

CAPITOLO 3. STRUTTURE DATI SOTTOSTANTI E ARCHITETTURE DI RIFERIMENTO

3.1 QUALITÀ DEL DATO: DATA, INFORMATION E KNOWLEDGE

Nella letteratura classica che si occupa della disciplina dei DSS, tre definizioni chiave descrivono il processo evolutivo del decision making: Data, Information e Knowledge. Nel prendere decisioni qualitative e quantitative, data, information e knowledge sono i pilastri e la base per condurre in modo sicuro i decisori alla corretta soluzione o azione da prendere.

- **Dati.** I dati sono fatti o eventi elementari che vengono regolarmente registrati con procedure automatiche e vengono immediatamente resi disponibili all'interno dell'organizzazione aziendale. I dati item sono utilizzati per qualificare e quantificare le azioni di business più operative.
- **Information.** L'analisi e la sintesi dei dati conducono all'informazione che ha la funzione di catturare lo stato dell'organizzazione aziendale rispetto alle attività prefissate, oltre ovviamente a supportare le esigenze manageriali rispetto a obiettivi e budget stimati.
- **Knowledge.** La combinazione tra data e informazione è utilizzata per attivare processi semplici e complessi di decisione nella forma di modelli quantitativi e qualitativi che portano alla knowledge, ossia la conoscenza profonda dell'organizzazione aziendale e del business previsto.

Risulta necessario suddividere i concetti di Data, Information e Knowledge e metterli a disposizione del processo decisionale che segue rispettivamente intelligenza, design e decisionale. Le attività quotidiane di un'organizzazione generano fatti o dati che devono essere registrati correttamente. Questi fatti o dati possono essere analizzati e sintetizzati in informazione che serve a fotografare attraverso opportuni indicatori la performance dell'azienda.

In conclusione, fissati i pilastri della trasformazione del dato in conoscenza risulta necessario considerare la qualità del dato e dei processi che guidano al risultato. Se il dato risulta non corretto in fase di registrazione, è prevedibile che la conoscenza che ne scaturisce diventi anch'essa non corretta, e il risultato poco stabile e comporta scelte sbagliate che portano talvolta a ingenti danni aziendali.

3.2 IL RUOLO DEL BUSINESS ANALYTICS ALL'INTERNO DEI SISTEMI DI INFORMAZIONE DELL'ORGANIZZAZIONE: L'APPROCCIO DATA-DRIVEN

I modelli analitici che si basano sulla teoria dell'informazione e del dato sono rispettivamente:

- Modelli descrittivi che sono definiti come un set di relazioni, indicatori di performance e algoritmi matematici che aiutano ad analizzare il comportamento di un'impresa o sistema. Sono la base dei modelli di supporto alla decisione.
- Modelli normativi che costituiscono la base delle decisioni quantitative che si fondano su elementi teorico-razionali, tali modelli non entrano nel merito dello sviluppo tecnologico-algoritmo.
- Modelli prescrittivi che coinvolgono analisi computazionale e sviluppo di problemi portati avanti da intelligenza umana che applicano intuito e giudizio, quindi la regola decisionale è impostata a priori.
- Modelli decisionali che sono la categoria derivata dai modelli normativi e modelli prescrittivi. Sono modelli intelligenti e automatizzati, forniscono la soluzione più adatta al problema presentato, agendo in modo attivo e presentando risultati attuabili; a loro afferiscono i nuovi modelli di machine learning e artificial intelligence.

All'interno del contesto organizzativo di impresa, lo sviluppo di tali modelli e il loro ruolo all'interno di un'infrastruttura IS è riconosciuto come Business Analytics. I sistemi di supporto alla decisione stanno velocemente assumendo un ruolo fondamentale in molti settori di mercato e di organizzazioni pubbliche. Nel contesto dei sistemi informativi aziendali i modelli quantitativi costituiscono un'importante componente all'interno dell'Information Value Chain. Tipicamente, i dati operazionali che sono disponibili in un'azienda sono analizzati, preparati e consolidati per creare dati di analisi o informazione, che compone il primo anello della IVC. I dati analitici vengono memorizzati in data mart, che sono dei sottoinsiemi della completa collezione dei dati a disposizione di un'azienda, chiamata datawarehouse.

3.3 SISTEMI DI MODELLAZIONE E DATABASE

I technical expert hanno accesso a tutti i componenti di sistema che includono i modelli sottostanti, i database e le piattaforme; quindi, possono variare strategia e scenari a seconda dei risultati ottenuti. I differenti technical expert collaborano a definire l'applicazione specifica di dominio integrando molteplici tool all'interno delle differenti fasi di processo di intelligenza di business. I decision maker sono gli attori che utilizzano le piattaforme e i modelli.

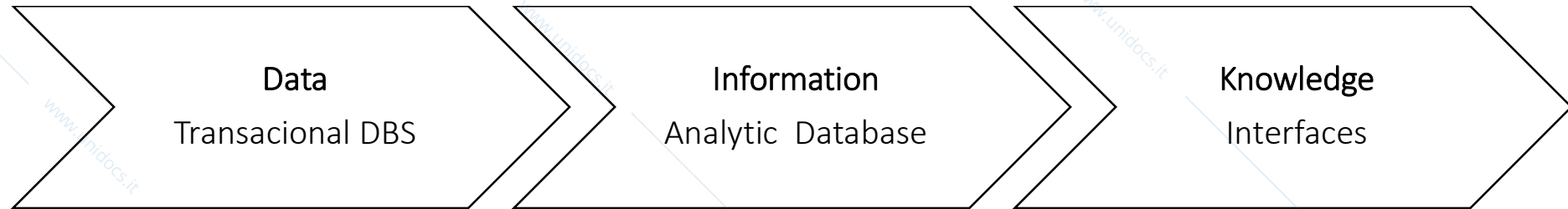
Tipicamente i decisori hanno accesso ai dati e le routine di analisi disponibili nei database, utilizzano quindi modelli sottostanti e ne valutano le performance e i risultati. Essi sono esclusivamente interessanti ai valori di output, che sono utilizzati ed espressi in termini decisionali. Una volta validati gli output e le decisioni proposte dal processo analitico, analizzano e creano dei dati decisionali all'interno di altri data mart; per raggiungere l'obiettivo prefissato, risulta necessario che le architetture di sistema prevedano l'integrazione delle piattaforme in forma stabile e a supporto di tutti gli attori previsti nel processo. L'architettura proposta si basa su sistemi di modellazione algebrica, la quale processa uno o più modelli algebrici per generare l'istanza dei modelli da elaborare. Una volta generata l'istanza dei modelli, la piattaforma tecnologica ha il compito di generare il risultato e storicizzarlo nel database.

3.4 DATI TRANSAZIONALI, DATI ANALITICI E INFORMATION VALUE CHAIN

Nel contesto di un'impresa, il ruolo delle piattaforme di gestione dato e di informazione è diventato uno dei pilastri fondanti dell'organizzazione dei processi aziendali. I dati utilizzati in azienda si classificano in due gruppi: dati transazionali e dati analitici. tipicamente i sistemi che comprendono i database transazionali sono ottimizzati per gestire le performance e l'efficienza di gestione dei valori in ingresso. I database analitici sono invece fortemente focalizzati verso l'utilizzo, l'analisi e il supporto alle decisioni. I dati analitici sono solitamente registrati e storicizzati in datawarehouse; questi repository subject oriented contengono informazioni non volatili orientate all'analisi della conoscenza per agire su decisioni in modo rapido e concreto. I dati contenuti nei datawarehouse sono modellati su schemi multidimensionali; questo approccio facilita le estrazioni e rendono rapidi i processi che eseguono query sulle suddette strutture, le quali mirano ad aggregare l'informazione su molteplici dimensioni per ottenere insight utili alle decisioni di business.

Una particolare prospettiva sul dato e le infrastrutture che lo gestiscono è la information value chain che parte dall'analisi e sintesi dei dati transazionali per arrivare a creare dati analitici o informazioni utili.

Figura 3.1 Information Value Chain



Data modeling, decisional modelling e model investigation sono fasi logiche che giocano un ruolo importante sia nell'interazione tra i tool di informazione sia nelle piattaforme di decisione; entrambe le strutture portano all'analisi del business che ricopre il ruolo dell'obiettivo ultimo da raggiungere. È facilmente intuibile che data modelling e decision modelling interagiscono in modo attivo e utile tra di loro.

- Data modeling si riferisce alla rappresentazione strutturata interna ed esterna dei fatti registrati nelle tabelle. Questo fornisce al decision maker l'informazione sul problema di decisione. Tipicamente data modelling coinvolge le relazioni tra i dati e il modello relazionale o identifica le categorie che sono utilizzate per definire le tabelle multidimensionali che portano al modello dati multidimensionali.
- Decision modelling è lo sviluppo di un modello o un set di modelli che cattura sia le decisioni che le strutture dati rispetto a un dato problema. Questi modelli sono utilizzati per valutare possibili decisioni o azioni in un determinato dominio di problemi e inoltre il probabile outcome di queste stesse azioni. Decision modelling include lo sviluppo dei modelli che sono utilizzati per il decision making.
- Model analysis e investigation si riferisce all'istanza di un modello con strutture dati e la valutazione sia dei parametri del modello che dei risultati per trovare la stabilità del modello stesso. Model analysis e investigation è spesso un'analisi descrittiva dei risultati ottenuti sulla quale vengono erogati degli insight di business o conoscenza rispetto a un dato problema.

3.4.1 DECISION DATABASE

Si riscontra facilmente che ogni sistema di supporto alle decisioni, modello dati, modello simbolico e tool algoritmico interagiscono tra loro in modo integrato. In ogni DSS – Decisional Support System esiste una forte correlazione e complicità tra i modelli dati, modelli algebrici e algoritmi, anche comunemente denominati motori di inferenza. In generale, è possibile considerare il decisional database come data mart, nel quale i modelli dati forniscono informazioni sul problema in questione, i modelli di decisione descrivono il problema di decisione o le relazioni tra l'informazione conosciuta e gli algoritmi che generano il set di soluzioni ottimali del problema in questione. Nelle applicazioni di programmazione stocastica, il decision database consiste in tre macro-tipologie di informazione:

- informazione deterministica o informazione invariata rispetto all'incertezza del futuro
- informazione stocastica ovvero possibili caratteristiche dell'incertezza legata al futuro e l'informazione che è soggetta all'incertezza
- soluzione o meglio possibili azioni ottimali nel futuro incerto

In sostanza, un optimisation based DSS è un sistema informativo contenente differenti applicazioni, le quali vengono utilizzate e combinate per individuare un ventaglio di soluzioni realizzabili rispetto a un particolare problema. Tra le applicazioni utilizzate ci possono essere anche sistemi analitici addizionali che spesso sono utilizzati per fornire ulteriori insight al problema. Questi tool addizionali possono consistere in On Line Analytical Processing, OLAP, che possono aiutare l'analista a raggiungere in modo rapido ed efficiente descrizioni complete del dato, per poter studiare il problema sotto esame e generare i report più appropriati per esporre la soluzione a terzi.

3.5 OLAP: VISTE MULTIDIMENSIONALI E DATACUBE

La tecnologia OLAP viene riconosciuta come un promettente e consolidato approccio per l'analisi e la navigazione dei datawarehouse e dati multidimensionali. I sistemi OLAP abilitano potenti sistemi di supporto alla decisione basati su analisi multidimensionali di grossi volumi di dati aggregati, comunemente presi da differenti sorgenti dati transazionali e database. Il modello dati multidimensionale OLAP e le tecniche proprietarie di aggregazione del dato organizzano ed espongono grossi volumi di dati per facilitare una veloce valutazione delle osservazioni utilizzando tool grafici e online in modo semplice.

I dati OLAP sono organizzati in cubi multidimensionali contenenti valori e misure che sono caratterizzati da un numero di dimensioni gerarchizzate. Le tipiche operazioni sui cubi di dati includono:

- roll-up, incrementando il livello di aggregazione
- drill-down, decrementando il livello di aggregazione o incrementando il dettaglio sulle diverse gerarchie
- slice-and-dice, selezione e proiezione
- pivot, re-orientamento della vista multidimensionale del dato

Per costruire efficienti sistemi di supporto della decisione attraverso la presenza di consistenti volumi di dati e datawarehouse che utilizzano query complesse, risulta necessario presentare e processare il dato in modo strutturato. Inoltre, i tool di supporto alla decisione necessitano di integrare paradigmi di modellazione, metodi di intelligenza artificiale e tecnologie database/multidimensionali.

L'architettura logica di un modello OLAP è basata su IVC: l'OLAP viewer è utilizzato per presentare e analizzare dati grezzi, dati analitici e dati decisionali.

L'approccio multidimensionale offre un discreto vantaggio rispetto alle tradizionali tipologie di database management, includendo applicazioni automatizzate di funzioni di aggregazione. In generale i cubi OLAP sono rappresentazioni multidimensionali del dato, nella quale ogni attributo di una data entità è considerato una dimensione separate. Un cubo OLAP consiste in una collezione di cuboidi, ciascuno rappresentante un determinato livello o gerarchia. Le procedure di aggregazione del cubo computano funzioni statistiche per determinati set di valori all'interno di ogni cuboide.

Figura 3.2 OLAP e Information Value Chain

