Università di Pisa

Dipartimento di Informatica
Corso di Laurea Magistrale in Informatica per l'Economia e per l'Azienda
(Business Informatics)

Tesi di Laurea Magistrale

PROGETTAZIONE E REALIZZAZIONE DI UN DATA WAREHOUSE PER UNA AZIENDA AGROALIMENTARE

Tutore Accademico:
Prof. Giorgio GHELLI

Tutore Aziendale:
Dott. Fabio MORSIANI

Candidato:
Giuseppe LO CONTE

Anno Accademico 2014-2015

Alla mia famiglia...

Riassunto

In questo lavoro di tesi viene mostrato come, a partire dai requisiti di analisi del cliente, viene progettato e realizzato un sistema di *data warehouse* e di *business intelligence*.
Successivamente ad una introduzione utile alla comprensione dei concetti fondamentali del caso aziendale vengono descritti gli strumenti utilizzati, le attività di progettazione e implementazione del data warehouse, soffermandosi sul processo di elaborazione dei dati in ingresso e sulla rappresentazione grafica delle informazioni offerta agli utenti.

Indice

Capitolo 1

INTRODUZIONE

1.1 Presentazione del problema

In questi anni di sviluppo delle nuove tecnologie, per le aziende moderne si è reso necessario raggiungere gli obiettivi prefissati con risultati economico-finanziari implicando la conoscenza profonda e sistematica dei propri processi di business. I manager aziendali oggi assumono il ruolo attivo nella presa delle decisioni (*knowledge worker*) richiedenti un accesso diretto ed immediato ai dati aziendali utile nel poter aumentare l'efficacia del proprio agire e la competitività delle proprie unità organizzative.

Le attività di pianificazione e controllo sono un prerequisito fondamentale per un'efficace attività di business volta nella maggior parte degli ambienti lavorativi a costituire un elemento indispensabile per mantenere i dati acquisiti. Talvolta le aziende sono costrette a gestire una elevata dimensionalità di dati informativi, infatti le normali basi di dati non offrono soluzioni adeguate per la gestione di questi nel momento del raggiungimento di una elevata capacità. L'enorme accumulo di dati e la pressante richiesta di utilizzarli attivamente per scopi che superino quelli di routine danno vita al fenomeno del data warehousing che è legato all'elaborazione giornaliera di quest'ultimi.

Il calcolatore risulta l'unico supporto adatto al processo decisionale, dato l'aumento esponenziale del volume dei dati operazionali e l'utilizzo massiccio di tecniche di analisi dei dati aziendali. Questa fase ha reso il sistema informativo un elemento strategico per la realizzazione del business, per tali motivi il ruolo dell'informatica è passato da passivo strumento per la registrazione delle

operazioni, a fattore decisivo per l'individuazione di elementi critici nell'organizzazione e nelle potenziali aree di business.

Si evidenzia la reale possibilità di conoscere e valutare su base oggettiva le performance di una funzione o di un processo legato alla comprensione delle informazioni ricavate dai dati, dimostrandosi indispensabile nel raggiungere decisioni strategiche e direzionali volte al conseguimento degli obiettivi prefissati dal management e finalizzati ad incrementare il valore aziendale.

Nel presente lavoro di tesi si descrivono tutte le fasi di progettazione e sviluppo di un sistema di supporto alle decisioni per una media azienda italiana operante nel settore agroalimentare. Si è manifestata l'esigenza di analizzare alcuni dei suoi processi principali al fine di monitorarne ed eventualmente migliorarne le prestazioni che nei processi sono risultanti. La soluzione proposta è un sistema di data warehouse e business intelligence realizzato nell'ambito di un progetto di aggiornamento del sistema di *Enterprise Resource Planning* (ERP), il quale svolge il ruolo di sorgente informativa unica.

In conclusione si è realizzato in fase preliminare un processo di creazione di un modello predittivo in riferimento al *Market Basket Analysis*.

1.2 Rassegna della letteratura

Una parte non trascurabile del lavoro è basata sul contenuto di Albano [1] particolarmente adatto come guida alla realizzazione di progetti di data warehousing in quanto concetti teorici e pratici sull'argomento sono correlati da indicazioni sulla presentazione formale di quanto progettato.

Il modello concettuale utilizzato è quello proposto da Golfarelli e Rizzi [2].

Le procedure di estrazione sono state perfezionate seguendo le indicazioni di Kimball e Caserta [3] [4] e manuale Oracle [5].

Di particolare aiuto per la realizzazione dell'applicazione di business intelligence sono stati i manuali MicroStrategy [6] [7] [8] .

1.3 Contenuto della tesi

Il presente documento ha l'obiettivo di presentare il lavoro svolto per il progetto di data warehousing per un'azienda del settore agroalimentare, introducendo il lettore alla tematica, mostrando successivamente come quanto richiesto sia stato progettato e realizzato, evidenziando tecniche e tecnologie utilizzate e descrivendo l'interazione con l'utente finale.

In particolare il Capitolo 2 descrive il caso reale di una azienda che ha richiesto lo sviluppo di un sistema di supporto decisionale per l'analisi dei dati gestiti da un sistema informatico integrato; sono inoltre presentate le aziende coinvolte ed il business di riferimento.

I Capitoli 3 e 4 descrivono le fasi di progettazione e modellazione concettuale e logica del *data warehouse*. Nella prima fase è descritta la prima parte della progettazione del sistema, in particolare le fasi di analisi dei requisiti e progettazione concettuale iniziale; seguono la fase di progettazione dei *data mart* sulla base dei requisiti raccolti e quella dai dati operazionali, dal confronto tra esse si arriva alla progettazione concettuale finale. Viene infine elaborato il modello logico dei *data mart* e del *data warehouse*.

Il Capitolo 5 è dedicato all'introduzione dei principali strumenti usati per lo sviluppo del sistema. Nel Capitolo 6 si presenta il processo di ETL, Vengono inoltre descritte le modalità di aggiornamento dei dati nel sistema e la gestione degli errori. Il Capitolo 7 presenta lo sviluppo della reportistica. Nel Capitolo 8 si descrive una introduzione sulla disciplina del data mining con le relative tipologie e il processo di creazione di un modello predittivo in riferimento al *Matket Basket Analysis*.

Infine il Capitolo 9 è dedicato alle conclusioni, alle considerazioni sul raggiungimento degli obiettivi e ai possibili sviluppi futuri.

Capitolo 2

CASO DI STUDIO

In questo capitolo, sono presentate le aziende coinvolte nel progetto, il business di riferimento e i processi di business del caso aziendale. Lo scopo è quello di delineare un quadro completo ed esaustivo per comprendere al meglio le esigenze informative manifestate dalla dirigenza durante la raccolta dei requisiti di analisi.

2.1 Presentazione del caso

Il caso aziendale oggetto del lavoro di tesi è relativo alla progettazione e realizzazione di un sistema che supporti le attività decisionali strategiche, direzionali e operative per una azienda del settore agroalimentare.

2.1.1 Deloitte

Deloitte riunisce più di 210.000 professionisti di 47 aziende indipendenti in oltre 150 paesi del mondo. Tramite le sue aziende, Deloitte fornisce servizi di:

- *consulting*
- *audit*
- *tax & legal*
- *enterprise risk*
- *financial advisory*

Pur essendo organizzate e governate localmente, le aziende del network condividono i valori, le metodologie e gli standard qualitativi per offrire ai clienti i propri servizi professionali a livello globale.

Negli ultimi anni Deloitte ha investito per rafforzare la propria presenza nel settore degli *analytics*, diventando partner specializzato delle più importanti aziende tecnologiche presenti sul mercato e ricevendo prestigiosi riconoscimenti per l'eccellenza delle soluzioni implementate.

Il progetto di sviluppo del sistema di supporto decisionale descritto in questo lavoro di tesi è stato realizzato nella sede di Parma della società Deloitte XBS e ha coinvolto un manager e un consultant. Durante le fasi progettuali e implementative, il team è stato affiancato dai specialisti responsabili del sistema informativo sorgente, i quali hanno fornito le indicazioni necessarie per l'estrazione dei dati.

2.1.2 Il business agroalimentare

Il settore agroalimentare comprende l'insieme di attività orientate alla produzione, trasformazione e distribuzione di prodotti alimentari.
I principali settori economici che costituiscono il settore agroalimentare sono:

- *Agricoltura*
- *Industrie fornitrici di mezzi tecnici per l'agricoltura*
- *Industria della trasformazione alimentare*
- *Settore del commercio* (distribuzione alimentare)

Ogni ambito di attività del settore ha una filiera produttiva, che comprende diverse fasi:

- produzione/acquisto delle materie prime
- lavorazione e trasformazione del prodotto
- confezionamento
- conservazione
- distribuzione

Il settore agroalimentare ha una filiera produttiva molto diversificata che va dall'agricoltore che lavora la terra, passando per l'operaio che utilizza macchinari per la lavorazione, la trasformazione e il confezionamento alimentare per arrivare, infine, al distributore finale del prodotto. Il settore agroalimentare prevede il trasporto di grandi quantità di merci per permettere la commercializzazione dei beni prodotti: per questo motivo esso è strettamente connesso al settore del commercio e al settore dei trasporti.

Il mercato agroalimentare mondiale è dominato da alcune grandi multinazionali che coprono elevate quote del mercato mondiale. L'elevata concentrazione sia dell'industria che del commercio alimentare fa sì che i prodotti trasformati e confezionati siano nelle mani di un numero ristretto di gruppi societari. La Tabella 2.1 riporta le prime imprese multinazionali che operano nel settore agroalimentare.

Philip Morris
Nestlé
Coca-Cola
Danone
Heinz

Tabella 2.1: Multinazionali

In Italia, invece, il ramo agroalimentare conferma la sua importanza all'interno dell'economia Italiana. Risulta essere un comparto trainante per il PIL nazionale spinto dal patrimonio ambientale e culturale del suo territorio e dalla creatività delle piccole e medie imprese. L'evoluzione degli stili alimentari ha portato ad alzare continuamente l'asticella della qualità e della sicurezza alimentare, diventata oggi un prerequisito. Per la carne, ad esempio, è sparita la bassa macellazione: tutti i tagli sono di prima scelta. L'olio d'oliva è quasi totalmente "extravergine".

L'agroalimentare italiano è inoltre recepito come produzione di qualità che non si limita solo alla bontà e alla genuinità, ma significa anche garanzie di sicurezza e origine. Ne sono testimoni il fatturato al consumo Agroalimentare Made in Italy sui mercati nazionale ed estero, dalle produzioni a denominazione di origine protetta (DOP) e a indicazione d'origine protetta (IGP). La

posizione del nostro Paese nell'ambito dei prodotti a marchio DOP o IGP è poi di assoluto predominio: 202 prodotti (130 Dop, 72 Igp). La qualità infatti ha un costo: aderire al sistema Dop/Igp implica maggiori costi di produzione per il rispetto del disciplinare, costi di certificazione e soprattutto per tutelarsi dalle imitazioni frequenti nei mercati extra-europei.

I prodotti maggiormente esportati sono:

Salumi
Formaggi e latticini
Vini
Frutta fresca e secca
Pasta

Tabella 2.2: Prodotti esportati

I principali paesi importatori dei prodotti italiani sono: Germania, Francia, Regno Unito, Stati Uniti, Svizzera, Spagna, Austria e paesi emergenti come la Cina.

Dalla Figura 2.1 si nota come il settore agroalimentare sia in crescita rispetto agli anni precedenti sulla maggior parte dei fronti, partendo dai consumi fino ad arrivare al fatturato.

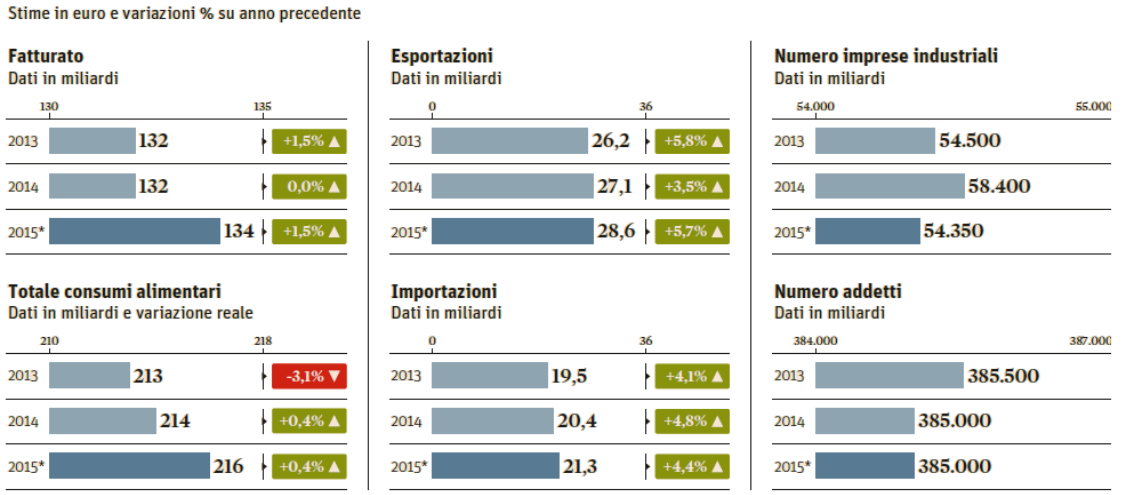

Figura 2.1: Dati ISTAT, Fonte ilSole24Ore

2.1.3 Azienda committente:

L'azienda fondata nei primi anni 60 opera nel settore agroalimentare, sia nel settore dei salumi, con una particolare specializzazione nel segmento del salame, che in quello delle carni fresche pollo, tacchino, bovino e suino. L'azienda opera principalmente in due stabilimenti e i suoi prodotti sono presenti in tutti i canali di vendita e in tutte le regioni Italiane. Negli ultimi anni ha sviluppato la sua presenza all'estero soprattutto nel comparto dei salumi.

Coltiva i terreni, gestisce gli allevamenti di proprietà e svolge internamente ogni fase della lavorazione, dalla macellazione alla produzione, controllando l'intera filiera delle carni suine e bovine, nonché dei prodotti di salumeria, garantendo produzioni di qualità e sicurezza alimentare.

Ad oggi l'azienda occupa circa 440 lavoratori. Inoltre ha creato nove punti vendita situati vicino agli stabilimenti produttivi per mantenere un contatto diretto con il consumatore e con il territorio. Alcuni negozi sono di proprietà, altri sono associati in partecipazione.

2.2 Analisi dei processi di business

Nella fase delle analisi dei requisiti le interazioni all'interno del team di lavoro e con i responsabili aziendali sono state molteplici e importanti, sia per comprendere la natura sia per conoscere a fondo le esigenze di analisi più adatte all' azienda. Sono stati accuratamente analizzati i processi aziendali, lavorando a stretto contatto con i responsabili di settore, intuendo le esigenze e le priorità di questi. Si è studiata la base di dati operazionale cercando di capire inizialmente la natura dei dati e successivamente se quest'ultimi fossero compatibili con le richieste di analisi fatte dai responsabili.

In questa fase preliminare si è rivelato fondamentale l'interazione con gli esperti della base di dati per chiarire dubbi e perplessità sorte analizzando l'elevata quantità di dati a disposizione. Una volta compresi tutti questi aspetti sono stati realizzati i requisiti di analisi. Si analizzano quindi i processi di business per il quale il cliente ha deciso di condurre le analisi e sono:

- *Processo Scontrini*
- *Processo Giacenze*
- *Processo Ricevimento Articoli*

2.2.1 Processo Scontrini

Analizzando l'obiettivo del processo degli scontrini è emerso nello studio, quello di fornire una fonte di informazione aggiornata e affidabile sulla situazione delle vendite giornaliera e sulle relative fasce orarie. Ogni emissione di scontrino avviene in maniera molto veloce e semplice, infatti il cliente sulla base delle proprie scelte personali seleziona gli articoli che vuole acquistare e una volta terminato questo iter può arrivare alla cassa oppure compiendo il pagamento presso il bancone dove effettua il pagamento in contanti, carta di credito o bancomat.

Nel caso di vendita effettuata ad un socio, quest'ultimo in qualità di membro ha diritto ad uno sconto aziendale. Durante lo studio preliminare del processo di vendita il responsabile del settore in questione ha espresso preferibilmente la volontà di poter stabilire e analizzare in maniera analitica per ogni scontrino emesso le informazioni di seguito qui elencate: gli articoli venduti per ogni

singolo scontrino emesso, l'incasso giornaliero per fasce orarie e per negozio e l'incasso per categoria merceologica.

2.2.2 Processo Giacenze

Sin dalla fondazione dell'azienda e con il suo successivo sviluppo, data una elevata quantità di prodotti movimentati e la presenza di depositi dislocati sul territorio, nati con lo scopo di rifornire i punti vendita, questi hanno fatto nascere l'esigenza e la possibilità concreta di studiare l'inventario di magazzino così da monitorare l'ammontare quantitativo delle scorte con il relativo importo monetario. Le quantità presenti all'interno di ogni deposito non sono costanti nel corso del tempo poiché giornalmente molti prodotti sono movimentati, affinché ciò accada il processo prevede le seguenti fasi: trasportati ai punti vendita, acquistati dai fornitori o semplicemente venduti.

Durante lo spostamento dei prodotti vengono monitorate sempre le quantità che nel processo di giacenza si dimostrano essenziali sia da acquistare che da essere vendute. Le quantità acquistate sono quelle che presenti nell'ingresso dei depositi sono dovuti a causa di acquisto o a causa di movimentazione produttiva, mentre quelle vendute sono quelle in uscita dai depositi causati da una vendita o da una movimentazione sempre collegata alla vendita dei prodotti.

Seguendo una modalità di ricerca qualitativa i responsabili intervistati nella loro sede di lavoro hanno espresso con convinzione il desiderio di poter osservare e successivamente di gestire al meglio le proprie rimanenze mensili di ogni deposito.

2.2.3 Processo Ricevimento Articoli

Durante lo studio è emerso che lo scenario del contesto di business dei ricevimenti merce si sviluppa attraverso il mezzo contenente la fornitura che arrivando presso uno dei depositi dell'azienda, in loco effettua il check-in e consegnando la bolla di accompagnamento, un operatore (buyer) inserisce i dati della fornitura nel gestionale con il quale il sistema in automatico aggiorna i dati relativi all'ordine corrispondente.

Selettivamente per ogni ordine di acquisto è possibile avere una quantità maggiore di consegne, quindi di conseguenza un numero di più ricevimenti ad ogni arrivo dove corrisponderà una bolla di accompagnamento. Il responsabile di settore viste le necessità del processo ha espresso la volontà di poter analizzare la quantità di merce ricevuta per fornitore, l'importo e la quantità di merce ricevuta per ogni mese riferita ad ogni articolo presente in ogni singolo deposito.

Capitolo 3

ANALISI DEI REQUISITI E PROGETTAZIONE CONCETTUALE INIZIALE

Si presentano nel seguito le prime fasi del processo di progettazione del data mart. sono descritte nel dettaglio le dimensioni, le gerarchie e le misure per ogni fatto individuato.

3.1 Introduzione al Data Warehousing

Definito da William H. Inmon nel 1990, un data warehouse è "orientato ai soggetti di interesse, integrato e consistente, rappresentativo dell'evoluzione temporale e non volatile". La costruzione di un sistema di data warehousing non comporta l'inserimento di nuove informazioni ma la riorganizzazione di quelle esistenti, e implica l'esistenza di un sistema informativo.

Mentre i dati operazionali coprono un arco temporale di solito limitato, poiché la maggior parte delle transazioni coinvolge i dati più recenti, il DW permette analisi che spazino su alcuni anni. Per questo motivo, esso è aggiornato ad intervalli regolari ed è in continua crescita. Proprio per il fatto che, in linea di principio, non vengano mai eliminati dati dal data warehouse e che gli aggiornamenti siano tipicamente eseguiti quando quest'ultimo è offline, fa sì che un DW possa essere considerato come un database di sola lettura.

Le fondamenti sono i *fatti* che riguardano particolari fenomeni aziendali (funzione o processo), per esempio, le vendite. Ogni fatto è caratterizzato da un insieme di *misure* che sono grandezze numerici che riguardano una prestazione o il comportamento di un fenomeno aziendale. Esempi di misure, nel caso di fatti sulle vendite, sono il prezzo del prodotto. Si analizzano le misure dei fatti secondo prospettive diverse di analisi, o *dimensioni*, per valutare i risultati del business nel contesto aziendale al fine di trovare soluzioni ai problemi critici o per cogliere nuove opportunità. Le dimensioni sono attributi che descrivono il contesto dei fatti. Per analizzare i fatti a diversi livelli di dettaglio è utile rappresentare non solo le dimensioni di analisi, ma anche le *gerarchie dimensionali* che interessano gli attribuiti delle dimensioni. Per esempio, la dimensione *tempo* è utile rappresentarla con attributi giorno, mese, trimestre e anno.

3.2 Processo Scontrini

Si descrivono la specifica dei requisiti e la modellazione concettuale iniziale del data mart degli scontrini.

3.2.1 Specifica dei requisiti

Dalla raccolta dei requisiti vengono derivate le dimensioni e le misure per ogni requisito di analisi.

Scontrini

N	Requisito di analisi	Dimensioni	Misure
1	Incasso per categoria merceologica, per negozio e per data	Data (Giorno, Mese, Anno), Categoria, Negozio	Importo, Scontrino Medio
2	Totale degli articoli venduti per negozio e per data	Data (Mese, Anno), Negozio	Scontrino Nr.
3	Incasso per fasce orarie per categoria merceologica e negozio	Data (Giorno, Mese, Anno), Orario (Fascia oraria), Categoria, Negozio	Importo, Scontrino Medio
4	Primi 15 articoli per mese e incasso	Data (Mese), Articolo	Importo
5	Sconto per categoria merceologica per negozio e per data	Data (Giorno, Mese, Anno), Categoria, Negozio	Sconto

Fatto Scontrini

Descrizione	Dimensioni preliminari	Misure preliminari
Il fatto è la singola riga di uno scontrino emesso	Data, Orario, Articolo, Categoria, Reparto cassa	Importo, Quantità, Sconto

Si elencano di seguito le dimensioni, gli attributi e le misure del Fatto Scontrini.

Dimensioni

Nome	**Descrizione**	**Granularità**
Data	Dimensione temporale	Un giorno
Orario	Dimensione temporale	Orario
Categoria	Classificazione merceologica aziendale	Una categoria
Articolo	Anagrafica articolo	Un articolo
Negozio	Anagrafica negozio	Un negozio

Data definisce il momento di emissione dello scontrino.

Data

Attributo	**Descrizione**
Giorno	Giorno in formato YYYYMMDD.
Settimana	Settimana in formato YYYYWW.
Giorno Settimana	Giorno della settimana (es. dal *Lunedì* alla *Domenica*).
Mese	Mese in formato YYYY MMM (es. 2016 Gennaio).
Trimestre	Trimestre in formato YYYYQ.
Anno	Anno in formato YYYY.

La dimensione orario definisce l'orario di emissione dello scontrino.

Orario

Attributo	**Descrizione**
Ora	Ora in formato HH:mm.
Fascia Oraria	Fascia oraria in formato HH.

Articolo

Attributo	Descrizione
Codice articolo	Codice articolo
Articolo	Descrizione dell'articolo (es. fettine pollo).
Codice categoria	Codice categoria
Categoria	Descrizione della categoria dell'articolo (es. Suino).

Negozio

Attributo	Descrizione
Codice negozio	Codice negozio
Negozio	Descrizione del negozio.

Reparto Cassa

Attributo	Descrizione
Reparto cassa	Descrizione del reparto cassa

Sono descritte le gerarchie dimensionali, con gli attributi che le compongono e il tipo di gerarchia.

Gerarchie dimensionali

Dimensione	Descrizione	Tipo
Data	Giorno → Mese → Trimestre → Anno	Bilanciata
Data	Giorno → Settimana → Anno	Bilanciata
Articolo	Codice Articolo → Categoria	Bilanciata

Sono analizzate le misure del fatto Scontrini e il loro tipo di aggregabilità.

Misure

Misure	Descrizione	Aggregabilità	Calcolata
Importo (Imp)	Importo speso	Additiva	No
Quantità (Qtà)	Quantità acquistata	Additiva	No
Sconto (Sc)	Sconto	Additiva	No
Scontrino Nr.(Nr)	Numero scontrini emessi	Semi Additiva	No
Scontrino Medio	$\frac{Imp}{Nr}$	Additiva	Si

3.2.2 Progettazione concettuale iniziale

La raccolta dei requisiti e la traduzione formale di questi ha permesso di identificare requisiti specifici sul fatto, sulle misure associate e sulle dimensioni, inclusa la definizione di gerarchie dimensionali. È possibile esprimere e rappresentare graficamente una sintesi delle informazioni raccolte attraverso il formalismo *DFM* (Dimensional Fact Model). Il modello ottenuto è presentato nella Figura 3.1.

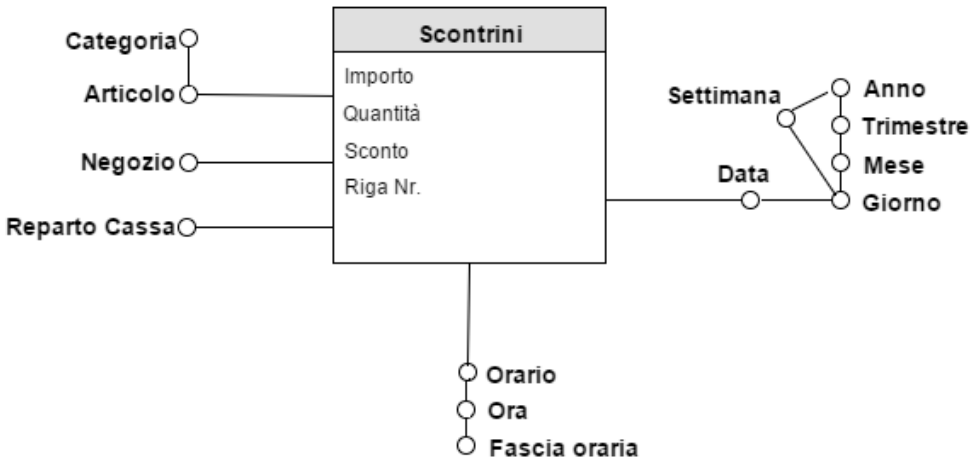

Figura 3.1: Schema concettuale iniziale

3.3 Processo Giacenze

Si descrivono la specifica dei requisiti e la modellazione concettuale iniziale del data mart delle giacenze.

3.3.1 Specifica dei requisiti

Dalla raccolta dei requisiti vengono derivate le dimensioni e le misure per ogni requisito di analisi.

Giacenze

N	Requisito di analisi	Dimensioni	Misure
1	Totale delle quantità acquistate e del loro valore per fornitore e categoria merceologica	Data (Giorno, Mese, Anno), Categoria, Fornitore	Importo, Quantità
2	Totale degli articoli in deposito	Data (Mese, Anno),	Importo, Quantità
3	Totale delle quantità acquistate e del loro valore per articolo e deposito	Data (Giorno, Mese, Anno), Deposito, Articolo	Importo, Quantità

Fatto Giacenze

Descrizione	Dimensioni preliminari	Misure preliminari
Il fatto è una rimanenza di un articolo	Data, Deposito, Articolo, Categoria, Società	Importo, Quantità

Si elencano di seguito le dimensioni, gli attributi e le misure del Fatto Giacenze.

Dimensioni

Nome	**Descrizione**	**Granularità**
Data	Dimensione temporale	Un giorno
Articolo	Classificazione merceologica aziendale	Un articolo
Categoria	Classificazione merceologica aziendale	Una categoria
Fornitore	Anagrafica fornitori	Un fornitore
Società	Anagrafica società	Una società

La dimensione Data definisce il momento di giacenza della merce.

Data

Attributo	**Descrizione**
Giorno	Giorno in formato YYYYMMDD.
Settimana	Settimana in formato YYYYWW.
Giorno Settimana	Giorno della settimana (es. dal *Lunedì* alla *Domenica*).
Mese	Mese in formato YYYY MMM (es. 2016 Gennaio).
Trimestre	Trimestre in formato YYYYQ.
Anno	Anno in formato YYYY.

Articolo

Attributo	**Descrizione**
Codice articolo	Codice articolo
Articolo	Descrizione dell'articolo (es. fettine pollo).
Codice categoria	Codice categoria
Categoria	Descrizione della categoria dell'articolo (es. Suino).

Deposito

Attributo	Descrizione
Codice deposito	Codice deposito
Deposito	Riferimento geografico del deposito.

Fornitore

Attributo	Descrizione
Codice fornitore	Codice fornitore
Fornitore	Descrizione del fornitore.

Società

Attributo	Descrizione
Codice società	Codice società
Società	Descrizione della società.

Sono descritte le gerarchie dimensionali, con gli attributi che le compongono e il tipo di gerarchia.

Gerarchie dimensionali

Dimensione	Descrizione	Tipo
Data	Giorno → Mese → Trimestre → Anno	Bilanciata
Data	Giorno → Settimana → Anno	Bilanciata
Articolo	Codice Articolo → Categoria	Bilanciata

Sono analizzate le misure del fatto Giacenze e il loro tipo di aggregabilità.

Misure

Misure	Descrizione	Aggregabilità	Calcolata
Importo	Importo	Additiva	No
Quantità	Quantità della merce	Additiva	No

3.3.2 Progettazione concettuale iniziale

In Figura 3.2 è mostrato il diagramma concettuale iniziale del fatto Giacenze, risultato della modellazione dei requisiti raccolti.

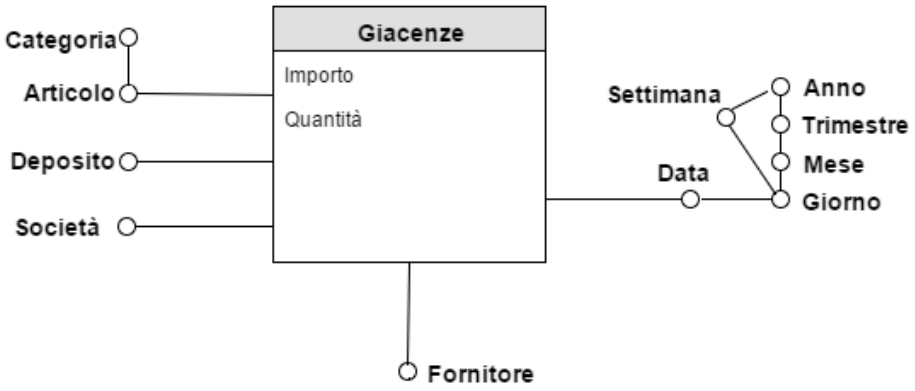

Figura 3.2: Schema concettuale iniziale

3.4 Processo Ricevimento Articoli

Si descrivono la specifica dei requisiti e la modellazione concettuale iniziale del data mart dei ricevimenti degli articoli.

3.4.1 Specifica dei requisiti

Dalla raccolta dei requisiti vengono derivate le dimensioni e le misure per ogni requisito di analisi.

Ricevimento

N	Requisito di analisi	Dimensioni	Misure
1	Quantità di merce ricevuta per fornitore, per mese.	Data (Mese, Anno), Categoria, Fornitore	Importo, Quantità
2	Importo e quantità di merce ricevuta per articolo e per deposito	Data (Mese, Anno), Articolo, Deposito	Importo, Quantità
3	Importo e quantità di merce accettata per buyer	Data (Mese, Anno), Articolo, Buyer	Importo, Quantità
4	Numero di articoli ricevuti per mese e per deposito	Data (Mese, Anno), Articolo, Deposito	Count

Fatto Articoli Ricevuti

Descrizione	Dimensioni preliminari	Misure preliminari
il fatto è la singola riga della bolla di accompagnamento per avere il dettaglio sull'articolo	Data, Deposito, Articolo, Categoria, Fornitore, Numero Ricevimento, Buyer	Importo, Quantità

Si descrivono di seguito le dimensioni, gli attributi e le misure del Fatto Ricevimento Articoli.

Dimensioni

Nome	**Descrizione**	**Granularità**
Data	Dimensione temporale	Un giorno
Articolo	Classificazione merceologica aziendale	Un articolo
Categoria	Classificazione merceologica aziendale	Una categoria
Fornitore	Anagrafica fornitori	Un fornitore
Deposito	Anagrafica dei depositi	Un deposito
Numero Ricevimento	Numero progressivo associato al ricevimento di merce	Un ricevimento
Buyer	Anagrafica buyer	Un buyer

La dimensione Data definisce il momento di ricevimento.

Data

Attributo	**Descrizione**
Giorno	Giorno in formato YYYYMMDD.
Settimana	Settimana in formato YYYYWW.
Giorno Settimana	Giorno della settimana (es. dal *Lunedì* alla *Domenica*).
Mese	Mese in formato YYYY MMM (es. 2016 Gennaio).
Trimestre	Trimestre in formato YYYYQ.
Anno	Anno in formato YYYY.

Articolo

Attributo	**Descrizione**
Codice articolo	Codice articolo
Articolo	Descrizione dell'articolo (es. fettine pollo).
Codice categoria	Codice categoria
Categoria	Descrizione della categoria dell'articolo (es. Suino).

Deposito

Attributo	Descrizione
Codice deposito	Codice deposito
Deposito	Riferimento geografico del deposito.

Fornitore

Attributo	Descrizione
Codice fornitore	Codice fornitore
Fornitore	Descrizione del fornitore.

Buyer

Attributo	Descrizione
Codice buyer	Codice buyer
Buyer	Dipendente dell'ufficio acquisti.

Numero Ricevimento

Attributo	Descrizione
Numero Ricevimento	Numero progressivo associato al ricevimento di merce

Sono elencate le gerarchie dimensionali, con gli attributi che le compongono e il tipo di gerarchia.

Gerarchie dimensionali

Dimensione	Descrizione	Tipo
Data	Giorno → Mese → Trimestre → Anno	Bilanciata
Data	Giorno → Settimana → Anno	Bilanciata
Articolo	Codice Articolo → Categoria	Bilanciata

Sono analizzate le misure del fatto Ricevimento Articoli e il loro tipo di aggregabilità.

Misure

Misure	Descrizione	Aggregabilità	Calcolata
Importo	Importo	Additiva	No
Quantità	Quantità della merce	Additiva	No

3.4.2 Progettazione concettuale iniziale

In Figura 3.3 è mostrato il modello concettuale del data mart per l'analisi degli articoli ricevuti. Le misure della tabella dei fatti sono Importo e Quantità, che contengono l'importo e la quantità di merce ricevuta. Per quanto riguarda le dimensioni invece, questo modello è composto da: Data, Articolo, Fornitore, Buyer, Deposito e della dimensione degenere Numero Ricevimento.

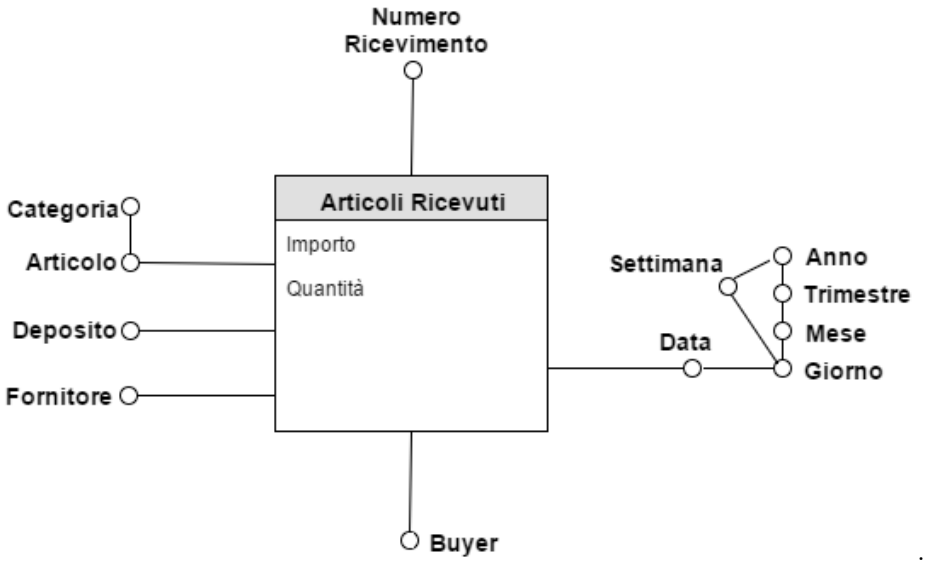

Figura 3.3: Schema concettuale iniziale

3.5 Riepilogo delle dimensioni e delle misure

Si mostrano le tabelle di riepilogo delle dimensioni e delle misure condivise tra i processi.

Dimensioni dei processi

Dimensione	**Scontrini**	**Giacenze**	**Ricevimento Articoli**
Data	X	X	X
Orario	X		
Articolo	X	X	X
Negozio	X		
Fornitore	X	X	
Deposito		X	X
Società		X	
Buyer			X
Numero Ricevimento			X
Reparto Cassa	X		

Misure dei processi

Misura	**Scontrini**	**Giacenze**	**Ricevimento Articoli**
Importo	X	X	X
Quantità	X	X	X
Scontrino Nr.	X		
Sconto	X		
Scontrino Medio	X		

3.6 Cambiamento delle dimensioni

Un attento confronto con i responsabili funzionali della società ha permesso di stabilire in che modo trattare le dimensioni con attributi che potrebbero cambiare nel tempo.

Esistono quattro tipologie per il trattamento delle dimensioni. In particolare, tre sono riferite alla dimensioni con attributi che cambiano raramente *slowly changing dimensions*, mentre la restante è riferita alle dimensioni con attributi che cambiano frequentemente. Di seguito vengono descritte le quattro tipologie appena menzionate:

- *Tipo 1 - Perdita della storia*: il valore dell'attributo dimensionale modificato viene sostituito con il nuovo valore. Questa è prevalentemente la soluzione più immediata, ma non permette di storicizzare i cambiamenti.
- *Tipo 2 - Storicizzazione*: in questo caso viene aggiunta una nuova riga alla tabella dimensionale, creando una nuova entità. Tutti i fatti verificati precedentemente alla modifica, continuano a far riferimento alla vecchia entità, mentre tutti i fatti che si verificano successivamente alla modifica faranno riferimento a quella nuova. Questa tipologia permette di storicizzare il cambiamento ma aumenta la mole di dati in fase di caricamento.
- *Tipo 3 - Storicizzazione con data cambiamento*: questa tipologia permette di mantenere lo storico così come il tipo 2, inoltre consente di memorizzare anche il momento temporale in cui avviene il cambiamento. Per ottenere questo risultato vi è la necessità di sostituire il vecchio attributo con tre campi: Attributo, Nuovo_Attributo, Data_modifica.
- *Tipo 4 - Elevata frequenza di cambiamento*: per gli attributi con un'alta frequenza di cambiamento è preferibile creare due tabelle dimensionali contenenti, in una gli attributi che rimangono invariati e nell'altra gli attributi che variano frequentemente.

In questo caso di studio non è necessario tenere traccia delle modifiche. Per questo motivo tutte le dimensioni sono trattate con il *Tipo 1*.

Capitolo 4

PROGETTAZIONE CONCETTUALE FINALE E LOGICA

Dopo la fase di modellazione concettuale iniziale guidata dalle analisi, realizzata nel precedente capitolo, il successivo passaggio è quello di affrontare le ultime fasi relative al design del data mart secondo il modello adottato. Il modello viene arricchito dagli schemi iniziali con eventuali altre informazioni interessanti, e presenta la base dati operazionale al fine di individuare le fonti dati ed arricchire gli schemi iniziali con eventuali altre informazioni.

4.1 Il sistema sorgente

La base dati operazionale che sarà la sorgente unica per il *data warehouse* è così descritta: l'analisi dei dati operazionali permetterà di estendere il modello concettuale iniziale presentato nel precedente capitolo, con nuovi attributi di interesse e di realizzazione del modello concettuale finale.

4.1.1 JD Edwards

JD Edwards, nota anche come JDE, è una piattaforma integrata di software ERP (*Enterprise Resource Planning*) prodotto di proprietà della Oracle Corporation. Questa conta un numero elevato di oltre 80 moduli che compongono l'offerta attuale del prodotto che sono in grado di fornire alle aziende di medie e grandi dimensioni, dimostrandosi uno strumento flessibile e capace di suppor-

tare un'ampia gamma di processi di business. JDE è utilizzato da circa 10.000 aziende nel mondo, di cui 500 in Italia. Il sistema gestionale può essere personalizzato per rispondere a specifiche esigenze con l'introduzione di set di valori validi per una categoria che si vuole descrivere.

Il modulo utilizzato per gestire i processi di scontrini, giacenze e ricevimento articolo è il *Financial Management*, componente per la gestione della dimensione finanziaria ed economica;

Alcune delle tabelle sorgenti dalle quali sono estratti i dati sono riassunte qui di seguito:

Tabella	**Prefisso**	**Descrizione**
F43121	PR	Ricevimenti di Acquisto
F4111	IL	Movimenti Magazzino
F0010	CC	Società
F0101	AB	Rubrica Indirizzi
F4101	IM	Articoli
F0006	MC	Società
F0401	A6	Fornitori

4.2 Modellazione concettuale finale

Dalla raccolta e specifica dei requisiti, passando per le fasi di modellazione concettuale iniziale, si arriva alla definizione del modello concettuale finale dei data mart. Questa fase evidenzia ulteriori dimensioni di analisi e misure fornite dal sistema sorgente che inizialmente non sono state incluse nella modellazione concettuale, in quanto non rilevate dalla raccolta dei requisiti.

Di seguito verranno descritte quelle incluse nel modello per completare l'offerta informativa del sistema o per prevedere eventuali esigenze informative future.
Successivamente sarà presentato il modello logico dei data mart e del data warehouse.

4.2.1 Processo Scontrini

La descrizione finale del processo scontrini è il seguente:

Fatto Scontrini

Descrizione	**Dimensioni**	**Misure**
Il fatto è la singola riga di uno scontrino emesso	Data, Orario, Articolo, Categoria, Reparto cassa, Unità di Misura	Importo, Quantità, Sconto, Riga Nr,

Si descrivono di seguito le dimensioni e la loro granularità.

Dimensioni

Nome	**Descrizione**	**Granularità**
Data	Dimensione temporale	Un giorno
Orario	Dimensione temporale	Orario
Categoria	Classificazione merceologica aziendale	Una categoria
Articolo	Anagrafica articolo	Un articolo
Negozio	Anagrafica negozio	Un negozio
Unità di misura	Anagrafica unità di misura	Un'unità di misura

Qui di seguito si elenca solo la nuova dimensione Unità di Misura. Essa è necessaria, per esempio, per distinguere gli articoli venduti in *Kg* oppure in *Pz*

Unità di misura

Attributo	**Descrizione**
Unità di Misura	Codice dell'unità di misura

Sono esposte le gerarchie dimensionali, con gli attributi che le compongono e il tipo di gerarchia.

Gerarchie dimensionali

Dimensione	**Descrizione**	**Tipo**
Data	Giorno → Mese → Trimestre → Anno	Bilanciata
Data	Giorno → Settimana → Anno	Bilanciata
Articolo	Codice Articolo → Categoria	Bilanciata

Sono analizzate le misure del fatto Scontrini, il loro tipo di aggregabilità e la loro formula in caso di misure calcolate.

Misure

Misure	**Descrizione**	**Aggregabilità**	**Calcolata**
Importo (Imp)	Importo speso	Additiva	No
Quantità (Qtà)	Quantità acquistata	Additiva	No
Sconto (Sc)	Sconto	Additiva	No
Scontrino Nr.(Nr)	Numero scontrini emessi	Semi Additiva	No
Scontrino Medio	$\frac{Imp}{Nr}$	Semi Additiva	Si
Sconto Medio	$\frac{Sc}{Nr}$	Semi Additiva	Si
Incidenza Sconto	$\frac{Sc}{Nr}$	Semi Additiva	Si
Riga Nr	Numero articoli acquistati	Additiva	No
Battuta Media	$\frac{Riga}{Nr}$	Semi Additiva	Si
Importo AP (Imp)	AGO(Imp,Year,1)	Additiva	Si
Quantità AP (Qtà)	AGO(Qtà,Year,1)	Additiva	Si
Giorni Apertura (Gr)	Numero giorni aperti	Semi Additiva	Si
Incasso Medio	$\frac{Imp}{Gr}$	Semi Additiva	Si
Nr Scontrini Medio per Giorni	$\frac{Nr}{Gr}$	Semi Additiva	Si

In Figura 4.1 è mostrato il diagramma concettuale finale del fatto Scontrini.

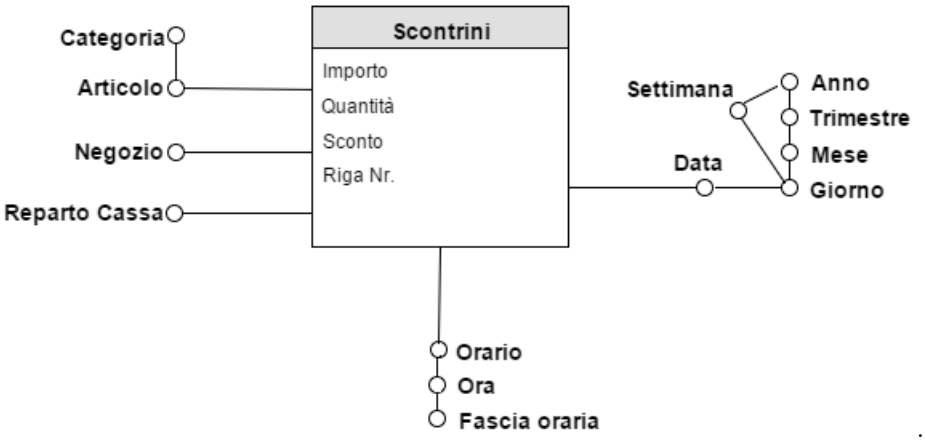

Figura 4.1: Schema concettuale finale

4.2.2 Processo Giacenze

La descrizione finale del processo giacenze è il seguente:

Fatto Giacenze

Descrizione	**Dimensioni**	**Misure**
Il fatto è una rimanenza di un articolo	Data, Deposito, Articolo, Categoria, Società, Unità di Misura	Importo, Quantità

Si descrivono di seguito le dimensioni e la loro granularità.

Dimensioni

Nome	**Descrizione**	**Granularità**
Data	Dimensione temporale	Un giorno
Orario	Dimensione temporale	Orario
Categoria	Classificazione merceologica aziendale	Una categoria
Articolo	Anagrafica articolo	Un articolo
Fornitore	Anagrafica fornitori	Un fornitore
Deposito	Anagrafica deposito	Un deposito
Società	Anagrafica società	Una società
Unità di misura	Anagrafica unità di misura	Un'unità di misura

Di seguito si elenca la nuova dimensione Unità di Misura già definita precedentemente.

Unità di misura

Attributo	**Descrizione**
Unità di Misura	Codice dell'unità di misura

Sono esposte le gerarchie dimensionali, con gli attributi che le compongono e il tipo di gerarchia.

Gerarchie dimensionali

Dimensione	**Descrizione**	**Tipo**
Data	Giorno → Mese → Trimestre → Anno	Bilanciata
Data	Giorno → Settimana → Anno	Bilanciata
Articolo	Codice Articolo → Categoria	Bilanciata

Sono descritte le misure del fatto Giacenze e il loro tipo di aggregabilità.

Misure

Misure	Descrizione	Aggregabilità	Calcolata
Importo	Importo	Additiva	No
Quantità	Quantità della merce	Additiva	No
Importo AP (Imp)	AGO(Imp,Year,1)	Additiva	Si
Quantità AP (Qtà)	AGO(Qtà,Year,1)	Additiva	Si

In Figura 4.2 è mostrato il diagramma concettuale finale del fatto Giacenze.

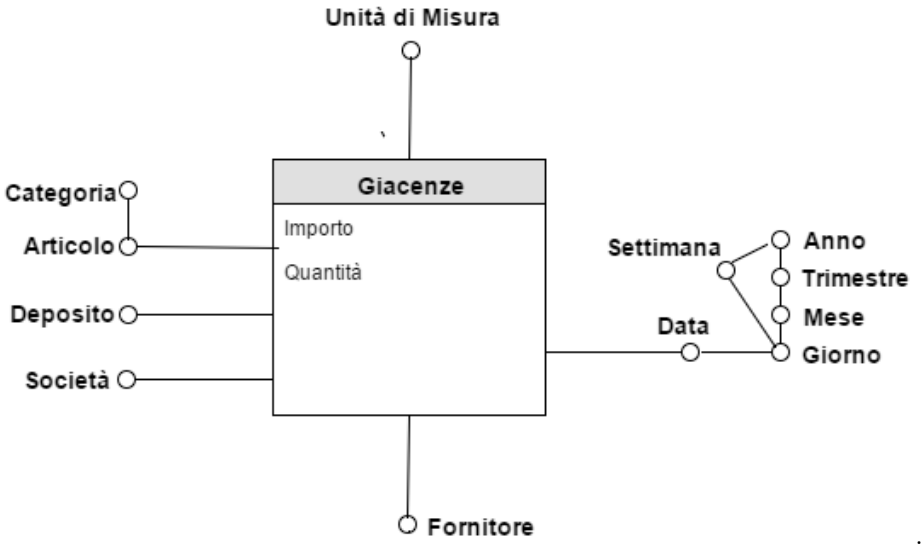

Figura 4.2: Schema concettuale finale

4.2.3 Processo Ricevimento Articoli

Fatto Articoli Ricevuti

Descrizione	**Dimensioni**	**Misure**
il fatto è la singola riga della bolla di accompagnamento per avere il dettaglio sull'articolo	Data, Deposito, Articolo, Categoria, Fornitore, Numero Ricevimento, Buyer, Unità di Misura	Importo, Quantità

Si descrivono di seguito le dimensioni, gli attributi e le misure del Fatto Ricevimento Articoli.

Dimensioni

Nome	**Descrizione**	**Granularità**
Data	Dimensione temporale	Un giorno
Categoria	Classificazione merceologica aziendale	Una categoria
Articolo	Classificazione merceologica aziendale	Un articolo
Fornitore	Anagrafica fornitori	Un fornitore
Deposito	Anagrafica dei depositi	Un deposito
Numero Ricevimento	Numero progressivo associato al ricevimento di merce	Un ricevimento
Buyer	Anagrafica buyer	Un buyer
Unità di misura	Anagrafica unità di misura	Un'unità di misura

Sono illustrate le gerarchie dimensionali, con gli attributi che le compongono e il tipo di gerarchia.

Gerarchie dimensionali

Dimensione	Descrizione	Tipo
Data	Giorno → Mese → Trimestre → Anno	Bilanciata
Data	Giorno → Settimana → Anno	Bilanciata
Articolo	Codice Articolo → Categoria	Bilanciata

Sono elencate le misure del fatto Articoli Ricevuti e il loro tipo di aggregabilità.

Misure

Misure	Descrizione	Aggregabilità	Calcolata
Importo	Importo	Additiva	No
Quantità	Quantità della merce	Additiva	No
MQ Ricevuti	Quantità in metri quadrati	Additiva	No
Importo AP (Imp)	AGO(Imp,Year,1)	Additiva	Si
Quantità AP (Qtà)	AGO(Qtà,Year,1)	Additiva	Si

In Figura 4.3 è mostrato il diagramma concettuale finale del fatto Articoli Ricevuti.

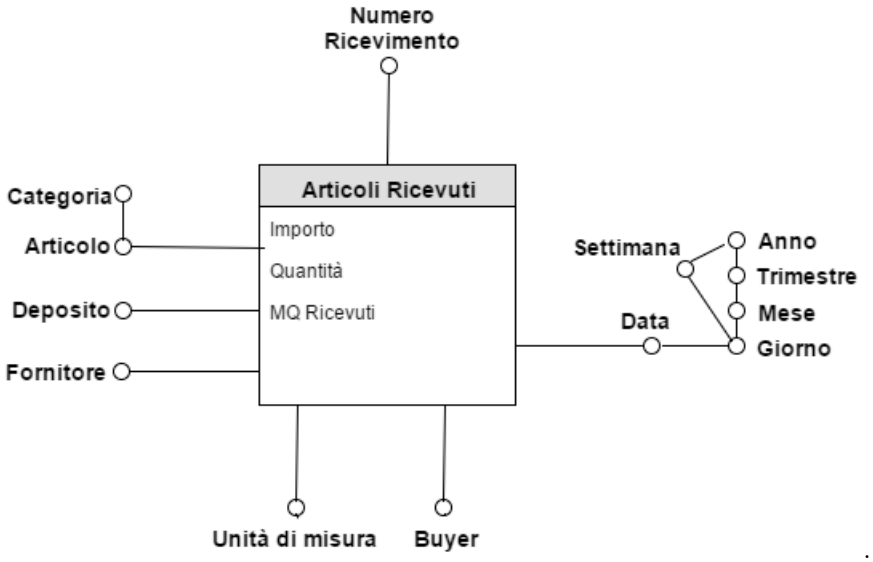

Figura 4.3: Schema concettuale finale

4.2.4 Riepilogo delle dimensioni e delle misure

Si mostrano le tabelle di riepilogo delle dimensioni e delle misure condivise tra i processi.

Dimensioni dei processi

Dimensione	**Scontrini**	**Giacenze**	**Ricevimento Articoli**
Data	X	X	X
Orario	X		
Articolo	X	X	X
Negozio	X		
Fornitore	X	X	
Deposito		X	X
Società		X	
Buyer			X
Numero Ricevimento			X
Reparto Cassa	X		
Unità di Misura	X	X	X

Misure dei processi

Misura	**Scontrini**	**Giacenze**	**Ricevimento Articoli**
Importo	X	X	X
Quantità	X	X	X
Scontrino Nr.	X		
Sconto	X		
Scontrino Medio	X		
Sconto Medio	X		
Incidenza Sconto	X		
Riga Nr	X		
Battuta Media	X		
Importo AP	X	X	X
Quantità AP	X	X	X

4.3 Modellazione logica del data mart

Tutte le dimensioni degeneri diventano tabelle dimensionali e non attributi delle tabelle dei fatti, in quanto rappresentano dimensioni interessanti condivise tra i data mart.

Le dimensioni relative alla data sono di tipo denormalizzato e hanno chiave primaria in formato *aaaammgg*; gli altri attributi sono rappresentazioni in formato testuale utili ai fini di presentazione nella reportistica.

Nessuna misura calcolata è presente nella progettazione logica, dato che i valori delle misure derivate non vengono memorizzati nel data warehouse. Esse sono state create direttamente nel tool di Business Intelligence, descritto nel Capitolo 7.

Non interessa mantenere lo storico per gli attributi dimensionali che sono stati modificati pertanto tutte le *Slowly Changing Dimension* sono gestite con la strategia di Tipo 1, ossia il nuovo valore sostituisce il vecchio valore.

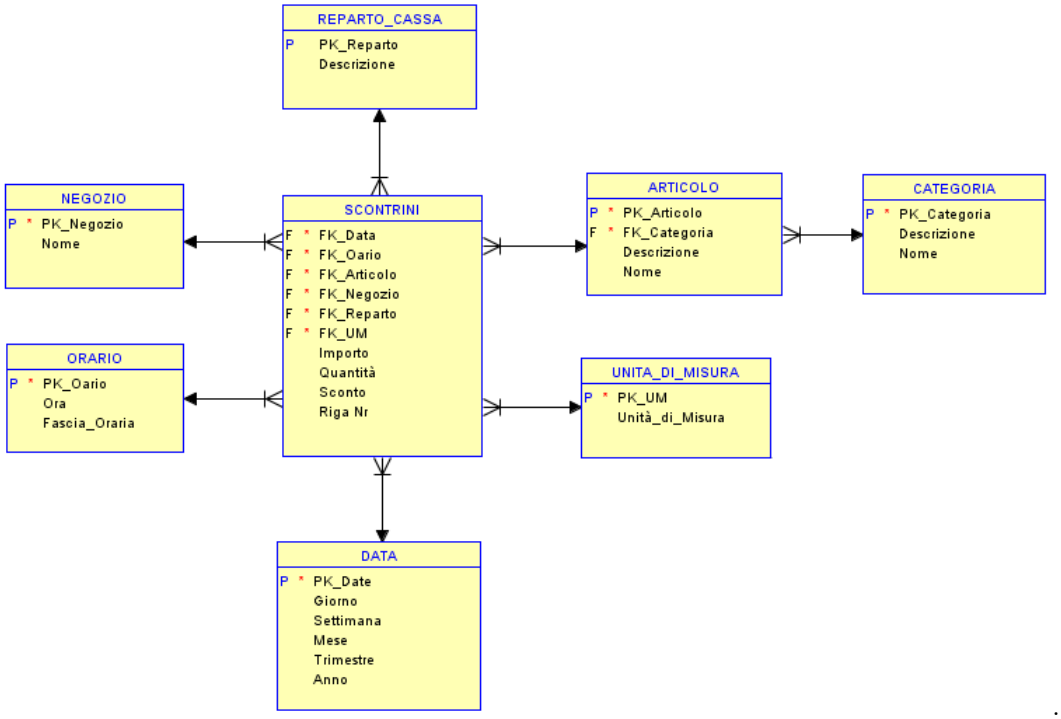

Figura 4.4: Modellazione logica del fatto Scontrini

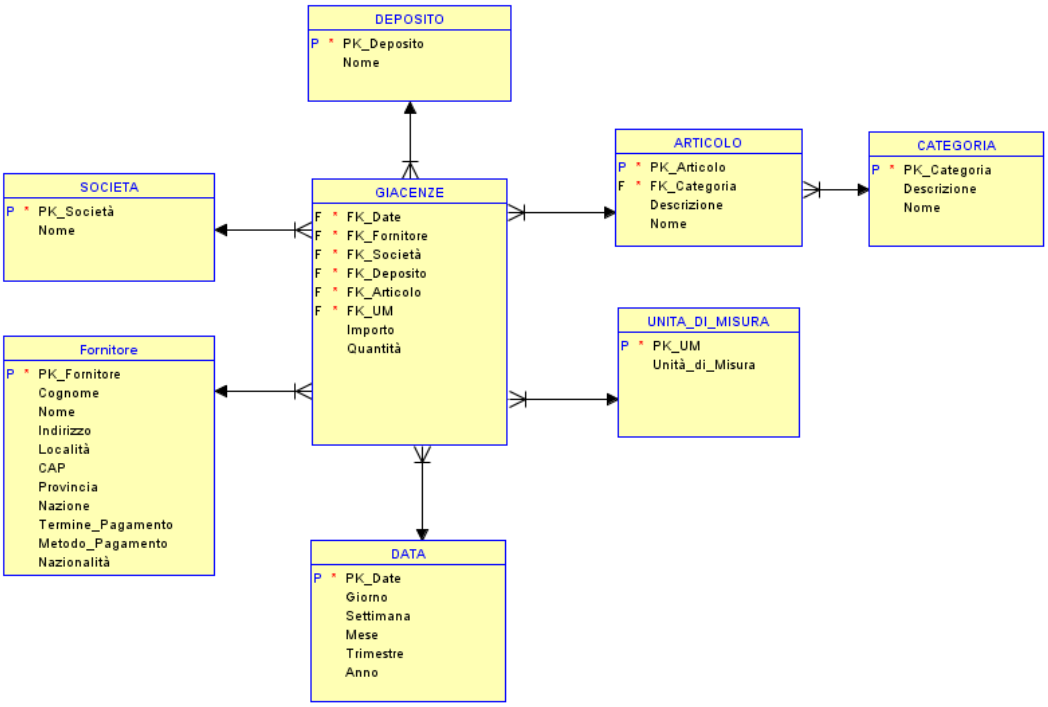

Figura 4.5: Modellazione logica del fatto Giacenze

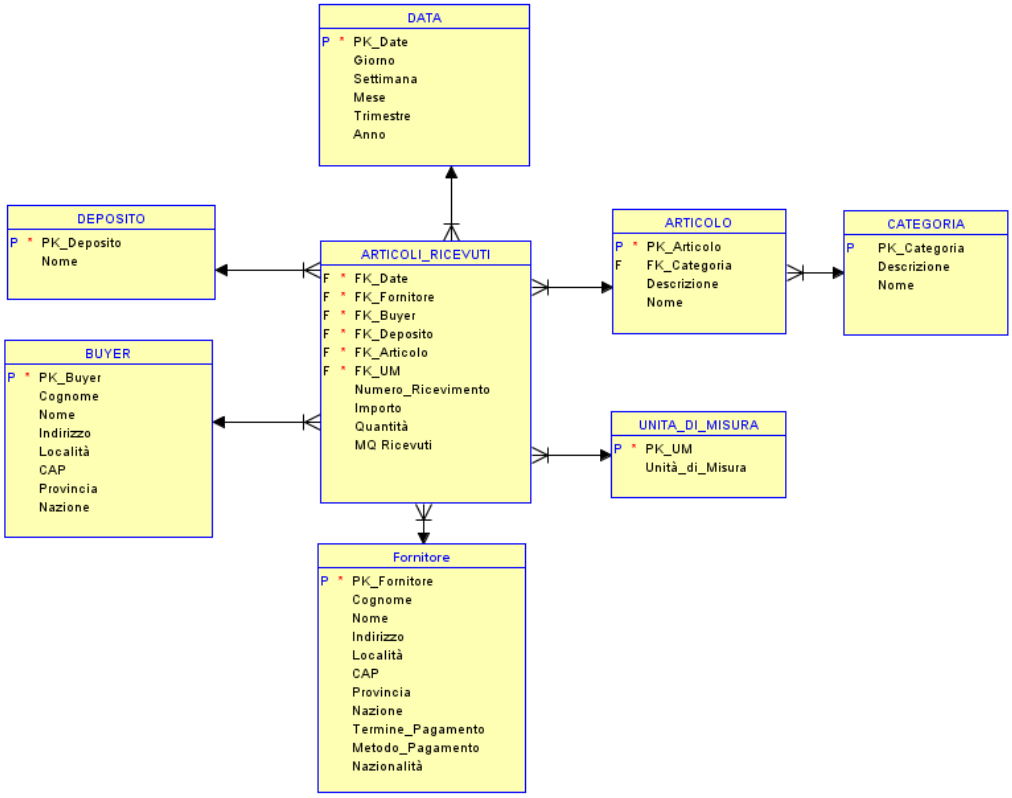

Figura 4.6: Modellazione logica del fatto Ricevimento Articoli

4.4 Modellazione logica del data warehouse

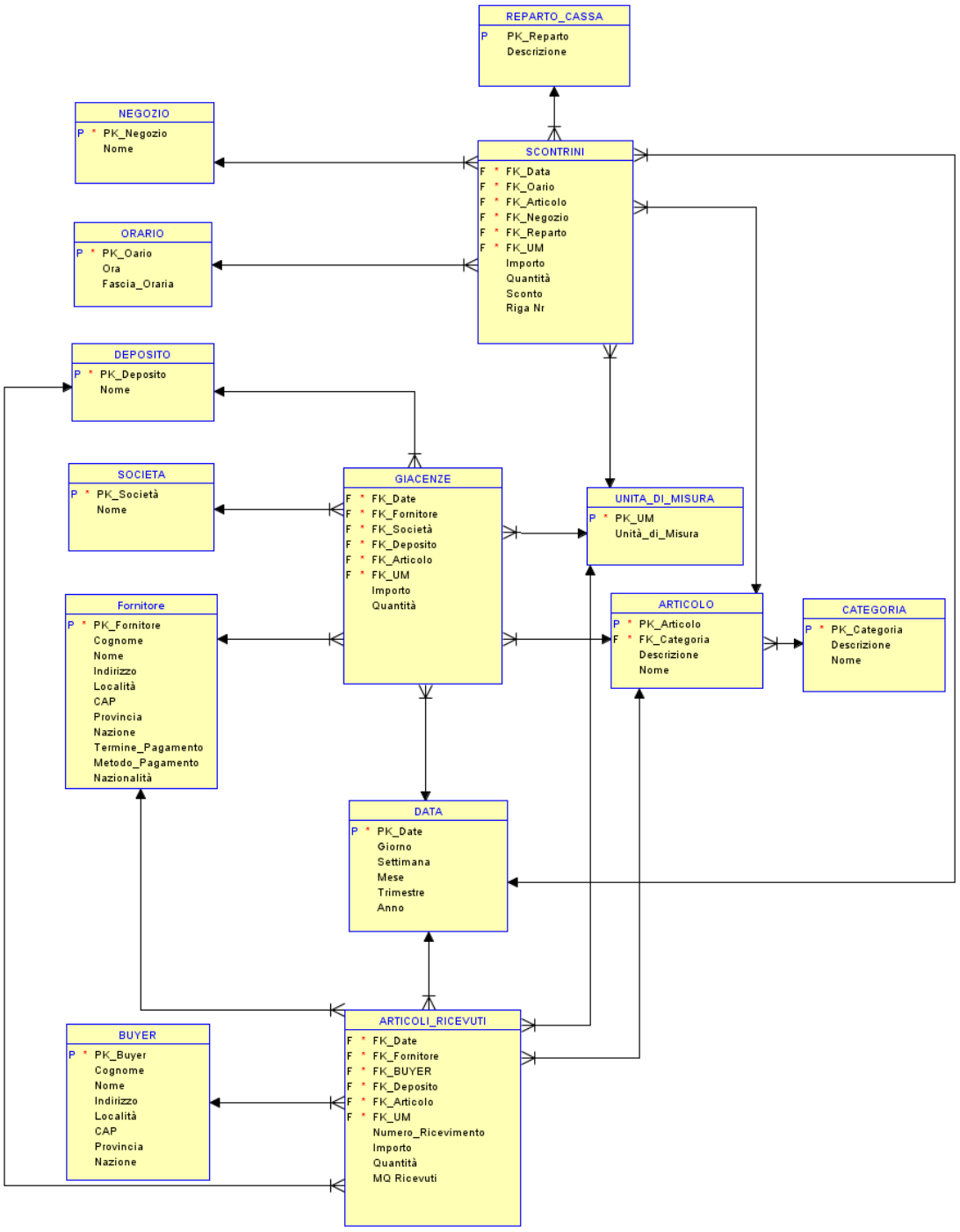

Figura 4.7: Modellazione logica del data warehouse

Capitolo 5

AMBIENTE DI SVILUPPO

Una volta compresa la struttura della base di dati per il supporto alle decisioni da realizzare si può passare allo studio dell'ambiente di sviluppo in cui si andrà a realizzare il data warehouse e la business intelligence.

5.1 Oracle

Oracle è uno tra i più famosi database management system (DBMS), cioè sistema di gestione di basi di dati, scritto in linguaggio C. Oracle fa parte dei cosiddetti RDBMS (*Relational DataBase Management System*) ossia di sistemi di database basati sul modello relazionale che si è affermato come lo standard dei database dell'ultimo decennio.

La società informatica che lo produce, la Oracle Corporation, è una delle più grandi del mondo. È stata fondata nel 1977 ed ha la sua sede centrale in California. La prima versione di Oracle risale anche nel 1977, da allora sono state introdotte numerose modifiche e miglioramenti per seguire gli sviluppi tecnologici, fino ad arrivare alla versione 12c R1.

Oracle, con i suoi prodotti, è il primo produttore di sistemi per data warehouse e basi di dati in termini di fatturato, e si posiziona come uno dei principali competitor nel Magic Quadrant 2015 della Gartner.

Il sistema utilizzato in questo progetto è Oracle Database 11g Release 2, un *database management system* (DBMS) relazionale a oggetti, che supporta nativamente le principali funzioni SQL analitiche.

Una delle sue risorse è la possibilità di poter memorizzare ed eseguire procedure e funzioni in linguaggio Oracle PL/SQL: questa funzionalità è stata utilizzata per lo sviluppo del processo ETL, come esposto a seguire nel Capitolo 6.

Oracle Database è un sistema di tipo OLAP-aware, che permette di essere estendibile a pagamento con strutture di accesso e di memorizzazione specifiche per la gestione di data warehouse. In merito, relativamente agli elevati costi e al susseguente volume di dati molto contenuto, si è deciso di affidarsi alla decisione responsabile di rinunciare a queste funzionalità ritenute eccedenti.

5.1.1 Oracle SQL Developer e Data Modeler

Oracle SQL Developer è l'ambiente di sviluppo integrato (*integrated development environment*) utilizzato per la realizzazione e la gestione del sistema di data warehouse su Oracle Database. Lo strumento consente a sviluppatori, progettisti e DBA di lavorare interfacciandosi con il database al fine di interagire con gli oggetti presenti nella banca dati.

Oracle SQL Developer supporta prodotti Oracle e una varietà di plugin di terze parti dove gli utenti possono connettersi al database non Oracle. Alcuni di questi database di terzi parti sono: *IBM DB2, Microsoft Access, Microsoft SQL Server, MySQL, e Teradata*.

Con il fine di implementare il data warehouse descritto in questo lavoro è emersa la necessità di utilizzare delle funzionalità di Oracle SQL Developer per:

- la creazione dell'interfaccia con la base di dati operazionale,
- la creazione e il debug delle procedure e delle funzioni in linguaggio PL/-SQL,
- l'esecuzione di interrogazioni SQL e la verifica dei piani di accesso,
- la creazione e gestione dei backup,
- la verifica delle prestazioni complessive del sistema.

Per la gestione del modello logico e del modello fisico, e per la generazione degli script SQL *Data Definition Language* (DDL) per i backup è stata invece utilizzata l'estensione di SQL Developer chiamata Data Modeler.

Si tratta di uno strumento stand-alone che supporta il modeling logico, relazionale e multidmensionale. Consente di aiutare gli sviluppatori a progettare i data model per Oracle Database in modo semplice e rapido. Esso fornisce capacità di reverse engineering e di progettazione e può essere utilizzato sia in ambiente tradizionali sia in ambienti cloud.

5.2 MicroStrategy

Analizzando il report Magic Quadrant del 2016, Figura 5.1, relativo ai fornitori di piattaforme per la business intelligence risulta evidente quanto sia cresciuta la competizione in tema di *analytics* a seguito di una sempre maggiore domanda da parte delle aziende.

MicroStrategy e gli altri maggiori player mondiali del settore (ad es. Microsoft), vista l'elevata concorrenza che nell'ultimo anno ha subito un'inaspettata impennata, questi sono state vittime di una notevole sofferenza, causata dall'incremento della competizione di fornitori emergenti quali Tableau e Qlik, che si sono rivelati in grado di rispondere più rapidamente alle esigenze del mercato.

Figura 5.1: Magic Quadrant 2016

Nonostante queste considerazioni, MicroStrategy fornisce soluzioni complete ed efficienti per gran parte delle applicazioni di business intelligence tra cui quella descritta nel presente lavoro.

Lo sviluppo dell'applicazione di business intelligence è stato infatti realizzato su MicroStrategy. La versione utilizzata è MicroStrategy 10, comprensiva di moltissime funzionalità dove l'utente può accedere a un'interfaccia di visualizzazione dei dati più fluida e rapida rispetto al passato, con la possibilità di aggiungere immagini, selezionare campi di testo, interagire con le dashboard e così via.

Fra le altre novità di Microstrategy 10, troviamo la presenza di un connettore diretto *Hadoop* e la possibilità di recuperare report decisionali che provengono da strumenti Bi di altri vendor, come *Sap Bo o Oracle Bi*. Il ventaglio di fonti supportati si allarga notevolmente, per includere database tradizionali o No-Sql, set di dati pubblici e anche dati provenienti da *Google Analytics, Drive, Dropbox, Facebook o Twitter*.

5.2.1 Visual Insight

Visual Insight è un componente MicroStrategy che utilizza la tecnologia *HTML 5* per creare facilmente dashboard e di analizzare le informazioni a partire da dati presenti nel database o prevenienti da fonti esterne, come Excel. Le dashboard sono progettate per essere immediate, flessibili e interattive, con particolare enfasi su:

- **Velocità**: abilità di cambiare istantaneamente dati in visualizzazioni grafiche
- **Visualizzazione**: fornire formati grafici avanzati per comunicare le informazioni in maniera più efficace e immediata.
- **Esplorazione**: possibilità di scoprire facilmente e velocemente anomalie e pattern nei dati.

Ogni dashboard è composta da uno o più dataset (insieme di dati/report di origine), dei layout (visualizzazioni), e dei filtri.

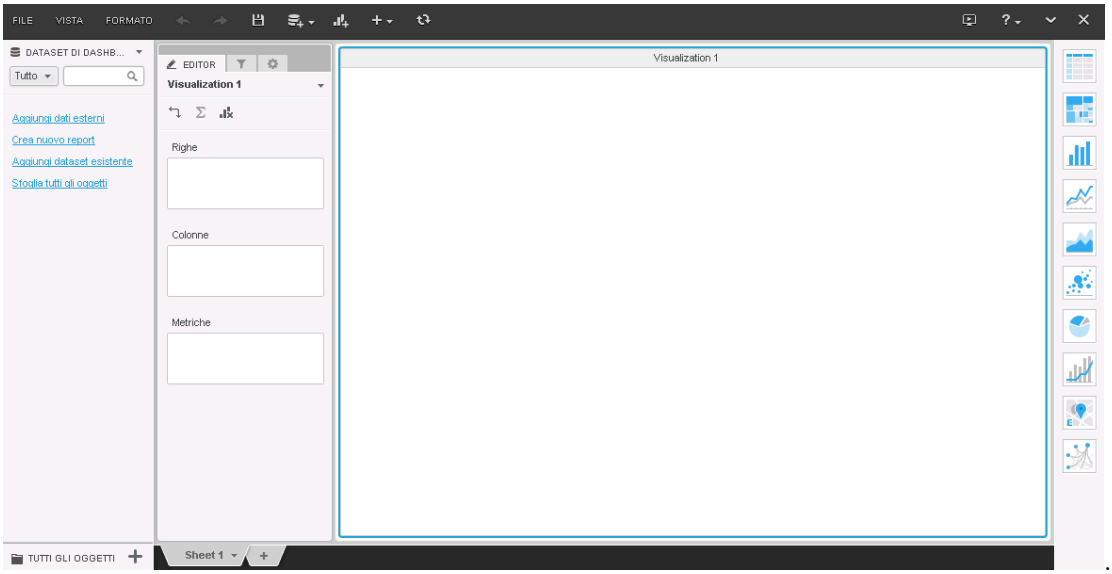

Figura 5.2: Visual Insight

Una dashboard è divisa in 5 pannelli o sezioni, posizionati nella zona sottostante la barra degli strumenti.

- *Pannello visualizzazioni*: Il pannello visualizzazioni contiene una o più visualizzazioni, utilizzate per mostrare i dati in tabelle, grafici o mappe. Si possono riorganizzare e spostare le visualizzazione semplicemente trascinandole con il mouse da un'altra parte. Visual Insight offre circa dieci tipi diversi di visualizzazione

Figura 5.3: Grafici Disponibili

- *Oggetti dataset*: Dopo aver selezionato una sorgente dati, gli oggetti del dataset scelto sono disponibili per l'uso nella dashboard. Questi oggetti sono mappati o come attributi o come metriche e possono essere spostati, manipolati e usati per crearne di nuovi. Gli attributi corrispondono ai livelli del business che si desidera analizzare. Invece, Le metriche sono calcoli che contengono i valori che si vuole visualizzare, per lo più aggregazioni di dati numerici.

- *Modifica visualizzazione*: Il pannello modifica visualizzazione controlla quali dati sono mostrati in una determinata visualizzazione. Applicando delle modifiche a questo pannello, la visualizzazione corrispondente viene aggiornata automaticamente.
- *Editor, Filtro, Proprietà*: Il pannello filtri permette la scelta di specifiche restrizioni da applicare ai dati. Il risultato di tale scelta si ripercuote poi nel pannello visualizzazioni. È possibile filtrare sia su attributi che su metriche, trascinando l'oggetto dataset desiderato nel pannello filtri. Per ogni filtro è possibile decidere se i valori selezionati sono tutti e soli quelli da includere o quelli da escludere dalle analisi. Inoltre è possibile scegliere lo stile di visualizzazione ed impostare le proprietà dell'attributo.
- *Raggruppamento a pagine*: Il pannello layout permette di spostarsi fra diversi layout della dashboard.

È possibile decidere quali pannelli mostrare o nascondere in base alle necessità degli utenti.

Capitolo 6

ESTRAZIONE, TRASFORMAZIONE, CARICAMENTO (ETL)

Si descrive il processo di estrazione, trasformazione e caricamento dei dati che include tutte le operazioni realizzate e le procedure implementate per integrare la sorgente di dati disponibile e caricare le informazioni richieste all'interno del data mart progettato.

6.1 Il processo ETL

Il processo *Extract-Transform-Load* (ETL) è fondamentale nei sistemi di data warehouse e consiste in un insieme di attività finalizzate a:

- estrarre i dati rilevanti da uno o più sistemi sorgenti
- trasformare i dati per migliorare la consistenza e la qualità
- rendere i dati utilizzabili e disponibili dalle applicazioni per le analisi decisionali.

Oltre a determinare la qualità dei dati per le analisi, il processo ETL utilizza una notevole quantità delle risorse di un progetto di data warehouse, soprattutto in termini di tempo. Infatti, nonostante il suo funzionamento sia

del tutto trasparente agli utenti finali, si stima che la sua implementazione e la sua manutenzione impieghino circa il 70% delle risorse totali di ogni progetto.

6.1.1 Le fasi del processo

La fase di estrazione consiste nella comunicazione con ciascuno dei sistemi sorgenti per la lettura e memorizzazione nel sistema di destinazione dei soli dati desiderati. Durante la fase di estrazione è di particolare importanza la conoscenza della qualità dei dati sorgenti, la quale consente di sapere se sono in grado di rispondere alle esigenze di business, ma soprattutto in quale modo essi sono capaci di fornire tale risposta; questo obiettivo può essere raggiunto attraverso una preventiva analisi detta *data-profiling*.

La trasformazione è la fase di riorganizzazione dei dati estratti. Le principali operazioni di trasformazione sono:

- la valorizzazione dei dati mancanti,
- la pulizia e correzione degli errori,
- l'integrazione dei dati provenienti da fonti disomogenee.

L'ultima fase del processo ETL è il caricamento dei dati nelle tabelle del data warehouse.

Il processo ETL può essere implementato attraverso diverse modalità che dipendono dalle analisi richieste e dalla complessità dei sistemi sorgenti: l'approccio utilizzato in questo progetto è quello a tre livelli (three-layer) mostrato in Figura 6.1.

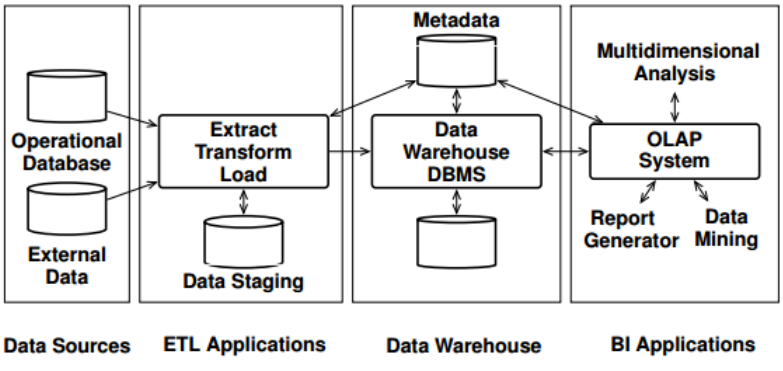

Figura 6.1: Architettura three-layer

Tale approccio prevede il livello delle sorgenti dati, il livello dei dati intermedi e il livello del data warehouse, separando il processo di estrazione dei dati dalla base di dati operazionale e il processo di caricamento nel data warehouse. Nel dettaglio, il flusso che interessa i tre livelli si articola come segue:

- i dati vengono letti dalla base dati operazionale e caricati in un'area di transizione del data warehouse, detta *staging area* o area temporanea,
- nella *staging area* avvengono le trasformazioni sui dati,
- i dati trasformati vengono caricati nell'area permanente del data warehouse.

Tutte le attività del processo sono realizzate per mezzo di procedure scritte in linguaggio Oracle PL/SQL e contenute nelle *stored procedure* del data warehouse. Si è preferito non far uso di un software ETL in quanto si disponeva già di un semi-lavorato sul quale sviluppare le procedure necessarie, secondo una metodologia consolidata per questo tipo di progetto.

6.1.2 Design della Staging Area

Il livello dei dati intermedi è stato implementato nella *staging area*, una parte di memoria che ha l'obiettivo di aggregare i dati provenienti da sorgenti diversi e di mantenerne le trasformazioni.

Come mostrato in Figura 6.2, la *staging area* è sviluppata in un database separato rispetto al database nel quale è implementata l'area permanente e ospita due tipologie di tabelle: le tabelle di estrazione, organizzate attraverso lo schema relazionale della base di dati operazionale, e le tabelle di trasformazione, sviluppate secondo lo schema relazionale dei data mart.

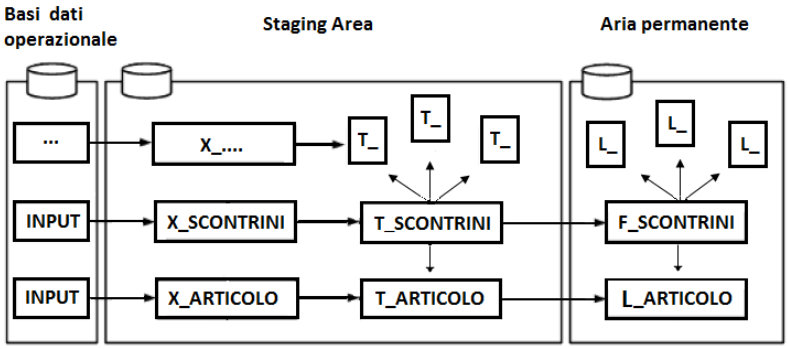

Figura 6.2: Staging Area

L'area dei dati intermedi non è un'area obbligatoria, però è fortemente consigliata qualora la fase di trasformazione sia complessa e/o interessi un elevato volume di dati; mantenere una *staging area* significa avere un punto di ripristino che vada ad evitare una nuova estrazione dei dati dai sistemi sorgenti. Questa operazione potrebbe presentarsi lenta qualora la fase di caricamento dovesse andare in errore per qualche causa.

6.1.3 Naming Convention

Per una maggiore manutenibilità, il nome di ogni oggetto del sistema adotta una naming convention che permette di identificarne il tipo e l'uso all'interno del flusso.

Tutti i nomi sono in maiuscolo e, se composti da più parole, esse sono separate dal carattere *underscore*. Le principali convenzioni utilizzate sono le seguenti:

Naming Convention

Prefisso	**Descrizione Oggetto**
X_	Tabella di estrazione in *staging area*
T_	Tabella di trasformazione in *staging area*
L_	Tabella di caricamento dimensione in area permanente
F_	Tabella di caricamento fatto in area permanente
ERR_	Tabella degli scarti

Inoltre le sequence, utilizzate per valorizzare le chiavi surrogate, hanno prefisso *SEQ* seguito dal nome della relativa tabella.

6.2 Organizzazione del flusso

Le procedure ETL sono organizzate in package, ciascuno dei quali implementa una fase del processo: CARICA_X per l'estrazione dei dati dal sistema sorgente, CARICA_T per la trasformazione e CARICA_L_F per il caricamento nelle tabelle dell'area permanente. L'ordine di esecuzione dell'intero flusso è composto da una procedura contenuta nel package CARICA_SA_DW.

Ogni procedura può scrivere su una sola tabella e questo consente la facile manutenibilità del codice in quanto:

- le cause di eventuali inconsistenze nei dati di una tabella, possono essere facilmente ricondotte ad una delle procedure del flusso abilitate alla modifica dei dati di quella tabella,
- in caso di errori in fase di esecuzione del caricamento/aggiornamento, le tabelle interessate sono rapidamente individuabili.

I nomi delle singole procedure hanno prefisso CARICA seguito dal nome della tabella sulla quale scrivono (ad es. CARICA_T_ARTICOLO è la procedura che scrive nella tabella della staging area T_ARTICOLO).

Il flusso appena descritto è mostrato in Figura 6.3 , dove si anticipano alcune caratteristiche delle fasi del processo.

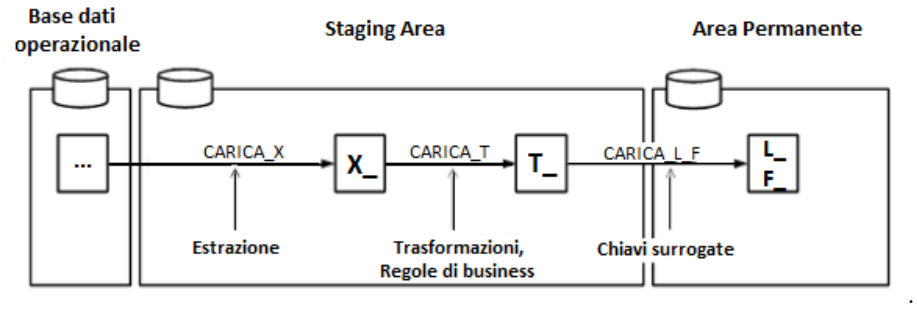

Figura 6.3: Flusso ETL

6.3 Estrazione

La fase di estrazione si effettua con le procedure contenute nel *package* CARICA_X e interessa le tabelle di estrazione X_. I dati necessari sono contenuti nelle tabelle di in un'unica base dati ospitata dal sistema: per accedervi è stato creato un *dblink* dal sistema Oracle 11gR2.

Tra gli approcci possibili per l'estrazione si è scelto quella completa rispetto a quella incrementale, data la ridotta dimensione dei volumi di dati esistenti e la semplicità di implementazione. Le tabelle X_ sono quindi alimentate in modalità *TRUNCATE/INSERT*, ossia ad ogni esecuzione del flusso vengono ripopolate interamente.

6.4 Trasformazione

La fase di trasformazione avviene con le procedure del *package* CARICA_T e interessa le tabelle di trasformazione T che hanno come nome e come nomi degli attributi le abbreviazioni dei relativi nomi di business. I dati presenti nelle tabelle X vengono modificati e caricati nelle tabelle T; le operazioni di trasformazione più comuni sono le seguenti:

- Sostituzione dei valori nulli: i codici di chiave esterna con valore *null* sono sostituiti dal valore 0.
- Eliminazione degli spazi: i campi delle tabelle della base dati operazionale hanno lunghezza fissa, per questo motivo le stringhe con caratteri inferiori rispetto alla lunghezza definita per il campo, vengono completate con caratteri di spazio; tali spazi possono provocare errori e di conseguenza sono rimossi sia in testa che in coda da tutti i campi di tipo stringa.
- Concatenazione dei campi: alcuni attributi nelle tabelle del data warehouse sono originati dalla concatenazione di più campi di una tabella della base dati operazionale.
- Conversione delle date: le date in formato giuliano sono trasformate nel formato giorno/mese/anno mediante funzione di conversione.
- Applicazione regole di business: i campi sono trasformati secondo le opportune regole di business e vengono aggiunti i campi delle misure calcolate di interesse.

6.5 Caricamento

Il caricamento è eseguito dalle procedure contenute nel package CARICA_L_F ed interessa sia le tabelle dimensionali L sia le tabelle dei fatti F, che hanno come nome e come nomi degli attributi le abbreviazioni dei relativi nomi di business. In questa fase non avviene nessuna trasformazione dei dati ma si valorizzano le chiavi surrogate e alcuni campi di controllo, identificabili dal prefisso W_.

Le procedure che intervengono sulle tabelle dimensionali L sono diverse da quelle che operano sulle tabelle dei fatti F per questo motivo saranno descritte separatamente; di seguito è riportata la struttura di una procedura CARICA_L:

```
MERGE INTO L_ARTICOLO l
USING (
        SELECT ...
        FROM T_ARTICOLO a
        LEFT JOIN L_ARTICOLO_CAT c ON (a.ARTICOLO_CAT_COD = c.ARTICOLO_CAT_COD)
) t
ON (l.ARTICOLO_COD = t.ARTICOLO_COD)

WHEN NOT MATCHED THEN
        INSERT VALUES (
        SEQ_NEXTVAL (' Seq_SCONTRINI '),
                ...
        )

WHEN MATCHED THEN
        UPDATE SET ...
        WHERE
                l.ARTICOLO_DES <> t.ARTICOLO_DES
                OR ...
;

COMMIT;
```

Le procedure di caricamento delle tabelle dimensionali vengono eseguite prima del caricamento delle tabelle dei fatti, in quanto è necessario che il vincolo di integrità referenziale tra la chiave primaria della tabella dimensionale e la chiave esterna della tabella dei fatti sia rispettato; motivo per il quale si deve rendere disponibile la chiave surrogata per le operazioni di lookup eseguite durante il caricamento delle tabelle dei fatti.

Il caricamento viene eseguito in modo incrementale con l'istruzione *MERGE*; si confronta il codice di ciascun record nella tabella di staging area con i codici dei record della tabella in area permanente:

- in caso di corrispondenza, il record è già presente nella tabella dimensionale e si verifica se anche gli altri campi sono invariati e, se quest'ultima ipotesi è falsa, si aggiornano mediante UPDATE;
- in caso di non corrispondenza, viene creata una chiave surrogata e viene inserito il nuovo record con la chiave appena creata.

Di seguito è riportata la struttura di una procedura CARICA_F_ :

```
-- DELETE
DELETE FROM F_SCONTRINI


-- INSERT
INSERT INTO F_SCONTRINI
SELECT ...
FROM T_SCONTRINI t
LEFT JOIN L_ARTICOLO v ON (v.ARTICOLO_COD = t.ARTICOLO_COD)
...

COMMIT;
```

Il caricamento viene eseguito in modo completo con la sequenza di istruzioni DELETE/INSERT. Il blocco INSERT inserisce i record; il blocco DELETE elimina la tabella dei fatti con i record.

Sia per quanto riguarda le dimensioni e sia nel caso dei fatti, la valorizzazione delle chiavi surrogate per gli attributi che sono chiave esterna avviene con una LEFT JOIN sulla chiave naturale tra la tabella in staging area e le tabelle in area permanente: nel caso in cui la LEFT JOIN restituisca il valore *null*, ossia il codice non sia presente nella tabella di *lookup*, la chiave surrogata viene impostata a 0. Questo impedisce che il riferimento errato possa generare errori, ma sia facilmente riconoscibile come un valore assegnato di default, anche in reportistica.

6.6 Aggiornamento e backup del data mart

Le operazioni di aggiornamento descritte di seguito sono implementate mediante l'oggetto Scheduler del database.

Il flusso ETL è programmato per l'esecuzione ogni giorno e ha una durata complessiva di circa 15 minuti. Inoltre viene eseguito giornalmente un backup del data warehouse con i dati, che sovrascrive quello del giorno precedente.

6.7 Memorizzazione e gestione degli errori

Si descrivono di seguito gli accorgimenti utilizzati per la memorizzazione dei dati e per la gestione degli errori.

6.7.1 Memorizzazione

Per la creazione delle tabelle sono stati scelti opportuni parametri di auto-espansione e percentuale di spazio libero riservata agli inserimenti. In base alla modalità di caricamento della tabella è stata scelta la percentuale di spazio libero per gli inserimenti:

- per le tabelle in TRUNCATE/INSERT si è scelto di non riservare spazio per gli inserimenti
- per le tabelle in MERGE si è assegnata la percentuale del 5%

Inoltre per le tabelle dei fatti la dimensione di auto-espansione è maggiore rispetto a quella delle tabelle dimensionali. Non sono state utilizzate strutture di memorizzazione e di accesso specifiche per data warehouse per questo motivo l'organizzazione fisica dei dati è di tipo seriale e non sono utilizzati indici particolari (ad es. indici bitmap).

Il sistema indicizza automaticamente tutte le chiavi primarie e tra l'altro sono stati creati gli indici sulle chiavi esterne per consentire al sistema di utilizzare l'operatore fisico NESTED LOOP in presenza di giunzioni, in particolare quelle con le tabelle dei fatti.

6.7.2 Gestione degli errori

Per tenere traccia degli esiti e dei tempi di esecuzione delle procedure ETL è stata predisposta una tabella di log dedicata, denominata W_ETL_LOG, che contiene le informazioni su:

- Numero sequenziale di esecuzione del flusso ETL
- Nome del package e dell'istruzione in esecuzione
- Data e ora di inizio esecuzione e durata dell'istruzione in ms
- Codice e messaggio di terminazione
- Numero di record elaborati

Inoltre è stata predisposta una procedura di invio e-mail con le notifiche sulla riuscita delle procedure programmate di caricamento.

Capitolo 7

REALIZZAZIONE DELL'APPLICAZIONE DI BUSINESS INTELLIGENCE E REPORTISTICA

In questo capitolo sono descritti l'applicazione e il modello di metadati utilizzati per lo sviluppo del livello di *business intelligence*. L'applicazione scelta per la realizzazione del livello di business intelligence è Microstrategy, un ambiente progettato allo scopo di soddisfare i sofisticati requisiti di business intelligence odierni. Esso fornisce funzioni integrate di query e reporting e analisi collaborative efficienti.

7.1 Architettura del sistema

L'architettura Microstrategy è schematizzata in Figura:

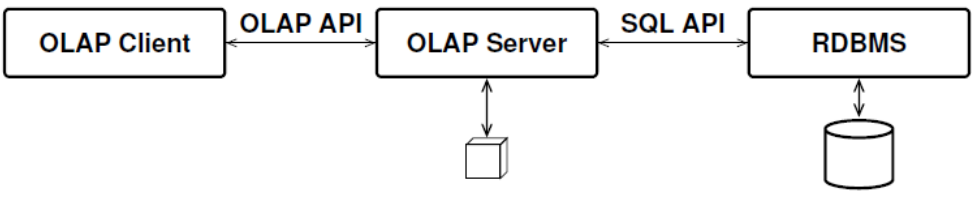

Figura 7.1: Architettura MicroStrategy

Il sistema si divide essenzialmente nelle seguenti componenti:

- **OLAP Client** l'architettura Microstrategy possiede due OLAP Client.
  - *Microstrategy Desktop*: offre un'interfaccia di tipo Windows e consente ai progettisti e agli amministratori di sistema di progettare, creare e mantenere un intero sistema di business intelligence tramite un'unica interfaccia.
  - *MicroStrategy Web*: è un'interfaccia Web che permette il controllo, il reporting e l'analisi del business. Grazie all'ampia gamma di opzioni di grafica e formattazione, con MicroStrategy Web è possibile creare report di tipo manageriale.
- **OLAP Server** in Microstrategy viene chiamato *Intelligence Server*. Il sistema fornisce una visione multidimensionale a cubo dei dati contenuti nel Data Server, inoltre contiene un motore analitico per il calcolo delle funzioni analitiche.
- **RDBMS** tra i tanti Data Server supportati abbiamo: *IBM DB2*, *Microsoft SQL Server*, *Oracle* e *Teradata*. Per questi RDBMS più avanzati lo strumento offre funzionalità di ottimizzazione delle query che consentono di generare codice SQL che sfrutta le funzionalità avanzate del particolare sistema di gestione dei dati con il quale interagisce.

Più dettagliata la Figura 7.2. Dove sono mostrati le componenti principali e i vari tipi di connessione per interfacciarsi (*ODBC, TCP/IP, HTTP*).

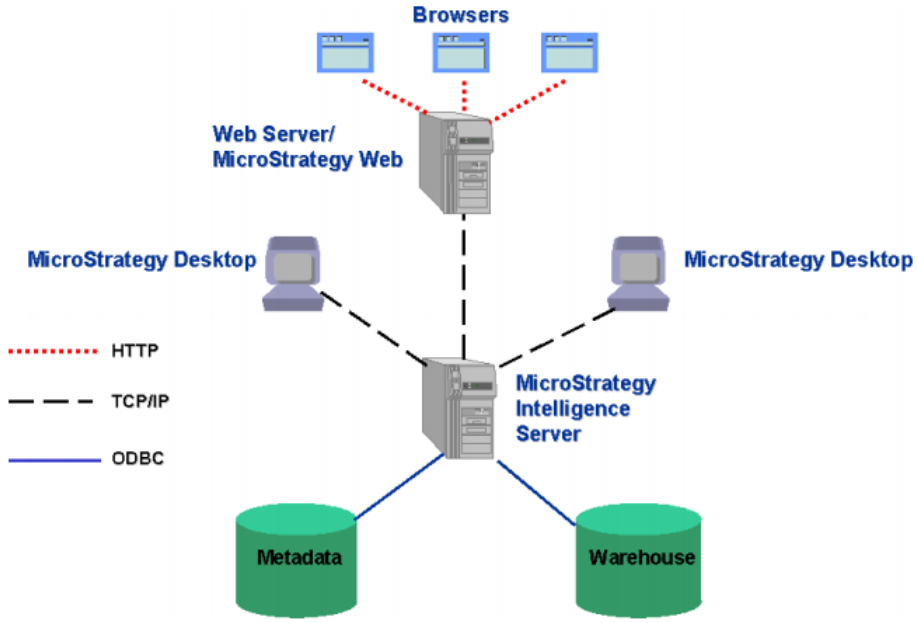

Figura 7.2: Architettura MicroStrategy

Altri prodotti utilizzati per la realizzazione del progetto sono:

- **Administrator**: è la componente che fornisce agli amministratori di sistema gli strumenti per gestire, monitorare e automatizzare l'infrastruttura del progetto, attraverso un monitoraggio centralizzato della gestione sia di utenti e oggetti sia delle prestazioni
- **Microstrategy Architect**: è lo strumento che mappa la struttura fisica della base dati in un modello di business logico. Le mappature sono memorizzate in un archivio centralizzato

7.2 I metadati

I metadati MicroStrategy sono archiviati in un repository centralizzato e la caratteristica principale dei metadati MicroStrategy è che sono memorizzati in un database relazionale standard. Questo approccio permette di aumentare la scalabilità e la facilità di gestione delle applicazioni BI.

I metadati memorizzano al loro interno lo schema fisico del data warehouse che rappresenta uno strumento indispensabile a mappare le informazioni in esso contenute. I metadati caricano tre tipi di oggetti:

- *Oggetti schema*: creati da un progettista e comprendono elementi come fatti, attributi e gerarchie
- *Oggetti applicazione*: sono utilizzati per creare report. Gli oggetti applicazione vengono generalmente creati da un progettista di report e sono basati sugli oggetti schema. Includono report, filtri, metriche ecc
- *Oggetti configurazione*: gestiti dall'amministratore, cambiando le configurazioni relative a utenti, gruppi di utenti e sicurezza.

7.3 Selezione degli oggetti del Data Warehouse

La selezione degli oggetti del data warehouse serve per indicare allo strumento quali sono le relazioni tra le tabelle e quali verranno utilizzate nei report. Di base lo strumento utilizza due categorie di oggetti: metriche e attributi. Le metriche sono di fatto le misure delle tabelle dei fatti e gli attributi sono degli oggetti utilizzati nell'ambiente di sviluppo MicroStrategy utilizzati per referenziare gli attributi dimensionali del data warehouse.

Il processo di selezione degli oggetti è importante, in funzione delle direttive date in questa fase lo strumento compone le interrogazioni sul data warehouse. Di seguito è mostrato il processo di selezione degli oggetti mediante un wizard incorporato nello strumento. Occorre procedere nel seguente modo:

- Indicare quali tabelle saranno utilizzate nei report

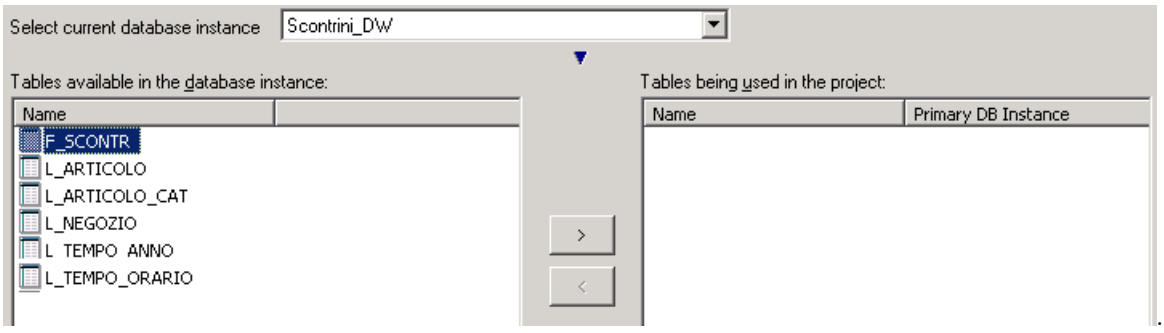

Figura 7.3: Warehouse Catalog Wizard

- Selezionare le misure che saranno utilizzate

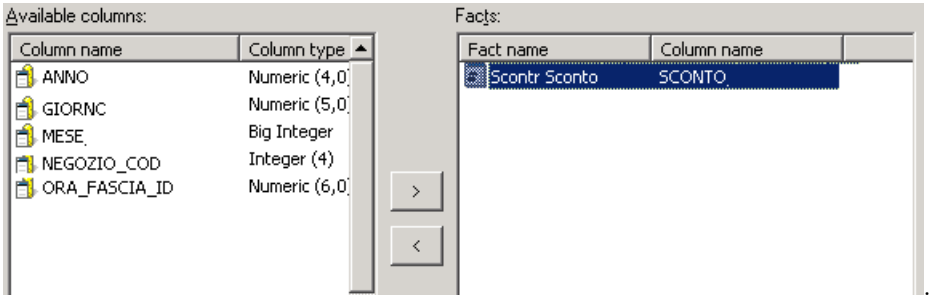

Figura 7.4: Fact Creation Wizard

- Indicare per ogni attributo:
  - il campo chiave esterna della tabella dei fatti che referenzia la chiave primaria della tabella contenente l'attributo dimensionale
  - il campo chiave primaria della tabella dimensionale contenente l'attributo
  - il campo descrittivo dell'attributo dimensionale da utilizzare per la visualizzazione dell'attributo nei report

Terminato il procedimento di selezione degli oggetti del data warehouse lo strumento crea gli attributi.

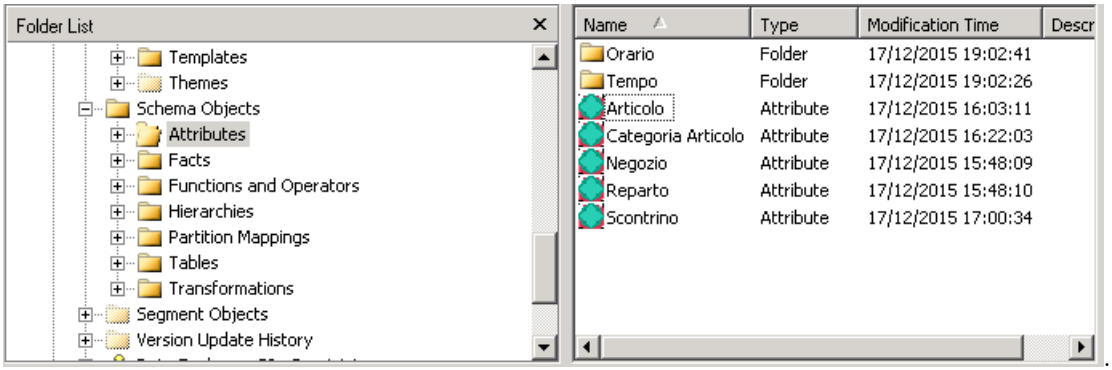

Figura 7.5: Attributi

7.3.1 Creazione delle metriche

Partendo dalla selezione delle misure fatta nella sezione precedente è possibile costruire metriche che utilizzano le misure per eseguire calcoli più complessi.

Le metriche possono essere classificate nei seguenti modi:

- *Metriche semplici*: devono contenere una formula per determinare i calcoli da eseguire sui dati. Un esempio è: *Sum(Scontrino)*.
- *Metriche composte*: vengono derivate e calcolate a partire dalle metriche semplici.

In Figura 7.6 è mostrata l'interfaccia grafica di MicroStrategy per la definizione delle metriche, in particolare della metrica composta Sconto Medio.

$$Sconto\ Medio = SUM(Sconto)\ /\ SUM(Scontrini\ Nr.)$$

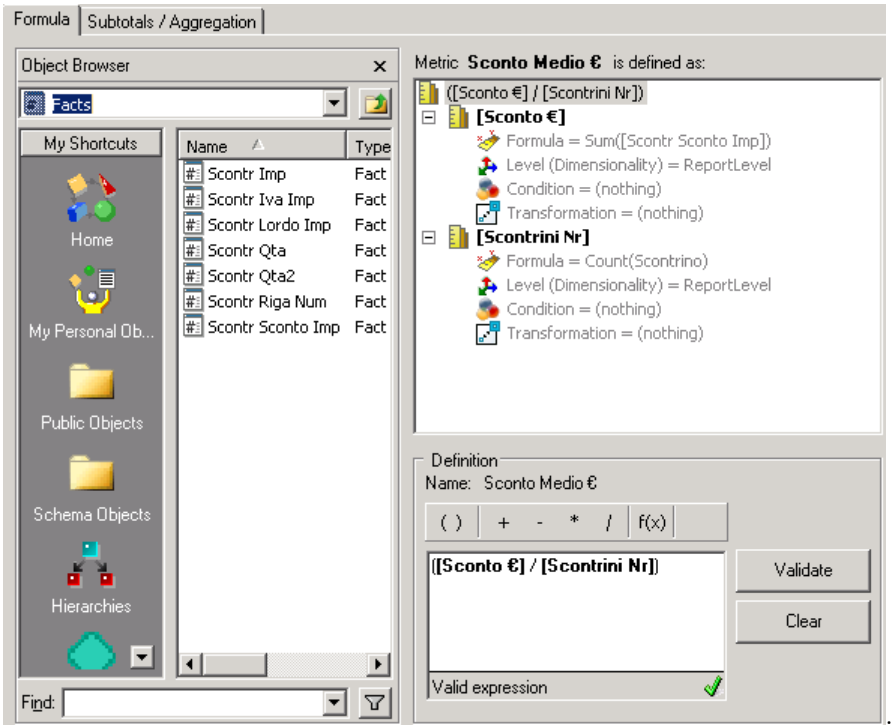

Figura 7.6: Metriche

Da notare che essendo un rapporto di misure quando vengono calcolate le sommatorie totali questa metrica potrebbe non essere calcolata in modo corretto dal momento che il totale calcolato come una somma di rapporti porterebbe ad un valore sbagliato. Il volare corretto deve essere calcolato come un rapporto di somme. Lo strumento permette di settare mediante un'apposita opzione se per i diversi livelli di aggregazione occorre eseguire la somma dei rapporti oppure il rapporto di somme.

Due proprietà fondamentali per le metriche sono:

- *Condizione*: si può pensare alla condizionalità come a un filtro di metrica che è indipendente dai filtri su qualsiasi report la metrica sia utilizzata. Una metrica condizionale consente di applicare un filtro a una sola metrica su un report pur non influenzando le altre metriche.
- *Trasformazione*: le trasformazioni temporali vengono utilizzate nelle metriche per confrontare i valori in momenti diversi, ad esempio anno in corso/anno precedente o data corrente/mese-a-oggi. MicroStrategy dispone di numerose trasformazioni precostituite, anche se è possibile crearne di personalizzate in base alle necessità. Nel nostro caso sono state create le misure *Importo AP* e *Quantità AP* che rappresentano l'importo e la quantità dell'anno precedente.

7.4 Report

I report consentono agli utenti di raccogliere informazioni di business attraverso l'analisi dei dati. I report sono costituiti da: attributi e fatti derivanti dal warehouse, filtri che determinano quali dati vengono visualizzati nel report e metriche per l'esecuzione di calcoli sui fatti.

In Microstrategy è possibile visualizzare un report in diverse prospettive, a seconda del tipo di lavoro da svolgere:

- *Report tabella*: rappresenta il tipo più utilizzato di report. La vista tabella offre una vista formattata e tabulare dei dati del report

Anno	Metrics / Negozio	Scontrini €					
2013							
2014							
2015							
Total							

Figura 7.7: Report tabella

- *Report grafico*: utilizza una rappresentazione dei dati in formato visivo. È possibile scegliere tra diversi tipi di grafici, al fine di visualizzare i dati nella maniera più efficiente possibile.

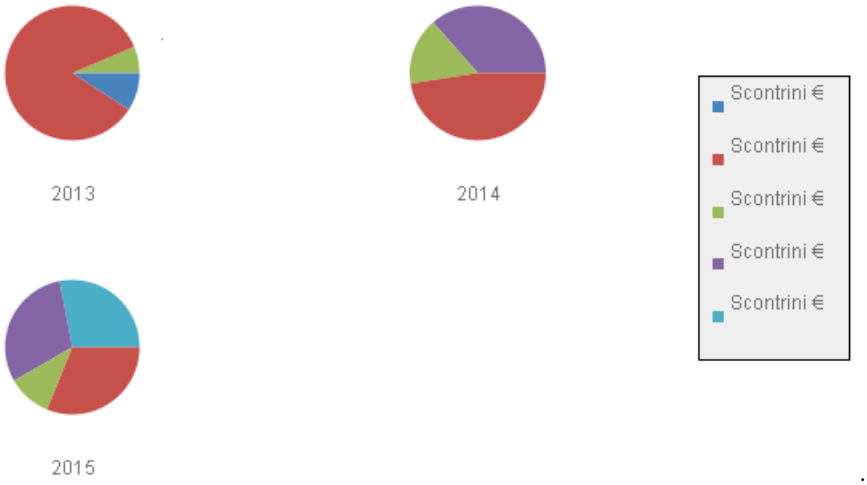

Figura 7.8: Report grafico

- *Report tabella/grafico*: rappresenta una visualizzazione combinata della vista tabella e di quella grafico

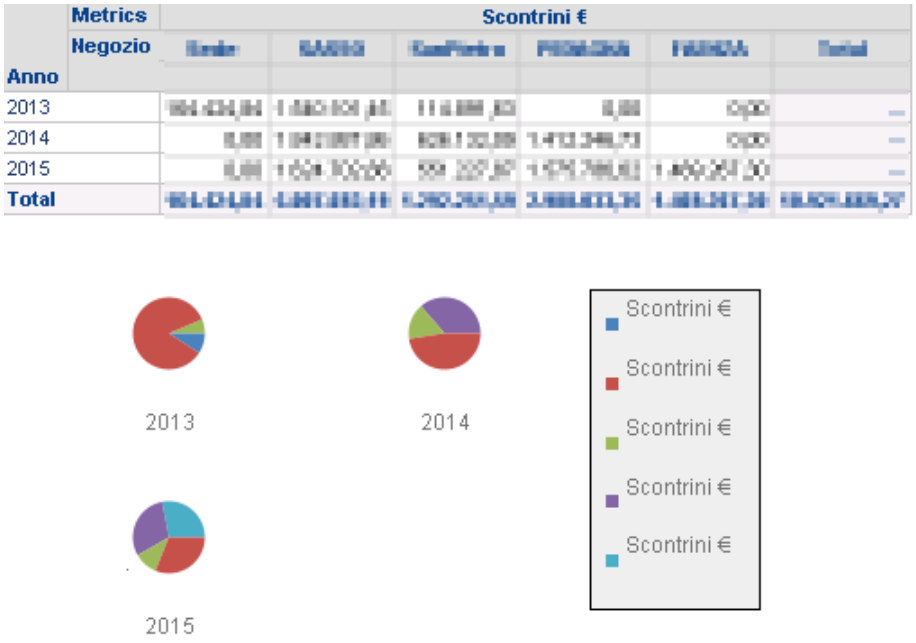

Figura 7.9: Report tabella/grafico

- *Report SQL*: visualizza l'istruzione SQL utilizzata per generare il report. Questo tipo di visualizzazione rappresenta un modo efficace per rilevare e risolvere gli eventuali problemi.

7.5 Dashboard

Utilizzare delle ottime visualizzazioni per poter esporre le informazioni celate dietro ai diversi sistemi informativi può risultare fondamentale per migliorare i processi decisionali, riducendo i tempi di acquisizione e comprensione dell'informazione. Grandi quantità di informazioni disponibili, sviluppano sensazioni di ansia e di maggiorazione cognitiva, in particolare per gli utenti che occasionalmente si affacciano a tale esperienza risultando così inesperti. Le rappresentazioni visuali offrono la possibilità di presentare le informazioni spostando direttamente il carico di lavoro, riferito sia alla sfera del sistema cognitivo dell'utente sia alla sfera del sistema percettivo.

L'uomo, in possesso di notevoli capacità percettive, dimostra di riconoscere

e richiamare immagini rapidamente, di rilevare variazioni in dimensioni, colore, forma e movimento. La presentazione dei dati in modo grafico permette all'utente di poter individuare nuove e utili proprietà, le loro connessioni e i possibili margini dai valori attesi. L'utilizzo del colore può far emergere aggregazioni presenti fra i dati, mentre l'uso di animazione può dare la capacità di districarsi rapidamente tra livelli di dettagli concatenati. I mezzi percepiti visivamente, creano le basi per fornire inevitabilmente un orientamento, un contesto, utile a descrivere dettagliatamente selezioni di regioni dello schermo che producono un feedback dinamico che si dimostra da ausilio per l'utente nell'identificare e monitorare i cambiamenti. [9]

Parlando di dashboard, questi assumono le vesti di documenti di sintesi comprendenti un insieme di report precedentemente realizzati, immagazzinando in un solo colpo d'occhio informazioni immediate sulle prestazioni di tutto il contesto aziendale. Le informazioni risultanti sono sfruttate dai manager e dai quadri che hanno bisogno di una visione generale delle performance di business, traendo così vantaggi enormi dalla visualizzazione tempestiva e immediatamente comprensiva di dati strategici sia dal punto di vista del tipo finanziario sia del tipo operativo.

I dashboard sono progettati per conseguire il massimo impatto visivo in una piattaforma, ideata per la comprensione rapida che sfrutta una combinazione di tabelle, grafici e altri indicatori. Attraverso i suddetti dashboard si giunge a una visione immediata della performance resasi attuale rispetto ai trend vigenti.

In questo lavoro di tesi è descritta e presentata mediante *screenshot* una dashboard realizzata per gli utenti del sistema di business intelligence; per ragioni di riservatezza i dati in essi contenuti sono stati mascherati.

La dashboard realizzata è stata sviluppata per il processo degli scontrini attraverso il tool *Microstrategy Visual Insight*, che consente di creare rapidamente e progettare un'analisi di esplorazione visiva con un display interattivo che può essere utilizzato per esplorare i dati aziendali. È possibile utilizzare i dati da un report esistente, eseguire manipolazioni sui dati per personalizzare le informazioni e aggiungere rappresentazioni visive, chiamate *visualizzazioni*, per rendere i dati di facile interpretazione. In particolare sono state creati cinque

layout diversi con differenti visualizzazioni dei dati:

- *Tempo*: visualizzazioni basate sull'andamento degli scontrini su un arco temporale (es. mese, settimana)
- *Giorno Settimana*: pagina contenente informazioni, in maniera distinta, dei giorni della settimana
- *Orario*: visualizzazioni centrate sull'andamento degli scontrini nelle diverse fasce orarie
- *Categoria*: layout sullo sviluppo delle diverse categorie merceologiche
- *Articolo*: pagina contenente informazioni sui singoli articoli delle categorie merceologiche

Questi layout sono aggregati in relazione a diverse dimensioni, presentando alcune caratteristiche e funzionalità comuni:

- si confrontano, a livello tabellare e grafico, le situazioni rilevate in periodi differenti;
- è possibile impostare dei filtri predefiniti per determinati attributi dimensionali;
- è possibile scegliere se visualizzare gli ultimi mesi consecutivi rispetto al periodo prescelto;
- è offerta la possibilità di approfondire il risultato ottenuto attraverso il drill-down lungo un'ulteriore dimensione a scelta e di ottenere il dettaglio sulle posizioni coinvolte.

7.5.1 Tempo

Fornisce una visione delle prestazioni dal punto di vista settimanale e mensile mostrando:

- gli incassi dei mesi selezionati con la variazione rispetto all'anno precedente per ogni singolo negozio;

- l'andamento degli incassi e dei numeri di scontrini emessi dal punto di vista grafico;
- l'andamento degli incassi confrontandoli rispetto all'anno precedente.

Per creare questa pagina, Figura 7.10, sono stati utilizzati una tabella e due grafici a linee, dove sulle righe è presente il periodo temporale (settimane, giorni) oppure i negozi e sulle colonne le metriche da analizzare. Selezionando il nome di un negozio oppure un periodo temporale è possibile vedere il dettaglio dell'andamento delle prestazioni delle applicazioni per quel negozio e/o periodo, così da ottenere il dettaglio sulle posizioni coinvolte.

Dalle varie visualizzazioni si nota che, come nel mese di gennaio, l'incasso sia più elevato rispetto a quello dell'anno precedente. Inoltre nella maggior parte dei casi il numero di scontrini rispetto all'incasso risulta avere un andamento lineare.

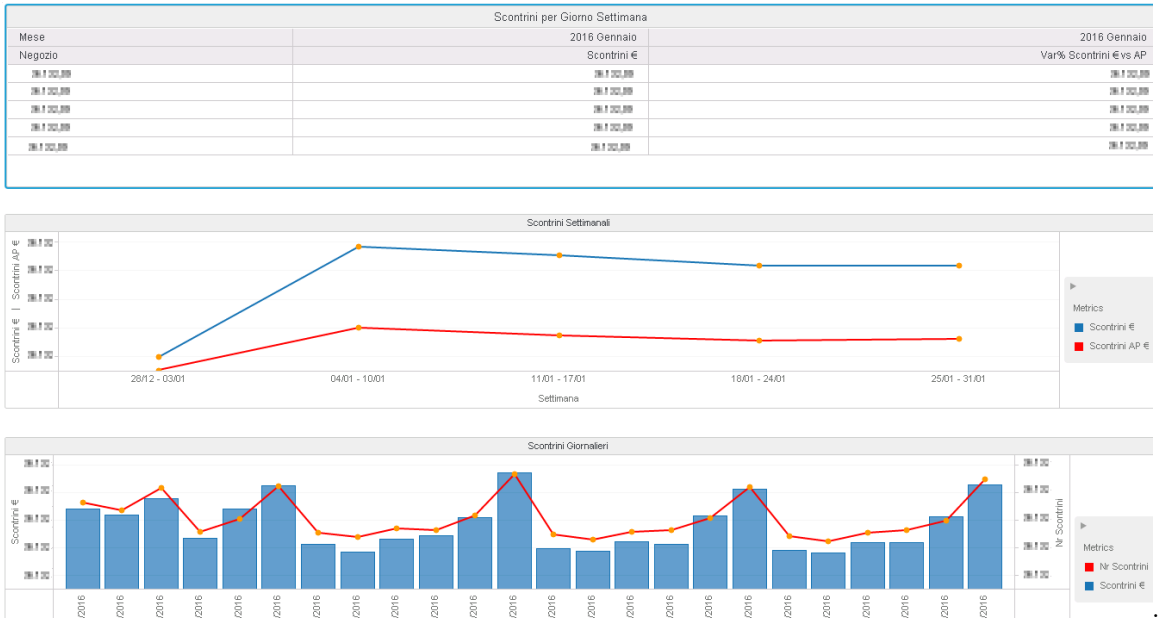

Figura 7.10: Layout Tempo

7.5.2 Giorno Settimana

Fornisce una visione delle prestazioni dal punto di vista giornaliero e si divide in quattro visualizzazioni rappresentate nella Figura 7.11:

- La tabella in alto a sinistra rappresenta lo scontrino medio giornaliero per negozio. Dalla tabella si evince come nel fine settimana lo scontrino medio sia più alto rispetto ad inizio settimana (lunedì, martedì), questo vale per tutti i negozi. Da ciò emerge che i clienti preferiscono fare le provviste settimanali nel week-end acquistando nel resto della settimana solo i prodotti utili alle situazioni di emergenza giornaliera.

- In alto a destra sono rappresentati due grafici a linea, il primo in base all'incasso medio e il secondo in base allo scontrino medio. Si nota come l'incasso medio sia crescente per tutti i negozi invece lo scontrino medio rimane stabile nei primi giorni della settimana ed aumenta nel fine settimana.

- Nel grafico a barre e a linea sono stati messi a confronto il numero di scontrini medio per giorno e l'incasso medio, ordinando in maniera decrescente quest'ultimo. Si vede, per esempio, che il venerdì il numero di scontrini è simile rispetto ai giorni di inizio settimana ma l'incasso medio è decisamente più elevato perché si acquistano molti più prodotti per singolo scontrino.

- Nel grafico a bolle (in basso a destra) si ha sull'asse delle ordinate lo scontrino medio, invece sull'asse delle ascisse il numero di scontrini medio per giorno. In questa visualizzazione si nota come, per esempio, il sabato il numero di scontrini è elevato ed ha un alto valore di scontrino medio invece il venerdì il numero di scontrino diminuisce ma lo scontrino medio aumenta.

  In un grafico a bolle le entità rappresentate possono essere confrontate tra loro in base alla loro dimensione e alla loro posizione rispetto agli assi numerici. Infatti gli assi X e Y di un grafico a bolle sono scale numeriche, quindi la posizione in cui viene rappresentato un dato descrive due valori numerici, mentre l'area del disegno dipende dal valore di un terzo parametro.

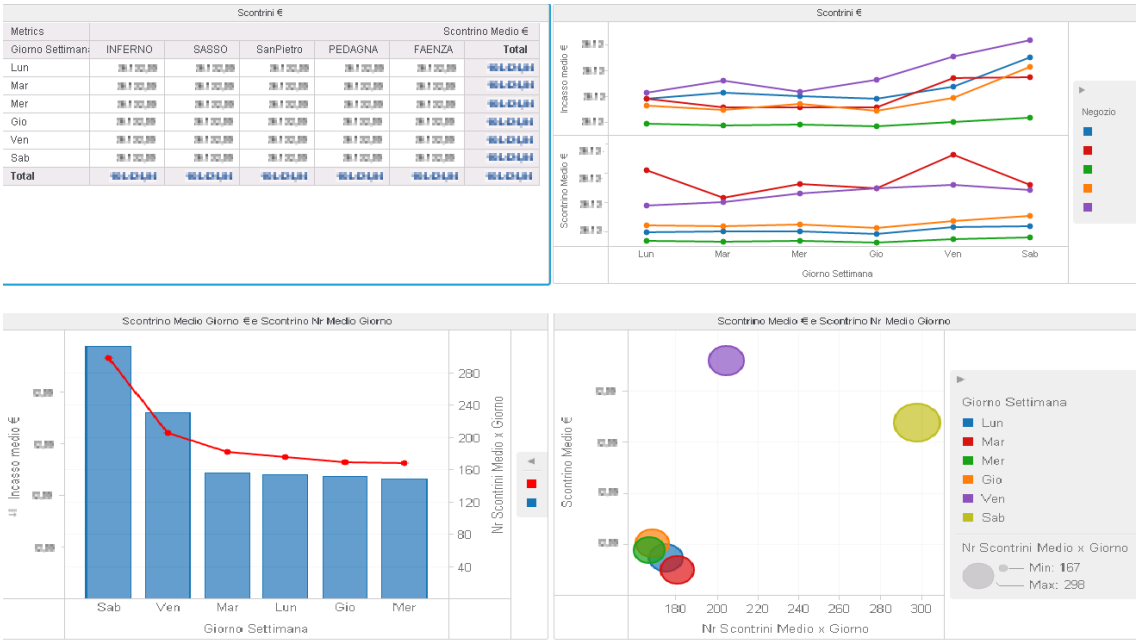

Figura 7.11: Layout Giorno

7.5.3 Orario

La pagina Orario, Figura 7.12, mostra il dettaglio delle fasce orarie del giorno selezionato, mostrando sia sotto forma tabellare sia sotto forma grafica l'andamento degli scontrini. Oltre a fornire informazioni generali presentate in forma tabellare, queste offrono anche delle viste di sintesi sotto forma di grafici allo scopo di rappresentare l'informazione in modo sintetico e immediato. La pagina si divide in tre visualizzazioni.

- Gli incassi per negozi e il totale complessivo in forma tabellare senza vista di sintesi;
- Gli incassi per negozi e per fascia oraria, rappresentati attraverso un grafico a linea (in alto a destra). Si osserva come l'andamento delle vendite raggiunge il picco in tutti i negozi nelle ore di tarda mattinata (10 e 11) e nelle ore di primo pomeriggio (16, 17) al contrario di quanto accade nelle

rimanenti fasce orarie dove l'afflusso di clienti si dimostra meno consistente rispetto a quanto accade nelle ore di maggiore concentramento;

- Si raffigura l'incasso medio con le barre e il numero di scontrino medio per giorno con la linea. Un'informazione che salta subito all'occhio è che durante le ore 16 e 17 abbiamo un numero di scontrini medio nella norma ma l'incasso è nettamente superiore, invece durante la mattinata le due misure seguono un andamento lineare. Da ciò emerge che i clienti preferiscono fare una spesa comprensiva di un maggior numero di prodotti utili al fabbisogno settimanale nelle ore pomeridiane.

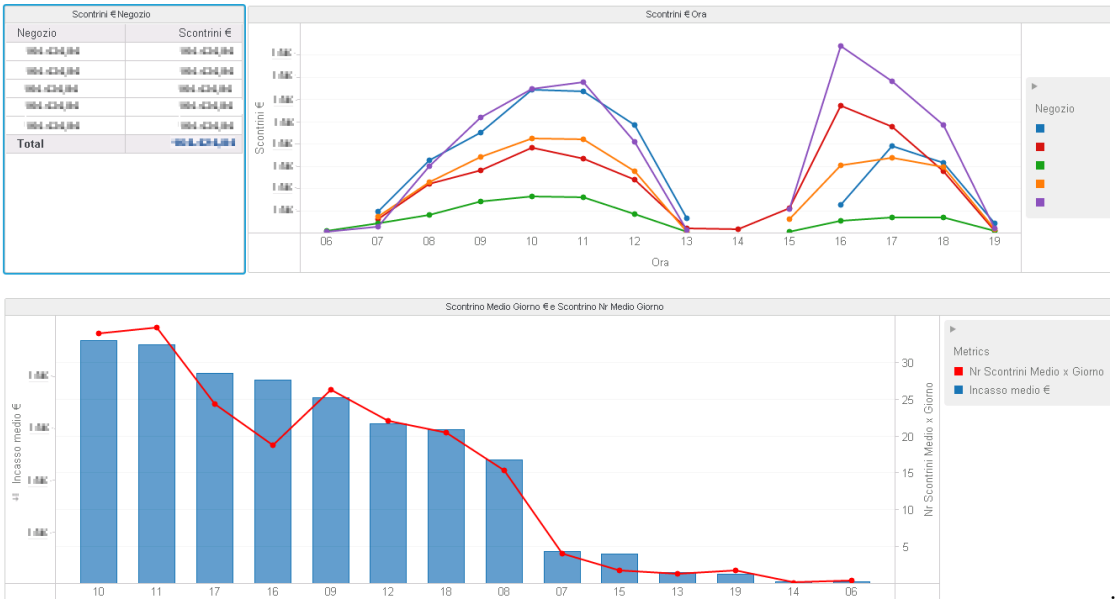

Figura 7.12: Layout Ora

7.5.4 Categoria

Fornisce una visione delle prestazioni dal punto di vista delle categorie merceologiche che si dividono in due visualizzazioni.

- L'andamento delle vendite delle singole categorie merceologiche in un determinato periodo confrontato con l'incasso dell'anno precedente;
- L'andamento delle vendite delle singole categorie merceologiche per negozio.

Dal grafico, Figura 7.13, emerge una diversa distribuzione degli incassi per categoria. In alcuni settori essa risulta fortemente influenzata dalla collocazione del punto vendita. È il caso, per esempio, della categoria "Gastronomia", dall'analisi da cui si evince una maggiore vendita in prossimità dei centri storici piuttosto che nelle periferie.

Questa tendenza può essere attribuita a diversi fattori. I punti vendita nel centro storico presentano una posizione strategica per gli impiegati degli uffici adiacenti che possono farvi capo sia nella pausa pranzo che al termine della giornata lavorativa. Per gli anziani che, facilitati dalla vicinanza da casa, possono recarsi a piedi all'occorrenza oppure per gli studenti di passaggio. Invece altre categorie merceologiche hanno una distribuzione delle vendite simile per tutti i negozi, per esempio, prosciutti, altri salumi, avicoli, vini e uova.

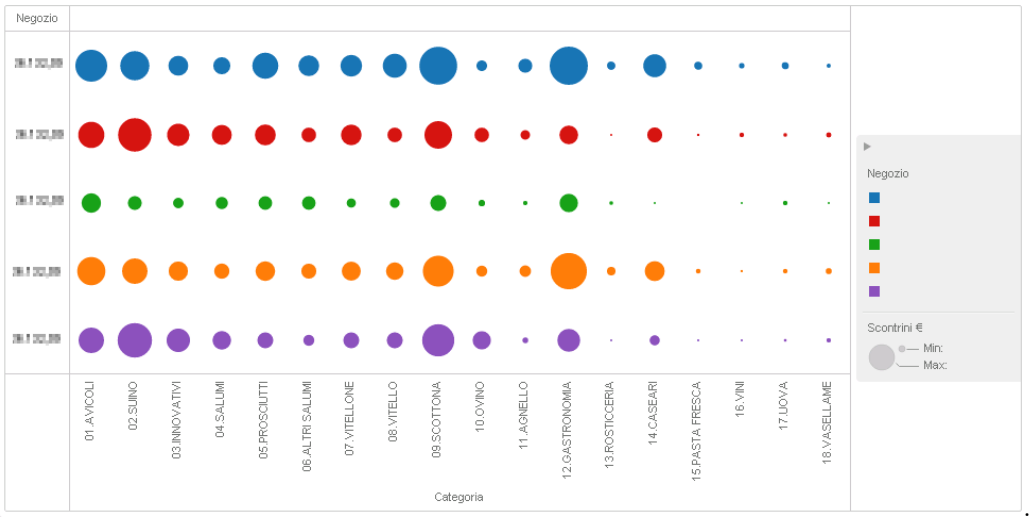

Figura 7.13: Layout Categoria

Invece dalla seconda visualizzazione, Figura 7.14, si nota che l'incremento degli incassi nella maggior parte delle categorie presenta un picco rispetto all'anno precedente.

Le ragioni di tale crescita risiedono probabilmente in una politica aziendale fortemente orientata al cliente e alle sue esigenze. Campagne promozionali e offerte stagionali hanno reso appetibili i prodotti per nuovi clienti. La ricollocazione dei prodotti sugli scaffali ed un attento marketing, hanno invogliato indubbiamente agli acquisti. Le offerte hanno contribuito invece alla fidelizza-

zione del cliente, soddisfatto dall'alto grado di qualità del prodotto. Invece altre tipologie di categorie, vini, uova, pasta fresca e vasellame hanno mantenuto lo stesso andamento, sicuramente, dovuto al poco consumo.

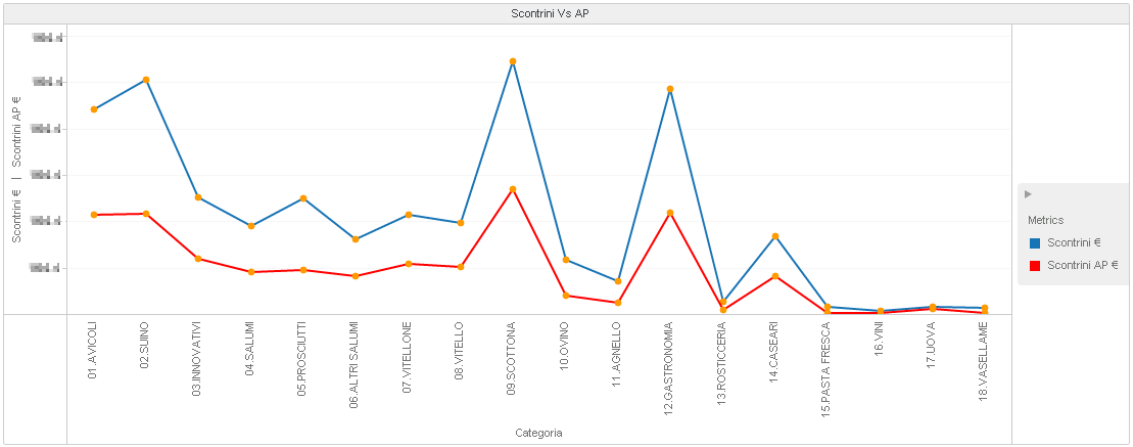

Figura 7.14: Layout Categoria

7.5.5 Articolo

Nel layout articolo è presente un raggruppamento per categorie che permette di scegliere la categoria da monitorare, mostrando il dettaglio degli articoli che ne fanno parte e indicando l'andamento delle vendite in un determinato arco temporale. Inoltre è presente un filtro che consente di impostare il range di articoli da selezionare in base alle vendite riscontrate.

Una delle due visualizzazioni contenute nella pagina rappresenta i migliori articoli per incasso filtrati per categoria. La seconda visualizzazione mostra in maniera tabellare l'elenco degli articoli venduti e le rispettive quantità della categoria selezionata. Selezionando il nome del negozio oppure un periodo temporale è possibile vedere il dettaglio dell'andamento degli articoli per quel negozio e/o periodo.

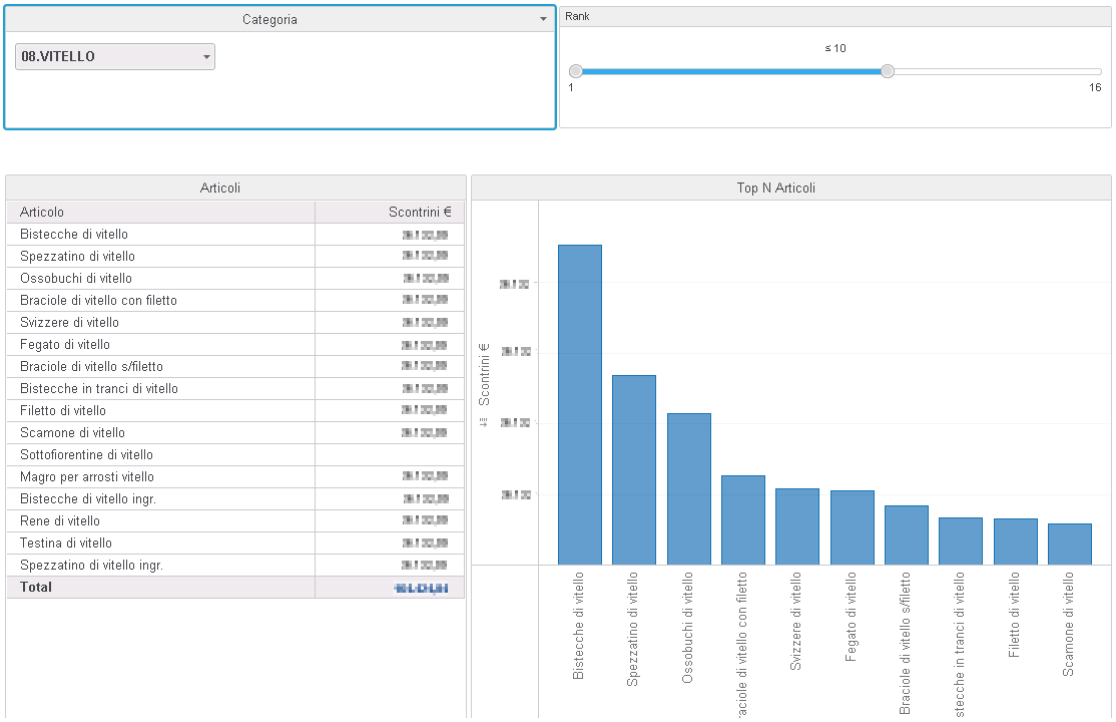

Figura 7.15: Layout Articolo

Capitolo 8

DATA MINING: REGOLE ASSOCIATIVE

In questo capitolo si descrive una introduzione sulla disciplina del data mining con le relative tipologie e il processo di creazione di un modello predittivo in riferimento al *Market Basket Analysis* sulla componente di Microstrategy (Data Mining Services).

8.1 Introduzione al Data Mining

Il data mining, letteralmente estrazione di dati, è un ambito dell'informatica il cui fine è di ricercare informazioni utili all'interno di grandi collezioni di dati, informazioni che nonostante ciò resterebbero non visibili. Negli ultimi anni il settore è in una fase di ampio sviluppo soprattutto dovuto incremento delle banche dati disponibili e all'interesse delle aziende che stanno prendendo in considerazione le potenzialità e i risultati che si ottengono con l'ausilio di tale disciplina.

Gli impieghi del data mining sono generalmente suddivisi in due grandi categorie:

- *Uso predittivo*: l'obiettivo è di predire un particolare attributo (funzione obiettivo) a partire da attributi conosciuti (predittori)
- *Uso descrittivo*: l'obiettivo è di individuare schemi ricorrenti (pattern frequenti), gruppi di dati simili (cluster), anomalie o pattern sequenziali che caratterizzino i dati analizzati. In seguito alla fase di mining è neces-

sario utilizzare tecniche di post-processing che consentono di validare e visualizzare i risultati acquisiti.

Molti sono gli elementi che hanno contribuito allo sviluppo del data mining. Tra questi:

- *Elevata dimensionalità*: il numero di attributi che caratterizzano un record è dell'ordine delle migliaia. Le tecniche di analisi dei dati tradizionali sono state sviluppate per dati con bassa dimensionalità e gli algoritmi hanno generalmente una complessità che cresce rapidamente con il crescere della dimensione. Il data mining offre algoritmi sviluppati per gestire in modo efficiente una notevole dimensionalità.

- *Differenza dalla statistica*: l'analisi di tipo statistico è caratterizzata da un metodo basato su ipotesi e test, quindi vengono proposte alcune ipotesi sui dati e per ognuna di queste si crea un test per verificarne la veridicità. Tale processo ha tempi di sviluppo molto lunghi in quanto nella maggior parte dei casi bisogna valutare migliaia di ipotesi. Il data mining permette di automatizzare il processo di generazione delle ipotesi e della loro valutazione.

- *Dati eterogenei e complessi*: le analisi dei dati tradizionali sono spesso individuate per lavorare su insiemi di dati con attributi dello stesso tipo, continui o categorici. Per le esigenze del data mining sono state sviluppate tecniche capaci di trattare adeguatamente dati con attributi eterogenei. Questo vale anche per l'analisi di strutture dati complesse come ad esempio ipertesto (es. pagine web) o dati relativi a serie temporali su diverse misure (es. clima).

- *Scalabilità*: molti dei dataset da analizzare hanno dimensioni molto elevate (nell'ordine dei gigabyte o terabyte). Per questo motivo sono stati studiati algoritmi con architetture di tipo parallelo e distribuito.

8.1.1 Tipologie di data mining

Il data mining offre differenti tipologie di analisi. Di seguito ne vengono descritte le principali con alcuni possibili esempi di utilizzo.

- *Analisi predittiva*: Questa analisi ha come obiettivo quello di costruire un modello predittivo partendo da un insieme di attributi conosciuti. Sono

presenti due differenti tecniche: la classificazione, usata per prevedere il valore di variabili discrete e la regressione, usata per variabili di tipo continuo. I modelli predittivi possono ad esempio essere usati per identificare utenti che stanno per abbandonare una determinata compagnia.

- *Analisi associativa*: Questa analisi viene utilizzata per identificare pattern frequenti che descrivono particolari caratteristiche dei dati. I pattern individuati sono generalmente mostrati sotto forma di regole associative. L'obiettivo è di trovare in modo efficiente i pattern più interessanti. Esempi sono l'analisi delle transazioni in un supermercato (market basket analysis). Quando nei record, invece, è presente una collocazione temporale è possibile utilizzare un'analisi per la ricerca di pattern sequenziali. Questi sono generalmente utilizzati per prevedere il verificarsi di eventi futuri, conoscendo l'ordine temporale di eventi già avvenuti. Esempi di applicazione possono essere lo studio dei comportamenti d'acquisto della clientela di una attività commerciale.

- *Analisi basata su cluster*: Questa analisi vuole identificare nei dati un certo numero di gruppi, chiamati cluster, in cui i dati siano molto simili all'interno dello stesso gruppo e significativamente differenti tra cluster diversi. Una possibile applicazione è per la segmentazione dei clienti al fine della profilazione.

- *Analisi delle anomalie*: Questa analisi si occupa di individuare piccoli gruppi di dati le cui caratteristiche sono significativamente differenti dagli altri. Usi tipici dell'analisi delle anomalie sono il riconoscimento di frodi.

8.1.2 Le fasi dell'estrazione della conoscenza

Il data mining è solo una delle componenti dell'intero processo di estrazione della conoscenza dai dati. In questo paragrafo vogliamo dare un'idea delle fasi che compongono l'intero processo.

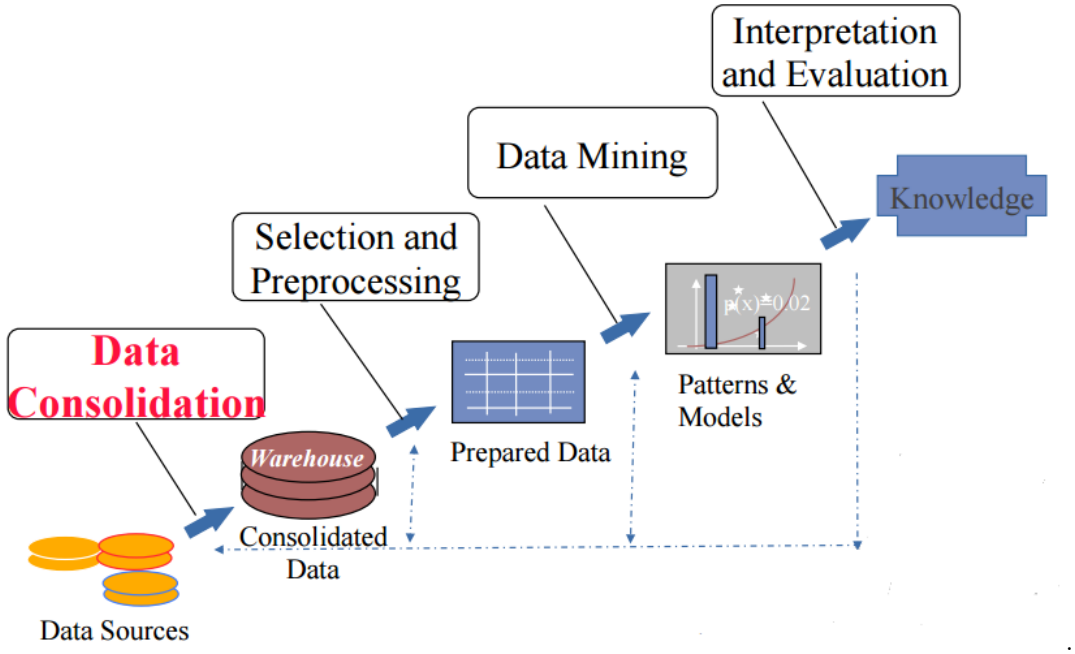

Figura 8.1: KDD Process

La prima fase del processo di estrazione della conoscenza prevede il consolidamento dei dati provenienti dalle diverse fonti esterne, come database relazionali o file esterni, in un'unica collezione. In questa fase si effettua la pulizia dei dati da eventuali rumori e da informazioni non corrette o non rilevanti, in modo da poter garantire un'adeguata qualità e affidabilità della collezione prodotta. Questo è un aspetto fondamentale in quanto non è possibile sperare di ottenere buoni risultati a partire da dati di scarsa qualità.

La seconda fase ha come obiettivo la produzione di una o più tabelle da usare come input per gli algoritmi di data mining. In questa fase i dati possono essere campionati per ridurre il numero di righe e sono eliminati gli attributi ridondanti o non ritenuti significativi. Infine, dopo l'utilizzo degli algoritmi di mining, l'ultima fase si occupa dell'interpretazione e valutazione dei risultati prodotti.

8.2 Microstrategy Data Mining Services

I servizi di data mining di MicroStrategy possono aiutare le aziende a utilizzare i loro dati per prevedere i risultati futuri. Data Mining Services può essere

ampiamente utilizzato in attività aziendali e settori diversi, quali la previsione di risultati e comportamenti futuri. Le sue caratteristiche sono:

- Utilizzo di MicroStrategy per creare regressioni lineari a più variabili, regressioni esponenziali a più variabili, regressioni logistiche, alberi decisionali, modelli predittivi di cluster, regole di associazione e modelli di serie temporali,
- Supporto per l'importazione di modelli predittivi di terzi grazie allo standard PMML (Predictive Model Markup Language), è uno standard XML che rappresenta i modelli di data mining,
- Funzionalità di visualizzazione di modelli predittivi.

8.2.1 Applicazione Modello Predittivo

Il processo della creazione di un modello predittivo e incorporazione nella piattaforma MicroStrategy di business intelligence comporta le seguenti azioni:

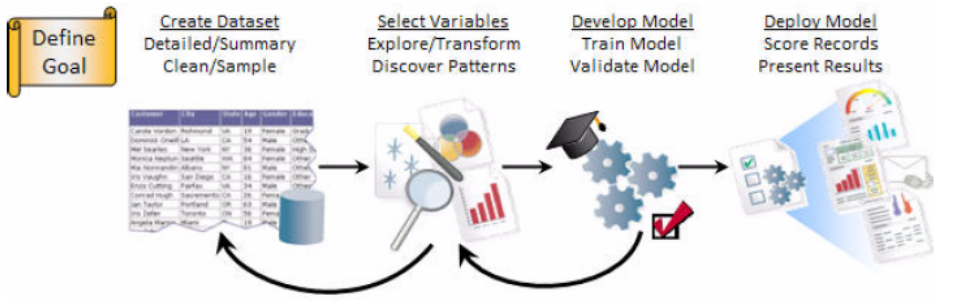

Figura 8.2: processo

Il processo ha inizio con l'importante fase di definizione di un obiettivo dell'analisi. L'obiettivo deve essere definito in termini aziendali, identificando i risultati desiderati per migliorare alcuni aspetti delle prestazioni dell'organizzazione. In presenza di un obiettivo accuratamente definito, è necessario attenersi ai passi elencati di seguito per la creazione e la distribuzione di un modello di data mining:

- Creazione di un report dataset da utilizzare per sviluppare il modello predittivo.

Nel lavoro svolto si è deciso di generare una quantità di dataset con l'intento di separare i singoli negozi. Si è giunti alla conclusione che ogni dataset si compone da tutti gli scontrini emessi nell'anno 2015 raggruppati per categoria merceologica.

- Creazione della metrica predittiva con MicroStrategy Desktop.

  In cui si sono impostati una serie di parametri tra i quali il nome dell'algoritmo da utilizzare, in questo caso Association, il supporto e la confidenza. Come si vede dall'immagine è stata impostata una soglia per il supporto minimo molto bassa, equivalente a 0,01 e una confidenza minima a 0,01. Valori così bassi sono stati necessari per la presenza di molte transazioni, ognuna delle quali contenente più categorie.

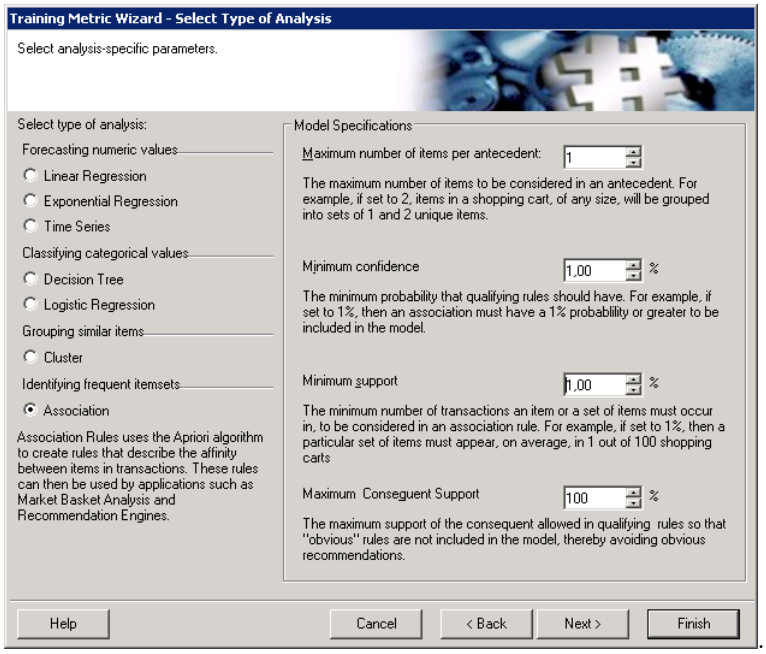

Figura 8.3: processo

In questo passo sono stati selezionati gli attributi da passare in ingresso all'algoritmo per l'estrazione delle regole associative ossia l'id dello scontrino e la categoria merceologica.

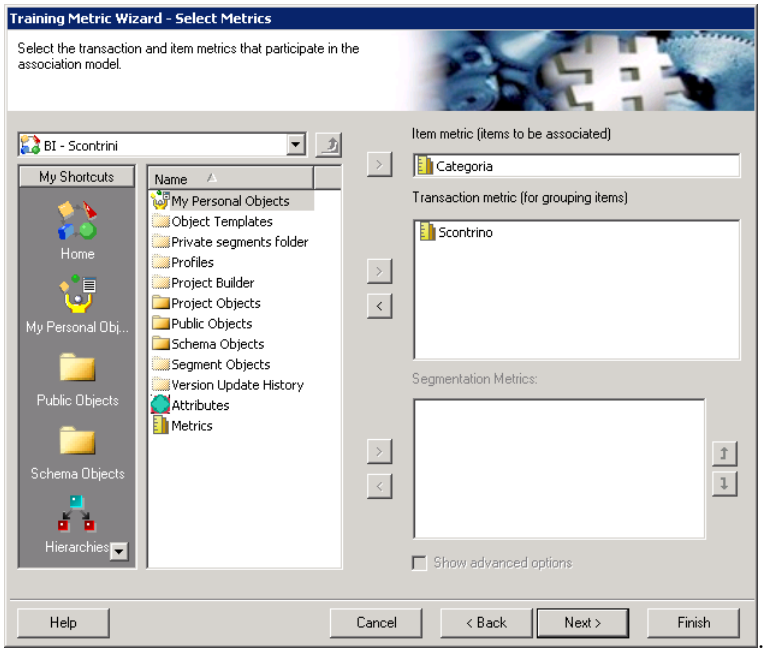

Figura 8.4: processo

Infine si imposta dove salvare il modello e quali caratteristiche far mostrate delle regole associative.

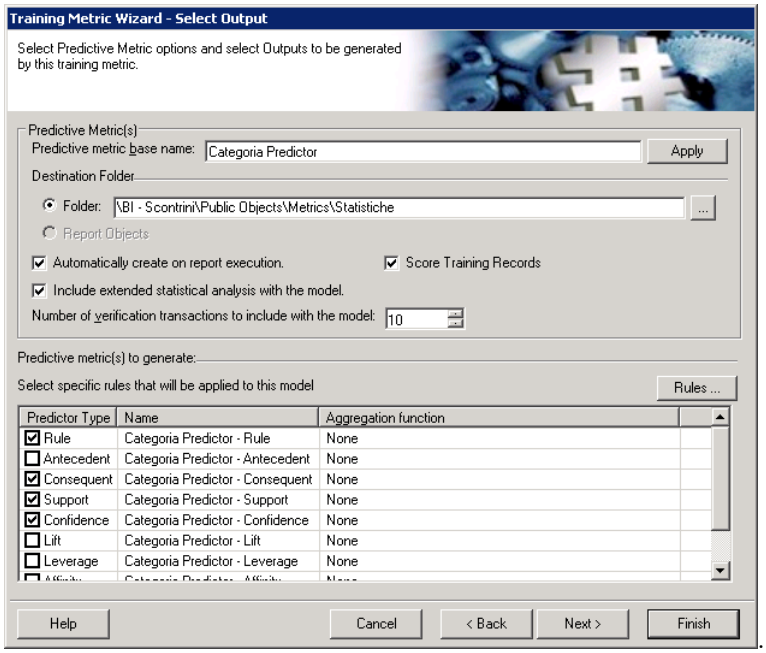

Figura 8.5: processo

- Distribuzione della metrica predittiva nei report MicroStrategy per prevedere nuovi risultati, un processo chiamato scoring.

Una volta creato il modello predittivo, è possibile visualizzarlo utilizzando il visualizzatore modello predittivo MicroStrategy.

Lo scopo dello studio era quello di trovare associazioni frequenti ed evidenziare quelle più interessanti. L'analisi effettuata è stata svolta su due negozi differenti: il primo negozio ubicato in centro paese l'altro in periferia, presso la sede della cooperativa. L'analisi ha riportato circa 35 regole associative per negozio, dopo essere state ordinate in base all'ordine decrescente dei valori di confidenza sono mostrate di seguito separatamente in base ai negozi.

Si procede a descrivere tali associazioni mostrando nella Figura 8.6 quelle relative al negozio in sede.

Antecedente	Conseguente	Confidenza	Supporto
OVINO	SUINO	73,40%	8,36%
INNOVATIVI	SUINO	71,80%	25,70%
VITELLO	SUINO	62,53%	6,93%
VITELLONE	SUINO	60,29%	8,46%
AVICOLI	SUINO	59,68%	19,12%
SCOTTONA	SUINO	59,56%	17,91%
OVINO	INNOVATIVI	59,19%	6,72%
ALTRI SALUMI	SUINO	55,64%	9,36%
AGNELLO	SUINO	55,49%	1,71%
GASTRONOMIA	SUINO	54,85%	8,88%
SUINO	INNOVATIVI	54,60%	25,70%
UOVA	SUINO	53,17%	1,93%
CASEARI	SUINO	51,46%	8,56%
VITELLO	AVICOLI	50,77%	5,62%

Figura 8.6: Regole Associative: Negozio in sede

- La regola con più alto valore di confidenza pari al 73.4% è quella che mette in relazione: `OVINO` → `SUINO`, ovvero quando è stato acquistato dell'ovino nel 73.4% dei casi è stato acquistato anche del suino.
- Altrettanto rilevante per valori statistici di confidenza (71,8%) e di supporto (25,7%) è l'associazione: `INNOVATIVI` → `SUINO`. La categoria degli

innovativi comprende tutti gli insaccati prodotti con.....I valori che abbiamo trovato sono da considerarsi abbastanza alti se paragonati all'enorme numero di elementi complessivi che compongono l'intero dataset.

- Un'altra regola significativa è `VITELLO` → `SUINO` con confidenza al 62.5%.

Si nota come nella maggior parte delle regole associative il conseguente sia la categoria del SUINO.

Invece nella Figura 8.7, sono mostrate le association rule del negozio ubicato in centro.

Antecedente	Conseguente	Confidenza	Supporto
PASTA FRESCA	GASTRONOMIA	60,55%	0,36%
OVINO	SUINO	60,54%	2,84%
	INNOVATIVI	54,79%	2,57%
ROSTICCERIA	GASTRONOMIA	53,65%	2,90%
.INNOVATIVI	SUINO	52,40%	13,18%
SUINO	INNOVATIVI	46,87%	13,18%
VITELLONE	SCOTTONA	46,23%	4,69%
.PROSCIUTTI	ALTRI SALUMI	42,36%	6,51%
	CASEARI	42,23%	6,49%
VASELLAME	GASTRONOMIA	41,94%	0,93%
ALTRI SALUMI	CASEARI	40,74%	8,41%

Figura 8.7: Regole Associative: Negozio in centro

Dall'ordinamento ascendente per confidence sono risultate interessanti 2 associazioni:

- la prima mostra la relazione `PASTA FRESCA` → `GASTRONOMIA` con confidenza (60,55%) e supporto (0,36%). Risulta molto interessante dal punto di vista del Data Mining perché non è immediato trovare un'associazione tra pasta fresca e gastronomia.
- la seconda descrive invece la relazione tra `OVINO` → `INNOVATIVI` con confidenza (54,79%).

Il negozio del centro si mostra come la soluzione più consona per i beni di prima necessità da acquistare all'occorrenza. I clienti scelgono questa soluzione per l'immediato utilizzo che, nella maggior parte dei casi, vista la freneticità

della vita nel centro della città, si rivela essere una combinazione di pasti veloci, precotti o a cottura rapida, come la scelta di acquistare nel reparto gastronomia, insieme all'acquisto di pasta fresca. Il negozio della periferia ubicato presso la sede della cooperativa, più grande e più fornito di quello in centro, è invece la soluzione ideale del consumatore attento non solo alla qualità del prodotto ma alla vastissima scelta. Infatti, si può osservare come in questa sede si prediliga l acquisto di carni di vario genere, tutte a km zero, non necessariamente per l'immediato utilizzo, ma presumibilmente come acquisto da provvista.

Capitolo 9

CONCLUSIONI

È stato presentato il lavoro svolto relativo alla progettazione e alla realizzazione di un sistema di data warehouse e business intelligence per l'analisi di tre processi di business aziendali, sviluppato all'interno di un'attività progettuale dell'azienda Deloitte rivolta ad una azienda italiana del settore agroalimentare.

L'attenzione e il tempo dedicati alle fasi di progettazione, in particolare alla raccolta e alla specifica dei requisiti, hanno consentito una realizzazione lineare che non ha richiesto modifiche sostanziali al modello formulato in prima istanza. L'approccio utilizzato è stato quello di sviluppare le implementazioni dei tre processi di business in sequenza, al fine di rendere disponibile nel più breve tempo possibile le parti quanto più esaudienti e rispondenti ai requisiti richiesti dall'azienda committente.

Il sistema realizzato ha complessivamente soddisfatto le aspettative, conseguendo una risposta positiva da parte degli utenti aziendali durante i periodi antecedenti la messa in produzione.

Migliorie aggiuntive potrebbero riguardare l'arricchimento delle informazioni rappresentate e la creazione di nuovi report per mostrare fenomeni non ancora evidenziati. Inoltre, potrebbe risultare davvero utile sfruttare in modo migliore i dati ora disponibili, potendo tentare delle esplorazioni conoscitive con tecniche di data mining sui singoli prodotti e non limitandosi alle categorie merceologiche.

L'esperienza di lavoro è stata particolarmente gratificante in quanto mi ha

dato l'opportunità di conoscere più da vicino la complessità sistemica della vita all'interno dell'azienda. Affrontando le difficoltà impreviste e analizzando al meglio i meccanismi di gestione di una azienda, ho potuto migliorare efficacemente le mie competenze nella realizzazione di un data warehouse adattato ad un caso reale.

Bibliografia

[1] Antonio Albano. Decision support databases essentials, Luglio 2015.

[2] Rizzi S. Golfarelli M. Data warehouse. teoria e pratica della progettazione, 2006.

[3] Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, and Bob Becker. The data warehouse lifecycle toolkit, 2008.

[4] Caserta J. Kimball R. The data warehouse etl toolkit, 2004.

[5] Oracle. Oracle database data warehousing guide, Novembre 2013.

[6] Microstrategy Inc. Microstrategy 10 basic reporting, d.n. 09491010, Dicembre 2015.

[7] Microstrategy Inc. Microstrategy 10 advanced reporting, d.n. 09451010, Dicembre 2015.

[8] Microstrategy Inc. Microstrategy 10 project design guide, d.n. 09331010, Dicembre 2015.

[9] Stephen Few. Information dashboard design: the effective visual communication of data, 2006.