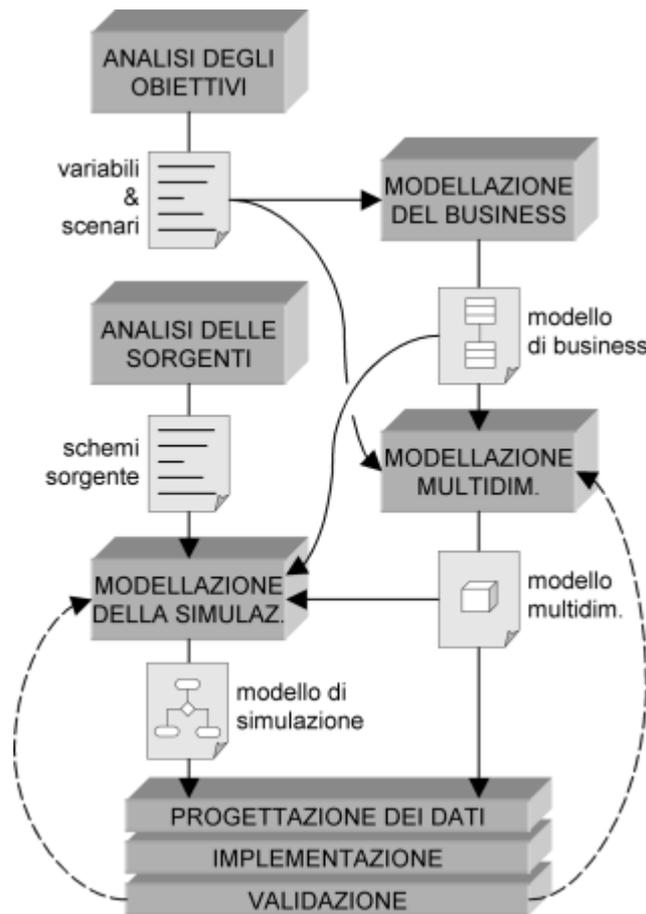


FIGURA 2. METODOLOGIA PER LO SVILUPPO DI APPLICAZIONI WHAT-IF.

Di seguito vengono spiegati i vari punti presenti nello schema:

- *Analisi degli obiettivi*: in questa prima fase avviene l'identificazione sia delle variabili di business, con annessa granularità, sia delle classi di scenari che l'utente vuole monitorare. Quindi lo scopo di questo primo step è relativo alla scelta e alla caratterizzazione dei fenomeni di business che devono essere simulati in un arco temporale ben definito dall'utente;
- *Modellazione di business*: l'obiettivo di questo passo è dare supporto al progettista nella comprensione del fenomeno di business aiutandolo nella scelta di quali indicazioni preliminari trascurabili o semplificabili, tipicamente tramite l'uso di diagrammi UML. Il modello finale ottenuto è costituito dall'insieme di tre sotto modelli, di cui uno predisposto alla modellazione statica delle principali entità coinvolte nel fenomeno di business e alle loro associazioni. Un secondo modello invece definisce come le variabili di business sono

funzionalmente derivate le une dalle altre, ed infine un ultimo modello per descrivere come le entità interagiscono dinamicamente tra loro;

- *Analisi delle sorgenti*: particolare attenzione sarà rivolta all'analisi delle sorgenti dalle quali attingere i dati al fine di valutarne la qualità e di capire la disponibilità e la struttura delle informazioni utili a guidare la simulazione. L'analisi delle sorgenti è una fase estremamente importante perché definisce il modello di simulazione da costruire;
- *Modellazione multidimensionale*: quanto definito nei primi due punti risulta cruciale in questa fase per la creazione dello schema multidimensionale, avente il compito di descrivere la predizione;
- Particolarmente importante è la scelta della granularità che definisce la dimensione della predizione che sarà utile per determinare il massimo livello di dettaglio per l'analisi dei risultati di simulazione;
- *Modellazione della simulazione*: l'obiettivo di questo punto è la costruzione di un modello funzionale/dinamico che permette di popolare per ciascun scenario, con i dati sorgente disponibili, la predizione. Questo passo rappresenta il nucleo principale per la progettazione e richiede di trovare il miglior compromesso tra precisione e complessità del modello di simulazione;
- *Progettazione dei dati e implementazione*: le modellazioni delle due fasi precedenti vengono implementate all'interno di questo step sulla piattaforma prescelta, al fine di creare un prototipo per il testing;
- *Validazione*: rappresenta l'ultima fase nella quale viene lanciata la simulazione su un periodo passato, per poi confrontare i risultati ottenuti con quelli verificatisi realmente. Nel caso i dati ottenuti non siano soddisfacenti occorre ripetere i passi relativi alla modellazione multidimensionale e modellazione della simulazione.

1.6.5 Analisi pratica di un'applicazione what-if

Un esempio pratico di svolgimento di un'analisi *what-if* è descritta in maniera piuttosto approfondita nel capitolo 6 del libro [8], del quale sono riportati i passi salienti.

Inizio, cosa serve. I primi passi nello svolgimento di un'analisi efficace includono identificare i limiti dello studio, coinvolgendo i soggetti giusti e avere le informazioni corrette. Tali limiti possono riferirsi a un singolo elemento, ad una raccolta di dati in relazione tra loro o ad un intero

complesso di informazioni. Un'attenta ricerca produce un'analisi più dettagliata e chiara nel definire i rischi e le verifiche richieste. Come i limiti dell'analisi si ampliano allora i vari risultati ottenuti acquistano significati correlati tra loro. I limiti possono includere i passi nella costruzione del sistema in esame, i passaggi necessari per il funzionamento del sistema o le fasi necessarie per mantenere il sistema stessa. Una chiara definizione dei limiti dell'analisi avvia la revisione in maniera efficace. Come accennato in precedenza la creazione di un team esperto è fondamentale. Il prossimo passo più importante è il reperimento delle informazioni necessarie. Un modo per raccogliere le informazioni riguardo ad uno specifico processo è quello di assistere al suo svolgimento, sia in prima persona o tramite la documentazione disponibile, come ad esempio informazioni relative alla progettazione o alle procedure operative. Se questi documenti non sono disponibili, la prima raccomandazione della revisione è quella di creare le informazioni necessarie: senza di esse non si possono fare analisi accettabili.

Fase di revisione. Una volta esaminate tutte le informazioni, il passo successivo è condurre l'analisi. In generale, un esperto nella gestione dei pericoli condurrà il team di sviluppo attraverso una serie di step *what-if*. Un figura di questo tipo è in grado di assicurare uno svogliamento produttivo ed efficace e può essere affiancato da qualcuno che prende appunti sullo sviluppo dell'analisi.

Sviluppare le domande *what-if*. Con le informazioni disponibili e la conoscenza del team di revisione, gli step *what-if* possono essere formulati riguardo a: errori umani e imprevisti vari. Tutti questi possono essere considerati durante tutte le possibili operazioni che coinvolgono il sistema in esame. Le domande da porsi potrebbero riguardare una delle seguenti situazioni:

- La mancata osservanza delle procedure o procedure seguite in modo non corretto;
- Procedure non corrette o mancato uso di procedure aggiornate;
- Operatore maldestro o non addestrato;
- Procedure modificate a causa di turbamenti;
- Condizioni di processo turbate;
- Malfunzionamento di eventuali apparati;
- Set up errato degli apparati;
- Errori di de-bugging;
- Combinazione dei precedenti eventi.

Man mano che le domande *what-if* sono generate bisogna assicurarsi che chiunque partecipi alla fase di analisi abbia la possibilità di segnalare possibili potenziali errori. Rispondere ad ogni domanda appena essa viene posta, crea il pericolo di terminare troppo presto la fase di verifica, rischiando quindi di non evidenziare tutte le possibili risposte. Nel caso si presentino domande troppo elaborate è opportuno suddividere l'analisi in pezzi più piccoli, per consentire un migliore gestione del problema.

Determinare le risposte. Dopo essersi assicurati che la squadra di revisione abbia esaurito gli scenari *what-if* più credibili, occorre che risponda alla seguente domanda: ”*che cosa sarebbe il risultato di tale situazione si avverasse?*”. Se le risposte sono sviluppate correttamente, rivedere i possibili errori può mettere in risalto svariati miglioramenti nella creazione del prodotto. Per ottenere un miglior risultato, in questa fase, può essere opportuno far partecipare il maggior numero di personale possibile per raggiungere una visione più ampia del problema.

Valutazione del rischio e relative raccomandazioni. Data una risposta a tutte le domande poste il prossimo compito è di giudicare la possibilità che tali situazioni accadano, valutandone la gravità: in poche parole, stabilire il rischio. Dopodiché occorre stabilire l'accettabilità del rischio. La discussione di ogni situazione di rischio comporta naturalmente delle raccomandazioni; la squadra proseguirà la revisione domanda per domanda finché non è stata analizzata tutta la casistica. A questo punto è essenziale fare un passo indietro e riesaminare il quadro della situazione per scovare eventuali dimenticanze.

Riportare i risultati a chi e come. Una volta che l'analisi è stata completata occorre riportare i risultati ottenuti e inviarla alla figura professionale di dovere. Di solito, il responsabile del reparto o di un impianto è il cliente della revisione. A questo punto è necessario generare una documentazione riportante lo scopo della revisione, i risultati più significativi e i dovuti suggerimenti. In alcuni casi è necessario includere nel documento i partecipanti al lavoro e il periodo temporale che è stato necessario per lo svolgimento.

1.6.6 Inconvenienti dell'analisi What-If

Nonostante i svariati aspetti positivi apportati dall'analisi *what-if*, non mancano comunque i problemi relativi nei contesti di business.

- Le variabili sono spesso interdipendenti, ovvero correlate tra loro, rendendo i singoli esami irrealistici. Ad esempio cambiando un fattore come il volume delle vendite, molto probabilmente si influenzeranno altri fattori come il prezzo di vendita;
- Spesso le ipotesi su cui si basa l'analisi sono realizzate utilizzando dati ed esperienza passati che potrebbero non essere più utili in futuro;
- L'assegnazione di un valore massimo e minimo, o ottimistico e pessimistico, è un'interpretazione strettamente soggettiva. Per esempio la previsione ottimistica di una persona può essere più conservativa di quella di un'altra persona che effettua una parte diversa dell'analisi. Questo tipo di soggettività può influenzare negativamente l'accuratezza e l'obiettività complessiva dell'analisi.

Ulteriori svantaggi derivano dallo studio degli scenari dove il maggior inconveniente di queste analisi, con risultati prestabiliti, sono le probabilità stimate e i set di risultati ottenuti, delimitati dai valori estremi in positivo e in negativo per ogni evento. Anche se ci sono poche probabilità che un evento si verifichi, la maggior parte degli investimenti o dei portafogli di investimenti, hanno il potenziale per ottenere degli elevati rendimenti positivi. Chi investe non deve ignorare questi casi, ma deve considerare che, anche se non capita spesso, questi eventi a bassa probabilità possono presentarsi e solamente attraverso l'analisi del rischio è possibile determinare se questi potenziali avvenimenti possono rientrare nella tolleranza del rischio. Un metodo per aggirare il problema è quello di eseguire un numero ingente di prove con un modello multivariato. Un'analisi fattoriale casuale si può considerare completata solamente eseguendo migliaia o addirittura centinaia di migliaia di prove indipendenti attraverso un computer per assegnare casualmente i valori agli elementi. Il tipo più comune di analisi fattore casuale è chiamato analisi Monte Carlo (già accennata precedentemente), in cui i valori dei fattori non sono stimati ma sono scelti a caso da un insieme limitato di variabili, ognuna con la propria distribuzione di probabilità.

Infine, come già è stato detto nei paragrafi precedenti, le performance passate non hanno direttamente alcuna incidenza sul rischio e sui risultati futuri, spetta agli investitori determinare il rischio futuro creando dei modelli pro-forma. Solo in questo modo sarà possibile determinare un profilo per il rischio degli investimenti e creare una base per il confronto degli investimenti futuri.

1.7 BI a supporto del budgeting

La *Business Intelligence* non è solo uno strumento per monitorare e prevedere il comportamento dei clienti ma può essere utile anche nella gestione del budgeting aziendale. Di norma la fine dell'anno fiscale costituisce un periodo stressante per l'impresa che potrebbe risultare maggiormente appesantito se ancora si utilizza Excel per la gestione del budget. In questo caso l'unica soluzione è abbandonare le vecchie abitudini per far fronte a metodi più sofisticati per gestire più accuratamente le pianificazioni, i processi critici di bilancio e migliorare la precisione delle previsioni, aiutando così anche le piccole medie imprese ad adattarsi alle varie condizioni mutevoli che si presentano nel corso dell'anno. La *BI* consente ai dipartimenti delle finanze di fornire bilanci tempestivi consolidati e prevede che i dati siano influenzati da una varietà di fonti in modo a capire come il proprio business aziendale può essere modellato su una serie di scenari.

Di seguito sono elencati alcuni suggerimenti per snellire la fase di budgeting e di pianificazione^[7].

- Nonostante Excel sia uno degli strumenti più utilizzati per la gestione del bilancio e per la pianificazione di applicazioni, risultando spesso anche un ottimo strumento di *BI* soprattutto nelle piccole aziende, prima o poi con la crescita dell'impresa nasce il bisogno di cambiare. Tale cambiamento collima con il raggiungimento di un fatturato annuo pari a 35 milioni di dollari (circa 31 milioni di euro) corrispondenti ad una marea di pagine Excel da gestire, che contribuiscono ad aumentare il rischio di errori oltre ad essere scomode e ingombranti da adoperare. Adottando un software di budgeting dedicato nella piattaforma *BI* è possibile eseguire in tempo reale predizioni *what-if* e monitorare le performance per migliorarle.
- Integrare un software *BI* che sia sufficientemente flessibile nei processi di budgeting già esistenti, piuttosto che partire da zero. Soprattutto nei casi delle piccole società non si ha né il tempo né le risorse per crearsi uno strumento ad hoc, occorre quindi cercare il miglior software che si integri con le fonti dati esistenti, includendo modelli personalizzabili per progettisti e gli utenti finali. Nel caso si trattasse di una piattaforma basata sul Web bisogna assicurarsi che la pianificazione sia disponibile e possibile anche offline.
- Se si crea un flusso di lavoro organizzato nel processo di *BI* di raccolta e pianificazione si ha il vantaggio di monitorare, analizzare e fare aggiustamenti tattici al processo di budgeting al meglio le per tutto il corso dell'anno. Così facendo si prevencono potenziali problemi prima

che si possano trasformare in grandi problemi permettendo anche di identificare le opportunità per migliorare il processo.

- La collaborazione all'interno di organizzazione può portare una maggiore efficienza a patto che siano garantiti gli accessi al sistema di budgeting solo alla persone giuste.
- Un approccio collaborativo ed integrato può rendere il business aziendale più agile a fronte dei cambiamenti delle condizioni di mercato che si evolvono nel corso di un anno o che possono cambiare nell'immediato. Bisogna considerare la pianificazione come un processo continuo eseguendo regolazioni lungo il percorso, cercando, dove possibile, di anticipare i cambiamenti e di includere anche scenari di casi peggiore per permettere all'azienda di adattarsi alle variazioni del mercato.
- Infine abbracciando la tecnologia mobile (tablet e smartphone) è possibile monitorare, analizzare, modificare e fare previsioni sui bilanci comodamente ovunque uno si trovi. Utilizzare quindi una soluzione *BI* in campo mobile non è solo conveniente ma permette di prendere decisioni critiche di business in qualsiasi momento.

CAPITOLO 2

Analisi del dominio applicativo

Nel seguente capitolo è svolta un'attenta analisi dei vari fattori di rilievo legati al mondo delle vendite, fondamentali per lo sviluppo della tesi. Le principali sorgenti da cui attingere i dati, utili per la realizzazione del progetto, sono il cliente, i prodotti, la fattura e i premi. Ognuno di questi aspetti è descritto presentando le principali caratteristiche che li compongono all'interno del dominio.

Una seconda parte del capitolo invece è incentrata sulla modellazione concettuale, resa possibile solamente dopo aver scremato i dati di rilievo inerenti agli aspetti presentati in precedenza. Sono quindi descritte le varie fasi che hanno portato alla costruzione dello schema di fatto, motivando e spiegando le varie scelte effettuate nell'individuazione del fatto, delle misure, delle dimensioni e delle gerarchie.

Le analisi conseguite sono fondamentali per lo sviluppo della modellazione concettuale in quanto costituiscono la base di partenza per lo sviluppo delle future applicazioni. La creazione di un buon modello è essenziale per offrire alla direzione commerciale di Orogel un strumento affidabile che operi su dati bene precisi per consentire valutazioni accurate sulla variazione del fatturato e di determinare quindi il possibile guadagno per l'azienda. Utilizzare dati corretti e ben definiti è cruciale per il processo decisionale di assegnazione dei premi, un solo dato errato può costituire un'ingente perdita monetaria che può variare da qualche centinaio di euro a persino migliaia di euro.

2.1 Analisi sorgenti dati

Nei seguenti sotto-paragrafi sono analizzate le principali sorgenti dati.

2.1.1 Soggetti

All'interno del dominio il termine soggetti identifica un insieme di clienti e persone di risalto nel dominio aziendale, a cui sono assegnate varie informazioni come: numero di telefono, indirizzo,

partita iva, codice fiscale e ragione sociale. Più nello specifico per cliente si intende un insieme di aziende, supermercati, negozi di alimentari e piccoli punti vendita che sono in relazione con Orogel. Nel dettaglio la figura del cliente è scindibile in due categorie:

- Cliente statistico: identifica univocamente lo stabile di vendita a prescindere dal proprietario, il codice che lo identifica resta invariato anche se cambia la ragione sociale. Questo può tornare utile ogni qual volta si vogliono analizzare statistiche di vendita inerenti all'esercizio vendita.
- Cliente di fatturazione: identifica univocamente il cliente a cui viene intestata la fattura, normalmente anche detto cliente intestatario.

Tra le altre informazioni allegate alla figura del cliente è presente il fido, ovvero il valore della merce venduta con pagamenti dilazionati che verranno effettuati dall'acquirente quando avrà già ricevuto la merce. Le note di riferimento indicano invece le informazioni per i contatti relativi al pagamento e comprendo gli orari per effettuare le telefonate, mail, fax ecc. I clienti inoltre possono essere raggruppati in diverse aree denominate aree Nielsen che si suddividono nel seguente modo:

- Area 1: Piemonte, Val d'Aosta, Liguria, Lombardia;
- Area 2: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna;
- Area 3: Toscana, Umbria, Marche, Lazio, Sardegna;
- Area 4: Abruzzo, Molise, Puglia, Campania, Basilicata, Calabria, Sicilia.

Altri tipi possibili di raggruppamento sono il canale, che specifica se si tratta di un cliente estero o meno, e i sottogruppi, dove avvengono i raggruppamenti per gruppo aziendale. Ulteriori specificazioni dell'entità soggetti si riferiscono all'identificazione degli agenti, che rappresentano le figure commerciali adibite a gestire e trovare i potenziali clienti per l'azienda, e dei capi che hanno funzioni simili agli agenti ma con livelli di responsabilità superiori. Le qualifiche e i ruoli sono altri due possibili tipi di raggruppamenti, le qualifiche identificano se un soggetto è una sede, un spedizioniere o una filiale. I ruoli invece danno un'informazione più dettagliata, ad esempio identificano le seguenti classi: fornitori agenti, fornitori trasporti, fornitori fresco, dipendente, filiali dirette e molte altre descrizioni. Nell'entità risiedono anche informazioni relative ad indicare se un cliente svolge il ruolo di fornitore o anche se è il responsabile di una holding, dove per holding si intende una società che detiene partecipazioni in altre società. Infine ad un cliente può essere

assegnato una tipologia listino canale utile ad indicare il listino applicato ad un soggetto. I listini possono essere di vari tipi e possono dipendere dalla stagione o semplicemente relativi a una determinata versione. L'abilitazione invece indica il cliente certificato per la fornitura e può assumere i valori di abilitato, non abilitato, temporaneamente abilitato, chiuso o cliente cessato.

2.1.2 Prodotti

La sorgente dati relativa ai prodotti comprende prevalentemente tutti gli articoli disponibili per gli acquisti effettuati dai clienti che Orogel ha fruibili per la vendita. Gli alimenti relativi alla vendita si dividono in: Orogel Surgelati, Orogel Confetture e Orogel Fresco. Il settore dei surgelati è il core business di Orogel, che propone una vasta gamma di prodotti comprendenti sia vegetali pronti all'uso, sia piatti pronti e monoporzioni. Le confetture, composte e marmellate Orogel si dividono nelle due linee "Virtù di frutta", per la vendita al dettaglio, e "Orofrutta", per il canale food service e industria. Orogel fresco è una delle principali Organizzazioni di produttori che commercializza in Italia e all'estero frutta e verdura coltivata con sistemi di produzione integrata e biologica. Nel dominio per ogni singolo prodotto è presente una descrizione e in alcuni casi può essere aggiunta un'indicazione che specifica se l'alimento appartiene alla categoria dei prodotti di tipo benessere o di tipo biologico. Ogni alimento è caratterizzato da un codice univoco che lo contraddistingue all'interno del dominio, da un codice EAN, un costo, da un fornitore ed appartiene ad un specie, ovvero un raggruppamento di prodotti deciso da Orogel. Gli alimenti possono essere prodotti direttamente dall'azienda o ottenuti tramite approvvigionamenti da fornitori esterni e sono poi racchiusi in confezioni, che a loro volta sono inserite in cartoni (identificati anch'essi da un codice EAN), imballati e caricati sui pallet. Di entrambi questi aspetti viene memorizzato il peso e l'altezza, utile per organizzare il trasporto della merce verso i relativi punti di consegna. Ulteriori informazioni aggiuntive si riferiscono all'indicazione dell'IVA d'acquisto e di vendita, alla tipologia del listino di cui un prodotto può fare parte e all'indicazione di promozione che serve a segnalare se l'alimento è venduto in promozione. Il costo industriale invece serve per segnalare i costi delle risorse che vengono destinate alla realizzazione di un prodotto con l'esclusione, quindi, dei costi "generali", quali, ad esempio, di amministrazione, assicurativi, di marketing, etc. e viene utilizzato come punto di riferimento per la determinazione del prezzo di vendita di un alimento. Rientrano in questa cerchia i costi direttamente imputabili al prodotto, ad esempio la materia prima e gli imballaggi, la cui somma è definita tecnicamente "costo primo", e i costi legati alla

produzione che, non essendo “direttamente” assegnabili ai singoli prodotti, vengono imputati a questi ultimi attraverso una operazione di riparto che utilizza, di volta in volta, indicatori “oggettivi”. Si pensi, ad esempio, al costo di ammortamento di un impianto industriale o al costo dell’energia elettrica necessaria per il suo funzionamento che, normalmente, vengono ripartiti utilizzando le “ore macchina” necessarie per la produzione dei singoli prodotti. A fini statistici, come per la gestione dei clienti, i prodotti sono raggruppati in gruppi detti gruppo commerciale e macro gruppo commerciale, mentre per una migliore organizzazione del dominio sono raccolti in famiglie denominate famiglia di vendita, famiglia d’acquisto e famiglia di produzione. Le famiglie raggruppano i prodotti secondo caratteristiche comuni sia dal punto di vista del prodotto in sé sia dal punto di vista commerciale. Infine sono presenti anche gruppi relativi ai prodotti su cui incentivare la vendita come, attualmente, spinaci, lunette, aromi e piselli. Altri possibili raggruppamenti possono essere i gruppi per listino di vendita e i gruppi vendita che contengono ad esempio: confetture foodservice, marchi privati, Orogel retail e primi prezzi. Per ogni alimento è inoltre indicato a che tipo di mercato è destinato, se catering, retail, confetture o discount, ed anche il marchio del prodotto stesso. Per esempio un prodotto può essere venduto con il marchio Orogel oppure può essere Orogel ha produrre un prodotto per un cliente a su cui applicare il marchio del cliente.

2.1.3 Premi

Con il termine premio si intende un corrispettivo in denaro da sottrarre all’acquisto di un determinato prodotto effettuato da un cliente, applicato di norma su base annuale. Per esempio se un acquirente nell’anno passato ha effettuato acquisti superiori ad una certa soglia, alla fine dell’anno gli viene riconosciuto un premio da sottrarre ad eventuali acquisti futuri su quella famiglia di alimenti. Tipicamente questa tipologia di premi assume il nome di premi quantitativi. Diversamente possono essere anche applicati dei premi in base agli incentivi per lanciare un prodotto. Nello specifico Orogel assegna due tipi di premi: premio nazionale e premio listing. Il primo è un premio che è legato ad un contratto quadro. Il contratto quadro è all’apice della gerarchia, quindi tutti i premi di questo tipo sono validi per tutti i clienti statistici legati alle holding. I contratti quadro sono accordi a lungo termine tra un’organizzazione di acquisto e un fornitore, o tra un’organizzazione di vendita e un cliente, per la fornitura di articoli o la distribuzione di servizi per un periodo di tempo in base a termini e condizioni predefiniti. I contratti quadro possono essere utilizzati come base per

previsioni di fatturato, riserva di quantità e pianificazione della capacità. I premi listing invece sono contributi definiti dai singoli capi area sui singoli clienti statistici. Esistono svariati premi che compongono le due tipologie, ad esempio i premi listing possono riferirsi a causali di premio che corrispondono a contributi per primo inserimento di un prodotto in vendita, contributi per mantenimento di un prodotto, contributi per volantini, contributi per aperture e molti altri. I premi nazionali invece possono comprendere premi a contributi promozionali per l'estero, merchandising, premi a concessionari o grossisti, premi sul reintegro sconti, premi per clienti delle grandi distribuzioni organizzate e altri. Per entrambe le categorie, le tipologie di premio non sono fisse ma possono aumentare nel tempo, per esempio un domani si potrebbe assegnare un nuovo premio per una nuova causale di vendita. All'interno del dominio il premio è espresso sia in percentuale sia come valore di importo, comprendendo in alcuni casi anche la valuta monetaria per ordini relativi a clienti esteri e specificando. Ovviamente per ogni prodotto presente sulla riga della fattura sarà esplicitamente indicato nel caso sia incluso nella promozione. I premi possono essere raggruppati in macro raggruppamenti, ad esempio:

- Premio ad agente;
- Premio a cliente;
- Premio a cliente di fatturazione;
- Premio a raggruppamento di clienti;
- Premio a gruppo di acquisto.

Possono inoltre essere raggruppati anche per categorie:

- Articolo con specifica;
- Articolo;
- Raggruppamento commerciale;
- Famiglia commerciale;
- Articolo convenzionale.

2.1.4 Fattura

La fattura è il documento fiscale obbligatorio emesso da un soggetto fiscale per comprovare l'avvenuta cessione di beni o prestazione di servizi ed il diritto a riscuoterne il prezzo. L'operazione di emissione di una fattura prende il nome di fatturazione. La fattura è identificata da un codice esadecimale univoco e da un numero variabile di righe, dove in ognuna è presente l'indicazione della spesa e di eventuali premi applicati. Oltre a questi viene specificato anche il tipo di movimento che può essere a quantità, riferito alla quantità in kg o al numero dei cartoni (l'importo finale viene calcolato quindi moltiplicando tale per l'importo di listino), o a valore. Gli aspetti importanti legati alla fattura sono la data di protocollo, la data di consegna e la data di competenza. La data di protocollo identifica la data in cui la fattura viene registrata presso i registri di Orogel, la data di consegna invece, come intuibile, si riferisce alla data di consegna della merce. La data di competenza invece si riferisce a fatture emesse con una data diversa dalla data di effettuazione della prestazione. Di norma questo tipo di fattura è utilizzata per la fatturazione a fine mese di merce consegnata nel corso del mese solare, in questo modo è possibile avere una sola fattura a fronte di molte spedizioni/consegne al cliente. Supponendo invece che un cliente effettui un ordine il 31 dicembre, la fattura che sarà registrata a gennaio avrà come data di competenza il mese dell'anno precedente.

2.1.5 Vendita

La vendita è l'evento di rilievo nel dominio aziendale che mette in relazione tutte le analisi fatte finora. Ad essa sono legati i vari aspetti della fattura come ad esempio l'importo finale e lordo, il cliente a cui è intestata, la quantità della merce acquistata, l'importo di eventuali provvigioni e le varie date di consegna e competenza.

2.2 Modellazione concettuale: Dimensional Fact Model

Un modello concettuale deve per definizione fornire una serie di strutture atte a descrivere la realtà di interesse, che sia facile da comprendere e che non dipenda dai criteri di organizzazione dei dati

nei calcolatori. Il modello E-R (Entity Relationship) è il modello più diffuso nelle imprese per la progettazione concettuale e documentazione dei data ad un alto livello di astrazione nelle basi di dati relazionali. Purtroppo la rappresentazione tramite schema E-R mal si presta ad un progettazione concettuale a partire dai requisiti utente, in quanto non è in grado di risaltare correttamente gli aspetti peculiari del modello multidimensionale, con l'aggravante inoltre che risulterebbe non adeguato da un punto di vista grafico-notazionale.

Il **Dimensional Fact Model (DFM)** è un modello concettuale grafico specificamente pensato per fungere da supporto alla progettazione di *data mart*, e può essere considerato come una specializzazione del modello multidimensionale per applicazioni di data warehousing. Il **DFM** consente al progettista e all'utente finale di confrontarsi su un modello intuitivo ed affinare le specifiche dei requisiti, risultando utile anche per creare una piattaforma di partenza stabile da cui partire per l'implementazione del modello logico ed ottenere una documentazione espressiva e priva di ambiguità. L'obiettivo del **DFM** è quello di generare schemi di fatto avente come elementi di base i fatti, le dimensioni, le misure e le gerarchie.

2.2.1 Identificazione del fatto

Per definizione, con il termine *fatto* si intende il concetto principale che descrive una categoria di eventi presenti nel dominio aziendale su cui gli utenti baseranno il processo decisionale. All'interno dell'evento risulta essenziale definire un intervallo di storicizzazione che rappresenta l'arco temporale che gli avvenimenti memorizzati dovranno coprire. Requisito fondamentale per un *fatto* è avere aspetti dinamici, che gli consentano di evolversi nel tempo.

L'identificazione del *fatto* nello sviluppo dell'applicazione d'analisi *what-if* per l'azienda Orogel è la vendita, evento fondamentale in quanto ad essa sono riconducibili un insieme di dati basilari quali: il cliente, l'importo, la merce e la quantità venduta. Nonostante la scelta del *fatto* possa apparire scontata non è sufficiente per esprimere un'informazione diretta, per costruire un'analisi più dettagliata sono necessari altri aspetti, nello specifico dimensioni e misure, che ne completino la visione globale.

2.2.2 Identificazione delle misure

Con il termine *misura* si intende una proprietà numerica di un fatto che ne descrive un aspetto quantitativo di interesse per l'analisi, ad esempio l'incasso relativo alla vendita.

Le misure che compongono il fatto sono, come facilmente intuibile, il prezzo finale, l'importo lordo, la quantità della merce acquistata, i premi nazionali e i premi listing. Nel dettaglio i premi nazionali e i premi listing sono ottenuti come somma di vari premi (come precedentemente spiegato).

2.2.3 Identificazione delle dimensioni

Per definizione con il termine *dimensione* si intende una proprietà con dominio finito di un fatto che ne descrive una coordinata di analisi. In uno scenario relativo al mondo vendita, come quello affrontato nello sviluppo della tesi, dimensioni d'esempio possono essere il negozio, il prodotto e la data.

Nella costruzione dello schema di fatto è risultato utile realizzare le seguenti dimensioni:

- Dimensione prodotti: che racchiude al suo interno le informazioni legate al prodotto;
- Dimensione famiglia di vendita: che identifica in famiglie i vari prodotti del dominio. Originalmente questa dimensione era compresa nella dimensione dei prodotti, ma a causa di un errore nel sistema aziendale è stato necessario separarla. Nei dati relativi alla famiglia di vendita presenti nella dimensione dei prodotti, una stessa famiglia era identificata da due codici diversi. Questo caso ha portato a diversi inconvenienti nella assegnazione dei premi simulati, fase presentata nel capitolo successivo.
- Dimensione tempo: utile per tenere la traccia temporale degli acquisti effettuati dai clienti;
- Dimensione punti vendita: dove sono racchiuse le informazioni dei vari clienti statistici;
- Dimensione premi simulati: nella quale sono salvate le simulazioni sui premi inseriti da parte dell'utente;
- Dimensione clienti fatturazioni: che racchiude gli intestatari delle fatture.
- Dimensioni area business: che contiene al suo interno le principali aree business del client

- Concessionari;
- Confetture;
- Estero;
- Grande distribuzione/distribuzione organizzata;
- Grande distribuzione organizzata;
- Industria.

2.2.4 Costruzione delle gerarchie

Come riportato in teoria, la gerarchia è un albero direzionato i cui nodi sono attributi dimensionali e i cui archi modello associazioni molti a uno, dove per attributo dimensionale si intendono altre dimensioni che descrivono in maniera dettagliata la dimensione padre. All'apice di un gerarchia vi è presente una dimensione e di seguito tutti gli attributi dimensionali che la descrivono. Ad esempio in un dominio relativo al prodotto alcuni attributi dimensionali possono essere la marca e il tipo.

Di seguito sono riportate nel dettaglio le gerarchie presenti nello schema di fatto.

La gerarchia relativa al tempo ospita al suo interno il mese e l'anno, questo perché la fattura viene gestita in base al mese di competenza. L'anno risulta fondamentale in quanto l'analisi viene svolta sulla base dell'anno corrente a confronto del precedente.

La gerarchia relativa al punto vendita, identificata dal codice del cliente statistico, presenta tutte le caratteristiche inerenti all'acquirente come il codice personale, descrizione, città e canale di vendita.

La gerarchia prodotti contiene al suo interno le informazioni necessarie al prodotto che comprendono i codici dei prodotti, il marchio, il gruppo della famiglia di vendita, la descrizione e la ragione commerciale.

La gerarchia delle famiglie di vendita ospita al solamente il codice e la descrizione della famiglia, per il motivo spiegato precedentemente.

La gerarchia pertinente alle aree business contiene al suo interno solamente il codice dell'area business del cliente e la descrizione della stessa.

La gerarchia relativa ai premi simulati, in relazione al codice del cliente di fatturazione e al codice della famiglia, racchiude le informazioni riguardanti le percentuali degli inserimenti effettuati dai

clienti e i flag relativi agli inserimenti stessi. Sebbene questi ultimi aspetti possano apparire come misure, comportando quindi la loro presenza direttamente all'interno del fatto, per motivi applicativi nello sviluppo dell'applicazione questo non è stato possibile. Inserendo questi campi come misure all'interno del fatto e lanciando lo script contenente la *query* su QlikView, si presentava un errore relativo a record duplicati. L'unico modo per arginarlo è stato quello di considerare questi campi come attributi dimensionali.

2.2.5 Schema di fatto

Le analisi svolte nei paragrafi precedenti hanno contribuito a determinare quello che è stato lo schema di *fatto* finale. Attraverso la seguente modellazione è possibile ottenere tutte le informazioni relative al fatturato dei clienti dell'anno precedente, in relazione alle famiglie dei prodotti acquistate, e sull'anno in corso, ovviamente ancora in forma parziale. Inoltre grazie alla modellazione effettuata la relazione che si viene a creare attraverso il *fatto*, tra clienti di fatturazione e i punti di vendita, consente di ottenere solamente i clienti statistici che hanno effettuato gli acquisti nel periodo di tempo selezionato. Nell'applicativo sviluppato risulterà quindi normale avere, nella visualizzazione dei dati per i due anni di analisi, delle famiglie di vendita non ancora valorizzate per i prodotti venduti nel 2015. Questa differenza è destinata ad assottigliarsi con il passare del tempo man mano che i clienti effettuano nuovi acquisti, che saranno poi registrati da Orogel nel sistema. La relazione presente tra la dimensione dei prodotti e la famiglia assicura di includere nell'analisi solamente le famiglie e gli alimenti effettivamente venduti. Attraverso la modellazione ottenuta è quindi possibile effettuare analisi diverse oltre alla gestione del cliente, sia sulla base del codice del prodotto sia sulla base del codice della famiglia di vendita, valutando premi e fatturato per questi due dati. E' inoltre possibile ottenere anche l'andamento dei premi sulla base del fatturato nell'anno precedente, che può essere d'aiuto all'analista nella scelta del futuro premio da assegnare durante lo studio d'analisi. Altri studi possono essere effettuati filtrando i dati per canale di vendita, holding di appartenenza del cliente, mercato di vendita o per i gruppi di famiglie o gruppi di vendita. Eventualmente un sottoinsieme di interesse di analisi dei clienti può essere selezionato sulla base degli agenti e dei capi area legati all'acquirente. Di seguito è mostrato lo schema di fatto ottenuto dall'analisi.

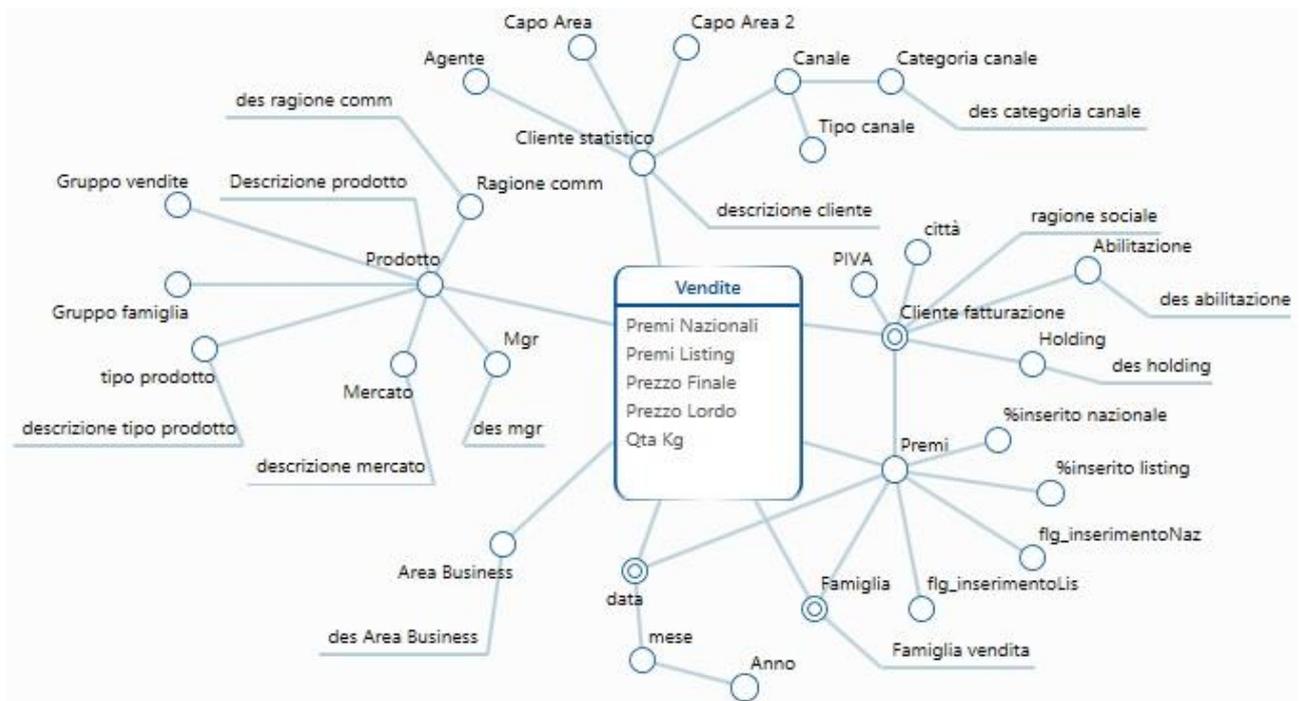


FIGURA 3. MODELLO CONCETTUALE DEL PROBLEMA ANALIZZATO

Attraverso le misure presenti nello schema è possibile calcolare i valori in percentuale delle due tipologie di premio dividendo i rispettivi premi del cliente per il suo fatturato complessivo in un determinato anno. La misura *Prezzo Finale* in relazione alla famiglia dei prodotti dà la possibilità di calcolare il fatturato complessivo per una determinata famiglia di vendita di un cliente, la somma degli importi di tutte le famiglia costituisce quindi il prezzo d'acquisto complessivo compiuto dal cliente. La misura della quantità della merce può essere utilizzata come informazione per analizzare il rapporto costo e quantità degli alimenti acquistati dal cliente.

CAPITOLO 3

MECCANISMO DI SIMULAZIONE

Nel corso del seguente capitolo sono presentati i passi che hanno contribuito alla creazione dell'applicazione. All'inizio della sezione è presente una breve introduzione dei principali strumenti utilizzati per lo svolgimento della tesi, strumenti che nell'organizzazione aziendale interagiscono fra loro come mostrato nella successiva immagine.

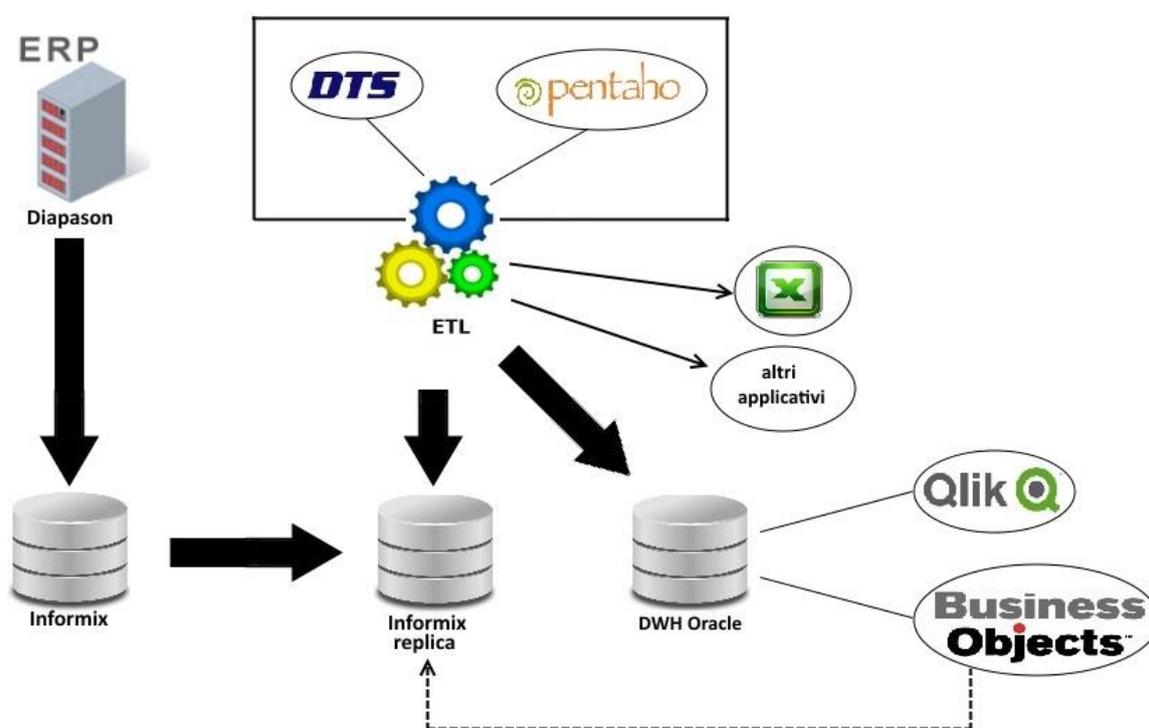


FIGURA 4. INTERAZIONE STRUMENTI AZIENDALI

Il sistema **ERP** funge da alimentatore per il database **Informix**, presente nel sistema in duplice copia. **Informix** costituisce per l'azienda **Onit Group srl** il cuore del sistema, dove sono archiviati tutti i dati usufruibili per le varie operazioni. Nel dominio aziendale ne esistono 2 entità, la seconda non è altro che una replica dell'originale, dove tutte le modifiche apportate su di essa non si ripercuotono sulla controparte reale. Lo strumento **ETL** procederà dunque a leggere e a caricare i dati sia sulla copia di **Informix** che sul **DWH Oracle** e inoltre, quando necessario, potrà esportare il risultato

di

una procedura su pagine Excel o altri applicativi. All'interno dell'azienda sono utilizzati due strumenti **ETL**, **DTS** e **Pentaho**. **DTS** è un vecchio strumento **ETL** di casa Microsoft incluso nelle precedenti versioni di SQL Server, utilizzato per eseguire all'incirca il 10% delle operazioni totali di estrazione, trasformazione e caricamento. Il restante 90% del lavoro è svolto da **Pentaho**. A livello di presentazione dei dati sono utilizzati due programmi: **QlikView** e **Business Objects**, con la particolarità che quest'ultimo può scrivere alcuni record (nella pratica pochi) sul DB **Informix**. **Business Objects** è un prodotto software dedicato alla business intelligence e, pertanto, permette di produrre una analisi ed una reportistica sui dati molto evoluta rispetto a prodotti generici. Differentemente da **QlikView**, che offre una rappresentazione grafica per la visualizzazione di elementi numerici come possono essere la marginalità, il fatturato o altri fattori come l'insieme dei clienti e prodotti coinvolti nell'analisi, **Business Object** offre una raffigurazione più tabellare mirata a descrivere le relazioni tra le dimensioni nel dominio aziendale. Gli strumenti usati per lo svolgimento della tesi sono **Pentaho**, **Toad for Oracle** e **Qlikview**.

La seconda parte del capitolo è incentrata sullo sviluppo dell'applicazione, analizzando nel dettaglio le varie operazioni che hanno contribuito ad ottenere il prodotto finale. La descrizione avviene anche per mezzo di immagini che mostrano alcuni passaggi dello sviluppo, usate soprattutto per la rappresentazione di alcuni passaggi come le procedure **ETL** e le *query* di generazione dei dati.

3.1 Strumenti utilizzati

Nei successivi sotto-paragrafi sono presentate e descritti i principali strumenti utilizzati per lo sviluppo della tesi.

3.1.1 Pentaho

Pentaho è una piattaforma di *Business intelligence* in grado di proporre una soluzione migliore rispetto alle offerte commerciali in termini di caratteristiche, funzioni e benefici, attraverso l'integrazione in un'unica piattaforma di popolari progetti Open Source come *JFreeReport*, *Kettle*, *Mondrian* e *Weka*. Le tecnologie utilizzate sostengono un'ampia serie di operazioni legate al business come l'analisi delle vendite e del profitto, l'analisi del cliente, *HR reporting*, *Financial reporting*, *KPI dashboards*, analisi della *Supply Chain* e segnalazioni operative. Differentemente

dai prodotti tradizionali in Pentaho le strutture ROLAP di un progetto vengono caricate in RAM in modo che possano essere gestite con prestazioni simili ai moderni motori *MOLAP* (*Multidimensional OLAP*). In questo caso si parla quindi *HOLAP*, sigla che indica un sistema ibrido (*Hybrid OLAP*) dal quale la stessa società che ha sviluppato il programma prende il nome, **Penta-ho** [9].

Il programma si collega ad **informix** per prelevare dei dati, offrendo un'interfaccia semplice che aiuta l'utente nella costruzione della procedura di caricamento delle informazioni su data base. Per alimentare le future tabelle del DB è necessario scegliere come primo passo la metodologia di importazione dei dati, selezionando dal riquadro degli oggetti a sinistra il tipo di caricamento. **Pentaho** mette a disposizione dell'utente diverse opzioni tra cui: caricamento dati da Access, da file CSV, Excel, Json e molti altri. Per quanto concerne l'applicativo svolto la maggior parte dei caricamenti sono stati svolti tramite l'operazione classica di lettura da DB denominata "Table input", attraverso la quale, una volta inserita nel piano di lavoro, è sufficiente digitare la propria *query* per selezionare i dati da leggere. Di norma, una volta caricate le informazioni, si procede ad apportare degli aggiustamenti ai dati letti attraverso operazioni di *scripting*, di selezione dei valori e operazioni di join per il recupero dei dati sulla base di determinate chiavi. Tramite script è possibile creare modifiche di alto livello non legate al solo linguaggio *query SQL*, mentre attraverso l'opzione di selezione dei valori è possibile cambiare i nomi dei campi ottenuti dal caricamento da DB. Ovviamente sono disponibili una moltitudine di altre funzioni che però non sono state utilizzate nello svolgimento della tesi. Una volta completate le modifiche si procede a popolare le tabelle del DB attraverso l'opzione di output preferita tra le tante offerte. Tra le svariate opzioni a disposizione dell'utente figurano ad esempio operazioni di "Table output" o operazioni di "Excel output". Indipendentemente dal tipo di oggetto scelto occorre sempre indicare all'interno del menù di opzioni degli stessi la connessione con il data base, specificando anche la relativa tabella di destinazione.

3.1.2 Toad for Oracle

Toad for Oracle è un software che permette di visualizzare i dati contenuti nel data base **Oracle** aziendale. **Oracle** è uno tra i più famosi DBMS esistenti e fa parte dei cosiddetti RDBMS, ovvero sistemi database basati sul modello relazione, un modello logico di rappresentazione o di strutturazione dei dati di un database. All'interno del dominio aziendale **Oracle** ospita tutti i dati ottenuti dalla procedura di **ETL** che saranno utili per i futuri svolgimenti. Attraverso Toad è quindi

possibile visualizzare tutte le tabelle presenti al suo interno, offrendo all'utente anche la possibilità di effettuare interrogazioni sui dati tramite *query SQL*, utili per la visualizzazione dei dati memorizzati e per controllarne la correttezza, o semplicemente visionare i valori dei campi scorrendo l'apposita finestra di visualizzazione.

3.1.3 QlikView

QlikView è una piattaforma di *Business Intelligence* leader mondiale nelle tecniche di analisi *in-memory* con il più rapido tasso di diffusione al mondo, che propone una nuova classe di soluzioni analitiche facili da utilizzare, veloci e flessibili. Il prodotto supporta gli individui nel miglioramento delle prestazioni delle organizzazioni nelle quali operano e nel governo dell'innovazione e del cambiamento. Un'altra caratteristica di rilievo è di permettere di utilizzare un'innovativa tecnologia associativa brevettata, che elabora dinamicamente i dati in memoria, per rendere drasticamente più semplice lo sviluppo, l'utilizzo e la manutenzione di sofisticate applicazioni di analisi e reporting.

In sostanza **QlikView** consente all'utente di ottenere una visualizzazione grafica dei dati memorizzati in un database attraverso tabelle, schemi e altri componenti. La lettura delle informazioni avviene tramite la scrittura, in linguaggio *query SQL*, delle "dimensioni" o dei dati che si desidera includere nella funzione di script. Una volta che lo script è eseguito e le tabelle scelte caricate in memoria, all'utente si presentano una moltitudine oggetti per la rappresentazione delle informazioni. Oltre alle comuni tabelle di semplice visualizzazione sono disponibili strumenti per la selezione e il filtraggio dei campi, pulsanti per interagire con l'applicativo, caselle per effettuare inserimenti e svariate altre funzioni. La particolarità di **Qlikview** non è solamente quella di limitarsi a una mera rappresentazione dei campi scelti, ma permette anche allo sviluppatore di poter interagire tramite la scrittura di codice con i dati. In questo modo è possibile ottenere una maggior personalizzazione di quello che si vuole sviluppare, per esempio in una tabella potrebbe essere utile far apparire determinati dati, ottenuti tramite una serie di calcoli in base alle selezioni effettuate dall'utente, o anche eseguire controlli su determinati valori presenti nell'applicativo. E' possibile quindi sviluppare tutto quello che si vuole avendo come limite solamente la conoscenza del linguaggio di programmazione.

Oltre a quanto detto è possibile assegnare determinate funzioni ai pulsanti. **QlikView** offre funzioni predefinite da assegnare ai tasti, come per esempio operazioni di selezione e cancellazione dei campi, attivazione di fogli e oggetti e svariate altre funzioni. Per operazioni più complesse invece è

possibile sviluppare delle macro sia in linguaggio Visual Basic sia in Java script. In questo modo si dà la possibilità allo sviluppatore di creare determinate funzioni che contribuiscono a migliorare le funzionalità offerte dal programma.

3.2 Sviluppo dell'applicazione

Lo scopo della tesi è stato quello di realizzare un'applicazione di analisi *what-if* per consentire alla direzione commerciale di Orogel di prendere decisioni rapide e precise sulla base di simulazioni effettuate tramite l'assegnamento di premi di vendita ai clienti. Il prodotto finale, come esplicitamente richiesto, è stato sviluppato attraverso il software Qlikview, programma largamente diffuso all'interno dell'azienda su cui l'utenza di Orogel ha già una buona familiarità. La simulazione è svolta assegnando le percentuali di premio sulla famiglia dei prodotti e visionando in tempo reale come variano i dati in base agli input inseriti dagli utenti. Il data entry è reso possibile per mezzo di determinate funzioni scritte all'interno delle colonne della tabella, dove oltre agli input dell'utente sono presenti altre informazioni relative al fatturato, ai premi listing e ai premi nazionali, sia dell'anno precedente che dell'anno in corso. Nello specifico l'utente dovrà inserire, nelle celle delle rispettive colonne, le percentuali dei premi nazionali e dei premi listing, utili per dare il via alla simulazione. Per una maggiore comprensione e revisione dell'analisi che si sta svolgendo sono presenti svariati grafici che mettono a confronto i valori dei due anni in esame e che aiutano gli utenti nell'individuare immediatamente i vari cambiamenti che si presentano. In questo modo è possibile attuare degli studi mirati ed efficaci. Mano a mano che l'utenza della direzione commerciale di Orogel esegue le simulazioni, per tenere traccia degli inserimenti è stato sviluppato una funzionalità in grado di mettere in comunicazione il programma Qlikview con il database Oracle, comunicazione che si traduce in un salvataggio delle prove eseguite su DB. Così facendo quando si presenterà il bisogno, gli utenti potranno caricare i dati precedentemente inseriti e continuare il lavoro dal punto dove l'avevano lasciato. Il programma sviluppato è pensato per essere usato in quel periodo di transizione in cui non sono stati assegnati ancora i premi per l'anno in corso, operazione che di solito dovrebbe concludersi intorno la fine mese di aprile/maggio di ogni anno. Dal momento che è possibile che alcuni premi non siano stati ancora valorizzati nell'anno in corso, è necessario per questi casi ribaltare, ovvero copiare, i premi dell'anno precedente in quello attuale. Nel successivi paragrafi è spiegato il funzionamento di questa procedura.

3.2.1 Proiezione premi

La procedura di proiezione dei premi avviene in due fasi, una prima fase è mirata a copiare i premi, in questo caso copiare i dati del 2014 nel 2015, nella apposita tabella presente sul data base Oracle, mentre un secondo step consiste nel prendere i premi, relativi al 2015, direttamente forniti dalla direzione commerciale Orogel e inserirli in tabella. Ovviamente se è presente per uno stesso cliente e una stessa famiglia un premio obsoleto questo andrà sostituito con i file Excel forniti da Orogel. Di seguito sono analizzate le due procedure adibite allo scopo appena descritto.

3.2.1.1 Fase ribaltamento dell'anno precedente

La procedura Pentaho che si occupa di ribaltare i premi dell'anno precedente in quello successivo è articolata su due vie, come mostrato nell'immagine sottostante.

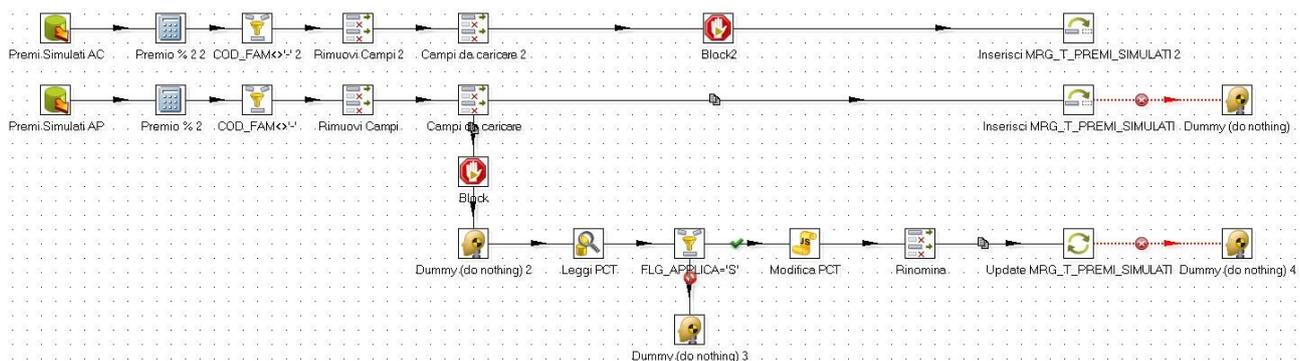


FIGURA 5. PROCEDURA RIBALTAMENTO PREMI

Il primo compito consiste nel recuperare le informazioni dal database attraverso l'oggetto *table input* dove tramite una *query* SQL si indicano i campi di una tabella presente sul data base da cui recuperare i dati. Come si può ben vedere sono presenti due operazioni di importazione dei dati, una per le informazioni dell'anno in corso (*premi simulati AC*) l'altra per quelle inerenti l'anno precedente (*premi simulati AP*). La *query* che ha il compito di prelevare le informazioni dal sistema è pressoché identica per entrambi gli oggetti, l'unica differenza è sul filtro per la data. Poiché la data è codificata come tipo "number" e non come tipo "date" è necessario sottrarre 10.000 per avere, dato un giorno preciso, il suo corrispondente relativo all'anno precedente. Ad esempio considerando come data 20150210 (10 febbraio 2015) e sottraendo a questo numero 10.000 il risultato è 20140210, ovvero 10 febbraio 2014. Nel caso di importazioni dei dati per l'anno

corrente, questa condizione non è presente nella *query*. Di seguito è possibile notare la *query* utilizzata per l'operazione.

```
-- Estrazione dati dal 2014 per irbaltare nel 2015
SELECT dimaziende.idazienda as COD_AZIENDA,

    TO_DATE(${p_data_simula_dal}, 'YYYYMMDD') AS DATA_INIZIO,
    TO_DATE(${p_data_simula_al}, 'YYYYMMDD') AS DATA_FINE,
    dimprodotti.idfamvendita AS COD_FAM_VENDITA,
    dimprodotti.idgrpvendita as COD_GRP_VENDITA,
    dimclientifatt.idclientefatt AS COD_CLIENTE,
    SUM(fctvendite.prezzofinale) as FATTURATO,
    SUM(fctvendite.premio1 + fctvendite.premio2 + fctvendite.premio3 + fctvendite.premio6
    + fctvendite.premio7 + fctvendite.premio8 + fctvendite.premio9 +
    fctvendite.premioA + fctvendite.premioALTRO + fctvendite.premioB +
    fctvendite.premioD + fctvendite.premioE + fctvendite.premioQ + fctvendite.premioR + fctvendite.premioS +
    fctvendite.premioT + fctvendite.premioU + fctvendite.premioZ + fctvendite.premioY + fctvendite.premioK) as PREMI_NAZIONALI,
    SUM(fctvendite.premio0+fctvendite.premio4 + fctvendite.premio5 + fctvendite.premioF +
    fctvendite.premioG + fctvendite.premioH + fctvendite.premioI + fctvendite.premioN + fctvendite.premioV +
    fctvendite.premioX) as PREMI_LISTING

FROM dimaziende,
    dimclientifatt,
    dimprodotti,
    fctvendite,
    dimpuntivendita
WHERE fctvendite.keyprodotto = dimprodotti.keyprodotto
AND fctvendite.keyazienda = dimaziende.keyazienda
AND fctvendite.keyclientefatt = dimclientifatt.keyclientefatt
AND fctvendite.keypuntovendita = dimpuntivendita.keypuntovendita
AND fctvendite.keyclientefatt != (SELECT keyclientefatt FROM dimclientifatt WHERE idclientefatt='3010445')
AND dimaziende.idazienda like '${p_azienda}'
AND keymese BETWEEN (SELECT keymese FROM dimgiorni WHERE keygiorno = ${p_data_simula_dal}-10000)
AND (SELECT keymese FROM dimgiorni WHERE keygiorno = ${p_data_simula_al}-10000)
AND '${p_simula}' = 'S'

GROUP BY dimaziende.idazienda,
    dimclientifatt.idclientefatt,
    dimprodotti.idfamvendita,
    dimprodotti.idgrpvendita
HAVING SUM(fctvendite.prezzofinale) != 0
```

FIGURA 6. QUERY PENTAHO RECUPERO PREMI ANNO 2014

Come si può notare, nella *query* sono stati recuperati, dalle varie dimensioni, i dati relativi al codice cliente, alla famiglia dei vendita, alla data, al fatturato e ai premi.

Per ottenere, dalla procedura Pentaho in figura 2, la percentuale del premio è stato necessario, attraverso l'oggetto che figura la calcolatrice adibito all'esecuzione di calcoli, effettuare la divisione fra il fatturato e i rispettivi premi (listing e nazionali). Di seguito i dati sono stati filtrati per consentire di operare solamente sulle famiglie di vendita che sono valorizzate, quindi escludendo quelle a valore nullo, e sono stati rimossi i campi relativi al codice del gruppo vendita, fatturato, premi nazionali e premi listing in quanto non più utili per la procedura. Tale scelta è utile ogni qualvolta sia necessario far risparmiare memoria al programma, evitando quindi di sovraccaricare le procedure e snellirne lo svolgimento. Nell'operazione seguente invece vengono esplicitamente indicati i campi da caricare utili al resto della procedura. L'oggetto raffigurante l'alt funge da semaforo, arbitrando l'esecuzione delle operazioni per evitare che si presentino sovrascrizioni dei dati indesiderate. L'alt denominato *block* resterà bloccato finché non sarà eseguita la parte arbitrata dal semaforo *block2*, a sua volta arrestato finché non viene eseguito il primo inserimento sulla tabella. In questo modo si procede dapprima a caricare i premi del 2014 e poi i premi del 2015, applicando le dovute sovrascrizioni nel caso siano presenti premi assegnati ad un famiglia di

prodotti nell'anno precedente. Gli oggetti di tipo dummy non hanno nessuna funzione, sono utilizzati esclusivamente a scopo visivo per organizzare al meglio il piano di lavoro. Il resto delle operazioni successive allo stop *block* consistono nel pescare i campi relativi alle percentuali inserite e ai flag degli inserimenti tramite un'operazione di lockup, filtrare i valori dove i flag siano uguali al carattere "S"(come spiegato nell'analisi delle gerarchie del capitolo precedente), aggiornare il valore della percentuali e scrivere sulla tabella del database il risultato dei seguenti passi.

3.2.1.2 Fase premi anno in corso

In questo step avviene il caricamento dei premi già stabiliti per l'anno in corso che Orogel ha fornito tramite due file excel al reparto *BU* di *Onit*. L'obiettivo è di inserire sulla tabella del database i premi aggiornati, sovrascrivendo nel caso siano presenti per una stessa famiglia i premi dell'anno precedente con quelli dell'anno attuale. Se non si trova un riscontro quindi, fintanto che non saranno definiti i nuovi premi, saranno utilizzati per il 2015 i premi del 2014. La procedura che si occupa di eseguire queste operazioni è mostrata di seguito.

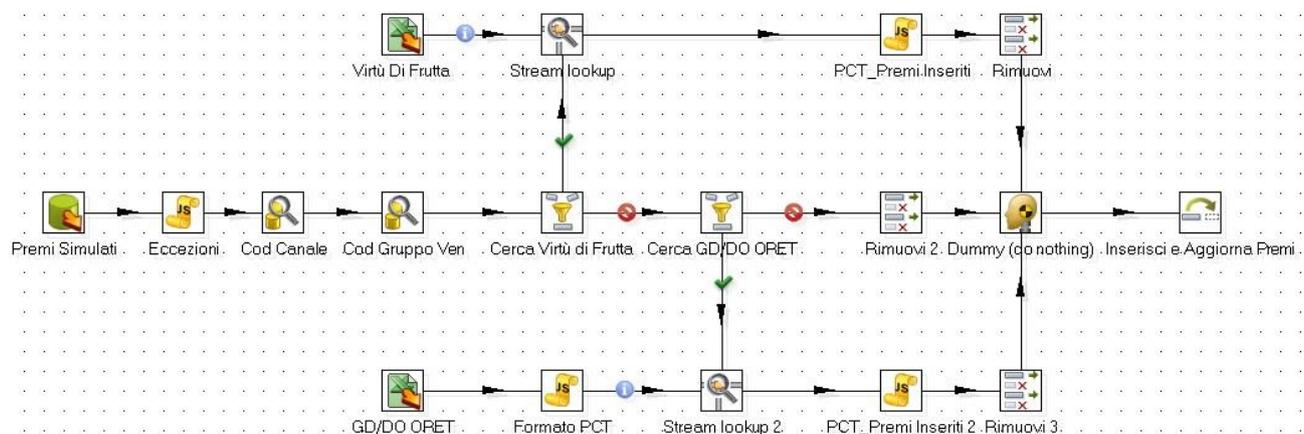


FIGURA 7. AGGIORNAMENTO PREMI DA FILE EXCEL

Attraverso l'operazione di "table input", denominata Premi Simulati, si recupereranno i dati precedentemente inseriti su DB dalla procedura vista nel paragrafo sovrastante, tramite la query mostrata di seguito.

```

SELECT  COD_AZIENDA,
        COD_CLIENTE,
        COD_FAM_VENDITA,
        DATA_INIZIO,
        DATA_FINE,
        PCT_SIMULATO_NAZ,
        PCT_INSERTITO_NAZ,
        FLG_APPLICA_NAZ,
        PCT_SIMULATO_LIS,
        PCT_INSERTITO_LIS,
        FLG_APPLICA_LIS

FROM    MRG_T_PREMI_SIMULATI

WHERE   DATA_INIZIO BETWEEN TO_DATE('${p_data_dal}', 'DD/MM/YYYY')
        and to_date('${p_data_al}', 'DD/MM/YYYY')

```

FIGURA 8. QUERY CARICAMENTO DATI PER AGGIORNAMENTO PREMI

L'oggetto di script denominato *eccezioni* è utilizzato per l'azzeramento dei premi di determinati clienti direttamente specificati dalla direzione commerciale di Orogel, in quanto non più di interesse per l'analisi. Una volta caricati i dati tramite gli oggetti di *lookup* si recupera il codice del canale e il codice del gruppo di vendita, utili nelle successive operazione di *filtering* dei dati. Attraverso l'operazione di filtraggio la procedura si divide in 3 strade, una parte si interfaccia con i dati provenienti dall'Excel inerenti alla linea della famiglia dei prodotti virtù di frutta, l'altra strada invece è messa in relazione con i dati provenienti dall'Excel per la linea di prodotti grandi distribuzioni e distribuzioni organizzate. Il *join* tra i dati caricati avviene per entrambe le vie con l'oggetto *stream lookup*. L'unica differenza fra le due vie è che i dati del file Excel, relativo alle grandi distribuzioni non sono forniti in percentuale, e quindi è necessario attraverso un semplice calcolo ottenere la dovuta informazione. Il resto della procedura continua nella medesima maniera per entrambe le vie, assegnando il nuovo valore dei premi e rimuovendo i campi non più utili. Alla fine entrambi i dati saranno inseriti in nella tabella del database.

3.3 Applicazione QlikView

Lo sviluppo dell'applicativo QlikView è fortemente basato sulle varie analisi condotte fin ora, in quanto per poter rappresentare le informazioni occorre importare tutti le strutture definite nella fase di modellazione concettuale. All'interno del programma quindi saranno presenti le stesse dimensioni che compongono lo schema di fatto. Il caricamento dei dati avviene tramite uno strumento di *scripting* offerto dal programma che consente all'utente di scrivere le *query* per la generazione dei dati, utili a per lo sviluppo dell'applicativo. Per permettere ciò è necessario stabilire un'apposita connessione con il database da cui si intende prelevare le informazioni, per lo

svolgimento della tesi è stato necessario stabilire una connessione di tipo **OLEDB**. Nel momento in cui si riportano le dimensioni nello script, bisogna tenere a mente che una volta lanciata l'esecuzione queste gestite in maniera sequenziale, occorre quindi prestare attenzione all'ordine disposto. Nel paragrafo successivo sono analizzate le *query* delle varie dimensioni importate.

3.3.1 Caricamento dimensioni

La *query* relativa al caricamento della dimensione tempo recupera i dati filtrando per l'anno maggiore uguale al 2014. L'istruzione di **Load**, propria del programma, specifica quali campi importati dalla *query* caricare in memoria, consentendo anche di modificarne i nomi. La chiave **keymesecompetenza** è utilizzata per mettere in relazione la dimensione con la *fact table*.

La *query* adibita al caricamento della famiglia di vendita preleva il codice e la descrizione, collegandosi alla *fact table* attraverso la chiave **keyfamvendita**. Nell'immagine sottostante è possibile visionare quanto detto fin ora.

<pre> DIMTEMPO: LOAD Distinct KEYMESE as KEYMESECOMPETENZA, MESE as Mese, Left(MESE,3)as "Mese Brv", ANNO as Anno, ; SQL SELECT * FROM DWHDB.DIMMESI WHERE ANNO >= 2014 ORDER BY KEYMESE; </pre>	<pre> DIMFAMVENDITA: LOAD KEYFAMVENDITA, FAMVENDITA as Fam_Vendita, IDFAMVENDITA as Cod_Famiglia, IDFAMVENDITA &' - '& FAMVENDITA as [Famiglia]; SQL SELECT KEYFAMVENDITA, IDFAMVENDITA, FAMVENDITA FROM DWHDB.DIMFAMVENDITA ORDER BY FAMVENDITA; </pre>
--	---

FIGURA 9. A SX IL CARICAMENTO DELLA DIMENSIONE TEMPO, A DX IL CARICAMENTO DELLA DIMENSIONE FAMIGLIA DI VENDITA

La *query* per il caricamento dei prodotti preleva dal database le informazioni inerenti al codice di un prodotto e alla sua descrizione, al marchio, al tipo del prodotto e molte altre informazioni visibili nelle immagine sottostante. Il collegamento alla *fact table* avviene attraverso la chiave **keyprodotto**.

<pre> DIMPRODOTTI: LOAD Distinct KEYPRODOTTO, COD_PRODOTTO as "Cod Prodotto", PRODOTTO as [Des Prodotto], PRODOTTO &' - '& COD_PRODOTTO as Prodotto, COD_MERCAIO as "Cod Mercato", MERCATO as Mercato, COD_RAGCOM,RAGCOM,COD_MRG,MGR, COD_MARCHIO as "Cod Marchio", MARCHIO as "Marchio", COD_TIPO_PRODOTTO as "Cod Tipo Prodotto", TIPO_PRODOTTO as "Tipo Prodotto", COD_GRUPPO_FAMIGLIA as "Cod Gruppo Famiglia", GRUPPO_FAMIGLIA as "GRUPPOVENDITE" ; </pre>	<pre> SQL SELECT KEYPRODOTTO, IDPRODOTTO AS COD_PRODOTTO, PRODOTTO, "FV_IDMERCATO" AS COD_MERCATO, FV_MERCATO AS MERCATO, FAMVENDITA, IDFAMVENDITA, IDRAGCOMMERCIALE AS COD_RAGCOM, RAGCOMMERCIALE AS RAGCOM, IDMACRORAGCOMMERCIALE AS COD_MRG, MACRORAGCOMMERCIALE AS MGR, IDMARCHIO AS COD_MARCHIO, MARCHIO AS MARCHIO, IDTIPOPRODOTTO AS COD_TIPO_PRODOTTO, TIPOPRODOTTO AS TIPO_PRODOTTO, IDGRPVENDITA AS COD_GRUPPO_FAMIGLIA, GRPVENDITA AS GRUPPO_FAMIGLIA, FROM DWHDB.DIMPRODOTTI ORDER BY PRODOTTO; </pre>
---	--

FIGURA 10. CARICAMENTO DIMENSIONI PRODOTTI, A SX OPERAZIONE LOAD A DX QUERY SQL

Il caricamento della dimensione dei clienti intestatari si occupa invece di importare pochi campi, tra i quali: il codice e la descrizione del cliente, la città, la partita IVA e le informazioni relative alla holding di appartenenza e all'abilitazione. La relazione con la *fact table* si ha tramite la chiave **keyclientefatt**.

<pre> DIMCLIENTEFATT: LOAD Distinct KEYCLIENTEFATT, IDCLIENTEFATT, RAGIONESOCIALE as "Ragione Sociale Cliente fatturazione" CITTA, PIVA, IDHOLDING, HOLDING as Holding, IDABILITAZIONE, ABILITAZIONE; </pre>	<pre> SQL SELECT KEYCLIENTEFATT, IDCLIENTEFATT, RAGIONESOCIALE, CITTA, PIVA, IDHOLDING, HOLDING, IDABILITAZIONE, ABILITAZIONE FROM DWHDB.DIMCLIENTEFATT; </pre>
--	---

FIGURA 11. CARICAMENTO DIMENSIONE CLIENTI FATTURAZIONE, A SX OPERAZIONE LOAD A DX QUERY SQL

La *query* che importa i campi della dimensione dei punti vendita nel sistema si occupa di prelevare tutte le informazioni inerenti al cliente statistico tra cui il canale, il codice e la descrizione del cliente, gli agenti e i capi area, la provincia e la nazione. Tramite la chiave **keypuntovendita** la dimensione dei punti vendita si collega alla *fact table*.

<pre> DIMPUNTIIVENDITA: LOAD Distinct KEYPUNTOVENDITA, AGENTE as Agente, CANALE, CAPOAREA as "Capo Area",CAPOAREA2 as "Capo Area2", COD_CLIENTE as Cod_Cliente, CLIENTESTAT as "Des_Cliente", CLIENTESTAT &' - '& COD_CLIENTE as Cliente, TIPOCANALE as "Tipo Canale", IDCATCANALE as "Cod Categoria Canale", CATCANALE as "CATEGORIACANALE", IDPROVINCIA as "Cod Provincia", PROVINCIA AS "Provincia", IDNAZIONE as "Cod Nazione",NAZIONE as "Nazione"; </pre>	<pre> SQL SELECT KEYPUNTOVENDITA, AGENTE,CANALE, CAPOAREA, CAPOAREA2, IDCLIENTESTAT AS COD_CLIENTE,CLIENTESTAT, TIPOCANALE, IDCATCANALE, CATCANALE, IDPROVINCIA, PROVINCIA, IDNAZIONE, NAZIONE FROM DWHDB.DIMPUNTIIVENDITA ORDER BY CLIENTESTAT; </pre>
---	---

FIGURA 12. CARICAMENTO DIMENSIONE CLIENTI FATTURAZIONE, A SX OPERAZIONE LOAD A DX QUERY SQL

Per quanto riguarda le aree business i campi da importare sono relativamente pochi, si caricano nel sistema le informazioni riguardanti il codice dell'area business con la sua descrizione e si imposta il collegamento alla fact table tramite la chiave **keyareabusiness**.

<pre> DIMAREEBUSINESS: LOAD KEYAREABUSINESS, IDAREABUSINESS as CODAREABUSINESS, AREABUSINESS as AreaBusiness; </pre>	<pre> SQL SELECT KEYAREABUSINESS, IDAREABUSINESS, AREABUSINESS FROM DIMAREEBUSINESS; </pre>
--	---

FIGURA 13. CARICAMENTO DIMENSIONE AREE BUSINESS

Differentemente dai casi precedenti la dimensione relativa ai premi simulati non ha un chiave esplicita per collegarsi alla *fact*, la chiave è stata quindi creata combinando i codici del cliente con il codice della famiglia di vendita attraverso il comando **AutoNumber**. Logicamente per consentire il collegamento alla *fact table* la stessa operazione è stata eseguita all'interno della rispettiva **Load**. I campi relativi agli inserimenti che possono effettuare gli utenti sono *pct_simulato_naz* e *pct_simulato_lis* che al primo caricamento contengono le percentuali dei premi inseriti dalla procedura creata su Pentaho. Le percentuali dei premi caricati sono troncate al secondo numero dopo la virgola perché non avrebbe senso visualizzare una percentuale con un maggiore precisione decimale. I due campi *flg_inserimentoNaz* e *flg_inserimentoLis* inizializzati a 1 fungono da *flag* e hanno il compito, all'interno dell'applicazione, di arbitrare la simulazione, tramite il valore 1 la simulazione userà i premi inseriti dall'utente altrimenti aggiornerà i valori con i premi reali. Per consentire all'utente di effettuare gli inserimenti dei dati, i campi relativi ai *flag* e alle immissioni delle percentuali sono stati resi modificabili attraverso la funzione **inputfield**. Ovviamente man

mano che la simulazione prosegue e gli inserimenti salvati, ad ogni nuovo caricamento dello script i campi *pct_simulato_naz* e *pct_simulato_lis* conterranno le informazioni salvate inserite dagli utenti.

```

INPUTFIELD PCT_INSERITO_LIS,PCT_INSERITO_NAZ,
flg_inserimentoNaz,flg_inserimentoLis,;
PremiSimulati:
LOAD Distinct COD_CLIENTE,
'1' as flg_inserimentoNaz,'1' as flg_inserimentoLis,
COD_FAM_VENDITA,DATA_INIZIO,
NUM(PCT_INSERITO_LIS,'#.##0,00') AS PCT_INSERITO_LIS,
NUM(PCT_INSERITO_NAZ,'#.##0,00') AS PCT_INSERITO_NAZ,
AutoNumber(COD_CLIENTE&COD_FAM_VENDITA) as ChiavePremi,
YEAR(DATA_INIZIO) as ANNO_INIZIO;

SQL
SELECT DISTINCTI
COD_CLIENTE,
COD_FAM_VENDITA,
PCT_INSERITO_LIS,
PCT_INSERITO_NAZ,
DATA_INIZIO,
YEAR(DATA_INIZIO) as ANNO_INIZIO
from DWO.MRG_T_PREMI_SIMULATI_TMP;

```

FIGURA 14. CARICAMENTO DIMENSIONE PREMI SIMULATI

3.3.2 Fact table

Una volta inserite le varie dimensioni è stato necessario riportare anche la *fact table*. All'interno di essa sono definiti il fatturato totale, i premi totali nazioni e listing ottenuti come la somma dei vari premi che li compongono (come riportato nel capitolo di analisi del dominio). I dati sono ovviamente filtrati per anno, più precisamente per mese come si può notare dall'immagine raffigurante la *query*. Il mese è impostato per essere maggiore e uguale a 169, valore numerico che identifica il primo gennaio 2014. Le altre selezioni dei dati consistono nell'escludere i clienti di fatturazione identificati dal codice 20 e 1, che corrispondono alla stessa Orogel, nel dettaglio alla cooperativa agricola e alla consortile Orogel. Così facendo si ottiene un relazione tra Orogel e tutti suoi clienti identificati dalla tipologia di mercato confetterie, retail, catering, discount e dalle famiglie di vendita sulle quali basare il processo di analisi *what-if*. Infine il filtro sulla provenienza è utile per risalire alle vendite che hanno fatture effettivamente contabilizzate. Di seguito è possibile analizzare la *query* relativa alla *fact table*.

SQL

```

SELECT
FCTVENDITE.KEYFAMVENDITA, FCTVENDITE.KEYCLIENTEFATT, FCTVENDITE.KEYMESE, FCTVENDITE.KEYMESECOMPETENZA,
FCTVENDITE.KEYPRODOTTO, DIMPRODOTTI.IDPRODOTTO, FCTVENDITE.KEYPUNTOVENDITA, FCTVENDITE.KEYDISTRIBUTORE,
FCTVENDITE.KEYAREABUSINESS, DIMFAMVENDITA.IDFAMVENDITA as IDFAMVENDITA,
PREMIO1+PREMIO2+PREMIO3+PREMIO6+PREMIO7+PREMIO8+PREMIO9+PREMIOA+PREMIOALTRO
+PREMIOB+PREMIOD+PREMIOE+PREMIOQ+PREMIOR+PREMIOS+PREMIOT+PREMIOU+PREMIOZ+PREMIOY+PREMIOK as PREMINAZIONALI,
PREMIO0+PREMIO4+PREMIO5+PREMIOV+PREMIOF+PREMIOG+PREMIOH+PREMIOI+PREMION+PREMIOX AS PREMILISTING,
PREZZOFINALE, PREZZOLORDO, PREZZONETTO, PREZZOSTANDARD,
PROVVIGIONI, QTAGG, QTACARTONI, FCTVENDITE.KEYAZIENDA, dimpuntivendita.idclientestat, FCTVENDITE.KEYAREABUSINESS

FROM FCTVENDITE INNER JOIN
DIMAREEBUSINESS ON (FCTVENDITE.KEYAREABUSINESS=DIMAREEBUSINESS.KEYAREABUSINESS) INNER JOIN
DIMPRODOTTI ON (FCTVENDITE.KEYPRODOTTO = DIMPRODOTTI.KEYPRODOTTO) INNER JOIN
DIMCLIENTIFATT ON (FCTVENDITE.KEYCLIENTEFATT = DIMCLIENTIFATT.KEYCLIENTEFATT) INNER JOIN
DIMPUNTIIVENDITA ON (FCTVENDITE.KEYPUNTOVENDITA = DIMPUNTIIVENDITA.KEYPUNTOVENDITA) INNER JOIN
DIMMESI ON (FCTVENDITE.KEYMESE=DIMMESI.KEYMESE) INNER JOIN
DIMFAMVENDITA ON (FCTVENDITE.keyfamvendita = DIMFAMVENDITA.keyfamvendita)

WHERE (FCTVENDITE.KEYMESE >= 169) AND (FCTVENDITE.KEYCLIENTEFATT != 20) AND (FCTVENDITE.KEYCLIENTEFATT != 1)
AND ( (DIMPRODOTTI.FV_IDMERCATO = 'D')
OR (DIMPRODOTTI.FV_IDMERCATO = 'C')
OR (DIMPRODOTTI.FV_IDMERCATO = 'R')
OR ( (DIMPRODOTTI.IDFAMVENDITA) IN ('FRUT03', 'Z32', 'Z95'))
OR ( (DIMPRODOTTI.FV_MERCATO) = 'CONFETTURE'))
AND (FCTVENDITE.KEYAZIENDA = 1
AND FCTVENDITE.IDPROVENIENZA IN ('A', '4')));

```

FIGURA 15. FACT TABLE UTILIZZATO ALL'INTERNO DI QLIKVIEW

Tramite il tasto di visualizzazione delle tabelle è possibile verificare come le dimensioni importate su QlikView siano in relazione alla *fact table*.

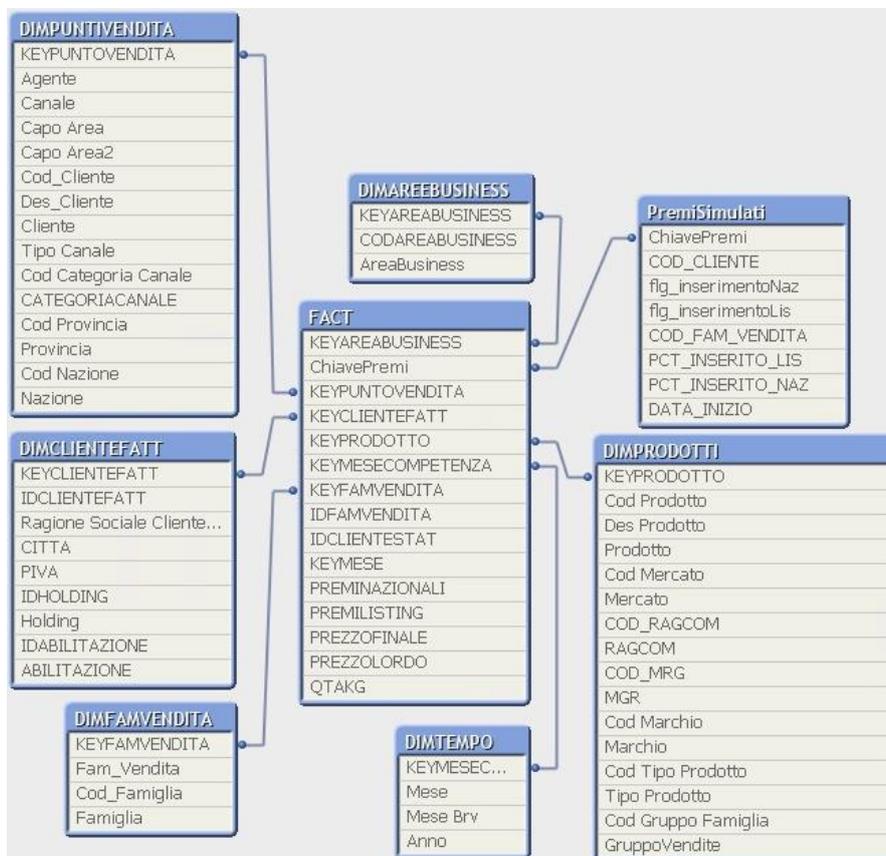


FIGURA 16. VISUALIZZAZIONE RELAZIONE TRA DIMENSIONI E FACT TABLE QLIKVIEW

3.3.3 Creazione tabella simulazione

Al fine di mostrare i dati caricati tramite script, contenente le *query* SQL, è stato necessario servirsi di tabelle grafiche, più precisamente chiamate tabelle lineari. Questo tipo di tabelle consentono di ospitare, all'interno di ogni colonna, particolari espressioni digitate dall'utente in un linguaggio proprietario di QlikView. Attraverso la scrittura di questi *script* è possibile rappresentare nelle celle di una colonna determinati calcoli. All'interno del progetto di tesi è stata utilizzato questo tipo di tabella per la realizzazione dello strumento di analisi *what-if*, dove una prima parte è adibita alla rappresentazione dei dati inerenti all'anno precedente, mentre il resto comprende la visualizzazione delle informazioni dell'anno corrente. Nello specifico la tabella presenta nelle prime due colonne le selezioni effettuate dall'utente, riguardanti il cliente e la relativa famiglia di vendita. Questi dati si riferiscono ai campi delle dimensioni dei punti vendita e delle famiglia di vendita. Adiacenti a questi sono presenti le informazioni memorizzate nel database relative al 2014, filtrate per essere conformi all'anno di analisi attraverso il comando $sum(\{ \$ < Anno = \{ \$ (=Max(Anno)-1) \} > \}$ "fattore da calcolare"). I dati come le percentuali dei premi reali nazionali e listing invece sono ottenuti dividendo il valore del premio per il fatturato, sempre sulla base di informazioni pertinenti al 2014. In questo modo è possibile ottenere per ogni cliente la relazione sul fatturato e sul premio relativa a una famiglia di vendita, come mostrato di seguito.

Fatturato finale 2014	Premi Nazionali 2014	% Premi Nazionali 2014	Premi Listing 2014	% Premi Listing 2014
€ 338.998,59	€ 94.786,63		€ 4.805,87	
€ 3.674,45	€ 1.326,74	36,11%	€ 67,95	1,85%
€ 180.685,45	€ 65.240,26	36,11%	€ 3.341,43	1,85%
€ 63.293,65	€ 22.853,49	36,11%	€ 1.170,50	1,85%
€ 8.982,98	€ 3.243,49	36,11%	€ 166,12	1,85%
€ 3.237,12	€ 1.168,83	36,11%	€ 59,86	1,85%
€ 79.124,95	€ 953,83	1,21%	€ 0,00	0,00%

FIGURA 17. RAFFIGURAZIONE DATI 2014, OGNI LINEA CORRISPONDE AD UNA FAMIGLIA

Per i dati inerenti al 2015 è stata adottata la medesima filosofia filtrano i dati sulla base dell'anno corrente, evitando quindi di decrementare il tempo. Quello che differenzia le due sezione della tabella è la gestione del data entry. Nella porzione dei valori del 2015 sono presenti quattro colonne dove è consentito l'inserimento dei dati, due di esse ospitano le percentuali inserite dall'utente per

la simulazione, le altre permettono all'utente di variare la simulazione in base ai premi reali o ai premi inseriti. Di seguito è mostrata la sezione relativa all'anno in corso per i soli premi nazionali dato che il procedimento è il medesimo anche per i premi listing.

Fatturato finale 2015	%Premi nazionali reali 2015	Premi nazionali reali 2015	pct_Premi_Nazionali_Simulati	Premi Nazionali Simulati 2015	Modificato_Naz	Applica_PremioNaz	%Premi Nazionali 2015	Premi Nazionali 2015
€ 1.347,36	21,11%	€ 284,43	2,29%	€ 39,75			2,67%	€ 39,75
€ 787,84	21,11%	€ 166,31	3,00%	€ 23,64	✓	1	3,00%	€ 23,64
€ 328,84	21,11%	€ 69,44	3,00%	€ 9,87	✓	1	3,00%	€ 9,87
€ 75,84	21,11%	€ 16,01	6,00%	€ 4,55	✓	1	6,00%	€ 4,55
€ 119,28	21,11%	€ 25,18	1,00%	€ 1,19	✓	1	1,00%	€ 1,19
€ 14,70	21,11%	€ 3,10	2,00%	€ 0,29	✓	1	2,00%	€ 0,29
€ 20,76	21,11%	€ 4,38	1,00%	€ 0,21	✓	1	1,00%	€ 0,21

FIGURA 18. SEZIONE DATI 2015. IN ROSSO DATA ENTRY UTENTE, IN GIALLO I RISULTATI DELLA SIMULAZIONE

La colonna Premi Nazionali Simulati 2015 rappresentata è ottenuta moltiplicando il valore inserito con il corrispettivo fatturato. *Applica_PremioNaz* consente all'utente di scegliere se utilizzare i dati digitati o quelli reali, tramite l'inserimento del valore 1 le successive colonne aggiornano le informazioni utilizzando nei calcoli i premi simulati, altrimenti tramite il valore 0 vengono usati premi reali. Inizialmente questa funzione doveva essere gestita attraverso i caratteri "S" e "N" ma a causa di un'incompatibilità, che non permette di modificare i caratteri in tabella, la soluzione finale è stata forzata alla scelta appena descritta. La colonna denominata *Modificato_Nazionale* dà un'informazione grafica all'utente sul tipo di premio che si sta utilizzando, nel caso si utilizzi un valore inserito sarà presente una spunta verde altrimenti, in caso contrario, una croce rossa. Infine nelle colonne *%Premi Nazionali 2015* e *Premi Nazionali 2015* sono riportati i dati aggiornati sulla base della simulazione effettuata. Lo stesso procedimento è equivalente per i premi listing. Infine le ultime due colonne della tabella riportano i totali delle percentuali e dei valori monetari dei premi nazionali e listing ottenuti dalla simulazione. A seconda di come la simulazione è stata affrontata, i totali possono riferirsi non solo ai premi digitati dall'utente ma anche ad una soluzione ibrida, dove il valore finale è calcolato attraverso le somme di premi simulati e reali arbitrati dal campo applica premio. Se in una riga tale valore risulterà zero, allora sarà utilizzato il corrispettivo premio reale per il calcolo.

%Premi 2015	Premi 2015
12,00%	€ 1.131,87
6,00%	€ 70,82
1,00%	€ 499,67
2,00%	€ 323,63
0,00%	€ 0,00
2,00%	€ 71,75
1,00%	€ 166,01

FIGURA 19. DATI TOTALI DELLE SIMULAZIONI

3.3.4 Salvataggio simulazione su data base

Il passo successivo all'inserimento in tabella consiste nel salvataggio dei dati digitati dall'utente. Il processo è stato reso possibile attraverso l'implementazione di diverse funzioni macro scritte in linguaggio Visual Basic script. Le fase di salvataggio è articolata in due step, il primo step consiste nel creare il file XML su cui memorizzare i dati presenti in tabella, il secondo invece recupera i dati dal file XML e li scrive sul database. Per la creazione del file XML sono stati prelevati dalla tabella il valore dei campi relativi al codice cliente, alla codice famiglia di vendita, alle percentuali inserite e ai *flag*, escludendo quindi tutti gli altri dati. La scrittura delle informazioni su file inizia dalla seconda riga della tabella, dato che nella prima sono presenti i totali dei valori e quindi non di interesse per la funzione. Una volta che tutte le righe sono state scorse il file creato presenterà la medesima struttura, dove ogni *tag* row corrisponde ad una linea della tabella.

```
<data table='CH39'>
  <row id='2'>
    <Cod_Cliente>306J423</Cod_Cliente>
    <Cod_Famiglia>R1A</Cod_Famiglia>
    <pct_Premi_Nazionali_Simulati>0.11</pct_Premi_Nazionali_Simulati>
    <Applica_PremioNaz>S</Applica_PremioNaz>
    <pct_Premi_Listing_Simulati>0.03</pct_Premi_Listing_Simulati>
    <Applica_PremioLis>S</Applica_PremioLis>
  </row>
```

FIGURA 20. ESEMPIO DI RIGA SCRITTA IN NOTAZIONE XML

Com'è possibile notare i premi sono memorizzati sul database in forma numerica, suddividendo il valore digitato dall'utente per cento, e i *flag* codificati in modo che nel DB i valori corrispondenti ad 1 e 0 siano memorizzati come caratteri "S" e "N". Una volta compiuta questa fase, la seconda funzione preleva il file XML e salva in delle apposite variabili il valore contenuto nei *tag*. A questo punto per poter scrivere su database è necessario, sempre via codice, impostare la connessione specificando il giusto provider, che nel caso dell'applicativo sviluppato è di tipo **OLEDB**. Oltre a stabilire la connessione occorre riportare il nome del database di destinazione e, nel caso siano necessari, anche il nome utente e la password di accesso al DB. Per eseguire il salvataggio è necessario impostare una *query* come mostrato nell'immagine sottostante.

```

sSQL= "UPDATE MRG_T_PREMI_SIMULATI_TMP SET PCT_INSERITO_NAZ="&inserisci_premnaz&
      ,FLG_APPLICA_NAZ="&premi_nazSN&", PCT_INSERITO_LIS= "&inserisci_premlis&"
      ,FLG_APPLICA_LIS = "&premi_lisSN&"
      WHERE COD_CLIENTE="&cli&" AND COD_FAM_VENDITA="&fam&"
set rsList = tempConn.Execute(sSQL)

```

FIGURA 21. QUERY PER AGGIORNAMENTO DATI SU DB

L'aggiornamento avviene tramite un'operazione di update. Un'altra possibile soluzione valutata nelle prime versioni dell'applicativo, consisteva nell'utilizzare il comando *insert* in accoppiata con il controllo *if exist* per inserire nella tabella dei premi simulati, presente sul DB, le famiglie con i rispettivi premi non presenti nella dimensione. Tale soluzione però è risultata inutile in quanto l'inserimento è stato poi eseguito della procedura Pentaho. Una volta stabilita la *query*, tramite il comando *Excute* è possibile lanciare, sulla connessione precedentemente impostata, l'operazione di aggiornamento.

All'interno dell'applicazione sono due i pulsanti che includono le macro adibite al salvataggio. Uno ha lo scopo di eseguire il salvataggio della tabella attualmente visualizzata, ovvero della tabella contenente i dati dei clienti che l'utente decide di analizzare. L'altro tasto invece salva tutte le modifiche effettuate, scorrendo tutti i clienti importati nell'applicativo filtrati per l'anno selezionato. Essendo questa soluzione la più dispendiosa in termini di risorse, il salvataggio e lo scorrimento di tutti i dati può richiedere qualche minuto, un messaggio a video informerà l'utente di attendere un po' di tempo fino al completamento della procedura.

3.3.5 Creazione grafici

Oltre alla tabella per gli inserimenti e le analisi dei premi, sono presenti anche svariati grafici che mettono a confronto i vari aspetti di interesse nei due anni di analisi. L'uso di questi schemi serve a dare all'utente un'informazione diretta e precisa sull'andamento della simulazione. Tra i tipi di grafici utilizzati vi sono, oltre i comunissimi schemi a barre, i cruscotti che offrono una rappresentazione simile ad un tachimetro. Per entrambe le tipologie è possibile utilizzare delle espressioni che permettono la visualizzazione dei dati.



FIGURA 22. TIPI DI GRAFICO A CONFRONTO

La differenza nelle due rappresentazioni è caratterizzata dal fatto che mentre i grafici a barre possono essere utilizzati per qualsiasi raffigurazione, l'uso dei tachimetri è di solito limitato alla presentazione di dati in forma percentuale. La lancetta principale ha il compito di indicare i valori dell'anno in corso mentre la tacchetta è relativa all'anno precedente. Ogni barra e ogni lancetta è determinata da espressioni diverse adibite alle rispettive rappresentazioni. Altri schemi presenti nell'applicativo mettono a confronto gli inserimenti effettuati, raggruppandoli per famiglie, con i premi delle famiglie dei prodotti del 2014, o relazionandoli per categoria di premio. Ad esempio sono presenti due schemi che hanno il compito di raffigurare, in valore monetario, le percentuali digitate dall'utente con le controparti reali dei due anni, ottenendo quindi come risultato finale un grafico a tre barre che si aggiorna dinamicamente con il proseguimento della soluzione.

3.3.6 Scelte scartate

Versioni iniziali dell'applicativo consistevano nella presenza di due tabelle, una tabella per la gestione del *data entry* effettuato dall'utente e una tabella lineare, nella quale includere i dati inseriti attraverso la pressione di un tasto. La presenza di una tabella per la gestione dell'inserimento era obbligatoria a causa dell'impossibilità di modificare i valori all'interno tabelle di tipo lineare, condizione che rappresenta una forte limitazione del software QlikView. Questa problematica è stata poi superata attraverso una funzione di inserimento non propriamente adatta allo scopo, ma che non ha comportato effetti indesiderati. L'inconveniente che è nato dall'utilizzo di due tabelle

riguardava soprattutto la gestione dei pulsanti, dove l'impossibilità di assegnare, come accade nei comuni linguaggi di programmazione, un evento *click* al bottone ha fatto propendere l'abbandono di questa scelta. I tasti che inizialmente comparivano nel progetto erano due e comprendevano l'inserimento dei dati in tabella e il ribaltamento dei premi del 2014 nelle colonne del 2015 (operazione poi gestita attraverso la procedura Pentaho). All'interno del software QlikView ogni pulsante deve essere gestito attraverso una variabile che si incrementa/decrementa ad ogni pressione e, in base al valore della stessa, la funzione da eseguire rimane attiva fino alla successiva interazione con il tasto. E' facile intuire come la gestione di più tasti che richiedano di essere premuti per attivare e disattivare le funzioni possa creare facilmente confusione. Per lo stesso motivo sono state eliminate anche gli inserimenti di uno stesso valore per colonna (funzione simile al trascinarsi delle celle su Excel che copia lo stesso valore per tutta la colonna/riga), che avevano lo scopo di prelevare il valore inserito dall'utente in un particolare oggetto, denominato casella di input, e di copiarlo per tutte le celle. Il tutto era gestito da quattro pulsanti, due per gli inserimenti delle percentuali dei premi e altri due per impostare i valori di applicazione della simulazione, aumentando notevolmente la complessità di utilizzo del programma per l'utenza.

CAPITOLO 4

Interfaccia grafica dell'applicativo

4.1 Introduzione

Nel seguente capitolo è presentata l'interfaccia grafica dell'applicativo sviluppato, con lo scopo di dare una rappresentazione visiva di quanto progettato e offrire una guida per l'utilizzo del progetto di tesi. La descrizione è organizzata per fogli di lavoro, dove per ognuno sono elencati gli oggetti che lo compongono con relative immagini.

4.2 Interfaccia grafica.

L'applicativo è organizzato su tre fogli di lavoro:

- Cruscotto premi;
- Simulazioni premi;
- Mappa selezioni.

4.2.1 Gestione premi.

Si sottolinea che l'applicativo è stato concepito per essere utilizzato dal corrente anno per gli anni successivi. Sulla parte sinistra del foglio di lavoro è indicato l'anno, una finestra di ricerca e un riquadro con le selezioni correnti che l'utilizzatore impiega per la gestione del programma. Nella parte inferiore del citato riquadro sono elencate le caselle di elenco che racchiudono le opzioni scelte dall'utente. Le stesse si suddividono in:

- Cliente: sono riportati i nominativi dei punto vendita che utilizza la direzione commerciale di Orogel;
- Canale: è indicato il canale di distribuzione dei prodotti;
- Gruppo vendite: al suo interno sono presenti i macro raggruppamenti dei prodotti scelti da Orogel;

- Holding: Indicano le società che detengono le quote di altre società;
- Famiglia: sono indicate le famiglie dei prodotti.

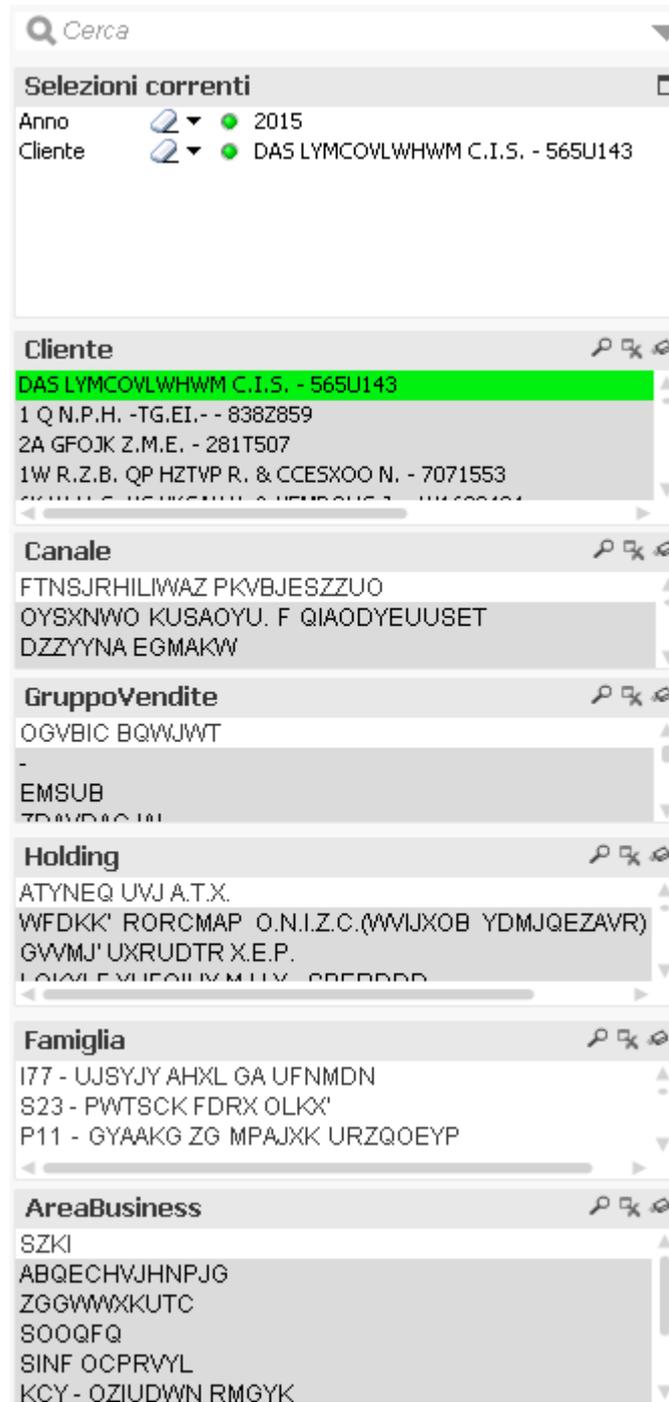


FIGURA 23.CASELLE DI SELEZIONE, DATI CRIPTATI PER PRIVACY

Nella parte centrale del foglio di lavoro sono rappresentati, oltre a un pulsante adibito all'eliminazione delle selezioni, quattro tachimetri, premi nazionali, premi listing, premi nazionali

simulati, premi listing simulati che riportano una scala in percentuale che va da una zona rossa (negativa) ad una zona verde(positiva) passando da una zona gialla (media):

- Il tachimetro premi nazionali e premi listing confronta le percentuali dei premi assegnati dell'anno in corso con l'anno precedente;

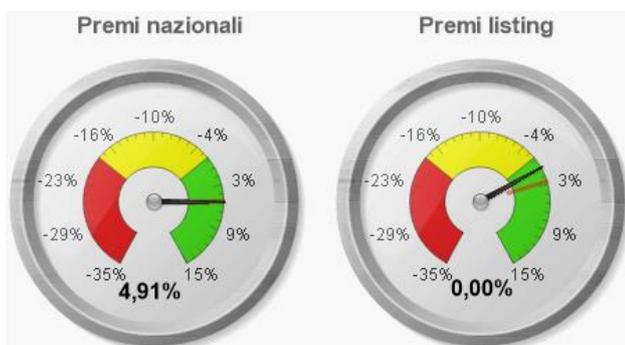


FIGURA 24. TACHIMETRI PERCENTUALI PREMI REALI

- I tachimetri premi nazionali simulati e premi listing simulati indicano le medie dei premi inseriti dall'utente per l'anno in corso.

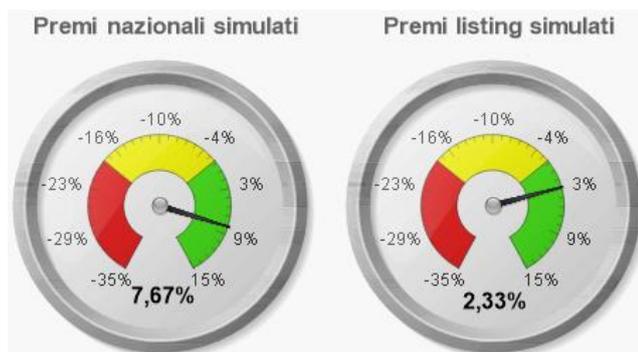


FIGURA 25. TACHIMETRI PREMI SIMULATI

Nella parte inferiore del foglio di lavoro sono riportati quattro grafici a barre che analizzano il fatturato del cliente e le due tipologie di premio nei due anni di analisi. Un ultimo grafico mette a confronto il totale dei premi assegnati nel 2014 con il totale dei premi dell'anno 2015. Il totale dei premi dell'anno corrente è ottenuto dalla somma dei premi reali e simulati utilizzati nella simulazione.

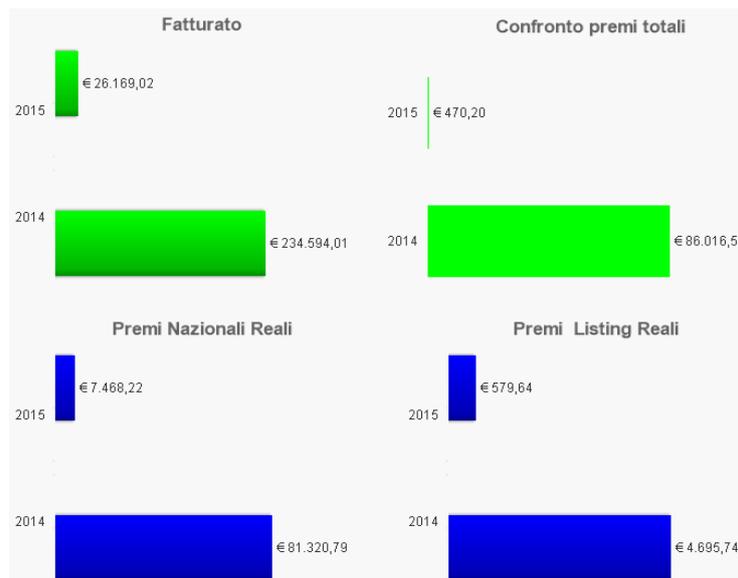


FIGURA 26. CONFRONTI GRAFICI A BARRE

4.2.2 Simulazione premi

La parte sinistra del foglio riporta i medesimi oggetti del foglio precedente.

Nella parte superiore sono riportati 5 tasti, rispettivamente: elimina selezioni, salva visualizzazione corrente, salva tutto, carica e elimina input.

- Elimina selezioni rimuove le selezioni nelle caselle d'elenco scelte dall'utente;
- Salva visualizzazione corrente memorizza sul database i dati riportati in tabella che si analizzano;
- Salva tutto: memorizza tutti i dati modificati nella tabella di lavoro;
- Carica esegue il caricamento dei dati dal database in tabella;
- Elimina input rimuove i dati inseriti dall'utente in tabelle



FIGURA 27. PULSANTI DELL'APPLICATIVO

Sotto i pulsanti è presente un oggetto che racchiude tre oggetti quali:

- Analisi premi: raffigura la tabella nella quale l'utente inserisce il valore del dato del premio per la simulazione:

- Le prime due colonne indicano il codice cliente e il codice della famiglia dei prodotti;
- Le colonne contraddistinte dall'anno 2014 riflettono i dati relativi al fatturato, ai premi nazionali e ai premi listing dell'anno;

Analisi Premi				Andamento premi nazionali reali		
Cod_Cliente	Cod_Famiglia	Fatturato finale 2014	Premi Nazionali 2014	% Premi Nazionali 2014	Premi Listing 2014	% Premi Listing 2014
		€ 338.998,59	€ 94.786,63	27,96%	€ 4.805,87	1,42%
306J423	R1A	€ 3.674,45	€ 1.326,74	36,11%	€ 67,95	1,85%
306J423	R03	€ 180.685,45	€ 65.240,26	36,11%	€ 3.341,43	1,85%
306J423	R13	€ 63.293,65	€ 22.853,49	36,11%	€ 1.170,50	1,85%
306J423	R18	€ 8.982,98	€ 3.243,49	36,11%	€ 166,12	1,85%
306J423	R19	€ 3.237,12	€ 1.168,83	36,11%	€ 59,86	1,85%
306J423	R59	€ 79.124,95	€ 953,83	1,21%	€ 0,00	0,00%

FIGURA 28. SEZIONE DELLA COLONNA RELATIVA AL 2014

- Le colonne relative all'anno 2015 inerenti al fatturato e ai premi nazionali e listing riportano i dati reali presenti nel database mentre nelle colonne evidenziate con il colore blu, *pct_Premi_Nazionali_Simulati* e *pct_Premi_Listing_Simulati*, l'utente effettua l'inserimento del dato del premio. La colonna *Premi Nazionali Simulati 2015* e *Premi Listing Simulati 2015* riportano il valore in euro del premio assegnato che è il prodotto tra il fatturato del 2015 e la percentuale del premio inserito. Nelle colonne *Applica_PremioNaz* e *Applica_PremioLis* l'utente sceglie se utilizzare, per la simulazione, il premio reale o il premio simulato, tale scelta viene notificata nelle colonne *Modificato_Naz* e *Modificato_List*. All'interno delle stesse se si utilizza il premio simulato rappresentato dal valore 1 risulterà una spunta verde altrimenti se si utilizza il premio reale rappresentato dal valore 0 apparirà una croce rossa.
- Nelle colonne *%Premi Nazionali 2015*, *Premi Nazionali 2015*, *%Premi Listing 2015*, *Premi Listing 2015* sono presenti i valori in percentuale ed in euro delle scelte effettuate nel punto precedente.

Cod_Cliente	Cod_Famiglia	Fatturato finale 2015	% Premi nazionali reali 2015	Premi nazionali reali 2015	pct_Premi_Nazionali_Simulati	Premi Nazionali Simulati 2015	Modificato_Naz	Applica_PremioNaz	% Premi Nazionali 2015	Premi Nazionali 2015
		€ 92.022,01	32,58%	€ 29.982,63	4,67%	€ 4.407,71			4,67%	€ 4.407,71
306J423	R1A	€ 1.180,33	39,50%	€ 466,23	11,00%	€ 129,84	✓	1	11,00%	€ 129,84
306J423	R03	€ 49.966,89	39,50%	€ 19.736,92	6,00%	€ 2.998,01	✓	1	6,00%	€ 2.998,01
306J423	R13	€ 15.850,36	39,50%	€ 6.260,89	5,00%	€ 792,52	✓	1	5,00%	€ 792,52
306J423	R18	€ 4.901,94	39,50%	€ 1.936,26	1,00%	€ 49,02	✓	1	1,00%	€ 49,02
306J423	R19	€ 3.587,25	39,50%	€ 1.418,96	3,00%	€ 107,62	✓	1	3,00%	€ 107,62
306J423	R59	€ 16.535,23	1,00%	€ 165,35	2,00%	€ 330,70	✓	1	2,00%	€ 330,70

FIGURA 29. SEZIONE PREMI NAZIONALI. IN BLU DATA ENTRY UTENTE, IN CHIARO RISULTATO DELLE SCELTE

Analisi Premi			Andamento premi nazionali reali				Andamento p		
Cod_Cliente	Cod_Famiglia	% Premi Listing Reali 2015	Premi Listing Reali 2015	pct_Premi_Listing_Simulati	Premi Listing Simulati 2015	Modificato_List	Applica_PremioLis	% Premi Listing 2015	Premi Listing 2015
		1,58%	€ 1.434,25	2,33%	€ 2.467,26			2,47%	€ 2.755,43
306J423	R1A	1,90%	€ 22,43	3,00%	€ 35,41	✗	0	1,90%	€ 22,43
306J423	R03	1,90%	€ 949,37	4,00%	€ 1.998,68	✓	1	4,00%	€ 1.998,68
306J423	R13	1,90%	€ 301,16	0,00%	€ 0,00	✗	0	1,90%	€ 301,16
306J423	R18	1,90%	€ 93,14	4,00%	€ 196,08	✓	1	4,00%	€ 196,08
306J423	R19	1,90%	€ 68,16	2,00%	€ 71,75	✓	1	2,00%	€ 71,75
306J423	R59	0,00%	€ 0,00	1,00%	€ 165,35	✓	1	1,00%	€ 165,35

FIGURA 30. SEZIONE PREMI LISTING. IN BLU DATA ENTRY UTENTE, IN CHIARO RISULTATO DELLE SCELTE

- Nelle colonne %Premio 2015 e Premio 2015 risultano i totali delle scelte effettuate.

Analisi Premi			
Cod_Cliente	Cod_Famiglia	%Premi 2015	Premi 2015
		7,00%	€ 6.874,97
306J423	R1A	14,00%	€ 165,25
306J423	R03	10,00%	€ 4.996,69
306J423	R13	5,00%	€ 792,52
306J423	R18	5,00%	€ 245,10
306J423	R19	5,00%	€ 179,36
306J423	R59	3,00%	€ 496,06

FIGURA 31. SEZIONE FINALE TABELLA, IN VERDE I TOTALI DELLA SIMULAZIONE

- Andamento premi nazionali e andamento premi reali: raffigurano un grafico che mostra l'andamento dei premi in relazione al fatturato per i due di anni di analisi. Di seguito è mostrato un esempio di grafico per i soli premi nazionali (lo schema è analogo per i premi listing).



FIGURA 32. GRAFICO ANDAMENTO PREMI NAZIONALI REALI PER I DUE ANNI A CONFRONTO

Nella parte inferiore sono rappresentati due raccoglitori, uno per i premi nazionali e uno per i premi listing, che raffigurano il confronto in euro tra i premi reali e i premi simulati per le famiglie di vendita e la comparazione tra i premi reali del 2014 e del 2015 con i premi simulati. Di seguito lo schema dei confronti per famiglia di vendita.

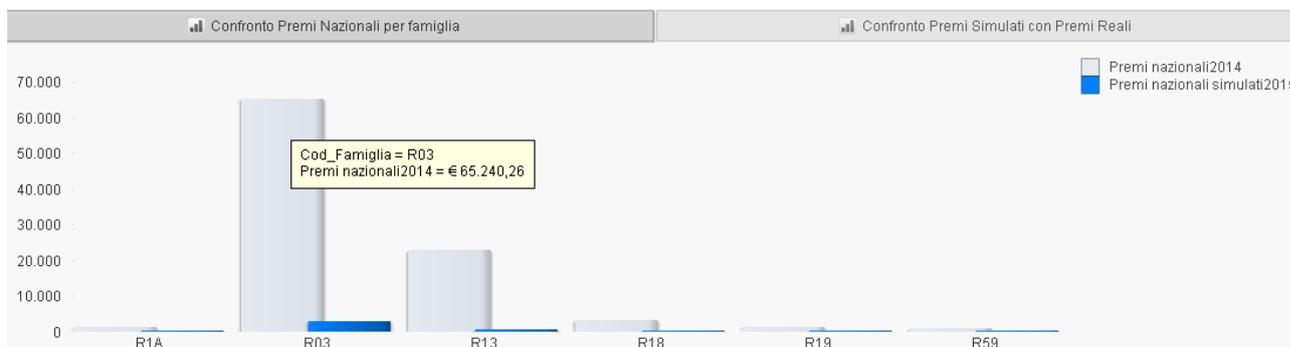


FIGURA 33. CONFRONTO PREMI NAZIONALI E PREMI USATI PER LA SIMULAZIONE

Nell'immagine sottostante invece è rappresentato il grafico che mette a confronto i premi nazionali reale del 2014 e 2015 con il totale dei premi inseriti dall'utente.

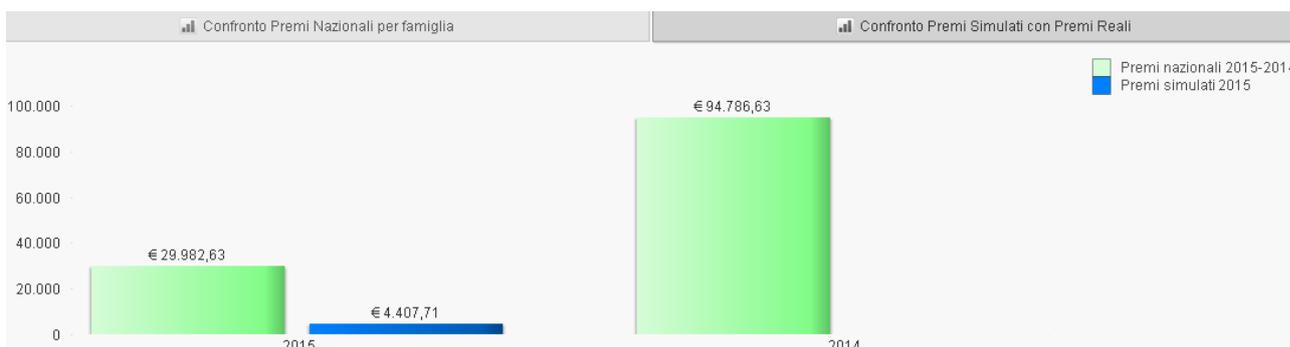


FIGURA 34. CONFRONTO PREMI REALI CON SIMULATI

Non sono presenti gli schemi dei premi listing in quanto simili ai grafici precedentemente mostrati dei premi nazionali.

4.2.3 Mappe selezioni

In questo foglio di lavoro sono indicati i dati, divisi per colore verde e blu, che l'utente può scegliere all'interno dell'applicativo. Il colore verde indica le selezioni inerenti la figura del cliente

mentre in blu sono raffigurate le selezioni relative alle famiglia di vendita dei prodotti. Nell'immagine sottostante è presente la sezione che riguarda il cliente

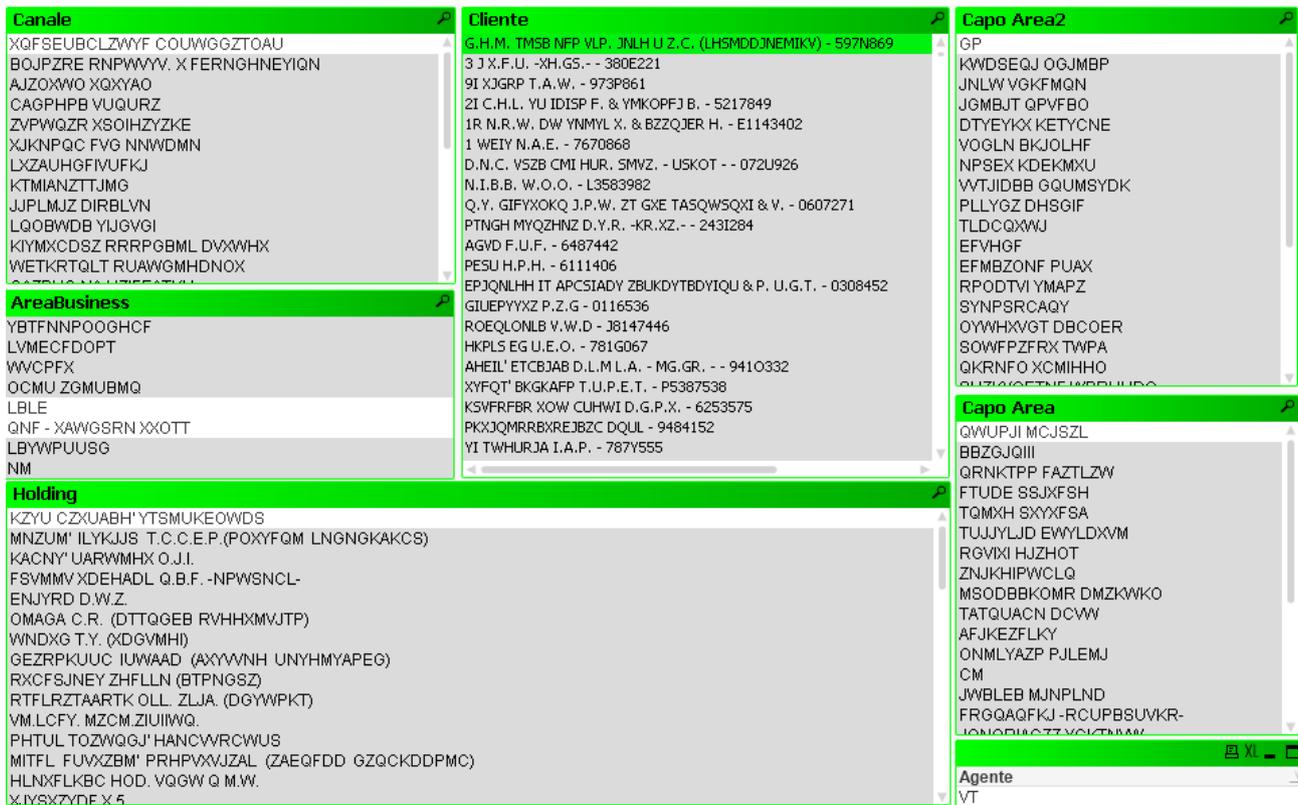


FIGURA 35. CASELLE DI SELEZIONE RELATIVE ALLE SELEZIONI DEL CLIENTE

Di seguito le selezioni inerenti alla famiglia di vendita.

Famiglia	RAGCOM	Marchio
I7H - PNPACT WCRXDW "IFOKMWBSQ"	LIOOX	GEZU QSELSC
A51 - ETUSIC EZRT 50 RMPUJA	DEYJP KPEKH.EZNO"BM"	EQ SWYYCB EIFVOCIM
I92 - FQMPXA OO WCRLCZ EFHJSOMB	JZKRXIZH PZL "JO"	GTQOCI MKXGQY
M88 - PRIPIR KJJKHY'	VITSRHNE DHIRG "W"	P.&.K.
J28 - LDXFUM WHJAF	OXHEJLGN CGZWEW "QG"	HVZ EOHUEY
R67 - IZYKBF GZMTCTA XINO	BLPWJRN GMZHJ	GLX YMGKMGWTDJIX
- - KG	JVZKUIKP TALZFBZDO	KXRC YQIET
JV - WMTOCJWNSRCD	RCCZJBKZ MTCHDPOM	HJZFFTPC
DA642 - AMBPKDWUWMI YZXKGTDTTII	VRVPJEZD JKJRWDSNWH	JPWWTB
SQ223 - BDAAKEFGYP MVUVNK/YQBSERF	ZQPQSF E CQOKHWLV	GEZMQQRER
HW566 - ILLBJXQVA DPCFU	INSQVHCISU SUMF"AL"	BYACUQ
NZ414 - ENINANORU YKSAARN	BJL DLUKAVPXAH	APDDTPJWB
OV749 - YVTGXQIN CYLCYDE	HTLRI."HEY"	KV-XRDBO
DT698 - ORKTRPIKN V5 KBDWDM	NXXKL BSJEC.TBKEB	NUBMD
TE706 - PDWVQ KVVXTGMQN XCKCW	MDPEELD	ORKAXMVOW
TI637 - QCUHXEFOEN VJDYB	AYRYFP UJO CELUUK	TVVDXKKF NOQDPEM
UL869 - FNQW5AK HJLKNWHRCZYOC	KDC.STKTOOMOX "TA"	TTELEZIPJSFV AZREX
CH634 - AFQDKJH CVBNMG	FREZLBDT	GRXL&FKNM YZJWJG
ZL436 - EIO. OJMNQP GU OJVJFFER	LKQQ	BPUJNG
GX531 - THV.UFP.CNFXH.BAI.CYLYWUN	KCAJWFMMXR'	RKWS UMLDBK
TY - ITAONL	UENWUQM SURL"KS"	PHFJQ WTIGMH
BP6 - PYGHPFY	XZSQR	EAPHD REWE.M
PH6 - GTBQUJP	(WXAY)WCEZYQ YYTLNL	LNYIG ZCLN.Q
QV8 - KSSMFOGBV	(JAJR)BSAGBFQ ELULOQ	ANHCH LCYX.LNG
AF1 - JETXGQON	(HCVX)CDZG. XKMWWW	JFAPR LJWU.LRY
QA9 - DDTWKOHRF QNCEWYSHQREIZ DZCSD	(JTOE)ZPDYD DCSXI	HWANP HPCY.JH
EO8 - JSSWZZ TQA YUFXOXFWOX	(PBNZ)PHOLL UWJSPPA	WHCMZSYO VDVHZE
RY4 - FMYDOLLQZ YMLZYD	(IUB)RZFLZGXF	SGZQ
GruppoVendite	(PUMY)CBBQFGNA	
-	(PZDC)WCAYR.LZUSTZ	
BJYSC	(LGQY)PFBEEN SVTYS	Mercato
YLPYPPFNOS	(ZZLQ)CNMMUYNLRX	RGOJX
AXWB ENBRIRH	(BQWN)JIA JCA	RDQLLQJF
HCKLJG PGRKPRJ	(CZFH)BVBGNCA	NLXXDGGQONZ
PG	(EGBP)DRIL HEAS VFP.	IAPQRXRT
SBWAOH MMXIH	(KQXX)SGNDSC	ZT
NNMKZN PEEFIE Q.YSVMP	(GIEN)KCNRMX WIYVK	XPPHUG
ZAINZ FDVGJK	(XDBA)V.VAZ KCL	

FIGURA 36. CASELLE DI SELEZIONE RELATIVE ALLA FAMIGLIA DI VENDITA

4.3 Rappresentazione dei fogli di lavoro

Di seguito sono mostrate le immagine complessive dei fogli di lavoro che compongono il progetto.

Foglio di lavoro cruscotto premi:



FIGURA 37. PRIMO FOGLIO DI LAVORO

Conclusioni e sviluppi futuri

Lo svolgimento della tesi ha permesso di interfacciarsi con nuovi strumenti per lo sviluppo di applicazioni di *business intelligence*. L'applicazione di analisi *what-if* sviluppata ha confermato la sua grande utilità nel supporto della gestione dei premi ai clienti. Tramite esso gli utenti possono monitorare concretamente il guadagno dell'azienda e la relativa variazione del fatturato in relazione con i premi assegnati. Le maggiori criticità riscontrate sono state relative alla gestione del *data entry* in quanto il programma QlikView non è propriamente pensato per questo tipo di funzioni. All'interno del reparto *BU* di *Onit* non era mai stato creato un prodotto che consentisse all'utente la possibilità di inserire i dati direttamente nell'applicativo. Le varie versioni del progetto sono state riviste e modificate più volte nel corso della tesi, soprattutto per andare incontro alle specifiche di Orogel. La direzione commerciale aveva espressamente richiesto di realizzare un prodotto contenente una sola tabella per la gestione della simulazione dei premi, al fine di ottenere in questo modo un utilizzo più immediato, più facile e un'interfaccia grafica maggiormente intuitiva. Gli obiettivi prefissati sono stati ampiamente raggiunti e i vari test effettuati hanno confermato l'affidabilità del prodotto sviluppato. A progetto concluso è stato svolto un attento lavoro di analisi per misurare la validità dell'applicativo realizzato. La fase di salvataggio dei premi, che costituisce il cuore del programma, adempie perfettamente al suo lavoro non mostrando nessun segno d'errore. La possibilità di avviare una simulazione dei premi in completa autonomia rappresenta un importante punto di svolta che consente all'utente di operare in modo autonomo e con un notevole risparmio di tempo per la gestione della procedura. Oltre ai vari vantaggi operativi, la progettazione dell'applicativo ha aperto nuove porte per lo ulteriori sviluppi che riguardano applicazioni di *data entry*.

Per quanto riguarda gli sviluppi futuri da apportare al programma, sono già prese in considerazione eventuali soluzioni per consentirne l'utilizzo attraverso un web service. Tramite questa implementazione, al programma sarà permessa la comunicazione con un web service mediante l'uso della libreria *jQuery*. Il web service quindi preleverà i dati da Qlikview e li trasformerà in *query SQL* da passare al data base di destinazione. In questo modo si crea un ciclo dove i dati scritti possono essere riletti e usati dall'applicazione in QlikView, con il vantaggio che una soluzione di questo tipo consentirebbe agli utenti di collegarsi all'applicazione anche senza essere connessi alla rete locale. Altri miglioramenti consistono nel valutare il cambiamento del fatturato dell'anno

precedente in relazione ai premi assegnato nell'anno in corso e nel trovare una strada più vantaggiosa nell'ottimizzazione del salvataggio totale delle modifiche effettuate. Tale funzione è stata fortemente limitata dalla poca conoscenza del linguaggio *VBscript* in relazione agli oggetti presenti all'interno del programma. Attraverso questa implementazione si otterrebbe quindi una procedura di archiviazione de dati più veloce.

Ringraziamenti

Un ringraziamento speciale l'azienda Onit Group ed al professore Matto Golfarelli per avermi dato la possibilità di intraprendere questo percorso di tesi e a tutta la mia famiglia e gli amici per avermi sostenuto in tutti questi anni.

Grazie.

Bibliografia

- [1] M. Golfarelli, A. Proli, S. Rizzi, Un sistema di analisi what-if per la redditività delle filiali, 2015
- [2] Monte Carlo Method,
<http://www.wolframalpha.com/input/?i=monte+carlo+method&lk=4&num=1>,2015
- [3] Stefano Rizzi, *What-If Analysis*, <http://www-db.deis.unibo.it/~srizzi/PDF/eds-WIA.pdf>, 2015
- [5] Lisa Jo Rudy, *What-If Tools for Analyzing Possible Management Outcomes*,
<http://business.tutsplus.com/articles/what-if-tools-for-analyzing-possible-management-outcomes--cms-21444>, 2014
- [4] Investopedia, *Complete Guide To Corporate Finance*,
<http://www.investopedia.com/walkthrough/corporate-finance/4/project-analysis/scenario.aspx>
- [6] Stan Brewer, 'What-if analysis' lets you assess changes before they happen, http://www.mrc-productivity.com/about/press_WhatIfAnalysis.html,2015
- [7] Dwight deVera, *Quick tips for Using Bi to Improve Budgeting&Planning*
,<http://www.arcplan.com/en/blog/2011/10/quick-tips-for-using-bi-to-improve-budgeting-planning/>
- [8] Lou J. Diberardinis, *Handbook of Occupational Safety and Health Second Edition*, Wiley-Interscience, 1999
- [9] Pentaho,
<http://wiki.pentaho.com/display/COM/Community+Wiki+Home>,2015.2015
- [10] Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, and Riccardo Torlone.
Basi di dati, modelli e linguaggi di interrogazione. McGraw-Hill, second edition, 2006
- [11] Guido Destri, *Sistemi informativi. Il pilastro digitale di servizi e organizzazioni*. Franco Angeli,2013.