

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

CAMPUS DI CESENA

SCUOLA DI INGEGNERIA E ARCHITETTURA

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA E SCIENZE INFORMATICHE

PROGETTAZIONE DI UNO STRUMENTO DI ANALISI DATI PER SISTEMI SCADA - ENERGY MANAGEMENT SYSTEM

Tesi in

Sistemi Informativi

Relatore

Chiar.mo Prof. Matteo Golfarelli

Presentata da

Cristiano Bragagni

Correlatore

Ing. Claudio Gambetti

Sessione II

Anno Accademico 2013/2014

Sommario

Introduzione	1
Capitolo 1	5
Business intelligence	5
1.1 Evoluzione dei sistemi informativi	5
1.2 Definizione di business intelligence	7
1.3 Piramide della business intelligence	8
1.4 Decision support systems	10
1.5 OLTP e OLAP	14
1.6 ETL	16
Capitolo 2	19
Energy Management System	19
2.1 Sistemi SCADA	19
2.2 SCADA/EMS	24
2.3 Sistemi presenti sul mercato: caratteristiche e funzionalità	27
2.4 SIRIUS	28
2.5 ABB	34
2.6 On.Energy management system	40
Capitolo 3	49

L'applicativo On.Energy Management System	49
3.1 Aspetti generali	49
3.2 Data Analysis	53
3.3 Limiti della data analysis	56
3.4 Analisi dei requisiti	58
3.5 Analisi del problema	60
Capitolo 4	63
Definizione dello schema dei metadati	63
4.1 Normalizzazione del modello relazionale	63
4.2 Metamodello del cubo	66
4.3 Modellazione DFM	72
4.4 Schema di fatto	77
4.5 Progetto di alimentazione	79
Capitolo 5	85
Implementazione del sistema	85
5.1 Panoramica generale sull'applicazione client	85
5.2 Interfaccia di configurazione dei metadati	86
5.3 Recupero dati dal data warehouse	91
5.4 Interfaccia di analisi dei dati	94

5.5	Analisi delle Prestazioni	98
5.6	Caso di studio Amadori	100
	Conclusioni e sviluppi futuri	109
	Bibliografia	111

Introduzione

I sistemi SCADA hanno assunto una notevole importanza nel campo dei sistemi informativi grazie alla loro capacità di poter operare su un numero crescente di realtà differenti. Inizialmente, i sistemi SCADA sono stati impiegati per controllare e supervisionare lo stato dei sistemi utilizzati in ambito industriale, ma nel corso del tempo hanno trovato impiego in altri settori, come nella generazione e distribuzione di energia, negli impianti siderurgici e chimici, fino ad essere utilizzati in impianti nucleari.

I sistemi SCADA, impiegati per rispondere alle esigenze di diverse applicazioni, hanno dovuto adattare le proprie funzionalità: ogni settore ha fatto emergere specifiche necessità, così da decretare una suddivisione dei sistemi SCADA in tanti sottosistemi, ognuno rivolto ad un campo d'applicazione diverso. È necessario specificare che le diverse funzionalità dei sistemi SCADA hanno risentito diversamente di questa suddivisione: se per le operazioni di controllo e supervisione dei sistemi non ci sono state modifiche rilevanti, la funzionalità che ha subito profonde trasformazioni è l'analisi dei dati. La motivazione di questo forte impatto è da attribuire ai differenti contesti di utilizzo di questi sistemi: ogni settore ha necessità di analizzare i dati con strumenti capaci di visualizzare informazioni di natura differente.

L'analisi dei dati è una funzionalità indispensabile per qualsiasi sistema informativo, perché capace di estrapolare informazioni cruciali per la riorganizzazione dei processi aziendali. Nell'ambito dei sistemi SCADA, le funzionalità di analisi dei dati sono progettate su esigenze espresse dal dominio del sistema informativo, il quale necessita di opportune tecniche per ricavare informazioni rilevanti per il contesto su cui opera. Ogni implementazione di sistema SCADA riesce ad emergere sul mercato in base alle funzionalità offerte.

A definire il sottosistema SCADA d'interesse di questa tesi è la collaborazione con l'azienda *Onit Group s.r.l.*, con la quale si decide di intraprendere un percorso di sviluppo che ha come scopo finale l'ampliamento delle tecniche di analisi dei dati del prodotto *On.Energy Management System*. Tale

soluzione offre un'implementazione di un sistema SCADA/EMS, cioè un'infrastruttura SCADA con funzionalità orientate a monitorare e migliorare l'efficiamento dei consumi energetici.

L'obiettivo di questa tesi è approfondire le competenze sulle funzionalità sviluppate nei sistemi SCADA/EMS presenti sul mercato, così da conoscerne le potenzialità offerte: tutte le conoscenze acquisite servono a progettare uno strumento di analisi dati flessibile e interattivo, con il quale è possibile svolgere analisi non proponibili con le altre soluzioni analizzate. La progettazione dello strumento di analisi dei dati è orientata a definire un modello multidimensionale per la rappresentazione delle informazioni: il percorso di progettazione richiede di individuare le informazioni d'interesse per l'utente, così da poterle reintrodurre in fase di progettazione della nuova base dati. L'infrastruttura finale di questa nuova funzionalità si concretizza in un data warehouse: tutte le informazioni di analisi sono memorizzate su una base dati diversa da quella di On.Energy, evitando di correlare le prestazioni dei due diversi sottosistemi. L'utilizzo di un data warehouse pone le basi per realizzare analisi su lunghi periodi temporali: tutte le tipologie di interrogazione dati comprendono un'enorme quantità d'informazioni, esattamente in linea con le caratteristiche delle interrogazioni OLAP.

Il risultato di questa tesi è uno strumento di analisi dati dinamico, che permette all'utente di definire la struttura del fatto oggetto dell'analisi: è prevista una fase di configurazione, in cui si definiscono le istanze dei dati di On.Energy che rientrano nel progetto di alimentazione del data warehouse. Una volta definita la base dati su cui operare, il sistema progetta uno schema multidimensionale alimentato sulla base dei parametri definiti dall'utente. Una volta materializzata la base dati, l'utente può interrogare il sistema per visualizzare dati d'interesse, sapendo di poter operare su un ambiente multidimensionale. Al termine della tesi è proposto un confronto tra le attuali possibilità d'analisi della soluzione On.Energy e il sistema progettato. Infine, la funzionalità è introdotta nel pacchetto software della soluzione On.Energy, migliorando le possibilità di analisi del prodotto. A valorizzare le potenzialità dello strumento, è riportato un caso di studio reale, effettuato sull'installazione On.Energy dell'azienda Amadori: è condotta una sessione di analisi dati, dove l'impiego dello strumento progettato evidenzia nuove possibilità nell'ambito dell'interrogazione dati, facendo emergere scenari precedentemente non realizzabili.

La struttura della tesi si divide in 5 capitoli. Nel primo capitolo è riportato un approfondimento sulle tecniche di business intelligence: sono elencate le motivazioni che hanno decretato la diffusione di questi sistemi, soprattutto per quanto riguarda l'analisi dei dati. Nel secondo capitolo sono trattate le architetture dei sistemi SCADA, con particolare attenzione all'applicazione di queste tecnologie nell'ambito del controllo dei consumi energetici, dove prendono il nome di SCADA/EMS. Nel corso dello stesso capitolo sono analizzate le funzionalità di tre soluzioni software che realizzano fisicamente queste tecnologie: lo scopo è di conoscere il know-how di questi sistemi, indispensabile per poter progettare uno strumento che porti un reale valore aggiunto all'analisi dei dati. Nel terzo capitolo è approfondito nel dettaglio il dominio del prodotto On.Energy Management System di Onit Group. Il percorso intrapreso per analizzare questa soluzione passa per la conoscenza dei concetti fondamentali del prodotto, per poi effettuare una valutazione dell'attuale possibilità d'analisi dei dati. Tale studio ha lo scopo di evidenziare i limiti strutturali degli strumenti d'analisi: da questa indagine è possibile trarre nuovi requisiti utili per la progettazione del nuovo strumento. Il capitolo conclude con un'analisi del problema. Il quarto capitolo definisce la progettazione dello strumento realizzato: dopo aver eseguito una ricognizione dello schema attuale di On.Energy, si definisce una base dati utile a configurare la struttura del cubo finale. Quindi, si procede con la realizzazione dello schema di fatto e la definizione del processo di alimentazione del data warehouse. Nel quinto capitolo sono riportate le interfacce fornite all'utente per realizzare un'analisi dei dati: attraverso l'interfaccia di configurazione si definiscono le istanze che alimentano il data warehouse, mentre l'interfaccia di interrogazione del cubo permette di eseguire sessioni di analisi sulla base dati. Al termine del capitolo è riportato il caso di studio Amadori, con il quale si delinea il percorso che passa dall'esigenza espressa dal cliente, fino alla realizzazione dello strumento capace di rispondere alle necessità di analisi dei dati.

Capitolo 1

Business intelligence

L'obiettivo di questo capitolo è di capire e conoscere le motivazioni che hanno permesso alla *business intelligence* di acquistare, nel corso degli anni, un crescente interesse nel campo dell'analisi dei dati. Allo scopo di analizzare questo argomento, verrà brevemente riassunto come i sistemi informativi si sono evoluti nel corso del tempo, con la prospettiva di migliorare la struttura dei processi aziendali. Successivamente, sarà approfondita l'architettura della business intelligence, dove verrà spiegata l'importanza della fase decisionale, al fine di introdurre i *decision support system*. A questo punto si evidenzieranno le differenze tra le due tipologie di interrogazione dati OLAP e OLTP. L'ultima parte del capitolo permetterà di esaminare una delle fasi maggiormente interessanti nell'attività di costruzione di un sistema OLAP, denominata ETL.

1.1 Evoluzione dei sistemi informativi

Prima di parlare del concetto di business intelligence, è opportuno fare un passo indietro e aprire una panoramica sulla nascita dei sistemi informativi e sulla loro successiva evoluzione nel corso del tempo. I sistemi informativi nascono nel momento in cui si ha la necessità di introdurre, all'interno dell'ambito aziendale, un nuovo elemento che possa creare un forte valore aggiunto nella gestione del processo produttivo: si vuole dare una forte spinta innovativa inserendo il concetto di informatica nell'ambito aziendale.

L'introduzione dell'informatica porta un primo contributo importante nel processo produttivo aziendale, infatti essa riesce a migliorare in maniera significativa la storicizzazione delle informazioni utili nelle fasi di produzione. L'informatica permette di organizzare le informazioni in maniera strutturata e di velocizzare il coordinamento delle operazioni tra i processi: queste nuove

opportunità rivoluzionano il modo di programmare i processi produttivi, anche se l'informatica viene vista solamente come uno strumento di supporto. Successivamente nasce l'esigenza di rivedere il ruolo dell'informatica all'interno dei sistemi informativi: l'obiettivo è di trasformare il ruolo dell'informatica, portandolo da passivo, cioè un semplice strumento che aiuta l'utente nella programmazione, ad attivo, cioè un mezzo per agire direttamente sull'intero processo informativo, rendendolo maggiormente efficiente.

Il nuovo ruolo dell'informatica diventa di tipo organizzativo, con i compiti di riscrivere i processi, di individuare elementi critici dell'organizzazione e di cercare nuove potenziali aree di business. Oltre alle motivazioni sopra citate, anche altri aspetti hanno richiesto l'intervento dell'informatica nella fase decisionale; basti pensare alla crescente mole di dati immagazzinati dai sistemi informatici, che ha reso impossibile una loro analisi senza l'adeguato supporto del calcolatore. La fase di analisi dei dati ha richiesto l'intervento di strumenti informatici nel momento in cui i processi aziendali hanno raggiunto la loro massima produttività grazie al supporto dell'informatica: l'enorme quantità dei dati da analizzare è diventata un'operazione troppo onerosa se non approcciata con calcolatori dedicati. Dunque, si rende necessario sviluppare una nuova serie di tecniche dedicate all'estrapolazione di dati d'interesse, in un formato appropriato a supportare l'attività decisionale da parte degli utenti.



Supporto tecnologico alla gestione del processo informativo



Strumento di analisi per la riorganizzazione dei processi aziendali

Figura 1 - Il duplice ruolo dell'informatica nei sistemi informativi

Come spiegato nella *figura 1*, l'informatica ha assunto nel corso degli anni un duplice utilizzo: da una parte deve essere uno strumento di supporto alla gestione del processo produttivo, dall'altra deve

assumere l'importante ruolo di pianificare la riorganizzazione dei sistemi informativi, valutando le scelte più opportune da prendere. È a questo punto che interviene la business intelligence, che si propone come disciplina di riferimento per la realizzazione dei nuovi strumenti di analisi.

1.2 Definizione di business intelligence

Il sentiero che porta alla definizione di *business intelligence* (per abbreviazione BI) deve essere affrontato per livelli, visto che tale termine mette in relazione diversi concetti che devono essere prima opportunamente compresi. E' possibile comunque dare una prima definizione molto naïf del termine: la Business intelligence è un insieme di tecniche e procedure che permettono di trasformare dati memorizzati dai sistemi informativi, in informazioni capaci di evidenziare dettagli non precedentemente disponibili. Ci sono due aspetti che è necessario specificare nella definizione appena data: il primo riguarda il contesto di utilizzo delle tecniche di BI, mentre il secondo si rivolge alla tipologia di informazioni coinvolte in questa trasformazione.

Il concetto di business intelligence trova la sua naturale collocazione nell'ambito dei sistemi informativi aziendali; lo scopo di questi sistemi è di raggiungere un determinato obiettivo, realizzabile mediante la pianificazione di processi che opportunamente coordinati, mirano alla trasformazione di un determinato insieme di risorse in un prodotto di business. I processi che fanno parte del sistema informativo aziendale sono classificabili in:

- **Processi operativi**, che hanno lo scopo di concorrere direttamente nel raggiungimento dell'obiettivo primario dell'azienda;
- **Processi gestionali**, i quali si occupano di trasformare gli obiettivi a lungo termine in una programmazione a breve termine, supportati da specifici indicatori dell'andamento produttivo generale.
- **Processi direzionali**, che concorrono nella programmazione a medio-lungo termine dell'organizzazione aziendale.

Se in un primo momento il processo di informatizzazione si è posto l'obiettivo di supportare i processi operativi e gestionali nella loro esecuzione, successivamente è nata l'esigenza di supportare, sulla base delle informazioni prodotte, le scelte dei processi decisionali.

Una volta fatta chiarezza sulle tipologie dei processi dei sistemi informativi, è possibile dare una seconda definizione di business intelligence, scritta con concetti più tecnici rispetto alla prima definizione: la business intelligence consiste in un insieme di strumenti e procedure capaci di trasformare i dati prodotti dai processi operativi in nuove informazioni, a loro volta utilizzate dai processi decisionali nella delicata fase di riorganizzazione dei sistemi informativi.

Il nuovo ruolo della business intelligence deve porsi come strumento rivolto a figure aziendali che non presentano necessariamente competenze informatiche specifiche: l'ulteriore sfida è quella di trasformare dati aziendali in un'informazione fruibile a precise figure professionali, che necessitano di recepire un'informazione completa, idonea e inerente alla realtà aziendale in cui operano.

Una volta emerse queste nuove esigenze, è diventato prioritario permettere ai sistemi informatici di trasformare i propri dati di business in informazioni utili al processo decisionale. La nuova esigenza richiesta ai sistemi informatici era chiara: garantire uno strumento che permetta all'utente di operare decisioni consapevoli e complete; l'obiettivo finale è di consentire all'utente di valutare le decisioni nella maniera più completa possibile, con una migliore consapevolezza rispetto al passato.

1.3 Piramide della business intelligence

La business intelligence è solitamente rappresentata da una piramide divisa in sezioni orizzontali, dove ogni sezione rappresenta una funzionalità specifica del sistema. L'utilizzo di questa figura serve ad esprimere meglio il concetto di dipendenza tra le sezioni; infatti, per poter svolgere il proprio compito, ogni funzionalità utilizza le informazioni messe a disposizione dal livello sottostante, per poi renderle disponibili al livello superiore.

Come è possibile vedere dalla *figura 2*, la piramide pone alla base della sua struttura le applicazioni gestionali: questa rappresentazione corrisponde a quanto riportato nella definizione della business intelligence, cioè l'utilizzo dei dati dei processi operazionali come punto di partenza per l'analisi. Una particolarità che ha reso estremamente interessante la BI è la possibilità di disporre di dati selezionati da sorgenti diverse: questo fattore ha permesso di introdurre in un unico strumento di analisi, dati che provengono da sistemi diversi, quindi non direttamente comparabili.



Figura 2- La piramide della business intelligence

Per quanto riguarda la punta della piramide, vi è una correlazione con quello che è lo scopo della business intelligence, cioè la creazione di uno strumento di supporto per assistere le fasi decisionali. La parte centrale della piramide può essere scomposta in due macro-funzionalità: la *generazione di informazioni* e la *creazione di conoscenza*.

La *generazione di informazioni* è rappresentata dall'analisi OLAP ed è un processo che permette di produrre nuove informazioni partendo dai dati prelevati dalle sorgenti: opportune tecniche di pulizia dei dati ed eliminazione dei duplicati hanno permesso di creare un nuovo dominio delle informazioni, maggiormente consistente e con una maggiore qualità dei dati.

La creazione di conoscenza è il cuore del sistema e ha permesso di ricavare nuove informazioni applicando deduzioni logiche sui dati a disposizione. Esso è composto da tre funzionalità: *l'esplorazione delle informazioni*, il *data mining* e *l'analisi what-if*. *L'esplorazione di informazioni* applica analisi statistiche sui dati, ricavando informazioni di sintesi utili nella macro-analisi dei dati. Il *data mining* offre tecniche di analisi di grandi quantità di dati con lo scopo di creare delle correlazioni tra il verificarsi di fenomeni diversi. *L'analisi what-if è una simulazione che opera sui dati con lo scopo di controllare il comportamento di un sistema complesso [...] In particolare, l'analisi what-if misura come i cambiamenti su un insieme di variabili indipendenti hanno influenza su un insieme di variabili dipendenti, utilizzando un sistema di simulazione [RIZ-AWI]*. In sostanza, l'analisi what-if offre la possibilità di avvalersi di sistemi predittivi, utili per verificare l'impatto che determinate scelte possono avere sul sistema.

1.4 Decision support systems

1.4.1 Definizione

Il *decision support system* (per abbreviazione DSS) è nato per rispondere alla crescente disponibilità di dati prodotti dai sistemi informativi. La generazione di questi dati sono il frutto delle operazioni eseguite dai sistemi informativi e sono utilizzati per coordinare i processi operativi. La realizzazione di sistemi ad hoc per l'analisi dei dati è spinta da due forti motivazioni: la richiesta di nuove tecniche per l'analisi di quest'enorme mole di dati e l'impossibilità di poter estrarre informazioni dai dati a disposizione. L'obiettivo dei DSS è fornire indicazioni ad un platea di utenti che hanno lo scopo di agire sui processi decisionali.

I decision support system sono sistemi software interattivi che permettono all'utente di valutare e scegliere tra diverse possibilità [MDF-DSS]. L'obiettivo dei DSS è quello di fornire un sistema informatico in grado di supportare il processo decisionale: i DSS sono impiegati in tutti i casi che prevedono di valutare l'impatto derivante da una determinata scelta, comparandola con le altre alternative disponibili.

Un sistema di questo tipo deve, in primo luogo, essere in possesso di informazioni che permettano di valutare lo stato attuale dell'azienda: le basi di dati dei sistemi operazionali si sono rivelate una buona base di partenza, soprattutto per poter ricostruire informazioni utili a raccontare il modo di operare dell'azienda. Una volta messe a disposizione tali informazioni, sono venute in soccorso del DSS le tecniche che permettono l'estrapolazione di informazioni partendo da dati grezzi: lo studio di modelli che hanno permesso di ricavare informazioni utili al processo decisionale sono alla base di questo sistema.

Assieme alla necessità di reperire informazioni da grosse mole di dati, si richiede che gli strumenti dedicati all'analisi siano anche flessibili e interattivi, così da poter fruire agevolmente del contenuto dei dati a seconda della tipologia di interrogazione che viene sottoposta al sistema.

Tra i requisiti alla base di questi sistemi è introdotta la richiesta di fornire uno strumento adattabile, cioè capace di sottoporsi a problemi diversi, con la possibilità di riuscire ad utilizzare le stesse tecniche di analisi anche su modelli di dati diversi.

1.4.2 Caratteristiche

Nel corso della storia, per rispondere alle diverse esigenze dell'utente i sistemi DSS sono stati applicati in ambiti differenti, generando diverse tecniche di ricerca della soluzione finale: questo avvenimento ha prodotto, nel tempo, una serie di sotto-sistemi diversi. Una prima classificazione è emersa sulla base della tipologia di supporto che il sistema è in grado di fornire all'utente. In particolare, sono emersi tre tipi di modalità capaci di assistere le analisi dell'utente:

- I **sistemi passivi** sono utilizzati per raccogliere informazioni sullo stato attuale del sistema, ma non comprendono tecniche per misurare come eventuali scelte possano pesare sull'efficienza del sistema informativo.
- I **sistemi attivi** hanno la caratteristica di riuscire a fornire informazioni previsionali, grazie alle quali hanno la possibilità di informare l'utente con suggerimenti e soluzioni, così da supportare l'utente nella decisione tra più alternative possibili.

- I **sistemi cooperativi** permettono maggiore flessibilità, in quanto danno la possibilità all'utente di agire attivamente, anche sui suggerimenti forniti in una prima analisi del sistema: la possibilità di un dialogo interattivo tra il sistema e l'utente fornisce uno strumento altamente flessibile.

Indipendentemente dal tipo di supporto offerto all'utente, un sistema DSS è principalmente composto da 3 elementi fondamentali: il *database management system*, che gestisce grandi quantità di dati relativi alla classe di problemi per cui il DSS è stato progettato, il *model-base management system*, che supporta il processo di applicazione di un modello teorico al problema reale da risolvere, e il *dialog generation and management system*, che costituisce il modulo dedicato all'interfacciamento utente, progettato anche per utenti non competenti in ambito informatico.

Oltre alla tipologia di supporto, altre classificazioni di dati dipendono da più fattori: dall'ambito in cui sono impiegati, sulla tipologia e quantità di informazioni su cui devono operare le proprie decisioni, sul tipo di servizio d'analisi che devono offrire all'utente per migliorare il supporto fornito, etc. La maggior parte delle specializzazioni dei sistemi DSS si sviluppa sulla base dell'ambito su cui sono applicati. Una possibile classificazione maggiormente dettagliata rispetto alla prima è composta da:

- Model-driven DSS
 - Questi sistemi nascono in ambito di accesso e manipolazione di dati in campo finanziario; l'obiettivo è di creare dei modelli di simulazione per l'applicazione di algoritmi di ottimizzazione. Si basano su modelli semplici, che prendono in considerazione dati limitati e che pongono l'accento sullo scambio di informazioni tra utente e sistema, con lo scopo di migliorare la ricerca.
- Data-driven DSS
 - Concentrano la loro attenzione su serie temporali di dati, con possibilità di operare perfino su dati real-time. L'analisi di dati storicizzati risulta quindi essere l'elemento di maggiore interesse per quanto riguarda l'analisi.
- Communication-driven DSS
 - Il focus di questi sistemi si pone sull'utilizzo di infrastrutture capaci di fornire il miglior supporto nell'ambito della comunicazione. Le analisi per definire le decisioni rilevanti

sono fortemente supportate da bacheche interattive, oltre ad audio e video conferenze.

- Document-driven DSS
 - Il modello fondamentale su cui si basa questo sistema è il recupero di informazioni direttamente da ogni forma di documento, sia esso un'immagine, un file audio, un video o un testo.
- Knowledge-driven DSS
 - Questi sistemi fungono da assistenti personali le cui maggiori competenze sono attinenti al dominio: la funzione principale è di capire i problemi del dominio e tentare di trovare soluzioni che li risolvano.
- Web-based DSS
 - Questi sistemi prendono il nome dal web che risulta essere il dominio su cui sviluppano il reperimento dei dati e la ricerca di informazioni.

L'elenco dei vari sistemi di decision support system riesce a dare un'idea chiara sui campi di applicazione di queste tecnologie, soprattutto sulla base della tipologia di sorgente dati disponibile. Successivamente verrà maggiormente analizzata l'architettura che compone questi sistemi, definendo come si strutturano per svolgere il tipo di analisi messa in campo.

1.4.3 Architettura

Per comprendere in maniera più completa il modo in cui queste tecnologie si strutturano è opportuno approfondire l'architettura con cui dividono le proprie funzionalità. Come detto in precedenza l'idea di questi sistemi è quella di fornire risultati, i quali non sono altro che informazioni che possono essere recuperate attraverso opportune tecniche d'analisi; tutte le operazioni si applicano su database, i quali contengono informazioni non strutturate.

Le componenti principali che fanno parte di un sistema DSS sono:

- *Data management system*. Questa parte dell'architettura si occupa della gestione e dell'elaborazione dei dati; i dati presi in considerazione sono informazioni che servono al sistema per fornire supporto all'attività del DSS.
- *Model management system*. Esso è il motore che permette tramite la modellazione dei problemi di poterli studiare e di recuperare informazioni utili per una risoluzione.
- *Motore della conoscenza*. Ha il compito di fornire conoscenza, esperienza e apprendimento continuo.
- *Interfaccia utente*. Supporta l'utente nella gestione delle interrogazioni e permette la visualizzazione dei dati d'interesse.
- *End-user*. L'utente ha il ruolo di determinare gli obiettivi del sistema e di intervenire, migliorando o indirizzando la soluzione, in ogni momento del processo di ricerca;

Definite le basi con cui le architetture dei processi decisionali si compongono, si tratteranno aspetti legati alla sfera di concetti compresi dai data warehouse.

1.5 OLTP e OLAP

Il mondo della business intelligence fornisce un insieme di tecniche che mirano a migliorare l'organizzazione dei sistemi informativi. Business intelligence e Sistemi informativi hanno due scopi diversi; la loro manipolazione dei dati del dominio dipende dalle operazioni che svolgono per raggiungere i loro obiettivi.

I sistemi informativi coordinano più processi, i quali operano su un insieme definito di dati da elaborare prima di fornirli al processo successivo. Le caratteristiche principali nella gestione dei dati sono:

- Un costante accesso ai dati, con un alto numero di transazioni: i dati che sono interessati da ogni operazione sono un piccolo insieme;
- Bassa interattività con i dati, infatti il flusso dei dati d'interesse è definito a priori, in linea con le funzionalità offerte dal quel determinato processo; l'interattività è congelata perché definita in fase di progettazione del processo;

Sistemi di questo tipo sono detti On-Line Transactional Processing (OLTP per abbreviazione): è facile dire che gli strumenti di business intelligence svolgono operazioni su dati di natura molto diversa, maggiormente orientati ad analisi di alto livello su grosse quantità di dati. Infatti, le peculiarità di uno strumento di BI possono essere così sintetizzate:

- Lettura di un enorme quantità di dati, utili a fotografare lo stato del sistema in diversi istanti temporali. Restituiscono in output dati di sintesi;
- Il carico di lavoro fortemente variabile è causato dalla forte interattività con il sistema; infatti, la generazione delle interrogazioni può comprendere carichi di lavoro molto diversi tra una operazione e l'altra. L'interattività è definita ad ogni interrogazione dall'utente, che definisce il tipo di analisi da effettuare sui dati.

Sistemi che presentano queste caratteristiche sono detti *On-Line Analytical Processing* (OLAP per abbreviazione).

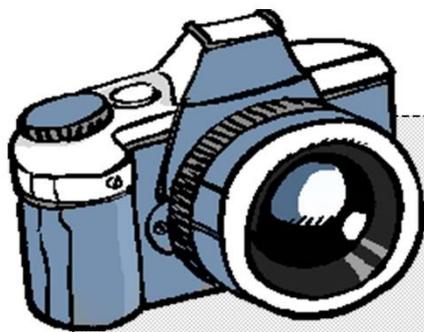
In letteratura, la rappresentazione di queste due diverse metodologie di interrogazione è associata in maniera esemplificativa a due figure che permettono di catturarne rapidamente le differenze. Per quanto riguarda le tecnologie OLTP, si prende come riferimento il database operativo, mentre per i sistemi OLAP, a fare da garante dei prerequisiti sopra esposti è il *data warehouse*.

Il Data warehouse è infatti lo strumento principale che consente di evidenziare tutte le caratteristiche con cui i sistemi OLAP si distinguono da quelli OLTP; infatti, come concetti principali dei sistemi data warehouse è possibile identificare:

- La manipolazione dei dati è concentrata sull'utente finale che dovrà visionarli;
- Pur prendendo i dati da più sorgenti differenti, determinate trasformazioni permettono di restituire una visione dei dati unificata.
- I dati sono storicizzati tenendo in considerazione un lungo intervallo temporale.
- La lettura di grandi volumi di dati prende in considerazione informazioni non volatili.

Per comprendere al meglio queste due grandi tipologie di interrogazione dati, ci si avvale della *figura 3*: i sistemi OLTP fotografano lo stato del sistema in un determinato momento, infatti vengono rappresentati dalla macchina fotografica che allo stesso modo, fotografa lo stato della realtà in un

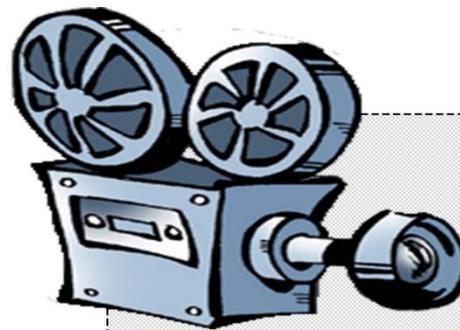
istante temporale definito; i sistemi OLAP permettono di catturare lo stato dei dati in diversi istanti temporali; la figura di riferimento per la rappresentazione di questi sistemi è la cinepresa.



OLTP

Database operazionali

- Migliaia di utenti
- Transazioni cablate negli applicativi
- Accesso lettura / scrittura
- Scopo legato all'applicazione
- Dati elementari di dettaglio
- Focus della qualità sulla integrità del dato
- Rappresentano lo stato attuale del sistema
- Aggiornamenti continui
- Modello normalizzato
- Accesso mirato a poche tabelle



OLAP

Data warehouse

- Centinaia di utenti
- Transazioni di analisi ad hoc
- Accesso in lettura su milioni di record
- Scopo supporto alle decisioni
- Dati di sintesi
- Focus della qualità sulla consistenza del dato
- Rappresentano lo stato attuale e storico del sistema
- Aggiornamenti periodici
- Modello multidimensionale
- Accesso su gran parte delle tabelle

Figura 3 - Differenze tra interrogazioni OLTP E OLAP

1.6 ETL

Come si è potuto evidenziare precedentemente, una delle caratteristiche che contraddistingue maggiormente le basi di dati OLAP dalle OLTP è la maggiore consistenza dei dati, cioè la presenza di informazioni che, seppur integrate da sorgenti diverse, risultano essere maggiormente complete e coerenti. La possibilità di avere dati di questo tipo è dovuta essenzialmente ad uno step compreso nella costruzione del data warehouse, denominato *extraction, transformation and loading* (abbreviato ETL): si occupa di caricare i dati all'interno del data warehouse, in maniera tale da avere una migliore qualità dei dati rispetto alle singole sorgenti.

La definizione ETL prende il nome direttamente delle tipologie di operazioni che vengono effettuate sui dati, con lo scopo di poterli rendere disponibili al sistema. Le tre macro-operazioni che sono effettuate sulle sorgenti comprendono **l'estrazione**, che identifica i database sorgenti e su cui viene fatta la lettura dei dati, **la trasformazione**, in cui i dati sono resi maggiormente coerenti tra loro con una qualità migliore rispetto a quelli di partenza e il caricamento, in cui i dati sono ribaltati sulla struttura del data warehouse. Di seguito forniamo un elenco delle attività che sono condotte sui dati in questa fase della creazione del data warehouse:

- L'estrazione permette di interrogare direttamente le basi di dati operazionali: in una prima lettura dalle sorgenti, si procede con un'importazione completa dei dati, mentre nelle successive interrogazioni, determinati meccanismi di controllo permettono di prendere in considerazione solo i nuovi dati inseriti;
- Un controllo dei dati duplicati permette di eliminare le informazioni ripetute più volte all'interno della stessa sorgente dati;
- Il controllo di validità logica nelle associazioni presenti permette di evitare di riportare, sulla base di dati di destinazione, associazioni che sono fisicamente presenti, ma che logicamente non devono essere prese in considerazione dal sistema. Un simile controllo viene applicato sui dati che presentano valori impossibili o errati.
- L'operazione di trasformazione si prende in carico la possibilità di valorizzare i campi mancanti; le informazioni non valorizzate sono dedotte da quelle presenti, ma solo nei casi in cui siano logicamente correlate.
- La pulizia dei dati comprende l'eliminazione dei campi che non sono usati: questa operazione riduce la complessità del modello, utilizzando meno risorse del sistema e migliorando la chiarezze del modello finale.
- Campi che presentano testi liberi, ma di cui si hanno informazioni in merito, sono resi maggiormente coerenti tramite trasformazioni opportune.
- Dati appartenenti a sorgenti diverse che rappresentano il medesimo dettaglio, sono trasformati in un unico concetto nel modello di destinazione.
- I dati sono opportunamente sintetizzati nel caso in cui la sorgente di destinazione abbia un livello meno dettagliato dei dati di origine. L'operazione di aggregazione permette di ridurre la cardinalità delle tabelle e di migliorare l'utilizzo delle risorse del sistema.

Una volta effettuate queste fasi, i dati vengono opportunamente caricati nel data warehouse; a questo punto, l'utente ha la possibilità di interrogare l'opportuna interfaccia per effettuare le analisi specifiche.

Capitolo 2

Energy Management System

Lo sviluppo di questo capitolo è stato pensato per portare il lettore a conoscere meglio il mondo delle tecnologie SCADA. Lo studio che sarà effettuato prenderà in considerazione un preciso ambito di applicazione dei sistemi SCADA, quello dell'efficiamento energetico: particolarmente legati a questo ambito sono i sistemi di produzione di energia e quelli di controllo dei consumi. I sistemi SCADA che operano funzionalità specifiche nel campo della gestione energetica sono chiamati SCADA/EMS. Questo capitolo permetterà di capire quali sono le funzionalità messe in campo dai sistemi per la supervisione energetica: per un'analisi completa, saranno prese in considerazione le soluzioni software di tre diverse aziende impiegate nel campo dell'ITC. Lo scopo di valutare diverse soluzioni informatiche è quello di capire i diversi campi di applicazione, oltre a esaminare le funzionalità che hanno portato il mercato a scommettere sul loro utilizzo.

2.1 Sistemi SCADA

I sistemi SCADA sono ampiamente usati in ambito industriale per la supervisione, il controllo e l'acquisizione dati da processi industriali [DSA-WIS]. L'acronimo SCADA sta per Supervisory Control And Data Acquisition. Il significato di tale acronimo propone le due attività principali di questi sistemi: l'acquisizione dati e il controllo utile alla supervisione del sistema che si propongono di gestire. Le soluzioni SCADA fungono principalmente da supporto ad altri sistemi, sui quali propongono di migliorare la gestione delle seguenti attività: la raccolta dei dati, la supervisione in tempo reale e il controllo diretto, cioè la possibilità da parte dell'operatore di cambiare attivamente lo stato del sistema, agendo direttamente sui dispositivi installati nell'impianto monitorato.

L'utilizzo di sistemi SCADA non è da considerarsi esclusiva di impianti industriali. Nel corso degli anni, sono state molteplici le realtà che hanno adottato queste soluzioni; infatti, tra i settori che utilizzano i sistemi SCADA si includono gli impianti di generazione e di distribuzione di energia, impianti siderurgici, impianti chimici fino ad arrivare a progetti sperimentali su impianti nucleari. Anche a livello domestico è possibile avere contatti con queste tecnologie: sono impiegate sia nell'ambito del monitoraggio per la verifica dei consumi energetici, sia nel controllo di apparecchiature come impianti di ventilazione e sistemi di illuminazione.

Un'altra caratteristica fondamentale dei sistemi SCADA è che la loro applicazione può comprendere sistemi che hanno una forte distribuzione geografica: la centralizzazione delle fasi di analisi e controllo dei dati permette di poter disporre di indubbi vantaggi. Questo permette ai sistemi SCADA di poter operare anche nel controllo del traffico (aereo, ferroviario, automobilistico), nella gestione di sistemi del trasporto dei fluidi (acquedotti, gasdotti, oleodotti, ...) e in strutture dedicate al telerilevamento ambientale.

La causa che hanno spinto le organizzazioni a dotarsi di soluzioni SCADA sono molteplici: la principale è da attribuire al crescente numero di apparecchiature adottate nei sistemi industriali. Infatti, le aziende sono arrivate a dotarsi di strumentazioni estremamente complesse: questo fattore ha portato a dover monitorare costantemente decine di migliaia di componenti per ogni impianto. Oltre alla complessità, un altro fattore molto importante è la supervisione di tali sistemi, soprattutto nel caso in cui determinati parametri nella lavorazione industriale diventino critici: l'obiettivo è di ridurre i tempi di controllo e di gestione delle criticità; tali fattori hanno un'importanza fondamentale nel processo industriale. Un'altra causa che ha facilitato la diffusione di queste soluzioni è la possibilità di analizzare i dati immagazzinati, permettendo all'utente di elaborare informazioni altrimenti inaccessibili.

Per una comprensione ampia di questi sistemi, è d'obbligo far notare che nello svolgimento delle loro funzioni, possono incorrere in molti problemi. Tra i principali è possibile evidenziare la distribuzione geografica del sistema, la distribuzione dell'intelligenza di controllo e i tempi di reazione del sistema ai comandi dell'operatore umano.

Le numerose problematiche, tecnologie e architetture che questi sistemi presentano, creano un divario tra la modellazione di un sistema di questo tipo e il suo effettivo campo di applicazione: sono numerosi i fattori che intervengono nella valutazione dell'applicabilità di un sistema di questo tipo sul problema in oggetto.

In sostanza, quello che contraddistingue tali sistemi, indipendentemente dal dominio di applicazione, sono le tre funzionalità base che ad essi sono richieste:

- **Acquisizione dei dati.** Questa caratteristica risulta di basilare importanza rispetto alle altre. L'acquisizione di tali dati risulta essere la prima attività di cui è indispensabile poter disporre; infatti, senza questa funzionalità le altre non hanno ragione di esistere. La difficoltà dell'implementazione consiste nello scambio di informazioni con altri sistemi: questa problematica è resa particolarmente complessa dall'estrema eterogeneità delle tecnologie utilizzate, oltre all'elevato numero dei diversi protocolli impiegati per l'interscambio delle informazioni. Va inoltre aggiunto che alla base della rilevazione dei dati ci sono comunque dei sensori, i quali possono essere gestiti da sistemi che si interpongono tra il sensore stesso e il sistema di recupero.
- **Supervisione dei dati.** La supervisione dei dati si rende necessaria per un costante controllo del sistema; tramite questa funzionalità, l'utente supervisore ha la possibilità di verificare il corretto modo di evolversi dei sistemi. Va considerato che la supervisione è un'attività eseguibile in due modalità: la prima permette di verificare lo stato attuale del sistema, mentre la seconda è un'analisi del sistema con una prospettiva temporale più ampia, in cui è possibile studiare l'andamento del sistema con un numero di informazioni molto maggiore.
- **Controllo dei dati.** Rappresenta la capacità di intervenire su un sistema e decidere come cambiarne l'evoluzione. Normalmente la stessa architettura che viene utilizzata per l'acquisizione del dato, è sfruttata anche per modificare il dato stesso. Il controllo dei dati è un'attività da eseguire in tempi rapidi, per rispondere in maniera immediata a funzionamenti anomali del sistema.

Le funzionalità appena definite sono alla base di questi sistemi e, per supportare le peculiarità di cui si fanno garanti, necessitano di un'infrastruttura che ne rispetti le caratteristiche, mantenendo una separazione distinta tra i diversi sottosistemi che devono integrare.

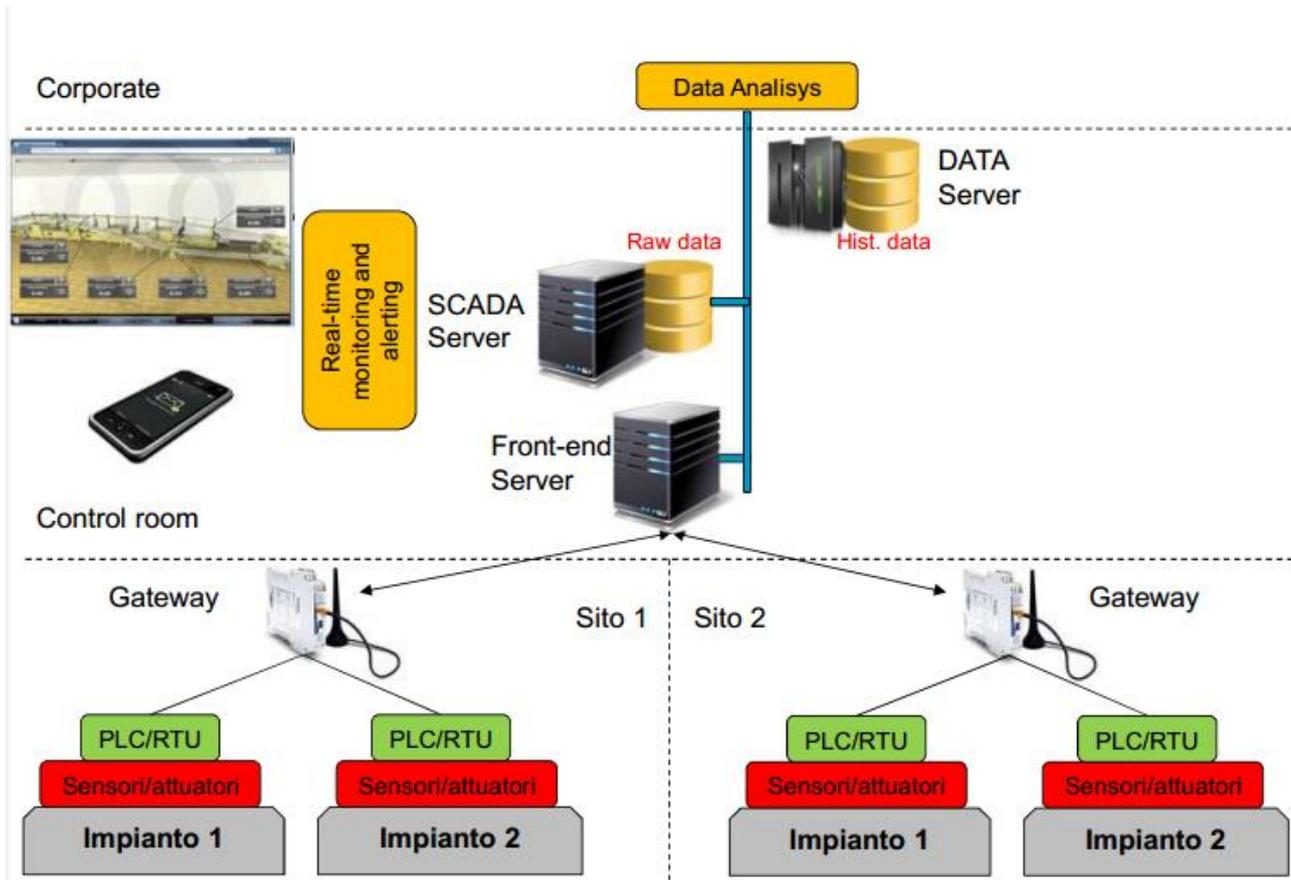


Figura 4 - Architettura dei sistemi SCADA

Come è possibile vedere dalla *figura 4*, l'implementazione di questi sistemi si compone essenzialmente di tre distinte parti che hanno la caratteristica di strutturarsi a livelli: un livello svolge un ruolo di interfacciamento con il mondo esterno, mentre i restanti sfruttano le funzionalità messe a disposizione dal primo. Tutti e tre i livelli hanno la caratteristica di poter essere allocati in punti differenti del sistema: questo fattore è direttamente collegato alla possibilità di essere geograficamente distribuiti.

Quando si descrive l'architettura di questi sistemi, è opportuno definire gli elementi esterni con il quale è necessario interfacciarsi per una prima trasformazione delle informazioni: i dati letti dai sistemi esterni assumono una precisa collocazione nel sistema SCADA. La parte esterna al software è rappresentata da diversi impianti, dove ogni impianto contiene la propria componentistica hardware che esegue la rilevazione dei dati d'interesse. Gli elementi che fanno parte di ogni impianto si differenziano per la logica con cui agiscono sui dati: si possono trovare i sensori, che rilevano il dato grezzo direttamente sul campo, i controllori, che permettono di gestire un insieme

di sensori e possono applicare logiche sui dati, e i gateway, che raccolgono dati dai controllori e permettono di poter gestire operazioni direttamente sui sensori.

Il primo livello che fa parte del sistemi SCADA è il **Front-End Server**; esso ha il compito di interfacciare i dati provenienti dai gateway con il sistema stesso: è sostanzialmente un ponte tra il sistema di controllo e gli impianti da supervisionare. Oltre a garantire la correttezza dei dati in termini di qualità, deve essere tempestivo nel controllo dei dati in ambito real-time. Come detto, il front-end server garantisce un set di funzionalità dedicate, rese successivamente disponibili dagli altri due moduli del software.

Il secondo livello è denominato **SCADA Server** e fornisce le funzionalità di monitoraggio e gestione dei dati: non si occupa di informazioni storicizzate, ma permette di gestire in prima persona gli allarmi e garantisce la gestione diretta dei dati sulla base di aggiornamenti in tempo reale, il tutto scambiando informazioni con il Front-End Server.

Il terzo livello è denominato **DATA Server** e include uno strumento di analisi dei dati, il quale può comprendere una serie di strumenti più o meno sofisticati, su cui vengono fatte le analisi maggiormente informative di tutto il sistema. I dati analizzati sono gli stessi di quelli rilevati dai sensori, ma storicizzati con intervalli di tempo che possono variare in base alle necessità: la storicizzazione dei dati implica una proporzionalità inversa tra l'ampiezza dell'intervallo di memorizzazione tra due dati consecutivi e il costo delle risorse di memorizzazione: infatti, ad un breve intervallo di tempo per la storicizzazione dei dati corrisponde un alto costo delle infrastrutture necessarie alla memorizzazione delle informazioni.

Nella prossima sezione della tesi verranno considerati prodotti che, nel campo delle soluzioni SCADA, presentano funzionalità specifiche per l'impiego di tali tecnologie nell'ambito del controllo dei consumi energetici; tale settore richiede funzionalità specifiche mirate allo studio, all'analisi e al monitoraggio di fenomeni caratteristici.

2.2 SCADA/EMS

Le soluzioni SCADA comprendono delle specializzazioni particolari a seconda del tipo di dominio e del settore su cui operano. L'ambito d'interesse per quanto riguarda questa tesi è quello relativo al consumo energetico, sia esso domestico o industriale. L'obiettivo principale, che porta ad inserire nel proprio processo aziendale una soluzione di controllo di questo tipo, è la verifica dei consumi energetici del proprio sistema produttivo, con il quale è possibile individuare eventuali perdite o malfunzionamenti degli impianti, allo scopo di limitare lo spreco delle risorse utilizzate.

Le soluzioni SCADA indirizzate ad analizzare dati relativi a consumi energetici sono chiamate *energy management system*: queste soluzioni mirano dunque a proporre una serie di strumenti di analisi dedicati al settore energetico, il quale possiede delle peculiarità che altri sistemi appartenenti al mondo delle soluzioni SCADA non hanno la necessità di gestire.

La crescente necessità di doversi affidare a strumenti di controllo energetico nasce dall'importanza che, negli ultimi decenni, ha spinto i processi produttivi a diminuire l'impatto ambientale sull'intero ecosistema: oltre al reperimento e smaltimento delle materie utili alla lavorazione, l'attenzione da



Figura 5 - Esempi di sistemi di produzione di energia

parte del comparto industriale è stata quella di migliorare il consumo energetico dei propri impianti, eliminando gli inutili sprechi di energia e avvalendosi di sistemi per la rilevazione e il controllo dei consumi energetici.

Non solo questi sistemi sono impiegati nel telecontrollo dei consumi in ambito industriale, ma hanno altre differenti applicazioni, come ad esempio nel campo delle energie rinnovabili (mostrate dalla *figura 5*), dove svolgono il ruolo di supervisore per quanto riguarda il corretto funzionamento nel complesso processo di produzione energetica. In questo ambito propongono funzionalità specifiche del dominio, come l'adozione di algoritmi che calcolano la produzione di energia teorica del proprio impianto: queste informazioni permettono di capire qualitativamente lo stato dei propri processi di produzione. Le stime sulla produzione teorica di energia sono generalmente fatte recuperando dati meteorologici, che servono come informazioni per calcolare i valori teorici di produzione. In sostanza, queste soluzioni sono preferite nell'ambito dei sistemi per la generazione di energia come fotovoltaici o eolici: la differenza principale di questi impianti sta soprattutto nell'analisi dei dati, essendo la parte di acquisizione e controllo dei dati identica a quella di una basilare piattaforma SCADA.

Tenendo in considerazione le tre funzionalità principali di un sistema SCADA, quella che include le maggiori modifiche dal punto di vista funzionale è il modulo di analisi dei dati: sia la parte di acquisizione che quella di controllo presentano caratteristiche comuni alle altre specializzazioni di queste tecnologie. Perciò, si parla normalmente di *advanced applications*, cioè di funzionalità aggiuntive ai sistemi SCADA che dipendono dalla tipologia di specializzazione fornita.

Per riuscire a valutare in maniera coerente un sistema SCADA / EMS, è opportuno definire quali siano le caratteristiche fondamentali di questi sistemi, così da accertarne la coerenza. Le funzionalità principali che contraddistinguono un sistema EMS riguardano la gestione software. Le caratteristiche di maggiore importanza sono:

- L'aderenza completa del sistema alle caratteristiche base di un sistema SCADA. In particolare è opportuno specificare eventuali dettagli che possono identificare e valorizzare i servizi messi in campo dal sistema:

- Protocolli di acquisizione dati supportati dal sistema, cioè l'interfacciamento con un numero di dispositivi il più ampio possibile.
- Possibilità di integrare dati al sistema con diverse metodologie.
- Intervallo di tempo minimo dedicato alla memorizzazione dei dati grezzi.
- Tempestività nel controllo real-time dei dispositivi.
- Sistema di accesso dati con verifica delle credenziali e assegnazione di ruoli e funzionalità specifiche per utente o per gruppi di utenti.
- Monitoring dei dati con diagrammi specifici, soprattutto per quanto riguarda la visualizzazione di serie temporali;
- Gestione di controllo allarmi, con pianificazione dei valori limite e gestione di un sistema di segnalazione, che può andare dal semplice pop-up a schermo ad un sistema di e-mailing e invio sms.
- Reporting dei dati flessibile e personalizzabile, che permetta di fornire dati di sintesi in relazione ad altre informazioni d'interesse.
- Creazione e visualizzazione personalizzata di KPI: i *Key Performance Indicators* sono indicatori che hanno lo scopo di misurare le prestazioni del processo aziendale; l'individuazione e lo studio di tali indicatori è complesso, ma quello che si richiede da una piattaforma EMS è di poterli calcolare sulla base degli altri dati, assieme ad una loro rapida consultazione.

Oltre alla descrizione delle funzionalità principali dei sistemi SCADA e di quelli EMS, non verranno presi in considerazione eventuali approfondimenti sulle tipologie di architetture che questi sistemi implementano, non essendo l'argomento principale di questa tesi. Sarà svolta una comparazione a grandi linee delle funzionalità implementate da diversi software, che competono sul mercato delle soluzioni SCADA/EMS.

Scopo della prossima sezione, dunque, sarà lo studio di soluzioni software sviluppate da diverse aziende operanti nel settore del telecontrollo di sistemi SCADA/EMS. I prodotti presi in considerazione saranno valutati sia per l'architettura che implementano, sia per le funzionalità di telecontrollo. Una particolare comparazione sarà fatta sulla parte di analisi del dato, che

rappresenta lo scopo principale: la gamma di funzionalità disponibili può variare fortemente da un prodotto all'altro.

2.3 Sistemi presenti sul mercato: caratteristiche e funzionalità

Al fine di valutare in maniera completa gli elementi che la soluzione On.Energy prende in considerazione, introduciamo all'interno di questa analisi il confronto con altre soluzioni SCADA presenti sul mercato, con l'obiettivo di comparare più soluzioni tra di loro. Questo confronto permette di avere una maggiore visione sulle implementazioni che questi sistemi propongono in soluzioni concrete. Va aggiunto che ogni sistema può presentare grosse differenze sia dal punto di vista dell'infrastruttura utilizzata, sia dalle funzionalità che permette di utilizzare. Quello che questo confronto cercherà di mettere in luce è, da una parte l'aderenza di ogni prodotto con le specifiche generali dei sistemi SCADA/EMS, e dall'altra le caratteristiche maggiormente innovative di ogni prodotto rispetto agli altri.

Per tutte le soluzioni che si analizzano all'interno di questa tesi, è specificato l'ambiente da cui provengono le aziende che le realizzano. Infatti, il settore di provenienza di un'azienda implica un insieme di conoscenze specifiche, arricchite dalle esperienze maturate sul campo. Questo dettaglio è fortemente riscontrabile analizzando le soluzioni realizzate: per esempio, Onit Group nasce principalmente come software house, quindi detiene delle conoscenze specifiche nell'ingegnerizzazione del software, mentre non è a disposizione di soluzioni hardware per il recupero dei dati sul campo. In questo caso, per adempiere a questa lacuna infrastrutturale, sono richieste forniture esterne di hardware; queste collaborazioni permettono di stabilire contratti di partnership con altre aziende complementari sul mercato (vedi ad esempio aziende come Casadei Pellizzaro, Sinergia, etc....).

Nelle prossime sezioni verranno illustrate le soluzioni di tre software che concorrono nella fetta di mercato dei sistemi SCADA/EMS: in particolare, sarà analizzata la soluzione di una piccola impresa

italiana, cioè la *SIRIUS* con il prodotto *GMPPower*, assieme ad una soluzione appartenente ad un'azienda multi-nazionale, rappresentata dalla svizzera *ABB* con il prodotto *Network Manager*. Alla fine di questi due approfondimenti, verrà presentata la soluzione *On.Energy management system* di *Onit Group s.r.l.*, comparando ogni caratteristica con quello che è stato possibile far emergere nelle soluzioni precedenti.

2.4 SIRIUS

Sirius è un'azienda italiana che sviluppa soluzioni per il telecontrollo elettrico, orientata a soddisfare le esigenze di sistemi destinati alla produzione di energia [SIR-SHP]; essa opera nel settore dell'automazione e del telecontrollo da più di 10 anni, settore in cui ha una profonda conoscenza in campo hardware.

La suite creata da questa società deriva fortemente dal campo dell'elettronica: infatti, le proposte messe in campo sono capaci di gestire un sistema che comprende sia la parte hardware, che quella di telecontrollo software. L'azienda pone l'attenzione sulla produzione di energie rinnovabili, in particolare eolico e fotovoltaico: le loro soluzioni includono l'installazione di sensori di misure meteorologiche; i dati prelevati dai sensori permettono di calcolare stime sulla produzione teorica di energia, al fine di confrontare la resa del proprio impianto.

L'analisi condotta nei prossimi paragrafi cercherà di verificare la completezza delle funzioni messe in campo dalla suite, confrontandola con le funzionalità richieste da questi software. Una prima parte metterà in luce gli aspetti legati all'architettura della soluzione, mentre una seconda parte evidenzierà le funzionalità software.

2.4.1 Architettura

La soluzione proposta permette di controllare più impianti con un unico sistema centralizzato: questo grazie al modulo di *GPM Local Front-End* che gestisce tutte le comunicazioni con le varie

installazioni sottoposte a controllo, centralizzando il recupero e la storizzazione dei dati, garantendo nel tempo anche un'ottimizzazione dei consumi di banda.

Come è possibile constatare dalla *figura 6*, il sistema si suddivide in due macro moduli: la parte di interfacciamento e il sistema centrale. La parte di interfacciamento è normalmente posta in prossimità della rete di controllo, mentre il sistema centrale è posizionato in base al tipo di installazione scelta: nel caso di **installazione centralizzata**, il sistema centrale non è sulla stessa rete di quello d'interfacciamento, ma è instaurato tra i due un tunnel di comunicazione affidabile e sicuro; invece nell'**installazione stand-alone**, i due moduli stanno praticamente sullo stesso server.

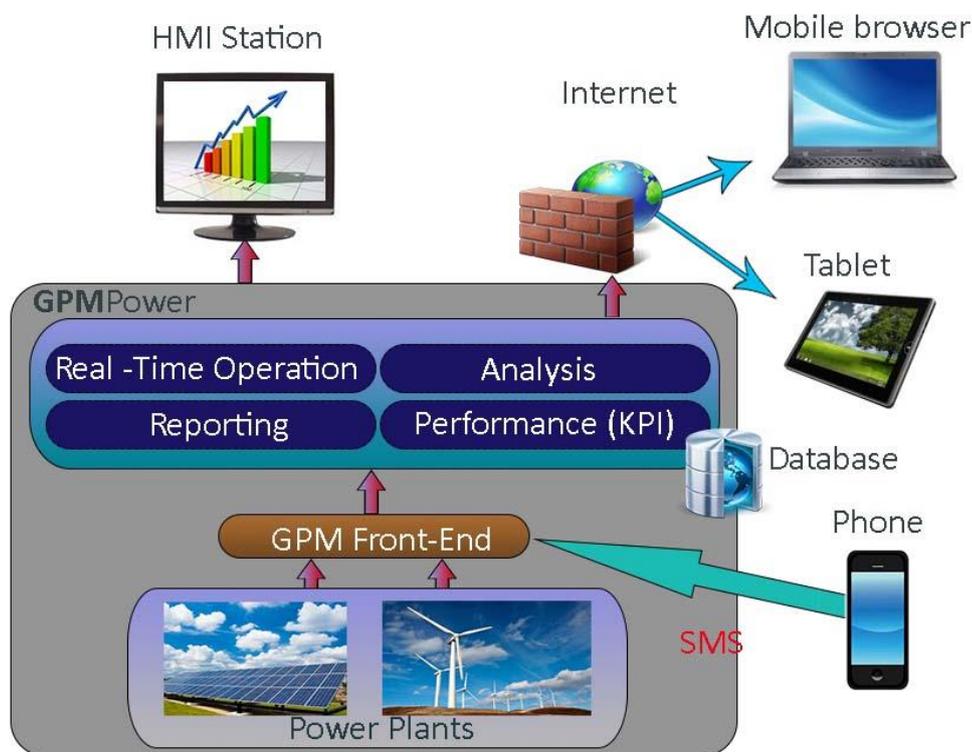


Figura 6 - Architettura di GMPPower di SIRIUS

Comparando questi aspetti con quanto detto nelle slide di presentazione delle tecnologie SCADA, la funzione di Front/End è compiuta da *GMP Front-End*, la parte di SCADA Server è rappresentata dal modulo *Real Time Operation* e il Data Server è composto dai moduli di *Reporting*, *Performance(KPI)* e *Analysis*.

Da notare che la soluzione di SIRIUS si compone direttamente di un pacchetto hardware che pone come funzionalità di particolare interesse un sistema di ridondanza dei dati, oltre a fornire un gruppo di continuità per il funzionamento del sistema anche in mancanza di elettricità.

2.4.2 Data acquisition

Per quanto riguarda l'acquisizione dei dati, c'è un numero importante di protocolli di comunicazione supportati. L'acquisizione dei dati è divisa tra le due diverse soluzioni GMPPower e SIRIUS. Tra i tanti protocolli supportati possiamo citare Modbus, Odbc, Opc, Xml SA, Profibus, Device NET, etc....

Per migliorare il sistema di storicizzazione dei dati per la *data analysis*, il sistema di SIRIUS mette in campo un insieme di accorgimenti che permettono di garantire ai dati storicizzati alcune proprietà, come la consistenza dei dati ed un minore utilizzo delle risorse hardware per la trasmissione.

Nella soluzione promossa da SIRIUS, è utilizzata una soluzione che può comprendere un sistema di storicizzazione dei dati temporaneo tra i dispositivi sul campo e il database di analisi; è quindi scelta una soluzione con un database mantenuto localmente. I motivi di una soluzione di questo tipo garantiscono due vantaggi: la possibilità di utilizzare il database locale come buffer temporaneo dei dati e il salvataggio temporaneo dei dati nel caso di problemi di rete. Questa soluzione garantisce al cliente un sistema di salvataggio di dati consistente, il quale evita la perdita accidentale dei dati.

Oltre ad un sistema che mantenga la coerenza dei dati, è realizzata una trasmissione delle informazioni con un sistema ad eventi, anziché un costante aggiornamento nel tempo dei dati: sono trasmessi dati solo nel caso in cui presentino scostamenti rilevanti rispetto al campione precedente, beneficiando di un minor spreco delle risorse sia a livello di trasmissione che di storicizzazione dei dati.

Rispetto a quanto detto precedentemente, SIRIUS mette in campo anche la consueta storicizzazione dei dati ad intervalli regolari: l'intervallo di tempo che è garantito per questa soluzione è di 10 minuti. Il recupero di informazioni dai dispositivi sul campo non si limita al solo valore dei parametri; infatti, sono registrati in real-time anche gli errori di stato o il malfunzionamento del sistema.

2.4.3 Real time operation

Al fine di limitare l'accesso alle funzionalità, è implementato un sistema di autenticazione che permette di configurare, per ogni utente, l'insieme di dati che può visionare e il controllo dei dispositivi di competenza. L'autenticazione viene verificata con username e password e permette di limitare il campo di azione di ogni utente.

Il sistema offre essenzialmente due interfacce per il monitoring dei dati: la *Electrical Operation* e la *Power Plant Operation*. La *Electrical Operation* permette di avere una rappresentazione grafica dell'impianto elettrico, in cui è possibile verificare dati della rete di cablaggio dei dispositivi; l'interfaccia permette di monitorare e controllare i singoli dispositivi dell'impianto, oltre ad offrire le misure aggregate di potenza attiva e i dati meteorologici raccolti; attraverso questa funzionalità è possibile verificare lo stato degli allarmi. La visualizzazione *Power plant operation* focalizza la sua attenzione ad un livello più generale, in cui è possibile avere un quadro generale sullo stato dei componenti dell'impianto, oltre a poter gestire comandi generici destinati a gruppi di componenti.

Alcune funzionalità di controllo permettono di gestire i parametri dell'impianto direttamente tramite SMS e e-mail; è opportuno puntualizzare che esistono metodi più efficienti per controllare il sistema direttamente da smartphone, ad esempio applicazioni dedicate; evidentemente questa funzionalità era maggiormente interessante una decina di anni fa, quando i telefoni non avevano le possibilità tecnologiche degli smartphone attuali.

2.4.4 Data Analysis

Il modulo di Data Analysis è dedicato all'analisi dei dati che sono immagazzinati dal sistema: lo scopo di questo modulo è di offrire potenti strumenti per poter rappresentare i dati a disposizione, avvalendosi di opportuni grafici. Il grafico utilizzato per la rappresentazione dei dati dipende da due fattori: la tipologia di dato disponibile e il tipo di analisi da condurre sui dati. Le possibili analisi che il prodotto è in grado di supportare sono:

- *Trends*. La gestione dei dati temporali si rende maggiormente monitorabile con un grafico a linee, in cui sulle ascisse si ha la dimensione temporale, mentre sulle ordinate è riportato il valore associato alla variabile analizzata.
- *Status manager*. La visualizzazione di un grafico spettrale permette di effettuare rapidi confronti, mettendo in corrispondenza giorno su giorno il dato in ogni istante. Il valore di ogni dato viene sostituito da una colorazione scelta a discrezione dell'utente.
- *Downtime manager*. Questa funzionalità si avvale di un grafico spettrale per la visualizzazione dei dati; la differenza rispetto a Status Manager sta nella tipologia di dato analizzato: infatti, il grafico rappresenta gli errori verificati dal sistema, permettendo di filtrare i dati per impianto, per singolo device, per causa di errore o per tipo di errore.
- *Event List*. Attraverso il supporto di una griglia, è possibile verificare gli eventi registrati dal sistema; i dati visualizzati possono essere filtrati dall'utente per intervallo temporale e dispositivo da analizzare.

Tutte le informazioni che il sistema storicizza sono esportabili nei formati più comuni quali Excel, PDF o eventuali formati rappresentanti i grafici come immagini (JPG, SVG...).

La parte di analisi dei dati non termina con queste funzionalità. Infatti, nella prossima sezione è presentato un altro insieme di strumenti che appartengono sempre alla data analysis, ma che è separato in fase di presentazione del prodotto. Il motivo di questa separazione è dovuto alla differente tipologia di analisi delle prossime funzionalità; infatti, gli strumenti che verranno presentati si focalizzano per supportare il controllo di dati nell'ambito della produzione energetica condotta con impianti eolici e fotovoltaici.

2.4.5 Performance and KPI

Il modulo *Performance and KPI* propone funzionalità studiate appositamente per sistemi fotovoltaici ed eolici, perché prendono in considerazione informazioni che, per altre tipologie di impianti, non danno alcun valore aggiunto all'analisi. Infatti, i dati meteorologici raccolti dal sistema, sono utili per calcolare l'energia teorica prodotta da un impianto che sfrutta luce solare o vento. Le funzionalità disponibili in questo modulo sono:

- **Curve di potenza.** *La curva di potenza di una macchina eolica mostra il rapporto tra la velocità del vento e la potenza elettrica istantanea erogata dal generatore [NEX-EEE].* Lo scopo di questa funzionalità è di valutare la possibile curva di potenza del proprio impianto, tenendo in considerazione nel calcolo di vari fattori, come la capacità della pala eolica e i parametri atmosferici di riferimento (densità dell'aria, forza del vento). Lo strumento permette quindi di comparare i dati reali di generazione di corrente con quelli teorici: l'utilità di questo grafico permette di stabilire la situazione sullo stato della propria produzione in base alle aspettative, con la possibilità di individuare eventuali problemi tecnici nel funzionamento dell'impianto;
- **Target / budget analysis.** Soluzioni dedicate alla quantificazione delle performance del sistema.
- **Energia prodotta ed energia potenzialmente persa.** Questi KPI valorizzano quantitativamente il guadagno effettivo e quello mancato derivante da malfunzionamenti del sistema.

Come detto precedentemente, l'azienda si è concentrata nello sviluppo di funzionalità dedicate a sistemi di produzione di energie rinnovabili: se da un lato questi sforzi portano l'azienda ad avere un prodotto fortemente specializzato in questo ambito, dall'altro ne limita fortemente l'efficacia in altre realtà aziendali.

2.4.6 Reporting

La parte di reportistica ricalca in maniera fedele le visualizzazioni messe in campo in fase di monitoring e analisi dei dati, con la possibilità di eseguire esportazioni di dati su periodi di tempo giornalieri, settimanali o mensili. Tra i dati che possono essere compresi nelle esportazioni sono presi in considerazione gli eventi registrati dal sistema o lo stato dei singoli device. Altre esportazioni permettono report generici su tutti gli eventi registrati dall'impianto, come l'energia prodotta e i guasti registrati dal sistema.

2.5 ABB

ABB è una multinazionale svizzera con sede a Zurigo [ABB-AHP]: l'azienda opera in maniera distribuita in tutto il mondo e il suo business è composto da una quantità tale di prodotti, da rendere necessaria un'ulteriore divisione dell'organizzazione. La compagnia risulta essere una tra le più grandi aziende al mondo operanti nell'ingegneria elettronica. Una recente riorganizzazione della società ha permesso di suddividere l'organizzazione aziendale in 5 settori di business costituiti dalle seguenti divisioni:

- **Power Products:** produce componenti per la trasmissione e la distribuzione di energia elettrica quali trasformatori, quadri, interruttori, cavi ed apparecchiature complementari.
- **Power Systems:** offre sistemi e servizi per reti di trasmissione e distribuzione di energia e per centrali elettriche. In questo settore offre tutta la strumentazione per il controllo e la gestione di impianti elettrici. Questa divisione è ulteriormente partizionata in *Grid Systems, Substations, Network Management and Power Generation*.
- **Discrete Automation and Motion:** fornisce prodotti, soluzioni e servizi per la produttività industriale e l'efficienza energetica come motori, generatori, controllori, logica programmabile e robots industriali. Negli ultimi anni ha spostato l'attenzione anche sulle energie rinnovabili, fornendo componentistica per pale eoliche e sistemi fotovoltaici.
- **Low Voltage Products:** fornisce prodotti e sistemi di controllo per gli impianti elettrici, accessori per il cablaggio di reti, sistemi di ventilazione e sistemi di sicurezza;
- **Process Automation:** soluzioni per la strumentazione, l'automazione e l'ottimizzazione dei processi industriali.

Chiaramente all'interno di questa sezione verrà esaminata con maggiore attenzione la suddivisione dedicata alla power system, la quale si occupa di fornire quanto necessario per supportare un sistema di tipo SCADA/EMS. Infatti, il prodotto Network Manager risulta fortemente modulabile, tanto che permette la gestione di sistemi SCADA/EMS, SCADA/GSM e SCADA/DMS. Nella prossima sezione si evidenzierà il modulo riferito al sistema SCADA/EMS, considerando comunque tutta l'infrastruttura generale.

2.5.1 Network manager

Network manager è la soluzione SCADA/EMS/GSM/DMS per sistemi indipendenti operanti su larga scala, per aziende di distribuzione e trasmissione e per strutture della pubblica amministrazione. Lo stesso software può anche essere utilizzato per la gestione di reti per impianti di gas, ma anche per fornire uno strumento di supervisione alla gestione industriale.

L'architettura presenta una soluzione modulabile, che prende in considerazione la possibilità di offrire i servizi di SCADA/EMS WAMS (WAMS, Wide Area Monitoring System, gestione per il monitoring di sistemi distribuiti su vasta scala), SCADA/GSM (Generation Management System, gestione dedicati a di impianti di generazione energia), SCADA/DMS (Distribution Management System, sistemi per la distribuzione di beni e servizi) e MMS (market management system, sistemi per la gestione del marketing). Fornitori esterni offrono soluzioni per l'organizzazione dell'helpdesk, la gestione degli aggiornamenti software, la formazione del personale e altri servizi dedicati al supporto della clientela.

2.5.2 Architettura

Come precedentemente introdotto, l'architettura di questo sistema permette di coprire le esigenze di molteplici soluzioni, soprattutto per quanto riguarda le specializzazioni dei sistemi SCADA. Questo tipo di soluzione ha una coerente modulazione del sistema: la tipologia del sistema da installare si esegue in base ai moduli compresi nel tipo di installazione selezionata.

Una riflessione che è possibile fare riguarda la metodologia con cui questo sistema è sviluppato; si pone l'attenzione su come ogni soluzione descritta condivide con le altre gran parte dell'architettura, differenziandosi solo per una serie di funzionalità specifiche che implementa. Infatti, l'infrastruttura di Network manager è stata progettata in modo che tutti i moduli specifici del prodotto si collochino sopra un elemento fondamentale del sistema, l'infrastruttura per l'interfacciamento con i sistemi esterni.

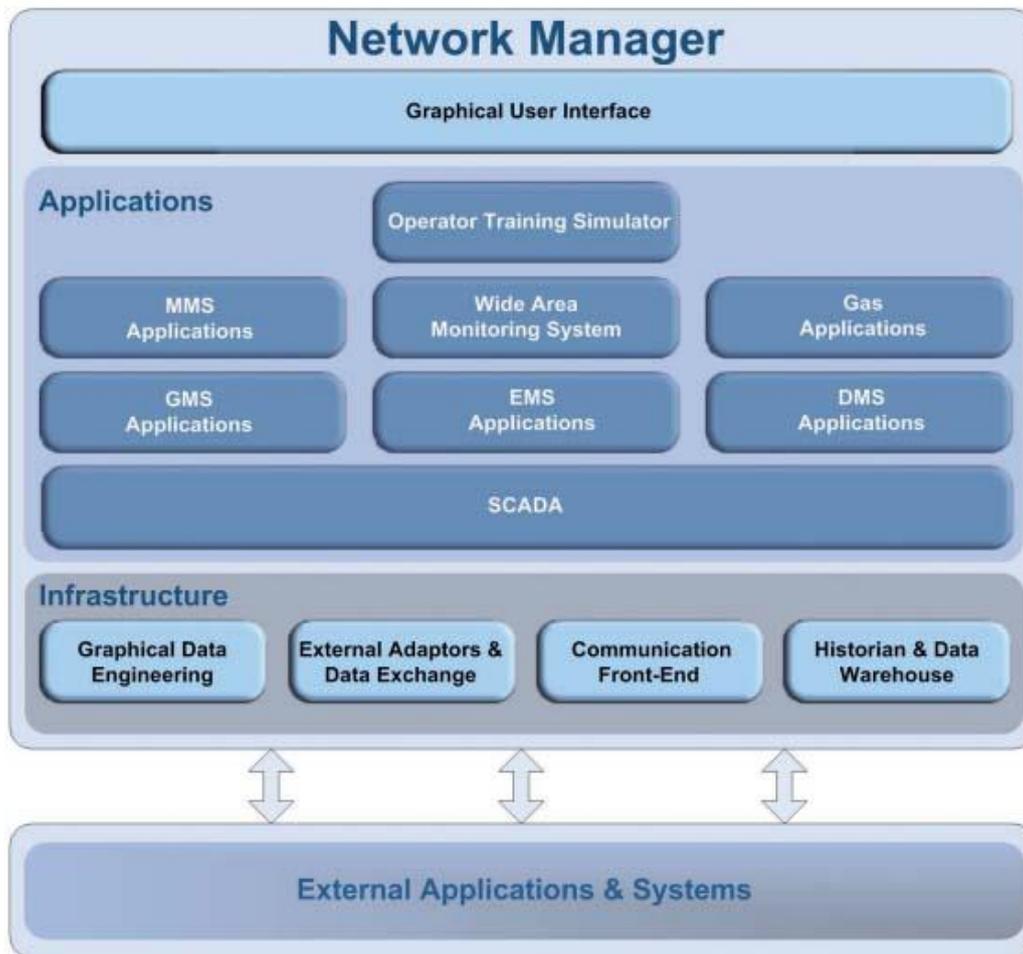


Figura 7 - Architettura del prodotto Network Manager di ABB

Come visibile in *figura 7*, l'architettura della soluzione Network Manager è composta da tre parti fondamentali: *Infrastructure*, *Applications* e *Graphical User Interface*. *Infrastructure* permette di interfacciare i moduli interni al prodotto con tutti i sistemi esterni, premettendo di isolare le problematiche derivanti dallo scambio dati con altre tecnologie. *Applications* contiene diversi moduli che hanno in comune la possibilità di dialogare con i sistemi esterni attraverso uno strato denominato SCADA, che mette a disposizione dei moduli, un set di funzionalità utili alla supervisione e al controllo dei dati. *Graphical User Interface* permette all'operatore di dialogare con il sistema, disponendo delle funzionalità offerte dai moduli installati.

2.5.3 Infrastructure

Questa parte del sistema è composta da un insieme di funzionalità diverse tra loro, ma che si accomunano in un obiettivo definito: risolvere il divario tecnologico tra la eterogeneità dei sistemi esterni e il cuore del sistema. Tra le problematiche principali da affrontare, troviamo la necessità di gestire il maggior numero di protocolli per interfacciarsi con i sistemi esterni. Il risultato di questo livello è destinato ad essere a disposizione agli strati successivi: infatti, questo livello permette di accedere ai dati dei componenti esterni indipendentemente dal divario tecnologico.

Il modulo *External adaptors & data exchange* svolge il compito di inter-scambiare dati con l'esterno. Le tipologie di sorgenti dati da cui è possibile estrarre informazioni è corposo: si passa da *file flat* a connessioni a data base come SQL*Net, ma anche da CORBA (OMG) a porte COM (Microsoft). Sono fornite infrastrutture dedicate allo scambio di messaggi con pattern *publish-subscribe*. Sono definiti standard dedicati allo scambio di dati con centri di controllo esterni. ABB si pone anche l'obiettivo di partecipare ad iniziative di standardizzazione per quanto riguarda lo scambio di dati: questo dimostra come, in maniera attiva, l'azienda rimane sempre all'avanguardia in queste tecnologie.

Diverso è il compito del secondo modulo preso in considerazione. Infatti, quello denominato *Graphical data engineering* fornisce un ambiente per l'ingegnerizzazione dei dati, comprendendo la creazione del database e di tutti i layout che rappresentano un qualsiasi oggetto elettronico della rete. Funzionalità di definizione del layout sono fornite per rendere più veloce la configurazione del sistema, che rimane estremamente complicata. Molti standard di modellazione di oggetti fisici o astratti sono utilizzati all'interno della soluzione, oltre all'utilizzo di interfacce per l'importazione o l'esportazione modelli attraverso *Common Information Model*. Sono offerte funzionalità grafiche per la configurazione dell'ambiente e un sistema di controllo di dati. Un'ulteriore funzionalità del sistema permette la creazione personalizzata di ogni tipologia di display: un'interfaccia easy-to-use con importazione di immagini e modellazione personalizzata di icone permette di costruire maschere dedicate.

Il terzo modulo preso in considerazione è dedicato alla comunicazione con i dispositivi sul campo. A questo proposito, il *Communication front-end*, specifica la tecnologia di comunicazione con RTUs,

permettendo di definire i tempi di raccolta dati diminuendo il numero di interrogazioni grazie ad operazioni atomiche. La gestione di questo modulo è garantita su reti WAN o LAN e protocollo TCP/IP. L'obiettivo principale è sempre quello di convertire il protocollo legato a RTU in uno scambio dati con un protocollo indipendente dal formato, in maniera tale da uniformare la gestione dei dati con il server.

L'ultimo modulo di questa parte è dedicato alla gestione data warehouse del prodotto: infatti il modulo *Historian* permette la storicizzazione dei dati, garantendo rimedi al pericolo di perdita di informazioni. Il numero dei dati che è possibile archiviare è a discrezione dell'utente e permette di disporre di una onerosa mole di dati. Alcuni esempi di dati disponibili sono lo stato del punto di controllo, il valore degli accumulatori, i valori analogici, i calcoli, le sequenze di eventi, gli allarmi e i risultati delle applicazioni. Ci si avvale dell'utilizzo di strumenti esterni per migliorare funzionalità particolarmente complesse: ad esempio, l'uso di Matlab permette di fornire calcoli avanzati e nuove formule; ogni nuova serie di dati può essere sottoposta a tecniche di *data mining*. Sempre in questo modulo si trovano le funzionalità di reportistica e definizione dei trend.

2.5.4 Applications

Dopo aver descritto come si comporta il livello infrastrutturale del sistema, si passa alla descrizione delle applicazioni del sistema: questa sezione include i moduli software più importanti per il sistema, che sono rappresentati principalmente dal modulo SCADA e EMS, ma che si compongono di strumenti come *WADS* e *operator training simulator*, riportati successivamente.

Il primo modulo che viene presentato in questo livello è legato alla parte logica dei sistemi **SCADA**: infatti, oltre ad avere un portale di autenticazione che permette la validazione degli utenti, esso pone alla base della sua implementazione un livello di affidabilità dati, che permette di gestire algoritmi per garantire la sicurezza e l'integrabilità delle comunicazioni. In questo modulo, sono introdotte tutte le funzionalità di visualizzazione grafici e gestione di recupero dei dati in caso di perdita.

Il secondo modulo **EMS applications** riguarda il sistema di gestione energia e si scompone in molte funzionalità, ognuna indipendente dall'altra. L'intero modulo è stato quindi suddiviso in funzioni per

il monitoring della rete, per la verifica della sicurezza, per il supporto alle decisioni e per il miglioramento delle condizioni di lavoro della rete. La prima funzionalità descritta da questo modulo è quella relativa al controllo della rete, la quale si occupa di valutare in ogni momento le condizioni operative: tra gli oggetti maggiormente monitorati si hanno il bus di scambio messaggi, lo stato del processo di recupero dati e una fotografia sulla telemetria del sistema. La funzionalità *Security Assessment* verifica le condizioni di sicurezza della rete attraverso una vasta gamma di indicatori, dedicati a fotografare lo stato della rete: tra i vari parametri, sono controllati il flusso della potenza, l'analisi dei corto circuiti e la verifica delle interruzioni. Oltre a questi strumenti di valutazione è disponibile un supporto per sostenere il processo decisionale: questo strumento assiste l'operatore nel valutare numericamente quanto possano incidere determinate scelte sul sistema, soprattutto negli ambiti della sicurezza, sullo stato della rete e sulle condizioni della base di dati. L'ultima funzionalità, denominata *Operation Enhancement*, assiste l'operatore negli spunti di miglioramento delle condizioni della rete, portando un contributo significativo per effettuare analisi di controllo del flusso della potenza e sulla stabilità del flusso di voltaggio del sistema.

Il terzo modulo del livello delle applicazioni riguarda valutazioni sul monitoraggio di sistemi su larga scala. Infatti, il ***Wide area monitoring system***, permette di fare valutazioni sul sistema di trasmissione, soprattutto per quanto riguarda grossi impianti, in cui possono incorrere problemi legati ai disturbi del segnale. Sono fornite informazioni real-time con processi di verifica dello stato dell'intero sistema. L'utilizzo di sensori GPS permette di conoscere il livello di sparsità dei componenti della rete, così da effettuare calcoli consapevoli sulle prestazioni della rete: questa funzionalità permette di proporre soluzioni topologiche diverse, con la definizione delle modifiche da introdurre nel caso di nuovi nodi inseriti nel sistema.

Il quarto modulo è ***Operator training simulator*** che si pone l'obiettivo di addestrare lo staff necessario nella complessa costruzione di reti ad alto voltaggio, fornendo esperienza e conoscenza. Permette anche di costruire simulazioni capaci di valutare la qualità della rete.

2.5.5 Graphical user interface

Il modulo definito all'interfaccia del sistema permette di essere altamente scalabile, fortemente configurabile anche in casi di grossi sistemi geograficamente distribuiti. Tra le molteplici funzionalità messe in campo da questo modulo è possibile riconoscere:

- Gestione di eventi e allarmi con possibilità di aggiungere o rimuovere le stazioni d'interesse;
- Un'interfaccia di comunicazione fornita con l'introduzione di allegati di qualsiasi formato dati;
- Uno strumento grafico di definizione di query;
- Funzionalità di drag and drop;
- Personalizzazione disponibile per ogni operatore;
- Configurazione delle interfacce d'interesse;
- Gestione di interfacce multi monitor.

Le informazioni fornite sulla parte relativa all'interfaccia non sono dettagliate come abbiamo potuto constatare con altre soluzioni, ma certamente, è possibile dedurre che la soluzione offerta da ABB eccelle sicuramente nell'aspetto legato all'ingegnerizzazione di questi sistemi.

2.6 On.Energy management system

All'interno di questa sezione si propone l'obiettivo di dare un quadro generale sulle specifiche del prodotto On.Energy sia da un punto di vista funzionale sulla base di quanto detto sui sistemi SCADA/EMS, sia dal punto di vista comparativo, scoprendo le funzionalità che questo strumento mette a disposizione; lo scopo finale è quello di riuscire a comparare, in maniera intuitiva, pregi e difetti delle tre soluzioni analizzate.

2.6.1 Aspetti generali

On.Energy Management System è il software sviluppato da *Onit Group s.r.l.* [ONI-OHP] dedicato alla supervisione, al controllo dei consumi energetici e all'acquisizione dei dati; questo prodotto è suddiviso principalmente in base alla tipologia di sistema su cui è impiegato: le possibilità d'installazione possono comprendere attività come realtà industriali, gestione di condomini o controllo di uffici.

Il software ha ricevuto un buon consenso in ambito internazionale. Le caratteristiche maggiormente apprezzate sono la gestione multilingua e le forti personalizzazioni applicabili. Come per le altre soluzioni analizzate, *On.Energy* gestisce l'accesso alle funzionalità attraverso uno step di autenticazione, che permette ad ogni utente di accedere a dati a cui è stato appositamente abilitato.

Il sistema offre 2 tipi di piattaforme disponibili per interagire con l'applicativo: la prima è la piattaforma *Silverlight*, fruibile direttamente dal proprio pc tramite browser, la seconda è fornita con *HTML5*, il quale permette anche ai possessori di piattaforme mobili di poter utilizzare il programma.

All'interno di questa sezione verranno principalmente analizzate le caratteristiche software del prodotto, proponendo dal punto di vista dei protocolli software supportati, quelli che presentano le principali applicazioni nel mondo dell'elettrotecnica. Come tutti i software SCADA/EMS, si compone essenzialmente in 3 caratteristiche fondamentali:

- *Acquisizione dei dati.* Il software presenta caratteristiche legate al recupero di valori direttamente da apposito hardware; come detto precedentemente, l'hardware in questione dipende principalmente da aziende con cui sono state legate particolari partnership. Inoltre, altre soluzioni di integrazione dati permettono di poter reperire informazioni anche da sistemi esterni, che non prevedono apposito hardware da interrogare.
- *Monitoring dei dati.* Una delle caratteristiche maggiormente evolute riguarda la visualizzazione dei dati storicizzati all'interno del sistema: la forte personalizzazione delle

interfacce e il numero di cruscotti utilizzati permette una ricca personalizzazione del software.

- *Controllo delle operazioni.* Lo stesso hardware che permette la lettura dei dati offre anche la possibilità di poterli impostare: la possibilità di modificare il valore dei dati è fornita anch'essa con grafiche ricche di iconografie.

2.6.2 Architettura

L'architettura proposta da Onit rispecchia quanto teoricamente analizzato nella parte di spiegazione dei sistemi SCADA di questa tesi. Infatti, l'organizzazione del sistema si compone di un *server di front/end* per l'interscambio dei dati, mentre il modulo principale del sistema svolge sia le attività specifiche del *SCADA Server*, sia quelle del *DATA Server*.



Figura 8 - Rappresentazione del sistema On.Energy di Onit Group

La *figura 8* permette di entrare nello specifico dell'infrastruttura; è possibile distinguere 3 parti del sistema: la prima è l'*accentratore di dati* chiamato *Data Retrieval*, la seconda rappresenta il fulcro di On.Energy e la terza è la *Web application*, che dà la possibilità all'utente di interfacciarsi con i propri dispositivi.

Nella parte relativa al *Data Retrieval* fanno parte tutte le funzionalità che riguardano l'integrazione dei dati, siano essi direttamente acquisiti dai dispositivi sul campo, oppure forniti con altri metodi supportati dal sistema. La parte di On.Energy permette di gestire la doppia funzionalità di sistema *Real-time monitoring* e *Data Analysis*: il software risulta essere molto più propenso alla parte di analisi del dato, che vanta un insieme di importanti funzionalità dedicate. La parte relativa alla consultazione dell'utente è accessibile con qualsiasi dispositivo; l'interfaccia resa disponibile viene

direttamente scaricata sul proprio device, in cui esegue il codice necessario alla visualizzazione dei dati.

Per quanto riguarda le installazioni offerte per questo tipo di soluzione, è ovvio che, per entrambi i core funzionali del sistema, è possibile installarli in *housing* dal cliente, oppure in *hosting* sull'infrastruttura Onit. I tre servizi d'installazione disponibili sono *cloud*, *semi-cloud* e *on-site*: la prima indica che l'onere dell'infrastruttura è interamente preso in carico da Onit; la seconda specifica che il Data Retrieval è tenuto internamente dal cliente mentre la parte di On.Energy resta in carico ad Onit; la terza opzione indica che tutta l'infrastruttura risiede direttamente nella rete del cliente.

Rispetto alle soluzioni viste in precedenza, Onit sembra essere la soluzione meno indipendente dal punto di vista infrastrutturale, visto che non detiene direttamente una soluzione propria anche dal punto di vista dell'hardware; però almeno in questa fase, grazie alle partnership legate esternamente, anche questa soluzione appare funzionalmente completa come le altre. Per quanto riguarda SIRIUS, ci sono forti somiglianze con il prodotto On.Energy, anche se è stato fatto un importante lavoro nell'ambito del recupero dei dati in caso di malfunzionamenti nello scambio dati. ABB, dal canto suo, offre un potente strumento di configurazione in una suite completamente compatibile con il proprio sistema: questa funzionalità, anche se non rivolta direttamente ad un utente medio, dà garanzie di veloce installazione del sistema anche su soluzioni riguardanti un numero elevato di impianti geograficamente sparsi. Al fine di valorizzare il prodotto di ABB, è possibile dire che la forte modularizzazione del sistema permette di garantire un set di soluzioni maggiore; infatti, la base del prodotto viene rivenduta come soluzione per altri settori, potenziando notevolmente il suo business.

2.6.3 Data acquisition

Per quanto riguarda l'acquisizione dei dati, il sistema sia la lettura dei dati direttamente da dispositivi di campo, sia l'integrazione ad hoc dei dati. Le principali modalità che permettono di acquisire dati sono:

- Lettura da dispositivi hardware. Tra i principali protocolli utilizzati possiamo distinguere:
 - *Modbus*. Questo protocollo permette di comunicare con i propri *conduttori logici programmabili* (per abbreviazione PLC). È utilizzato come uno standard *de facto* nella comunicazione di tipo industriale e risulta uno dei protocolli più utilizzati al mondo tra i dispositivi elettronici industriali. Modbus è spesso usato per connettere un computer supervisore con un'*unità terminale remota* (in inglese *Remote terminal unit*, abbreviato RTU) nel controllo di supervisione e sistemi di acquisizione dati. In sostanza, funge da accentratore per ulteriori dispositivi impiegati nella raccolta dati.
 - *S7*, una tipologia di PLC appositamente sviluppata da Siemens, anch'essa ampiamente utilizzata in ambito industriale.
 - *LonWorks*, una piattaforma di collegamento utilizzata soprattutto nel campo del telecontrollo di sistemi di illuminazione, sistemi di ventilazione, sistemi di sorveglianza, ecc.
- Lettura di dati tramite importazione da file. I formati supportati per queste integrazioni sono *csv*, *xls* e *xlsx*.
- Integrazione dei dati attraverso sistemi di comunicazione, come l'integrazione di dati attraverso mail o tramite scambio di file in formato XML.

Il controllo dei dati risulta estremamente legato ai dispositivi impiegati all'interno della propria installazione: questa funzionalità non presenta essenziali differenze sulle tecnologie hardware coinvolte nella fase di acquisizione, dove troviamo appunto gli stessi protocolli supportati.

Comparandolo con gli altri software presi in esame, questa soluzione presenta un minor numero di protocolli supportati per l'interscambio dei dati: questo dettaglio risulta abbastanza comprensibile comparando la tipologia di azienda: infatti si tiene conto che sia ABB, sia SIRIUS, nascono direttamente dal campo della elettronica, dove hanno maturato delle competenze specifiche nel settore dei dispositivi elettronici. La soluzione di ABB offre un potente strumento di integrazione configurabile per la lettura su altri sistemi. SIRIUS non sembra avere funzionalità di integrazione dati attraverso file o con applicativi esterni.

2.6.4 Monitoring & Real-time operation

La soluzione di Onit si avvale di strumenti dal forte impatto grafico che permettono all'utente di svolgere le attività di supervisione dei dati. Nella *figura 9* è possibile vedere come Onit ha interpretato la funzionalità di monitoring dei dati, proponendo elementi grafici interessanti: l'utente ha la possibilità di selezionare il tipo di dato da visualizzare e lo strumento grafico con cui mostrarlo.

Oltre ad una visualizzazione grafica dei dati, On.Energy permette di caricare immagini che riportano il layout del sistema osservato: l'utente può personalizzare a propria discrezione queste mappe, visto che può aggiungere sia elementi che visualizzano lo stato dei dispositivi, sia strumenti che modificano i parametri dei componenti dell'impianto.



Figura 9 - Widget per il monitoring in tempo reale dei dati

Altri elementi che appartengono al modulo di supervisione real-time sono gli allarmi: la loro funzione è di avvisare tempestivamente l'utente quando i valori di una variabile superano l'intervallo consentito. Gli allarmi sono visualizzati nella schermata principale dell'interfaccia, permettendo all'utente una rapida identificazione delle anomalie. In aggiunta, queste segnalazioni possono essere notificate attraverso SMS ed e-mail. Gli sfioramenti delle soglie di ogni variabile sono registrati dal sistema, che poi renderli disponibili nella creazione di report periodici.

Comparando le funzionalità di On.Energy con gli altri prodotti analizzati, notiamo che le funzionalità di supervisione e controllo sono sostanzialmente le stesse; la gestione degli allarmi presenta caratteristiche simili, anche se la segnalazione degli errori è proposta solo attraverso le interfacce di supervisione, e non attraverso SMS o e-mail come per la soluzione On.Energy. Le maggiori differenze sono visibili nell'ambiente di configurazione del sistema: infatti, il prodotto di ABB offre

uno strumento più dedicato ad una figura specifica, rispetto ad un utente di supervisione dati. SIRIUS si distingue per la possibilità di controllare i parametri del sistema tramite SMS ed e-mail.

2.6.5 Data Analysis

Lo scopo principale della data analysis è quello di fornire strumenti utili alla visualizzazione dei dati storicizzati: in questa fase è importante proporre all'utente una vasta scelta del grafico da utilizzare, oltre che permettere di personalizzare il modo in cui sono visualizzati i dati.

On.Energy offre una serie di grafici per eseguire analisi approfondite sulle serie temporali: infatti, sia i grafici a linee, che i grafici spettrali visualizzano in modo differente la stessa informazione. I grafici a linee forniti sono particolarmente flessibili: permettono di inserire nella stessa analisi più serie di dati differenti, offrendo uno strumento di comparazione altamente configurabile.

L'analisi dei valori può essere pressochè inutile in molti casi reali; infatti, i dispositivi sul campo possono registrare informazioni non significative; un esempio sono i contatori che misurano l'energia elettrica erogata totale: il dato risulta interessante solo se viene considerata la differenza tra due letture consecutive, avendo così calcolato il consumo di elettricità in un intervallo di tempo. On.Energy dà la possibilità di calcolare qualsiasi funzione tra più variabili, così da costruire nuove serie temporali che presentano dati maggiormente interessanti. Quindi, per una buona analisi dei dati oltre ad uno strumento potente, è necessario poter disporre di dati significativi.

On.Energy utilizza un sistema di generazione di etichette temporali, create sulla base di schedulazioni. Queste informazioni possono essere utilizzate per aggregare i dati in corrispondenza dell'intervallo di tempo in cui ogni etichetta è valida. Alcuni controlli permettono di migliorare l'analisi dei dati aggiungendo aggregazioni per etichetta. Un esempio reale di utilizzo può essere la creazione di un grafico a torta che raggruppa i dati sulla base del valore delle etichette.

La *figura 10* mostra i risultati dell'elaborazione di alcuni componenti: la visualizzazione del grafico spettrale è molto utile nell'analisi di valori variabili, come può essere il valore della temperatura di un oggetto.

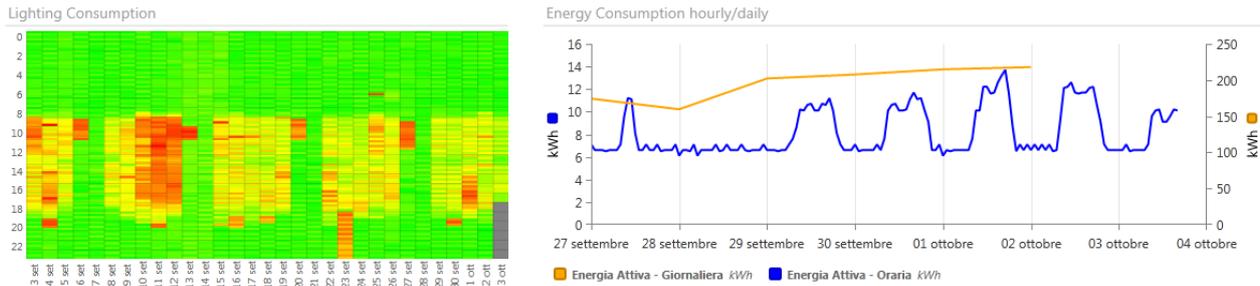


Figura 10 - Strumenti per l'analisi dei dati

Comparando le differenze nella storicizzazione dei dati, SIRIUS permette di registrare dati ad intervalli di 10 minuti, con la possibilità di recuperare i dati solamente quando cambiano in maniera significativa. Mentre per la soluzione di ABB, la gestione della storicizzazione dei dati è garantita da una piattaforma di data warehouse sviluppata esternamente. Onit, invece offre un intervallo di storicizzazione dei dati ogni 15 minuti tra due rilevazioni consecutive e un performante sistema di duplicazione dei dati su file.

In rapporto alle altre soluzioni, Onit sembra fornire un set di cruscotti particolarmente ricco, oltre alla possibilità di calcolo di funzioni di dati e al sistema di generazione etichette. La soluzione SIRIUS è fortemente specializzata nell'analisi dei dati per impianti di produzione di energie rinnovabili: la loro soluzione include funzionalità specifiche, come il calcolo della curva di potenza. La soluzione ABB non riporta molte informazioni sull'analisi dei dati, anche se la gestione di queste informazioni è demandata ad un ambiente di data warehouse dedicato.

2.6.6 Report

La generazione dei report è una funzionalità che permette di riepilogare informazioni che il sistema propone già in analisi dei dati. In On.Energy, i resoconti mensili mettono in corrispondenza le informazioni dei trend, raggruppandole in base alle etichette create e riportando la lista degli allarmi generati.

Grazie al supporto di altre, On.Energy riesce a proporre all'utente funzionalità come il calcolo della bolletta elettrica, con risultati praticamente identici rispetto a quanto emerge dai resoconti mensili.

Per quanto riguarda i report, possiamo dire che queste funzionalità rimangono poco evidenziate all'interno dei vari applicativi a causa di vari fattori: ad esempio, le potenti analisi del dato mettono in secondo piano la creazione di report, riducendone l'importanza. Inoltre, i report riutilizzano le informazioni già presenti nel sistema; infatti, la creazione di nuova reportistica è generalmente richiesta su misura dai cliente.

Capitolo 3

L'applicativo On.Energy Management System

L'obiettivo di questo capitolo è presentare in maniera più dettagliata l'applicativo *On.Energy management system*: si evidenzieranno meglio i concetti alla base di questo software e come è modellata la logica di questo sistema. Successivamente, saranno analizzate le modalità di analisi dei dati della soluzione, con lo scopo di capire le potenzialità offerte al cliente; a questo punto si delineeranno le lacune funzionali che affliggono il sistema. Una volta fatte emergere le nuove necessità nel campo dell'analisi dei dati, sarà eseguita l'analisi dei nuovi requisiti, che sono il punto di partenza su cui si dovrà basare la progettazione del nuovo strumento di analisi dei dati. L'ultimo argomento del capitolo tratterà l'analisi del problema.

3.1 Aspetti generali

In questa sezione sono analizzati tutti i concetti presenti nel software On.Energy, i quali hanno il merito di svolgere un ruolo chiave nella composizione delle funzionalità offerte da questi sistemi. In sostanza, sarà analizzato il dominio dei dati di questo software, al fine di comprendere l'importanza di ogni concetto, anche in relazione a ciò che rappresenta nella realtà. I concetti fondamentali che fanno parte della soluzione sono la *struttura della gerarchia*, gli *stream*, gli *indicatori*, le *categorie*, le *etichette*, le *tariffe* e le *fasce tariffarie*.

Il primo concetto preso in considerazione è la *struttura gerarchica*: questa modellazione si pone come basilare nella strutturazione dell'intero software; tale organizzazione permette di avvicinare

l'utente ad un concetto di divisione gerarchica di diverse parti di un sistema; infatti, è possibile rappresentare sia una divisione spaziale di un territorio, oppure un'organizzazione aziendale del personale: il modo di organizzare la gerarchia dipende da quale dato si ha interesse a fare emergere. In concreto, questa modellazione si trasforma in diversi esempi: la strutturazione di una azienda in stabilimenti-reparti-linee-zone, la divisione di un insieme di uffici, la scomposizione di un condominio in piani e appartamenti. La struttura della gerarchia è rappresentata da un albero che collega nodi; ogni singolo nodo della gerarchia corrisponde ad un elemento nella realtà; come in ogni gerarchia, i concetti di dipendenza tra nodi permettono di esprimere significati d'appartenenza e di poter associare ad ogni livello della struttura un determinato ruolo. Ai nodi della gerarchia è possibile associare le altre informazioni del sistema: stream, indicatori, categorie e tariffe; ognuno di questi concetti ha una specifica visibilità all'interno del sistema determinata dal nodo a cui è associato.

Tra gli elementi che possono essere associati ad un nodo, il più importante è lo *stream*: esso è il frutto dell'interscambio delle informazioni con gli accentratori o direttamente con i dispositivi di rilevazione. Semplificando tale concetto, ogni stream rappresenta i valori che ha assunto nel tempo una variabile o un dispositivo installato sull'impianto; ogni serie temporale di dati è storicizzata ad intervalli regolari sul sistema. In aggiunta, lo stream è associato ad un insieme di informazioni che aiutano l'utente a specificarne alcuni dettagli: è possibile associare la tipologia di misura che rappresentano, scegliendo tra misure di energia, calore, frequenza, corrente elettrica, etc. Per la visualizzazione dei dati, è possibile definire l'unità di misura che verrà impostata di default, a seconda della misura selezionata. Come detto in precedenza, i valori degli stream sono informazioni che registrano il valore di un parametro del sistema in un intervallo di tempo. Ad esempio, le tipologie di dati letti in un condominio sono generalmente legate ai consumi, come quello di acqua, di energia elettrica, di gas, etc. Ci sono anche valori che rappresentano informazioni differenti, come la potenza istantanea erogata. In base alla tipologia di dato storicizzato, ha senso applicare un certo insieme di operazioni, con lo scopo di creare nuove informazioni utilizzabili dall'utente.

Approcciando questo tema da un punto di vista teorico, le misure che possono essere registrate in questi sistemi sono sostanzialmente di tre tipologie diverse:

- **Misure di flusso.** Sono quelle misure che si riferiscono ad un determinato intervallo di tempo e sono lette in maniera cumulativa. Esempi di queste misure sono i consumi effettivi di energia (acqua, luce, gas), il numero di prodotti venduti in un giorno, il numero di nati in un anno.
 - Le analisi dei dati consentono di raggruppare i valori sia per la gerarchie temporali, sia per quelle non temporali. Per esempio, ha senso sommare il consumo energetico sia visualizzando il dato aggregato mensilmente, sia controllando il consumo di energia di un'azienda aggregando i consumi degli stabilimenti che la compongono.
- **Misure di livello.** Sono misure che fotografano il dato in un istante di tempo; esempi di queste misure sono i contatori di consumo energetico (acqua, luce, gas), il voltaggio applicato in un determinato componente, il numero di prodotti in inventario, il numero di abitanti di una città.
 - Le analisi su questi dati non hanno senso se sommate per gerarchia temporale. Quindi, ha senso sommare gli abitanti di più città per calcolare il totale di una regione, ma non aggiunge informazione sommare gli abitanti di una città in mesi differenti.
- **Misure unitarie.** Sono valutate in istanti di tempo, ma rappresentano valori relativi. Esempi di queste misure sono il prezzo unitario di un prodotto, la percentuale di sconto, il cambio di una valuta, il costo dell'energia per unità di misura.
 - Non ha senso eseguire nessuna somma sulle gerarchie: ad esempio, sommare il costo della benzina non porta nessuna informazione sia eseguendolo su diversi distributori, sia raggruppandolo per differenti anni.

In alcuni casi, è possibile trasformare una misura di livello in una di flusso: per riuscire a estrarre maggiori informazioni da una misura di livello, si rende necessario eseguire delle operazioni sui dati. Infatti, l'informazione più interessante si ricava sottraendo ad una rilevazione il valore precedentemente recuperato: tale differenza registra lo scostamento del valore nell'intervallo di tempo; un esempio di come questa operazione possa creare nuova informazione è applicabile nel consumo energetico: avendo a disposizione la lettura del contatore, il modo per capire il consumo di energia di un determinato periodo è di sottrarre il valore finale con quello iniziale. Questo esempio ricalca la trasformazione di una misura di livello in una di flusso.

Per riuscire ad estrapolare informazioni indirette dagli stream, On.Energy offre la possibilità di eseguire trasformazioni sui dati delle serie temporali: il risultato di questi calcoli è utilizzato per creare nuove serie temporali; questi nuovi dati hanno la capacità di far emergere informazioni maggiormente interessanti, proprio come accade nel caso delle misure di flusso, quando si calcola la differenza tra una lettura e quella successiva. In On.Energy, gli oggetti che modellano le trasformazioni dei dati degli stream sono chiamati indicatori: come per gli stream, gli indicatori sono associati a informazioni che indicano la tipologia di misura d'appartenenza del dato storicizzato.

Per quanto riguarda la modellazione di serie temporali di valori numerici, stream e indicatori sono le uniche modalità di rappresentazione disponibili. Le entità presentate successivamente permettono di aggiungere informazioni testuali a specifici intervalli temporali; esse mirano a valorizzare ulteriormente le informazioni delle serie temporali, etichettando ogni rilevazione con le informazioni che derivano dall'incrocio tra l'istante di registrazione del dato e gli intervalli temporali che lo contengono.

La prima entità che permette di associare informazioni testuali ad intervalli temporali è la *categoria*: essa è composta da un insieme di *etichette*. La creazione delle etichette deve rispettare alcuni vincoli: per esempio, per ogni nodo, non possono comparire etichette della medesima categoria in uno stesso istante di tempo. Esempi di questo concetto sono i turni di lavoro (es. Turno 1 06.00-14.00, Turno 2 14.00-22.00, Turno 3 22.00-06.00), l'assegnazione della tipologia di giorno (feriale, festivo), la divisione del giorno in mattino, pomeriggio e sera. L'associazione di una pianificazione di etichette è strettamente legata ad un nodo: in questo modo, si permette di poter gestire pianificazioni diverse. Un esempio può essere la divisione del proprio impianto in reparti, in cui i turni di lavorazione hanno orari diversi.

Un particolare tipo di categorie sono le tariffe, che permettono di catturare informazioni aggiuntive. Esse includono *fasce tariffarie*, le quali hanno le stesse caratteristiche delle etichette: la differenza sta nell'aggiungere ad ogni fascia tariffaria informazioni di costo. Le tariffe hanno lo scopo di quantificare il costo di una serie temporale in un certo intervallo di tempo; per calcolare questo dato eseguono una serie di passaggi: per primo raggruppano i dati di ogni serie temporale in base agli intervalli delle fasce tariffarie, successivamente aggregano i dati di ogni intervallo ed infine calcolano il costo totale, eseguendo il prodotto del dato aggregato con il costo associato alla fascia tariffaria.

3.2 Data Analysis

In questa sezione sono approfonditi gli strumenti di analisi dei dati, resi disponibili dalla soluzione On.Energy. Lo scopo è far emergere i limiti di ogni tipologia di analisi, sia dal punto di vista dell'implementazione, sia da quello di visualizzazione e flessibilità nella gestione dei dati.

Il primo strumento riguarda la visualizzazione di serie temporali di dati numerici (nello specifico stream e indicatori). Il senso più ampio di questo strumento è vedere, in diverse modalità, l'evoluzione temporale dei dati. Le serie temporali possono essere analizzate con tre strumenti: il *formato tabellare*, *grafico a linee* e il *diagramma spettrale*. La modalità di visualizzazione tabellare è semplice e analitica, ma per questo scarsamente informativa: permette di consultare i dati allo stesso dettaglio di come sono registrati dal sistema, quindi senza alcun raggruppamento; l'utente può decidere di comparare più serie storiche, allineando i valori in uno stesso asse temporale. È inoltre possibile aggiungere le etichette, anch'esse recuperate sulla base del periodo temporale analizzato.



Figura 11 - Grafico a linee di due serie temporali

Il grafico a linee evidenzia l'andamento dei valori sull'asse del tempo; con questo strumento è facile vedere i trend delle serie temporali, soprattutto grazie a funzionalità pensate per facilitare la consultazione dei dati sia spostando il periodo da analizzare, sia modificando il dettaglio temporale con cui sono presentati. Come è possibile vedere dalla *figura 11*, si possono visualizzare valori di più

serie sullo stesso grafico: grazie a questa opportunità è possibile evidenziare la presenza di dipendenze tra serie diverse.

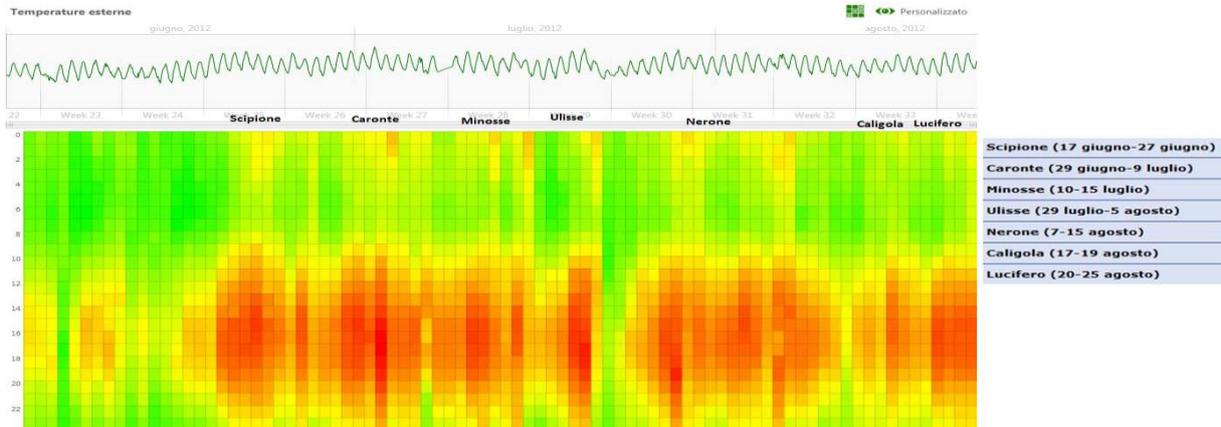


Figura 12 - I 7 anticicloni africani dell'estate 2012

Il grafico spettrale risulta molto utile nell'analisi di misure legate alla temperatura: come è possibile vedere dalla *figura 12*, lo strumento permette di riscontrare agevolmente valori critici di un impianto: la personalizzazione del tipo di colorazione dà la possibilità di creare maschere dal forte impatto grafico.

Power Distribution

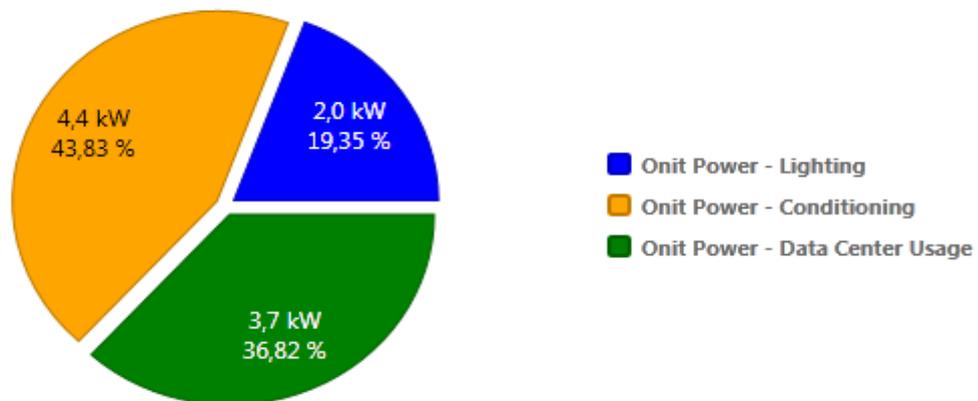


Figura 13 - Grafico a torta di dati aggregati per serie temporale d'appartenenza

Un ulteriore strumento per l'analisi dei dati è quello offerto dal *grafico a torta*: esso permette di analizzare il valore aggregato di più serie storiche. Come visibile dalla *figura 13*, un esempio è determinare quanto incide ogni sotto-sistema elettrico nel calcolo del consumo totale di energia. Anche in questo grafico, i concetti che possono essere introdotti nell'analisi si riducono a stream e indicatori: lo strumento non dispone della capacità di introdurre nel grafico altri concetti. Il periodo

di tempo da analizzare è definito a priori e non c'è possibilità di interagire con lo strumento: anche questo tool è utile nella visualizzazione sintetica dei dati ed ha un forte impatto grafico, ma non permette nessun tipo di interattività per modificare il tipo di analisi da condurre sui dati.

Un altro strumento d'analisi del dato è il *widget costi*, che permette di mettere in corrispondenza il valore delle serie temporali con le tariffe. Come è possibile vedere dalla *figura 14*, questo strumento è utile nella costificazione dei consumi elettrici; i valori di costo sono calcolati attraverso due concetti distinti del modello: le serie temporali forniscono il consumo di energia totale, reso disponibile dalla interrogazione dei corrispondenti contatori dell'impianto, mentre le fasce tariffarie attribuiscono ad ogni intervallo temporale un costo unitario dell'energia elettrica. Anche in questo caso, la flessibilità delle manipolazioni dei dati risulta inespressa.



Figura 14 - Costificazione delle serie temporali con le tariffe

In un certo senso, è possibile mettere nell'ambito dell'analisi del dato anche tutta la parte di reportistica, la quale offre uno strumento capace di fornire spunti di analisi sul comportamento dei dati. In particolare, i report permettono di fare delle aggregazioni delle serie temporali, raggruppando i dati in base alle etichette create dal sistema. Anche nel caso dei report, lo strumento risulta poco flessibile, legato ad un calcolo statico dei dati ed una visualizzazione non personalizzabile, se non con opportuni sviluppi ad hoc.

Nella prossima sezione si evidenzieranno i limiti delle funzionalità presenti all'interno del prodotto per quanto riguarda la data analysis: questa critica costruttiva, aiuterà il raggiungimento dell'obiettivo di uno strumento capace di essere flessibile, interattivo e di gestire tutti i concetti che costituiscono informazione per l'utente.

3.3 Limiti della data analysis

Lo scopo di questa sezione è di evidenziare le lacune funzionali degli strumenti d'analisi del dato della piattaforma On.Energy. La mancanza di uno strumento flessibile e potente vanifica tutta l'infrastruttura del sistema, sia per quanto riguarda il recupero dei valori, sia per la generazione delle informazioni, come etichette e fasce tariffarie: tutti gli strumenti realizzati non hanno la possibilità di analizzare il set completo d'informazioni messo a disposizione dal sistema.

Avvalendosi della sezione precedente, i maggiori deficit funzionali emersi dall'analisi degli strumenti di analisi dei dati sono:

- Impossibilità di effettuare analisi comprendendo tutte le informazioni rilevanti del sistema
 - Ogni strumento attualmente disponibile considera uno o due concetti diversi; informazioni come etichette e fasce tariffarie non compaiono contemporaneamente in nessuna analisi.
 - L'introduzione di nuovi concetti negli strumenti disponibili richiede grossi sforzi implementativi, oltre a non garantire un supporto adeguato nella visualizzazione del dato;
- Mancanza di personalizzazione nelle aggregazione dei dati degli strumenti di analisi:
 - Non è possibile scegliere in maniera arbitraria le informazioni da inserire nell'analisi: nel caso ottimale, ogni analisi può comprendere una categoria o una fascia tariffaria.
 - La granularità dei dati messi a disposizione è cablata dallo strumento: in alcuni casi si tiene in considerazione solo il valore finale di calcoli complessi, mentre in altri vengono riportati tutti i record presenti sulla base dati, costringendo il sistema a onerosi calcoli.
 - Per le maschere che aggregano dati, non è possibile definire la funzione di aggregazione da utilizzare. Tutte le funzioni di calcolo e aggregazione sono cablate nelle maschere di visualizzazione.
- Nessuna deduzione logica viene esplicitata tra i nodi della gerarchia e i vari concetti.

- I concetti di stream, indicatori, tariffe, categorie, fasce tariffarie e etichette sono legate ad un nodo: questa dipendenza non è visibile in nessuno strumento di analisi.
- Si sottolinea che ogni concetto, essendo legato a un nodo, ha anche delle dipendenze indirette con gli altri nodi della gerarchia: anche questo dettaglio dovrebbe essere incluso in un buon strumento di analisi.
- La creazione di una struttura gerarchica per la propria azienda non viene presa in considerazione da nessuno strumento di analisi: l'utente, che modella l'organizzazione dell'impianto con la gerarchia, non ha la possibilità analizzare i dati sulla base di queste informazioni.
- I concetti del sistema sono presentati in analisi, ma risultano scarsamente informativi
 - Nelle analisi dei dati non è presente un sistema di normalizzazione del dominio. Succede così che informazioni come stream e indicatori sono riproposti all'utente in analisi senza essere normalizzati: queste informazioni non hanno valore per l'utente e creano confusione nella comprensione dei dati.
 - Stream e indicatori devono essere presentati in analisi per quello che rappresentano, cioè delle serie temporali che registrano misure fisiche: quindi, i due concetti si devono fondere, ponendo in primo piano le informazioni di misure di energia, calore, corrente elettrica, luminosità, peso, ecc.
- Organizzazione della struttura gerarchica per livelli
 - Come già detto, nessuno strumento di analisi prende in considerazione le informazioni della gerarchia aziendale.
 - In realtà, il software non permette di raggruppare e catalogare i nodi del sistema: la divisione dell'organizzazione del sistema rimane un concetto astratto;
 - Non esiste un'analisi che permetta di comparare due nodi della gerarchia che rappresentano lo stesso concetto. Per esempio, non esistendo un concetto di reparto, non è possibile mettere a confronto i dati di due reparti diversi.
- Staticità dell'analisi dei dati
 - Le possibilità di eseguire analisi sui dati è vincolata dalle possibilità dello strumento.
 - Nessuno strumento permette di comparare i dati su un dettaglio temporale specifico, come ad esempio il confronto tra anni diversi di uno stesso fenomeno.

Il compito della prossima sezione sarà quello di razionalizzare l'analisi dei requisiti, definendo in maniera chiara quali sono le caratteristiche che andranno a comporre il nostro sistema; è necessario argomentare ogni requisito in maniera precisa, così da gettare le basi per l'analisi del problema.

3.4 Analisi dei requisiti

In questa sezione sono definiti i requisiti della nuova funzionalità di analisi dei dati da realizzare; l'obiettivo di questo strumento è di fornire nuove possibilità all'utente per eseguire analisi sui dati, superando i problemi evidenziati dai tool esistenti. Tale strumento sarà inserito all'interno della soluzione On.Energy.

I requisiti di questa analisi emergono dallo studio delle funzionalità attualmente offerte dal prodotto di Onit: infatti, tutti gli strumenti di analisi del dato mostrano molte criticità; tutti i punti deboli evidenziati saranno messi al centro del nuovo sviluppo. Le nuove caratteristiche del progetto promuovono il miglioramento di vari aspetti, che interessano sia la riprogettazione del modello dei dati, sia le possibilità di analisi che l'utente deve essere in grado di svolgere.

Il primo punto dell'analisi dei requisiti parte dalla gerarchia; a questo proposito, si vuole permettere di creare un nuovo modello di dati, che metta in relazione i nodi con il medesimo significato. Quindi, si vuole creare un nuovo dominio che permetta di gestire un insieme di informazioni, definite dinamicamente dall'utente in fase di configurazione. Ad esempio, nel caso di un condominio, l'utente deve poter mettere in corrispondenza il consumo energetico dei vari piani dell'edificio.

Il secondo punto di questa analisi è la nuova modellazione dei concetti di stream e indicatori; come detto precedentemente, quello che si vuole realizzare è uno strumento che metta in risalto l'informazione maggiormente rilevante, tralasciando i dettagli tecnici del modello. Stream e indicatori sono fusi in un unico concetto di serie temporale, dove le informazioni rilevanti sono specifiche della misurazione rappresentata, come la tipologia e l'unità di misura utilizzata. Questo requisito permette di porre al centro del nuovo strumento di analisi solamente informazioni utili per l'utente, eliminando tutto ciò che risulta superfluo.

Al terzo punto di questo nuovo strumento si pone la gestione dei dati maggiormente rilevanti del modello sorgente: gerarchia, stream, indicatori, categorie e tariffe saranno tutti a disposizione dell'analisi; a questo punto, sarà l'utente a scegliere quale tipo di dato inserire. L'introduzione di tutti i concetti all'interno di un nuovo modello pone il problema di come inserire nella progettazione questi dati, con lo scopo di avere un nuovo dominio coerente e affidabile.

Come quarto punto dell'analisi dei requisiti, si evidenzia la possibilità di aggregare i dati in base alle scelte dell'utente; quindi, il livello di aggregazione non è definito a priori. Per un'efficiente analisi, non è sempre necessario disporre di dati particolarmente dettagliati. Per evitare di limitare il campo di azione dell'utente, la definizione della granularità deve essere gestita secondo il tipo di analisi che deve condurre. Anche la definizione delle funzioni di aggregazione sui dati deve essere impostata dall'operatore.

Il quinto punto si concentra sull'aspetto dell'usabilità del nuovo strumento di analisi dei dati; infatti, rimane di fondamentale importanza avere uno strumento facile da usare, flessibile e interattivo; un punto fondamentale è poter analizzare informazioni sul maggior numero di dettagli possibile, come ad esempio sfruttare i concetti impliciti della gerarchia temporale, oppure quelli della struttura aziendale. Inoltre, si richiede che lo strumento sia progettato per consentire la gestione di un grosso quantitativo di dati: questo punto è centrale per tutti gli strumenti di analisi dei dati.

Come sesto punto si richiede al software di garantire la coerenza dei dati: in questo caso, si vuole avere uno strumento che gestisca in automatico l'aggiornamento dei dati: questa richiesta nasce con l'introduzione di una nuova modellazione del dominio, in aggiunta a quella già presente; la presenza di due modelli, che hanno il dovere di rappresentare le stesse informazioni, introduce problematiche di inconsistenza tra i dati che rappresentano.

Una volta effettuata l'analisi dei requisiti (esposta in *figura 15*), si è certi che qualsiasi altro step successivo dovrà rispettare le disposizioni definite in questa sezione. Infatti, il prodotto finale deve rispettare fedelmente l'analisi dei requisiti, per limitare al massimo la divergenza tra aspettative del cliente e quanto effettivamente implementato.

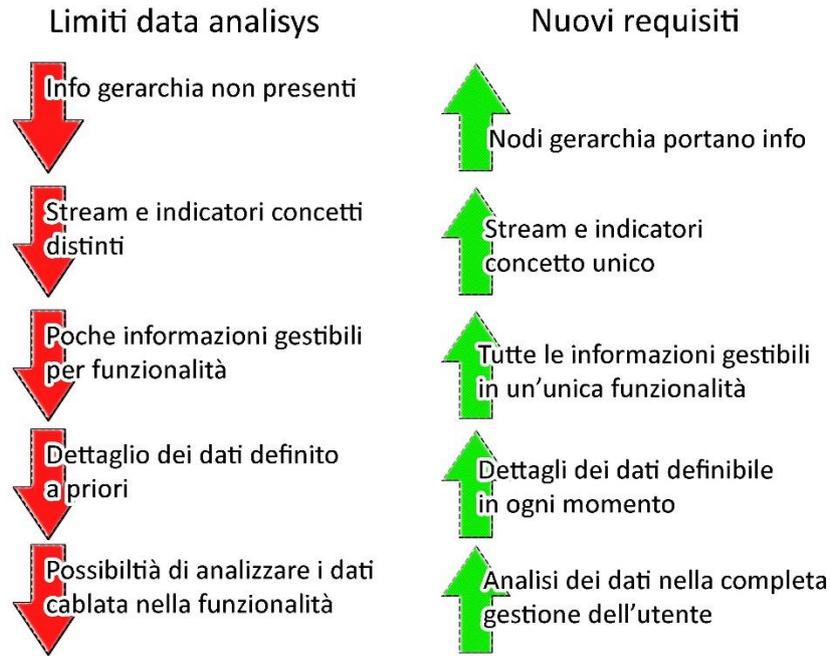


Figura 15 - Analisi dei requisiti

Nella prossima sezione si analizzeranno i problemi relativi all'analisi appena descritta: si evidenzieranno le criticità emerse, ponendo un approccio capace di definire le strategie di risoluzione delle problematiche emerse.

3.5 Analisi del problema

L'obiettivo di questa sezione è di effettuare una possibile analisi del problema legata a questo sviluppo. Saranno analizzate tutte le richieste pervenute in fase di analisi dei requisiti e si proporrà una possibile soluzione di quanto emerso.

La richiesta di gestire stream e indicatori come un'unica entità suggerisce un'indicazione: la nuova funzionalità utilizzerà un modello diverso rispetto al sistema sorgente; la necessità di creare un nuovo modello di dati risiede nell'esigenza di razionalizzare nuovi concetti: in questo caso, stream e indicatori saranno un concetto unico definito dal valore, dall'istante di tempo e dalle informazioni di misura rappresentata.

Sostanzialmente, il processo di estrazione dei dati dovrà mantenere una relazione tra i dati dello schema sorgente e quelli del nuovo modello. Ogni volta che sono introdotte modifiche sui dati del

modello sorgente, si devono ripercuotere anche su quello di destinazione. Anche la gerarchia sarà modellata tramite una nuova rappresentazione, dove ogni livello rappresenterà un nuovo concetto definito dall'utente. Nel nuovo modello emerge il problema di come codificare il nuovo dominio: infatti, si deve rappresentare in un unico oggetto tutte le informazioni che l'utente decide di inserire tra stream, indicatori, categorie e tariffe. La necessità di rappresentare un dato in questa maniera fa emergere un concetto di dimensioni già affrontato nel primo capitolo, in merito alle metodologie OLAP.

Il problema di gestire dati a livelli di aggregazione diversi, invece, risulta essere uno dei principali motivi che ha reso i sistemi OLAP uno strumento di riferimento per le analisi dati. A causa della necessità di gestire i dati in questo modo, è evidente che la struttura di riferimento sarà proprio un modello OLAP. Per quanto riguarda lo strumento integrato nell'interfaccia utente si pone il problema di offrire un tool che permetta di sfruttare dati multidimensionali, così da interrogare opportunamente la base dati e permettere all'utente una facilità di utilizzo dello strumento.

L'infrastruttura del progetto deve essere in grado di coordinare i diversi moduli del sistema; ci deve essere una organizzazione del software tale da permettere di aggiornare il modello multidimensionale ad ogni modifica dei dati del sistema sorgente.

In definitiva, lo strumento realizzato per l'implementazione di questa funzionalità comprende tutte le caratteristiche fondamentali di una base di dati OLAP, sia dal punto di vista della modellazione, sia dalle modalità di interrogazione della base dati.

Capitolo 4

Definizione dello schema dei metadati

L'obiettivo di questo capitolo è la creazione dello schema concettuale necessario alla realizzazione del data warehouse, utile al recupero delle informazioni per l'analisi dei dati. La realizzazione dello schema concettuale richiede delle operazioni preliminari utili a facilitare le varie fasi della progettazione. Il primo passo è individuare la porzione dello schema d'interesse di On.Energy: si evidenzieranno le classi interessanti per l'analisi dei dati. In seguito sarà necessario definire come sono organizzati i metadati: essi specificano quali istanze dei dati saranno utilizzati per alimentare le informazioni del cubo. Con le informazioni dello schema di On.Energy e il sottoinsieme dei dati definiti nei metadati, è possibile costruirsi un prospetto semplificato dello schema di On.Energy: una volta ottenuto questo schema, si adotterà la metodologia *dimensional fact model* per progettare lo schema concettuale; sarà quindi necessario identificare fatti, misure, dimensioni e gerarchie. Identificati questi concetti, sarà naturale progettare lo schema di fatto del data warehouse. L'ultimo argomento analizzerà tutti i processi che compongono il progetto di alimentazione dei dati del data warehouse.

4.1 Normalizzazione del modello relazionale

In questa capitolo è definito il processo di modellazione dello schema del cubo; tale processo di creazione inizia con la definizione di due basi di dati: la base di dati utilizzata attualmente da On.Energy e quella del metamodello del cubo, utile a configurare le informazioni necessarie alla materializzazione finale dei dati.

Punto iniziale di questo studio è la ricognizione delle basi di dati sorgenti. Tali modelli risultano in una forma che non è interessante per la funzionalità che sarà progettata: la motivazione principale è dovuta alla sostanziale diversità dei casi d'uso tra l'ambiente On.Energy, che si occupa del rilevamento dei dati dai dispositivi sul campo, e quello d'analisi dei dati, a cui non interessano le informazioni definite per il reperimento dei dati. Si prenda l'esempio di stream e indicatori, modellati diversamente dall'ambiente SCADA: se si escludono le informazioni utili alla generazione dei dati, essi risultano concetti simili. L'altro schema da esaminare è quello del metamodello del cubo, che serve a mappare le istanze di On.Energy nello schema finale del cubo. Dopo la ricognizione delle sorgenti, si cercherà di dedurre un prospetto dello schema di On.Energy, il quale presenta una semplificazione dovuta all'applicazione delle regole definite nei metadati del cubo: si deve tenere conto che alcune informazioni del modello precedente non sono interessanti nella nuova analisi, quindi non c'è la necessità di rappresentarle. Il prospetto dello schema di On.Energy avrà quindi una serie di vantaggi se confrontato con quello di partenza: tali benefici spiegano la necessità di creare un modello semplificato, il quale ridefinisce le relazioni tra le entità maggiormente interessanti. Inoltre, i nuovi concetti esposti presentano delle caratteristiche che prima non erano riscontrabili, permettendo di trarre benefici nella loro modellazione, oltre ad una migliore qualità dei dati e ad una loro maggiore coerenza. L'utilità di creare un prospetto dello schema di On.Energy sta nel facilitare la fase di progettazione: infatti, tale schema permette di spiegare meglio il processo di definizione dello schema di fatto. Nel processo di alimentazione dati del cubo, si inseriscono le logiche per ricavare le informazioni non presenti. Tra le attività rivolte alla creazione del cubo, c'è anche l'evidenziazione delle nuove relazioni che emergono nel prospetto dello schema di On.Energy: anch'esse dovranno essere riportate, soprattutto nel caso in cui la loro presenza sia d'interesse per l'utilizzatore finale. Nella costruzione del cubo finale, si ha l'opportunità di agire oltre che sulla struttura dei dati, anche sulla loro qualità; infatti, quando si procede al popolamento dei dati, si può migliorare la qualità delle informazioni: possono essere incomplete per dati mancanti o parzialmente presenti. L'utilizzo di una nuova base dati per la materializzazione del cubo è un'operazione nota nella realizzazione dei sistemi OLAP: duplicando i dati su due ambienti differenti, si evita di vincolare le prestazioni dei singoli sistemi. Anche l'operazione di trasformazione dei dati è comunemente nota nella progettazione dei data warehouse: infatti le varie operazioni descritte compongono la fase ETL, il cui scopo è fornire un modello di partenza per il sistema di analisi dei dati.

4.1.1 Metamodello di On.Energy

Il primo processo di ricognizione è eseguito sul metamodello dei dati del dominio On.Energy; tale processo nasce con lo scopo di evidenziare le informazioni d'interesse, cioè quelle prese in considerazione nella creazione dello schema del cubo: è chiaro che solo una parte delle informazioni del sistema sorgente è effettivamente utilizzata dal modello di destinazione. Detto questo, evitiamo di riportare nel diagramma delle classi del sistema On.Energy i dettagli non interessanti in questa tesi: si riporta solamente un sottoinsieme del dominio.

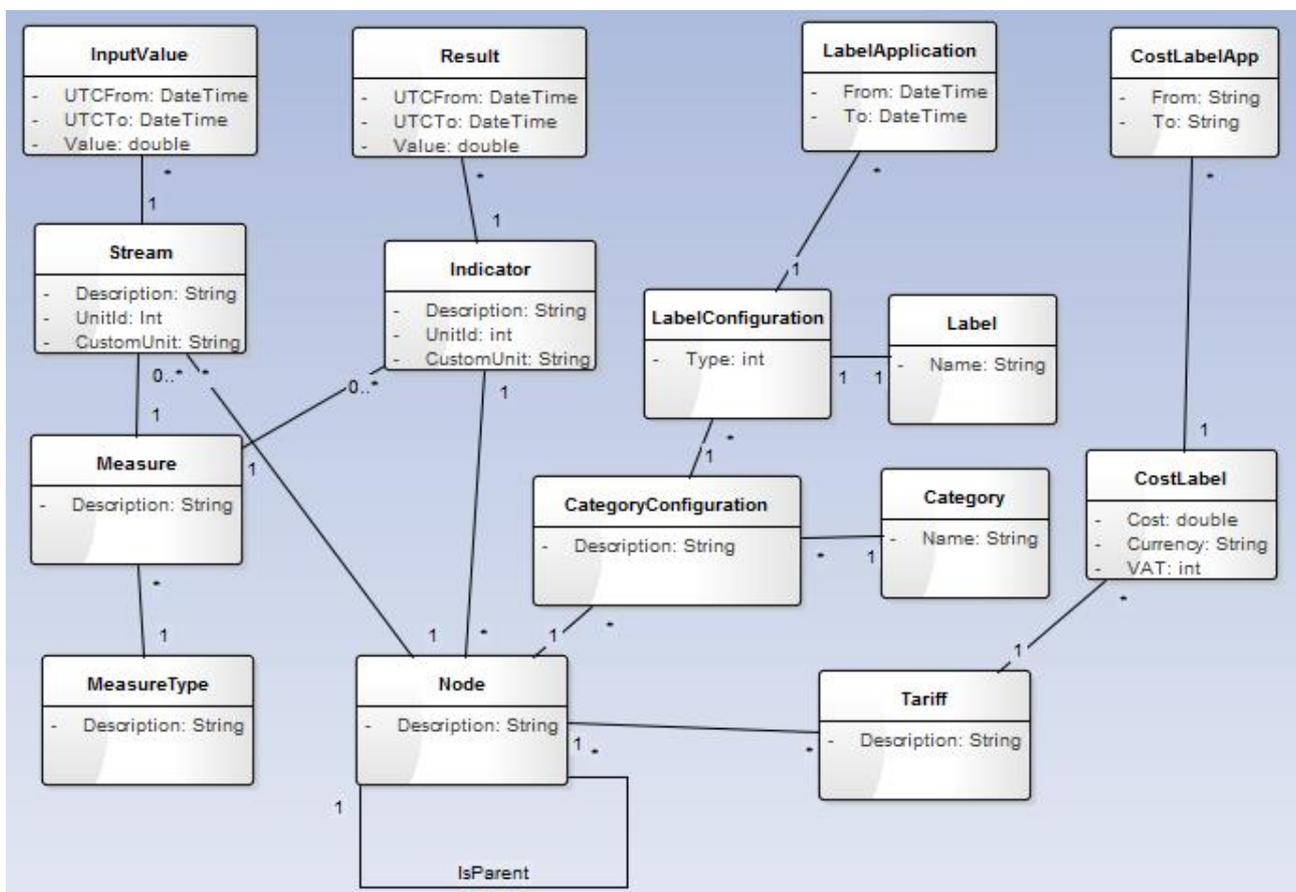


Figura 16 - Diagramma delle classi del metamodello di On.Energy

Ciò che si vuole mettere in risalto del modello di partenza è presentato dalla *figura 16*: il processo di ricognizione si occupa proprio di isolare il sotto-insieme d'interesse del dominio, riportando solamente i campi utilizzati nell'alimentazione dei dati dello schema del cubo. Infatti, l'esclusione dei campi inutili permette di ridurre la complessità del modello, evidenziando la presenza di più

classi rappresentanti il medesimo concetto; è il caso delle classi *InputValue* e *Result*, le quali una volta tolte le informazioni relative al recupero dei dati, presentano lo stesso set di attributi.

4.2 Metamodello del cubo

Insieme alla base di dati di On.Energy vista nel precedente paragrafo, vi è un altro dominio che deve essere considerato nella definizione dello schema del cubo: quello dei metadati. I metadati del cubo hanno un doppio compito da realizzare:

- Definire le istanze dei dati del modello On.Energy che faranno parte della creazione del cubo: le istanze esplicitamente inserite dall'utente sono gli stream, gli indicatori, le categorie, le tariffe e la proiezione della struttura gerarchica.
- Definire i dettagli di visualizzazione del cubo: specificare tutte le informazioni utili a completare lo schema del cubo. I parametri che è possibile impostare sono i nomi delle dimensioni e alcune specifiche delle misure, come il nome e l'unità di misura predefinita per la visualizzazione dei dati.

Anche se sono realizzati due casi d'uso differenti, lo schema dei dati utilizzato per memorizzare queste informazioni è unico. La motivazione che ha spinto a mantenere un unico modello sta nel forte legame tra questi dati: in particolare, la definizione delle istanze dei dati costituisce una prima fase di definizione di questo schema. Una volta definiti questi dati, il sistema applica una serie di logiche per costruirsi lo schema di fatto. Solamente a questo punto è possibile definire i dettagli di dimensioni e misure. Nelle prossime sezioni sono trattati in maniera separata i dettagli che differenziano i due sottoinsiemi di classi dello stesso schema.

4.2.1 Definire le istanze dei dati del modello On.Energy

Questa porzione di schema mantiene le correlazioni tra la configurazione del cubo e l'insieme di istanze che fanno parte di On.Energy. In sostanza, si definiscono i dati dello schema presi in

considerazione per la materializzazione dei dati del cubo. Una volta scelte le istanze, il sistema propone lo schema del cubo utilizzando le logiche della fase di progettazione.

Tramite l'analisi dei requisiti si sono potute identificare le informazioni d'interesse per l'utente, con l'obiettivo di determinare i parametri su cui può delineare la propria analisi. Il seguente elenco esprime le possibili scelte che può definire l'utente:

- Per definire le istanze dei nodi da introdurre nell'analisi, l'utente filtra le possibili istanze definendo il nodo base e il numero di livelli necessari ad effettuare il "taglio" della struttura gerarchica;
- La granularità dei dettagli della dimensione tempo è decisa dall'utente: infatti, si può decidere di agire sul livello di dettaglio temporale, lasciandolo alla precisione massima o arrotondandolo all'ora, giorno, mese o anno dell'istante di registrazione della serie temporale;
- Stream e indicatori da introdurre nell'analisi sono selezionati dall'utente; le istanze che si possono selezionare dipendono dalla visibilità del nodo di partenza.
- Il mantenimento del nome dello Stream/indicatore può essere omissso dal modello; togliendo quest'informazione, i valori delle serie temporali perdono l'informazione sulla loro provenienza.
- I campi che fungono da chiave di ogni rilevazione sono il dettaglio temporale e il riferimento di stream/indicatore: dal momento in cui si sceglie di eliminare o ridurre la cardinalità di uno di questi due concetti, l'utente ha l'obbligo di specificare quale funzione di aggregazione applicare sui dati che collidono su una stessa chiave. L'utente può scegliere di raggruppare uno stesso dato anche con più funzioni di aggregazione diverse.
- Categorie e tariffe sono selezionate in base al tipo di dettaglio che si vuole fare emergere nell'analisi; questa scelta incide notevolmente sulle relazioni del modello finale.

4.2.2 Definire i dettagli di visualizzazione del cubo

Una volta definite le istanze dei dati di On.Energy da inserire nella costruzione del cubo, sono applicate delle logiche di progettazione del fatto; il fatto ricavato presenta una definizione di

dimensioni e misure ricavate dalle istanze introdotte nel passo precedente. In questo paragrafo non sono trattate le logiche di creazione di dimensioni e misure, poiché rientrano nella progettazione del fatto. In questo paragrafo si specificano quali parametri sono definiti per completare la costruzione dello schema dei metadati del cubo. Questo sottoinsieme dello schema dei metadati è realizzato per memorizzare i seguenti parametri:

- Per ogni dimensione creata, si richiede di specificare una descrizione. Tale valore è proposto all'utente in fase di definizione delle dimensioni da inserire in analisi;
- Per ogni misura creata, si devono definire il nome e l'unità di misura utilizzata per mostrare i dati.
- Un dettaglio che non ha effetti sul modello finale, ma che interessa l'analisi dei dati è la definizione delle misure calcolate: l'utente costruisce nuove informazioni da quelle presenti, specificando quali misure sono oggetto dell'operazione scelta; il numero delle misure da scegliere dipende dal numero di operandi della funzione selezionata.
- La periodicità dell'aggiornamento dei dati è modificato in base alla necessità di mantenere i dati di analisi allineati con quelli delle sorgenti;

4.2.3 Realizzazione del metamodello del cubo

Una volta definiti i due sottoinsiemi che compongono lo schema dei metadati, è possibile presentare la sua modellazione finale. Nella *figura 17* è presentato il diagramma delle classi dello schema del metamodello del cubo: ogni entità *Cube* contiene delle informazioni legate al suo stato attuale di aggiornamento ed un'associazione con la sua definizione. La definizione del cubo si compone di parametri utili alla sua materializzazione, come il periodo di riferimento dei dati, la parte di struttura gerarchica da considerare, la granularità della dimensione temporale dei dati storicizzati, gli stream, gli indicatori, le dimensioni e le misure. Stream e indicatori riportano due dettagli: l'id dell'istanza a cui fanno riferimento e la tipologia di aggregazione quando necessaria. Le dimensioni nascono per collegare ai dati delle serie temporali nuove informazioni e si specializzano in 4 tipologie: dimensioni derivanti da etichette di tariffe o categorie, dimensioni create da etichette con data di applicazione di tariffe o categorie, dimensioni che rappresentano il nome del nodo di un livello della struttura gerarchica e dimensioni che riportano il nome di stream/indicator d'appartenenza del dato. La

definizione delle misure può comprendere sia misure semplici che misure calcolate: per le misure semplici si memorizzano le informazioni sulla misura (tipologia, unità di misura, etc...), mentre le misure calcolate sono caratterizzate da un'operazione e dalle misure semplici nel ruolo di operandi; le operazioni applicabili sono somma, media, massimo e minimo.

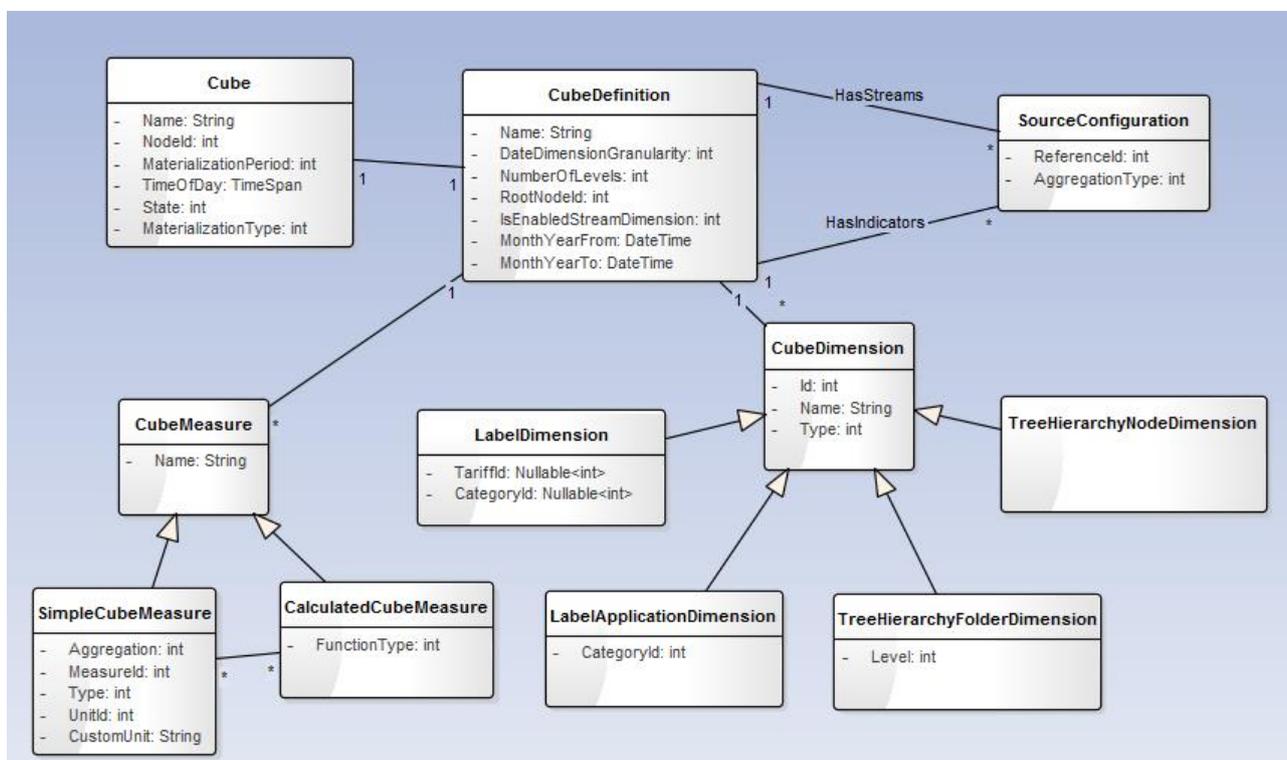


Figura 17 - Diagramma delle classi dei metadati del cubo

A questo punto, dopo aver inserito in analisi una porzione dello schema di On.Energy e il modello utile alla storicizzazione delle informazioni dei metadati, si rende opportuno creare un prospetto che semplifica lo schema di On.Energy. Tale semplificazione avviene grazie alle informazioni del metamodello del cubo. È riportato questo passaggio perché semplifica molto il processo di creazione dello schema di fatto, evidenziando molti dettagli che nascono dall'unione delle logiche dei due schemi sorgenti.

4.2.4 Prospetto del modello di On.Energy

Prima di affrontare la modellazione del fatto, è opportuno fare chiarezza sulle ripercussioni dei metadati sulle relazioni del modello On.Energy. Infatti, la definizione delle istanze permette di porre una sorta di parametrizzazione nelle relazioni del modello On.Energy: questo permette di costruire

un prospetto utile alla costruzione del fatto. Si possono elencare un insieme di operazioni che mirano a semplificare il dominio dei dati sorgente:

- Eliminando le informazioni che permettono di tenere traccia dei parametri relativi al recupero dati, i concetti di InputValue e Result assumono lo stesso significato: entrambi sono creati per rappresentare dati delle serie temporali. Nel prospetto riportato sono modellati in un'unica classe, dato che non ha senso mantenerli distinti.
- Stream e indicator forniscono le informazioni necessarie per la generazione di InputValue e Result, quindi derivano anch'essi dallo stesso concetto e sono ridotti ad un'unica entità nel prospetto.
- Categorie e tariffe subiscono una sostanziale modifica: si lascia il compito allo strumento di ETL di recuperare, per ogni valore InputValue e Result, l'etichetta valida nell'istante di rilevazione del dato. Non esistendo una relazione diretta tra InputValue/Result e etichetta/fascia tariffaria, si associano i dati sulla base dell'unico elemento in comune, il nodo d'appartenenza. Praticamente, le etichette e le fasce tariffarie perdono i riferimenti temporali, ma sono inserite come informazioni dei valori delle serie temporali. Tutte le informazioni utili alla configurazione delle etichette non sono riportate nel nuovo modello.
- Anche il numero dei nodi avrà una modellazione sostanzialmente diversa rispetto al modello di partenza; infatti i dati dello schema precedente sono limitati ad un numero di livelli definito dall'utente.
- Le informazioni provenienti dallo schema dei metadati mirano a semplificare le associazioni tra i concetti. Il numero di categorie/tariffe è definito a priori: questo permette di risolvere le relazioni molti-a-molti con un numero finito di relazioni molti-a-uno, dato che l'utente ha già scelto quali istanze faranno parte dell'analisi dei dati.

Il risultato di questa trasformazione è visibile nello schema in *figura 18*. Lo schema proposto può essere inteso come una prospettiva dello schema di On.Energy, a cui sono aggiunte le regole definite nello schema dei metadati. Infatti, la differenza tra questa prospettiva e l'originale modello di On.Energy sta proprio nella selezione di determinati parametri, intenti a ridurre la complessità del modello sorgente. Alcune associazioni molti-a-molti nello schema di On.Energy diventano

parametrizzate: in particolare, il numero dei nodi (indicato con M), il numero delle categorie (indicate con K) e il numero delle tariffe (indicate con J).

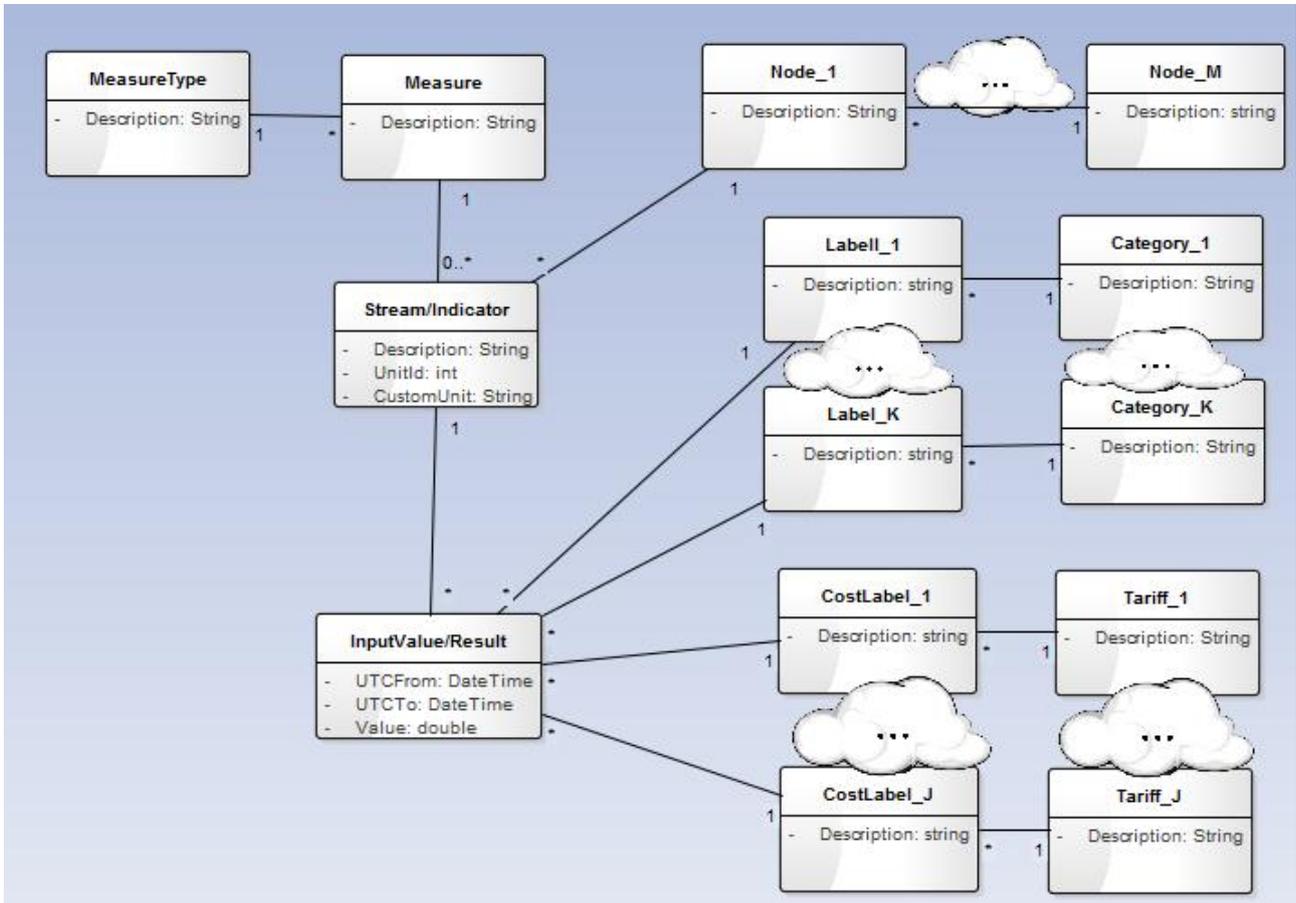


Figura 18 – Prospetto dello schema di On.Energy dopo aver applicato le scelte dei metadati

Il prospetto presentato permette di facilitare con maggiore chiarezza le prossime fasi di costruzione del fatto. Infatti, sono stati analizzati due processi fondamentali: il primo definisce come le etichette siano valorizzate per ogni valore delle serie temporali, mentre il secondo spiega come si semplifica il modello grazie alla parametrizzazione dei dati. Si rende noto che questo modello non è mai realmente storicizzato. A questo punto, si può analizzare il processo che porta a definire lo schema concettuale.

4.3 Modellazione DFM

Il *Dimensional Fact Model* (abbreviato DFM) è un modello concettuale grafico, pensato per assistere la fase di progettazione dello schema concettuale di un data warehouse; l'utilizzo del DFM permette all'utente finale e al progettista di confrontarsi su un modello intuitivo, utile a creare le fondamenta per la costruzione del modello logico e permettere di generare una documentazione espressiva e priva di ambiguità. Scopo del DFM è di generare schemi di fatto, che sono il risultato di questo processo di modellazione concettuale. Gli elementi base degli schemi di fatto sono i fatti, le dimensioni, le misure e le gerarchie.

Scopo di questa sezione è analizzare il prospetto dello schema di On.Energy, con la finalità di creare uno schema di fatto. Il progetto di modellazione concettuale comprende le fasi di scelta del fatto, selezione delle misure, identificazione delle dimensioni e costruzione delle gerarchie.

4.3.1 Identificazione del fatto

Estrapolando la definizione dalla teoria, il fatto si descrive come il concetto principale di ogni schema di fatto ed ha il compito di modellare un insieme di eventi che accadono nel contesto del sistema. Nel concetto di fatto deve sempre essere presente un aspetto dinamico, ovvero un insieme di informazioni utili a raccogliere un dettaglio del sistema in diversi istanti temporali. È necessario sottolineare l'importanza di questo elemento: sul fatto si concentra l'attenzione del processo decisionale. Un esempio di fatto è il concetto di vendita effettuata da un'azienda: la vendita è solitamente un dettaglio di importanza basilare, visto che ad essa è riconducibile un insieme di dati fondamentali come il cliente beneficiario, l'importo totale, i prodotti venduti e le rispettive quantità.

L'identificazione del fatto nel prospetto di On.Energy in esame è un'operazione abbastanza semplice: l'unica entità contenente un aspetto dinamico è la classe `InputValue/Result` dello schema prospetto di On.Energy. L'oggetto rappresentante le serie temporali è abbastanza semplice e contiene, oltre alla data di rilevamento, un valore numerico che esprime il dato specificato dall'associazione con gli `Stream/Indicator`. Anche se l'identificazione del fatto appare ovvia, è

possibile aggiungere una considerazione importante: il valore espresso dal fatto preso in considerazione non contiene un'informazione diretta, come nel caso del concetto delle vendite. Infatti, sia InputValue che Result registrano un valore numerico, che assume un determinato significato grazie all'associazione con Stream/Indicator, i quali forniscono informazioni specifiche sulla misura registrata.

4.3.2 Identificazione delle misure

Grazie alla teoria è possibile affermare che la misura è un attributo del fatto ed è un valore numerico: descrive l'aspetto quantitativo, principale oggetto d'interesse per l'analisi. Un esempio di misura è l'incasso della vendita.

Dopo aver assistito alla rapida individuazione del fatto nel prospetto dello schema di On.Energy, anche per la ricerca delle misure non ci sono margini d'errore: le misure dello schema di fatto sono direttamente legate al valore numerico dell'entità InputValue/Result; infatti, non esistono altri valori numerici che permettono un loro utilizzo nello schema di fatto finale. Maggiormente interessante è l'analisi di quante misure possono derivare dall'unica presente nell'entità; questo perché lo strumento che si vuole creare ha la necessità di distinguere opportunamente ogni valore sulla base della misura rappresentata, evitando di gestire un valore unico per ogni tipo di misura modellata dal sistema.

4.3.3 Regole di costruzione delle misure

Il numero delle misure create dipende direttamente dalla tipologia di dato che le serie temporali storicizzano; infatti, le informazioni sul tipo di misura, associate all'entità Stream/Indicator, sono determinanti per collocare adeguatamente ogni valore. Una modellazione di questo tipo evita di appiattire il modello in un solo campo, che sarebbe dunque privato di qualsiasi informazione; infatti, esso acquista semantica solamente grazie all'associazione con Stream/Indicator. Un'ulteriore frammentazione delle misure si ha escludendo alcune dimensioni o riducendone il dettaglio informativo: infatti, se si riducono informazioni relative al fatto, si rende necessario specificare quale tipo di aggregazione si applica ai record che memorizzano le medesime informazioni.

In sostanza, è possibile creare le misure solamente dopo aver definito le istanze di stream e indicatori che fanno parte del cubo. Partendo dalle istanze selezionate, si creano dei raggruppamenti sulla base di 5 informazioni: la tipologia della misura (es. energia, volume), la tipologia della serie (es. bolletta, appartamento, edificio), l'unità di misura (es. kWh, minuti, m³), l'unità di misura personalizzata (definita dall'utente) e il tipo di aggregazione (somma, conteggio, minimo, massimo). Per ogni raggruppamento è creata una misura che contiene i valori della serie temporale corrispondente. Si fa notare un aspetto importante: le informazioni che servono a creare i raggruppamenti non sono utilizzate per creare altre dimensioni, dato che la loro espressività è già stata utilizzata nella costruzione delle misure.

4.3.4 Identificazione delle dimensioni

Da un punto di vista teorico, il concetto di dimensione si definisce come una proprietà del fatto che assume un valore di un dominio finito: permette di dettagliare con informazioni il fatto, descrivendone una coordinata dell'analisi. Nell'esempio delle vendite, possibili dimensioni sono il prodotto, il magazzino e il cliente.

La ricerca delle dimensioni che compongono il fatto è chiaramente visibile nel prospetto del modello di On.Energy. Le potenziali dimensioni sono create da tre relazioni: la prima è costituita dallo stream/indicator, la seconda è dovuta all'assegnazione di ogni serie temporale ad una etichetta/fascia tariffaria e la terza deriva dall'associazione con un nodo della struttura gerarchica. L'associazione con lo stream/indicator permette di caratterizzare il fatto di informazioni sulla tipologia della serie temporale rappresentata, in particolare aggiunge informazioni sullo stream/indicator di appartenenza della misura. L'associazione con i nodi della struttura gerarchica permette di ricostruire l'insieme di nodi che l'utente ha scelto di inserire nell'analisi. Invece, l'associazione con le etichette è maggiormente complicata a causa della parametrizzazione del dato da parte dell'utente: la relazione non evidenzia un legame multi-a-molti, ma il numero di associazioni è definito a priori grazie allo schema sorgente dei metadati. I metadati obbligano l'utente a determinare le effettive relazioni con le etichette, selezionando esplicitamente le categorie/tariffe da inserire nella realizzazione del modello; tale dettaglio risulta determinante nella

costruzione dello schema di fatto, dato che ne semplifica la modellazione introducendo un numero di associazioni molti-a-uno tra il fatto e le categorie/tariffe.

4.3.1 Regole di costruzione delle dimensioni

Si definiscono le regole seguite dal sistema per creare fisicamente le dimensioni del fatto:

- È definita una dimensione per memorizzare lo stream o l'indicatore di origine per ogni valore della serie temporale; questa dimensione può essere omessa se opportunamente specificato nei metadati.
- Sono create tante dimensioni quanti sono i livelli di struttura gerarchica richiesti dall'utente; sono presi in considerazione i livelli del sottoalbero del nodo scelto per la definizione del cubo: i nomi dei livelli di una specifica misura sono inseriti in base alla relazione del corrispondente stream/indicatore con i nodi della gerarchia. Questo significa che per ogni valore della serie temporale, i nomi dei nodi riportati sulle dimensioni sono antenati dello stream/indicatore corrispondente; chiaramente, sono riportati i nodi che rientrano nella proiezione della struttura gerarchica definita dall'utente.
- Per ogni istanza di categoria selezionata tra i dati di On.Energy, è creata una dimensione che raccoglie il nome dell'etichetta, assegnata sulla base dell'istante di registrazione del valore della misura.
- Come nel caso delle categorie, per ogni tariffa è creata una dimensione: anche in questo caso, ad ogni istante di tempo di registrazione del valore della misura, è memorizzato il nome della fascia tariffaria applicata.
- Infine, sono create un numero fisso di dimensioni derivanti dall'informazione sul tempo di registrazione dei dati.

4.3.2 Gerarchie

Prendendo la definizione dalla teoria, la gerarchia si rappresenta con un albero direzionato, in cui tutti i nodi sono attributi dimensionali, cioè informazioni che descrivono in maniera dettagliata la

dimensione. Nel dominio delle vendite, esempi di attributi dimensionali sono la marca del prodotto, la sua categoria e il nome del produttore. Tutti gli archi della gerarchia modellano associazioni molti-a-uno; la gerarchia rappresenta informazioni riconducibili direttamente ad una dimensione, infatti ogni nodo è un dettaglio della dimensione.

Nella modellazione dello schema di fatto sono menzionate le tre tipologie di dimensioni selezionate dall'analisi: gli stream/indicator, i nodi della gerarchia e le etichette. Oltre alle gerarchie derivate da queste dimensioni, ce n'è un'altra indirettamente presente nel modello: è la gerarchia temporale, presente in qualsiasi schema di fatto. La gerarchia generata dall'informazione sul tempo di registrazione della misura diventerà un oggetto determinante nello strumento di analisi; infatti, ogni gerarchia temporale viene scomposta in tante informazioni ricavate dalla dimensione tempo: queste informazioni saranno poi rese disponibili in fase di analisi, in modo da poter aggregare le misure su valori specifici; un esempio di scomposizione della misura tempo è la frammentazione in anno, semestre, quadrimestre, trimestre, etc.

La gerarchia ottenuta dalla dimensione Stream/Indicator contiene informazioni sulla tipologia di misura registrata dalla serie temporale, oltre a dettagli sul nodo di appartenenza. L'originale modello di On.Energy mostra i nodi con un'associazione a sé stessi; tale modellazione rappresenta la parentela di un nodo con un nodo padre: nella progettazione di schemi di fatto è possibile gestire queste situazioni ricorsive, scegliendo quanti livelli riportare nello schema. In questo caso, quest'informazione è definita nei metadati e consente di capire quanti nodi saranno riportati nel data warehouse. Inoltre, ogni schema di fatto è creato sulla base di un nodo; questa scelta specifica il campo d'analisi, dando la possibilità all'utente di considerare solo nodi appartenenti al sottoalbero del nodo scelto per la creazione del modello. Una volta scelti i nodi base e il numero dei livelli della struttura da riportare nello schema finale, la struttura è "tagliata" e riportata sul numero di livelli scelti dall'utente: tutte le informazioni collegate a nodi non riportati nello schema finale, sono proiettati sul primo nodo antenato presente sulla proiezione della struttura.

La gerarchia creata dalla relazione con lo stream è molto semplice: permette di inserire il nome del dispositivo di riferimento da cui provengono i dati della serie temporale. Il numero di gerarchie create dalle etichette si esprime con la somma di istanze selezionate dalle categorie e dalle tariffe:

ognuna di queste istanze andrà a formare una nuova dimensione, in cui potranno comparire descrizioni di etichette appartenenti alla corrispondente categoria/tariffa.

4.4 Schema di fatto

Le precedenti sezioni hanno contribuito a determinare tutte le informazioni utili per modellare efficacemente lo schema di fatto. In questa sezione non resta che applicare le tecniche di progettazione DFM per ricavare lo schema di fatto. La struttura del fatto si ricava partendo dal prospetto dello schema di On.Energy e dalle scelte definite per la definizione di fatto, misure, dimensioni e gerarchie.

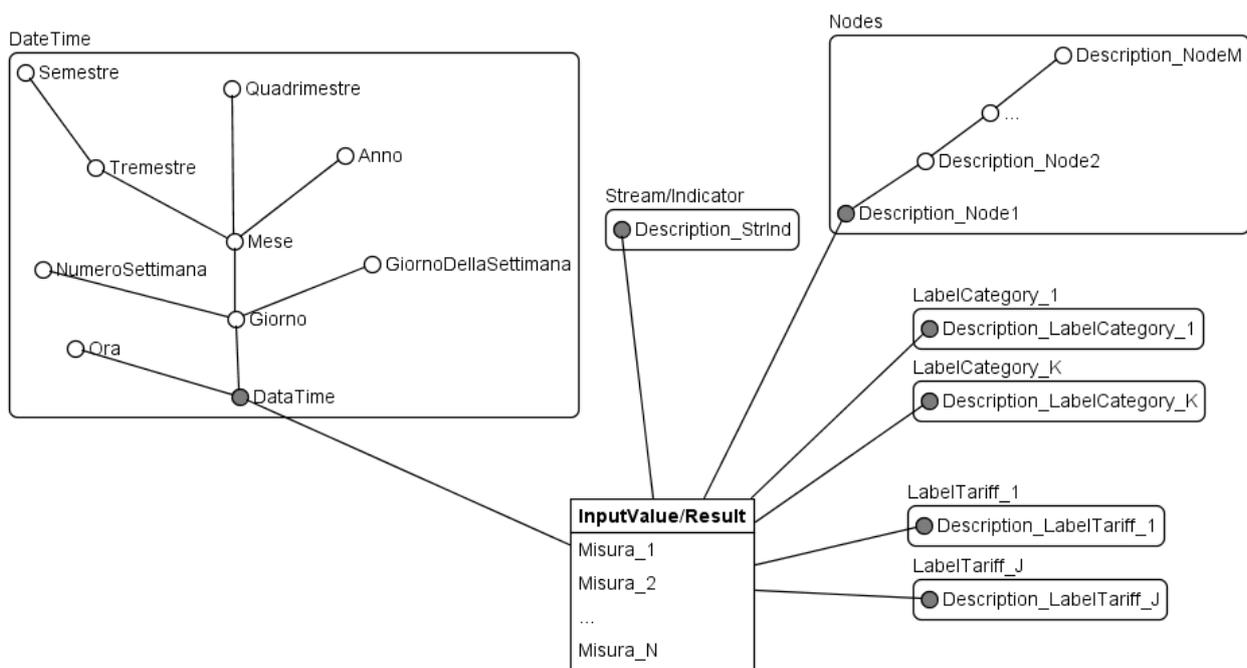


Figura 19 - Schema di fatto

In *figura 19* è modellato lo schema concettuale del data warehouse. La realizzazione dello schema di fatto si avvale dello strumento *BI Modeler* [SCA-BIM]. L'appunto da far emergere in questo schema di fatto riguarda le molteplici associazioni con le gerarchie delle etichette: tale molteplicità deriva dalla definizione dei metadati: permettono di definire, in fase di costruzione dello schema, le categorie/tariffe che faranno parte del sistema. Il medesimo ragionamento vale anche per il numero

dei nodi da riportare nello schema. Come detto, il numero delle misure è stabilito dalla logica definita nelle regole di creazione.

La definizione dello schema fisico dedotto da questo schema di fatto presenta una modellazione atipica, almeno per quanto riguarda la procedura consueta di costruzione di un data warehouse. Infatti, il risultato finale non è uno schema fisico su database, ma bensì utilizza una serializzazione su file dei dati. L'oggetto serializzato è rappresentato da una classe: le caratteristiche della classe e l'insieme di informazioni aggiuntive memorizzate, hanno l'obiettivo di rappresentare un dato multidimensionale.

La modellazione fisica consueta di uno schema di fatto è comunemente eseguita con la creazione di tabelle: un insieme di *dimension table* rappresentano le dimensioni dello schema di fatto, mentre una tabella *fact table* è utilizzata per rappresentare il fatto. Il fatto è composto da misure ed identificatori dei record delle dimension table. La costruzione del modello fisico su database può comprendere delle varianti in base al tipo di schema adottato nella costruzione. La modellazione attraverso una classe è simile a quella fatta con le tabelle: si costruiscono una serie di attributi che permettono di memorizzare sia le misure, sia gli identificatori dei concetti collegati. Altre strutture dati si occupano di mantenere l'accoppiamento tra gli identificatori e le opportune informazioni correlate.

La modellazione di uno schema fisico con delle classi serializzate su file è molto differente rispetto all'utilizzo di un database: è possibile dire che l'utilizzo di classi permette una migliore flessibilità in fase di progettazione della base-dati, mentre in fase di interrogazione ci sono grossi svantaggi che gravano sulle prestazioni del sistema. La scelta di utilizzare file serializzati è dovuta principalmente alla dinamicità delle dimensioni dello schema di fatto; infatti, il numero delle dimensioni può variare in base alle informazioni dello schema dei metadati. Soluzioni dinamiche sono gestibili anche con i database, ma la scelta di propendere per l'utilizzo di file permette di mantenere invariata la complessità della base dati; inoltre, la soluzione On.Energy contiene già una buona infrastruttura utilizzabile per gestire la serializzazione su file.

Per quanto riguarda la memorizzazione dei dati su file, è stato deciso di partizionare la serializzazione dei dati in base al mese-anno di riferimento: questa scelta è in linea con il tipo di

analisi svolta dall'utente, il quale decide di selezionare un periodo temporale di alcuni mesi per il caricamento dei dati. Questa partizione dei dati su mese suddivide le classi, utili alla memorizzazione dei dati, in tre categorie:

- **Dati indipendenti dal tempo:** mantengono il mapping tra gli id e le informazioni dei concetti. Non dipendono dal tempo, quindi vengono memorizzate una sola volta.
- **Dati strettamente dipendenti dal mese:** storicizzano gli intervalli di validità delle etichette per ogni categoria/tariffa.
- **Dati atomici della fact table:** sono paragonabili ai record della fact table e sono costituiti dalle misure e dagli id delle dimensioni a cui ogni rilevazione fa riferimento; diversamente dai database, la loro modellazione si avvale di vettori la cui dimensione dipende dai metadati inseriti nell'analisi. Sono composti dal dato temporale, da un vettore per registrare i valori delle misure e un vettore per memorizzare gli identificatori delle dimensioni.

Definito il modello fisico dei dati, è necessario spiegare come avviene il processo di alimentazione delle classi che costituiscono i dati del data-warehouse.

4.5 Progetto di alimentazione

Il progetto di alimentazione ha lo scopo di caricare i dati dalle sorgenti, eseguire le dovute trasformazioni sulla base delle scelte prese in esame e alimentare la nuova base dati di riferimento, necessaria a rispondere alle interrogazioni che l'utente sottoporrà al sistema in analisi dei dati. In questa sezione si specificano i passi che questo step dovrà prendere in considerazione.

Il modello grafico utilizzato per la rappresentazione del progetto di alimentazione si chiama *Data Flow Diagram* (abbreviato DFD): questa modellazione enfatizza le operazioni effettuate sulle informazioni, mostrando le dipendenze funzionali tra i processi; inoltre, permette di approfondire il sistema modellato grazie alla specifica di più livelli di astrazione [GOL-DFD].

La definizione del progetto di alimentazione segue un approccio top-down: inizialmente, sono illustrati dei DFD che mettono in relazione le macro-funzionalità del sistema; successivamente, le

macro-funzionalità più complesse saranno ulteriormente approfondite, riportando una modellazione DFD dedicata.

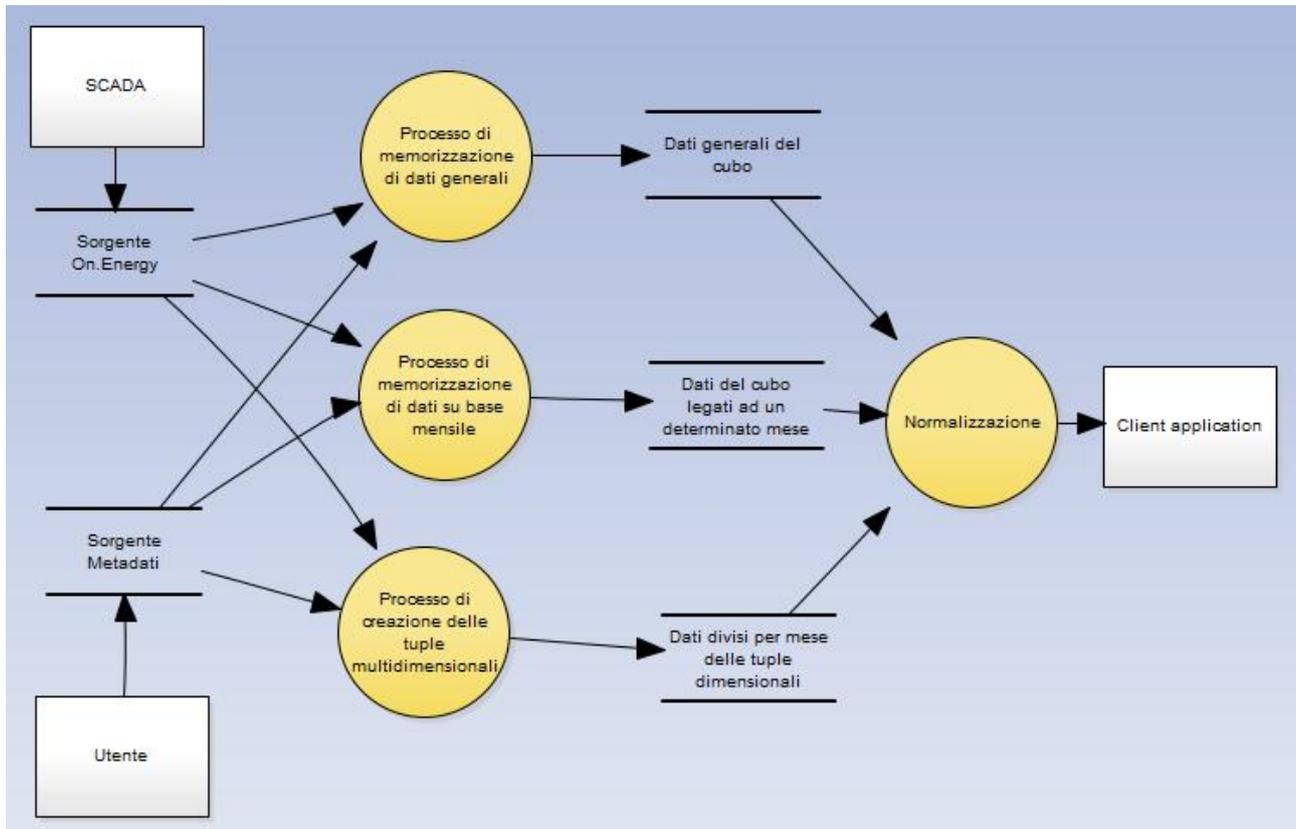


Figura 20 - Macro-step del progetto di alimentazione

La figura 20 mostra le tre principali macro-funzionalità del sistema. Ogni funzionalità prende i dati dagli schemi sorgenti, esegue la propria elaborazione e storicizza i dati su un insieme di file isolati. Ogni base di dati descrive delle informazioni diverse: solamente rileggendo interamente i dati dei tre diversi schemi, il processo di **Normalizzazione** ricompone i dati del record dimensionale, allo scopo di proporli all'interfaccia client.

Il *processo di memorizzazione di dati generali* riporta nel proprio modello le informazioni utili a mappare gli identificatori dei vari concetti con le relative descrizioni: fondamentalmente ricopia i dati dal modello sorgente, ma solamente quelli d'interesse per l'analisi. Anche il *processo di memorizzazione di dati su base mensile* esegue le stesse operazioni, ma opera una memorizzazione su file divisi mensilmente, in base al numero di mesi definiti dall'analisi. Il processo di *Normalizzazione* ricompone i dati in un'unica lista di record, associando gli identificatori con le opportune informazioni a cui fanno riferimento. Per quanto riguarda il *processo di creazione delle*

tuple dimensionali è necessario aggiungere un livello di astrazione dello schema DFD dedicato, dato che questa funzionalità è maggiormente complessa rispetto alle altre.

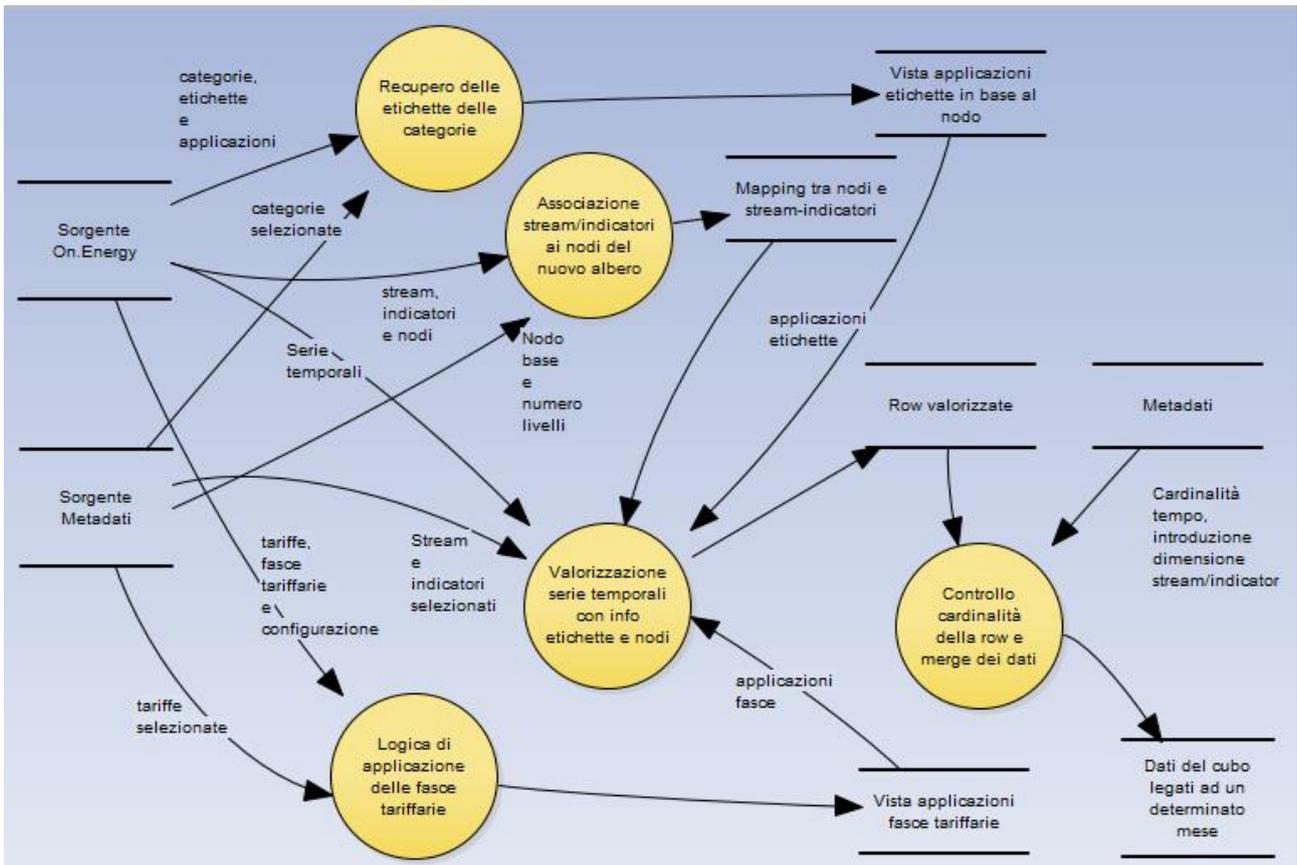


Figura 21 - Focus del processo di definizione delle tuple dimensionali

Con il *data flow diagram* della *figura 21* si concentra il cuore di alimentazione dei dati del sistema. Il processo di *recupero delle etichette delle categorie* raccoglie i dati delle etichette da inserire in analisi e sfrutta l'API messa a disposizione dal sistema per recuperare gli intervalli di validità alle etichette: l'output del processo è composto dalle applicazioni delle etichette. Anche la *logica di applicazione delle fasce tariffarie* è un processo simile, ma utilizza le fasce tariffarie per la costruzione delle loro applicazioni temporali: anch'esso recupera i dati di interesse e li rende disponibili per i processi successivi.

Il processo *associazione stream/indicatori ai nodi del nuovo albero* permette di assegnare gli stream/indicator ad un nodo della nuova porzione di struttura gerarchica, ma solo nei casi in cui i nodi-padre reali siano esclusi dalla nuova struttura.

I tre processi appena descritti servono a fornire le informazioni necessarie alla valorizzazione delle tuple dimensionali. Il processo *valorizzazione delle serie temporali con info etichette e nodi* applica le logiche per associare, ad ogni serie temporale, un'etichetta per ogni categoria/fascia inserita in analisi. Inoltre, alimenta le informazioni sul nodo d'appartenenza per ogni stream/indicatore. L'output del processo non è altro che la lista delle *row*, ognuna composta dal valore della serie temporale, dallo stream e da tante colonne quante sono le informazioni di etichette/nodi da inserire nella tupla.

Si fanno notare due aspetti importanti, che rendono questa progettazione molto dissimile rispetto ad una progettazione con database. Per prima cosa, la gerarchia temporale mantiene un unico dato legato alla dimensione tempo, cioè non sono denormalizzate anche le altre informazioni come avviene nel caso dei database. Tutta la scomposizione dell'unico dato temporale sarà ricavata in fase di visualizzazione del dato. La seconda riguarda la possibilità di gestire valori non presenti, come può essere la mancanza di un nodo d'appartenenza; nel caso delle tabelle, tale situazione richiede tecniche di valorizzazione obbligatoria dei dati, normalmente copiando quelli presenti.

L'ultimo processo, denominato *controllo cardinalità della row e merge dei dati*, si occupa di creare un numero di attributi sulla base delle tipologie di misure presenti e le unità di misura precisate. Successivamente, sulla base delle specifiche dei metadati, verifica se mantenere il nome dello stream/indicator, e se necessario, riduce la granularità del dato temporale. Questa operazione elimina alcune informazioni delle row: si creano così gruppi di tuple con le stesse informazioni; il numero di tuple per ogni gruppo è direttamente proporzionale al numero di informazioni ridotte. L'ultimo step aggrega i valore delle misure delle righe che presentano il medesimo dettaglio dei dati: i tipi di aggregazione da effettuare sui dati dipendono dai metadati.

L'output finale di tutto il processo di trasformazione è appunto un insieme di classi serializzate, costituite da dati generici e liste delle row suddivise per il mese di riferimento. Tutto il processo descritto è eseguito solamente la prima volta: nelle successive iterazioni si prendono in considerazione una porzione di dati ridotta. Detto ciò, ogni aggiornamento dei dati prende in considerazione la ricostruzione dei file che contengono i dati delle rispettive sorgenti modificate. Infatti, l'aggiornamento dei dati avviene per tre motivi principali:

- Allineamento dei dati aggiunti alle serie temporali. La ricostruzione giornaliera inserisce i dati non processati, valutando il periodo di ricostruzione dei dati in base alle modifiche intervenute nelle 24 ore appena trascorse.
- Modifica dei dati di un'istanza On.Energy. Se interviene una qualsiasi modifica sui dati sorgenti di On.Energy cooperanti nella definizione di un cubo, è necessario materializzare nuovamente i dati memorizzati dal data warehouse. Un esempio possibile è la modifica del nome di un'etichetta o l'eliminazione di uno stream di dati.
- Modifica dei metadati. L'utente decide di cambiare dei parametri del metamodello del cubo: questa operazione obbliga a ricreare ex-novo il data warehouse.

Capitolo 5

Implementazione del sistema

L'obiettivo del capitolo è di presentare le interfacce necessarie per permettere all'utente di realizzare un'analisi dei dati. I capitoli precedenti hanno delineato due diversi casi d'uso rivolti all'utente: il primo riguarda la definizione dei metadati, mentre il secondo è relativo all'analisi dei dati. Per la configurazione dei metadati, è fornito all'utente un wizard per definire i parametri che realizzeranno fisicamente il data warehouse. Una volta definiti i parametri per la costruzione del modello dei dati, l'utente dispone di uno strumento per l'interrogazione del sistema, grazie ad una maschera che permette di condurre l'analisi utilizzando una tabella pivot. Successivamente, sarà affrontato il tema delle prestazioni per verificare l'effettiva validità del programma, anche per analisi su un'importante set di dati. Il capitolo si conclude con il caso di studio condotto con l'azienda Amadori.

5.1 Panoramica generale sull'applicazione client

Questa sezione ha lo scopo di orientare il lettore tra i diversi casi d'uso definiti in questa tesi: da una parte si vuole permettere all'utente di predisporre il dominio necessario all'analisi dei dati, mentre dall'altra si deve fornire uno strumento interattivo necessario ad analizzare i dati recuperati del sistema SCADA/EMS realizzato.

Come primo punto è necessario riassumere le tecnologie impiegate per permettere all'utente di utilizzare il prodotto On.Energy. Infatti, l'interfacciamento col sistema è reso disponibile all'utente attraverso un'applicazione client sviluppata con Silverlight: questa tecnologia permette di essere utilizzata da browser tramite opportuna installazione di un plug-in. Il limite della piattaforma

Silverlight è di poter rendere disponibile il suo utilizzo solamente a PC e Mac, escludendo le piattaforme mobili tra i sistemi compatibili; bisogna tenere conto che l'analisi dei dati è comunque un'attività fortemente interattiva, tale fattore ne riduce le potenzialità sui dispositivi mobili. La comunicazione attraverso la piattaforma Silverlight e il sistema On.Energy avviene attraverso la tecnologia *Windows Communication Foundation* (abbreviato WCF): grazie ad essa, il client dell'utente esegue il recupero dei dati una sola volta dal server, per poi analizzarli localmente nella propria macchina.

L'interfaccia di configurazione del data warehouse permette la definizione dei metadati dello schema: attraverso l'opportuna maschera sono definiti i parametri ampiamente discussi nei capitoli precedenti; i metadati restringono la sfera di competenza dell'analisi dei dati e il sistema permette di creare diversi basi di dati su cui effettuare le proprie interrogazioni.

L'interfaccia di analisi dei dati è uno strumento sviluppato sui concetti della tabella Pivot: tale controllo guida l'analisi dei dati permettendo di raggruppare su colonne e su righe i valori delle dimensioni, presentando nelle celle i valori aggregati delle misure; le funzioni di aggregazione dei dati sono scelte in base alle esigenze dell'utente. Tale controllo è sviluppato da una multinazionale del settore. La predisposizione del controllo, ad essere esteso con funzionalità proprie, permette di adattare il suo utilizzo alla realtà del dominio.

5.2 Interfaccia di configurazione dei metadati

La sezione in esame si occupa di definire come è implementato il wizard di configurazione dei metadati. Come detto, la configurazione dei metadati permette di definire i parametri utili all'implementazione fisica del data warehouse, necessario al sistema per consentire un veloce reperimento dei dati richiesti in fase di analisi.

La funzionalità di configurazione dei metadati è opportunamente suddivisa in step per fornire all'utente una costruzione assistita della base di dati. Infatti, la struttura a step permette di costruire gradualmente l'analisi, supportando l'operatore nella scelta dei dati da visionare. In questa sezione si evita di riportare nuovamente i metadati definibili dall'utente, ma si verificherà direttamente

come viene guidato l'utente attraverso il wizard di configurazione. Si cercherà di porre attenzione nei passaggi tra uno step e l'altro, visto che ognuno di essi applica delle logiche che indirizzano l'utente ad una costruzione coerente dei metadati.

Il sistema preso in esempio nella descrizione degli step di configurazione rappresenta il controllo di un insieme di condomini: per ogni condominio si hanno più piani, che a loro volta hanno appartamenti che ospitano famiglie. Ogni famiglia detiene un contatore della luce, uno dell'acqua e uno del gas.

5.2.1 Definizione dei dati di persistenza

Il primo step di costruzione del data warehouse ha il compito di richiedere all'utente dati generici: è richiesto di specificare il tempo di storicizzazione dei dati, il numero di livelli da mantenere della struttura gerarchica, le gerarchie e le tariffe da includere nell'alimentazione dei dati. Va detto che ogni modello è creato su un nodo scelto in partenza: tale associazione incide sia sulla creazione della struttura della gerarchia, sia sulla visibilità globale dei dati che si possono inserire in analisi.

Cubo Id 0 Stato Da inizializzare Prossimo aggiornamento n.d. ← →
Ultimo aggiornamento iniziato n.d. Ultimo aggiornamento completato n.d.

Step 1 - Dettagli cubo

Numero livelli gerarchia

Includi dimensione stream/formula

Granularità dimensione data/ora

Categorie

Stagione	<input checked="" type="checkbox"/> Etichetta	<input type="checkbox"/> Applicazione etichetta
Giorno-Notte	<input checked="" type="checkbox"/> Etichetta	<input type="checkbox"/> Applicazione etichetta
Mattino-Pomeriggio-Sera	<input checked="" type="checkbox"/> Etichetta	<input type="checkbox"/> Applicazione etichetta

Tariffe Tariffa Energia

Data inizio persistenza

Data fine persistenza

Figura 22 - Step 1: configurazione dati di analisi

Nella *figura 22* sono riportati i parametri di configurazione dell'analisi dei dati: il significato di ogni parametro è stato approfondito più volte nel corso dei capitoli precedenti. Le note da aggiungere

sono due: la prima che l'utente può definire un periodo temporale specifico per limitare la memorizzazione dei dati, la seconda che per ogni categoria è possibile associare un'etichetta oppure un'applicazione etichetta; se per l'etichetta abbiamo ampiamente discusso, per l'applicazione etichetta non è stato ancora detto niente. L'applicazione etichetta intesa come dimensione non è altro che il nome dell'etichetta concatenato al periodo di validità; la modellazione di questa dimensione è stata prevista per poter effettuare calcoli sulla base di una particolare istanza di etichetta, e non raggruppare dati di etichette diverse: visto che questa dimensione è stata forzosamente richiesta, non sarà ulteriormente approfondita, poiché non apporta nessun valore all'analisi dei dati.

Confermando i parametri impostati in questo step, il sistema applica alcune logiche utili a supportare l'utente nella creazione dei dati d'analisi. Nell'esempio riportato in figura, i parametri definiti portano a ridurre la qualità dei dati, visto che è arrotondato il tempo di rilevazione al dettaglio dell'ora: questo fattore implica che ogni misura creata dovrà tenere conto di una funzione di aggregazione per i valori con lo stesso livello di dettaglio. Un'altra operazione che svolge il sistema è costruire la proiezione di struttura gerarchica: nello step successivo sarà proposto il set di nodi derivanti dal numero di livelli introdotto nell'analisi dei dati.

5.2.2 Introduzione serie temporali

Prima di poter passare alla creazione di dimensioni e misure, è necessario introdurre gli stream e indicatori che faranno parte dell'analisi dei dati. Nel secondo step l'utente può scegliere quali stream e indicatori inserire in analisi: per ogni serie temporale selezionata, è richiesto di specificare le funzioni di aggregazione da applicare ai dati raggruppati per il medesimo livello di dettaglio.

Si fa notare che nella *figura 23* sono riportati solamente i nodi relativi al numero dei livelli introdotti. Ogni volta che è inserito uno stream o un indicatore, il sistema riassegna tale oggetto ad un nodo della proiezione della gerarchia disponibile. Inoltre, ad ogni serie temporale selezionata è assegnato un colore che rappresenta una determinata misura di destinazione.

5.2.3 Struttura dello schema multidimensionale

Il terzo step applica tutte le regole di implementazione del cubo definite nel capitolo precedente; infatti, grazie ai primi due step sono definite le istanze che andranno a comporre il cubo; queste informazioni, assieme alle regole di costruzione del cubo, permettono di creare le misure e le dimensioni del fatto.

Nella *figura 24* si vede come l'utente può intervenire nell'inserimento dei metadati utili a definire i nomi delle dimensioni e delle misure del fatto. Per ogni dimensione sono riportate le informazioni dei dati che ne faranno parte, al fine di aiutare l'utente nella definizione di un appropriato nome per la colonna. Per ogni misura l'utente deve definire l'unità di misura comune a tutti i valori visualizzati.

 Nome Stagione(etichetta) Esempio	 Nome Edifici Esempio Dei Tulipani, Delle Rose	 Nome Stream / Formula Esempio Contatore acqua, Contatore elettrico
 Nome Giorno-Notte(etichetta) Esempio	 Nome Piani Esempio Primo Piano, Secondo Piano, Terzo Piano, ...	Nome Misura Volume Tipo Unità di misura m Tipo di aggregazione Somma
 Nome Mattino-Pomeriggio-Sera(etichetta) Esempio	 Nome Appartamenti Esempio Bragagni, Rossi, Bianchi, ...	Nome Misura Energia Tipo Unità di misura W Tipo di aggregazione Somma
 Nome Tariffa Energia Esempio		Nome Misura Volume Tipo Bolletta Unità di misura m Tipo di aggregazione Somma

Figura 24 - Step 3: definizione dei nomi delle colonne per dimensioni e misure

Al termine di questo step si ha la definizione di ciò che sarà memorizzato sul data warehouse: con la realizzazione del modello dei dati finale termina la parte di configurazione delle informazioni che influenzano il data warehouse. Gli step successivi includono la definizione di informazioni utili in fase di visualizzazione del dato: tali informazioni non influenzano la creazione del modello fisico del data warehouse.

5.2.4 Definizione delle misure calcolate

Lo step di definizione delle misure calcolate permette di aggiungere delle misure ulteriori allo schema del fatto; il modo per definire tali misure è costruirsi dei calcoli: gli operandi messi in relazione sono le misure ricavate nello step precedente. La gestione di queste informazioni influenza solo la parte di visualizzazione dei dati, visto che tali calcoli sono eseguiti nel momento in cui è richiesta la visualizzazione.

5.2.5 Dati generali

L'ultimo step richiede di assegnare una descrizione alla configurazione e di definire la periodicità nell'aggiornamento dei dati. Per adesso, il periodo di aggiornamento dei dati è solamente giornaliero, quindi il sistema richiede l'orario in cui effettuare la ricognizione dei dati aggiunti e inserirli adeguatamente nel data warehouse. Una volta terminato questo step il sistema prende in carico la creazione del data warehouse: le modalità con cui i dati sono recuperati e salvati sono descritte nella sezione del progetto di alimentazione del capitolo precedente.

5.3 Recupero dati dal data warehouse

Il progetto di alimentazione illustrato nel precedente capitolo si occupa di valorizzare il data warehouse: tale ambiente si compone di una strutturazione a classi, persistite in memoria attraverso la serializzazione di array di byte compressi in file. Come detto, i file contenenti informazioni generiche sono memorizzati in un unico file, mentre i dati di dettaglio contenenti riferimenti temporali, sono divisi sulla base del mese-anno a cui fanno riferimento. In questa sezione si descrive in quale modo sono recuperati i dati e come alimentano lo strumento di analisi utilizzato dall'utente.

A seconda del periodo di analisi che l'utente chiede di analizzare, il sistema recupera i file necessari alla ricomposizione dei dati: i file generici contenenti informazioni sulle dimensioni sono recuperati

ogni volta, mentre le tuple contenenti i valori delle serie temporali sono recuperate solamente se appartenenti al periodo di analisi richiesto.

Prima di iniziare un'analisi dei dati, l'utente deve sottoporre al sistema il set di dati da visualizzare: deve quindi identificare il data warehouse da interrogare scegliendo tra le configurazioni create, oltre ad impostare il periodo da visualizzare, composto da mese-anno di inizio e mese-anno di fine. Una volta fornite queste informazioni, il sistema prende in carico la richiesta e recupera i dati di configurazione dell'analisi d'interesse. Assieme a queste informazioni, introduce anche tutti i file mensili del periodo di visualizzazione richiesto: tali file risultano già compressi in fase di storicizzazione, così sono pronti per essere trasmessi da server a client, lasciando a quest'ultimo il compito di scompattarli. Una volta eseguita questa operazione, è necessario valorizzare l'oggetto utile a contenere tutte le informazioni d'interesse per l'analisi: per ogni tupla ricevuta dal server, si sostituiscono gli identificatori contenuti nelle dimensioni con le loro descrizioni; il mapping tra l'identificatore e la descrizione è definito nei file generici memorizzati per ogni analisi.

Una volta costruita la lista delle classi dinamiche, è possibile valorizzare l'interfaccia dedicata all'analisi dei dati. Lo strumento di visualizzazione dei dati si chiama *Pivot Grid* [TEL-PGR] ed è sviluppato dalla *Telerik*: questo controllo offre la possibilità di utilizzare le dimensioni dell'oggetto per eseguire delle analisi su di esso. Il numero delle dimensioni e quello delle misure dipende dagli attributi definiti nell'oggetto. Lo strumento permette le seguenti funzionalità:

- Definire le dimensioni e le misure fornendo una lista di elementi;
- Scomporre ogni misura temporale in tante dimensioni dipendenti ad essa: il tempo di rilevazione dei dati è scomposto in anno, quadrimestre, trimestre, mese, settimana, giorno della settimana, giorno e ora.
- Ricalcando le funzionalità di una tabella pivot, è possibile eseguire analisi specifiche, impostando i valori di righe e colonne sulla quale raggruppare i dati: ogni misura è raggruppata sulla base delle dimensioni definite dall'utente.
- Ogni volta che si esegue un'analisi, è possibile filtrare a monte i dati, impostando i valori da selezionare per ogni dimensione: si può scegliere di tenere in considerazione solo valori che rispettano determinate condizioni.

- Ad ogni raggruppamento è possibile indicare una o più funzioni di aggregazione dei valori interessati; quindi, per ogni misura si specifica in che modo raggruppare i valori che presentano le medesime caratteristiche.
- Le visualizzazioni dei dati sono disponibili in due modalità: tabellare e grafica. Nel formato tabellare i dati sono riportati con i loro valori numerici, mentre in quella grafica è lo strumento *Telerik* che dispone di diagrammi per rappresentare i dati in maniera sintetica; le tipologie di diagrammi messe a disposizione dal controllo sono ad istogrammi, a linee ed ad aree.

Prima di mostrare la visualizzazione dei dati interrogati, si esplicitano le attività di lavoro necessarie a personalizzare l'utilizzo del controllo nella maniera più naturale possibile. Infatti, sono state sviluppate molte personalizzazioni sia per carenze del controllo nativo, sia per la necessità di fornire funzionalità ad hoc, utili al miglioramento dello strumento. In particolare, le attività che hanno riguardato l'estensione di funzionalità del controllo sono:

- Ridefinizione di tutte le funzioni di aggregazione. Il motivo di questo sviluppo deriva dall'impossibilità del controllo di gestire dati non valorizzati. Infatti, la costruzione del fatto così come analizzata, può generare misure che non presentano alcun dettaglio, cioè senza valore. Il controllo non è in grado di gestire tale situazione, così trasforma le non-valorizzazioni in valori 0. Nella maggior parte delle funzioni di aggregazione presenti, la gestione di un valore nullo come 0 dà risultati errati. Ad esempio, nella funzione di minimo i valori che non andrebbero presi in considerazione sono trattati come 0: tale gestione cambia i risultati dell'aggregazione.
- Ridefinizione di tutti i filtri a monte dei valori delle dimensioni. Anche in questo caso, il motivo che ha obbligato ad intervenire con sviluppi specifici è dovuto al mancato supporto della lingua italiana. Infatti, i filtri su dati temporali richiedevano l'inserimento di date nel formato americano.
- Salvataggio dei parametri di visualizzazione del controllo. Grazie a questo sviluppo, l'utente può decidere di salvare tutte le impostazioni dell'analisi che sta conducendo: questa funzionalità permette di memorizzare una determinata analisi dei dati, così da poterla consultare per analisi successive.

- Espansione della scomposizione del dettaglio temporale. Per le dimensioni della tipologia data, il controllo forniva una scomposizione costituita da pochi dettagli, impossibilitando l'utente ad eseguire alcune analisi come ad esempio sui dati trimestrali. Questa gestione è stata rivista e ampliata.
- Inserimento di misure personalizzate. Uno sviluppo specifico ha permesso di realizzare la creazione di misure ulteriori rispetto a quelle di base, permettendo di definire calcoli personalizzati e riproporli come misure del fatto.

Dopo aver analizzato come sono ricostruiti i dati e quali sono gli sviluppi che hanno interessato questa parte del progetto, non resta che verificare come operare realmente sui dati, utilizzando lo strumento di analisi appositamente progettato.

5.4 Interfaccia di analisi dei dati

Nelle sezioni precedenti sono state definite le potenzialità dello strumento di analisi dei dati: le possibilità fornite dallo strumento sono l'unione di ciò che originariamente offre il controllo, assieme alle personalizzazioni sviluppate per coprirne le carenze funzionali, così da migliorarne le potenzialità.

In questo paragrafo si vuole mostrare come questo strumento di analisi porta effettivi vantaggi. In particolare, si definiscono delle domande, che pongono lo strumento a creare una configurazione capace di rispondere alle esigenze di analisi dei dati. Tutte le domande sono poste in relazione all'ambiente modellato dalla soluzione On.Energy: il sistema rappresenta una suddivisione di più condomini, ogni condominio è composto da piani ed ogni piano ha due appartamenti. Ogni condominio registra i consumi orari di energia elettrica, acqua e gas. La configurazione dei metadati è stata presentata nella sezione di configurazione del dominio del data warehouse.

5.4.1 Caso di analisi dei dati 1

La prima richiesta, a cui il cliente vuole rispondere attraverso lo strumento di analisi dei dati, è la seguente: *sono interessato a comparare i consumi di gas per i piani di tutti gli edifici, per i mesi da gennaio a giugno dell'anno 2013, raggruppando i dati in base alla stagione.*

La realizzazione della risposta richiede una serie di passi utili a configurare adeguatamente la visualizzazione dei dati. Come primo passo è necessario limitare la lettura dei dati al periodo richiesto dall'analisi; successivamente, si filtrano a monte i dati dei mesi d'interesse, si inserisce nelle colonne la dimensione stagione e nelle righe quella dei piani; la misura da selezionare è il consumo di gas; infine, s'impone la visualizzazione del grafico a istogrammi, con tipologia a stack.

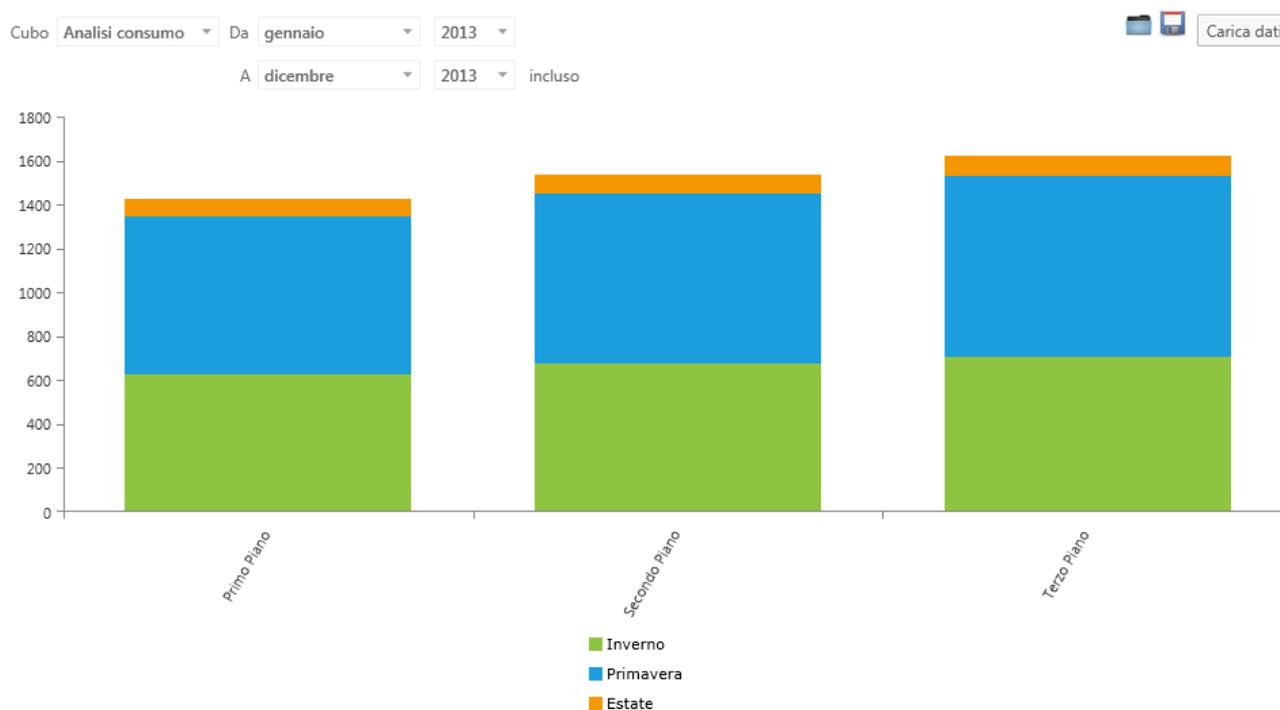


Figura 25 – Grafico ad istogrammi su stack dell'analisi 1

In *figura 25* è visibile il risultato dell'analisi dei dati. Le valutazioni che si possono fare osservando un grafico di questo tipo sono:

- In linea generale non è possibile fare delle comparazioni tra stagioni, dato che il filtro sui dati non tiene conto di un numero di giorni uguale per le tre stagioni in esame.

- Essendo il numero di appartamenti uguale per i tre piani, si può affermare che il consumo di gas aumenta al crescere del numero del piano: ciò può essere dovuto alla maggiore esposizione agli agenti atmosferici dei piani superiori rispetto a quelli inferiori, condizione che implica una maggiore dispersione di calore; tale fattore richiede un maggior consumo di gas per riscaldare adeguatamente il proprio appartamento.

È necessario specificare che la forte interattività che si ha con lo strumento permette di modificare la propria analisi in maniera continua, evolvendo in maniera interessante, analisi che presentano delle condizioni d'interesse per l'utente.

5.4.2 Caso di analisi dei dati 2

La seconda richiesta, a cui il cliente vuole rispondere attraverso lo strumento di analisi dei dati, è la seguente: *sono interessato a comparare il consumo medio orario di energia elettrica nell'estate del 2013, mettendo a confronto i giorni della settimana e raggruppando i valori in base alla fascia tariffaria d'appartenenza.*

Prima di passare alla configurazione dell'analisi è opportuno spiegare brevemente come si compongono le fasce tariffarie. Le fasce tariffarie nascono per distribuire maggiormente il consumo di energia elettrica durante tutta la giornata: il motivo deriva dal maggiore costo di produzione dell'energia quando essa è più richiesta: sostanzialmente, nelle ore centrali del giorno è richiesta molta più energia rispetto alle ore notturne. Così, per favorire un consumo maggiormente distribuito, si pensa di far pagare diversamente il costo dell'energia elettrica dividendola in tre fasce diverse: la F1 è la fascia tariffaria più costosa e viene applicata dal lunedì al venerdì dalle 8 alle 19, la F2 è la fascia di costo intermedio e viene applicata dal lunedì al venerdì negli orari 07.00-08.00, 19.00-23.00 e il sabato dalle 07.00 alle 23.00; invece per la F3 sono applicate le fasce orarie rimanenti con l'aggiunta dei giorni festivi. L'applicativo On.Energy permette di gestire le tariffe, grazie al quale è possibile costruirsi un'analisi in cui tutti gli istanti di rilevazione del consumo sono attribuiti ad una diversa fascia tariffaria.

Riesaminando la richiesta, la realizzazione di una risposta di questo tipo si compone di vari passaggi: prima di tutto è bene specificare che la cardinalità dei dati deve essere portata al dettaglio orario, impostando la somma come funzione di aggregazione. Successivamente si passa all'analisi dei dati: si caricano i dati sulla base temporale d'interesse, s'impostano le coordinate di analisi, che sono le fasce tariffarie e i giorni della settimana, infine si definisce come filtro a monte dei dati la stagione *estate*. La misura da selezionare è il consumo elettrico, ma la funzione di aggregazione da impostare sul controllo è la media.

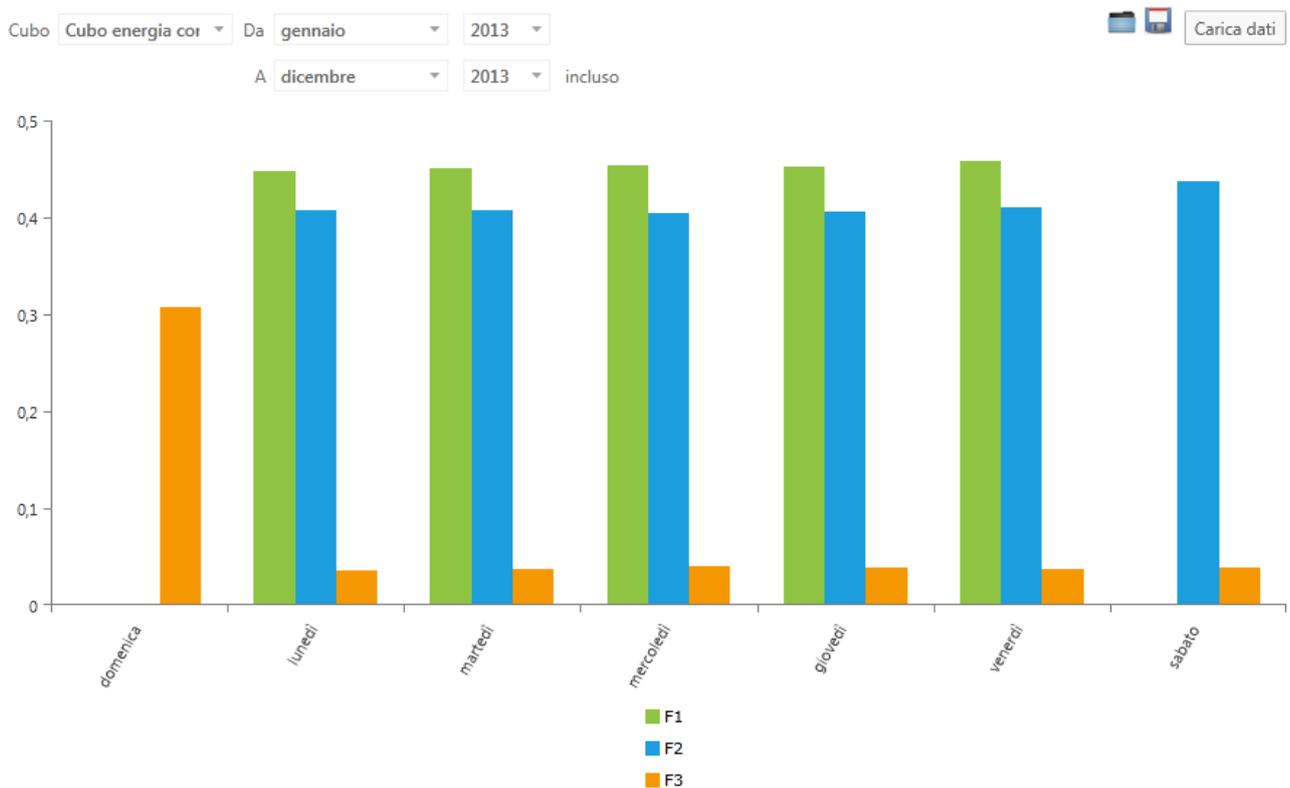


Figura 26 – Grafico ad istogrammi dell'analisi 2

In *figura 26* è visibile il risultato dell'analisi dei dati. Le valutazioni che si possono fare osservando il grafico sono:

- Il consumo medio orario della domenica presenta solamente la fascia F3: succede così che il valore medio include sia le ore diurne che quelle notturne, mostrando un valore medio più

basso rispetto alle altre fasce, dato che esse non comprendono orari notturni nelle loro definizioni.

- Anche il sabato risulta un giorno poco interessante, visto che è configurata la fascia F2 per gran parte della giornata.
- Il consumo delle fasce nei giorni feriali è sostanzialmente più interessante: la fascia F1 mostra un consumo medio molto più alto rispetto alle F3: questo denota uno scarso utilizzo di energia nelle ore notturne, dettaglio che incide negativamente sul totale della bolletta. Di poco inferiore il consumo di energia nella fascia F2 rispetto alla F1.

5.5 Analisi delle Prestazioni

In questa sezione si valuteranno le prestazioni del software, tenendo in considerazione la natura della base di dati del data warehouse. Infatti, come spiegato nei capitoli precedenti, i dati non appartengono ad un database definito con delle tabelle, ma sono la rappresentazione di classi serializzate su file. Quindi, l'interrogazione delle sorgenti non utilizzerà un protocollo d'interrogazione dati multidimensionali, ma si limiterà a scaricare per intero i dati d'interesse dal data warehouse, per poi mantenerli in memoria dove lo strumento di visualizzazione si occuperà di aggregarli per le varie interrogazioni.

L'analisi delle prestazioni condotta in questo paragrafo s'interesserà solamente delle fasi che richiedono maggiori risorse hardware per la loro computazione. Nel corso della progettazione, è stato possibile evidenziare 4 fasi che richiedono grosse computazioni sui dati: la prima specifica tutta la parte di alimentazione dei dati nel data warehouse, la seconda riguarda il trasferimento dei dati dal server al client, la terza comprende la decompressione e la materializzazione dei dati su rows, mentre la quarta e ultima fase si focalizza sulla manipolazione dei dati dello strumento di analisi.

La misurazione delle prestazioni del software valuterà nella maniera più completa le tempistiche di acquisizione e manipolazione dei dati, variando la quantità dei dati d'interesse. L'esempio che prendiamo in considerazione nell'analisi delle prestazioni si compone di tre livelli della gerarchia, 3 categorie, 1 tariffa e 5 stream diversi, ognuno composto da 200.000 rows. Analizzeremo le prestazioni dello strumento di analisi aggiungendo uno stream alla volta nella configurazione dei

metadati, in modo da monitorare le prestazioni al crescere della cardinalità del fatto. Per mantenere invariati i test sui diversi set di dati, il tempo di manipolazione dei dati dello strumento sarà cronometro sul raggruppamento della dimensione mese: è scelta proprio questa dimensione perché fa parte delle dimensioni ricavate a tempo di esecuzione, quindi non è un valore effettivamente memorizzato, perciò presenta dei tempi di manipolazione più lunghi.

Le quattro fasi cruciali per la valutazione delle prestazioni dello strumento sono misurate con delle stampe a schermo dei tempi di esecuzione. Inoltre, è opportuno specificare le caratteristiche dei pc con il quale viene fatto il calcolo delle prestazioni: per quanto riguarda il pc server, quello dedicato alla materializzazione dei dati su data warehouse, si avvale di un processore Intel Core 2 da 2.10GHz, mentre il pc client, che si occupa di richiedere i dati, materializzare le rows ed eseguire il raggruppamento, utilizza un processore Intel Core i3 2.53GHz. Entrambi i pc sono dotati di CPU a 64 bit e 4 GB di RAM. Sono connessi tramite Wi-Fi ad una LAN con router domestico.

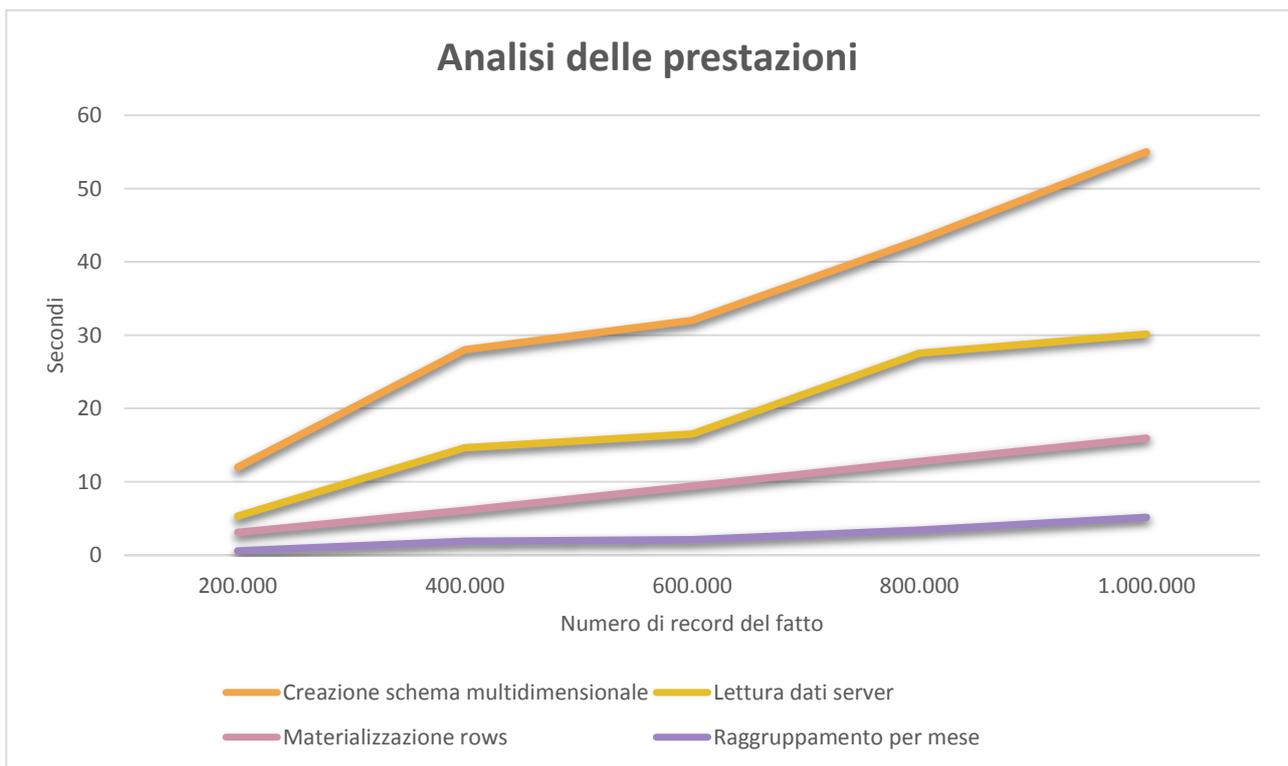


Figura 27 - Grafico a linee delle prestazioni dello strumento di analisi dei dati

Nella *figura 27* sono presentati i tempi di esecuzione delle fasi caratterizzanti il ciclo completo di analisi dei dati. In merito ai dati esposti, si possono aggiungere una serie di valutazioni importanti:

- La creazione del database è l'attività più onerosa in termini di prestazioni: si fa notare che tale operazione è fatta una sola volta e non incide sulla responsività dell'interfaccia del sistema.
- Le altre tre operazioni obbligano l'utente ad attendere il loro completamento. Di queste operazioni, quelle più onerose sono eseguite all'inizio dell'analisi, per caricare i dati; quindi, solamente la manipolazione dei dati in memoria è eseguita continuamente dall'utente.
- Tutte le operazioni hanno tempi che crescono in maniera lineare al crescere del numero di dati dell'analisi.

L'analisi delle prestazioni ha fornito delle risposte in linea con quanto si poteva dedurre in fase di realizzazione del sistema. La possibilità di definire i metadati implica indirettamente il numero di elementi da gestire ad ogni analisi: questo fattore è fondamentale perché permette all'utente di introdurre solamente dati d'interesse, riducendo il carico generale del sistema ed evitando di importare dati che non interessano.

5.6 Caso di studio Amadori

Il caso di studio oggetto di questa sezione è stato determinante nella definizione dei requisiti che hanno gettato le basi per la realizzazione di questa tesi: lo sviluppo di tali specifiche nasce dalla collaborazione del *team Energia* di Onit con il cliente Amadori.

L'azienda Amadori è una dei principali leader nel settore agroalimentare italiano, specializzata nel settore avicolo: alla base del successo c'è la decisione di supportare direttamente l'intera filiera integrata, fattore che gli permette di certificare il ciclo di lavorazione, dalla scelta della materia prima fino alla distribuzione del prodotto [AMA-AHP].

Lo scopo di questa sezione è di riportare le fasi salienti che hanno contraddistinto il ciclo di progettazione del nuovo strumento di analisi dei dati; saranno riportate le fasi che hanno maggiormente caratterizzato il percorso di progettazione dello strumento descritto in questa tesi.

5.6.1 Nuovi requisiti

On.Energy fa il suo ingresso nell'ambiente Amadori all'inizio di settembre del 2013: come detto, l'adozione di tale progetto mira a controllare il consumo energetico dell'azienda, monitorando i consumi (idrici, elettrici, ecc.) grazie all'utilizzo di specifico hardware, capace di comunicare al sistema On.Energy il valore dei principali contatori installati nei vari reparti.

Dopo una prima fase di configurazione, il sistema entra pienamente nel ciclo di vita della produzione e col passare dei mesi nascono nuove esigenze, a cui seguono analisi e nuovi sviluppi da parte di Onit. Ad esempio, è stato sviluppato un sistema di allarmistica capace di monitorare eventuali consumi anomali delle risorse: questo sviluppo permette di confrontare il consumo medio di risorse per ogni turno, evidenziando situazioni anomale. Nel tempo queste informazioni diventano indicazioni fondamentali: per esempio, aiutano ad identificare malfunzionamenti negli impianti di distribuzione delle risorse, come nel caso di perdite nell'impianto idrico.

Dopo questa prima fase, nasce l'esigenza di aggiungere nuove funzionalità al sistema. Infatti, oltre alle informazioni relative al consumo energetico, si ha la necessità di memorizzare informazioni relative alla produzione: sono introdotti nel sistema opportuni strumenti capaci di calcolare la quantità di prodotto lavorato. Nei successivi incontri tra Onit e Amadori, emerge la necessità di mettere in relazione i dati di consumo energetico con quelli di lavorazione.

La prima risposta per mettere in relazione il consumo energetico con la quantità di prodotto lavorato si concretizza con il concetto di consumo energetico medio: tale valore si ricava dal rapporto tra il consumo di una determinata risorsa e la quantità di prodotto lavorato. Dopo aver definito tale dato come nuovo requisito, Amadori lascia il compito ad Onit di analizzare il problema: si tratta di definire il metodo migliore per realizzare uno strumento di analisi dei dati, capace di mettere in relazione i dati del consumo con quelli della produzione.

5.6.2 Analisi del problema

Una prima analisi del problema consiste nella valutazione dei dati a disposizione, assieme ad una maschera idonea a permettere un'analisi dei dati. Sostanzialmente, i valori di consumo energetico e di produzione sono rappresentati da due stream: il primo calcola il consumo di energia elettrica, mentre il secondo storicizza un dato incrementale che conteggia le unità di prodotto lavorate. Onit valuta di utilizzare le tecniche di analisi dei dati già disponibili nel prodotto, introducendo solamente un elemento capace di mettere in relazione i valori di due stream.

Onit presenta la soluzione visibile nella *figura 28*. Per ogni istante di rilevazione, sono eseguiti i calcoli di consumo medio energetico tra i dati precedentemente analizzati (definito con il termine *resa* nella colonna in figura). Inoltre, è disponibile anche il raggruppamento del dato per ogni turno di lavorazione. L'utente ha così un calcolo del dato raggruppato sulla base del turno: la visualizzazione dei dati è solamente realizzata con una vista di tipo tabellare.

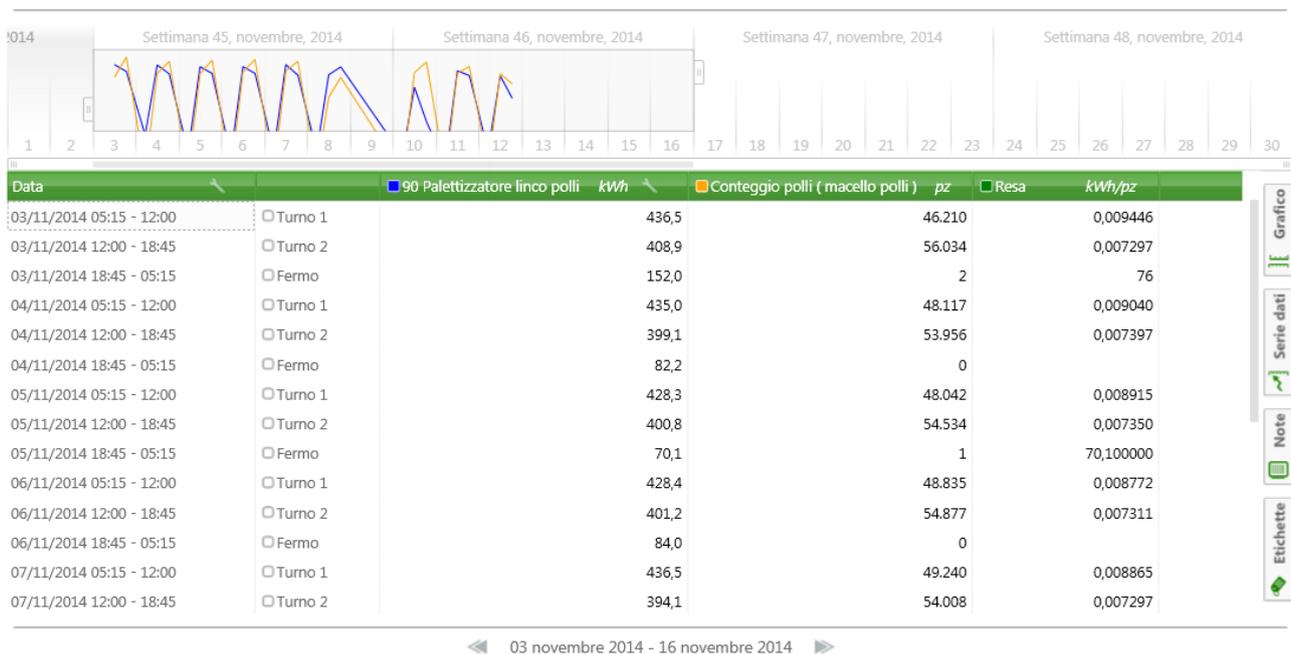


Figura 28 - Analisi della resa raggruppata per turno

Nei successivi incontri tra Onit e Amadori, si gettano le basi per espandere le possibilità di analisi dei dati del prodotto On.Energy. Tra tutte le richieste avanzate dal team Amadori, sono riportate quelle maggiormente interessanti:

- Si vuole analizzare il dato sulla base delle altre etichette memorizzate dal sistema, ad esempio sulla base della dimensione del prodotto lavorato; infatti, in ogni reparto si associano tre tipologie di etichette: i valori possibili sono piccolo, medio e grande.
- Lo strumento deve anche poter mettere in relazione i valori di più etichette diverse; ad esempio, in alcune analisi è interessante visualizzare i dati mantenendo sia le informazioni del prodotto lavorato, sia quelle relative al turno di lavorazione.
- Si vuole analizzare il consumo medio anche per serie temporali diverse, come possono essere quelle relative agli stream che rilevano, ad intervalli regolari, il consumo di acqua.
- È importante puntualizzare che ci sono più stream che memorizzano la medesima tipologia d'informazione, ma riguardanti zone diverse: è opportuno prevedere analisi che permettano di sommare più serie temporali diverse, così da calcolare i consumi totali di un determinato impianto o reparto.
- Si vuole analizzare come il consumo medio cambia nel corso del tempo: analisi interessanti possono essere condotte confrontando i valori ad un dettaglio temporale minore: ad esempio, i consumi energetici medi raggruppati per trimestre.

Dopo aver sottoposto tali requisiti ad Onit, il team Energia analizza il problema e formula una prima conclusione: continuare ad inserire modifiche alle maschere esistenti non porterebbe a rispondere a tutti i casi d'uso elencati. Inoltre le modifiche agli attuali strumenti di analisi dei dati, aggiungendo scenari non previsti in fase di progettazione, aumenterebbero i costi di manutenzione del prodotto.

Onit decide di costruire un nuovo strumento di analisi dei dati: infatti, per rispondere ai nuovi requisiti di analisi, è necessario avvalersi di architetture di data warehouse. Inoltre, emerge l'esigenza di realizzare il nuovo progetto all'interno del contesto On.Energy; questa soluzione deriva essenzialmente da due motivazioni: per primo si vuole fornire all'utente uno strumento integrato, facile da usare e direttamente accessibile dal prodotto; il secondo è motivato dal fatto che l'ambiente On.Energy dispone di una quantità tale di informazioni, che se opportunamente coese, forniscono un valore aggiunto e completo per eseguire un'appropriata analisi dei dati.

Nasce così l'obiettivo di progettare quanto definito in questa tesi: nella prossima sezione si evitano di riportare le fasi di analisi (descritte nel capitolo 3 e 4), ma si focalizza l'attenzione su come lo strumento è stato capace di distinguersi del dominio di Amadori, costituendo un valore aggiunto nell'analisi dei dati.

Successivamente alla sua realizzazione, lo strumento di analisi dei dati realizzato dalla tesi è stato rilasciato all'interno del prodotto On.Energy in Amadori. Anche se la versione del controllo non è aggiornata da tempo, lo strumento ha permesso di raggiungere lo scopo per cui è stato ideato, cioè eseguire le analisi dei dati definite nel paragrafo precedente.

5.6.3 Benefici del nuovo strumento di analisi

Riportiamo un esempio di analisi dei dati, realizzata con il supporto dello strumento nel dominio di Amadori. Il requisito a cui si vuole rispondere è la seguente: *analizzare l'andamento mensile del valore del consumo energetico medio (energia elettrica / numero di pezzi lavorati) per l'anno 2014 (gennaio-ottobre), limitando i dati al turno 1 e alla lavorazione del prodotto Grande. Tale valutazione deve prendere in considerazione l'area Macello del reparto Pollo dello stabilimento di Cesena.*

La *figura 29* mostra la realizzazione del grafico a linee corrispondente al requisito sottoposto. A prima vista il valore del consumo energetico medio mostra un andamento fortemente altalenante. Sono proposte di seguite alcune valutazioni:

- Il costo in termini di energia elettrica necessario a produrre un pezzo risulta avere il suo picco nel mese di luglio: questo dettaglio porta a pensare che la temperatura esterna incida direttamente sul costo dell'energia elettrica, valutazione confermata dalla necessità di mantenere nei reparti una temperatura refrigerata per le fasi di lavorazione.

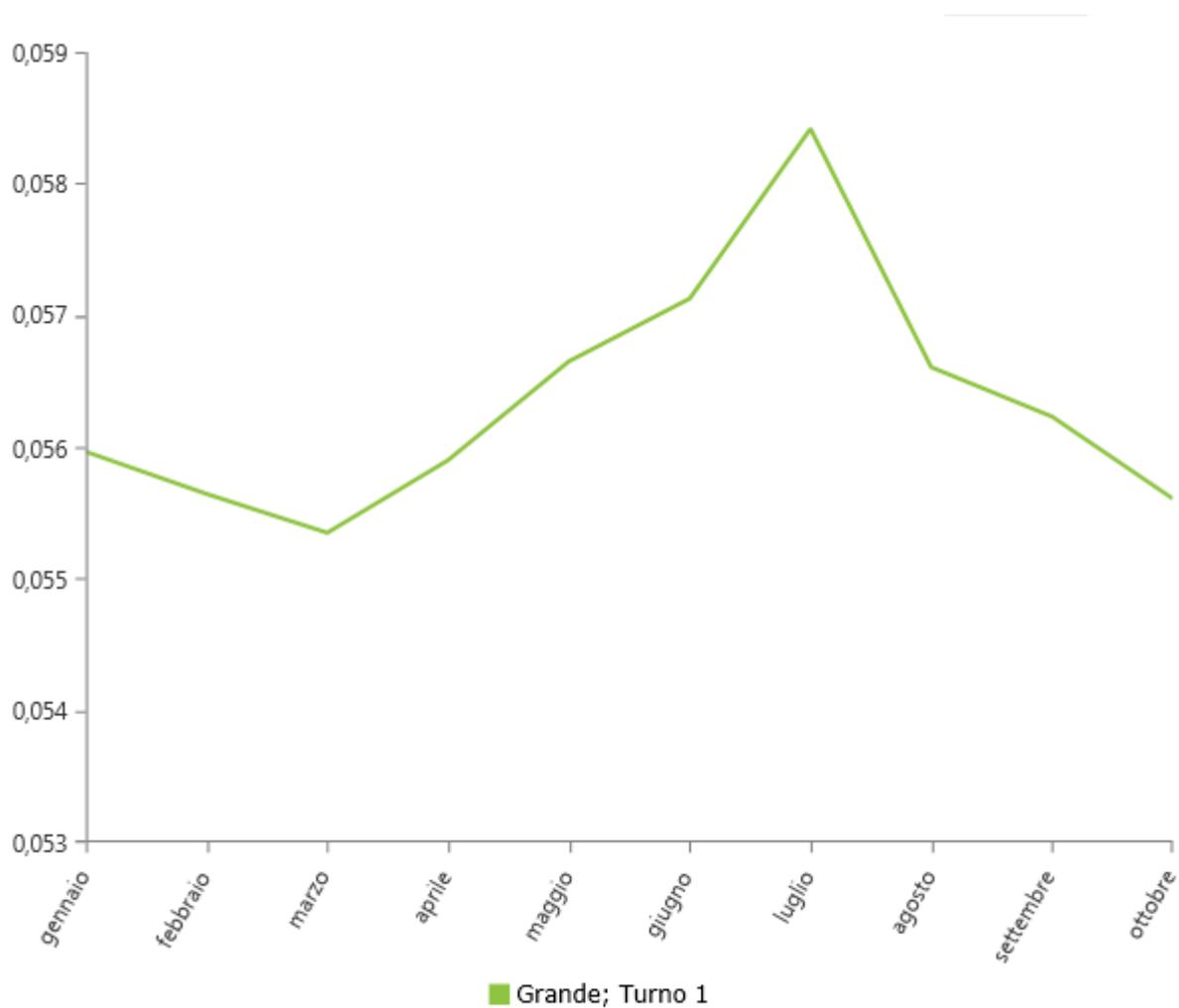


Figura 29 - Trend consumo energetico medio su base mensile

- È presente una notevole differenza tra il mese di marzo e gli altri mesi dell'anno: in merito a questo dato, è necessario approfondire maggiormente quali siano le motivazioni che fanno emergere un risultato di questo tipo.

Il grafico del consumo energetico mette a confronto le possibilità di analisi del nuovo strumento con quanto precedentemente offerto dal prodotto: l'analisi svolta è solamente una tra le infinite che l'utente ha la possibilità di costruirsi autonomamente.

Per effettuare un giudizio più analitico, è proposta la *tabella 1*: si confrontano le possibilità di analisi offerte precedentemente da On.Energy rispetto allo strumento progettato e realizzato in questa tesi.

Funzionalità di analisi dati di On.Energy **Strumento di analisi dati oggetto della tesi**

Analisi con stream, indicatori, categorie, tariffe e struttura gerarchica	<i>Impossibile eseguire analisi con tutte le informazioni di On.Energy</i>	<i>Tutti i principali concetti possono essere introdotti all'interno di qualunque analisi senza dover sviluppare funzionalità ad hoc</i>
Grado di flessibilità dello strumento	<i>Il numero dei dati da inserire in analisi è cablato nel codice</i>	<i>I concetti da inserire in analisi sono impostati in fase di configurazione dei metadati</i>
Filtri sui dati	<i>Filtri sui dati non impostabili</i>	<i>Ogni dimensione può essere filtrata, definendo un sotto-insieme di valori da includere nell'analisi</i>
Raggruppamenti sui dati	<i>Le possibilità di raggruppare i dati sono cablate dagli strumenti</i>	<i>La forte interattività con lo strumento permette di raggruppare i dati su qualsiasi dimensione definita</i>
Creazione di informazioni derivate utilizzabili per l'analisi	<i>Livello di dettaglio dei dati non modificabile</i>	<i>Livello di dettaglio sui dati definibile in ogni momento, con possibilità di poter operare su informazioni derivate dalla dimensione tempo</i>
Inserimento di nuove misure	<i>Le operazioni che si possono effettuare sui dati sono cablate dal codice, oppure create dalle nuove serie temporali per effettuare calcoli personalizzati</i>	<i>Attraverso la definizione delle misure calcolate è possibile creare nuove misure del fatto</i>

Tabella 1 - Benefici introdotti dall'attuale strumento di analisi dei dati

In conclusione, si può affermare che la progettazione dello strumento realizzato in questa tesi è una scelta vincente: sicuramente lo strumento realizzato mostra buoni margini di flessibilità,

permettendo di rispondere ad un numero di domande potenzialmente infinito per quanto riguarda l'analisi dei dati. L'analisi realizzata ha permesso di contenere i costi relativi alla manutenzione del software: infatti, se Onit non avesse cambiato il metodo di approccio nell'implementazione di analisi dati, probabilmente i costi sarebbero cresciuti esponenzialmente nel tempo.

Conclusioni e sviluppi futuri

La progettazione dello strumento oggetto della tesi ha permesso di approfondire il campo delle tecnologie SCADA/EMS, realizzando una funzionalità di analisi dei dati sviluppata con le tecniche di business intelligence. La progettazione dello strumento ha dovuto colmare una lacuna funzionale di On.Energy: collegare le principali informazioni della base dati; la nuova funzionalità utilizza la modellazione multidimensionale per crearsi un nuovo schema dall'attuale dominio.

Una volta concluso il progetto, si è condotto un preciso lavoro di analisi dei dati per misurare la validità dello strumento in termini di supporto all'utente. Gli eccellenti risultati dello strumento sono stati confermati da fonti diverse. Il primo risultato che indica l'effettiva utilità dello strumento, è la comparazione rispetto agli attuali limiti di analisi dati di On.Energy. Le possibilità di creare autonomamente il proprio schema e di definire i termini della propria analisi sono innovative; nessuno strumento di analisi dei dati del prodotto On.Energy permette tanta flessibilità nelle mani dell'utente. Il secondo risultato che conferma la validità del progetto, è la scelta da parte di Onit di introdurlo nella soluzione software del prodotto On.Energy. Il terzo e ultimo risultato che convalida la capacità dello strumento di rispondere ad esigenze reali, è rappresentato dal caso di studio condotto sull'azienda Amadori: tale approfondimento ha permesso di misurare la proprie potenzialità direttamente sul campo, facendo emergere informazioni che le altre analisi dei dati non potevano certamente mostrare.

La progettazione di questo strumento ha aperto le porte ad ulteriori sviluppi, incentivati dai buoni risultati dimostrati nella tesi. Infatti, sempre nel contesto del dominio On.Energy, si sta pensando di utilizzare la modellazione multidimensionale anche per realizzare un'analisi dati degli allarmi: essi presentano un fenomeno che si evolve nel tempo, a cui è possibile correlare un grosso insieme di informazioni. Un altro sviluppo può interessare il sistema di materializzazione dei dati: si potrebbe progettare un sistema dinamico basato sulla creazione di tabelle, così da sfruttare le tecniche di interrogazione di dati multidimensionali, anziché la struttura a file proposta in questa tesi.

Oltre alla progettazione del sistema di creazione del data warehouse, altri sviluppi possono interessare la parte di visualizzazione dell'analisi dei dati. In particolare, può essere interessante realizzare interfacce fruibili dai dispositivi mobili, così da migliorare ulteriormente il prodotto On.Energy in termini di compatibilità. Inoltre, può essere interessante utilizzare una progettazione che permetta di esportare i dati d'interesse in formato multidimensionale, demandando il compito di supportare l'analisi dei dati ad uno strumento già presente sul mercato.

Bibliografia

[ABB-AHP] ABB, *ABB Homepage*, <http://www.abb.com/>, 2014

[ABB-ESO] ABB, *Energy System Operation*,

[http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/08dc5d47449c7b30852575fa0055d0b9/\\$file/BR_Network_Manager.pdf](http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/08dc5d47449c7b30852575fa0055d0b9/$file/BR_Network_Manager.pdf), 2014

[ABB-NMS] ABB, *Network Manager Services*,

[http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/bbccc97cad590ebc12577070021a88f/\\$file/BR_Network_Manager_Service.pdf](http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/bbccc97cad590ebc12577070021a88f/$file/BR_Network_Manager_Service.pdf), 2014

[ABB-SEA] ABB, *SCADA/EMS and SCADA/GSM*,

[http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/0d9220cf797fa8a0852575fa0057038d/\\$file/BR_SCADA_EMS_GMS.pdf](http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/0d9220cf797fa8a0852575fa0057038d/$file/BR_SCADA_EMS_GMS.pdf), 2014

[AMA-AHP] Amadori, *Amadori Home page*, <http://www.amadori.it/>, 2014

[DSA-WIS] A. Daneels, W. Salter, *What is SCADA?*, in *International Conference on Accelerator and Large Experimental Physics Control Systems*, 1999

[GOL-DFD] Matteo Golfarelli, *Modellazione funzionale con data flow diagram*,

<http://bias.csr.unibo.it/golfarelli/SISPEC/dispense/DFD.pdf>, 2014

[MDF-DSS] Marek J. Druzdzel and Roger R. Flynn, *Decision Support Systems*, University of Pittsburgh, Pittsburgh PA, 2002.

[NEX-EEE] NextVille, *Eolico: Efficienza e prestazioni dei generatori eolici*,

<http://www.nextville.it/index/524>, 2014

[ONI-OHP] Onit Group s.r.l., *Onit Home Page*, <http://www.onit.it/>, 2014

[RIZ-AWI] Stefano Rizzi, *What-If Analysis*, <http://www-db.deis.unibo.it/~srizzi/PDF/eds-WIA.pdf>, 2014

[SCA-BIM] Stefano Cazzella, *Business Intelligence Modeler*,

<http://caccio.blogdns.net/software>, 2014

[SIR-GPM] SIRIUS, *GPM System*, <http://www.sirius.to.it/it/files/SIRIUS%20-%20Catalogo%202011-03.pdf>, 2014

[SIR-SHP] SIRIUS; *Sirius Home Page*, <http://www.sirius.to.it/it/index.html>, 2014

[TEL-PGR] Telerik, *Pivot Grid*, <http://demos.telerik.com/silverlight/#PivotGrid/FirstLook>,
2014

[TDT-KRO] Ralph Kimball, Margy Ross, *The Data Warehouse Toolkit*, Wiley Computer
Publishing, 2nd edition, 2002