Alma Mater Studiorum · Università di Bologna

SCUOLA DI INGEGNERIA E ARCHITETTURA
Corso di Laurea Magistrale in Ingegneria e Scienze Informatiche

DATA WAREHOUSE E CRUSCOTTO DIREZIONALE PER L'ANALISI DEL PERSONALE IN UN'AZIENDA DI SERVIZI

Tesi di Laurea in Sistemi Informativi e Business Intelligence

Relatore:
Chiar.mo Prof.
Stefano Rizzi

Correlatore:
Dott. Daniele Corda

Presentata da:
Davide Solazzi

Sessione III
Anno Accademico 2013/2014

Alla mia famiglia
, ai miei amici ed a
chi mi vuole bene ...

Indice

Elenco delle figure

Elenco delle tabelle

Introduzione

In un contesto aziendale risulta di fondamentale importanza la possibilità di analizzare grandi quantità di dati prodotti dai processi di business. Questo è proprio l'obiettivo che si prefigge la *business intelligence* (BI). Il termine BI può assumere i seguenti significati:

- Un insieme di processi aziendali per raccogliere dati ed analizzare informazioni strategiche;
- La tecnologia utilizzata per realizzare questi processi;
- Le informazioni ottenute come risultato di questi processi.

Il principale obiettivo dei sistemi di BI consiste nel fornire supporto durante i processi decisionali, raccogliendo le informazioni generate durante lo svolgimento delle attività aziendali e mettendo a disposizione strumenti per l'analisi dei dati. Al giorno d'oggi ogni azienda è dotata di diversi sistemi operazionali, utilizzati per gestire, standardizzare ed automatizzare il flusso delle informazioni prodotte durante l'esecuzione delle attività. Ognuno di questi sistemi possiede un proprio database nella quale memorizzare le informazioni di dominio, mantenendo così separato ogni contesto. Da qui nasce l'esigenza di integrare i dati provenienti da sistemi differenti, al fine di consentire ai proprietari del business di effettuare analisi sulle integrazioni e di prendere di conseguenza delle decisioni in base ai risultati ottenuti. Poiché i sistemi sono costituiti da database realizzati con tecnologie differenti, che nativamente non si integrano tra loro, sono necessarie operazioni e trasfor-

mazioni dei dati al fine di ottenerne un'integrazione sulla quale costruire il sistema.

L'obiettivo di questa tesi è proprio quello di descrivere il lavoro da me svolto presso Iconsulting S.r.l per la progettazione e realizzazione di un sistema di BI. Il progetto riguarda l'analisi del personale per un'azienda di servizi di grandi dimensioni ed è stato svolto utilizzando gli strumenti di BI messi a disposizione da Microsoft.

La tesi sarà così strutturata:

- Capitolo 1: verrà fornita un'introduzione storica della BI e sarà descritta la struttura generica di un sistema;
- Capitolo 2: esaminerà l'infrastruttura hardware e gli strumenti software utilizzati per la realizzazione del progetto;
- Capitolo 3: tratterà l'intero caso di studio e la relativa implementazione del sistema;
- Conclusioni: verranno tratte le conclusioni e descritti gli sviluppi futuri.

Capitolo 1

Business Intelligence & Data Warehousing

1.1 Storia della BI

Il termine Business Intelligence è stato introdotto per la prima volta nel 1868 da Richard Millar Devens' [4]. Egli lo utilizzò per descrivere il modo con cui un banchiere, *Sir Henry Furnese*, era riuscito ad avere successo nella propria carriera. Furnese riuscì a comprendere la situazione economica, politica e del mercato prima dei suoi concorrenti: *"attraverso l'Olanda, le Fiandre, la Francia e la Germania creò una perfetta organizzazione di business intelligence"*, scrive Devens, pertanto *"le notizie...furono ricevute da lui per primo"*. Furnese utilizzò le informazioni in suo possesso con fine fraudolento, passando quindi alla storia come banchiere corrotto. Quella appena proposta rappresenta una prima idea di raccolta di informazioni per la valutazione del business.

Per avere importanti sviluppi nel settore si dovette attendere fino alla metà del XX secolo, periodo in cui la tecnologia iniziò ad essere considerata di supporto alla BI. Nel 1958 l'informatico Hans Peter Luhn con l'articolo [8] descrisse *"un sistema automatico...sviluppato per diffondere informazioni alle varie sezioni di un'organizzazione industriale, scientifica o di governo"*.

Egli inoltre riportò la definizione di intelligenza presente sul dizionario di Websters come *"l'abilità di cogliere le interrelazioni tra fatti presentati in modo da guidare l'azione verso un obiettivo desiderato"*. Quest'ultima definizione si avvicina molto a quella della BI: un modo per capire velocemente e rapidamente grandi quantità di dati, così da poter intraprendere la miglior decisione possibile. Le ricerche di Luhn non fornirono solo argomentazioni teoriche, egli infatti sviluppò metodi che furono utilizzati nei primi sistemi analitici realizzati da IBM.

La nascita dei computer ed il relativo utilizzo nel mondo del business fornì alle aziende un metodo alternativo alla memorizzazione di dati su carta. L'invenzione dell'hard disk da parte di IBM nel 1956 rappresentò una rivoluzione per il salvataggio dei dati. Vennero introdotti i floppy disc, i laser disk ed altre tecnologie che permisero la produzione di una maggiore quantità di informazioni dato che vi era posto disponibile in cui salvarle. Questo ha portato alla creazione dei primi *Database Management Systems* (DBMS), rinominati genericamente anche *Decision Support Systems* (DSS). A partire dagli anni '70 iniziarono a spuntare i primi sviluppatori di sistemi di BI, che realizzarono strumenti per la gestione e l'organizzazione dei dati. Tuttavia la tecnologia era nuova e difficile da utilizzare, inoltre a causa del suo elevato costo aveva un mercato ristretto alle grandi aziende.

Con l'introduzione dei DBMS le compagnie iniziarono a memorizzare i dati prodotti dalle attività di business in sorgenti operazionali. L'ultima fase per ottenere informazioni significative dai dati consiste nel fare il reporting degli stessi. Le applicazioni di business, tuttavia, producono dati che riguardano settori differenti, di conseguenza gli stessi vengono memorizzati in sorgenti differenti, ognuna delle quali è completamente separata dalle altre. Le organizzazioni avevano però l'esigenza di eseguire report su un'unica versione dei dati, nasce quindi il problema dell'integrazione delle sorgenti.

Nei primi anni '80 Ralph Kimball e Bill Inmon individuarono la soluzione nel data warehouse, struttura che memorizza i dati provenienti da sorgenti differenti. Nonostante gli approcci utilizzati fossero differenti, Kimball e In-

mon avevano la stessa idea di base. Con il data warehouse le aziende furono in grado di memorizzare le informazioni prodotte dalle attività di business in un unico repository centralizzato. Grazie ad esso venne superata l'architettura a silos esistente in favore di una soluzione contenente dati integrati, non volatili, variabili nel tempo e orientati ai soggetti.

Il termine business intelligence iniziò a diffondersi su larga scala nei tardi anni '90 e primi anni 2000, grazie all'inserimento nel mercato di nuovi fornitori di software. Durante questo periodo la BI aveva due funzioni: produzione di dati e reporting, organizzazione dei dati e presentazione. La tecnologia adottata aveva però un problema principale, la complessità d'utilizzo. Molti dei progetti aziendali venivano ancora gestiti dal dipartimento IT, facendo emergere che gli utenti finali non erano ancora capaci di eseguire attività di BI in modo indipendente. Gli strumenti esistenti erano stati pensati per gli esperti, ed era necessaria un'intensa formazione analitica per l'acquisizione delle conoscenze. Col passare degli anni iniziarono ad essere sviluppati tools anche per gli utenti meno tecnici, ma questo cambiamento avvenne lentamente. Questa fase di sviluppo venne anche chiamata BI 1.0.

L'inizio del XXI secolo rappresenta una svolta nel mondo della BI. Vengono sviluppate nuove tecnologie che introducono una maggiore semplicità d'utilizzo. I nuovi strumenti consentono l'elaborazione real-time, con i dati che vengono inseriti all'interno del data warehouse quando generati dalle attività di business, consentendo alle aziende di prendere decisioni con le più recenti informazioni a disposizione. Altre tecnologie permettevano un accesso self service ai sistemi agli utenti meno esperti, liberando i dipartimenti dall'onere di gestione dei progetti. Durante questo periodo, soprannominato anche BI 2.0, ha contribuito fortemente la crescita esponenziale di Internet. Venne utilizzato un approccio maggiormente orientato al web ed ai browser, in contrapposizione agli strumenti proprietari che avevano caratterizzato la generazione precedente. La nascita dei *social network* quali Facebook, Twitter e dei Blog fornì agli utenti un nuovo modo per condividere idee ed opinioni. La crescente interconnessione del mondo imprenditoriale portò le

compagnie ad avere la necessità di informazioni in tempo reale. Per mantenere il passo della concorrenza dovevano capire le opinioni e le esigenze dei consumatori. La BI non veniva più considerata come uno strumento aggiunto o un vantaggio, ma stava diventando un obbligo per le imprese che volevano rimanere competitive ed appetibili su di un nuovo mercato orientato ai dati.

Attualmente il periodo di grande innovazione e sviluppo degli anni 2000 si è trasformato in un intenso processo di raffinazione. Alcune delle caratteristiche che si vogliono migliorare sono la presentazione dei dati e l'ampliamento delle opzioni di self service. Alcuni dei nuovi strumenti di visualizzazione si sono evoluti ed avvicinati agli utenti finali. L'obiettivo è fornire ad essi un potente strumento per l'accesso completo ai dati, così che possano esplorarli in modo autonomo, senza alcun tipo di formazione.

Gli sviluppi futuri in ambito BI sono orientati al cloud ed al settore del mobile. Grazie al cloud è possibile delocalizzare il software su Internet, riducendo il costo di archiviazione e rendendo l'accesso ai dati più facile e veloce (BI as-a-service). Altro aspetto di grande importanza è l'aumento delle piattaforme mobile, che consentono agli utenti di svolgere attività di BI in modo portabile, tramite smartphone, tablet o altri dispositivi (BI pervasiva).

1.2 Architettura di un sistema di BI

In figura 1.1 viene mostrata la generica architettura di un sistema di BI, i suoi principali componenti sono i seguenti:

- Sorgenti operazionali;
- Processo ETL;
- Data Warehouse (DW);
- Modello dei dati;
- Analisi dei dati.

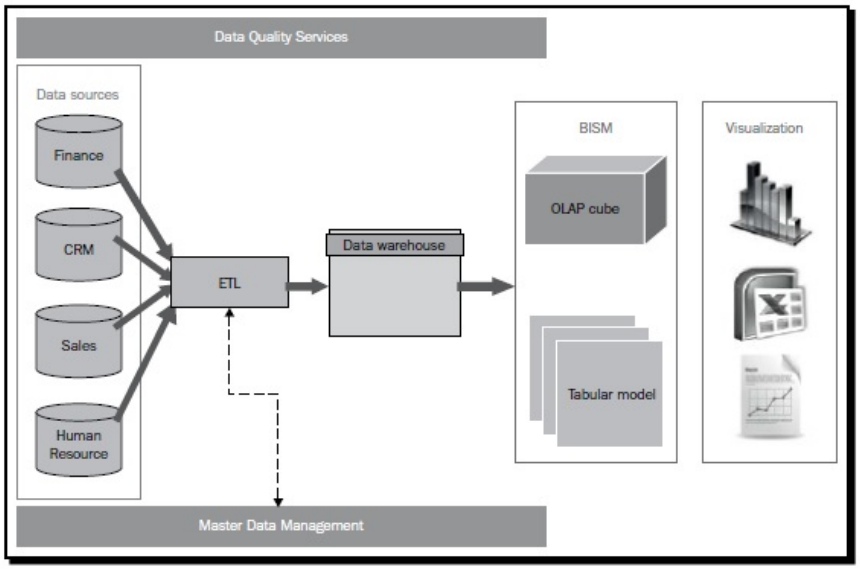


Figura 1.1: Architettura di un generico sistema di BI. BISM è l'acronimo di Business Intelligence Semantic Model e rappresenta il modello con il quale verranno memorizzati i dati di un DW

A seconda dell'ambiente e del software utilizzato può variare l'architettura del sistema, tuttavia i componenti mostrati in figura sono quelli solitamente sempre presenti. Per ognuno di essi i vari fornitori di software mettono a disposizione strumenti che ne consentono la gestione. Nei seguenti paragrafi verrà fornita una descrizione per ogni componente elencato.

1.3 Sorgenti operazionali

Rappresentano il punto di partenza dell'intera architettura. All'interno di un'organizzazione sono presenti diverse sorgenti operazionali, ognuna delle quali memorizza informazioni appartenenti a contesti differenti. Questi sistemi hanno il principale scopo di fornire supporto per l'esecuzione dei processi di business, per esempio il sistema delle vendite mantiene informazioni relative a ordini, spedizioni e restituzioni, oppure quello delle risorse umane cattura informazioni relative a promozioni o licenziamenti di dipendenti. Le

	Operational System	Analytic System
Purpose	Execution of a business process	Measurement of a business process
Primary Interaction Style	Insert, Update, Query, Delete	Query
Scope of Interaction	Individual transaction	Aggregated transactions
Query Patterns	Predictable and stable	Unpredictable and changing
Temporal Focus	Current	Current and historic
Design Optimization	Update concurrency	High-performance query
Design Principle	Entity-relationship (ER) design in third normal form (3NF)	Dimensional design (Star Schema or Cube)
Also Known As	Transaction System On Line Transaction Processing (OLTP) System Source System	Data Warehouse System Data Mart

Figura 1.2: Sistemi operazionali vs sistemi analitici

attività generate da questi sistemi sono dette transazioni e gli stessi prendono il nome di sistemi OLTP (*Online Transaction Processing*). Per facilitare l'esecuzione del business essi devono consentire differenti forme di interazione con il database, come inserimenti, aggiornamenti o cancellazioni. Durante ognuna di queste attività viene generata una transazione che il sistema avrà il compito di tenere memorizzata.

I sistemi OLTP sono adatti per l'esecuzione dei processi di business, ma non per la valutazione degli stessi. Le sorgenti sono infatti organizzate in modo completamente separato, ognuna con la propria tecnologia e contenente i dati del dominio applicativo trattato, per tale motivo è necessario passare ai sistemi analitici. In Figura 1.2 viene mostrato un confronto tra i sistemi OLTP e quelli OLAP. Quest'ultimi sono organizzati ed ottimizzati per effettuare l'analisi dei dati e forniscono strumenti per eseguire il reporting degli stessi. Prima di poter arrivare a questi sistemi è necessario svolgere un

ulteriore passo intermedio. I dati contenuti nei database operazionali sono in terza forma normale ed hanno una granularità d'informazione di molto superiore rispetto a quella necessaria ai possessori del business per prendere le decisioni. In secondo luogo risulta essere necessario integrare le diverse sorgenti, in modo tale da consentire valutazioni sui dati integrati, quest'ultimo obiettivo principale dei sistemi di BI. Il processo che consente il passaggio dai sistemi OLTP a quelli analitici prende il nome di ETL (*Extract, Transform Load*) e verrà analizzato nel paragrafo successivo.

1.4 Processo ETL

Tale processo rappresenta un passaggio fondamentale per la costruzione del DW. All'interno di un'azienda sono presenti più sistemi che svolgono il ruolo di sorgente per il sistema di BI. Risulta quindi necessario effettuare l'estrazione dei dati da ognuno di essi, trasformarli in una forma che sia adatta per il DW, ed infine caricarli all'interno dello stesso. Questo processo è detto *Extract Transform and Load* (ETL). Esistono numerosi tools in commercio che consentono di eseguire tali operazioni, in ambito Microsoft lo strumento ETL messo a disposizione prende il nome di SQL Server Integration Services (SSIS). Esso è incluso con il DBMS SQL Server ed è in grado di gestire l'integrazione di differenti tipi di sorgenti, come DB Oracle, file di testo, XML, servizi Web e DB dello stesso SQL Server. Durante lo svolgimento di questo processo viene solitamente impiegata un'area intermedia, situata tra le sorgenti ed il DW, nella quale memorizzare i dati elaborati. Tale area prende il nome di Staging e può essere utilizzata anche per garantire un livello di fault tollerance. Se, per esempio, fallisse la fase di trasformazione non sarebbe necessario eseguire nuovamente anche l'estrazione in quanto i dati sono già stati riportati all'interno della staging. Ad eccezione di eventuali fallimenti durante la fase di ETL l'accesso alla staging area deve essere effettuato esclusivamente per il caricamento dei dati all'interno del DW. Tale area contiene infatti dati intermedi dell'elaborazione, per questo motivo se

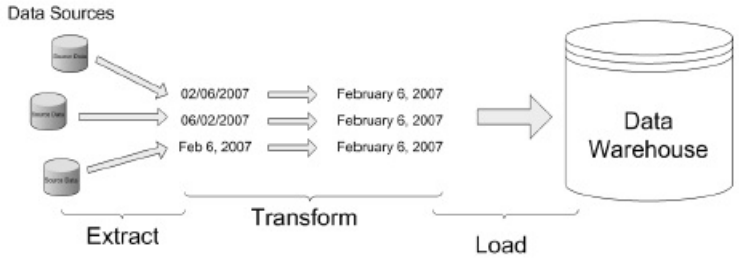


Figura 1.3: Esempio di processo ETL

ne deve impedire l'accesso agli utenti finali. Di seguito verranno analizzate le quattro principali operazioni che costituiscono la fase di ETL.

1.4.1 Estrazione

Durante questa fase si effettua l'estrazione dei dati dalle sorgenti, per poi renderli disponibili per le successive elaborazioni. L'obiettivo principale del processo è quello di estrarre solamente i dati d'interesse con le minori risorse possibili. Come già anticipato il dettaglio delle informazioni presente in un database è molto maggiore rispetto a quelle necessarie per la valutazione del business, inoltre per garantire un certo livello di efficienza ed elevata interattività durante l'analisi è necessario estrarre solamente una ridotta parte delle informazioni. Tale operazione dovrebbe essere realizzata in modo tale da evitare delle ripercussioni negative sui sistemi in termini di performance o tempi di risposta. I tipi d'estrazione possibili sono i seguenti:

- *Statica*: vengono prelevati tutti i dati. Rappresenta una fotografia delle informazioni contenute nelle sorgenti operazionali e viene solitamente eseguita per il primo popolamento del DW;

- *Incrementale*: vengono estratti solamente i record che hanno subito modifiche dalla data di ultima estrazione. In questo caso è necessario mantenere informazioni sulla precedente estrazione, in modo tale da

poter determinare quali record sono stati modificati. L'esecuzione può essere effettuata in modo immediato o ritardato.

1.4.2 Pulitura

La pulitura dei dati rappresenta una fase intermedia del processo ETL. Essa gioca un ruolo molto importante in quanto ha l'obiettivo di aumentare la qualità dei dati. Durante questa fase vengono individuati tutti quei record che contengono dati "sporchi" e ne viene effettuata la correzione, così da garantire la consistenza delle sorgenti. I principali problemi che vengono riscontrati nei dati operazionali sono i seguenti:

- Dati duplicati;
- Dati mancanti;
- Dati errati;
- Inconsistenze tra campi correlati: per esempio regione e provincia.

1.4.3 Trasformazione

Consiste nell'applicare usa serie di trasformazioni ai dati, in modo tale da convertirli dal formato sorgente a quello destinazione. Le diverse sorgenti memorizzano informazioni in formato differente, di conseguenza per poter effettuare l'integrazione risulta necessario trasformare i dati in un formato uniforme. Le principali operazioni realizzate durante questa fase sono le seguenti:

- Conversione: le sorgenti relazionali potrebbero utilizzare formati differenti per la memorizzazione delle informazioni, è quindi necessaria un'operazione di conversione al fine di ottenere un formato uniforme;
- Integrazione: si effettua l'integrazione dei dati provenienti da sorgenti differenti;

- Aggregazione: i dati vengono aggregati ad un livello di dettaglio differente (solitamente su base mensile);
- Misure derivate: vengono calcolate nuove misure a partire da quelle già esistenti;
- Selezione: si seleziona un sottoinsieme dei campi contenuti nelle sorgenti, in modo tale da ridurre la quantità di dati da elaborare.

1.4.4 Caricamento

Rappresenta l'ultima fase del processo di ETL. I dati vengono caricati dal livello riconciliato, realizzato successivamente alla fase di trasformazione, al DW finale. Le modalità di caricamento sono principalmente due:

- *Refresh*: Il DW viene ricaricato completamente. I vecchi dati vengono cancellati per lasciare spazio ai nuovi. Questa modalità viene solitamente utilizzata per effettuare il primo caricamento del DW;
- *Update*: Vengono caricati nel DW solamente quei dati che hanno subito una modifica dalla data dell'ultimo caricamento. I dati che sono stati modificati non vengono cancellati o aggiornati, garantendo così la storicizzazione del DW.

1.5 Data Warehouse

Elemento fondamentale di un sistema di BI, il suo principale obiettivo è quello di consentire l'analisi dei dati ed il reporting degli stessi. Questi obiettivi portano alla definizione di una struttura dedicata ai DW. Come già anticipato i sistemi operazionali vengono costruiti rispettando i criteri della normalizzazione, che garantiscono una maggior efficienza riducendo per esempio la ridondanza dei dati. Un database progettato in 3NF potrebbe avere tabelle con un numero elevato di relazioni. Di conseguenza la realizzazione di un report su un sistema di questo tipo potrebbe rallentare l'esecuzione della

query dato l'elevato numero di join necessari per recuperare le informazioni. La struttura di un DW viene appositamente progettata per evitare questo tipo di inconvenienti, riducendo il tempo di risposta ed incrementando le performance delle query per il reporting e l'analisi dei dati. Il modello con il quale viene costruito un DW prende il nome di modello multidimensionale. Nel paragrafo successivo verranno descritti nel dettaglio i componenti principali di tale modello.

1.5.1 Modello multidimensionale

Il modello *Entity-Relationship* (ER) tradizionalmente utilizzato per la progettazione concettuale di database, non può essere adottato come fondamento per la realizzazione di DW. Tale modello non risulta adatto in quanto "*è di difficile comprensione agli utenti e non può essere navigato efficacemente dai DBMS*" [6]. Esso descrive la struttura del dominio applicativo e le relative associazioni, ma non esprime concetti come la multidimensionalità o la gerarchia dei livelli di aggregazione. Per risolvere tali problematiche viene utilizzato un formalismo che prende il nome di Dimensional Fact Model (DFM). Il DFM è un modello concettuale grafico, pensato appositamente per la modellazione multidimensionale, con l'obiettivo di:

- Fornire pieno supporto alla progettazione concettuale;
- Rendere disponibile un ambiente nel quale effettuare query in modo intuitivo;
- Facilitare la comunicazione tra progettista ed utente finale;
- Costruire una piattaforma per la progettazione logica. Tale modello è infatti indipendente dal modello logico utilizzato.

In Figura 1.4 viene mostrata la rappresentazione grafica di un DFM. Esso consiste in un insieme di schemi di fatto, i cui elementi costitutivi sono i seguenti:

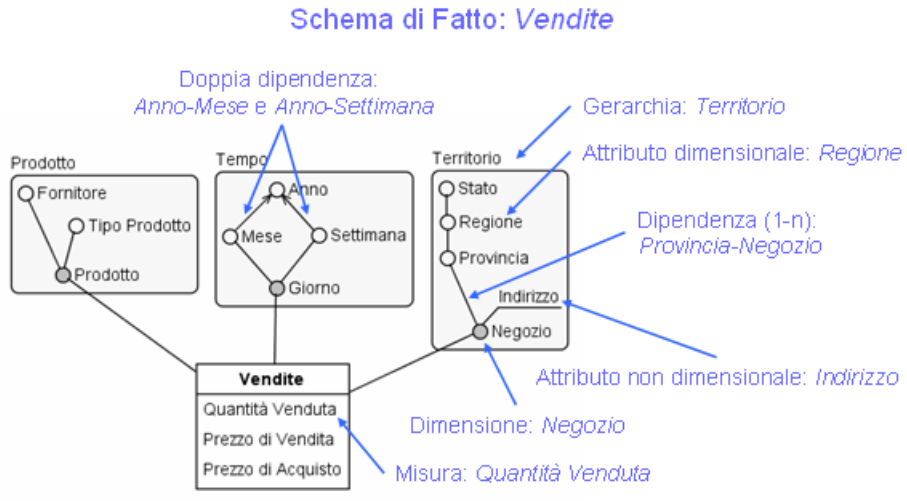


Figura 1.4: Esempio di schema di fatto per il business delle vendite

- *Fatto*: concetto di interesse per il processo decisionale che modella un insieme di eventi che si verificano all'interno della realtà aziendale. Ogni fatto è descritto da un insieme di misure ed esprime un'associazione molti-a-molti tra le dimensioni. Questa relazione è espressa da un Evento Primario, ovvero da un'occorrenza del fatto. Nei DFM viene rappresentato con un box rettangolare che ne specifica il nome e con all'interno le misure d'interesse;

- *Misura*: proprietà numerica di un fatto che descrive un aspetto quantitativo di interesse per l'analisi. Un fatto può anche non contenere alcuna misura, in questo caso si registra solamente il verificarsi dell'evento;

- *Dimensione*: proprietà, con dominio finito, che descrive una coordinata d'analisi di un fatto. Ogni fatto contiene generalmente più dimensioni che ne definiscono la granularità, ovvero l'evento di massimo dettaglio analizzabile. Nella Figura 1.4 le dimensioni sono prodotto, giorno e negozio. L'informazione elementare rappresentabile riguarda le vendite di un prodotto effettuate in un negozio in un dato giorno;

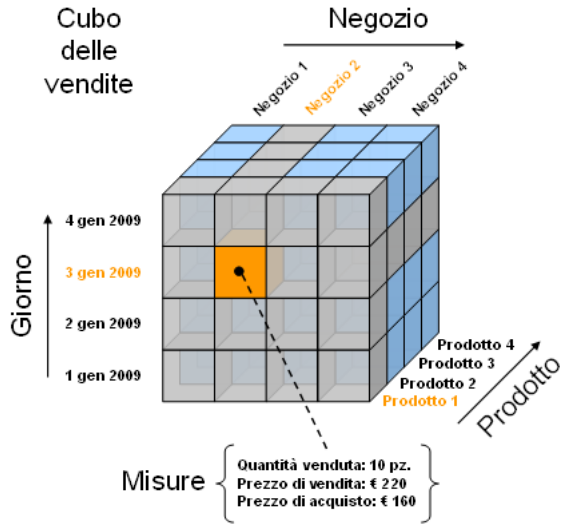


Figura 1.5: Esempio di cubo

- *Attributo dimensionale*: proprietà, con dominio finito, di una dimensione. Un prodotto può avere un fornitore ed un tipo. Le relazioni tra attributi dimensionali sono espresse dalle gerarchie;

- *Gerarchia*: albero direzionato in cui i nodi sono attributi dimensionali e gli archi rappresentano associazioni molti-a-uno tra coppie di attributi dimensionali. Una gerarchia include la dimensione, come radice dell'albero e tutti gli attributi che la descrivono. Essa definisce il modo in cui i dati possono essere aggregati per fornire supporto durante il processo decisionale.

La struttura che meglio si adatta alla rappresentazione di dati multidimensionali è il cubo. In Figura 1.5 viene mostrato il cubo per il settore delle vendite. Tale modello utilizza le dimensioni come coordinate d'analisi, mentre ogni cella contiene le misure del fatto. Queste ultime registrano i valori delle misure per ogni occorrenza di un evento primario. Ogni cubo che ha un numero di dimensioni superiore a tre prende il nome di ipercubo.

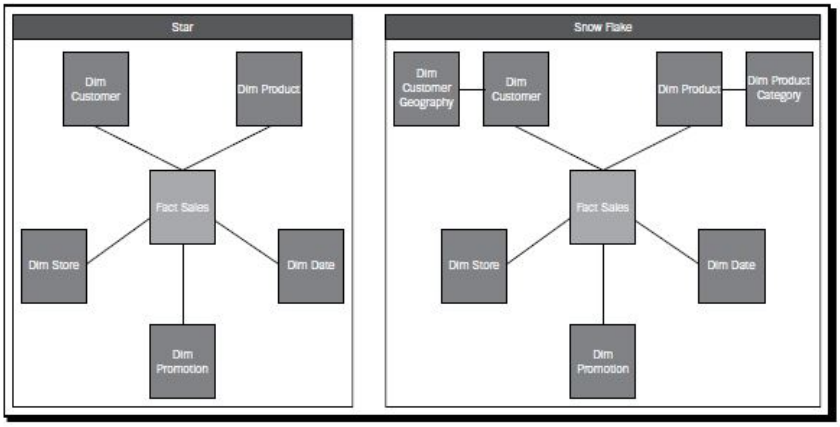


Figura 1.6: Star schema vs snowflake schma

La modellazione multidimensionale contrasta con il formato relazionale utilizzato nelle sorgenti. Per rappresentare dati multidimensionali in database relazionali esistono due differenti schemi:

- *Star schema*: la tabella dei fatti si trova al centro dello schema e le dimensioni sono collegate ad essa tramite relazioni di un solo livello;

- *Snowflake schema*: può contenere delle relazioni in cui una dimensione è collegata ad una tabella dei fatti per mezzo di una dimensione intermedia. Tale schema risulta similare alla forma normalizzata, pertanto saranno richiesti un maggior numero di join per rispondere ad una query, rendendolo più lento e meno preferibile rispetto allo schema a stella. Ovviamente non sempre è possibile realizzare uno schema a stella senza operazioni di snowflake, tuttavia come best practice sarebbe preferibile evitare lo snowflake quando non strettamente necessario.

In Figura 1.6 viene mostrato un esempio di schema a stella e di quello snowflake.

1.6 Modello dei dati

Un DW svolge il ruolo di sorgente per l'analisi dei dati ed il reporting, di conseguenza opera più velocemente di un normale sistema relazionale. Esso tuttavia non risulta essere così veloce da soddisfare tutte le esigenze, perché rimane comunque un database relazionale. Per risolvere questo problema e garantire un ottimo rapporto tra velocità d'elaborazione e tempo di risposta alle query è necessario introdurre un ulteriore livello in un sistema di BI. Questo nuovo livello prende il nome di modello dei dati, contiene un modello basato su file o su memoria dei dati ed ha lo scopo di fornire risposte veloci durante l'esecuzione delle query. La suite Microsoft utilizzata nel caso di studio offre due diversi modelli dei dati:

- *Cubo OLAP*: struttura che archivia i dati su file caricandoli dal DW in un modello dimensionale. Tale struttura cerca di precalcolare le differenti combinazioni di dimensioni e fatti in modo tale da consentire un'elevata interattività, consentendo agli utenti di aggiungere o rimuovere attributi in base alle analisi necessarie. Questo modello mette a disposizione diverse operazioni come il *roll-up* o *drill-down*, grazie al quale è possibile navigare i dati da punti di vista differenti. Il processo d'analisi è anche detto Online Analytical Processing (OLAP), in quanto garantisce un elevato livello d'interattività agli utenti;
- *Formato tabulare in memoria*: questo secondo modello consiste nel caricare i record delle tabelle d'analisi in memoria, eseguendo la query direttamente su quest'ultima. Tale struttura risulta essere molto veloce dal punto di vista dei tempi di risposta, ma è richiesta anche un'elevata capacità di memorizzazione che non sempre è disponibile.

1.7 Analisi dei dati

Rappresenta la parte front-end di un sistema di BI. Rimanendo in ambito Microsoft esistono differenti modalità con la quale analizzare le informazioni

provenienti dal sistema. Una prima tecnica consiste nell'utilizzare lo strumento Excel, in esso è presente la possibilità di connettersi ad un cubo OLAP e di effettuare analisi libere, inserendo indicatori e informazioni descrittive a piacimento o variando il livello di dettaglio dell'analisi. La suite Performance Point, come parte dello strumento Microsoft SharePoint, consente di creare dashboard avanzate e garantisce ottime prestazioni se utilizzato in accoppiata con un cubo OLAP. Un altro tool molto importante è Microsoft SQL Server Reporting Services, che consente la creazione avanzata di report a partire da diverse origini dati.

Capitolo 2

Aspetti tecnici e strumenti utilizzati

Per la realizzazione del progetto si è resa necessaria un'apposita infrastruttura hardware e software interna all'azienda. Grazie alla partnership esistente tra Iconsulting ed il cliente è stato possibile utilizzare la stessa configurazione hardware già adottata in altri progetti. La stessa strategia è stata applicata per la scelta dello strumento software, adottando la suite Microsoft. I motivi di tale scelta sono molteplici: essendo gli altri progetti realizzati con strumenti Microsoft risulta più facile gestire i differenti DM e le relative integrazioni. In secondo luogo sono semplificati gli sviluppi per i dipendenti di Iconsulting, che potranno lavorare con strumenti di cui conoscono le funzionalità, garantendo una migliore qualità del servizio. Infine si forniscono agli utenti finali strumenti standard per la reportistica, in modo tale da rendere le analisi più veloci ed efficienti.

2.1 Configurazione Hardware

L'infrastruttura hardware utilizzata per il progetto è replicata per due differenti ambienti:

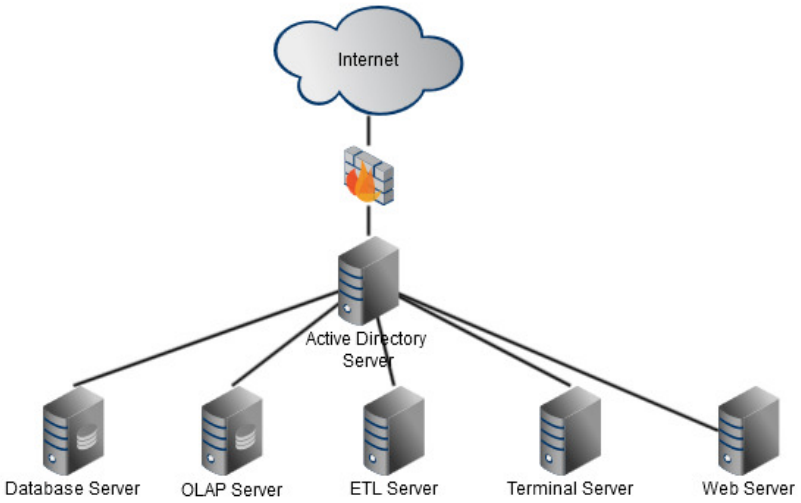


Figura 2.1: Infrastruttura di rete dell'azienda cliente

- Sviluppo: ambiente che contiene tutti i server per effettuare lo sviluppo dei progetti ed il testing. Ad esso accedono principalmente gli sviluppatori di Iconsulting, ma possono accedere anche gli utenti finali per testare modifiche particolarmente critiche prima che vengano portate in produzione. Periodicamente vengono mandati in esecuzione i sistemi di BI esistenti per effettuare l'allineamento dei dati rispetto all'ambiente di produzione. In questo ambiente viene anche mantenuto lo storico dei rilasci in modo tale da riportare i sistemi ad uno stato consistente in caso di errori;

- Produzione: contiene tutti i server che ospitano i sistemi attualmente utilizzati dall'azienda cliente. L'aggiornamento dei dati viene effettuato con periodo giornaliero, le attività vengono schedulate in precisi intervalli di tempo in modo tale da evitarne la sovrapposizione ed il conseguente rallentamento nell'esecuzione. Dato che questo ambiente viene utilizzato dagli utenti finali i dati hanno un'elevata criticità, pertanto è fondamentale garantirne sempre la consistenza.

In Figura 2.1 viene mostrata l'infrastruttura utilizzata per entrambi gli ambienti. La connessione alla rete aziendale interna viene effettuata per mezzo di una VPN ad accesso remoto. Questo meccanismo consente agli utenti che lavorano da casa o in movimento di accedere ai server presenti in una rete privata utilizzando l'infrastruttura resa disponibile da una rete pubblica come ad esempio internet. L'organizzazione di quest'ultima è irrilevante, dal punto di vista logico è come se i dati venissero inviati su un collegamento privato dedicato. I server utilizzati per il progetto sono i seguenti:

- *Active Directory Server*: macchina che gestisce l'autenticazione degli utenti alla rete privata ed ai relativi server. Una volta che l'utente è autenticato può accedere alle risorse disponibili in rete. Vengono utilizzati due domini differenti a seconda dell'ambiente a cui si accede. Per quello di sviluppo il dominio è *DOMTST*, mentre per quello di produzione è *DOMZPO*. In questo modo è possibile separare non solo fisicamente ma anche logicamente i due ambienti;

- *Database Server*: macchina che gestisce le sorgenti relazionali di tutti i progetti con l'azienda cliente. Essa contiene i db di staging area realizzati durante la fase di ETL ed i DW sulla quale vengono costruiti i cubi OLAP;

- *ETL Server*: gestisce tutti i flussi ETL dei progetti con l'azienda cliente. Sulla macchina i flussi vengono schedulati con modalità giornaliera e precisi intervalli temporali, sulla base delle esigenze progettuali e del momento della giornata nella quale i dati da caricare vengono resi disponibili. Tutte le modifiche ad un flusso vengono svolte e testate in ambiente di sviluppo, pertanto affinché le sorgenti siano aggiornate è necessario effettuarne il deploy in produzione;

- *OLAP Server*: macchina che ospita il modello dei dati ottimizzato per l'analisi. Come nel caso dei server precedenti in esso sono contenuti tutti i progetti attivi con l'azienda cliente. Il server contiene due diverse istanze per l'analisi dei dati: multidimensionale (cubo OLAP) o

tabulare. A seconda delle necessità è quindi possibile scegliere l'istanza più adatta, per il progetto in esame è stato utilizzato il modello multidimensionale. Per effettuare le analisi gli utenti possono utilizzare gli strumenti che supportano la connessione all'istanza. Il software utilizzato dall'azienda cliente per la reportistica e per l'analisi è Microsoft Excel, che si integra nativamente con gli altri strumenti Microsoft utilizzati nel progetto;

- *Terminal Server*: macchina alla quale i dipendenti Iconsulting si collegano in accesso remoto. Tale server viene utilizzato per l'amministrazione, la gestione e lo sviluppo di tutti i progetti con il cliente. Da esso, utilizzando l'IDE SQL Server Management Studio (SSMS), è possibile connettersi agli altri server e svolgere le operazioni opportune;

- *Web Server*: macchina che ospita la piattaforma Sharepoint realizzata da Microsoft. Essa consente la creazione e la gestione di siti web realizzati per scopi aziendali. Per i progetti tra Iconsulting e l'azienda cliente è stato costruito un apposito sito web nel quale vengono effettuati i rilasci della reportistica finale e della relativa documentazione.

2.2 Configurazione Software

Il progetto concordato tra Iconsulting ed il cliente prevede la realizzazione di un sistema di BI che, dopo aver recuperato i dati delle sorgenti operazionali, esegua la fase di ETL e costruisca il rispetto DW ed il cubo OLAP. Il primo verrà utilizzato per costruire report in Excel che sfruttano una sorgente relazionale, mentre con il secondo gli utenti potranno effettuare analisi interattive dei dati tramite il modello multidimensionale.

Il cliente per i motivi già anticipati all'inizio del capitolo, ha deciso di utilizzare come strumento per la realizzazione del sistema la suite Microsoft Business Intelligence 2012.

Figura 2.2: Organizzazione logica di Microsoft BI

I tre principali componenti della suite vengono mostrati in figura 2.2 e sono i seguenti:

- Parte core: include tutti quegli strumenti che consentono di eseguire la fase di ETL, la creazione del cubo ed il reporting e che sono inclusi con il DBMS Microsoft SQL Server;
- Presentazione: I prodotti Microsoft Office e la tecnologia Sharepoint che svolgono il ruolo di presentazione dei dati;
- Personalizzazioni: strumenti aggiuntivi che gli sviluppatori possono realizzare sfruttando le potenzialità di Microsoft BI.

Nei paragrafi seguenti verrà fornita una descrizione dei componenti utilizzati all'interno del progetto.

2.2.1 Microsoft SQL Server

SQL Server è stato inizialmente sviluppato come prodotto per la gestione di database, tuttavia col passare degli anni è cresciuto includendo numerose funzionalità aggiuntive, quali ad esempio quelle relative alla BI. In Figura 2.3 viene mostrata la sua organizzazione funzionale:

- *SQL Server Database Engine*: è il componente core di SQL Server, consente la creazione di database relazionali, inclusi DW e data mart.

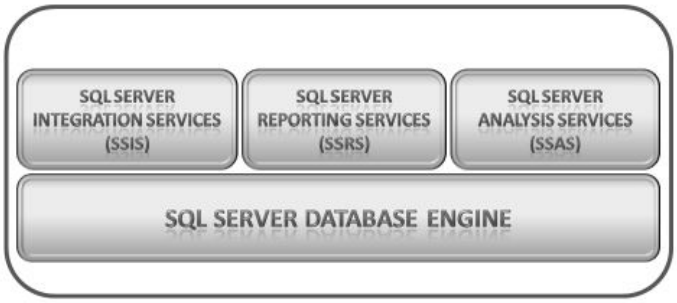


Figura 2.3: Componenti di Microsoft SQL Server

Offre una serie di strumenti per la modifica, l'aggiornamento, la cancellazione di record in sorgenti relazionali e per l'interrogazione degli stessi mediante query;

- *SQL Server Integration Services (SSIS)*: software che consente la connessione a diverse sorgenti, la trasformazione dei dati in base alle esigenze ed il caricamento in un database SQL Server. Tale prodotto permette la realizzazione della fase di ETL;
- *SQL Server Analysis Services (SSAS)*: software che, a partire da sorgenti relazionali carica i dati in database che hanno la struttura di cubi OLAP, per un'analisi interattiva ed efficiente;
- *SQL Server Reporting Services (SSRS)*: realizza la parte di frontend di un sistema di BI e consente la creazione di report che si basano su dati provenienti da sorgenti di diversa natura.

2.2.2 SQL Server Database Engine

Vi sono differenti strumenti che la suite mette a disposizione per lavorare con SQL Server, ma quello che consente la gestione dell'intero sistema è SQL Server Management Studio (SSMS). Tale strumento comprende diversi componenti che vengono utilizzati per creare, amministrare e gestire il sistema. I principali, mostrati in Figura 2.4, sono i seguenti:

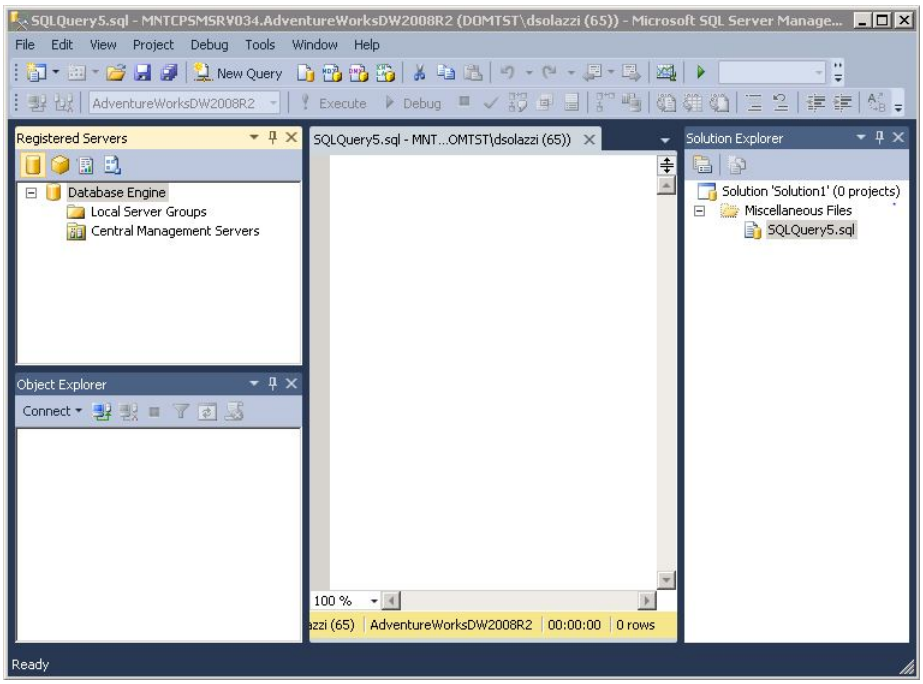


Figura 2.4: Schermata iniziale di SQL Server Management Studio

- *Registered Servers*;
- *Object Explorer*;
- *Query Editor*;
- *Solution Explorer*.

Registered Servers

Tale pannello mantiene le connessioni ai server che sono stati utilizzati. Attraverso ogni connessione è possibile controllare lo stato del server o gestire i suoi oggetti. Per ogni utente viene mantenuta una lista locale dei server alla quale si è connesso. Le principali operazioni che è possibile eseguire tramite tale interfaccia sono le seguenti:

- Registrazione Server: per poter utilizzare gli oggetti contenuti in un server è necessario effettuarne la registrazione. Durante tale processo

si specifica il nome del server che si vuole registrare ed il tipo di autenticazione utilizzata dallo stesso. SSMS separa il task di connessione al server da quello di registrazione, pertanto la registrazione di un server non comporta anche la connessione, che dovrà essere eseguita in modo separato;

- Creazione di gruppi di server: i server già registrati possono essere raggruppati logicamente o in base al tipo (Database Engine, Integration Services, Analysis Services, Reporting Services). In questo caso è necessario creare il gruppo, assegnandogli un nome ed una descrizione, in seguito è possibile aggiungere i server desiderati.

Object Explorer

Componente che fornisce una vista ad albero di tutti i database contenuti all'interno di un server. L'albero è organizzato in forma gerarchica, di conseguenza espandendo ogni ramo è possibile scorrere la struttura logica del server. All'interno della stessa interfaccia è possibile connettersi a più server contemporaneamente, anche a quelli di tipo differente rispetto all'attuale connessione. Le operazioni principali che consente di realizzare tale strumento sono:

- Connessione al server: mentre dall'interfaccia Registered Servers si effettua la registrazione, in questo caso è possibile connettersi ai server correttamente registrati;

- Gestione server: SSMS consente la contemporanea gestione di più database server. Ogni istanza di database possiede i propri oggetti locali che non sono condivisi con gli altri. Risulta quindi possibile gestire separatamente la configurazione di ogni database;

- Esecuzione e arresto server: selezionando il server è possibile avviarlo o arrestarne l'esecuzione;

- Creazione database: selezionando un server al quale si è già connessi è possibile creare un nuovo database. Nella form di creazione è necessario specificare il nome del db ed eventualmente il proprietario;

- Modifica database: è possibile applicare delle modifiche ad un database esistente. Oltre ad effettuare la cancellazione dello stesso è possibile inserire nuovi file dati per il db in esame, oppure aggiungere un filegroup secondario. Quest'ultimi sono insiemi di file che forniscono supporto alle attività amministrative, come il backup o il ripristino;

- Gestione tabelle: completata la creazione del database è necessario passare alla creazione delle tabelle. Per ogni colonna che si vuole inserire è necessario specificarne il nome, il tipo, stabilire se potrà contenere o meno valori nulli ed eventuali valori di default. Per completare la creazione della tabella si deve scegliere la chiave primaria. Le stesse tabelle create possono poi essere modificate o cancellate a seconda delle necessità.

Query Editor

Tramite questo componente è possibile eseguire, sui database ai quali si è collegati, differenti tipi di query. Per default la nuova query che viene creata è di tipo relazionale, ma ne esistono di altri tipi, come MDX o XMLA. Per potere eseguire una nuova interrogazione è necessario connettersi al server e specificare il database che si vuole interrogare. Il linguaggio utilizzato per le interrogazioni è detto Transact-SQL (T-SQL), estensione proprietaria del linguaggio SQL sviluppata da Microsoft in collaborazone con Sybase. L'editor di query consente di realizzare numerose attività:

- Generazione ed esecuzione di statement T-SQL: completata la scrittura della query viene eseguita l'interrogazione sul db scelto ed al termine dell'elaborazione vengono mostrati i risultati nell'apposito pannello;

- Memorizzazione delle query su file: le interrogazioni possono essere memorizzate su file ed in seguito importate nell'editor;
- Analisi del piano di esecuzione delle query: è possibile analizzare e mostrare graficamente il piano scelto dall'ottimizzatore per l'esecuzione di una data query.

Solution Explorer

Questo componente consente di organizzare le interrogazioni effettuate sui database sotto forma di progetti. Quando un nuovo progetto viene creato esso memorizza informazioni relative alle connessioni ed alle query effettuate. In questo modo si organizza il lavoro svolto in progetti, con la possibilità di unire in un'unica soluzione quelli logicamente correlati.

2.2.3 SQL Server Integration Services

SQL Server Integration Services è il componente di SQL Server che consente la realizzazione del processo di ETL. Questo tool nacque inizialmente con il nome di Data Transformation Services (DTS), ma con il passare degli anni Microsoft ha aggiunto ulteriori feature e aumentato le sue potenzialità, per tale motivo è stato nominato Integration Services a partire da SQL Server 2005. Per lo sviluppo di applicazioni mediante SSIS viene messo a disposizione un potente IDE detto SQL Server Data Tools (SSDT) che è anche formalmente conosciuto come Business Intelligence Development Studio (BIDS). Microsoft ha integrato tale ambiente con l'IDE principale per lo sviluppo dei suoi prodotti, ovvero Visual Studio, riunendo in un'unica posizione tutti gli strumenti necessari per la realizzazione di un sistema di BI completo. L'IDE SSDT, infatti, può essere utilizzato anche per lo sviluppo di applicazioni di tipo SSAS e SSRS.

I pannelli di SSIS che rivestono maggiore importanza sono:

- *Package Designer*: è il pannello che si trova in posizione centrale. In esso verranno sviluppati tutti i flussi ETL del progetto. Nel corso

di questo paragrafo verrà fornita una descrizione più dettagliata delle modalità di sviluppo;

- *Solution Explorer*: mostra la cartella del progetto. Ogni progetto SSIS contiene al suo interno tre sotto cartelle:
  - *SSIS Packages*: contiene i pacchetti SSIS che sono stati sviluppati. Ognuno costituisce un flusso ETL;
  - *Connection Managers*: contiene tutte le connessioni che sono state stabilite e che possono essere usate dai pacchetti appartenenti al progetto;
  - *Miscellaneous*: contiene file correlati al progetto, come documenti o immagini;
  - *Project Parameters*: oltre alle tre cartelle elencate è presente un'ulteriore voce in cui è possibile specificare dei parametri del progetto.
- *SSIS Toolbox*: può avere un contenuto differente a seconda che nel Package Designer sia selezionato Data Flow o Control Flow. La distinzione tra le due voci verrà descritta nel seguito del paragrafo. Il pannello contiene tutti i componenti ed i task che possono essere utilizzati per la realizzazione di un flusso ETL.

Gli elementi di principale importanza di un progetto SSIS sono i pacchetti. Ad ogni pacchetto corrisponde un flusso di esecuzione che svolge una determinata attività. I pacchetti possono essere relazionati tra loro, è possibile definire delle precedenze, passare parametri in input e utilizzare variabili definite globalmente o localmente. Lo sviluppo di ogni pacchetto viene effettuato utilizzando i componenti disponibili nella SSIS Toolbox. Essi rappresentano gli elementi di base di un'applicazione SSIS e sono già programmati per essere utilizzati. Per inserirli nella soluzione corrente è sufficiente effettuare il drag and drop del componente desiderato dalla toolbox. L'intera soluzione

viene realizzata importando i componenti ed effettuandone la configurazione in base alle necessità. Lo sviluppo di ogni package è organizzato in due fasi:

- *Control Flow Tab*: è un pannello che risiede all'interno del Package Designer e consente la definizione del flusso di esecuzione. Tale pannello ha dei componenti per lo sviluppo dedicati, che prendono il nome di task. Tutti i componenti che fanno parte di questa categoria non operano direttamente sui dati ma consentono la gestione del flusso di esecuzione. Per esempio è possibile raggruppare task in contenitori o definire un ordine di precedenza per gli stessi task o tra pacchetti differenti;

- *Data Flow Tab*: anche questo pannello, come il precedente, risiede all'interno del Package Designer. Per poterlo utilizzare è necessario inserire all'interno della soluzione del pacchetto un Data Flow Task. Quest'ultimo rientra tra quelli per il controllo di flusso e costituisce uno dei componenti più importanti di un progetto SSIS. Esso infatti consente la realizzazione della parte più importante di un processo ETL, ovvero l'estrazione dei dati da differenti sorgenti, la loro trasformazione ed il caricamento nella destinazione. Per l'implementazione di un Data Flow Task la SSIS Toolbox fornisce dei componenti dedicati, che consentono l'esecuzione di operazioni sui dati.

Control Flow Task

I task di questa categoria consentono di gestire il flusso di esecuzione ETL di un pacchetto. Come già descritto in precedenza è possibile creare contenitori di pacchetti, relazionarli tra loro o definire vincoli di precedenza. Solamente il Data Flow Task consente lo svolgimento di operazioni sui dati. I principali task sono elencati nella tabella seguente:

Task	**Descrizione**
Execute SQL Task	Esegue degli statement SQL su un database ed eventualmente ne restituisce il risultato

File System Task	Esegue operazioni sul file system come sposta, copia, cancella o altro
Data Flow Task	Consente l'esecuzione di operazioni sui dati. Per esso vengono dedicati appositi componenti di sviluppo
FTP Task	Invia e riceve file tramite sessioni FTP
Send Mail Task	Invia e-mail
Web Service Task	Utilizza un servizio web e carica il risultato in un file o variabile
XML Task	Esegue operazioni XML come la validazione di file XML
Execute Process Task	Consente l'avvio di file eseguibili o applicazioni (con o senza parametri)
Execute Package Task	Consente l'esecuzione di altri pacchetti SSIS che si trovano all'interno o all'esterno del progetto corrente
Expression Task	Questo task esegue un'espressione SSIS ed inserisce il risultato in una variabile
Bulk Insert Task	Inserisce o carica dati da un file flat ad un database
Analysis Service Processing Task	Elabora un oggetto di tipo SSAS come un cubo, un database, una dimensione o una partizione

Tabella 2.1: Componenti per il controllo di flusso

Data Flow Task

I componenti di questa categoria effettuano elaborazioni sui dati. Possono essere suddivisi in tre categorie:

- Sorgente: forniscono solamente degli output. Alcuni tipi di sorgente possono avere più output contemporanei;
- Trasformazione: applicano delle trasformazioni ai dati, solitamente hanno almeno un input ed un output;
- Destinazione: sono i componenti che si occupano di memorizzare i dati elaborati. Solitamente ricevono degli input e non hanno output.

Sorgente

I principali componenti di questa categoria si collegano a sorgenti di database o sorgenti flat:

Componente	**Descrizione**
OLE DB Source	Una qualsiasi sorgente che fornisce una connessione OLE DB
ADO.NET Source	Una qualsiasi sorgente che fornisce una connessione ADO.NET
ODBC Source	Una qualsiasi sorgente che fornisce una connessione ODBC
Flat File Source	File di testo o CSV che hanno delimitatori o una lunghezza fissa
Excel Source	Connessione ad un foglio Excel
Raw File Source	File con struttura binaria che consente il passaggio di dati tra differenti Data Flow

Tabella 2.2: Componenti sorgente per il flusso dei dati

Trasformazione

Sono possibili numerose trasformazioni in un Data Flow SSIS. Nella tabella seguente vengono elencate quelle più importanti:

Componente	Descrizione
Derived Column	Crea una nuova colonna
Data Conversion	Effettua una conversione di tipo
Aggregate	Effettua un'aggregazione su una o più colonne del flusso di dati
Conditional Split	Suddivide le righe del flusso di dati sulla base di una o più espressioni
Lookup	Ricerca dei valori nella tabella specificata
Merge Join	Unisce due flussi provenienti da differenti sorgenti
Multicast	Crea una copia del flusso di dati corrente
OLE DB Command	Esegue un'istruzione SQL su una connessione OLE DB
Row Count	Conta il numero di righe del flusso ed inserisce il risultato in una variabile
Script Component	Esegue uno script sul flusso di dati
Sort	Effettua l'ordinamento di un flusso
Union All	Unisce i flussi di dati
Pivot	Sposta i valori dalle righe alle colonne
Unpivot	Scambia le colonne con le righe

Tabella 2.3: Componenti di trasformazione del flusso dei dati

Destinazione

I componenti di questa categoria sono equivalenti a quelli sorgenti, con l'unica differenza che questi ricevono dati in input invece che inviarli in output.

I principali sono i seguenti:

Componente	Descrizione
OLE DB Destination	Una qualsiasi destinazione che consente una connessione OLE DB
ADO.NET Destinaton	Una qualsiasi destinazione che consente una connessione ADO.NET
ODBC Destination	Una qualsiasi destinazione che consente una connessione ODBC
Flat File Destination	File di testo o CSV che hanno delimitatori o una lunghezza fissa
Excel Destination	Foglio di Microsoft Excel
Raw File Destination	File con struttura binaria che consente il passaggio di dati tra differenti Data Flow
Recordset Destination	Variabile di tipo oggetto che può essere utilizzata per ulteriori elaborazioni
SQL Server Destination	I dati vengono caricati in un db SQL Server. È possibile solamente con un'istanza locale di SQL Server

Tabella 2.4: Compoonenti destinazione per il flusso dei dati

2.2.4 SQL Server Analysis Services

SSAS è un servizio che viene usato per gestire dati memorizzati in un DW o data mart. La struttura usata per l'organizzazione dei dati è il cubo multidimensionale che effettua aggregazioni e consente l'esecuzione, in modo efficiente, di report e query complesse. I sistemi analitici utilizzano solitamente tre differenti tipi di architettura per la memorizzazione dei dati multidimensionali:

- *Relational OLAP (ROLAP)*;
- *Multidimensional OLAP (MOLAP)*;
- *Hybrid OLAP (HOLAP)*.

Le tre architetture si differenziano per il modo in cui memorizzano i dati a livello foglia e precalcolano gli aggregati. In ROLAP non vengono mantenuti dati precalcolati. Durante l'esecuzione delle query si effettua l'accesso ai dati della sorgente relazionale e si recuperano quelli d'interesse. MOLAP è un formato di memorizzazione in cui i nodi foglia e le aggregazioni vengono mantenute in un cubo multidimensionale. In caso di memorizzazione in un cubo una certa quantità di dati dovrà essere duplicata. Con ROLAP non sarà necessario spazio addizionale per dati replicati. Inoltre il calcolo delle aggregazioni può essere eseguito in modo rapido utilizzando viste indicizzate. Utilizzando MOLAP alcune aggregazioni vengono precalcolate e memorizzate in formato multidimensionale. Il sistema non dovrà impiegare altro tempo per calcolare tali aggregazioni. Inoltre in MOLAP il database ed il relativo motore sono ottimizzati per lavorare insieme, di conseguenza la risposta alle query sarà generalmente più veloce. HOLAP è un formato ibrido che combina le due architetture descritte in precedenza. Le aggregazioni vengono memorizzate come in HOLAP, mentre le foglie vengono lasciate in formato relazionale. Il principale vantaggio di HOLAP è la non ridondanza del livello foglia.

Per la creazione di un progetto SSAS è necessario utilizzare l'IDE BIDS, il quale fornisce una piattaforma integrata con Visual Studio per la realizzazione e gestione di cubi multidimensionali. I passi da seguire per lo sviluppo di un cubo a partire dal DW relazionale sono i seguenti:

- Definizione delle sorgenti dati;
- Creazione delle Data Source Views;
- Creazione delle dimensioni;
- Modellazione del cubo.

Definizione sorgente dati

Per la realizzazione del cubo è necessario specificare all'interno del progetto SSAS il DW da utilizzare come sorgente dati. Nel modello multidimensionale deve essere presente almeno una sorgente, ma se ne possono utilizzare anche più di una contemporaneamente.

Creazione Data Source Views

Una *Data Source View (DSV)* è un'astrazione di una sorgente che agisce in maniera similare ad una vista relazionale e diventa la base per il cubo e le sue dimensioni. L'obiettivo di una DSV è quello di consentire il controllo sui dati presenti nel progetto in modo indipendente dalle sorgenti sottostanti. Per esempio sulla vista si possono rinominare o concatenare colonne senza andare a modificare le sorgenti originali. È possibile costruire più DSV su una stessa sorgente dati, in modo tale da soddisfare diverse esigenze in uno stesso progetto.

Creazione dimensioni

Prima di arrivare allo sviluppo del cubo si deve effettuare la creazione delle dimensioni d'analisi. Durante questo processo è necessario scegliere, tramite la DSV, la tabella per la quale si vuole realizzare la dimensione. Completata la creazione viene messa a disposizione un'apposita interfaccia per la modellazione della dimensione. Le operazioni realizzabili sono le seguenti:

- Modifica degli attributi: è possibile modificare le proprietà di ogni attributo, rinominarli o effettuarne la cancellazione;
- Definizione dipendenze funzionali: si possono definire dipendenze funzionali tra attributi;
- Creazione gerarchie: è possibile creare gerarchie con gli attributi della dimensione. Per gli attributi presenti in gerarchia verranno aggiunte in automatico da SSAS le relative dipendenze funzionali.

Modellazione del cubo

Per la creazione del cubo è necessario avere definito in precedenza la sorgente dati e la vista su di essa. Il processo di creazione consiste nello scegliere, dalla vista interessata, la tabella dei fatti per la quale dovrà essere creato il gruppo di misure. In automatico SSAS rileverà tutti quegli attributi che hanno valori numerici all'interno della tabella e li suggerirà come indicatori. Si potranno quindi scegliere tutti gli indicatori che dovranno comparire nel gruppo di misure. Nel caso le dimensioni correlate al fatto non fossero state create, SSAS provvederà alla loro creazione in modo automatico. Per ogni cubo è possibile selezionare o aggiungere più tabelle dei fatti, ovvero più gruppi di misure. Le operazioni di modellazione disponibili, successivamente alla creazione, sono le seguenti:

- Modifica struttura del cubo: è possibile aggiungere o rimuovere gruppi di misure o dimensioni;
- Utilizzo delle dimensioni: si specificano come le dimensioni devono essere utilizzate sul cubo;
- Calcolo indicatori: utilizzando la sintassi MDX si possono definire nuovi indicatori calcolati;
- Partizioni: si possono modificare o creare partizioni sul cubo. Le partizioni vengono utilizzate da SSAS per gestire i dati e le aggregazioni dei gruppi di misure del cubo;
- Aggregazioni: si possono modificare o creare aggregazioni sul cubo.

2.2.5 Strumenti di reporting

Per la fase finale del sistema si è deciso di non utilizzare il componente SSRS incluso nella suite Microsoft. Questa scelta è giustificata dal tipo di reportistica richiesta dall'azienda cliente. Gli utenti finali hanno infatti richiesto la possibilità di analizzare i dati secondo due differenti modalità:

- Analisi tramite cubo;
- Report in formato Excel.

Il componente SSRS potrebbe essere utilizzato per soddisfare il secondo dei due punti sopra elencati, tuttavia analizzando il modello dei report richiesti si è giunti alla conclusione che l'utilizzo di Microsoft Excel avrebbe permesso una realizzazione più facile e rapida. SSRS è infatti uno strumento che consente la creazione di report a livello avanzato, mentre la reportistica richiesta ha una complessità che può essere facilment gestita in Excel. A seguito di tale scelta si è deciso di utilizzare lo stesso strumento anche per l'analisi da cubo in quanto Excel integra nativamente la possibilità di interrogare cubi OLAP.

Analisi da cubo

Con questa modalità si predispone il foglio Excel per la connessione al server OLAP utilizzando la configurazione guidata e specificando il cubo che si vuole analizzare. Completata la procedura di connessione Excel costruirà in automatico una *pivot table* dei dati contenuti nel cubo. Quest'ultima è una particolare tabella riassuntiva utilizzata in particolar modo nei sistemi di BI come strumento d'analisi e di reporting. In questo caso la pivot viene costruita direttamente sul cubo, pertanto l'utente può sfruttare le aggregazioni precalcolate per navigarla in modo interattivo inserendo, rimuovendo campi o applicando filtri. Una pivot viene solitamente divisa in quattro aree:

- Righe: si riportano i campi interessati in riga;
- Colonne: si riportano i campi interessati in colonna;
- Valori: specifica quali sono i dati della tabella;
- Filtri: specifica i campi che vengono utilizzati per filtrare i dati della tabella.

Report

Per la creazione dei report l'azienda cliente ha predisposto dei modelli d'esempio. Ognuno di essi prevede delle pivot table costruite a partire dai dati estratti dal DW relazionale e filtrati secondo determinati campi. A differenza dell'analisi da cubo questi report risultano statici, ovvero essi contengono solo i campi specificati nel modello e non è possibile inserirne di nuovi. Per realizzare tali report in Excel è stato utilizzato il linguaggio Visual Basic for Applications (VBA). VBA è un'implementazione di Visual Basic inserita da Microsoft all'interno delle sue applicazioni. Il linguaggio opera sugli oggetti presenti in Excel e permette la creazione di automazioni o l'aggiunta di nuove funzionalità. Per i report del progetto VBA è stato utilizzato per automatizzare lo scaricamento dei dati, il caricamento delle pivot table e la formattazione finale.

2.2.6 Il portale Sharepoint

Tale piattaforma viene utilizzata in entrambi gli ambienti di sviluppo e produzione. Il suo obiettivo è quello di fornire uno spazio comune tra Iconsulting ed il cliente nel quale consegnare i report. Il portale viene realizzato sotto forma di sito web interno e fornisce numerose funzionalità per la gestione dei documenti elettronici. L'organizzazione attuale del sito prevede una cartella per ogni progetto in corso, per il quale è possibile gestirne contenuto ed autorizzazioni. Quando i report vengono completati ed accettati dal cliente vengono rilasciati sull'opportuna cartella del portale, dal quale gli utenti potranno scaricarne una copia per l'utilizzo. La piattaforma fornisce anche la possibilità di utilizzare i report direttamente da browser.

Capitolo 3

Caso di studio: un'azienda di servizi

3.1 Analisi del caso

Prima di passare ad analizzare le fasi che riguardano la realizzazione del progetto viene effettuata una generica introduzione del caso di studio. In questo paragrafo verrà quindi descritta l'organizzazione dell'azienda cliente e successivamente si analizzeranno le esigenze che hanno portato alla progettazione di un sistema di BI. Per concludere si descriverà il processo di analisi del personale, elencando i KPI (*Key Performance Indicators*) utilizzati ed il relativo modello di controllo.

3.1.1 Il profilo aziendale

L'azienda per la quale è stato realizzato il progetto è una società di servizi operante in diversi campi. Essa è nata come cooperativa ed è cresciuta tramite l'acquisizione, nel corso degli anni, di altre società. Il modello di business è quello seguito dalla società capogruppo della cooperativa e riguarda la gestione e l'erogazione in *outsourcing* delle attività "non core" di imprese, enti pubblici o strutture sanitarie. Vengono integrati i più tradizionali servizi di *facility management* quali pulizie, servizi tecnico-manutentivi e servizi lo-

gistici con l'offerta di servizi specialistici come la gestione e manutenzione di impianti di sollevamento o la progettazione, gestione e manutenzione di impianti di illuminazione e di impianti antincendio e per la sicurezza. L'azienda attualmente sta centralizzando i servizi erogati (acquisti, amministrazione, controllo di gestione, personale,...) sulla società capogruppo, con l'obiettivo di semplificarne la gestione e l'esecuzione. Questo processo viene svolto per le società di grandi o medie dimensioni, mentre quelle più piccole rimarranno slegate e manterranno i propri processi interni. Il motivo di tale scelta è giustificato dal fatto che lo sforzo necessario per integrare i servizi delle aziende più piccole è molto più alto del vantaggio che si potrebbe ottenere con l'integrazione. Tra i vari servizi in corso di centralizzazione è presente anche quello di gestione del personale, per il quale è stato sviluppato il sistema di BI. Di conseguenza il progetto di tesi consente l'analisi dei dipendenti di tutte le società per le quali è stata centralizzata la gestione del personale.

3.1.2 Esigenza di progetto

L'esigenza di progetto è partita dall'ufficio personale della società capogruppo. L'ufficio distribuisce mensilmente, all'interno dell'azienda, il reporting che permette di effettuare l'analisi del personale con evidenza dell'andamento storico. In precedenza questi report venivano generati a partire dai sistemi pregressi gestiti da ogni singola società, attraverso scaricamento di file Excel ed integrazione di basi dati personali. Un processo di questo tipo risultava difficoltoso per l'ufficio del personale, in quanto doveva integrare dati diversi, in formati diversi e che arrivavano con periodicità e cadenza diversa. In seguito al processo di centralizzazione e di migrazione verso una coppia di gestionali unici per tutte le società, l'ufficio del personale ha apportato la richiesta di un sistema di BI all'ufficio IT. L'obiettivo di quest'ultimo era quello di integrare automaticamente i dati, creare delle logiche condivise di interpretazione ed interrogazione delle informazioni, fornire delle automatizzazioni che consentissero di diminuire la manualità dell'ufficio personale sui dati stessi e di automatizzare la produzione del reporting mensile per tutta

la società. Parallelamente la BI moderna permette la costruzione di sistemi che forniscono la possibilità d'eseguire analisi self service. Di conseguenza per tutte le esigenze non istituzionali, come l'analisi di una particolare situazione di un dipendente o di un fenomeno relativo al mondo del personale, è possibile utilizzare i sistemi di BI.

3.1.3 Analisi del personale

L'azienda cliente si è strutturata con un ufficio del personale che non svolge solo le attività legate al mondo dei dipendenti, ma gestisce anche il monitoraggio dell'organico, pertanto esso si occupa anche dell'analisi dei dati. Questo è giustificato dalla grande dimensione della struttura organizzativa e dall'elevata capillarità delle responsabilità dei gruppi organizzativi. La struttura dell'azienda si articola, infatti, fino ai gruppi operativi, responsabili generalmente di circa un centinaio di dipendenti. Questi gruppi si occupano principalmente del business e non hanno conoscenza delle logiche relative alla gestione dei dati. L'ufficio personale è stato quindi strutturato anche con l'obiettivo di eseguire analisi dei dipendenti ed inviare la relativa reportistica alle parti interessate dell'azienda. Di conseguenza il processo di analisi del personale risulta essere fortemente centralizzato, con un ufficio preposto a tale scopo che, nel corso del tempo, ha maturato anche l'*ownership* del dato. Uno degli obiettivi dei progetti di BI è quello di riportare a fattore comune la conoscenza, ovvero fare in modo che quest'ultima non dipenda più dall'interpretazione data da una singola persona, ma che venga riportata su un sistema e resa a fattore comune in modo tale che, qualora il processo debba subire delle revisioni, non si dipenda più da una singola persona.

Il processo di analisi del personale viene svolto con cadenza mensile ed ha l'obiettivo di tenere monitorati tutta una serie di KPI. Il principale di essi è quello relativo ai costi, ma ne sono presenti anche degli altri:

- Assenze;
- Ferie;

- Straordinari;
- Trasferte;
- Diaria;
- Infortuni.

Di quelli appena elencati, molti sono ancora legati al monitoraggio dei costi. Per esempio è importante monitorare i costi di trasferta o i costi relativi al monte ferie. Quest'ultimo caso riveste particolare importanza, in quanto un dipendente che accumula un numero elevato di ferie rappresenta un costo per l'azienda, che dovrà poi essere erogato in futuro.

Il modello di controllo dei KPI varia a seconda degli indicatori analizzati. Per tutti quelli di alto livello, quali assenze, ferie o infortuni non viene utilizzato un modello vero e proprio. Per questi tipi di indicatori l'analisi è fortemente legata alla sensibilità dell'ufficio personale o dei responsabili dei gruppi operativi. Viceversa, per l'allocazione dei costi sulle commesse, viene utilizzato un modello di controllo più strutturato. In questo caso si realizza un processo di budgeting delle ore sulle commesse, di conseguenza ogni responsabile di commessa riceve un preventivo di quante ore di lavoro dovrebbero essere erogate sulla determinata commessa. Al termine di ogni mese viene quindi svolto un processo di analisi degli scostamenti tra le ore effettivamente erogate e quelle di budget. Per chiarire la procedura con la quale l'azienda effettua l'allocazione dei costi sulle commesse è necessario chiarire i seguenti concetti:

- *Struttura organizzativa*;
- *WBS (Work Breakdown Structure)*;
- *Timesheet*.

I primi due punti costituiscono le due realtà ortogonali di qualsiasi analisi del conto economico di una società di servizi. Il timesheet è il fenomeno che

permette di raccordare le due strutture. Nei paragrafi seguenti verrà fornita una descrizione più dettagliata dei concetti elencati.

Struttura organizzativa

La materia prima con la quale lavora una qualsiasi società di servizi sono le persone. Queste ultime vengono organizzate su strutture gerarchiche, che poi lavorano per i clienti. In questo caso la struttura delle persone costituisce la gerarchia dei costi. Una struttura organizzativa permette di eseguire un controllo puntale del personale ed anche di verificare che ci sia una consistenza organica adeguata al lavoro che la particolare struttura deve svolgere. Essa è quindi basata sull'organizzazione interna della società. L'azienda cliente ha deciso di organizzarsi su una struttura molto flessibile, contenente dodici livelli, che è stata adattata al modello di business.

WBS

Una WBS è un sistema logico di decomposizione di un'attività in componenti additive. La definizione di insiemi contenenti attività atomiche (*Work Breakdown Elements*) è volto a facilitare l'analisi di dettaglio ed a definire, in modo preciso, il perimetro delle attività più articolate. La WBS fornisce una base organizzativa solida per lo sviluppo, la pianificazione ed il controllo delle attività. Grazie alla suddivisione in elementi incrementali risulta più facile raggiungere condizioni di modularità sia in termini organizzativi, che in quelli contabili e di reportistica. Per il progetto in esame la WBS viene utilizzata per realizzare la strutturazione dei clienti e costituisce la gerarchia che permette l'analisi dei ricavi. La strutturazione adottata è quella mostrata in Figura 3.1.

Timesheet

Il timesheet costituisce il metodo che permette di registrare le ore di lavoro svolte da un dipendente per una certa attività. Per il progetto in esame

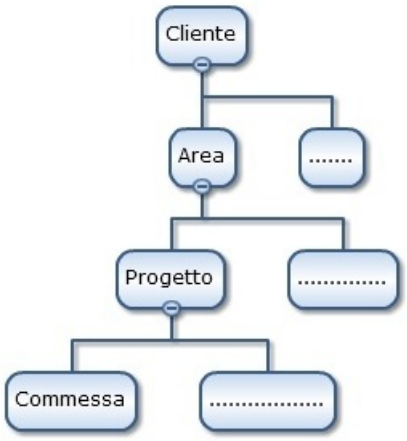


Figura 3.1: WBS rappresentate la strutturazione dei clienti

questo concetto viene utilizzato per definire le ore svolte dal dipendente su di una commessa, esso permette quindi di raccordare la struttura organizzativa con la WBS. Attraverso il timesheet è quindi possibile allocare i costi del personale sulle commesse dei clienti, ad essi dovranno poi essere aggiunti gli altri costi, quali i materiali, per redigere il conto economico finale. Principalmente una società di servizi chiede ai propri dipendenti di compilare un timesheet per poterne allocare i costi, il secondo fine è quello del monitoraggio degli altri KPI, quali assenze, straordinari, ferie o trasferte.

3.2 Modellazione back end

In questo paragrafo verrà introdotta l'architettura del sistema e saranno esaminati i componenti del progetto relativi alla parte di back end. La prima fase da cui si è partiti per realizzare il sistema è stata l'analisi delle sorgenti. L'obiettivo era quello di capire le modalità con la quale i dati sarebbero stati messi a disposizione, le logiche correlate ad essi e la relativa strutturazione. Nel mondo del personale le casistiche possibili sono numerose, pertanto si sono resi necessari più incontri tra Iconsulting e l'azienda cliente per concordare le modalità di calcolo dei KPI e le relative logiche di gestione. Oltre a questa

prima complicazione ne è presente una seconda dovuta all'integrazione delle differenti sorgenti coinvolte, che verrà trattata nel successivo sottoparagrafo. Completata la prima fase d'analisi viene svolta la vera e propria realizzazione del sistema di BI per la parte di back end, che comprende:

- Import;
- Staging Area (SA);
- Datamart (DM);
- Cubo multidimensionale.

L'ultimo punto ha una duplice classificazione. Il cubo viene considerato come back end, perché comunque include il lavoro di modellazione multidimensionale, ma allo stesso tempo deve anche essere considerato come front end, in quanto può essere utilizzato per svolgere analisi libere.

3.2.1 Architettura del sistema

Nella Figura 3.2 viene mostrata l'architettura adottata per il progetto. Inizialmente si procede importando le estrazioni Zucchetti e SAP all'interno della SA, successivamente si effettua la pulizia dei dati attraverso la fase di ETL ed infine si costruisce il DM, che contiene i fatti e le dimensioni d'interesse oltre alla vista con la quale verranno costruiti i report. A partire dal DM si effettua la creazione del cubo, realizzando il relativo modello multidimensionale e quella dei report, sfruttando la vista precedentemente costruita. Questi ultimi due componenti costituiscono la parte di front end per gli utenti finali che potranno scegliere se interrogare il cubo, o analizzare i dati dei report.

3.2.2 Le sorgenti transazionali

Il processo di centralizzazione dei servizi dell'azienda cliente, ha portato ad una migrazione dei gestionali utilizzati dalle varie società. Per quanto

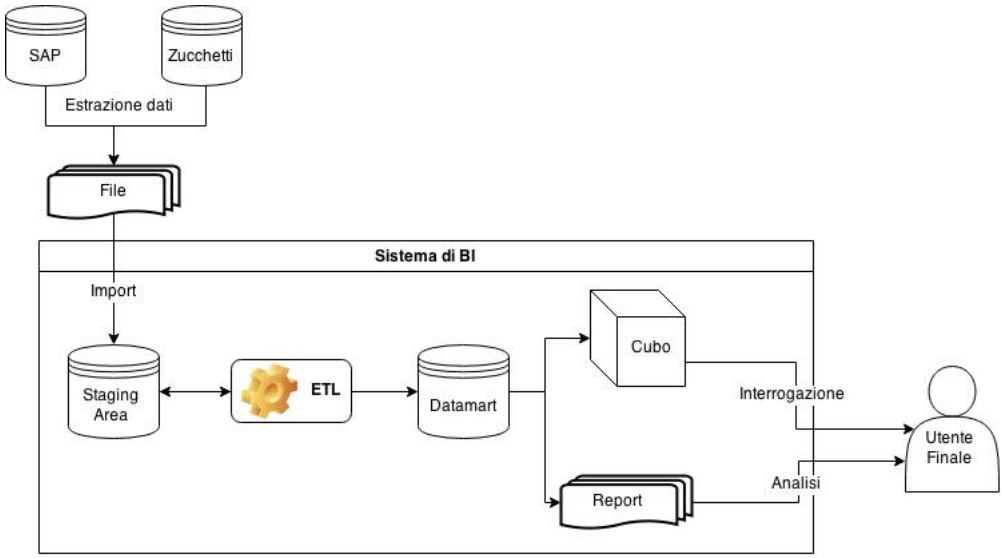


Figura 3.2: Architettura del sistema di BI

riguarda l'analisi del personale, in precedenza il processo veniva realizzato singolarmente dalle società, ognuna delle quali utilizzava i propri gestionali. A valle della centralizzazione dei servizi si è deciso di utilizzare una coppia di gestionali unici:

- *Zucchetti*;
- *SAP*.

Il primo viene utilizzato dall'azienda per la gestione delle buste paghe dei dipendenti e, per il progetto in esame, fornisce i dati per la valorizzazione di tutti i KPI considerati. Il secondo fornisce le informazioni anagrafiche dei dipendenti, la relativa struttura organizzativa e la gerarchia delle commesse per le quali hanno lavorato. I due sistemi lavorano con logiche differenti, di conseguenza si è resa necessaria una complessa fase d'integrazione, per raccordare le informazioni dei dipendenti presenti nei due mondi. Per entrambi i sistemi i file, in formato CSV, vengono messi a disposizione in delle apposite share di rete, dalle quali vengono recuperati e caricati tramite flussi ETL. La modalità di caricamento è giornaliera, i file vengono resi disponibili

durante la notte e caricati dai processi eseguiti al mattino. L'esecuzione di questi processi viene schedulata con orari precisi, per evitare la sovrapposizione con gli altri progetti di BI e con l'obiettivo di rendere disponibili, prima dell'inizio della giornata lavorativa, i dati al cliente. Nei sottoparagrafi seguenti verranno esaminate le due sorgenti e la tecnica utilizzata per la loro integrazione.

Zucchetti

Zucchetti è un'azienda italiana che produce soluzioni software e servizi per imprese. I prodotti forniti sono soluzioni gestionali e contabili che consentono di svolgere le seguenti operazioni:

- Gestione delle paghe;
- Gestione documentale;
- Gestione del personale e delle risorse umane in aziende di grandi dimensioni;
- Gestione contabile e fiscale;
- Business intelligence.

L'azienda cliente utilizza principalmente il software per la gestione delle paghe e del relativo personale. Da questa sorgente, infatti, vengono presi tutti gli indicatori per la costruzione del sistema di BI. Zucchetti offre servizio all'azienda cliente con modalità cloud, di conseguenza il software è esterno all'azienda e viene gestito dalla stessa Zucchetti. Quest'ultima utilizza un proprio DW interno contenente tutte le informazioni d'interesse dell'azienda cliente. Da questo DW Zucchetti estrae i dati che vengono poi messi a disposizione di Iconsulting su apposito sito FTP. Di seguito vengono elencate le estrazioni ed il relativo contenuto:

- *Soggetti*: elenca i dipendenti dell'azienda cliente. Contiene il codice che identifica il dipendente, il mese di validità dei dati ed alcune informazioni anagrafiche. Queste ultime vengono prese direttamente da SAP tramite un automatismo realizzato dalla stessa Zucchetti;

- *Rapporti*: elenca i rapporti di lavoro dei dipendenti con le società. Contiene il mese di validità, il codice del dipendente, il codice della società per la quale lavora ed ulteriori informazioni aggiuntive sul rapporto quali la qualifica, la regione di lavoro, lo stato del dipendente (in forza o cessato), il tipo di rapporto, il tipo di retribuzione e la data d'inizio ed eventuale fine del rapporto. Ogni dipendente può avere più rapporti di lavoro con società differenti;

- *Agenti di rischio*: specifica l'agente di rischio che ha portato all'infortunio. Contiene la data dell'infortunio, il codice del dipendente che l'ha subito, codice e descrizione dell'agente di rischio. Questa estrazione viene utilizzata per la costruzione del fatto relativo agli infortuni;

- *Forme di accadimento*: specifica la forma di accadimento dell'infortunio. Contiene la data dell'infortunio, il codice del dipendente che l'ha subito, codice e descrizione della forma di accadimento. Anche questa estrazione, come la precedente, viene utilizzata per la costruzione del fatto relativo agli infortuni;

- *Infortuni*: elenca gli infortuni subiti dai dipendenti. Contiene la data dell'infortunio, il codice del dipendente che l'ha subito ed alcune informazioni sull'infortunio quali la natura e la sede dell'infortunio. In questa estrazione oltre alle informazioni anagrafiche sono presenti due indicatori utilizzati per la costruzione del fatto degli infortuni: i giorni d'infortunio calcolati e gli effettivi giorni d'assenza del dipendente;

- *Budget*: elenca le ore stimate di lavoro sulle commesse. Le informazioni contenute sono il mese di validità, la commessa, codice e descrizione

dell'attività che deve essere svolta, la versione della stima e le ore di lavoro pianificate;

- *Timesheet*: elenca le ore di lavoro svolte dai dipendenti sulle commesse. Contiene il codice del dipendente, quello dell'azienda, la commessa per la quale ha lavorato, il mese ed il giorno di validità, il tipo di attività svolta ed il relativo numero di ore. Le informazioni contenute in questa estrazione hanno granularità giornaliera, ma sul DW finale verranno aggregate mensilmente;

- *Costi*: elenca tutte le voci di costo di un dipendente. È l'estrazione principale e contiene il mese di validità, il codice del dipendente, un campo che viene valorizzato a seconda della voce di costo ed il relativo valore numerico. I principali indicatori in esso contenuti sono le ore di assenza, quelle lavorate, ferie maturate e godute, costi del dipendente ed eventuali oneri aggiuntivi.

Nella tabella seguente vengono invece mostrate la frequenza di aggiornamento delle estrazioni elencate in precedenza:

Estrazione	**Frequenza aggiornamento**
Soggetti	Da Lunedì a Venerdì
Rapporti	Da Lunedì a Venerdì
Agenti di rischio	Da Lunedì a Venerdì
Forme di accadimento	Da Lunedì a Venerdì
Infortuni	Da Lunedì a Venerdì
Budget	Sabato
Timesheet	Sabato
Costi	11 e 20 del mese

Tabella 3.1: Dettaglio aggiornamenti estrazioni Zucchetti

Le estrazioni specificate vengono rese disponibili giornalmente, ma contengono dati solo per i giorni indicati in tabella.

SAP

SAP è una multinazionale europea leader nel settore degli ERP e nell'offrire soluzioni software alle imprese. Nell'azienda cliente viene utilizzata una soluzione di tale fornitore per gestire le anagrafiche dei dipendenti, la loro struttura organizzativa e le commesse per le quali hanno lavorato. Il servizio viene offerto e gestito da una società esterna ma, a differenza di Zucchetti, il sistema è collocato internamente all'azienda cliente. Di conseguenza le estrazioni vengono pubblicate su una share di rete interna. Il processo ETL che si occupa di caricare i dati provenienti da SAP è suddiviso in due parti:

- Caricamento anagrafiche dei dipendenti;
- Caricamento della struttura organizzativa e collegamento con il dipendente.

Per l'anagrafica dei dipendenti vengono messe a disposizione delle apposite estrazioni che prendono il nome di *Infotype*. SAP utilizza gli infotype per memorizzare tutte le informazioni relative ai dipendenti utili per scopi di amministrazione. Ogni file viene identificato da un nome e da un codice a quattro cifre, in particolare la suddivisione numerica adottata è la seguente:

- Infotype 0000 a 0999: contengono tutte le informazioni per la gestione del personale;
- Infotype 1000 a 1999: contengono informazioni per la gestione delle organizzazioni;
- Infotype 2000 a 2999: contengono informazioni per la gestione del tempo;
- Infotype 4000 a 4999: contengono informazioni per la gestione delle assunzioni;

- Infotype 9000 a 9999: range riservato per la creazione di infotype personalizzati.

Nello specifico gli infotype utilizzati per la costruzione dell'anagrafica dei dipendenti sono i seguenti:

- *IT0000*: elenca le azioni eseguite sui dipendenti. Contiene il codice identificativo del dipendente, il periodo di validità ed una descrizione dell'azione intrapresa;
- *IT0001*: elenca l'allocazione organizzativa dei dipendenti. Contiene il codice del dipendente, la società a cui appartiene, il periodo di validità ed altre informazioni anagrafiche;
- *IT0002*: contiene i dati personali dei dipendenti;
- *IT0006*: elenca gli indirizzi dei dipendenti;
- *IT0007*: specifica il tipo d'orario di lavoro dei dipendenti. In particolare in esso vengono riportati la percentuale orario di lavoro, il numero di ore e di giorni lavorativi previsti settimanalmente;
- *IT0008*: specifica il tipo di contratto dei dipendenti;
- *IT0022*: memorizza informazioni sul tipo di formazione dei dipendenti;
- *IT0050*: indica in che modo viene effettuata la rilevazione delle presenze. Per il progetto in esame i dipendenti vengono gestiti con il timesheet Zucchetti;
- *IT0105*: gestisce le informazioni di contatto dei dipendenti;
- *IT0155*: contiene dati amministrativi aggiuntivi;
- *IT0315*: specifica i centri di costo dei dipendenti.

Per quanto riguarda i dati relativi a struttura organizzativa e gerarchia delle commesse, le estrazioni non vengono fornite sotto forma di infotype ma viene utilizzato un protocollo di comunicazione standard detto IDOC (*Intermediate document*). Quest'ultimo è uno standard per lo scambio di documenti elettronici tra le applicazioni scritte per il business SAP o tra applicazioni SAP e programmi esterni. Il protocollo viene utilizzato per le trasmissioni asincrone di file. Ogni documento può essere trasmesso al destinatario senza che venga richiesta alcuna connessione ad un database centrale. Ai fini del progetto vengono messe a disposizione due estrazioni che presentano il seguente contenuto:

- *Struttura organizzativa*: descrive la struttura organizzativa. La costruzione della gerarchia viene effettuata riportando, per ogni oggetto, le sue relazioni con gli altri oggetti esistenti ed il periodo di validità. La Tabella seguente riporta le relazioni possibili ed il loro significato:

Oggetto	**Def. Relazione**	**Ogg. destinatario**
O	è il superiore di	O
O	comprende	S
O	è diretto da	S
O	Delegato autorizzatore	S
S	Titolare	P
S	Somministrato	P
S	Resp. Ad Interim	P
S	Delegato autorizzatore	P
O	Autorizzatore	S
O	Compilatore	S
O	Gestisce la Commessa	9S
9S	Autorizzatore	S

9S	Compilatore	S
S	viene descritto da	C

Tabella 3.2: Relazioni tra i vari oggetti della struttura organizzativa

Dove O è l'unità organizzativa, S la posizione, P la persona, 9S la commessa e C la mansione;

- *Commessa*: elenca la gerarchia delle commesse. Per ogni commessa viene riportato il periodo di validità ed i relativi livelli fino all'area.

Come per Zucchetti anche in SAP i file vengono resi disponibili con modalità giornaliera. La frequenza d'aggiornamento è anch'essa giornaliera, ovvero queste estrazioni potrebbero contenere ogni giorno nuovi dati.

Integrazione Zucchetti-SAP

Come già introdotto in precedenza, l'azienda cliente utilizza i due sistemi per scopi diversi. SAP si occupa della gestione della pianta organica e delle relative commesse, mentre Zucchetti gestisce tutto il front end relativo ai dipendenti. Per poter operare correttamente, Zucchetti deve conoscere la struttura della pianta organica. In seguito ad una nuova assunzione, il dipendente viene registrato su SAP ed entra in Zucchetti con un automatismo. Questa procedura viene svolta da Zucchetti con l'unico scopo di gestire la propria anagrafica, di conseguenza non è prevista la riconciliazione in seguito alle estrazioni messe a disposizione per l'azienda cliente. Si è resa quindi necessaria un'analisi per definire la procedura da adottare per l'integrazione. La principale complicazione era dovuta alla diversa logica utilizzata dai due sistemi per la gestione dei dati. In SAP, infatti, il codice identificativo del dipendente (*CID SAP*) è univoco, mentre in Zucchetti il codice è univoco per società (*CID + società*). Analizzando le varie estrazioni messe a disposizione da Zucchetti è stato individuato un campo denominato *codice soggetto*

esterno nell'estrazione relativa ai *Soggetti*. Tale campo si riferisce al codice del dipendente presente in SAP, che risulta essere univoco, pertanto si è partiti da quest'ultimo per effettuare l'integrazione. La procedura adottata è la seguente:

1. Collegamento Zucchetti-SAP: vengono integrati i dipendenti dei due mondi attraverso il *codice soggetto esterno*;

2. Dati SAP: utilizzando la chiave *CID SAP* si recuperano tutte le informazioni anagrafiche dei dipendenti si crea la struttura organizzativa ed infine la gerarchia delle commesse;

3. Dati Zucchetti: utilizzando la chiave *CID + società* si recuperano tutte le informazioni relative a dipendenti, timesheet, costi, budget e infortuni.

Le chiavi utilizzate dai due sistemi vengono portate fino al DM finale in modo tale da utilizzare l'una o l'altra a seconda dei dati con cui si lavora.

3.2.3 Import

Completata la descrizione della struttura e dell'organizzazione delle sorgenti transazionali si passa ad analizzare la procedura, realizzata mediante SSIS, con cui i dati sono stati importati nella sorgente relazionale. Nel seguito di questo paragrafo verrà quindi analizzato l'algoritmo utilizzato per l'import dei dati.

Modalità d'aggiornamento dei dati

Come descritto nel Capitolo 1 esistono due modalità differenti per l'aggiornamento dei dati di un sistema di BI: full-refresh o incrementale. Risulta quindi necessario, durante la fase di import, stabilire le modalità con cui le varie tabelle relazionali dovranno essere aggiornate. I criteri da utilizzare per la scelta sono:

- Storicizzazione dei dati;
- Performance del caricamento.

Una tabella che non subisce modifiche frequenti nel corso del tempo, offre prestazioni di caricamento migliori se aggiornata in modo incrementale. Viceversa una tabella che evolve velocemente, offre prestazioni di caricamento migliori se aggiornata in full-refresh, dato che il tempo impiegato per controllare tutte le righe modificate sarebbe maggiore rispetto a quello per la creazione dell'intera tabella. Oltre a questo aspetto è necessario considerare anche l'esigenza o meno della storicizzazione dei dati. Con la modalità incrementale la storicizzazione viene garantita, mentre per poter mantenere la storicizzazione in quella full-refresh è necessario ricevere lo storico dei dati. In virtù delle considerazioni fatte, per il progetto in esame sono state adottate le seguenti modalità:

- Dati Zucchetti: tutte le estrazioni vengono importate in modalità incrementale;
- Anagrafica SAP: estrazioni importate in modo incrementale;
- Struttura organizzativa SAP: estrazione gestita in modalità full-refresh. Giornalmente vengono caricati tutti i dati per ricostruire la struttura organizzativa valida alla data;
- Commesse SAP: questo tipo di estrazione viene gestita in modo ibrido. L'import viene svolto in modalità full-refresh, ma la relativa tabella di SA viene poi aggiornata in modo incrementale.

Algoritmo di import dei dati

In questo paragrafo verrà mostrata la procedura adottata per l'import incrementale dei dati. Con *upsert* si intende l'operazione di aggiornamento di una tabella incrementale. Essa consiste nell'inserimento di nuovi record o nella modifica di quelli già esistenti. La procedura utilizza nel progetto è composta da due fasi:

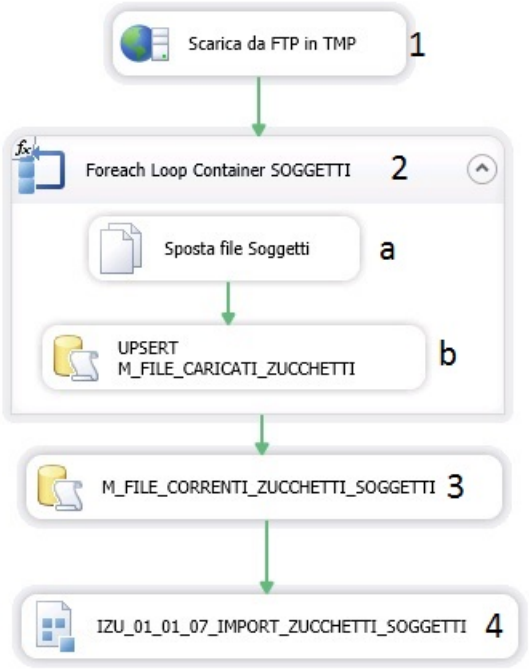


Figura 3.3: Esempio di fase 1 dell'import per i dipendenti Zucchetti

1. Scaricamento delle estrazioni sul server locale;
2. Caricamento delle tabelle.

In Figura 3.3 viene mostrata la prima fase della procedura, svolta in questo caso per i file dei *soggetti*:

1. Da FTP (Zucchetti) o dalla share di rete (SAP) vengono scaricati tutti i file e salvati in una cartella temporanea. Il pacchetto SSIS viene configurato con il percorso dal quale scaricare i dati;
2. Per ogni file dei *Soggetti* scaricato in precedenza:
   (a) Lo si sposta nella cartella denominata *Soggetti*;
   (b) Utilizzando l'upsert, si inserisce il file considerato nella tabella di debug *M_FILE_CARICATI_ZUCCHETTI*. Essa contiene il nome del file, un flag *CARICATO* per indicare se il file è stato caricato a

sistema, la data di caricamento ed una colonna che specifica il tipo di file (soggetti, rapporti,...). Questa tabella viene storicizzata, pertanto conterrà l'elenco di tutti i file caricati.

3. In *M_FILE_CORRENTI_ZUCCHETTI_SOGGETTI* vengono caricati tutti i file, di tipo *Soggetti*, che sono stati esaminati. La tabella contiene il nome del file, la data di caricamento e due flag *CARICATO_T* e *CARICATO_TT* che verranno utilizzati nella fase successiva. A differenza della precedente, questa tabella viene svuotata ad ogni import;

4. Si esegue il pacchetto SSIS che contiene la seconda fase dell'import.

Dal punto 2 al 4 la procedura risulta la medesima per tutti i file Zucchetti e SAP. Per lo svolgimento della seconda fase dell'import sono essenziali due tabelle:

- *T_ZUCCHETTI_SOGGETTI*: in essa vengono caricate, una alla volta, le estrazioni relative ai *soggetti*. Questa tabella viene aggiornata in modalità full-refresh, pertanto è svuotata ad ogni caricamento;
- *TT_ZUCCHETTI_SOGGETTI*: contiene tutte le estrazioni dei *soggetti*. La tabella viene aggiornata in upsert e risulta quindi storicizzata.

Il suo funzionamento viene mostrato in Figura 3.4:

1. Si inseriscono nella variabile *vTabelle_Z_Soggetti* tutti i nomi dei file contenuti nella tabella *M_FILE_CORRENTI_ZUCCHETTI_SOGGETTI*;

2. Per ogni nome di file contenuto nella variabile:

   (a) Si aggiorna la connessione puntando al file da caricare;

   (b) Si svuota la *T_ZUCCHETTI_SOGGETTI*;

   (c) Si inserisce il file considerato nella *T_ZUCCHETTI_SOGGETTI*;

   (d) Nella *M_FILE_CORRENTI_ZUCCHETTI_SOGGETTI* viene settato il flag *CARICATO_T*, utilizzato per indicare che il file è stato caricato nella *T_ZUCCHETTI_SOGGETTI*;

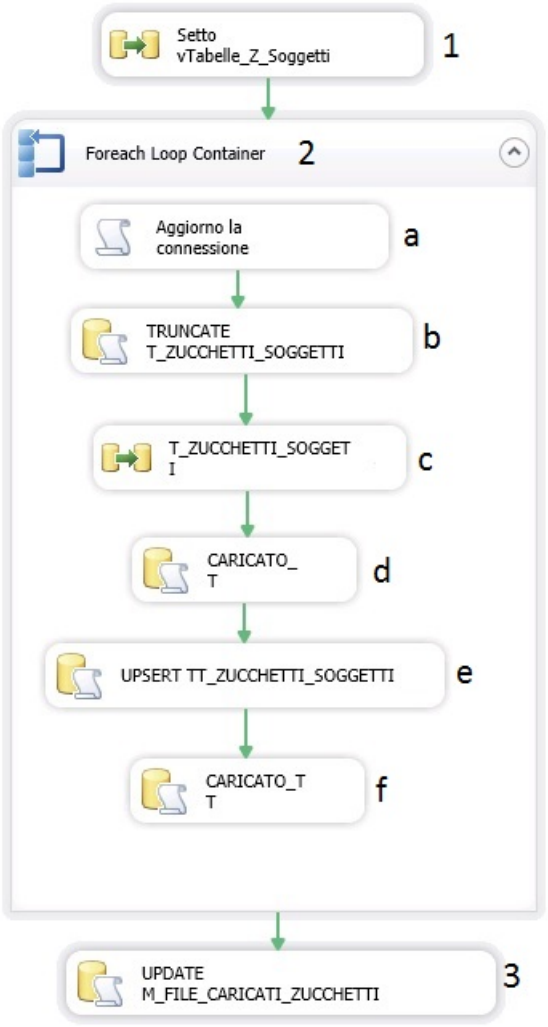


Figura 3.4: Esempio di fase 2 dell'import per i dipendenti Zucchetti

(e) Si aggiorna in upsert la *TT_ZUCCHETTI_SOGGETTI*, inserendo le nuove righe non presenti o modificando quelle già esistenti;

(f) Nella *M_FILE_CORRENTI_ZUCCHETTI_SOGGETTI* viene settato il flag *CARICATO_TT*, utilizzato per indicare che il file è stato caricato nella *TT_ZUCCHETTI_SOGGETTI*;

3. Nella *M_FILE_CARICATI_ZUCCHETTI* viene settato il flag *CARICATO*. Quest'ultimo viene valorizzato per ogni file che è stato corret-

tamente caricato nella *TT_ZUCCHETTI_SOGGETTI*.

Il comportamento del pacchetto SSIS che svolge la fase 2 viene replicato per tutte le altre estrazioni. Per i dati SAP che vengono aggiornati in full-refresh (struttura organizzativa e commesse) la fase 1 viene svolta allo stesso modo, mentre nella fase 2 si effettua un *Drop And Create* delle tabelle interessate.

3.2.4 Staging Area

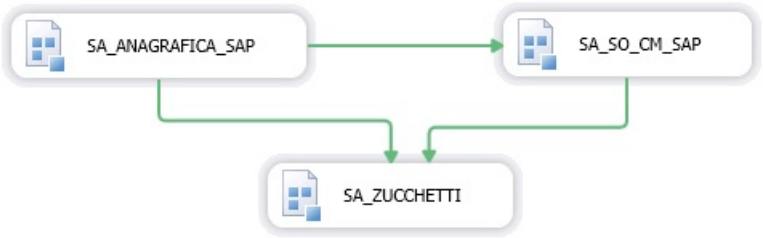


Figura 3.5: Organizzazione della staging area del progetto

I dati risultanti da questa fase sono quelli che hanno subito il processo di ETL. In questa fase viene anche effettuata l'integrazione delle due sorgenti Zucchetti e SAP, nello specifico la strutturazione adottata è quella mostrata in Figura 3.5. Si parte realizzando la fase di staging per l'anagrafica SAP dei dipendenti (*SA_ANAGRAFICA_SAP*), nella fase successiva si costruisce la struttura organizzativa, la gerarchia delle commesse e si collega il dipendente alla struttura (*SA_SO_CM_SAP*). Nell'ultima fase, a partire dal dipendente SAP, si costruiscono tutte le tabelle per i dati Zucchetti (*SA_ZUCCHETTI*). Nei paragrafi seguenti verranno esaminate tutte e tre le fasi di staging.

SA_ANAGRAFICA_SAP

In Figura 3.6 viene mostrato il flusso ETL che realizza la SA per l'anagrafica dei dipendenti. I passi svolti sono i seguenti:

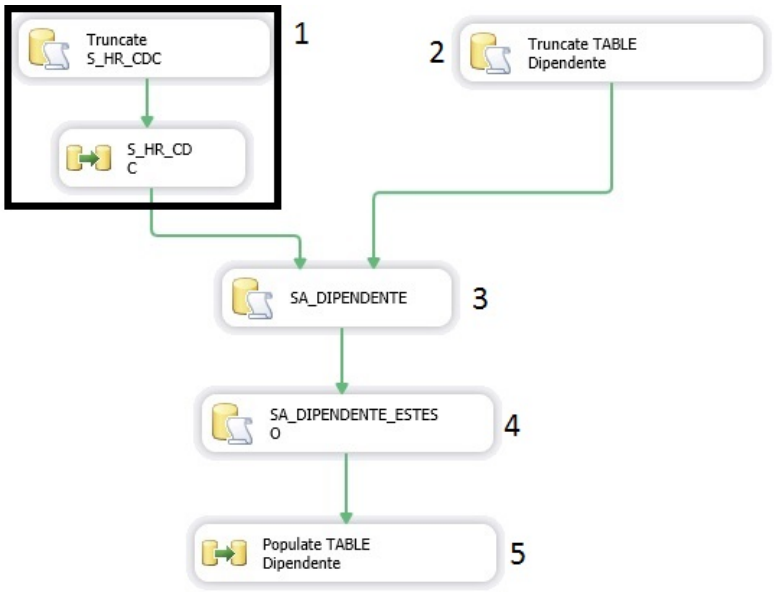


Figura 3.6: Organizzazione della staging area per l'anagrafica dei dipendenti

1. Si aggiorna la *S_HR_CDC* in modalità full-refresh. Questa tabella viene alimentata a partire dai dati provenienti dal controllo di gestione e contiene i centri di costo a cui i dipendenti afferiscono. Questi ultimi vengono inseriti nell'anagrafica del dipendente;

2. Si svuota la tabella *Dipendente*. Essa viene utilizzata per alimentare gli altri DM che possiedono una dimensione relativa al dipendente;

3. Viene creata la *SA_DIPENDENTE* utilizzando le informazioni dagli infotype importati in precedenza. Durante la creazione viene anche realizzata la fase di ETL, nella quale si selezionano i dati d'interesse, si effettua la pulizia degli stessi e si applicano delle regole per individuare i dipendenti assunti e cessati. Questa tabella è quella effettivamente utilizzata nel progetto per la gestione dei dipendenti;

4. Si crea la *SA_DIPENDENTE_ESTESO*. Questa tabella è equivalente a quella creata in precedenza, ma possiede alcune informazioni anagrafiche aggiuntive;

5. Si popola la *Dipendente* tramite la *SA_DIPENDENTE*.

SA_SO_CM_SAP

Lo svolgimento di questa SA viene effettuato a partire dalle commesse, per poi passare alla creazione della struttura organizzativa ed al collegamento con il dipendente. Il flusso per le commesse opera nel modo seguente:

1. Si applica la procedura di ETL alla *T_SEDE_TECNICA*, tabella che contiene le commesse appartenenti al file importato in precedenza (modalità full-refresh). Durante questa fase vengono scartati tutti i record che presentano chiavi replicate o con dei valori nulli e viene inoltre effettuata la conversione dei dati ad un formato standard;

2. Si aggiorna in upsert la *S_SEDE_TECNICA*. Essa contiene lo storico di tutte le commesse.

Per quanto riguarda la struttura organizzativa la procedura adottata consiste nel:

1. Creare la *S_OGG_STRUTTURA_ORGANIZZATIVA* a partire dai dati importati in precedenza. Essa contiene l'elenco di tutti gli oggetti che compongono la struttura organizzativa. Durante questa fase vengono applicate le classiche operazioni di ETL per la pulizia e selezione dei dati;

2. Creare la *S_REL_STRUTTURA_ORGANIZZATIVA* a partire dai dati importati in precedenza. Essa contiene l'elenco di tutte le relazioni tra gli oggetti della struttura organizzativa. Anche in questo caso, come in precedenza, si realizzano le operazioni di ETL dei dati;

3. Attraverso una procedura ricorsiva si legano gli oggetti e le relazioni ed infine si aggiungono i dipendenti per completare la struttura organizzativa valida alla data.

Figura 3.7: Organizzazione della staging area per i dati Zucchetti

SA_ZUCCHETTI

La Figura 3.6 mostra la fase di staging per la parte Zucchetti. In essa vengono legati i dati appartenenti alle due sorgenti transazionali. Le operazioni realizzate sono le seguenti:

1. Si costruisce la tabella dei rapporti che contiene l'elenco di tutti i rapporti dei dipendenti in Zucchetti;

2. Viene costruita la tabella dei soggetti, sfruttando le informazioni anagrafiche recuperate da SAP;

3. Si effettua il collegamento rapporti-dipendenti;

4. Si integrano i dipendenti Zucchetti con quelli SAP. Nello specifico si parte dalla tabella dei soggetti e si effettua l'integrazione con il dipendente costruito in SAP tramite il *codice soggetto esterno*;

5. Si crea la tabella relativa ai costi dei dipendenti, contenente tutti i relativi indicatori;

6. Si crea la tabella contenente tutti gli indicatori relativi al timesheet dei dipendenti;

7. Si crea la tabella che contiene le ore di lavoro stimate per le commesse;

8. Si crea la tabella degli infortuni. Essa include anche le informazioni relative alle forme di accadimento e agli agenti di rischio.

Anche in questa fase, come nelle precedenti, si applicano le classiche operazioni di ETL per eliminare i record che hanno valori inconsistenti in chiave e per esprimere i dati in un formato uniforme. Al termine della staging viene certificata la qualità dei dati, ovvero è possibile garantirne la consistenza. Tutte le tabelle che riportano il prefisso *SF* diventeranno dei fatti in fase di costruzione del DM.

3.2.5 Datamart

Lo scopo di questa fase è quello di completare la definizione delle tabelle sulle quali verranno costruiti il cubo ed i report. Per la creazione del DM vengono raffinate le tabelle create in fase di staging con l'obiettivo di definire i fatti e le dimensioni utilizzate. Le principali operazioni svolte in questa fase sono:

- Selezione dei soli record d'interesse per fatti e dimensioni a partire dalle tabelle di staging. Queste ultime contengono più dati di quelli necessari per facilitare la gestione di nuove richieste da parte dei clienti;

- Calcolo degli indicatori per i fatti;

- Aggregazione dei dati al corretto livello di dettaglio;
- Importazione nei fatti delle chiavi surrogate delle dimensioni.

Come in tutti i progetti di BI, la realizzazione del DM viene svolta a partire dalle dimensioni. In particolare quelle utilizzate per il progetto in esame sono le seguenti:

- *Tempo*: memorizza l'informazione temporale;
- *Dipendente*: contiene tutte le informazioni relative al dipendente;
- *Struttura organizzativa*: contiene le informazioni sull'organizzazione dei dipendenti;
- *Commessa*: memorizza le informazioni relative alle commesse.

Completata la definizione delle dimensioni è possibile passare a quella dei fatti. Per ognuno di essi, come descritto in precedenza, vengono svolte le aggregazioni opportune, calcolati i relativi indicatori ed aggiunte le chiavi delle dimensioni. I fatti realizzati sono:

- *Costi*: contiene tutti i costi dei dipendenti;
- *Timesheet*: memorizza il timesheet dei dipendenti;
- *Budget*: contiene le ore stimate di lavoro sulle commesse;
- *Infortuni*: memorizza gli infortuni dei dipendenti.

Oltre ai fatti sopra elencati, in fase di creazione del DM viene anche creata una vista sulla quale si appoggeranno i report finali. Essa ha l'unico scopo di integrare le informazioni provenienti dai differenti fatti, in modo tale da utilizzare un'unica sorgente con la quale alimentare i report.

3.2.6 Cubo

Rappresenta l'ultima fase per la modellazione back end. Un cubo svolge un duplice ruolo, deve essere considerato come back end, in quanto include la fase di modellazione dimensionale con la quale si definisce la sua struttura, ma allo stesso modo fa parte anche del front end, dato che può essere utilizzato dagli utenti finali per svolgere analisi libere. La modellazione del cubo è stata realizzata direttamente da SSAS a partire dalle tabelle presenti nel DM. Grazie allo strumento è stato possibile creare gerarchie altrimenti non definibili sul modello relazionale. Nei paragrafi seguenti verranno mostrate, attraverso DFM, le modellazioni multidimensionali realizzate per i fatti d'interesse.

Costi

In Figura 3.8 viene mostrato il DFM relativo ai costi dei dipendenti. Il livello di granularità utilizzato per il fatto permette di tenere traccia di tutti gli indicatori dei costi relativi ai dipendenti, afferenti ad un certo livello della struttura organizzativa, che hanno lavorato per una data commessa in un dato giorno. Questo rappresenta il fatto principale dell'intero progetto, dato che su esso sono allocati la maggior parte degli indicatori d'interesse. Tra le misure presenti il *costo totale* viene derivato a partire da *costo ordinario + costo straordinari + costo supplementari*, tuttavia si è deciso di inserire direttamente il valore calcolato per migliorare l'efficienza durante la navigazione del cubo.

Timesheet

Grazie a questo fatto è possibile analizzare le ore svolte dai dipendenti, afferenti ad un certo livello della struttura organizzativa per una data commessa in un dato giorno. Oltre alle ore lavorate si tiene traccia anche di quelle arretrate. È presente un'ulteriore dimensione *Prestazione e Lavora-*

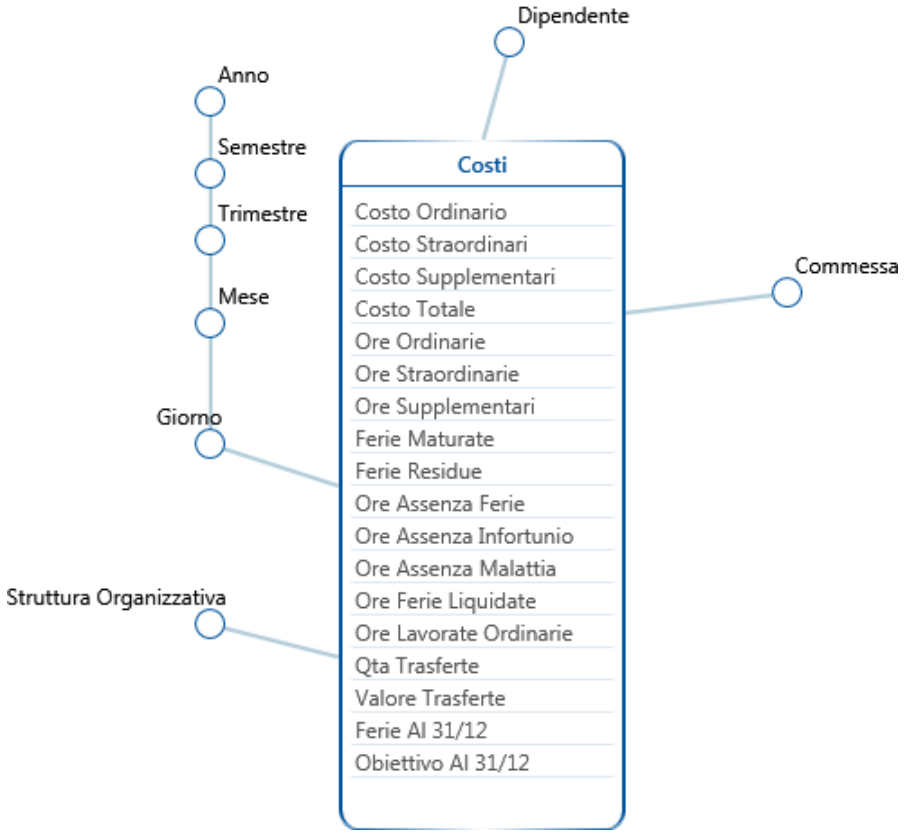


Figura 3.8: DFM per i costi dei dipendenti

zione, realizzata su SSAS, che fornisce informazioni sul tipo di lavoro svolto dal dipendente.

Budget

In Figura 3.10 viene mostrato il DFM che permette di analizzare le ore di budget e quelle di rettifica stimate per le commesse. Come per il timesheet è presente la dimensione *Prestazione e Lavorazione* che fornisce informazioni sul lavoro svolto per la commessa. In questo caso la dimensione temporale è aggregata per mese, in quanto la stima delle ore per le commesse viene effettuata mensilmente. L'indicatore relativo alle ore di rettifica viene valorizzato durante il corso del mese, sulla base delle variazioni applicate alle ore di budget. Grazie a questo fatto è possibile effettuare il confronto tra le ore

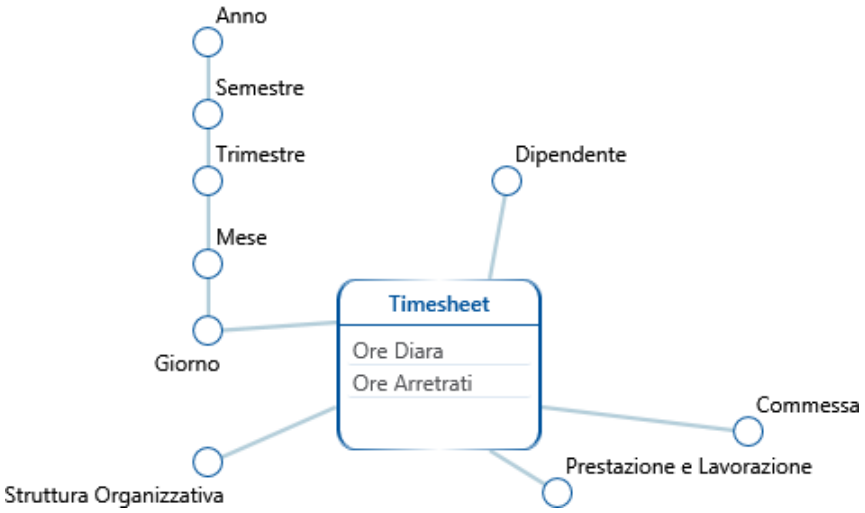


Figura 3.9: DFM per il timesheet dei dipendenti

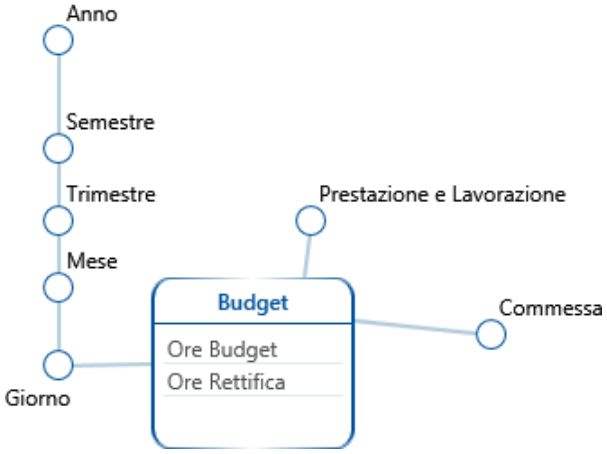


Figura 3.10: DFM per le ore di lavoro stimate per le commesse

stimate all'inizio del mese e quelle effettivamente lavorate al termine dello stesso, così da poter verificare gli scostamenti.

Infortuni

In Figura 3.11 viene mostrato il DFM degli infortuni. In questo caso le dimensioni d'analisi sono il dipendente, la sua struttura organizzativa, il giorno dell'infortunio e l'infortunio stesso. Quest'ultima dimensione fornisce

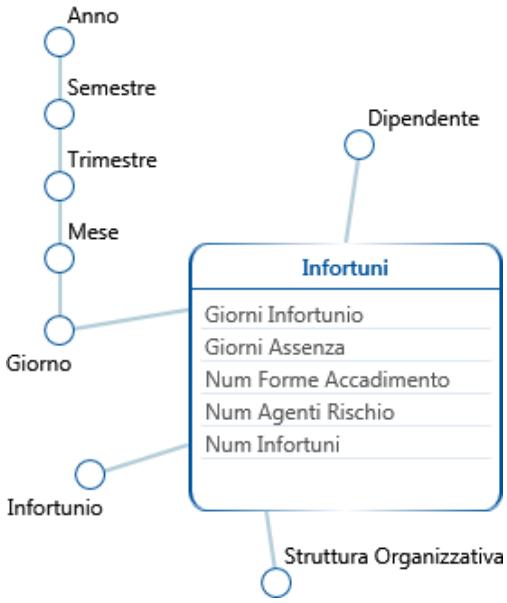


Figura 3.11: DFM per gli infortuni dei dipendenti

informazioni descrittive aggiuntive come la forma di accadimento o l'agente di rischio e viene creata direttamente in SSAS. Tramite il fatto è quindi possibile analizzare l'andamento degli infortuni nel corso del tempo ed evidenziarne le forme di accadimento e gli agenti di rischio più frequenti, in modo tale da poter intraprendere eventuali azioni preventive.

3.3 Modellazione front end

Rappresenta l'ultima fase di un sistema di BI ed è quella utilizzata direttamente dagli utenti finali. L'obiettivo del front end è quello di rendere disponibile una reportistica dettagliata ed accurata che permetta l'analisi dei principali indicatori d'interesse. Analizzando le misure è possibile ottenere un riscontro su tutti gli aspetti relativi al personale ed intraprendere le dovute azioni. Nello specifico l'azienda cliente ha richiesto due diverse modalità per la navigazione dei dati:

- Analisi libera: si interagisce direttamente con il cubo costruito in SSAS;

- Analisi tramite report: vengono costruiti dei report sfruttando le pivot table disponibili in Excel.

In entrambi i casi i report vengono rilasciati sul portale Sharepoint dell'ambiente di produzione, che rappresenta lo spazio ufficiale per i rilasci del front end. Gli utenti accedono al portale e si scaricano i report interessati in locale o in alternativa effettuano la navigazione direttamente da browser.

3.3.1 Analisi libera

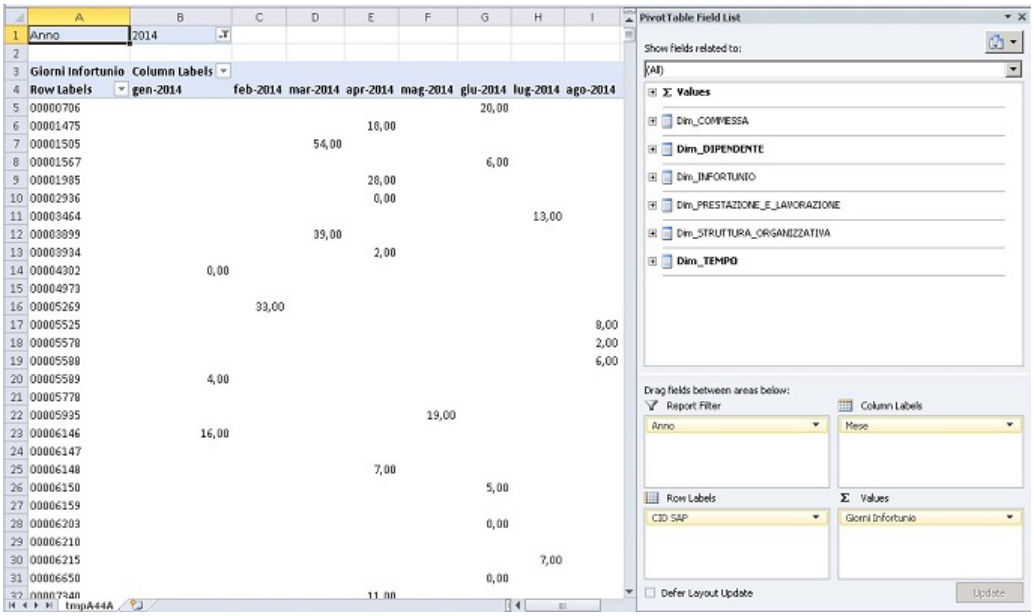


Figura 3.12: Analisi del personale mediante cubo

Questo tipo d'analisi viene solitamente realizzata per esigenze meno istituzionali e garantisce una completa possibilità di personalizzazione del report. Lo strumento utilizzato per la visualizzazione dei dati è Microsoft Excel, che fornisce un'integrazione nativa con i cubi sviluppati in SSAS. L'azienda cliente ha richiesto di poter effettuare l'analisi dei dati tramite pivot table, tuttavia lo strumento mette a disposizione numerosi altri componenti come diagrammi, slicer o tabelle.

In Figura 3.12 viene mostrato un esempio d'analisi mediante cubo, nella quale vengono esaminati i giorni di infortunio dei dipendenti. Sul pannello di destra vengono mostrate le dimensioni a disposizione e l'insieme degli indicatori (*Values*). L'utente può inserire o rimuovere elementi delle dimensioni a piacimento ed effettuare le classiche operazioni che il cubo mette a disposizione (*roll up, drill down,...*). Per l'analisi d'esempio si filtra l'anno 2014, si riportano sulle colonne della pivot tutti i mesi e sulle righe i codici SAP dei dipendenti, infine si effettua l'analisi sul numero giorni d'infortunio inserendo nella pivot l'indicatore correlato.

3.3.2 Analisi tramite report

Rappresenta il secondo tipo d'analisi resa disponibile per il sistema di BI. Questa modalità, a differenza della precedente, ha una formalità maggiore, ovvero questo è il caso in cui i report devono essere distribuiti a tutta l'azienda che farà le opportune valutazioni. Il cliente ha richiesto la creazione di quattro versioni differenti di report:

- Assenze: contiene gli indicatori correlati alle assenze;
- Straordinari e supplementari: contiene tutti gli indicatori relativi al lavoro straordinario e supplementare;
- Ferie: mostra tutti gli indicatori correlati alle ferie;
- Trasferte: mostra i costi relativi alle trasferte dei dipendenti.

Per la loro costruzione si è deciso di creare un'unica vista che integra gli indicatori per tutti e quattro i report, in modo tale da utilizzare una sola sorgente per l'alimentazione degli stessi.

La Figura 3.13 mostra il foglio iniziale del report. Esso mette a disposizione un pulsante che esegue uno script VB ed effettua le seguenti operazioni:

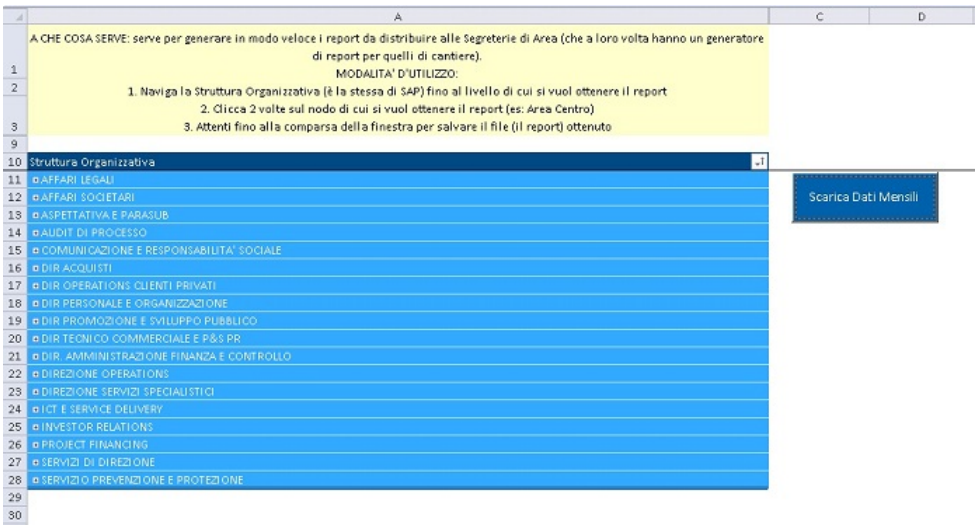


Figura 3.13: Foglio iniziale dei report

- Scaricamento dei dati: vengono scaricati, mediante query, i dati necessari per il report su di un foglio d'appoggio;
- Copia dei dati: i dati contenuti nel foglio d'appoggio vengono copiati sul foglio utilizzato come sorgente della pivot table;
- Aggiornamento delle pivot: si aggiorna la pivot che compone il report;
- Formattazione del foglio: si applica la formattazione richiesta dal cliente al foglio contenente il report.

In questo modo gli utenti possono aggiornare in automatico i report quando necessario. Lo stesso foglio elenca una serie di sottoaree che possono essere utilizzate per la generazione di un report equivalente contenente tutti i dati della sottoarea selezionata. Quest'ultima operazione viene realizzata semplicemente cliccando sulla sottoarea d'interesse.

Nelle pagine seguenti viene mostrata la pivot table per ognuno dei report realizzati per l'azienda cliente. In ogni foglio vengono inoltre messi a disposizione dei filtri, presenti nella parte superiore della pivot, per selezionare solo i dati d'interesse.

	A	B	M	N	O	P	AA	AB	AC	FN	FY	FZ	GA
1	**REPORT ASSENZE**												
2													
3	Inquadramento Dipendente	(All)											
4	Macroclasse	(All)											
5	Classe	(All)											
6	Stato (all'ultimo giorno dell'ultimo mese estratto)	(All)											
7	Data Ultima Cessazione	(All)											
8													
9													
10		2014-01-31				2014-02-28							
11	**Struttura Organizzativa - Dipendente**	**TOTALE ASSENZE (Inf,Mat,Mal)**	**TOTALE ASSENZE (Inf,Mat,Mal) su Ore Teoriche lavorabili**	**TOTALE ASSENZE (Inf, Mat, Mal) + Altre Assenze**	**TOTALE ASSENZE (Inf,Mat,Mal) + ALTRE Assenze su Ore Teoriche lavorabili**	**TOTALE ASSENZE (Inf,Mat,Mal)**	**TOTALE ASSENZE (Inf,Mat,Mal) su Ore Teoriche lavorabili**	**TOTALE ASSENZE (Inf, Mat, Mal) + Altre Assenze**	**TOTALE ASSENZE (Inf,Mat,Mal) + ALTRE Assenze su Ore Teoriche lavorabili**	**Total TOTALE ASSENZE (Inf,Mat,Mal)**	**Total TOTALE ASSENZE (Inf,Mat,Mal) su Ore Teoriche lavorabili**	**Total TOTALE ASSENZE (Inf, Mat, Mal) + Altre Assenze**	**Total TOTALE ASSENZE (Inf,Mat,Mal) + ALTRE Assenze su Ore Teoriche lavorabili**
12	⊕ AFFARI LEGALI	32	2,0%	68	4,3%	32	2,1%	42	2,8%	417	2,2%	666	3,5%
13	⊕ AFFARI SOCIETARI	0	0,0%	12	2,6%	0	0,0%	6	1,4%	60	1,1%	133	2,4%
14	⊕ ASPETTATIVA E PARASUB	0	0,0%	1.171	100,0%	0	0,0%	1.116	100,0%	0	0,0%	13.967	100,0%
15	⊕ AUDIT DI PROCESSO	8	1,6%	9	1,8%	0	0,0%	8	1,7%	24	0,4%	58	1,0%
16	⊕ COMUNICAZIONE E RESPONSABILITA' SOCIALE	173	12,9%	384	28,6%	173	13,5%	160	12,5%	5.215	37,1%	5.631	40,0%
17	⊕ DIR ACQUISTI	288	6,8%	423	9,9%	301	7,4%	376	9,3%	1.582	3,1%	2.235	4,4%
18	⊕ DIR OPERATIONS CLIENTI PRIVATI	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%
19	⊕ DIR PERSONALE E ORGANIZZAZIONE	784	4,3%	10.646	57,8%	213	1,2%	12.601	72,5%	2.732	1,2%	166.718	73,3%
20	⊕ DIR PROMOZIONE E SVILUPPO PUBBLICO	181	5,0%	184	5,1%	219	6,4%	227	6,7%	2.590	6,2%	2.883	6,9%
21	⊕ DIR TECNICO COMMERCIALE E P&S PR	543	6,6%	716	8,7%	658	8,3%	735	9,3%	6.743	6,6%	8.416	8,2%
22	⊕ DIR. AMMINISTRAZIONE FINANZA E CONTROLL	1.033	6,0%	1.349	7,9%	1.210	7,4%	1.447	8,9%	10.390	5,0%	13.075	6,3%
23	⊕ DIREZIONE OPERATIONS	113.156	7,1%	171.353	10,8%	121.425	8,0%	170.783	11,2%	1.304.367	6,9%	2.373.539	12,5%
24	⊕ DIREZIONE SERVIZI SPECIALISTICI	80	1,7%	256	5,5%	0	0,0%	160	3,6%	318	0,8%	1.394	3,3%
25	⊕ ICT E SERVICE DELIVERY	197	5,6%	268	7,6%	173	5,1%	248	7,4%	1.948	4,4%	2.509	5,6%
26	⊕ INVESTOR RELATIONS	32	9,5%	34	10,1%	0	0,0%	0	0,0%	32	0,8%	-2	0,0%
27	⊕ PROJECT FINANCING		0,0%		0,0%		0,0%		0,0%	0	0,0%	0	0,0%
28	⊕ SERVIZI DI DIREZIONE	37	2,1%	144	8,1%	0	0,0%	13	0,8%	296	1,4%	715	3,4%
29	⊕ SERVIZIO PREVENZIONE E PROTEZIONE	304	14,2%	375	17,5%	317	15,5%	362	17,7%	2.952	11,7%	3.992	15,8%
30	**Grand Total**	**116.848**	**7,1%**	**187.392**	**11,3%**	**124.721**	**7,8%**	**188.282**	**11,8%**	**1.339.664**	**6,8%**	**2.595.929**	**13,1%**
31													

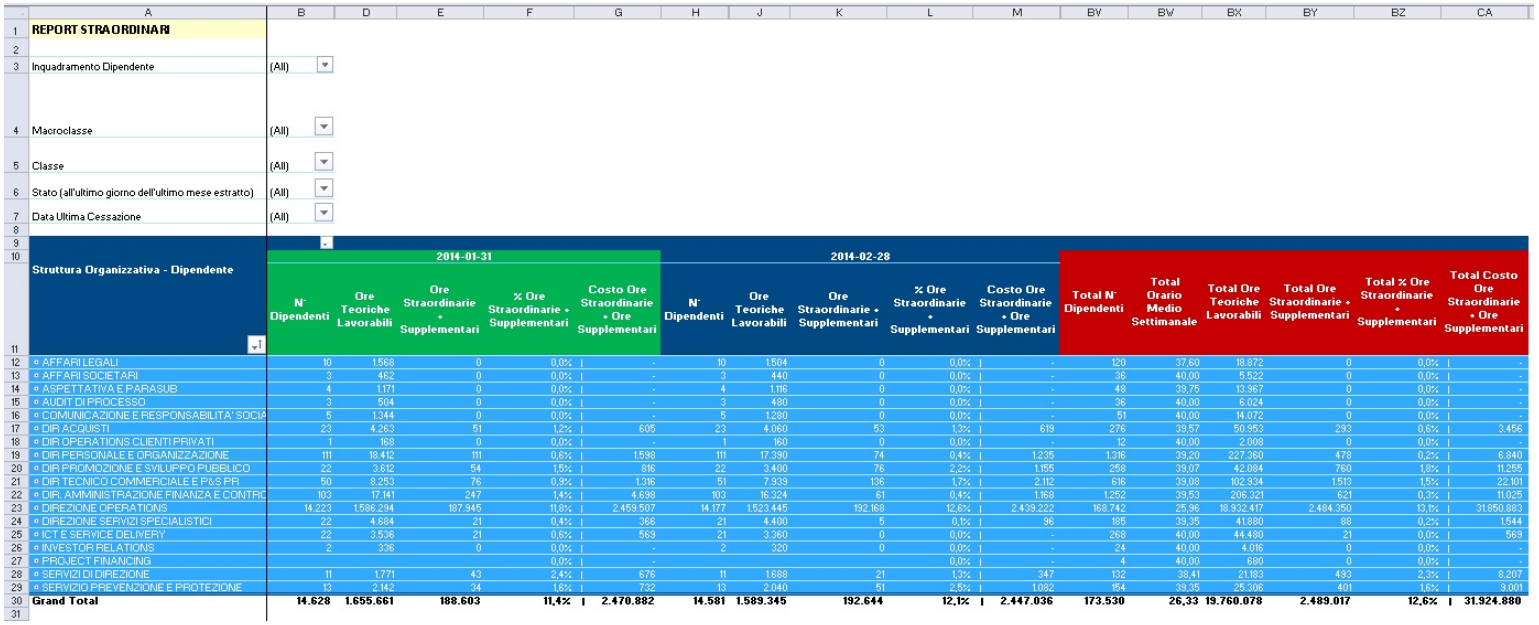
Figura 3.14: Report Assenze

REPORT FERIE

Inquadramento Dipendente	(All)
Macroclasse	(All)
Classe	(All)
Stato (all'ultimo giorno dell'ultimo mese estratto)	(All)
Data Ultima Cessazione	(All)

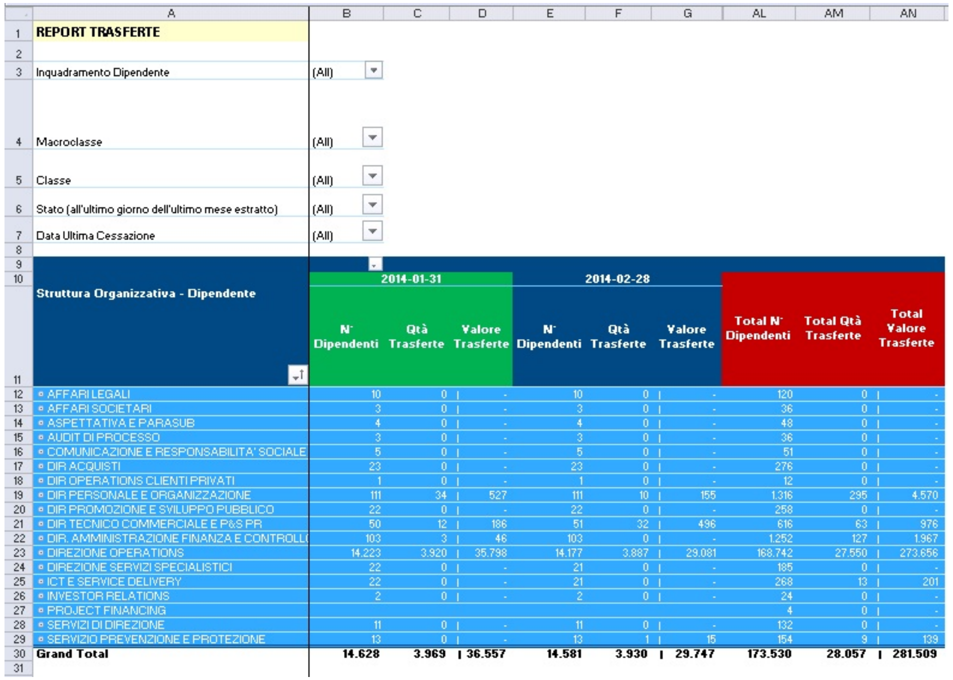
Struttura Organizzativa - Dipendente	2014-01-31 N° Dipendenti	2014-01-31 Ferie Maturate	2014-01-31 Ferie Liquidate	2014-01-31 Ferie Godute	2014-01-31 % Ferie Godute su teorico lavorabile	2014-01-31 Smaltimento Ferie (Goduto AC / Maturato AC)	2014-02-28 N° Dipendenti	2014-02-28 Ferie Maturate	2014-02-28 Ferie Liquidate	2014-02-28 Ferie Godute	2014-02-28 % Ferie Godute su teorico lavorabile	2014-02-28 Smaltimento Ferie (Goduto AC / Maturato AC)	Total N° Dipendenti	Total Ferie Maturate	Total Ferie Liquidate	Total Ferie Godute	Total % Ferie Godute su teorico lavorabile	Total Smaltimento Ferie (Goduto AC / Maturato AC)
AFFARI LEGALI	10	195	0	159	10,1%	81,6%	10	195	0	71	4,7%	36,4%	120	2.313	-89	2.397	12,7%	103,6%
AFFARI SOCIETARI	3	60	0	49	10,7%	81,6%	3	60	0	30	6,8%	49,7%	36	735	0	839	15,2%	114,2%
ASPETTATIVA E PARASUB	4	0	0	0	0,0%	0,0%	4	0	0	0	0,0%	0,0%	48	0	0	0	0,0%	0,0%
AUDIT DI PROCESSO	3	62	0	13	2,5%	20,7%	3	62	0	3	0,6%	4,8%	36	744	0	650	10,8%	87,3%
COMUNICAZIONE E RESPONSABILITA' SOCI	5	165	0	-88	-6,5%	-53,2%	5	165	0	4	0,3%	2,2%	51	1.674	-127	837	5,9%	50,0%
DIR ACQUISTI	23	506	35	321	7,5%	63,4%	23	506	82	167	4,1%	33,0%	276	6.279	117	5.783	11,3%	92,1%
DIR OPERATIONS CLIENTI PRIVATI	1	19	0	28	16,7%	148,4%	1	19	0	8	5,0%	42,4%	12	226	0	280	13,9%	123,7%
DIR PERSONALE E ORGANIZZAZIONE	111	723	602	3.027	16,4%	418,9%	111	627	0	604	3,5%	96,2%	1.316	7.489	1.164	9.404	4,1%	125,6%
DIR PROMOZIONE E SVILUPPO PUBBLICO	22	444	0	289	8,0%	65,2%	22	444	176	107	3,1%	24,1%	258	5.136	379	4.221	10,0%	82,2%
DIR TECNICO COMMERCIALE E P&S PR	50	1.016	0	826	10,0%	81,3%	51	996	0	230	2,9%	23,1%	616	12.477	485	11.457	11,1%	91,8%
DIR. AMMINISTRAZIONE FINANZA E CONTR	103	2.099	0	1.141	6,7%	54,4%	103	2.120	0	446	2,7%	21,1%	1.252	25.687	439	24.254	11,8%	94,4%
DIREZIONE OPERATIONS	14.223	193.429	16.047	124.318	7,8%	64,3%	14.177	193.006	46.871	90.425	5,9%	46,9%	168.742	2.262.458	191.003	1.895.359	10,0%	83,8%
DIREZIONE SERVIZI SPECIALISTICI	22	539	0	343	7,3%	63,7%	21	539	0	104	2,4%	19,3%	185	4.816	411	3.622	8,6%	75,2%
ICT E SERVICE DELIVERY	22	434	-29	216	6,1%	49,7%	21	434	0	81	2,4%	18,7%	268	5.397	6	4.789	10,8%	88,7%
INVESTOR RELATIONS	2	40	0	32	9,5%	80,9%	2	40	0	16	5,0%	40,5%	24	474	0	412	10,3%	86,8%
PROJECT FINANCING					0,0%	0,0%					0,0%	0,0%	4	75	0	0	0,0%	0,0%
SERVIZI DI DIREZIONE	11	230	0	138	7,8%	59,8%	11	230	0	107	6,3%	46,6%	132	2.761	0	2.942	13,9%	106,5%
SERVIZIO PREVENZIONE E PROTEZIONE	13	280	0	270	12,6%	96,5%	13	259	0	16	0,8%	6,3%	154	3.158	90	2.692	10,6%	85,2%
Grand Total	**14.628**	**200.239**	**16.656**	**131.081**	**7,9%**	**65,5%**	**14.581**	**199.701**	**47.128**	**92.418**	**5,8%**	**46,3%**	**173.530**	**2.341.902**	**193.877**	**1.969.937**	**10,0%**	**84,1%**

Figura 3.15: Report Ferie



	A	B	D	E	F	G	H	J	K	L	M	BV	BW	BX	BY	BZ	CA
1	**REPORT STRAORDINARI**																
2																	
3	Inquadramento Dipendente	(All)															
4	Macroclasse	(All)															
5	Classe	(All)															
6	Stato (all'ultimo giorno dell'ultimo mese estratto)	(All)															
7	Data Ultima Cessazione	(All)															
8																	
9			2014-01-31				2014-02-28										
10	**Struttura Organizzativa - Dipendente**	**N° Dipendenti**	**Ore Teoriche Lavorabili**	**Ore Straordinarie + Supplementari**	**% Ore Straordinarie + Supplementari**	**Costo Ore Straordinarie + Ore Supplementari**	**N° Dipendenti**	**Ore Teoriche Lavorabili**	**Ore Straordinarie + Supplementari**	**% Ore Straordinarie + Supplementari**	**Costo Ore Straordinarie + Ore Supplementari**	**Total N° Dipendenti**	**Total Orario Medio Settimanale**	**Total Ore Teoriche Lavorabili**	**Total Ore Straordinarie + Supplementari**	**Total % Ore Straordinarie + Supplementari**	**Total Costo Ore Straordinarie + Ore Supplementari**
11																	
12	⊕ AFFARI LEGALI	10	1.568	0	0,0%	-	10	1.504	0	0,0%	-	120	37,60	18.872	0	0,0%	-
13	⊕ AFFARI SOCIETARI	3	462	0	0,0%	-	3	440	0	0,0%	-	36	40,00	5.522	0	0,0%	-
14	⊕ ASPETTATIVA E PARASUB	4	1.171	0	0,0%	-	4	1.116	0	0,0%	-	48	39,75	13.967	0	0,0%	-
15	⊕ AUDIT DI PROCESSO	3	504	0	0,0%	-	3	480	0	0,0%	-	36	40,00	6.024	0	0,0%	-
16	⊕ COMUNICAZIONE E RESPONSABILITA' SOCIA	5	1.344	0	0,0%	-	5	1.280	0	0,0%	-	51	40,00	14.072	0	0,0%	-
17	⊕ DIR ACQUISTI	23	4.263	51	1,2%	605	23	4.060	53	1,3%	619	276	39,57	50.953	293	0,6%	3.456
18	⊕ DIR OPERATIONS CLIENTI PRIVATI	1	168	0	0,0%	-	1	160	0	0,0%	-	12	40,00	2.008	0	0,0%	-
19	⊕ DIR PERSONALE E ORGANIZZAZIONE	111	18.412	111	0,6%	1.598	111	17.390	74	0,4%	1.235	1.316	39,20	227.360	478	0,2%	6.840
20	⊕ DIR PROMOZIONE E SVILUPPO PUBBLICO	22	3.612	54	1,5%	816	22	3.400	76	2,2%	1.155	258	39,07	42.084	760	1,8%	11.255
21	⊕ DIR TECNICO COMMERCIALE E P&S PR	50	8.253	76	0,9%	1.316	51	7.939	136	1,7%	2.112	616	39,08	102.934	1.513	1,5%	22.101
22	⊕ DIR. AMMINISTRAZIONE FINANZA E CONTRO	103	17.141	247	1,4%	4.698	103	16.324	61	0,4%	1.168	1.252	39,53	206.321	621	0,3%	11.025
23	⊕ DIREZIONE OPERATIONS	14.223	1.586.294	187.945	11,8%	2.459.507	14.177	1.523.445	192.168	12,6%	2.439.222	168.742	25,96	18.932.417	2.484.350	13,1%	31.850.883
24	⊕ DIREZIONE SERVIZI SPECIALISTICI	22	4.684	21	0,4%	366	21	4.400	5	0,1%	96	185	39,35	41.880	88	0,2%	1.544
25	⊕ ICT E SERVICE DELIVERY	22	3.536	21	0,6%	569	21	3.360	0	0,0%	-	268	40,00	44.480	21	0,0%	569
26	⊕ INVESTOR RELATIONS	2	336	0	0,0%	-	2	320	0	0,0%	-	24	40,00	4.016	0	0,0%	-
27	⊕ PROJECT FINANCING				0,0%	-				0,0%	-	4	40,00	680	0	0,0%	
28	⊕ SERVIZI DI DIREZIONE	11	1.771	43	2,4%	676	11	1.688	21	1,3%	347	132	38,41	21.183	493	2,3%	8.207
29	⊕ SERVIZIO PREVENZIONE E PROTEZIONE	13	2.142	24	1,6%	732	13	2.040	51	2,5%	1.082	154	39,26	26.066	401	1,6%	9.001
30	**Grand Total**	**14.628**	**1.655.661**	**188.603**	**11,4%**	**2.470.882**	**14.581**	**1.589.345**	**192.644**	**12,1%**	**2.447.036**	**173.530**	**26,33**	**19.760.078**	**2.489.017**	**12,6%**	**31.924.880**
31																	

Figura 3.16: Report Straordinari e Supplementari



REPORT TRASFERTE

Inquadramento Dipendente	(All)
Macroclasse	(All)
Classe	(All)
Stato (all'ultimo giorno dell'ultimo mese estratto)	(All)
Data Ultima Cessazione	(All)

Struttura Organizzativa - Dipendente	2014-01-31			2014-02-28					
	N° Dipendenti	Qtà Trasferte	Valore Trasferte	N° Dipendenti	Qtà Trasferte	Valore Trasferte	Total N° Dipendenti	Total Qtà Trasferte	Total Valore Trasferte
AFFARI LEGALI	10	0	-	10	0	-	120	0	-
AFFARI SOCIETARI	3	0	-	3	0	-	36	0	-
ASPETTATIVA E PARASUB	4	0	-	4	0	-	48	0	-
AUDIT DI PROCESSO	3	0	-	3	0	-	36	0	-
COMUNICAZIONE E RESPONSABILITA' SOCIALE	5	0	-	5	0	-	51	0	-
DIR ACQUISTI	23	0	-	23	0	-	276	0	-
DIR OPERATIONS CLIENTI PRIVATI	1	0	-	1	0	-	12	0	-
DIR PERSONALE E ORGANIZZAZIONE	111	34	527	111	10	155	1.316	295	4.570
DIR PROMOZIONE E SVILUPPO PUBBLICO	22	0	-	22	0	-	258	0	-
DIR TECNICO COMMERCIALE E P&S PR	50	12	186	51	32	496	616	63	976
DIR. AMMINISTRAZIONE FINANZA E CONTROLL	103	3	46	103	0	-	1.252	127	1.967
DIREZIONE OPERATIONS	14.223	3.920	35.798	14.177	3.887	29.081	168.742	27.550	273.656
DIREZIONE SERVIZI SPECIALISTICI	22	0	-	21	0	-	185	0	-
ICT E SERVICE DELIVERY	22	0	-	21	0	-	268	13	201
INVESTOR RELATIONS	2	0	-	2	0	-	24	0	-
PROJECT FINANCING							4	0	-
SERVIZI DI DIREZIONE	11	0	-	11	0	-	132	0	-
SERVIZIO PREVENZIONE E PROTEZIONE	13	0	-	13	1	15	154	9	139
Grand Total	**14.628**	**3.969**	**36.557**	**14.581**	**3.930**	**29.747**	**173.530**	**28.057**	**281.509**

Figura 3.17: Report Trasferte

Conclusioni

Questo progetto di tesi ha come obiettivo la progettazione e realizzazione di un sistema di BI per l'analisi del personale in un'azienda di servizi. L'implementazione ha seguito le classiche fasi di un progetto di BI: import, staging, datamart e cubo. Inizialmente vengono importati i dati delle sorgenti transazionali, si realizza la fase di ETL ed infine si crea il datamart modellando in formato relazionale i fatti e le dimensioni d'interesse. A partire da quest'ultimo si crea il modello multidimensionale utilizzato per la costruzione del cubo. Gli utenti possono analizzare i dati semplicemente interrogando il cubo o sfruttando la reportistica creata appositamente per il progetto.

Per la fase di testing sono stati messi a disposizione dell'azienda cliente manuali che spiegassero le modalità d'utilizzo del cubo e della relativa reportistica. Come sempre accade in progetti di questo tipo, durante la fase di test sono emerse squadrature sui dati che hanno portato a delle modifiche al modello ed alle logiche di calcolo degli indicatori. Dove possibile le richieste sono state inserite nel modello attuale, mentre quelle non immediate sono state rimandate agli sviluppi successivi.

I benefici introdotti dal progetto sono molteplici. In precedenza l'ufficio del personale distribuiva report che venivano generati a partire dai sistemi pregressi gestiti da ogni singola società. L'ufficio doveva quindi svolgere complesse ed onerose operazioni d'integrazione di basi dati personali. Con l'introduzione del sistema di BI molti di questi colli di bottiglia vengono superati: si automatizza l'intera fase di integrazione dei dati, così come la

produzione dei report da distribuire all'interno dell'azienda. In secondo luogo è presente anche la possibilità di svolgere analisi mirate, sfruttando le potenzialità messe a disposizione dai sistemi di BI.

Per quanto riguarda gli sviluppi futuri, l'azienda cliente ha richiesto la realizzazione di cruscotti e l'introduzione della sicurezza sui dati. Il primo sviluppo rappresenta una nuova modalità di presentazione dei dati. Al momento, infatti, i report hanno un formato tabellare, mentre a tendere la necessità sarà la consultazione di cruscotti direzionali. Questi ultimi mostrano i principali indicatori d'interesse in maniera ergonomica, al fine di rappresentare con immediatezza l'andamento dei fenomeni rispetto agli obiettivi prefissati o agli standard aziendali. Il secondo aspetto è quello correlato alla sicurezza. Attualmente sia per il cubo, che per la reportistica, non è presente profilazione. Tutti gli utenti che hanno l'accesso al portale Sharepoint possono recuperarsi una copia del report e visualizzare tutti i dati. L'azienda richiede che ogni utente possa visualizzare solamente i dati pertinenti alla sua area, nascondendo tutti quelli per i quali non è autorizzato. Entrambi gli sviluppi richiesti sono in corso d'analisi e verranno progettati ed implementati nei prossimi mesi.

Bibliografia

[1] `http://msdn.microsoft.com`.

[2] `http://www.indyco.com/`.

[3] C. Adamson. *Star Schema*. McGraw-Hill, 2010.

[4] R. M. Devens. *Cyclopaedia of Commercial and Business Anecdotes*. D. Appleton and Company, 1868.

[5] Heinze J. History of business intelligence. `http://www.bisoftwareinsight.com/history-of-business-intelligence/`.

[6] R. Kimball and M. Ross. *The Data Warehouse Toolkit*. Wiley, 2013.

[7] R. Kimball, W. Thornthwaite, and J. Mundy. *The Microsoft Data Warehouse Toolkit*. Wiley, 2011.

[8] H. P. Luhn. A business intelligence system. *IBM Journal*, 1958.

[9] D. Petkovic. *Microsoft SQL Server 2012 A Beginners Guide*. McGraw-Hill, 2012.

[10] R. Reza. *Microsoft SQL Server 2014 Business Intelligence Development*. Packt Publishing, 2014.

[11] S. Rizzi and M. Golfarelli. *Data Warehouse Design: Modern Principles and Methodologies*. McGraw-Hill, 2009.

[12] A. Vaisman and E. Zimányi. *Data Warehouse Systems: Design and Implementation*. Springer-Verlag, 2014.