

*Università degli Studi di Modena e
Reggio Emilia*

Dipartimento di Ingegneria “Enzo Ferrari”
Corso di Laurea in Ingegneria Informatica

Analisi comparativa degli Strumenti di
Business Intelligence per le PMI

Relatore:
Prof.ssa Sonia Bergamaschi

Candidato:
Stefano Pini

Anno Accademico 2014/2015

Keywords:

PMI

Business Intelligence (BI)

BI tools benchmark

SpagoBI

Talend

Indice

1	Introduzione	5
2	La Business Intelligence (BI): concetti generali	6
2.1	Data Warehouse.....	6
2.1.1	Modelli	7
2.1.2	ETL.....	8
2.2	Analisi e presentazione dei dati	9
2.2.1	OLAP (OnLine Analytical Process)	9
2.2.2	Reporting	11
2.2.3	Grafici.....	12
2.2.4	Cruscotti Interattivi	12
3	Analisi Comparativa di strumenti di BI	14
3.1	Introduzione	14
3.2	Prodotti open source	14
3.2.1	Pentaho Data Integration	14
3.2.2	Talend Open Studio	15
3.2.3	Jaspersoft Community Edition.....	16
3.2.4	SpagoBI Suite	18
3.3	Prodotti Commerciali	19
3.3.1	Jaspersoft Studio Commercial Edition	19
3.3.2	Tableau.....	20
3.3.3	Pentaho Enterprise Edition	21
3.4	Tabella riassuntiva.....	23
3.4.1	Prodotti Open Source	23
3.4.2	Prodotti Commerciali.....	23
4	La Suite SpagoBI.....	24
4.1	Introduzione	24
4.2	Componenti.....	24
4.3	SpagoBI Server	26
4.3.1	Installazione e configurazione	26
4.3.2	Funzionalità ed Utilizzo	27
4.4	SpagoBI Studio	31
4.4.1	Installazione di SpagoBI Studio su Windows 8.1 64bit	31
4.4.2	Funzionalità ed Utilizzo	31

4.4.3	Grafici	32
4.4.4	Report Statici.....	33
4.4.5	Cruscotti (Cockpit)	34
4.5	Talend Open Studio for Data Integration.....	37
5	Conclusioni e Sviluppi Futuri	40
5.1	Risultati.....	40
5.2	Sviluppi futuri	40
6	Indice delle Figure	42
7	Bibliografia.....	43
8	Sitografia.....	44

1 Introduzione

La tesi qui proposta ha lo scopo di presentare il lavoro svolto all'interno dell'azienda Soft-Net Srl durante il tirocinio formativo della durata di tre mesi.

Soft-Net Srl è un'azienda di consulenza ed assistenza informatica operante nel territorio di Modena e Reggio Emilia, ma anche una Software House specializzata nello sviluppo di applicazioni web.



Il prodotto principale è un software CRM (Customer Relationship Management) completamente web-based fruibile tramite browser da qualsiasi PC connesso ad internet. I servizi, invece, spaziano dallo sviluppo di siti web/e-commerce alla realizzazione di campagne di web marketing e SEO (Search Engine Optimization), dall'assistenza tecnica alla realizzazione di architetture di rete e all'installazione di server aziendali. Soft-Net Srl è certificata Cisco, partner Dell e, dal 2014, partner Nethesis.

In un contesto di mercato sempre più competitivo e in rapido cambiamento, le aziende sono obbligate ad intraprendere scelte decisive per il loro futuro in tempi sempre minori. Diventa quindi necessario poter valutare rapidamente ed efficacemente i trend di mercato, i risultati della produzione e i molteplici fattori che influenzano l'andamento della propria impresa, come anche le previsioni di crescita a breve e a lungo termine. Un indispensabile supporto per questo tipo di analisi è un sistema informativo che integri i dati provenienti dai diversi settori dell'azienda e che ne permetta una facile consultazione.

È seguendo queste considerazioni che Soft-Net Srl ha deciso affacciarsi nel mondo della Business Intelligence, innanzitutto offrendo un tirocinio incentrato sulla ricerca e l'analisi di software di Business Intelligence per la PMI (cioè con costi affrontabili da quest'ultime, principali clienti dell'azienda) che potesse integrarsi all'interno del proprio CRM.

L'elaborato si articola in cinque sezioni: questa prima sezione introduttiva, un capitolo dedicato ai concetti della Business Intelligence, un'analisi comparativa dei prodotti disponibili sul mercato, un'analisi approfondita della suite SpagoBI, un ultimo capitolo contenente gli eventuali sviluppi futuri.

2 La Business Intelligence (BI): concetti generali

La Business Intelligence (spesso abbreviata in BI) è l'insieme di metodi, processi, strumenti e tecnologie che permette di ricavare informazioni utili e significative da dati grezzi altrimenti non utilizzabili, allo scopo di sviluppare una conoscenza globale e facilmente accessibile della propria impresa che supporti e migliori i processi decisionali, di comunicazione e di controllo della stessa.

In altre parole, possiamo definire la Business Intelligence come un servizio che raccoglie dati dai vari settori dell'azienda e li rende informazioni accessibili a supporto del gruppo decisionale dell'impresa.

Le funzionalità principali offerte da un sistema di BI sono:

- Elaborazione ed integrazione di dati eterogenei, tramite strumenti specifici che permettono di leggere ed elaborare dati raccolti da diverse fonti e raccogliarli, trasformati e ristrutturati, in un nuova base di dati ottimizzata per le operazioni di lettura.
- Navigazione ed analisi dei dati elaborati ed integrati, tramite tool che consentono una rapida visualizzazione ed esplorazione, anche approfondita e complessa, dei dati di interesse.
- Presentazione dei dati, tramite strumenti chiari, facilmente accessibili e personalizzabili, come report, grafici e cruscotti interattivi.

I possibili utilizzatori di questi strumenti sono potenzialmente tutte le persone tenute a prendere decisioni all'interno dell'impresa, dal caporeparto al manager aziendale. Questo è possibile poiché le funzionalità offerte dalla BI sono adattabili ai diversi livelli aziendali e ai diversi ruoli ricoperti dagli utilizzatori: in base al campo di competenza dell'utente è possibile limitare gli accessi ai dati, permettere l'utilizzo dei soli strumenti necessari e mostrare report personalizzati focalizzati sui dati di interesse dell'utente specifico.

In questo modo si può mettere a disposizione di tutti coloro che devono compiere scelte all'interno dell'azienda uno strumento di supporto alle decisioni realizzato su misura in base ai dati ed alle necessità di loro competenza e interesse.

2.1 Data Warehouse

Secondo la definizione di Bill Inmon, considerato da molti "il padre dei Data Warehouse", il Data Warehouse (abbreviato DW) è una collezione di dati integrata, orientata al soggetto, tempo variante e non volatile posta a supporto del processo decisionale manageriale. Ralph Kimball invece, altro importante ricercatore nell'ambito dei DW, lo definisce sinteticamente come una copia specificatamente strutturata per le interrogazioni e l'analisi di dati transazionali.

Il DW è la struttura dati che meglio si adatta alle necessità della Business Intelligence in quanto permette di ottenere tempi di esecuzioni ridotti per le operazioni di select anche su grandi quantità di dati, permette di integrare più sorgenti dati e mantenere la storia completa dei dati caricati.

Nella seguente tabella è mostrato un confronto fra la struttura dati tradizionale (OLTP, On Line Transaction Processing) e la struttura di un Data Warehouse (che sposa un modello

OLAP, On Line Analytical Processing).

I primi sono sistemi ottimizzati per eseguire rapidamente operazioni di inserimento/aggiornamento e lettura su un ristretto volume di dati, mentre i secondi sono ottimizzati per permettere di eseguire query complesse su grandi quantità di dati in tempi ragionevoli.

	Base di dati transazionale – OLTP	Data Warehouse – OLAP
Sorgente dati	Singola, dati operazionali	Multipla, dati provenienti da più sorgenti OLTP
Operazioni predominanti	Insert/update	Select
Volume dati per transizione	Piccolo	Grande
Storicità dei dati	Solitamente non presente	Presente
Dimensione	Generalmente contenuta	Elevata, in particolare a causa dei dati storici da mantenere
Utenza tipica	Dipendenti, molte connessioni contemporanee	Dirigenti e manager aziendali, poche connessioni contemporanee

2.1.1 Modelli

Il modello su cui si basa la struttura di un DW, chiamato modello multidimensionale, è differente dal classico modello relazionale. Esso offre una struttura che supporta operazioni del tipo OLAP, la base per la rappresentazione e l'interrogazione dei dati dei data warehouse.

Il modello multidimensionale stesso si può dividere in due modelli teorici leggermente differenti: lo star schema e lo snowflake schema. Entrambi contengono al centro dello schema la tabella dei fatti, a cui sono collegate le tabelle delle dimensioni. I fatti sono gli argomenti su cui sarà condotta l'analisi e si compongono di valori reali detti misure, mentre le dimensioni sono strutture che caratterizzano le misure presenti nella tabella dei fatti.

2.1.1.1 Star Schema

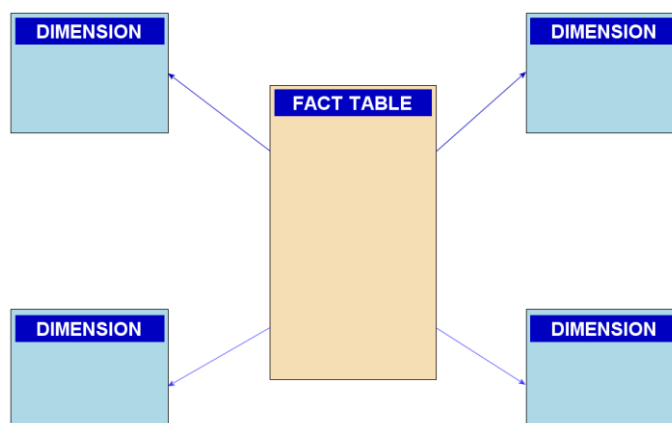


Figura 1 - Schema a stella (Star schema)

Lo schema a stella è composto da un insieme di relazioni $DT_1...DT_N$, ognuna delle quali è chiamata Dimension Table e rappresenta una dimensione, e una relazione FT, chiamata Fact Table. Ogni dimensione DT_i è caratterizzata da una chiave primaria (tipicamente surrogata) d_i e da un insieme di attributi che descrivono le dimensioni di analisi a diversi livelli di aggregazione. La Fact Table importa le chiavi di tutte le tabelle delle Dimension Table: la sua chiave primaria è data dall'insieme delle chiavi esterne delle tabelle $DT_1...DT_N$, $d_1...d_N$, in più contiene un attributo per ogni misura.

In questo tipo di schema, le tabelle delle dimensioni sono completamente denormalizzate: con un solo join si possono recuperare tutti i dati relativi ad una dimensione. Questo porta da un lato ad un notevole incremento delle prestazioni nelle operazioni di lettura, dall'altro ad una forte ridondanza dei dati, che può portare ad un notevole aumento delle dimensioni del Data Warehouse.

2.1.1.2 Snowflake schema

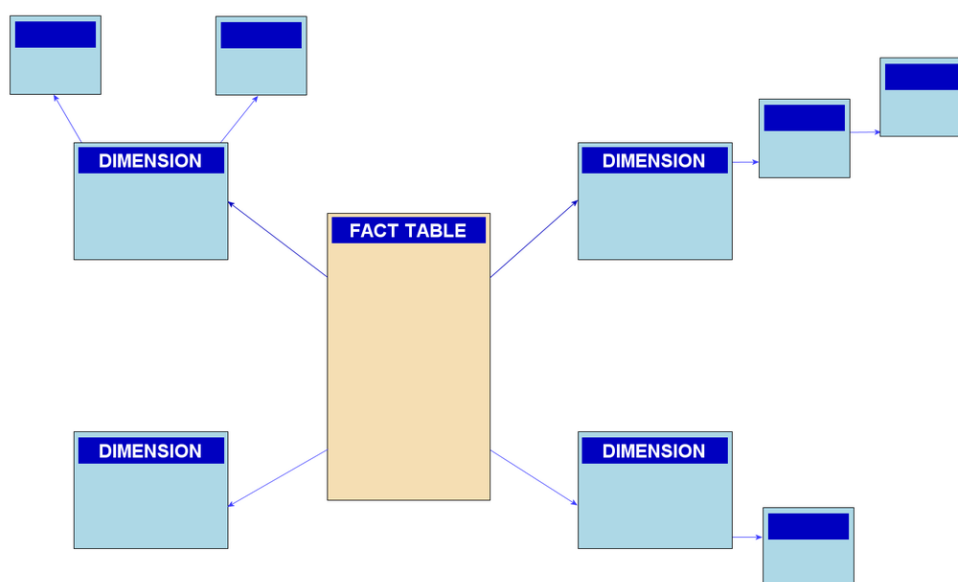


Figura 2 - Schema a fiocco di neve (Snowflake schema)

Lo schema a fiocco di neve, rispetto allo star schema, riduce la denormalizzazione delle Dimension Table DT_i eliminando tutte le dipendenze funzionali da attributi non chiave. Questo tipo di schema può essere ottenuto direttamente oppure può essere ricavato dallo star schema attraverso un processo di normalizzazione delle tabelle delle dimensioni che elimini la ridondanza dei dati.

Il vantaggio di questo schema è l'eliminazione della ridondanza dei dati e una minore occupazione di spazio da parte del DW. Di contro, la normalizzazione obbliga l'esecuzione di più join per recuperare tutti i dati relativi ad una dimensione (problema arginabile tramite l'utilizzo di viste che rappresentino la dimensione nella sua interezza).

2.1.2 ETL

ETL è un acronimo inglese che sta per Extract, Transform and Load: queste sono le tre operazioni necessarie a popolare di dati il Data Warehouse. Per la realizzazione di un DW valido e funzionale questa è una fase fondamentale in quanto occorre recuperare, trasformare ed integrare i dati in modo ottimale (per esempio eliminando i duplicati, rimuovendo tutti i dati non d'interesse, controllando i dati già presenti nel DW ed

eventualmente integrandoli con quelli nuovi) per rispettare i requisiti di velocità e semplicità di lettura necessari.

Analizziamo nel dettaglio le tre fasi dell'ETL:

- Estrazione: in questa fase è necessario connettersi alle diverse sorgenti dati per estrarne i contenuti e renderli disponibili per la fase di trasformazione.
Le sorgenti dati possono essere di molteplici tipi: database (relazionali e non), file di testo, file in forma tabellare (excel, csv), file xml, servizi web.
La possibilità di connettersi a tali sorgenti dipende dal software ETL utilizzato.
- Trasformazione: in questa fase i dati estratti dal passo precedente vengono elaborati e preparati per il caricamento all'interno del DW.
Possono essere necessarie operazioni di validazione, confronto, modifica, integrazione ed eliminazione dei dati o della loro struttura e calcolo di valori aggregati.
- Caricamento: in quest'ultima fase i dati, estratti dal primo passaggio e trasformati dal secondo, vengono caricati nel DW.

2.2 Analisi e presentazione dei dati

Una volta che la creazione del DW è stata completata, è necessario disporre di strumenti che permettano di visualizzare i dati e di effettuare analisi su di essi.

In base al suo ruolo all'interno dell'azienda, l'utente avrà accesso solo a determinati strumenti di visualizzazione ed analisi su un gruppo ristretto di dati o perfino solo a documenti realizzati in precedenza con i soli dati di suo interesse. In questo modo, un utente avanzato e con un ruolo di rilievo potrà realizzare in completa autonomia i documenti e le analisi di suo interesse, mentre un utente con compiti e competenze limitate avrà un accesso ridotto ai dati e agli strumenti presenti nella suite di BI.

Nei seguenti capitoli analizzeremo le peculiarità dei principali strumenti messi a disposizione dai prodotti attualmente presenti sul mercato.

2.2.1 OLAP (OnLine Analytical Process)

Sotto al nome di OLAP vengono solitamente fatti confluire tutti i sistemi che permettono l'analisi interattiva e veloce di grandi quantità di dati. Nell'ambito della BI, con OLAP vengono indicati sia la struttura del DW (schema a stella o a fiocco di neve) che gli strumenti che presentano i dati all'utente finale.

Secondo E.F. Codd, inventore del modello relazionale e del termine OLAP, sotto a questo nome possiamo racchiudere tutti gli strumenti che permettono di consolidare, visualizzare e analizzare dati rispetto a più dimensioni, cioè che permettono di effettuare analisi di dati multidimensionali. Fra le funzionalità indicate da Codd come maggiormente rilevanti troviamo:

1. Schema concettuale multidimensionale (*Multidimensional Conceptual View*)
2. Trasparenza
3. Performance consistenti
4. Manipolazione (Esplorazione) dei dati intuitiva

5. Reporting flessibile
6. Numero di dimensioni e di livelli di aggregazione illimitato

La struttura OLAP più comune è chiamata *cubo multidimensionale* (o *iper-cubo*). Questa struttura logica si basa su un cubo di dimensione N (che può essere un qualsiasi valore maggiore o uguale a 2) nel quale ogni asse corrisponde a una dimensione ed ogni intersezione fra i valori delle varie dimensioni corrisponde ai fatti relativi a quei valori.

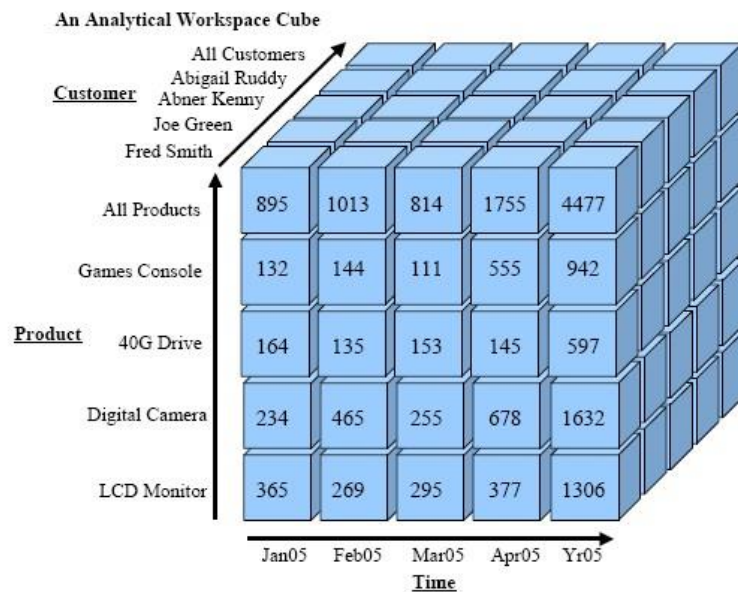


Figura 3 - Cubo Multidimensionale (o Cubo OLAP)

Dal cubo OLAP è possibile realizzare tabelle pivot dinamiche, tramite le quali è possibile effettuare analisi e ricerche su un grande numero di dati in tempi relativamente ridotti.

Queste attività si effettuano solitamente tramite le seguenti operazioni:

1. Slicing, cioè l'estrazione di un subset di informazioni dal cubo che si sta analizzando fissando uno specifico valore per una delle dimensioni del cubo stesso
2. Dicing, cioè l'estrazione di un subset di informazioni dal cubo che si sta analizzando limitando una o più dimensioni del cubo ad un set finito di valori
3. Drill-Down, cioè l'espansione del dato nelle sue componenti (o determinanti) tramite l'esplosione di una dimensione nelle sue componenti o di un dato calcolato nei suoi determinanti
4. Drill-Up, cioè l'operazione inversa rispetto al drill-down, permette di visualizzare i dati aggregati rispetto ad una dimensione invece che esplosi nelle sue componenti
5. Pivoting, cioè la rotazione del cubo per visualizzare facce differenti, così da visualizzare i dati aggregati rispetto ad altre dimensioni

La maggior parte degli strumenti di BI analizzati permette di realizzare tabelle pivot dinamiche, basate sulla struttura logica dei cubi OLAP, partendo da schemi a stella o a fiocco di neve.

2.2.2 Reporting

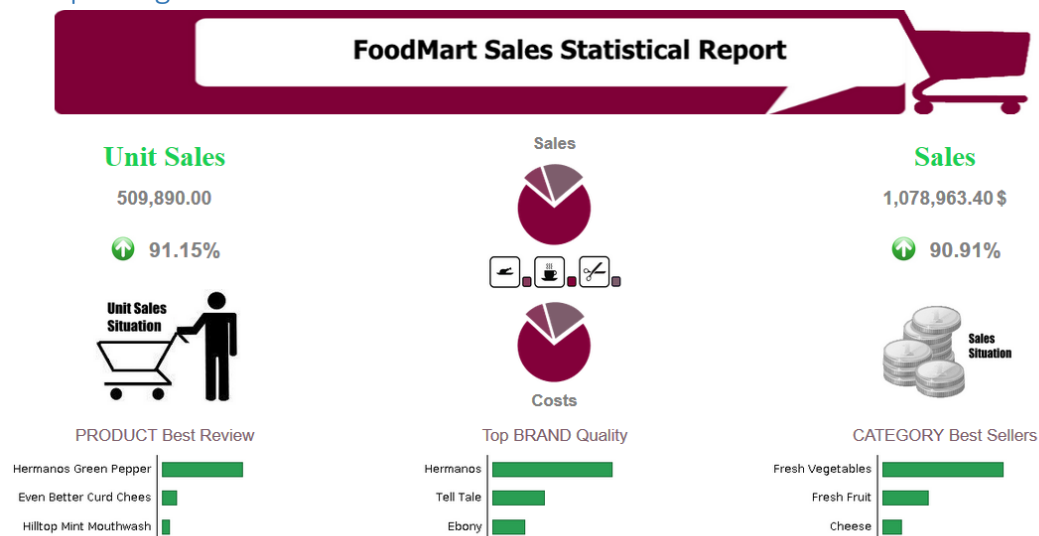


Figura 4 - Esempio di report (realizzato con BIRT)

Il *Reporting* è il processo di preparazione di un documento informativo e la sua presentazione ai soggetti interessati. In particolare, nell'ambito della BI con il termine *Reporting* si racchiudono gli strumenti e le operazioni volte alla preparazione di documenti informativi contenenti i dati di interesse estratti dal DW e alla loro presentazione/invio/condivisione con gli utenti interessati.

Solitamente le suite di BI offrono editor grafici specifici per realizzare documenti di report (o meglio modelli di report da cui produrre documenti sempre aggiornati). Al loro interno è possibile inserire intestazioni e piè di pagina personalizzati, testo libero, testo risultante da interrogazioni del data warehouse, immagini, grafici, tabelle ed altri documenti realizzati in precedenza.

Altri componenti spesso presenti sono un scheduler, che permette di realizzare report automaticamente partendo da modelli realizzati precedentemente ad intervalli regolari, e un componente adibito alla pubblicazione o all'invio (per esempio tramite mail) del report generato in automatico allo scheduler. In questo modo, una volta completata la realizzazione dei modelli dei report necessari, il prodotto di BI provvederà automaticamente a generare i report e a renderli disponibili agli utenti che ne hanno bisogno.

2.2.3 Grafici

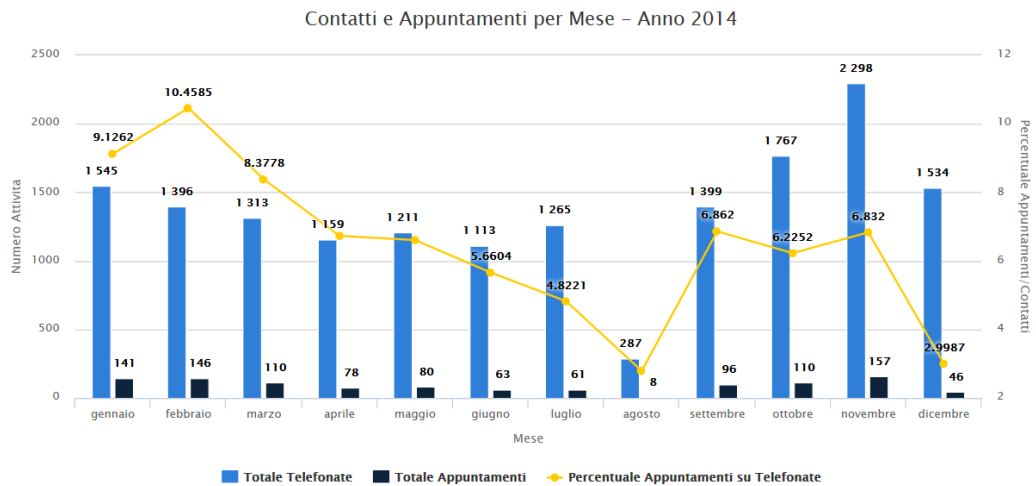


Figura 5 - Esempio di grafico (realizzato con SpagoBI, motore HighCharts)

I grafici, fra gli strumenti di presentazione dei dati, sono particolarmente utilizzati per la loro immediatezza e facilità di comprensione. Senza la necessità di avere specifiche competenze tecniche o metodi di analisi, un qualsiasi utente è in grado di estrarre informazione da un grafico correttamente realizzato.

Per questo motivo, la totalità degli strumenti di presentazione dei dati offrono la possibilità di rappresentare i dati attraverso la realizzazione di grafici di diverso tipo (per esempio grafici a colonne, a linee, a torta, ecc.). Inoltre, in molti strumenti è possibile effettuare operazioni dinamiche sui grafici realizzati, per esempio applicare filtri, ridurre il dominio dei dati e nascondere/mostrare misurazioni.

2.2.4 Cruscotti Interattivi

I cruscotti interattivi, o dashboard, sono uno dei metodi di visualizzazione più interessanti delle recenti suite di Business Intelligence. Essenzialmente sono documenti che al loro interno contengono altri documenti (report, grafici, tabelle, ecc.) il cui punto di forza è la possibilità di interazione: effettuando un filtro o una selezione su uno dei documenti presenti, anche gli altri documenti saranno aggiornati in base all'azione effettuata. In alcuni prodotti di BI è anche possibile impostare l'apertura di un nuovo documento (anche di un nuovo cruscotto) al click su determinate sezioni del documento; anche in questo caso il nuovo documento aperto potrà visualizzare dati differenti in base all'oggetto sui cui si è premuto (per esempio, cliccando su una barra di un grafico potrà aprirsi un nuovo documento con i dettagli della sezione dei dati su cui si è premuto).

Attraverso questa funzionalità, non necessariamente delegata ai soli cruscotti, si può creare un documento, formato da una serie di documenti appositamente collegati fra loro, che sia navigabile e filtrabile in tempo reale.

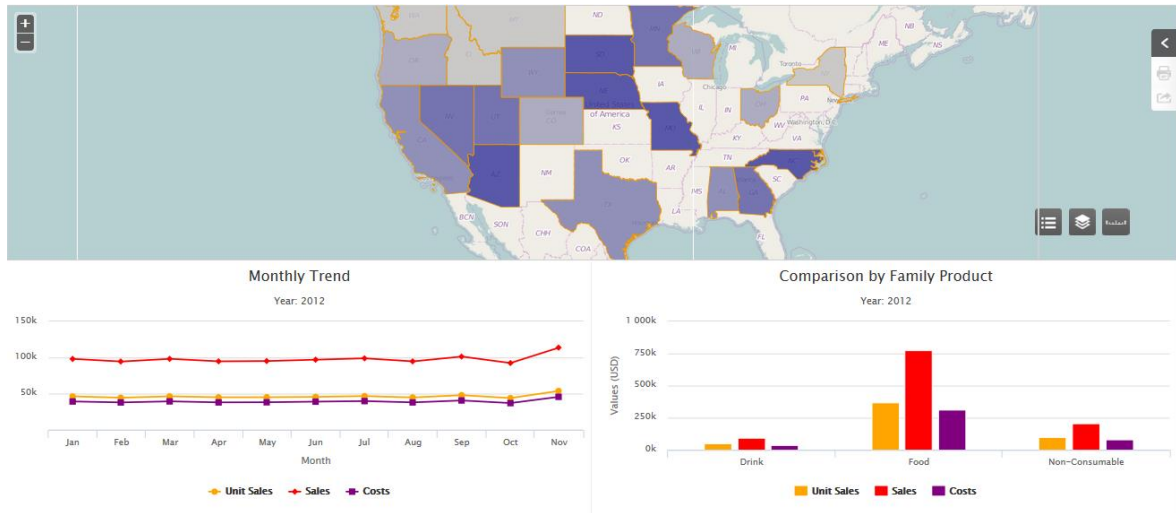


Figura 6 - Esempio di cruscotto interattivo (realizzato con SpagoBI)

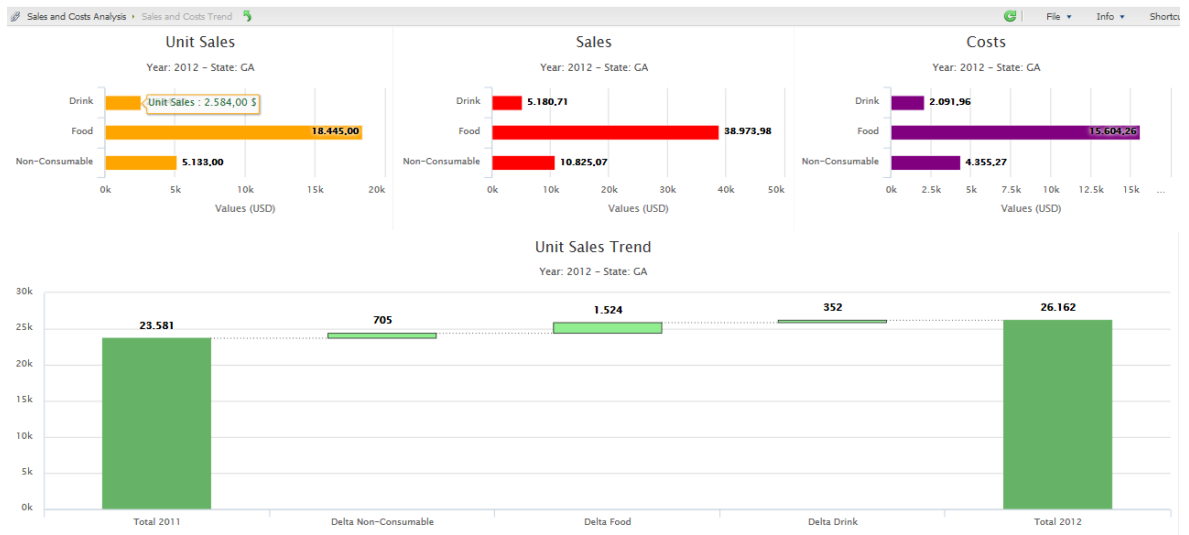


Figura 7 - Esempio di cruscotto interattivo (realizzato con SpagoBI) che si apre in seguito al click di uno stato all'interno del cockpit mostrato nella figura precedente

3 Analisi Comparativa di strumenti di BI

3.1 Introduzione

In questo capitolo verranno illustrati gli strumenti di Business Intelligence analizzati durante il tirocinio, mostrando le potenzialità e i limiti di ognuno di essi. Il prodotto cercato da Soft-Net Srl doveva rispettare principalmente il seguente requisito: doveva offrire la possibilità di integrare i documenti prodotti all'interno del loro CRM web-based o in una qualunque pagina web.

Tutti gli strumenti analizzati nel seguito sono utilizzabili per soluzioni di BI per PMI in quanto sono offerti o in versione gratuita, o a prezzi abbordabili per aziende di questo tipo (che ovviamente non possono affrontare i prezzi proposti dai prodotti di aziende del calibro di Microsoft, SAP, IBM, Oracle o MicroStrategy).

Per questo motivo è stata effettuata una suddivisione fra i prodotti commerciali (cioè legati all'acquisto di una o più licenze per poter essere utilizzati) e i prodotti open source (con licenze che permettano il libero uso per scopi commerciali).

3.2 Prodotti open source

3.2.1 Pentaho Data Integration

Pentaho Data Integration (abbreviato PDI) è lo strumento ETL, precedentemente conosciuto come Kettle, della Hitachi Data Systems. Ogni prodotto Pentaho è offerto in due versioni: la versione community, gratuita ed open source, e la versione enterprise, con prezzi personalizzati in base alle necessità del cliente. Inoltre i prodotti sono disponibili sia per il download individuale che per il download all'interno della suite Pentaho Community/Enterprise Edition che racchiude tutti gli strumenti sviluppati dall'azienda. Il prodotto testato è Pentaho Data Integration nella versione Community Edition.

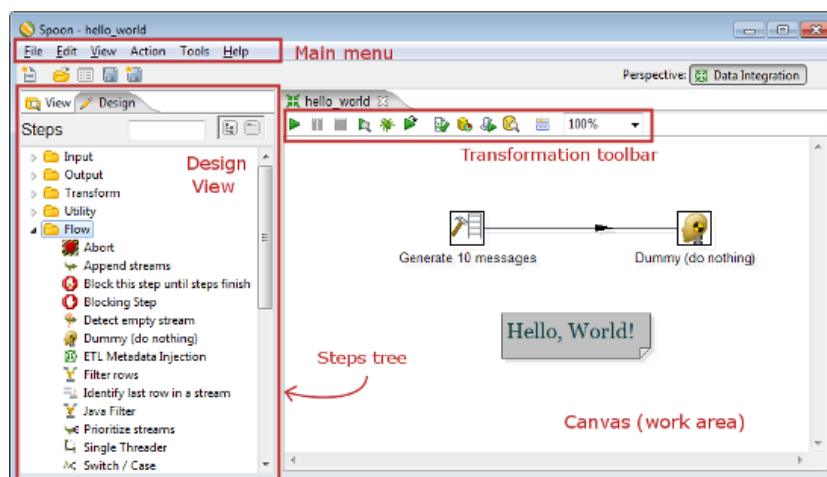


Figura 8 - Pentaho Data Integration - IDE Spoon

PDI offre un'intuitiva interfaccia grafica drag-and-drop, chiamata Spoon, con la quale si possono sviluppare processi ETL molto rapidamente. È possibile realizzare delle Transformations, che racchiudono le principali operazioni di ETL, o dei Jobs, che invece sono dei controllori di flusso all'interno dei quali è possibile inserire transformations, logica di controllo od operazioni di avviso o di invio di file semplicemente trascinando all'interno della pagina di sviluppo i blocchi corrispondenti alle operazioni da eseguire.

Le procedure così realizzate sono salvate in formato XML e vengono interpretate dal motore Kettle al momento dell'esecuzione.

È possibile estendere le funzionalità del programma attraverso l'installazione di plugin, recuperabili dal Marketplace presente sul sito di Pentaho gratuitamente, oppure tramite lo sviluppo di propri plug-in tramite l'utilizzo del software Pentaho Marketplace.

Pro

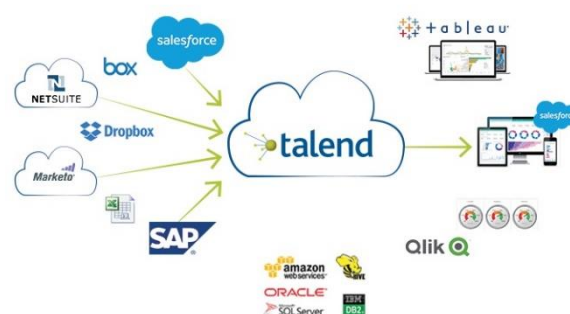
- Interfaccia Intuitiva
- Breve tempo di apprendimento
- Community numerosa
- Disponibile per Windows e Unix

Contro

- Basata su un interprete di file XML
- Più lento di altre soluzioni

3.2.2 Talend Open Studio

Talend Open Studio for Data Integration (TOS) è lo strumento ETL della software house Talend. Anche questo prodotto è offerto in più versioni: una versione base, open source e liberamente scaricabile e utilizzabile, una versione Enterprise, a canone annuale, e una versione Platform, che include anche altri strumenti, anch'essa a canone annuale.



Talend offre inoltre altri strumenti per l'integrazione dei dati come Big Data Integration, Cloud Integration, Application Integration, Data Quality e Master Data Management. Il prodotto provato è Talend Open Studio for Data Integration nella versione open source.

Come PDI, TOS offre un'intuitiva interfaccia visuale drag-and-drop con la quale si possono realizzare processi ETL, anche molto complessi. In questo caso non è presente la distinzione tra Transformation e Job, ma il programma è diviso in sezioni all'interno delle quali si possono compiere diverse azioni.

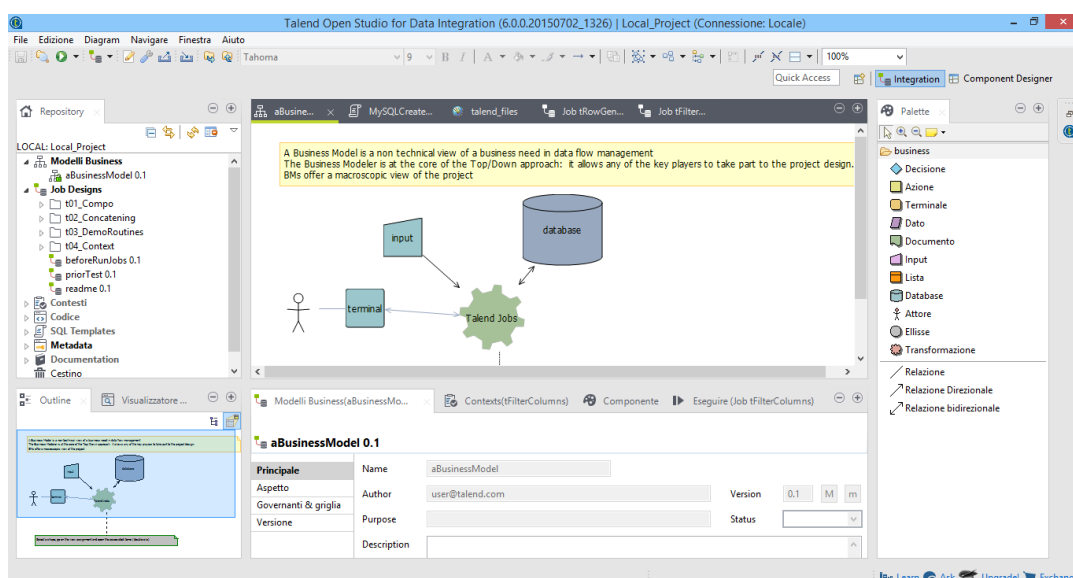


Figura 9 - Talend Open Studio for Data Integration. – IDE con esempio di Business Model

Nella prima sezione, denominata “Business Models”, è possibile realizzare modelli grafici di supporto al lavoro di ETL, così che ogni persona che prende parte al progetto possa avere una visione complessiva e dettagliata delle operazioni che dovranno essere eseguite.

Nella seconda parte, denominata “Job Designs” e vero cuore del programma, è possibile sviluppare, tramite drag-and-drop, i Job, cioè sequenze di operazioni di ETL, logiche di controllo, operazioni di avviso o invio di file, esecuzioni di altri job. I componenti messi a disposizione sono davvero numerosi e comprendono oltre 800 connettori che permettono di connettersi nativamente ad altrettanti tipi di database differenti. Inoltre sul sito talendforge.org, che raggruppa l’intera community di Talend, è possibile condividere e scaricare ulteriori componenti, connettori, temi, jobs e altri moduli gratuitamente, oltre a trovare tutorial, traduzioni, il bugtracker e il forum ufficiale Talend.

Si trova poi la sezione “Contexts”, nella quale si possono definire dei “Contesti”, contenitori di variabili (che possono essere valori numerici, stringhe, liste di valori, directory e altri oggetti) da utilizzare all’interno dei Job o delle procedure java.

Quest’ultime si possono definire nella sezione “Code”, nella quale è possibile realizzare delle procedure (scritte in java) da inserire all’interno dei Job.

Altre sezioni sono la “SQL Templates”, in cui si possono sviluppare query predefinite (per esempio per la creazione o l’eliminazione di tabelle e database), la sezione “Metadata”, in cui vengono definiti i metadata delle sorgenti dati utilizzate, e la sezione “Documentation”, in cui si può inserire la documentazione del progetto.

Al contrario di PDI, TOS non genera codice XML da eseguire tramite un interprete, ma genera direttamente il codice Java da eseguire per effettuare le operazioni inserite all’interno dei Job. In questo modo, a fronte di un codice generato generalmente più complesso, si hanno una maggiore velocità e la portabilità offerta da Java rispetto all’utilizzo di Job in formato XML, che invece necessita obbligatoriamente di un interprete (Kettle nel caso di PDI) e che, data la natura interpretata, risulta più lento nella maggior parte dei casi, in particolare a fronte di Job molto complessi.

Pro

- Esecuzione di Job rapida e portabile grazie all’utilizzo di codice Java nativo
- Elevata disponibilità di componenti aggiuntivi
- Interfaccia Intuitiva
- Community numerosa
- Disponibile per Windows e Unix

Contro

- Tempo di apprendimento più lungo rispetto ad altre soluzioni
- Funzioni avanzate disponibili solo nella versione Enterprise

3.2.3 Jaspersoft Community Edition

Jaspersoft è la suite di Business Intelligence di Tibco. È formata da diversi componenti: JasperReports Server, JasperReports Library, Jaspersoft Studio e Jaspersoft ETL.

Con la versione Community, gratuita e totalmente open source, è possibile realizzare report, grafici, tabelle dinamiche e/o pivot, ma non è possibile realizzare dashboard e, nonostante la visualizzazione sia possibile dall’interfaccia web integrata in JasperReports Server, non si possono integrare i documenti realizzati all’interno di altre web application.

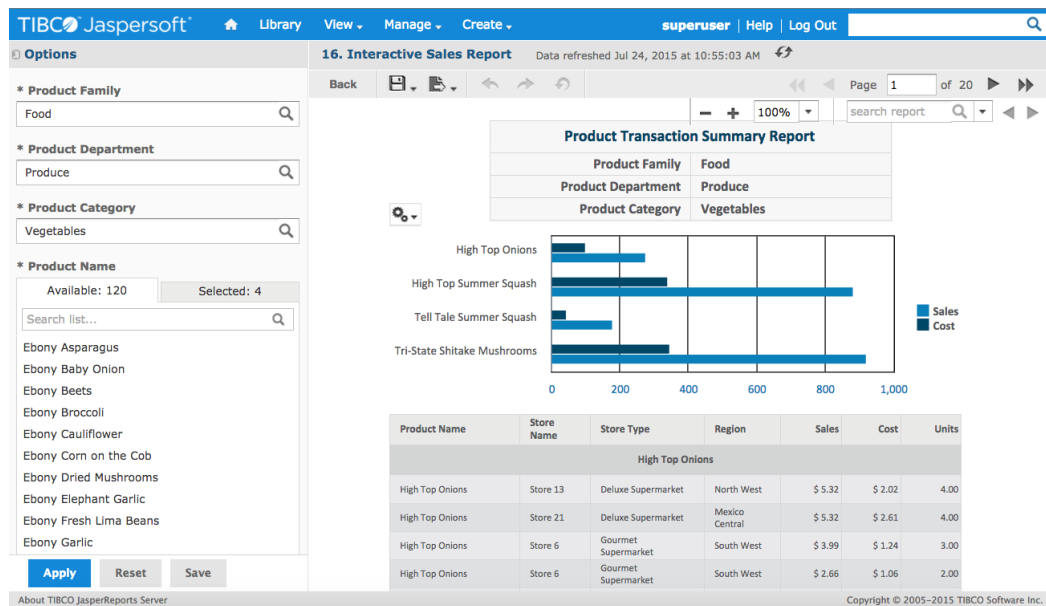


Figura 10 - JasperReports Server - Esempio di report

Scendendo nel dettaglio, JasperReports Server è il motore di rendering alla base della suite: gestisce l'esecuzione e la visualizzazione di report e altri documenti di analisi, permette la visualizzazione e la navigazione dei documenti direttamente da browser tramite la propria interfaccia web, consente di impostare l'automatica esecuzione e distribuzione di report tramite lo scheduler integrato e gestisce l'accesso ai documenti presenti sul server in base a utente e ruolo. Ulteriori funzioni sono presenti nelle versioni commerciali.

I documenti sono realizzati con Jaspersoft Studio; l'editor è basato su Eclipse ed offre un'interfaccia intuitiva e completa che permette di realizzare la maggior parte dei documenti senza dover metter mano al codice XML generato. I documenti, infatti, sono salvati in un unico file in formato XML che viene automaticamente generato dall'editor visuale. È comunque possibile modificare il codice XML dei documenti realizzati tramite il tool grafico come anche scrivere direttamente il codice XML del documento che si vuole realizzare.

Con JasperReports Library è possibile eseguire, all'interno dell'applicazione, i documenti realizzati con Jaspersoft Studio senza bisogno di installare l'intero server. Applicazione utile per chi deve realizzare nuovi documenti e testarli velocemente prima del caricamento sul server.

Jaspersoft ETL, invece, è lo strumento di Extract, Transform and Load messo a disposizione all'interno della suite. È basato su Talend Open Studio (quindi per ogni considerazione dello strumento rimando al paragrafo precedente) a cui aggiunge qualche tool specifico per l'esecuzione di operazioni su JasperReports Server.

Pro

Suite Completa
Editor, basato su Eclipse, intuitivo e completo
Community numerosa

Disponibile per Windows, Mac e Linux

Contro

Cruscotti (Dashboards) non presenti
Impossibile integrare i documenti realizzati in pagine web esterne
Molte funzionalità presenti solo nella versione commerciale

3.2.4 SpagoBI Suite

SpagoBI è la suite open source di Business Intelligence dell'italiana Engineering Ingegneria Informatica S.p.a.. Nella suite sono inclusi SpagoBI Server, SpagoBI SDK, SpagoBI Studio e SpagoBI Meta. Il prodotto è open source e gratuito e non esistono versioni commerciali poiché il business è stato incentrato sulla vendita di corsi e libri relativi all'utilizzo del programma. Il programma scaricabile dal sito, quindi, è la versione completa comprendente tutte le funzionalità sviluppate ed è liberamente utilizzabile.



Figura 11 - SpagoBI Server - Pagina Iniziale

Include al suo interno numerosi prodotti open source esterni, quali il motore di rendering JasperReports, il motore di rendering BIRT e la libreria javascript HighCharts per la visualizzazione di grafici all'interno delle pagine web. Offre la possibilità di realizzare documenti di numerose tipologie differenti, tra cui report, grafici interattivi, dashboard interattive, tabelle pivot (OLAP cubes) e documenti KPI (Key Performance Indicator). È inoltre possibile integrare i documenti realizzati all'interno di pagine web esterne mediante l'utilizzo di iframe.

Per questi motivi e per le notevoli possibilità di personalizzazione, questo è il prodotto che è stato scelto come strumento per la Business Intelligence all'interno di Soft-Net Srl. Per la trattazione dettagliata del programma rimando quindi al capitolo 4 che è interamente dedicato alla suite SpagoBI.

Pro

- Suite Completa
- Nessuna versione commerciale
- Interamente gratuito ed open source
- Personalizzabile e integrabile in web application esterne
- Disponibile per Windows e Unix

Contro

- Community poco numerosa
- Documentazione scarsa

3.3 Prodotti Commerciali

3.3.1 Jaspersoft Studio Commercial Edition

Jaspersoft è la versione commerciale della suite Jaspersoft Community Edition. È disponibile in diverse varianti: Reporting, Professional ed Enterprise, più la versione AWS (versione cloud hostata su Amazon Web Services).

Rispetto alla versione gratuita, la versione Reporting aggiunge solamente il supporto premium dedicato, mentre la versione Professional aggiunge anche la possibilità di integrare i documenti realizzati in pagine web esterne (tramite il framework proprietario Virtualize.js), la visualizzazione di tutti documenti tramite browser con codice HTML5 (senza l'utilizzo di Flash), l'Ad-Hoc Report Designer (strumento drag-and-drop per la realizzazione di report da parte di utenti non tecnici), le dashboard ed alcune funzioni per l'analisi di dati. Infine la versione Enterprise aggiunge a queste funzionalità la gestione multi-utente con permessi specifici e l'Audit Logging (analisi di come gli utenti utilizzano la piattaforma e le prestazioni della stessa).

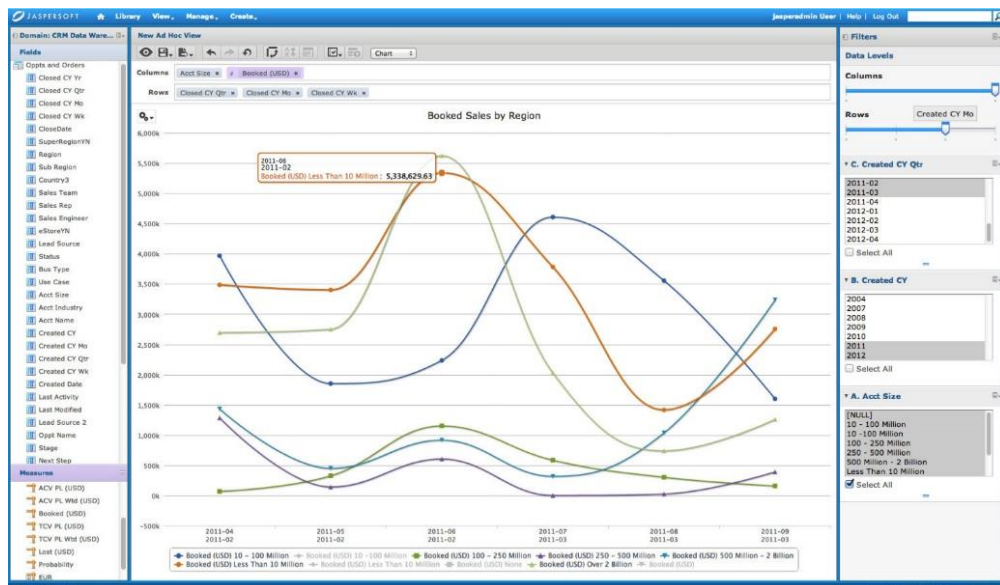


Figura 12 - Jaspersoft - Esempio di grafico

Non è disponibile un listino prezzi sul sito, ma bisogna contattare un commerciale di Tibco per ricevere informazioni riguardanti il prezzo. Purtroppo, per il target scelto il prezzo di partenza proposto è risultato troppo alto.

Pro

- Uguali a quelli della versione Community
- Dashboard disponibili e grafica accattivante
- Supporto tecnico incluso

Contro

- Non sono presenti informazioni riguardanti il prezzo sul sito
- Riscontrate difficoltà nel trovare un commerciale disponibile
- Prezzo troppo elevato per le PMI: per la versione Professional si parla di circa 20k € di sottoscrizione annuale per CPU core

3.3.2 Tableau

Tableau non è una suite completa di BI, ma un ottimo tool per la visualizzazione e l'analisi interattiva di dati. Il prodotto è disponibile in diverse varianti: Tableau Desktop, che permette di realizzare e visualizzare dashboard interattive sul proprio PC o Mac, Tableau Server, che permette di condividere le dashboard realizzate tramite Tableau Desktop con altri utenti e di visualizzarle e modificarle online tramite browser, e Tableau Online, versione cloud di Tableau Server. È disponibile anche una versione aggiuntiva, chiamata Tableau Public, che equivale ad una versione gratuita e pubblica di Tableau Server.

Data la necessità di condividere i documenti realizzati e, in particolare, di integrarli all'interno di altre pagine web, l'analisi si è focalizzata sulla coppia Tableau Desktop e Tableau Cloud.

Tableau Desktop è un potente software di analisi dei dati che permette, tramite un'intuitiva interfaccia grafica, di connettersi a numerosi tipi di sorgenti dati e di realizzare analisi dei dati tramite grafici, tabelle e tabelle pivot, in maniera estremamente semplice e veloce, dinamica ed attraverso una grafica accattivante. Una volta che la connessione con la base di dati è stata effettuata, tramite un intuitivo tool drag-and-drop vanno selezionate le tabelle e/o i file necessari per il documento da realizzare e le relazioni fra di essi. A questo punto è possibile iniziare l'analisi, sempre con un semplice e completo tool drag-and-drop, impostando righe, colonne, ulteriori misure, tipo di grafico ed eventuali filtri in un IDE rapido ed interattivo. Si possono anche realizzare dashboard interattive contenenti più documenti (filtrabili a vicenda) al loro interno.

Row ID	Alt Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Alt Customer Name	Alt Ship Mode	Alt Customer Segment	Alt Product Category	Alt Product
1319	Low	0.050000	5.850	2.2700	21	Tony Wilkins Winters	Regular Air	Small Business	Office Supplies	Pens &
1910	Medium	0.010000	4.910	0.5000	117	Linda Weiss	Regular Air	Home Office	Office Supplies	Labels
1911	Medium	0.090000	4.000	1.3000	117	Linda Weiss	Express Air	Home Office	Office Supplies	Paper
5055	High	0.020000	3.280	5.0000	117	Linda Weiss	Regular Air	Home Office	Office Supplies	Pens &
1573	Low	0.020000	11.970	1.8100	272	Eleanor Swain	Regular Air	Small Business	Office Supplies	Pens &
1574	Low	0.030000	159.310	60.0000	272	Eleanor Swain	Delivery Truck	Small Business	Furniture	Tables

Figura 13 - Tableau - Esempio di Join fra due gruppi di dati

Una volta realizzato, il documento può essere salvato sul computer e/o caricato su Tableau Server/Tableau Cloud. Tramite l'interfaccia web-based, è possibile visualizzare i documenti presenti sul server Tableau e modificarli direttamente online (con funzionalità però molto limitate rispetto all'editor Desktop). È inoltre possibile impostare determinati permessi agli utenti, in modo che possano accedere solo a determinate aree.



Figura 14 - Tableau - Esempio di dashboard su differenti dispositivi

È possibile integrare i documenti realizzati in altre pagine web, ma con alcune limitazioni: è possibile inserire dei documenti in altre web application, ma è necessario che l'utente acceda, se non già autenticato, da una schermata di login specifica. Non si possono quindi inserire tali documenti in altre applicazioni dedicate già dotate di proprio login, a meno di richiedere all'utente una doppia autenticazione: quella della web application e quella di Tableau. Questa applicazione è stata scartata proprio per questo motivo.

Pro

Diverse versioni fra cui poter scegliere

IDE completo e ben realizzato

Grafica accattivante e funzionale, probabilmente una delle migliori disponibili attualmente

Contro

Non sono presenti informazioni riguardanti il prezzo sul sito

Prezzo elevato, ma ancora accessibile (2k € per Tableau Desktop e 10k € per Tableau Server con 10 utenti il primo anno, poi il 20% di manutenzione annuale oppure 500€/anno/utente per Tableau Online)

Impossibile integrare i documenti realizzati in pagine web esterne senza richiedere l'autenticazione ad ogni accesso

Sia Tableau Desktop che Tableau Server sono disponibili solo per Windows e MacOS (le distribuzioni Linux non sono supportate)

3.3.3 Pentaho Enterprise Edition

Pentaho (Enterprise Edition) è la versione commerciale e completa della suite Pentaho.

Offre strumenti completi per la gestione dei Big Data, l'integrazione dei dati e la loro analisi e visualizzazione.

Il software è sviluppato in Java e quindi compatibile sia con Windows che con Unix, inoltre l'architettura permette l'estensione delle funzionalità tramite plug-in e l'integrazione in altre applicazioni web tramite API dedicate (Application Programming Interface).

Sono disponibili separatamente per l'acquisto Pentaho Data Integration, versione

commerciale di Pentaho Data Integration Community Edition, e Pentaho Business Analytics, che comprende PDI e tutti gli altri strumenti messi a disposizione per l'analisi e l'integrazione dei (big) data.

Con Pentaho Business Analytics, è possibile realizzare documenti di diverso tipo, fra cui report, grafici, tabelle pivot, mappe dinamiche e dashboard interattive. Se caricati sul server Pentaho, è possibile visualizzare i documenti tramite l'interfaccia web del server (che sfrutta le web API messe a disposizione anche per l'integrazione con altri prodotti) e modificarli direttamente online. È anche possibile passare dalla visualizzazione dei dati tramite grafico alla visualizzazione tabellare in tempo reale, tramite pochi semplici passaggi. Si possono inoltre inserire i documenti realizzati in altre pagine web con due modalità differenti: tramite iframe, che garantisce la corretta esecuzione e visualizzazione dei documenti scelti a costo di avere un ambiente completamente separato dalla pagina principale, e tramite web object, che permette di avere un ambiente condiviso fra pagina web e documento realizzato a costo di maggiori rischi di incompatibilità fra i due oggetti (in particolare del codice javascript).

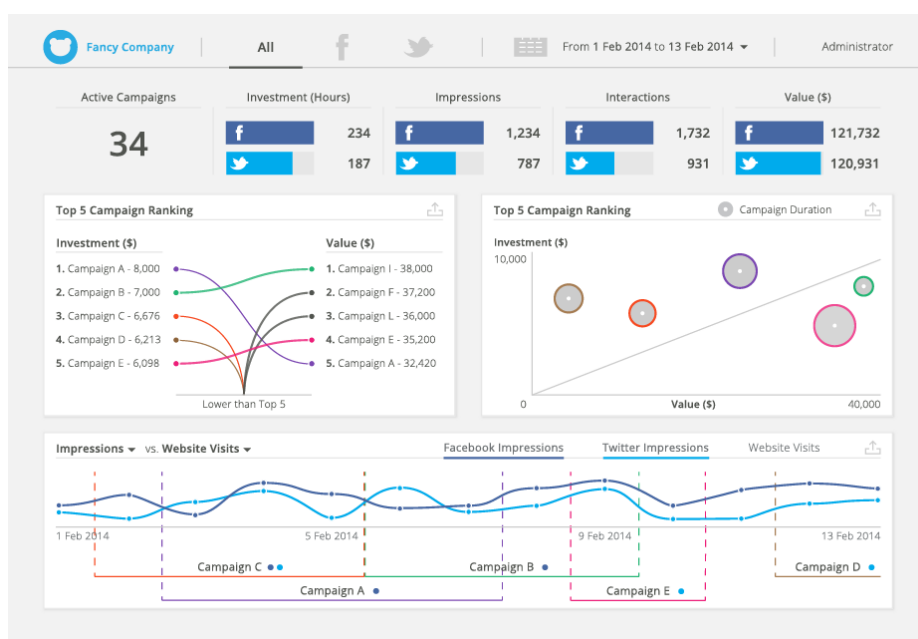


Figura 15 - Pentaho Enterprise Edition - Esempio di dashboard

Non sono disponibili informazioni riguardanti il prezzo sul sito dell'azienda, ma è possibile scaricare una versione di prova della durata di 30 giorni e richiedere una quotazione. Purtroppo, contattati per una valutazione, i prezzi proposti sono risultati eccessivi per il target scelto (PMI) quindi, nonostante la sua completezza, il prodotto è stato scartato.

Pro

- Disponibili API per la connessione da web application esterne
- Community numerosa e supporto tecnico incluso nella licenza
- IDE dedicato sviluppato ad hoc

Supporto sia per Windows che Unix

Contro

- Grafica sopra la media, ma non sempre all'altezza
- Non sono presenti informazioni riguardanti il prezzo sul sito
- Prezzo eccessivo per il target scelto (PMI): sottoscrizione annuale a partire da 30k €

3.4 Tabella riassuntiva

3.4.1 Prodotti Open Source

	Funzionalità	Interfaccia	IDE	Compatibilità	Documentaz.	Community
<i>PDI</i>	●●●●	-	●●●●	●●●	●●●●	●●●●
<i>TOS</i>	●●●●●	-	●●●●●	●●●●●	●●●●●	●●●●●
<i>Jaspersoft CE</i>	●●	●●	●●●●	●●●●●	●●●●●	●●●●●
<i>SpagoBI</i>	●●●●	●●●	●●●	●●●●●	●●	●●●

3.4.2 Prodotti Commerciali

	Funzionalità	Interfaccia	IDE	Compatibilità	Documentaz. e community	Supporto	Prezzo
<i>Jaspersoft</i>	●●●●●	●●●●	●●●●	●●●●●	●●●●●	●●●●	●●
<i>Tableau</i>	●●●●●	●●●●●	●●●●●	●●●	●●●●	●●●●	●●●
<i>Pentaho</i>	●●●●	●●●	●●●●	●●●●●	●●●●	●●●●●	●

4 La Suite SpagoBI

4.1 Introduzione

SpagoBI suite è una suite di Business Intelligence totalmente gratuita e open source, particolarmente flessibile ed adattabile alle necessità dell'utente. È la suite scelta durante il tirocinio per la realizzazione di documenti di reportistica.

La suite è strutturata a moduli il cui insieme forma un prodotto completo e funzionale che copre i settori che vanno dalla trasformazione dei dati (ETL) alla realizzazione di analisi multidimensionali (OLAP), dalla realizzazione di report all'analisi dei dati in tempo reale. La struttura a moduli permette, inoltre, di installare solo le parti necessarie ai propri scopi, senza l'obbligo di utilizzare la suite completa, mentre la disponibilità dei codici sorgenti e la realizzazione della parte Server in HTML, CSS e Java consente un ampio margine di personalizzazione del prodotto.

Inoltre, tutti i componenti della suite sono basati su Java quindi sono, in linea di massima, compatibili con qualsiasi sistema sul quale è possibile installare la Java Virtual Machine. Per l'esecuzione è necessario avere una versione funzionante di Java 7 o superiore.

Nota: all'interno del capitolo, con il termine "documenti" o "documenti analitici" si fa riferimento all'insieme di report, grafici, cruscotti e tabelle pivot ed in generale tutti i documenti di reportistica realizzabili con SpagoBI Studio e visualizzabili tramite SpagoBI Server in quanto questo è il termine utilizzato dalla suite.

4.2 Componenti



Figura 16 - Infrastruttura della suite SpagoBI

I moduli principali di cui si compone sono:

- SpagoBI Server: è il cuore dell'intera suite e integra tutti i motori e le funzionalità necessarie al funzionamento di SpagoBI
- SpagoBI Studio: è l'IDE della suite, è basato su Eclipse e consente di configurare connessioni a database e dataset, realizzare operazioni ETL e creare documenti analitici (report, grafici, cruscotti)
- SpagoBI Meta: è l'IDE, basato su Eclipse, che comprende i tool di gestione di database e dataset e trasformazione dei dati (è incluso in SpagoBI Studio)
- SpagoBI SDK: è il layer che permette di connettere SpagoBI con tool esterni
- SpagoBI Applications: sono una collezione di "Analytical Models" sviluppati utilizzando SpagoBI e liberamente scaricabili

Per le operazioni ETL (Extract, Transform, Load), cioè le operazioni di estrazione e trasformazione dei dati, SpagoBI consiglia l'utilizzo di Talend Open Studio for Data Integration (TOS).

4.3 SpagoBI Server

SpagoBI Server è il cuore dell'intera suite. Questo componente gestisce e memorizza al suo interno tutti i dati relativi ad utenti, permessi, reportistica, connessione alle sorgenti dati e impostazioni del server stesso.

La gestione del server avviene tramite un'interfaccia web mediante la quale è possibile non solo visualizzare la reportistica realizzata tramite SpagoBI Studio e caricata sul server, ma anche definire Sorgenti Dati, Data set e parametri e modificare le impostazioni di visualizzazione e accesso degli utenti.



Figura 17 - Interfaccia Web di SpagoBI Server - Pagina Iniziale

4.3.1 Installazione e configurazione

4.3.1.1 Download

SpagoBI si appoggia sull'hosting offerto da OW2 Forge, un organizzazione no-profit internazionale, per garantire trasparenza, sostenibilità e disponibilità del software. Per questo motivo, tutti i sorgenti e i file compilati della suite dalla versione 3 in poi si trovano all'indirizzo http://forge.ow2.org/project/showfiles.php?group_id=204.

I file utilizzati per l'installazione sono i seguenti:

- "All-In-One-SpagoBI-5.1-21012015.zip" che racchiude tutti i file eseguibili di SpagoBI Server per Windows e Linux per la parte Server
- "SpagoBIStudio_5.1.0_win64_19012015.zip" che contiene SpagoBI Studio compilato per Windows, versione a 64bit.

4.3.1.2 Installazione di SpagoBI Server su Debian 8 64bit

Una volta scaricato il file compresso, è sufficiente estrarre il contenuto in una cartella accessibile dall'utente (nel nostro caso /opt) e procedere alla configurazione del server.

4.3.1.3 Configurazione di SpagoBI Server

Innanzitutto è necessario modificare il file di configurazione base del server /opt/SpagoBI-5.1.0/conf/server.xml impostando il nome dell'host su cui si sta installando la macchina, la porta su cui avviare il servizio ed eventuali risorse aggiuntive (per es. database hsqldb).

All'interno del file /opt/SpagoBI-5.1.0/conf/web.xml si possono personalizzare le impostazioni del web server, in particolare è stato necessario modificare il tempo di timeout delle sessioni ("session-timeout") per portarlo da 20 minuti a 10 ore, così da coprire l'intera giornata lavorativa.

È stato inoltre necessario modificare alcuni file del template html utilizzato per risolvere alcuni problemi di visualizzazione presenti all'interno del programma nella versione 5.1.

4.3.1.4 Avvio di SpagoBI Server

Se non sono stati commessi errori nella configurazione e non sono presenti errori relativi ai permessi di lettura ed esecuzione, lanciando il comando "sh /opt/SpagoBI-5.1.0/bin/SpagoBIStartup.sh" verrà avviato il server e l'interfaccia web dell'applicazione.

4.3.2 Funzionalità ed Utilizzo

4.3.2.1 Accesso

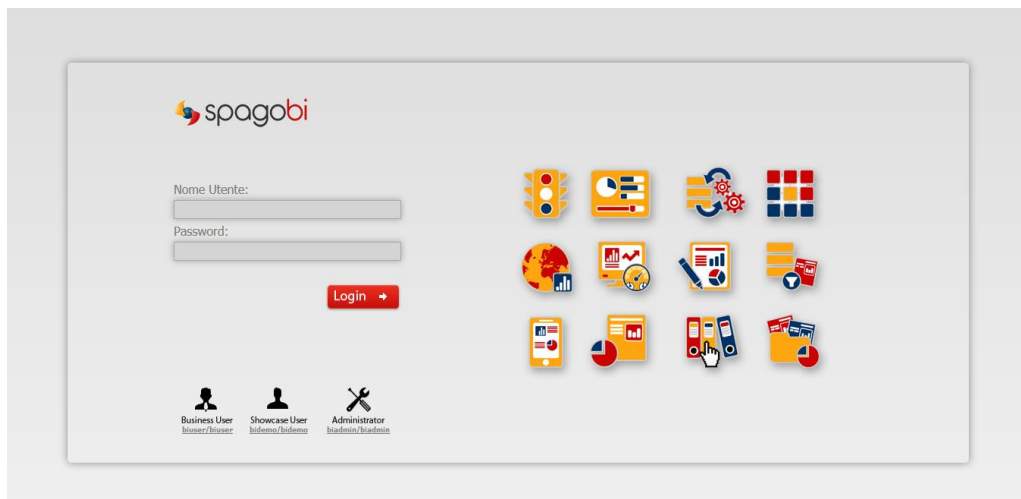


Figura 18 - Interfaccia Web di SpagoBI Server - Pagina di login

Aperto la pagina [http://\[indirizzo web server\]:\[porta\]/SpagoBI/](http://[indirizzo web server]:[porta]/SpagoBI/) all'interno di un browser, si accede all'interfaccia web del prodotto. La prima volta si può accedere utilizzando una delle credenziali predefinite presenti nella parte bassa della pagina per poi creare un proprio utente e relativa password ed eliminare gli utenti predefiniti.

4.3.2.2 Gestione Utenti

I permessi degli utenti sono gestiti tramite Utenti e Ruoli. Dalla pagina Gestione Ruoli si possono configurare i ruoli che saranno successivamente associati all'utente, impostando nome del ruolo, il tipo e le autorizzazioni di cui dispone.

Dalla schermata Gestione Utenti, invece, è possibile aggiungere o eliminare utenti e modificarne le proprietà, tra cui il nome utente, il nome completo, la password, altri campi

personalizzati e i ruoli di quell'utente. In questo modo l'utente avrà accesso solo alle funzioni e ai documenti a cui possono accedere i ruoli che possiede.

4.3.2.3 Data Source

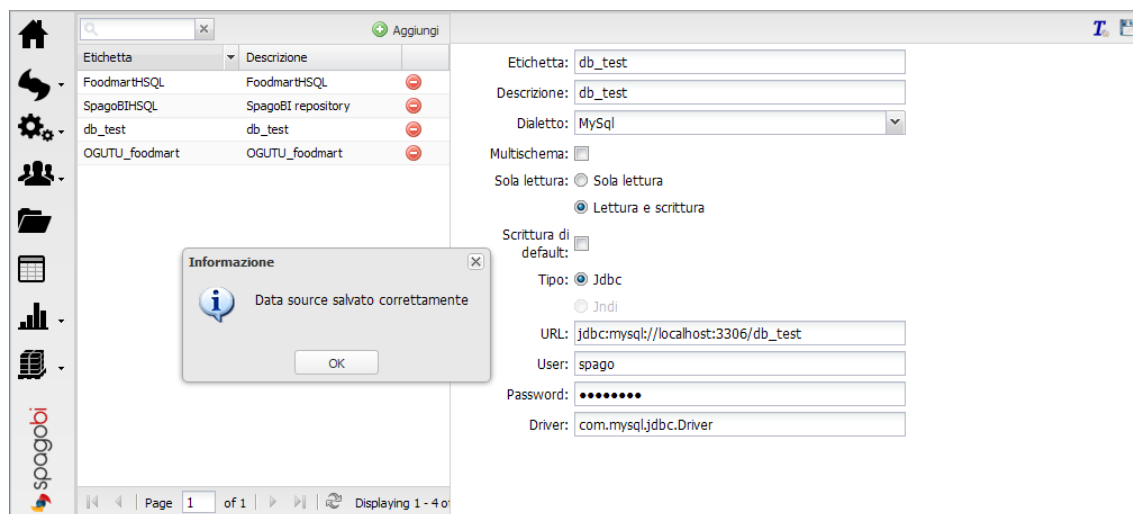


Figura 19 - Interfaccia Web di SpagoBI Server - Data Source, esempio di configurazione

Uno dei punti di forza dell'applicazione è la capacità di connettersi a molteplici tipologie di database server fra cui troviamo Oracle, SQL Server, HSQL, MySQL, PostgreSQL, DB2 (AS400), MongoDB, Apache HBase, OrientDB.

All'interno dei Data Source sono salvati i parametri di connessione al database. Dalla sezione dedicata è possibile inserire e modificare Data Source impostando un nome univoco e i parametri specifici, in particolare il tipo di dialetto, l'url del server DBMS, i dati di accesso, il driver da utilizzare ed il tipo di accesso (Sola Lettura, Lettura/Scrittura). È possibile testare la connessione tramite il pulsante "T" (test), presente in alto a destra a fianco del pulsante Salva con il quale invece si salva definitivamente la configurazione.

4.3.2.4 Data Set

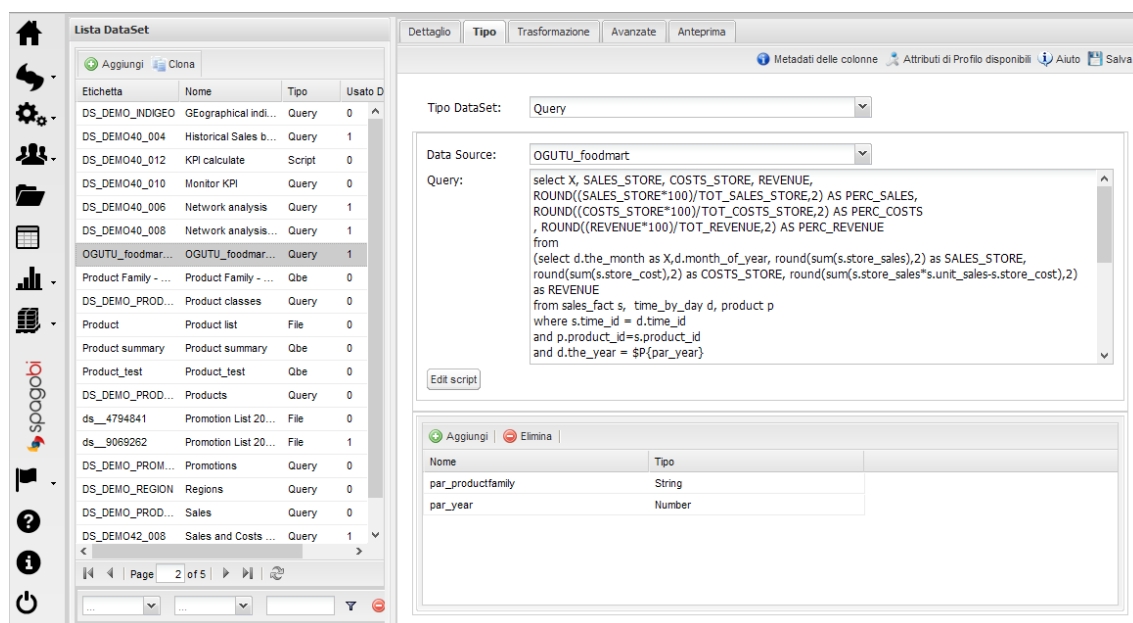


Figura 20 - Interfaccia Web di SpagoBI Server - Data Set

I Data Set sono set di dati che saranno utilizzati come sorgente dati nei documenti analitici caricati sul server. Possiamo considerare i Data Set come delle viste a cui accedono reportistica e cruscotti per recuperare i dati da visualizzare.

Dalla schermata Dettaglio è possibile configurare i parametri generali (nome, descrizione, visibilità, ecc.) e visualizzare la cronologia delle modifiche, mentre dalla schermata Tipo si definiscono la natura e le proprietà specifiche del data set. Esso può essere di diverse tipologie e, in base a quella scelta, saranno visualizzate impostazioni di configurazione differenti. Inoltre, nella stessa scheda è presente un elenco contenente la lista di parametri utilizzabili all'interno del Data Set e che saranno passati, durante l'esecuzione, dal documento analitico.

I tipi di Data Set più rilevanti sono:

- “File”, recupera i dati da un elenco contenuto nel file di testo indicato
- “Query”, recupera i dati eseguendo la query specificata sul Data Source indicato
- “Web Service”, recupera i dati da un servizio di rete remoto
- “QbE” (Query by Editor), strumento che permette di realizzare query SQL su un database tramite editor visuale e di memorizzarle come Data Set

4.3.2.5 Document Browser

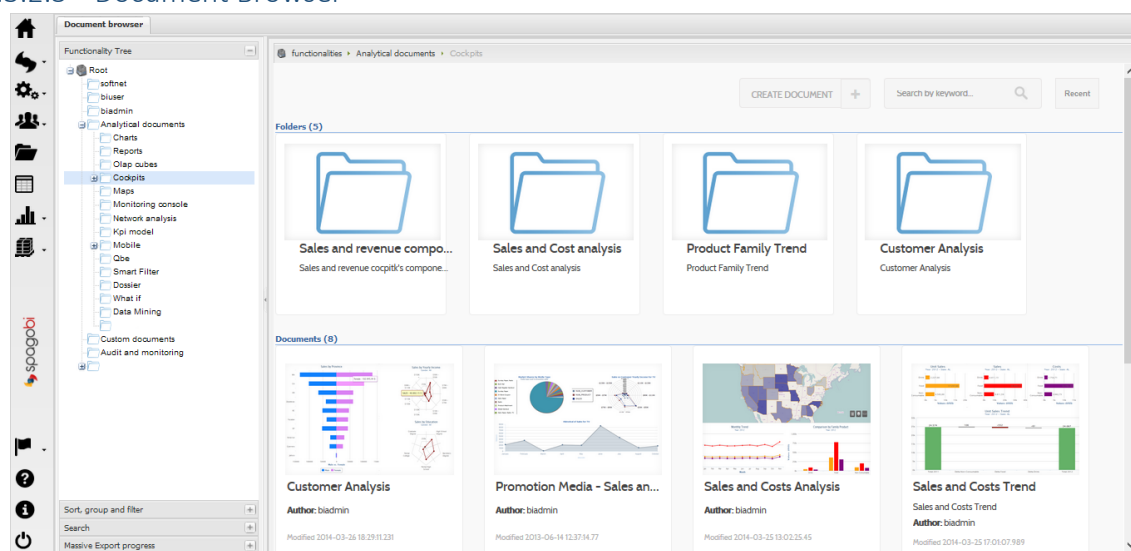




Figura 21 - Interfaccia Web di SpagoBI Server - Document Browser

Questa sezione consente di visualizzare i documenti analitici (report, grafici, cruscotti) caricati sul server e di navigare fra le cartelle disponibili tramite un'interfaccia che ricorda la navigazione con Esplora Risorse di Windows (o Finder di MacOS).

Se l'utente dispone di sufficienti permessi, per ogni documento è possibile eseguire più azioni. Cliccando sul nome o sull'immagine di anteprima del documento, esso viene caricato e visualizzato all'interno del browser in una scheda adiacente a quella del Document Browser dando la possibilità di aprire più documenti analitici contemporaneamente.

Cliccando sull'icona  si apre, invece, un popup con i metadati del documento, mentre cliccando sull'icona  si apre la pagina di configurazione del documento. Da questa pagina è possibile modificare tutte le impostazioni del documento, in particolare si può modificare il nome, impostare il tipo di documento e il relativo motore di rendering

(Engine), caricare un nuovo file (XML) di template, modificare le impostazioni di visibilità e impostare i parametri del documento stesso. Ogni parametro è caratterizzato da un titolo, un nome identificativo (che deve essere uguale al nome utilizzato all'interno del file XML e nella definizione del Data Set), da alcune impostazioni di visualizzazione e da un Driver Analitico (quest'ultimo verrà trattato nel dettaglio in seguito).

Tutti i documenti sono contraddistinti da file di template in formato XML, un tipo (per es. Report, Olap, Cockpit) e un motore (Engine) il cui compito è quello di elaborare il file XML contenente il template del documento e generare il codice HTML/CSS/Javascript necessario per la sua visualizzazione. Il template in formato XML viene solitamente realizzato e caricato sul server tramite SpagoBI Studio.

Per quanto riguarda i parametri, invece, essi svolgono solitamente una funzione di filtro sui dati caricati dal Data Set e permettono di realizzare documenti dinamici nei quali i dati mostrati possono variare durante la visualizzazione. Inoltre, i parametri permettono di realizzare documenti multi-livello (detti anche navigabili) cioè documenti nei quali, la selezione di una parte specifica dello stesso (per esempio una porzione di dati di un grafico a torta) apre un nuovo documento che mostra i dettagli della sezione di dati su cui si è premuto. Anche nei cruscotti è possibile aggiornare, in seguito alla selezione di una porzione di dati di un documento, tutti gli altri documenti impostando come parametro il valore o il dominio della porzione di dati selezionata.

4.3.2.6 Analytical Driver

I Driver Analitici sono strutture dati indipendenti dai documenti analitici che modellano il concetto di parametri. Ogni driver analitico si compone di un nome, del tipo di dato che vuole rappresentare e di diversi LOV, ovvero List Of Values, che si possono associare al driver analitico configurandone le impostazioni che comprendono il nome della variabile, il metodo di visualizzazione dei valori del LOV e un eventuale valore di default.

I LOV contengono una lista di valori validi per i driver analitici a cui sono collegati. Essi possono essere definiti tramite una lista statica di valori, tramite una query che estrae i valori dal data source specificato, da uno script javascript che restituisce una lista di valori oppure da un data set configurato precedentemente nell'apposita sezione.

Dopo aver definito Driver Analitici e relativi LOV, è possibile associare i driver ai parametri di uno o più documenti. In questo modo, i driver forniscono ai parametri una lista di valori, indipendenti dal documento in cui il driver è utilizzato, che essi possono assumere.

4.4 SpagoBI Studio

SpagoBI Studio è l'ambiente di sviluppo della suite. È basato su Eclipse, un IDE "famigliare" e conosciuto per molti sviluppatori.

Con SpagoBI Studio si possono creare e modificare tutti i tipi documenti presenti nella suite tramite un interfaccia grafica che risulta non sempre intuitiva e che spesso richiede la modifica diretta del codice sorgente dei documenti di reportistica realizzati. Il codice sorgente utilizzato da SpagoBI per la definizione di report e visualizzazioni è il codice XML il quale viene interpretato dal Server tramite diversi motori, uno per ogni tipo diverso di documento realizzabile.

4.4.1 Installazione di SpagoBI Studio su Windows 8.1 64bit

Una volta scaricato il file compresso, è sufficiente estrarlo in una cartella accessibile all'utente; in questo caso non sono necessarie ulteriori operazioni e per avviare il programma basta eseguire il file SpagoBI.exe presente all'interno della cartella estratta.

4.4.2 Funzionalità ed Utilizzo



Figura 22 - SpagoBI Studio - Pagina iniziale

Come indicato in precedenza, SpagoBI Studio è utilizzato principalmente per lo sviluppo dei documenti analitici. I principali documenti realizzabili (saranno analizzati in seguito) sono documenti di report, grafici, grafici geografici e cruscotti interattivi (chiamati Cockpit). È anche utilizzato per la definizione di metadata relativi a data source e data set e i relativi modelli OLAP, sfruttati da SpagoBI Server per la gestione dei cubi OLAP e tabelle pivot. Tramite le funzioni integrate in SpagoBI Studio, inoltre, è possibile connettersi ad un Server SpagoBI e pubblicare e/o aggiornare i documenti creati in SpagoBI Studio, scaricare i documenti presenti sul server, scaricare modelli di data set e definire ulteriori sorgenti dati senza abbandonare lo strumento di sviluppo.

I file all'interno del progetto sono divisi in diverse sezioni: la sezione "Business Analysis" contiene tutti i documenti di analisi realizzati, mentre le sezioni "Business Models", "Business Queries" e "Olap Templates", presenti anche in SpagoBI Meta, contengono

rispettivamente i modelli delle sorgenti dati, le interrogazioni alle sorgenti dati e i template dei cubi OLAP. Infine, la sezione “Resources” contiene i parametri di connessione con i server SpagoBI.

Sono presenti diversi “Perspective”, ognuno associato ad una funzione differente e con pannelli e funzioni specifiche.

Il perspective principale è quello chiamato “SpagoBI Studio”, che permette la creazione e la modifica di documenti analitici e il loro caricamento sul server. All’interno di questo perspective ogni file può essere aperto in due modi: con l’editor di testo di Eclipse, che permette di modificare direttamente il codice XML dei documenti, o tramite l’editor di SpagoBI, che mette a disposizione un’interfaccia grafica con la quale configurare e realizzare report, grafici e cruscotti (di un livello di complessità non elevato).

Un altro perspective particolarmente utilizzato è quello chiamato “Report Designer”, che mette a disposizione una comoda interfaccia drag-and-drop per la realizzazione di report.

4.4.3 Grafici

SpagoBI Studio permette di realizzare grafici di diverse tipologie, come gli istogrammi, i grafici a torta, a linee, ad area, a radar e a dispersione. Il motore su cui sono basati è la libreria javascript HighCharts per cui si può integrare la (scarsa) documentazione di SpagoBI con quella presente sul sito di HighCharts. Sono disponibili altri motori per la realizzazione di grafici, ma presentano funzionalità e grafica di livello inferiore.

L’interfaccia grafica messa a disposizione per la realizzazione dei grafici è abbastanza completa e consente, nella maggior parte dei casi, di raggiungere risultati di buon livello; anche la grafica a livello di visualizzazione su SpagoBI Server è molto gradevole. Per realizzazioni più avanzate o con impostazioni personalizzate, invece, è necessario affidarsi alla modifica manuale del codice sorgente.

Nelle immagini qui proposte è possibile visualizzare la schermata visuale di configurazione e il relativo codice sorgente. Spesso, data la brevità del codice, è più conveniente (in termini di tempo) realizzare i grafici utilizzando l’editor testuale rispetto all’utilizzo dell’interfaccia grafica dedicata.

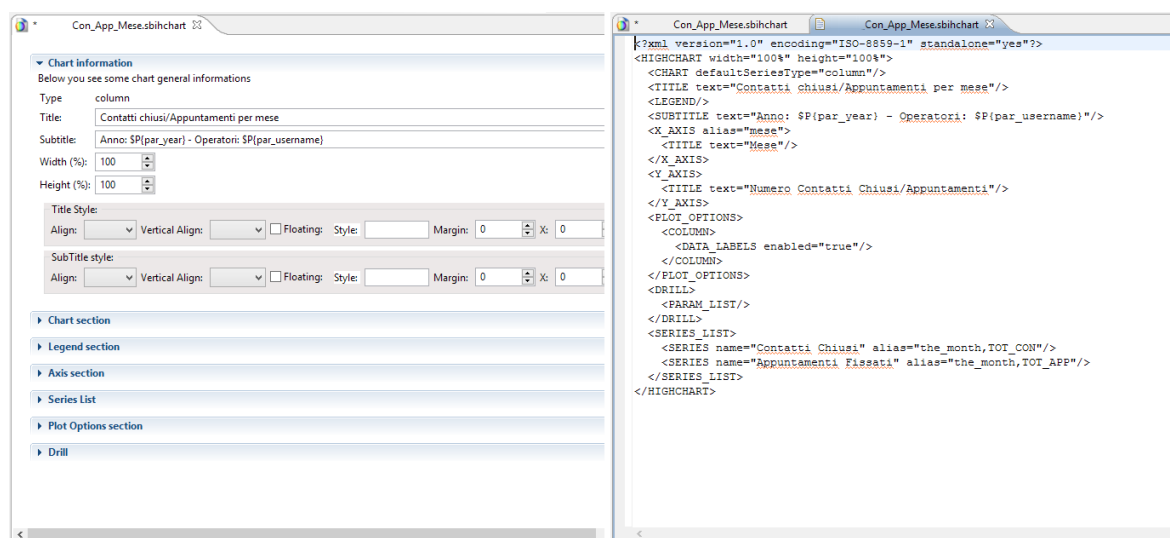


Figura 23 - Interfaccia per la realizzazione di grafici confrontata con il codice sorgente

Nella seguente immagine è possibile visualizzare il grafico risultante dal template mostrato nelle figure precedenti. Sono da notare i parametri (Driver Analitici) posizionati sulla destra.

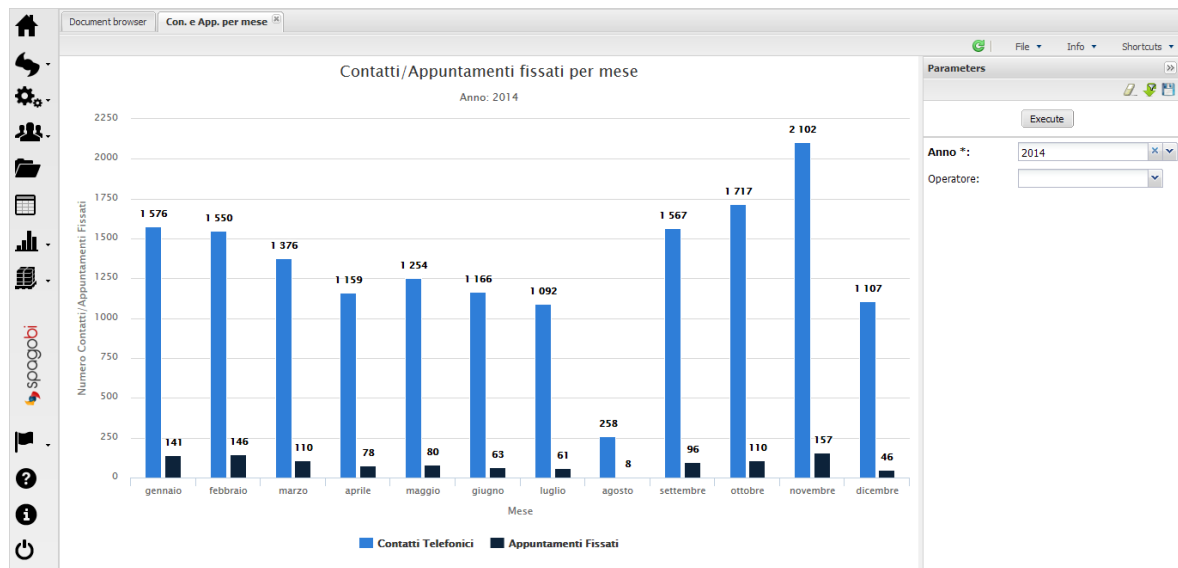


Figura 24 - Grafico risultante dal template della figura precedente

4.4.4 Report Statici

È possibile realizzare documenti di report (statici) con due differenti motori di rendering: ci si può basare sul motore Birt (open source e sviluppato da Eclipse) oppure sul motore JasperReports (versione appartenente alla suite Jaspersoft Community Edition analizzata in precedenza). In entrambi i casi è possibile inserire testi, immagini, tabelle e semplici grafici, ma la grafica risultante è sempre abbastanza scadente.

Nelle immagini seguenti si può vedere un esempio dell'editor presente per realizzare report basati sul motore Birt e il report risultante.

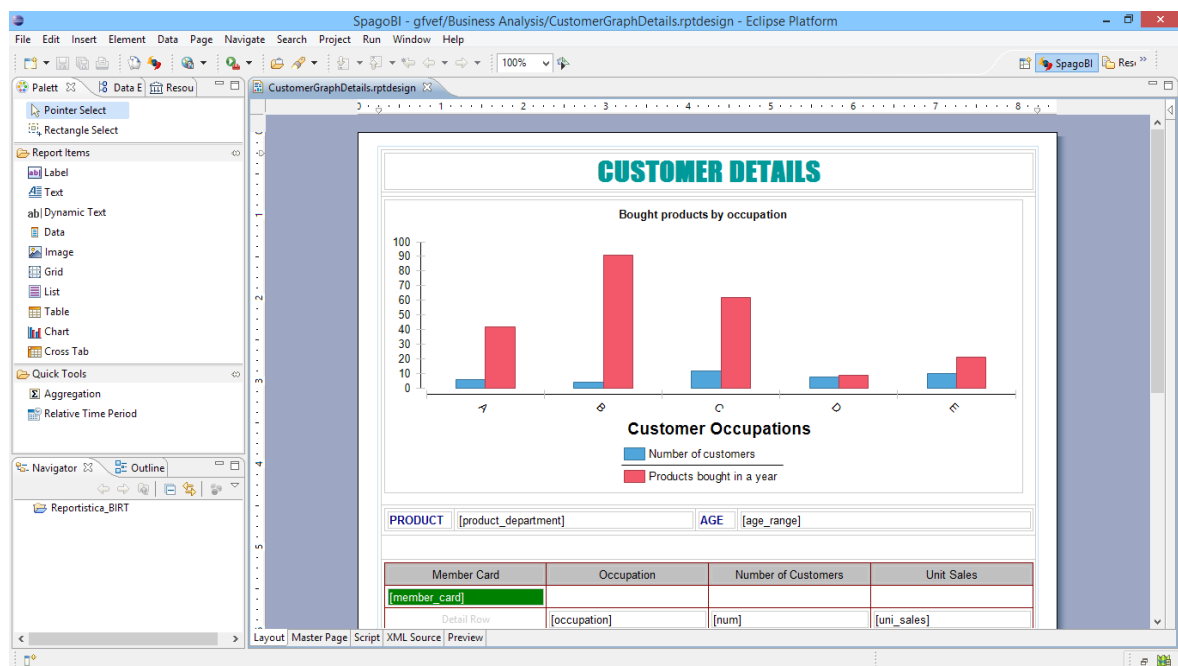


Figura 25 - Esempio di report realizzato con motore BIRT

4.4.5 Cruscotti (Cockpit)

Nella suite SpagoBI i cruscotti interattivi sono chiamati cockpit. Insieme alle tabelle pivot sono uno dei metodi di visualizzazione più utili presenti all'interno della suite poiché permettono di interagire e navigare dinamicamente fra i dati visualizzati.

Il motore utilizzato è il motore proprietario "Document Composition Engine" che, oltre alla visualizzazione di più documenti in un'unica schermata e all'utilizzo di parametri globali (comuni, cioè, a tutti i documenti presenti nel cockpit), permette di effettuare filtri e selezioni incrociate da un documento ad un altro direttamente all'interno del cockpit selezionando una porzione di dati da uno dei documenti presenti. Questa è una delle funzioni più utili dei cockpit perché estende la possibilità offerta dai cockpit di visualizzare più documenti (solitamente grafici) in un'unica schermata con la possibilità di filtrare i dati visualizzati in tempo reale.

Per spiegare meglio tale funzione, consideriamo la seguente figura che mostra un cockpit relativo alle offerte realizzate.

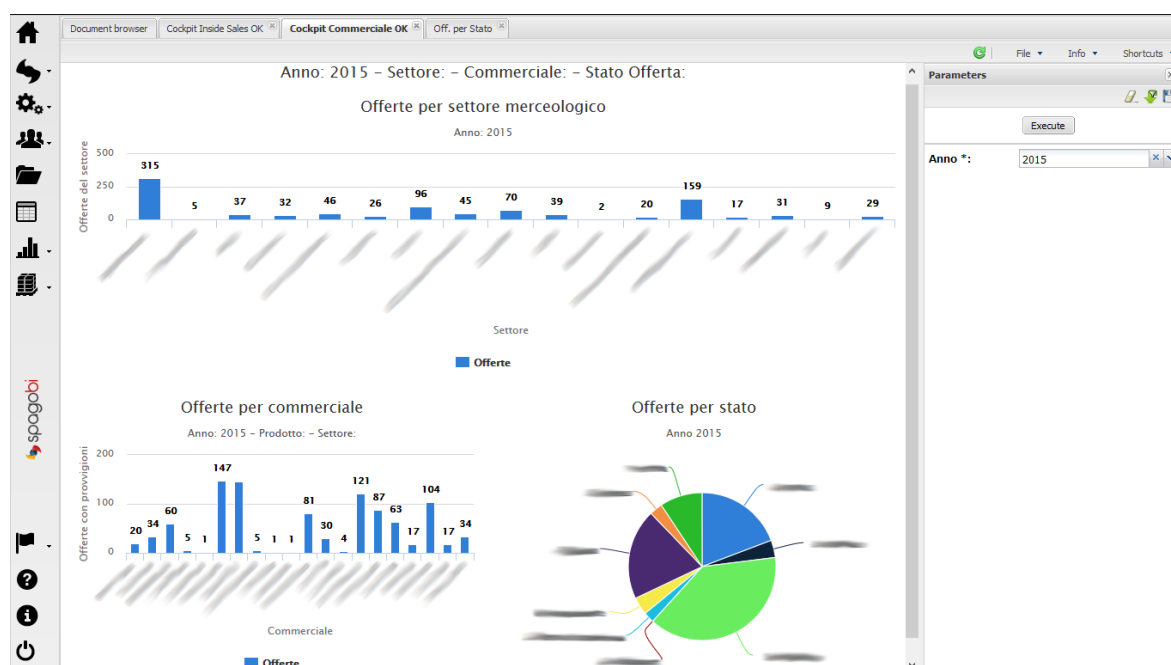


Figura 26 - Cockpit "Offerte"

La colonna di destra contiene i parametri del cockpit, cioè parametri globali comuni a tutti i grafici presenti, che in questo caso sono solamente uno, corrispondente all'anno da analizzare. A questo punto, è possibile cliccare su una delle colonne del grafico "Offerte per settore merceologico" per aggiornare i dati sottostanti con i dati relativi unicamente al settore merceologico selezionato (che viene anche riportato nella prima riga del cockpit). Per esempio, cliccando sulla colonna del settore "Edilizia" del grafico "Offerte per settore merceologico" i due grafici sottostanti e la prima riga del cockpit vengono aggiornati con i dati relativi all'anno 2015 e al settore Edilizia, come mostrato nella figura seguente.

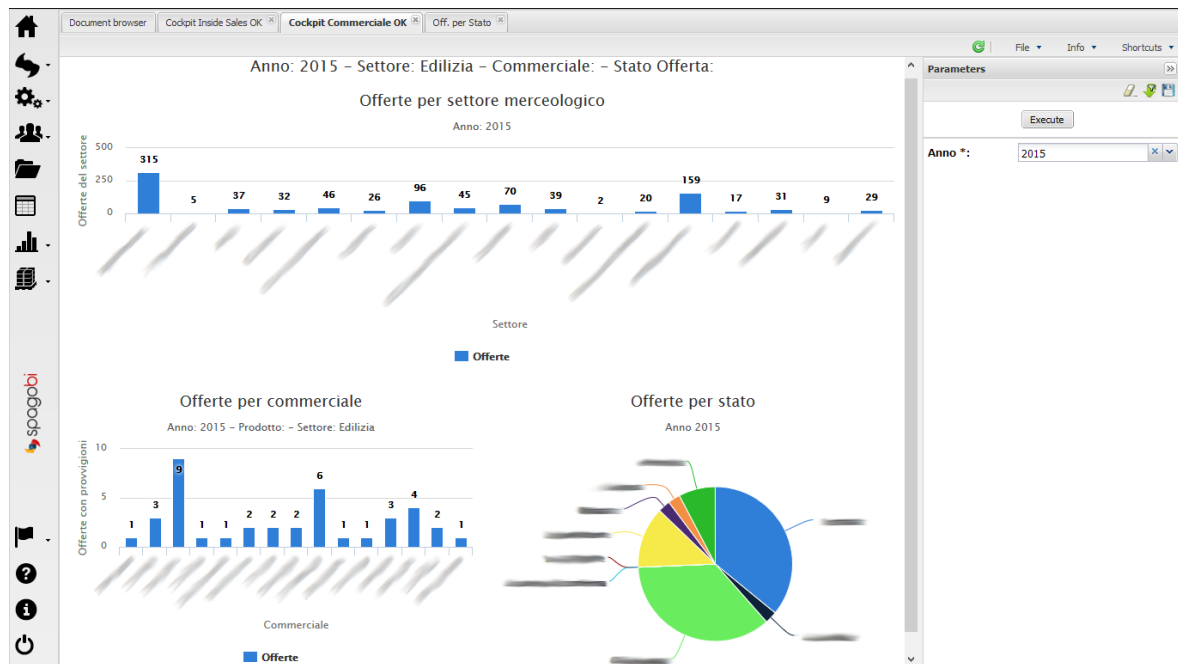


Figura 27 - Cockpit "Offerte" dopo il click sulla colonna "Edilizia"

È possibile modificare il filtro sui dati selezionando una differente colonna del grafico "Offerte per settore merceologico" oppure resettarlo cliccando sull'icona verde "Ricarica" presente in alto a destra. Stesso discorso è valido per gli altri grafici presenti in questo cockpit.

I template dei Cockpit, compreso quello mostrato pocanzi come esempio, sono realizzati tramite SpagoBI Studio. L'interfaccia grafica offerta permette di impostare la disposizione dei diversi documenti all'interno del cockpit, di gestire i parametri di ogni singolo documento e di impostare il passaggio di parametri da un documento ad un altro. In questo caso, però, l'interfaccia grafica risulta particolarmente scomoda e non completa: mancano infatti alcune funzionalità messe a disposizione dal server e configurabili soltanto tramite l'editing manuale del codice XML del template. Per questo tipo di documenti, quindi, è consigliabile la scrittura del template direttamente in formato XML in quanto permette una realizzazione più rapida ed efficace dello stesso.

Nelle figure seguenti sono mostrati il codice sorgente ed il risultato visivo di un altro Cockpit realizzato durante il periodo di tirocinio.

```

CP_JS_APTC.sbidoccomp X
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes"?>
<DOCUMENTS_COMPOSITION template_value="/jsp/engine/documentcomposition/template/dynamicTemplate.jsp">
<DOCUMENTS_CONFIGURATION video_width="1920" video_height="1080">

<DOCUMENT sbi_obj_label="      App_Mes_YOY" local_file_name="      App_Mes_YOY.sbihchart">
  <PARAMETERS>
    <PARAMETER type="IN" sbi_par_label="par_year_new" default_value="2015" />
    <PARAMETER type="IN" sbi_par_label="par_year_old" default_value="2014" />
    <PARAMETER type="IN" sbi_par_label="par_username" default_value="" />
  </PARAMETERS>
  <STYLE style="position:absolute;margin:0px;left:0px;top:0px;width:100%;height:350px;" mode="auto"/>
  <id>      App_Mes_YOY</id>
</DOCUMENT>

<DOCUMENT sbi_obj_label="      Con_App_Mese" local_file_name="      Con_App_Mese.sbihchart">
  <PARAMETERS>
    <PARAMETER type="IN" sbi_par_label="par_year" default_value="2015" />
    <PARAMETER type="IN" sbi_par_label="par_username" default_value="" />
  </PARAMETERS>
  <STYLE style="position:absolute;margin:0px;left:0px;top:350px;width:50%;height:400px;" mode="auto"/>
  <id>      Con_App_Mese</id>
</DOCUMENT>

<DOCUMENT sbi_obj_label="      Con_App_MeseP" local_file_name="      Con_App_MeseP.sbihchart">
  <PARAMETERS>
    <PARAMETER type="IN" sbi_par_label="par_year_new" default_value="2015" />
    <PARAMETER type="IN" sbi_par_label="par_year_old" default_value="2014" />
    <PARAMETER type="IN" sbi_par_label="par_username" default_value="" />
  </PARAMETERS>
  <STYLE style="position:absolute;margin:0px;left:50%;top:350px;width:50%;height:400px;" mode="auto"/>
  <id>      Con_App_MeseP</id>
</DOCUMENT>

</DOCUMENTS_CONFIGURATION>
</DOCUMENTS_COMPOSITION>

```

Figura 28 - Codice sorgente del Cockpit "Confronto Numero di Appuntamenti Mensili"

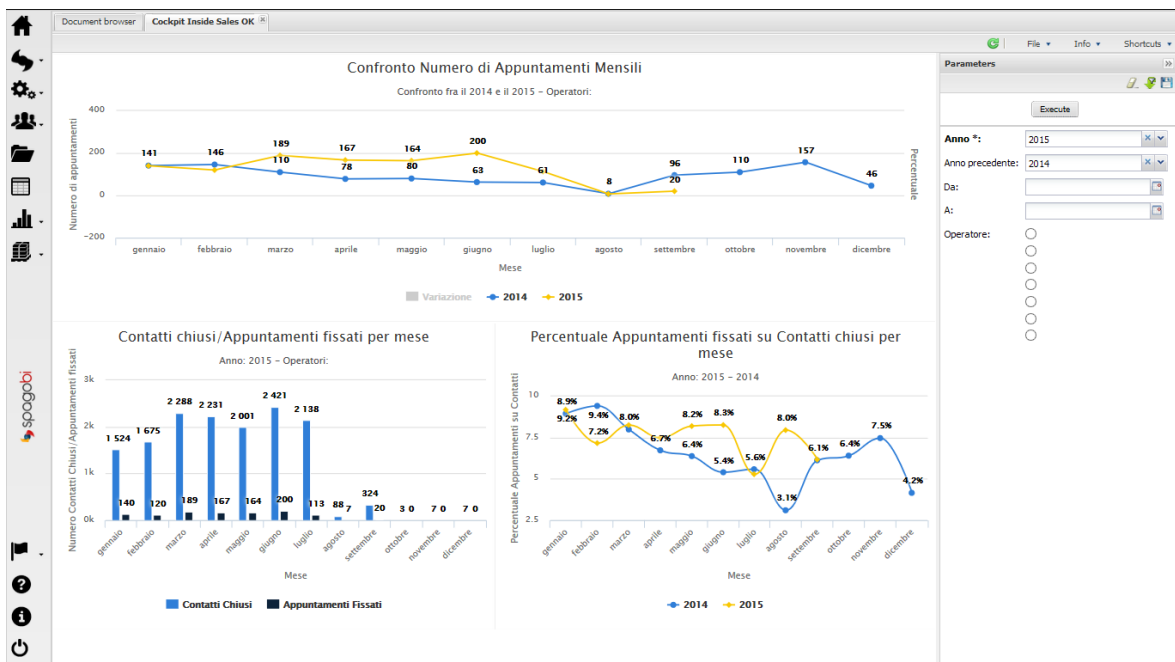


Figura 29 - Cockpit "Confronto Numero di Appuntamenti Mensili"

4.5 Talend Open Studio for Data Integration

Talend Open Studio for Data Integration (abbreviato TOS) è lo strumento di integrazione dei dati scelto per effettuare operazioni di ETL.

L'installazione e l'utilizzo sono più semplici rispetto a quelli della suite SpagoBI. Per l'installazione (su Windows) è sufficiente eseguire l'installer scaricabile dal sito www.talend.com e seguire la procedura. Al primo avvio, come negli altri IDE basati su Eclipse, viene richiesto un percorso da utilizzare come workspace e il nome del progetto.

In seguito, la prima operazione da effettuare, fortemente consigliata anche se non obbligatoria, è quella di configurare le connessioni alle sorgenti dati che saranno utilizzate. Questa operazione si effettua dalla sezione Metadata, all'interno della quale è possibile configurare connessioni a numerose tipologie di sorgenti tra cui database, file di testo (csv, xml, json), file Excel, sistemi Salesforce, server FTP e connessioni a servizi web. Una volta configurate, queste sorgenti possono essere usate all'interno dei moduli che vanno a definire i nostri Job, senza dover riconfigurare la connessione ad ogni utilizzo. Per determinati tipi di connessione, inoltre, si può definire la struttura dei dati come se fosse una tabella di un database relazionale (quindi tramite la definizione di tabelle con i relativi attributi). In questo modo, i componenti che, all'interno di un Job, riceveranno dati da un componente che utilizza una di queste connessioni caricheranno in automatico dai metadati la struttura dei dati che si attendono in input. Anche in questo caso, l'operazione serve a ridurre il tempo di scrittura dei Job ed aumentare la loro manutenibilità.

Dopo aver definito tutti i Metadata necessari ed eventualmente aver realizzato un Business Model, si può procedere alla progettazione dei Job. Ogni Job è composto da uno o più componenti connessi fra loro che permettono di impostare ed eseguire operazioni di ETL. Ogni componente può avere connessioni con altri componenti in ingresso, in uscita o in entrambe le direzioni in base alle funzioni che svolge. Sono presenti più di 800 connettori (cioè componenti che permettono di accedere a sorgenti dati) che permettono di connettersi ad altrettante tipologie di basi di dati sia in lettura che in scrittura, oltre a decine di componenti che permettono di gestire il flusso di dati e compiere operazioni di integrazione e trasformazione.

Fra i connettori, durante il periodo di tirocinio si sono rivelati particolarmente utili il connettore `tFileInputDelimiter`, che permette di leggere il contenuto di file CSV, il connettore `tMySQLConnection`, che permette di connettersi ad un server MySQL e i connettori `tMySQLInput` e `tMySQLOutput`, che offrono rispettivamente la possibilità di effettuare query su un database MySQL e la possibilità di salvare le righe ricevute in ingresso all'interno di una tabella di un database MySQL (entrambi i connettori utilizzando la connessione aperta con il componente `tMySQLConnection` o, in alternativa, una connessione specificata direttamente all'interno del componente).

Fra gli altri componenti, invece, sono molto utilizzati i seguenti componenti: il componente `tLogRow`, che stampa le righe ricevute in ingresso nella console di output (molto utile in fase di test e debug); il componente `tRowGenerator`, che genera automaticamente una sequenza di righe contenenti i campi ed i valori impostati (utile, per esempio, per la popolazione delle dimensioni temporali); il componente `tAggregateRow`, che aggrega le righe ricevute in ingresso per i campi specificati; il componente `tFilterRow`, che restituisce le

righe ricevute in ingresso dopo averle filtrate secondo le condizioni impostate; infine il componente tMap, che analizziamo nel dettaglio.

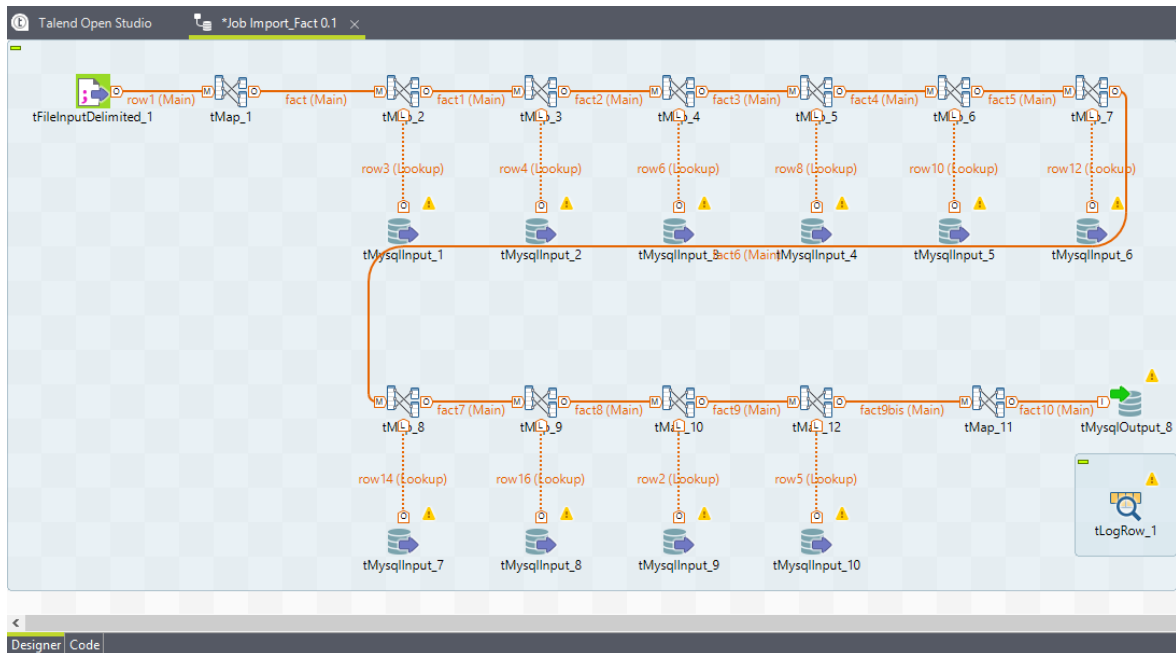


Figura 30 - Esempio di Job in Talend

Il componente tMap permette di selezionare, filtrare, modificare, trasformare e formattare le righe ricevute in ingresso. Permette, inoltre, di effettuare operazioni di join fra le righe ricevute in input e le righe ricevute in ingresso come lookup e di selezionare tutte le righe risultanti dal join oppure tutte le righe che non fanno parte del risultato dell'operazione di join. Anche per queste operazioni, Talend offre una comoda interfaccia visuale drag-and-drop mostrata nella seguente immagine.

Espressione	Tipo	Variabile

Espressione	Column
TalendDate.parseDate("MM/dd/yyyy HH:...	Date
row2.date_id	date_id
fact8.area_id	area_id
fact8.user_id	user_id
fact8.line_id	line_id
fact8.carrier_id	carrier_id
fact8.gateway_id	gateway_id
fact8.closing_cause_id	closing_cause_id
fact8.status_id	status_id
fact8.type_id	type_id
fact8.Duration	Duration
fact8.Number	Number
fact8.Cost1	Cost1
fact8.Cost2	Cost2
fact8.Ring_Time	Ring_Time
fact8.Datetime	Datetime

Colonna	C...	Tipo	A...	Modello data (...)	Lunghe...	Precisi...	Ness...	Comm...
date_id	<input checked="" type="checkbox"/>	int	<input checked="" type="checkbox"/>					
date_year	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					
date_month	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					

Colonna	C...	Tipo	A...	Modello data (...)	Lunghe...	Precisi...	Ness...	Comm...
Date	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"dd-MM-yyyy"	18	0		
date_id	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		18	0		
area_id	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		15	0		
user_id	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		3	0		

Figura 31 - Esempio di configurazione del componente tMap in Talend

Nella sezione di sinistra si trovano i campi delle righe in input e, se presenti, i campi delle righe di lookup, mentre a destra si trovano i campi delle righe in output. Tramite il tasto “+” in alto a destra è possibile aggiungere ulteriori output: in questo modo, partendo da un singolo blocco di dati in input, si possono realizzare più set di dati di output (ognuno con i propri campi) da collegare ad altrettanti componenti.

Per collegare i campi di input a quelli di output, è sufficiente trascinare ogni campo che si vuole in uscita dalla tabella di input a quella di output. Anche nel caso sia necessario effettuare un join fra la tabella di input e la tabella di lookup, è sufficiente trascinare ogni campo di join dalla prima al relativo campo della seconda.

Inoltre, cliccando sull'icona “...” presente a fianco di ogni espressione della tabella di output, è possibile definire delle trasformazioni da effettuare sui dati. Queste trasformazioni sono scritte in java e si possono sfruttare le funzioni già presenti all'interno di Talend (funzioni matematiche, binarie, sulle stringhe, sulle date, ecc.) oppure utilizzare delle funzioni personalizzate definite in precedenza nella sezione “Codice”.

Infine, nella parte bassa della schermata è possibile definire il tipo di dato dei differenti campi di input e di output, se possono contenere valori NULL e se sono chiavi primarie.

Nella figura n. 30 si può vedere il Job utilizzato per l'inserimento dei dati all'interno della tabella dei fatti del data warehouse. Questo Job, dopo aver caricato i dati da un file csv tramite il componente tFileInputDelimiter ed aver selezionato i campi di interesse tramite il componente tMap, sostituisce i valori delle dimensioni (inseriti precedentemente nelle relative tabelle attraverso un altro job) con le relative chiavi surrogate tramite i componenti tMap posizionati a cascata. Ogni componente tMap, infatti, effettua il join fra i dati in input e i dati in lookup (estratti da una differente tabella delle dimensioni dal componente tMysqlInput) e li invia al successivo tMap fino al definitivo inserimento nella tabella dei fatti. Nella figura n. 31 si può vedere in dettaglio come è definito il componente tMap_10 che effettua la sostituzione della data con il relativo identificatore della tabella delle dimensioni.

Tutti i Job realizzati possono essere esportati, tramite il comando “Build Job”, in formato eseguibile .jar all'interno di un pacchetto zip contenente, fra l'altro, i file da cui dipendono i job esportati (come, per esempio, le librerie java per connettersi alle sorgenti dati), i sorgenti dei job e la loro versione in linguaggio java. Grazie a questa funzione è possibile eseguire i Job realizzati in una qualsiasi macchina su cui è installata la Java Virtual Machine senza bisogno di installare Talend Open Studio.

5 Conclusioni e Sviluppi Futuri

5.1 Risultati

Durante i 3 mesi di tirocinio svolti presso Soft-Net Srl di Sassuolo (MO) è stato possibile studiare ed approfondire diverse tematiche riguardanti i Data Warehouse e la Business Intelligence, partendo da un'analisi dei prodotti disponibili sul mercato per arrivare a uno studio approfondito di alcuni strumenti open-source e allo sviluppo dei primi cruscotti. In particolare, il tirocinio è stato incentrato sulla comparazione di diversi strumenti di ETL e di BI e sullo studio della suite SpagoBI e di Talend Open Studio.

Per quanto riguarda SpagoBI, data la scarsa documentazione disponibile (per via del modello di business, basato sulla vendita di corsi di formazione, scelto da Engineering Spa) e le poche guide presenti online, in un primo periodo lo studio dell'applicazione è stato molto complicato e ha prodotto scarsi risultati.

Durante il percorso di apprendimento però, la stessa Engineering Spa ci ha consigliato di iniziare lo studio utilizzando il sito learn-spagobi.com della Ogutu Foundation (fondata da Stephen Ogutu) che, a seguito di una donazione alla fondazione, mette a disposizione vari tutorial che mostrano passo per passo come configurare Data Source e Data Set, come realizzare varie tipologie di documenti e come impostarne i parametri e i permessi. Grazie a questo prezioso strumento, è stato possibile apprendere le nozioni base di utilizzo di SpagoBI e realizzare autonomamente sia report che visualizzazioni di difficoltà sempre maggiore.

Per quanto riguarda Talend invece, grazie alla numerosa community e all'esaustiva documentazione, il percorso di apprendimento è stato più breve ed in poco tempo è stato possibile realizzare un primo piccolo Data Warehouse popolato con i dati estratti dai file csv prodotti dal centralino telefonico di un'azienda collaboratrice di Soft-Net Srl.

Sono stati inoltre realizzati alcuni grafici e cruscotti d'esempio relativi ai dati presenti all'interno del CRM di Soft-Net Srl, in particolare alla sezione dedicata all'ufficio di telemarketing, accedendo direttamente ad una copia fisica del database in quanto il tempo disponibile (3 mesi) non è stato sufficiente alla progettazione e alla realizzazione di un Data Warehouse in cui far confluire i dati provenienti dal CRM, dal centralino telefonico e dal gestionale.

5.2 Sviluppi futuri

Il primo passo da compiere nel proseguimento dello sviluppo è la realizzazione di un Data Warehouse completo contenente tutti i dati di interesse provenienti dalle diverse fonti aziendali seguendo lo schema a stella o quello a fiocco di neve. Questo sarà possibile grazie a Talend Open Studio for Data Integration, il quale, come abbiamo visto, permette l'estrazione e l'integrazione dei dati provenienti da sorgenti differenti e il caricamento degli stessi all'interno del DW. Si potrà inoltre configurare, tramite Talend Open Studio e SpagoBI, un'operazione pianificata che aggiorni automaticamente il DW collegandosi ad intervalli regolari alle sorgenti necessarie ed eseguendo i Job necessari.

Inoltre, si potranno realizzare ulteriori documenti (grafici, tabelle pivot basate su cubi OLAP, cruscotti, mappe) ed aumentare la conoscenza delle funzionalità offerte da SpagoBI, seguendo i corsi di formazione tenuti da Engineering Spa.

Infine, si potranno sfruttare le capacità acquisite e le possibilità offerte da SpagoBI per realizzare ed integrare grafici e cruscotti all'interno di una sezione dedicata del CRM di Soft-Net Srl.

6 Indice delle Figure

Figura 1 - Schema a stella (Star schema)	7
Figura 2 - Schema a fiocco di neve (Snowflake schema)	8
Figura 3 - Cubo Multidimensionale (o Cubo OLAP)	10
Figura 4 - Esempio di report (realizzato con BIRT)	11
Figura 5 - Esempio di grafico (realizzato con SpagoBI, motore HighCharts)	12
Figura 6 - Esempio di cruscotto interattivo (realizzato con SpagoBI).....	13
Figura 7 - Esempio di cruscotto interattivo (realizzato con SpagoBI) che si apre in seguito al click di uno stato all'interno del cockpit mostrato nella figura precedente	13
Figura 8 - Pentaho Data Integration - IDE Spoon.....	14
Figura 9 - Talend Open Studio for Data Integration. – IDE con esempio di Business Model .	15
Figura 10 - JasperReports Server - Esempio di report	17
Figura 11 - SpagoBI Server - Pagina Iniziale	18
Figura 12 - Jaspersoft - Esempio di grafico	19
Figura 13 - Tableau - Esempio di Join fra due gruppi di dati	20
Figura 14 - Tableau - Esempio di dashboard su differenti dispositivi.....	21
Figura 15 - Pentaho Enterprise Edition - Esempio di dashboard.....	22
Figura 16 - Infrastruttura della suite SpagoBI	24
Figura 17 - Interfaccia Web di SpagoBI Server - Pagina Iniziale	26
Figura 18 - Interfaccia Web di SpagoBI Server - Pagina di login	27
Figura 19 - Interfaccia Web di SpagoBI Server - Data Source, esempio di configurazione	28
Figura 20 - Interfaccia Web di SpagoBI Server - Data Set.....	28
Figura 21 - Interfaccia Web di SpagoBI Server - Document Browser	29
Figura 22 - SpagoBI Studio - Pagina iniziale	31
Figura 23 - Interfaccia per la realizzazione di grafici confrontata con il codice sorgente	32
Figura 24 - Grafico risultante dal template della figura precedente	33
Figura 25 - Esempio di report realizzato con motore BIRT	33
Figura 26 - Cockpit "Offerte"	34
Figura 27 - Cockpit "Offerte" dopo il click sulla colonna "Edilizia"	35
Figura 28 - Codice sorgente del Cockpit "Confronto Numero di Appuntamenti Mensili"	36
Figura 29 - Cockpit "Confronto Numero di Appuntamenti Mensili"	36
Figura 30 - Esempio di Job in Talend	38
Figura 31 - Esempio di configurazione del componente tMap in Talend.....	38

7 Bibliografia

W. H. Inmon, "What is a Data Warehouse?", Prism, 1995.

R. Kimball e M. Ross, "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", Wiley, 2013.

M. Golfarelli, S. Rizzi, "Data Warehouse. Teoria e pratica della progettazione", McGraw-Hill Education, 2006.

D. Beneventano, S. Bergamaschi, F. Guerra e M. Vincini, "Progetto di basi di dati relazionali. Lezioni ed esercizi", Pitagora, 2007.

E.F. Codd, S.B. Codd and C.T. Salley, "Providing OLAP to User-Analysts: An IT Mandate", E.F. Codd Associates, 1993.

8 Sitografia

<http://wiki.spagobi.org/>

<http://learn-spagobi.com/>

<http://www.talendforge.org/>

<http://help.talend.com/>

<https://en.wikipedia.org/>

<http://www.1keydata.com/>

<https://www.forrester.com/Business-Intelligence>

http://www.robortomarchetto.com/talend_studio_vs_kettle_pentao_pdi_comparison

<http://www.kimballgroup.com/2004/02/design-tip-51-latest-thinking-on-time-dimension-tables/>