



La suite Pentaho Community Edition

Enzo Strazzera





Cosa è la Business Intelligence

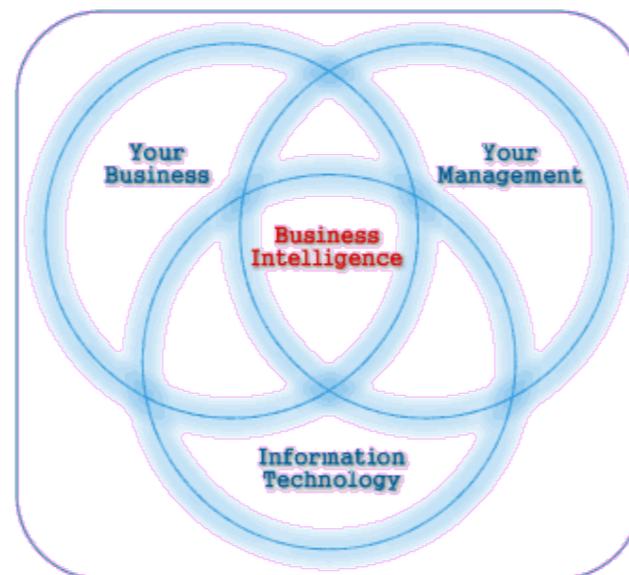
Con la locuzione **business intelligence (BI)** ci si può solitamente riferire a:

un insieme di processi aziendali per raccogliere ed analizzare informazioni strategiche.

la tecnologia utilizzata per realizzare questi processi,

le informazioni ottenute come risultato di questi processi.

Ogni sistema di business intelligence ha un obiettivo preciso che deriva dalla vision aziendale e dagli obiettivi della gestione strategica di un'azienda.





La Business Intelligence

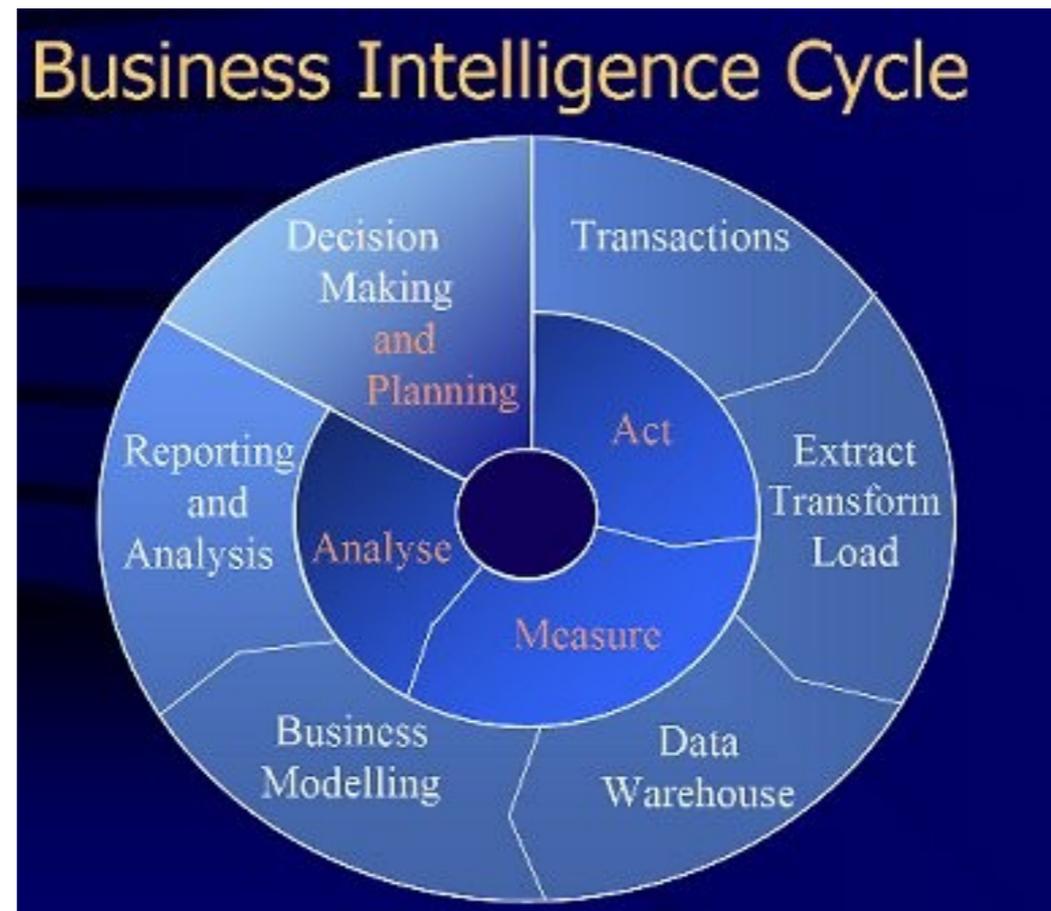
E' un insieme di modelli matematici e di metodologie di analisi che esplorano i dati per ricavare informazioni e conoscenze utilizzabili nel corso dei processi decisionali. (Carlo Vercellis)





A cosa serve la Business Intelligence

Attraverso i sistemi di BI intendiamo fornire ai diversi portatori di interesse informazioni mirate: si definiranno i livelli di accesso e la tipologia di dati che intendiamo distribuire.





A chi serve la BI

Enti, imprese, associazioni ... tutte quelle entità complesse che fanno uso di sistemi di raccolta ed elaborazione di dati informatizzata.

I diversi livelli di management, la forza di vendita, ma anche clienti o utenti esterni possono trarre vantaggio dalla rappresentazione dei dati e acquisire informazioni sulla realtà a cui si riferiscono

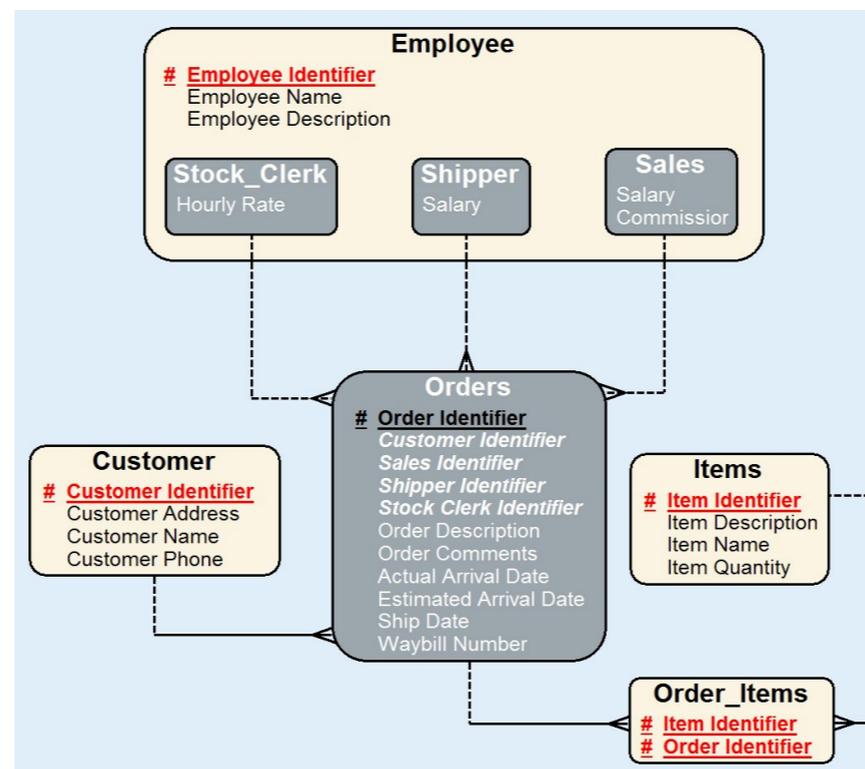




La BI - come usarla in azienda

Report, cubi Olap, Dashboards, sono strumenti sempre più diffusi e sono il risultato di un processo di Business Intelligence.

Occorre in primo luogo avere un modello di business: estrapolare i dati significativi, metterli in relazione in uno schema teorico, costruire quindi un modello di dati attraverso cui ottenere gli indicatori essenziali e produrre gli elaborati finali

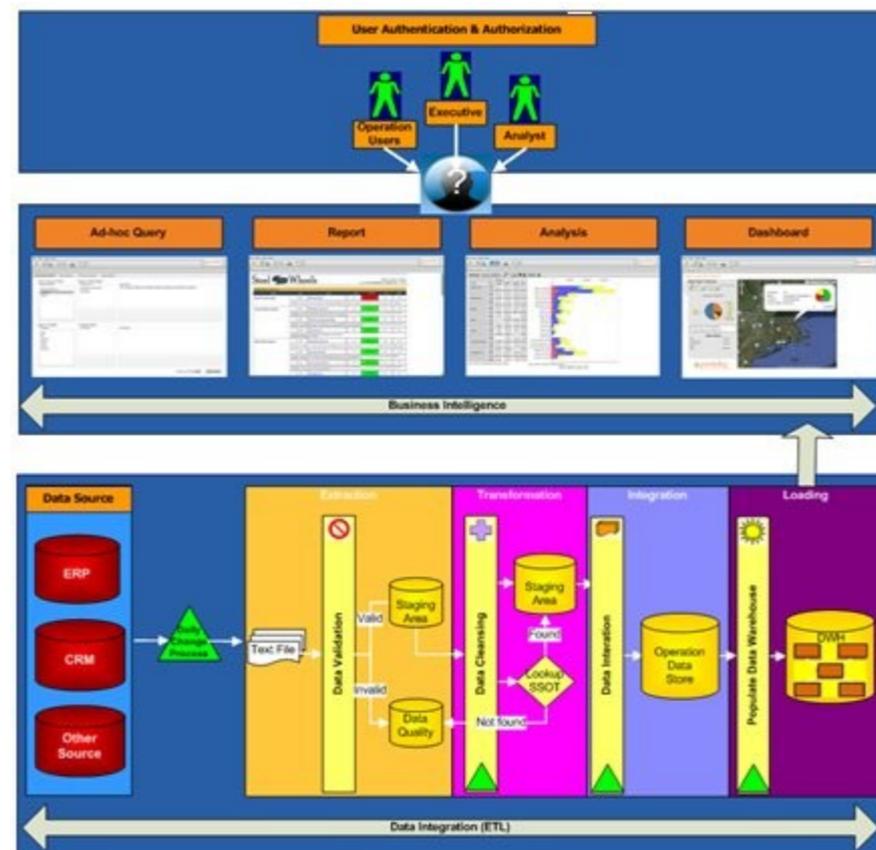




Strumenti della BI

Olap (on line analytical processing) è in sintesi una matrice di dati (o tabella pivot multidimensionale) che può essere interrogata interattivamente

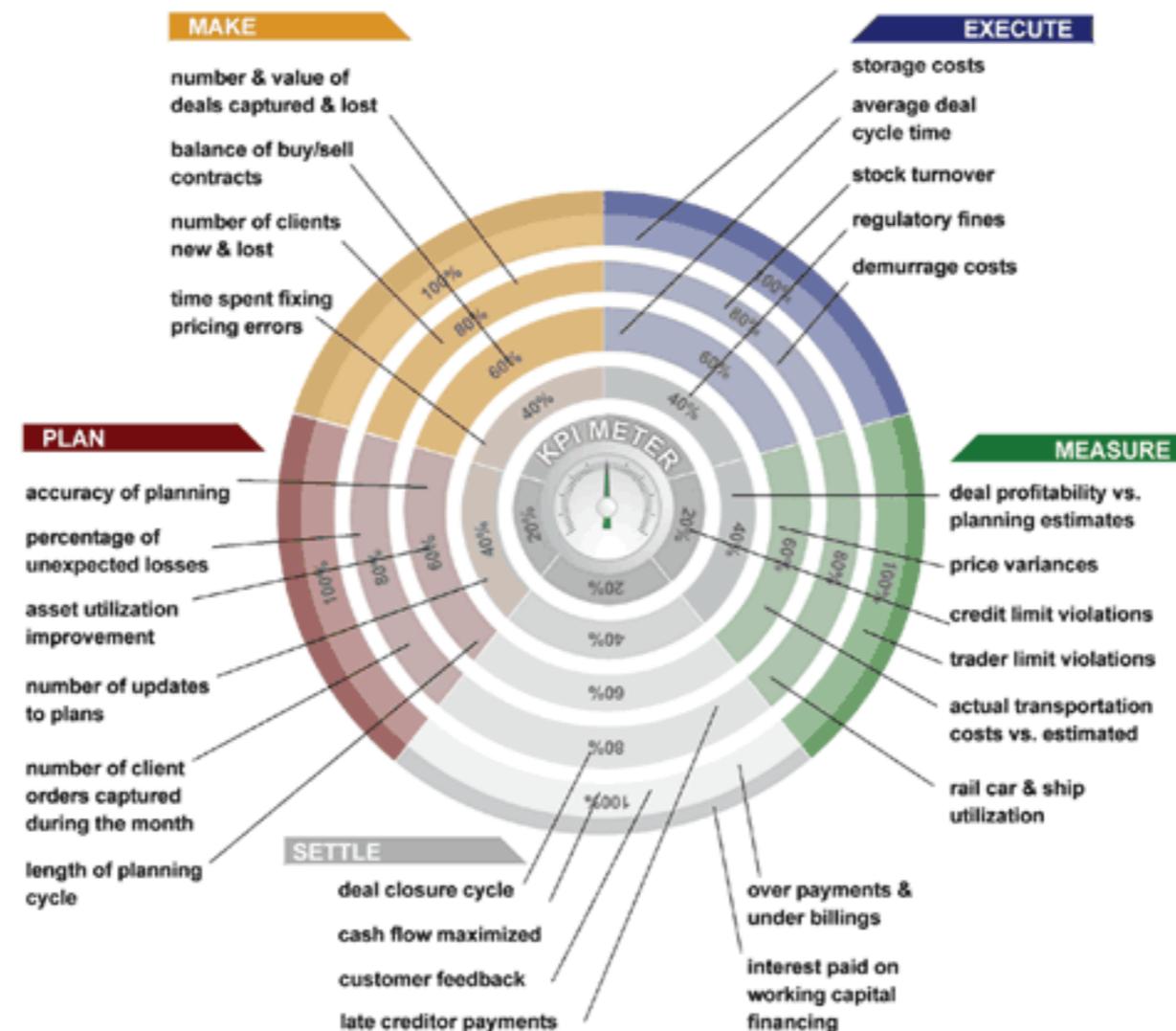
Dashboard o cruscotto è una pagina html alimentata dal sistema che attraverso rappresentazioni grafiche mostra in modo intuitivo l'andamento dei fenomeni che si intende tenere sotto controllo





I numeri: Key Performance Indexes

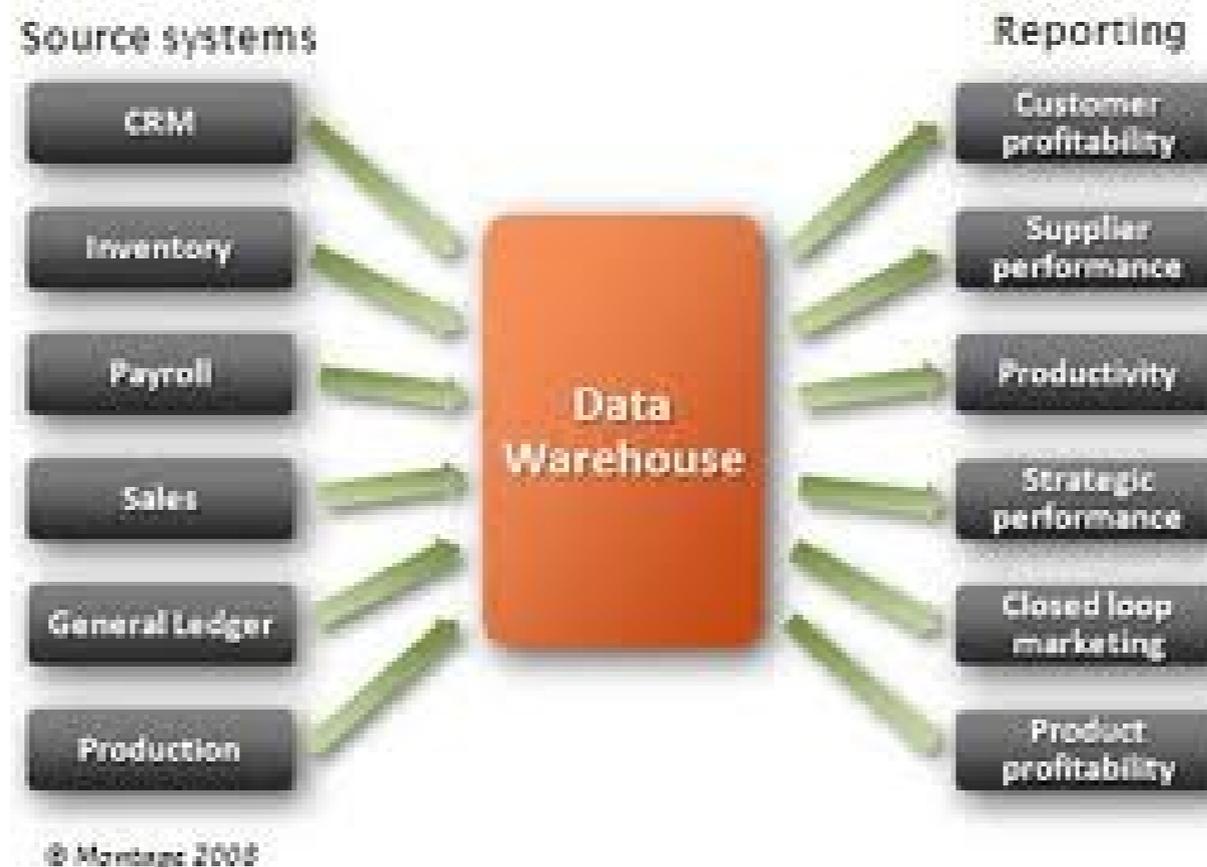
- Sono ovviamente legati alla specifica area in cui ci si muove: ad esempio una squadra di calcio misurerà il numero di gol, passaggi, palle perse, falli, ecc. per capire quali sono gli aspetti da migliorare, i punti di forza, le debolezze





Cosa è un data warehouse

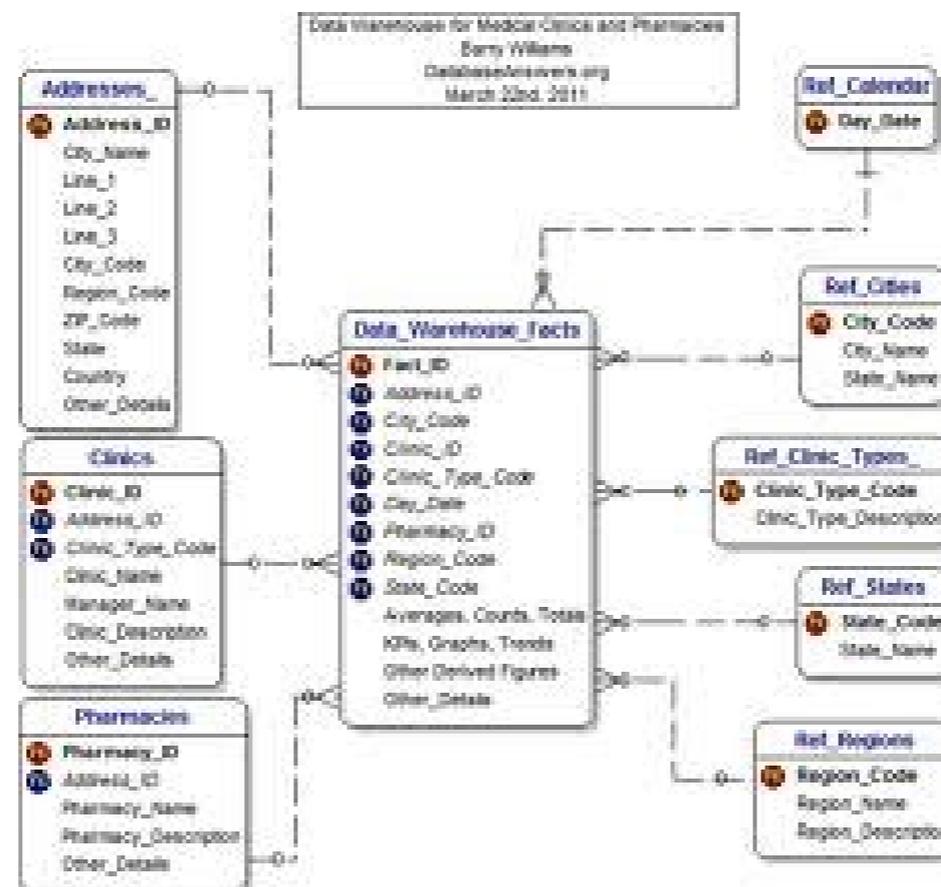
- Per lavorare sui dati occorre costruire un modello (generalmente avrà una forma a stella - star schema) ed alimentarlo con dati estrapolati dagli archivi operativi, che possono essere di diversa natura, e da fonti esterne, ripulirli, renderli compatibili, attraverso un processo detto Extract, Transform and Load. Il data Warehouse sarà progettato in modo da rendere semplice e veloce l'analisi e la rappresentazione dei dati.





Cenni sul data modeling

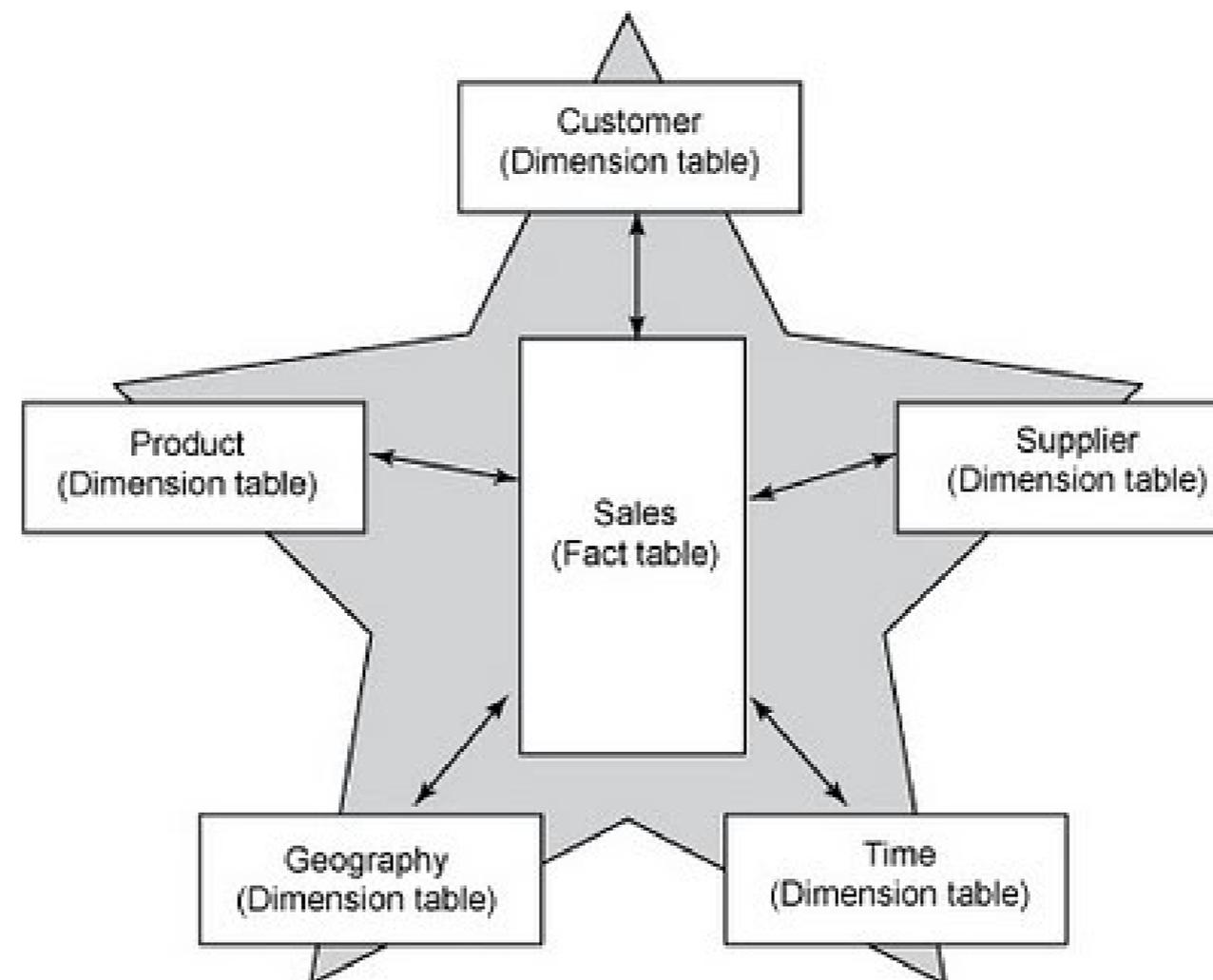
- La modellizzazione deve essere disegnata collaborando con un esperto di dominio, vale a dire da chi ha competenza sullo specifico campo di applicazione e che indicherà quali fenomeni devono essere controllati, la tipologia dei report, quali grandezze devono essere messe in relazione, i periodi di tempo da osservare, ecc. Tutti aspetti specifici di quel settore.





Star Schema

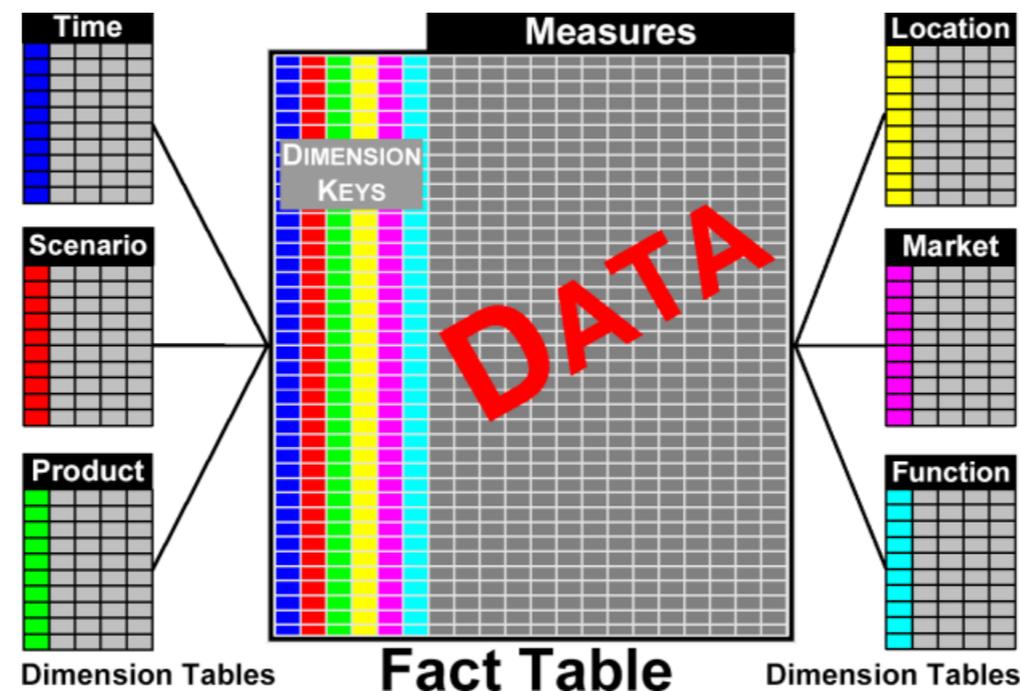
- Il risultato sarà - semplificando - uno star schema che conterrà una tabella centrale di fatti collegata a più tabelle periferiche di dimensioni attraverso dei campi chiave. Il modello di riferimento è (in questo caso) quello formalizzato da Ralph Kimball.





Cenni sul data modeling

- Le tabelle di dimensione devono essere complete cioè contenere tutte le tipologie che possono essere contenute nei fatti
- Esempio: la dimensione tempo_data (completa per il periodo temporale considerato, con tutte le informazioni necessarie: festività, giorni particolari etc.). Un aspetto importante sulle tabelle tempo è l'internazionalizzazione (possibilità di usare tabelle diverse per diversi paesi)





Cosa è Pentaho

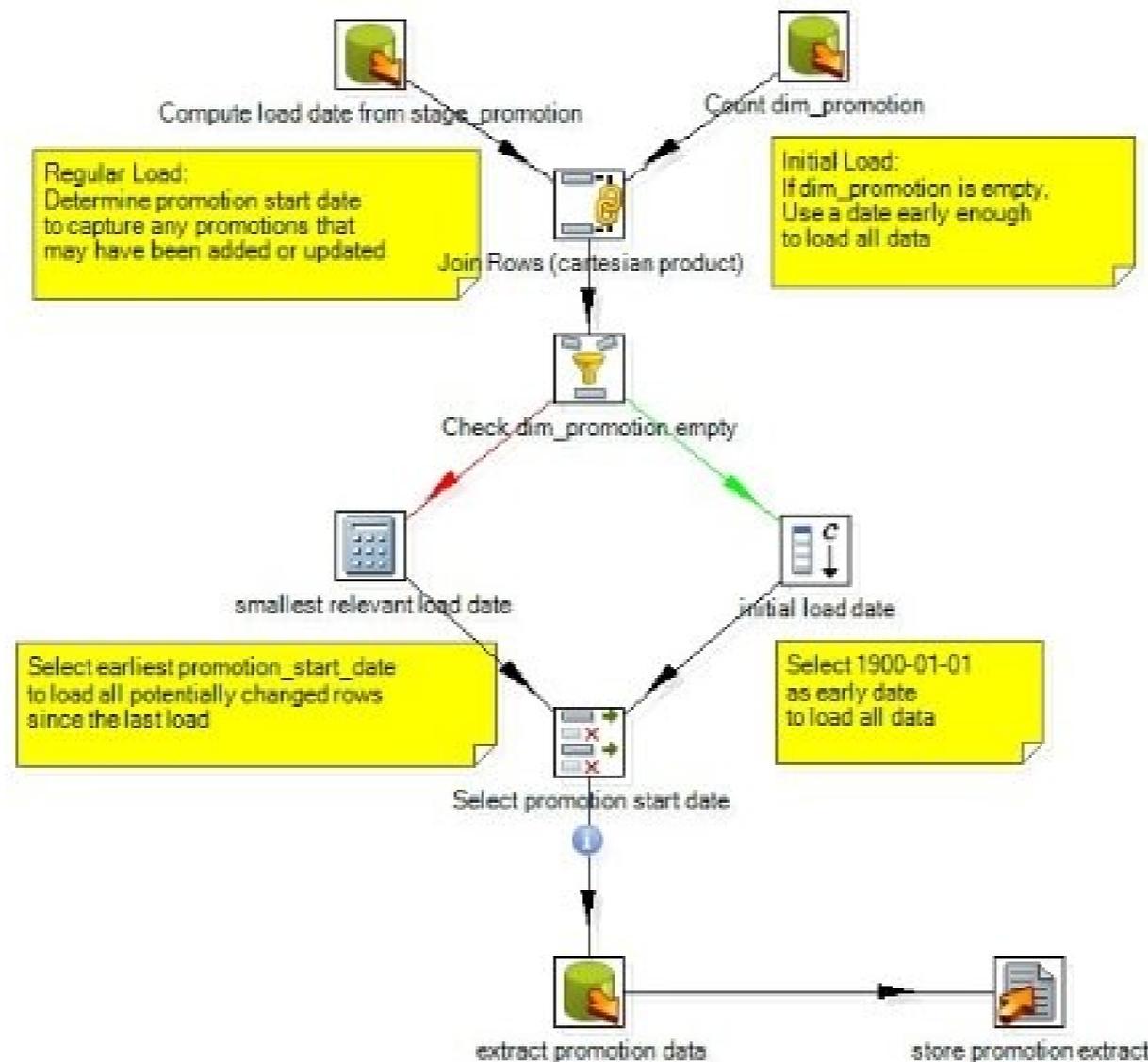
- La suite Pentaho è un sistema completo di OSBI (open source BI) composta da diversi tools che possono essere usati indipendentemente:
- Kettle (ETL); Mondrian (Olap); Report Designer; Metadata Editor; Community Dashboard Framework; Weka (Data mining), oltre agli strumenti di gestione e di amministrazione (bi server, design studio, administration console)





ETL - Pentaho Data Integration (Kettle)

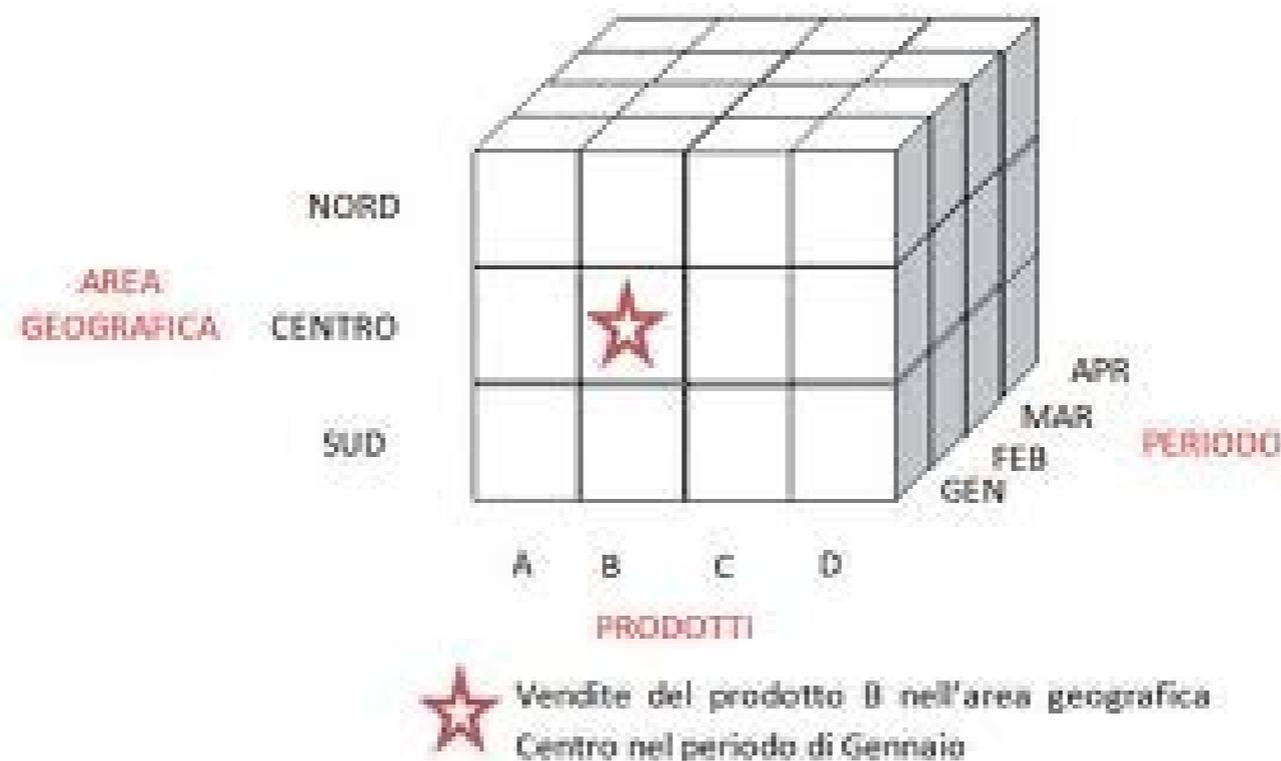
- Un ambiente grafico dove programmare i passi di acquisizione, trasformazione e carico dei dati. Gli schemi semplici possono essere inseriti in processi (job) che verranno eseguiti in batch in modo parallelo.





OLAP - Pentaho Mondrian

- I cubi, definiti attraverso un testo XML, possono essere progettati con un editor di testo o con l'aiuto dello schema workbench o del tool del server. Occorre definire lo star schema e selezionare i campi da visualizzare. E' possibile ottenere campi calcolati, usare altre colonne per definire le intestazioni o l'ordinamento (importante per la lettura e per la localizzazione)





Report

- Il modulo report permette di realizzare report interattivi da stampare o pubblicare via web

The screenshot shows a web browser displaying a report titled "Steel Wheels". The report is a table with columns: Product, Price, Quantity, Status, and Amount. The table is divided into three sections: "Advanced Wheels Design", "Common (Basic) Design", and "Other Metal Products". The "Status" column contains color-coded cells: red for "Out of Stock" and green for "In Stock".

| Product | Price | Quantity | Status | Amount |
|---|--|----------|--------------|-------------|
| Advanced Wheels Design | 100,000 | 1000000 | Out of Stock | 100,000,000 |
| Description: High-end steel wheels with multi-spoke design. Features: Strong steel wheels with the Pentacoil design. Call for price. | | | | |
| Common (Basic) Design | 100,000 | 1000000 | In Stock | 100,000,000 |
| | Description: For mass production of steel wheels. Call for price. Features: Strong steel wheels with the Pentacoil design. Call for price. | | | |
| | 100,000 | 1000000 | In Stock | 100,000,000 |
| Description: The standard steel wheels for the Pentacoil design. Features: Strong steel wheels with the Pentacoil design. Call for price. | | | | |
| Other Metal Products | 100,000 | 1000000 | In Stock | 100,000,000 |
| Description: Various metal products including wheels. Features: Strong steel wheels with the Pentacoil design. Call for price. | | | | |
| 100,000 | 1000000 | 1000000 | In Stock | 100,000,000 |
| Description: Various metal products including wheels. Features: Strong steel wheels with the Pentacoil design. Call for price. | | | | |



Dashboards

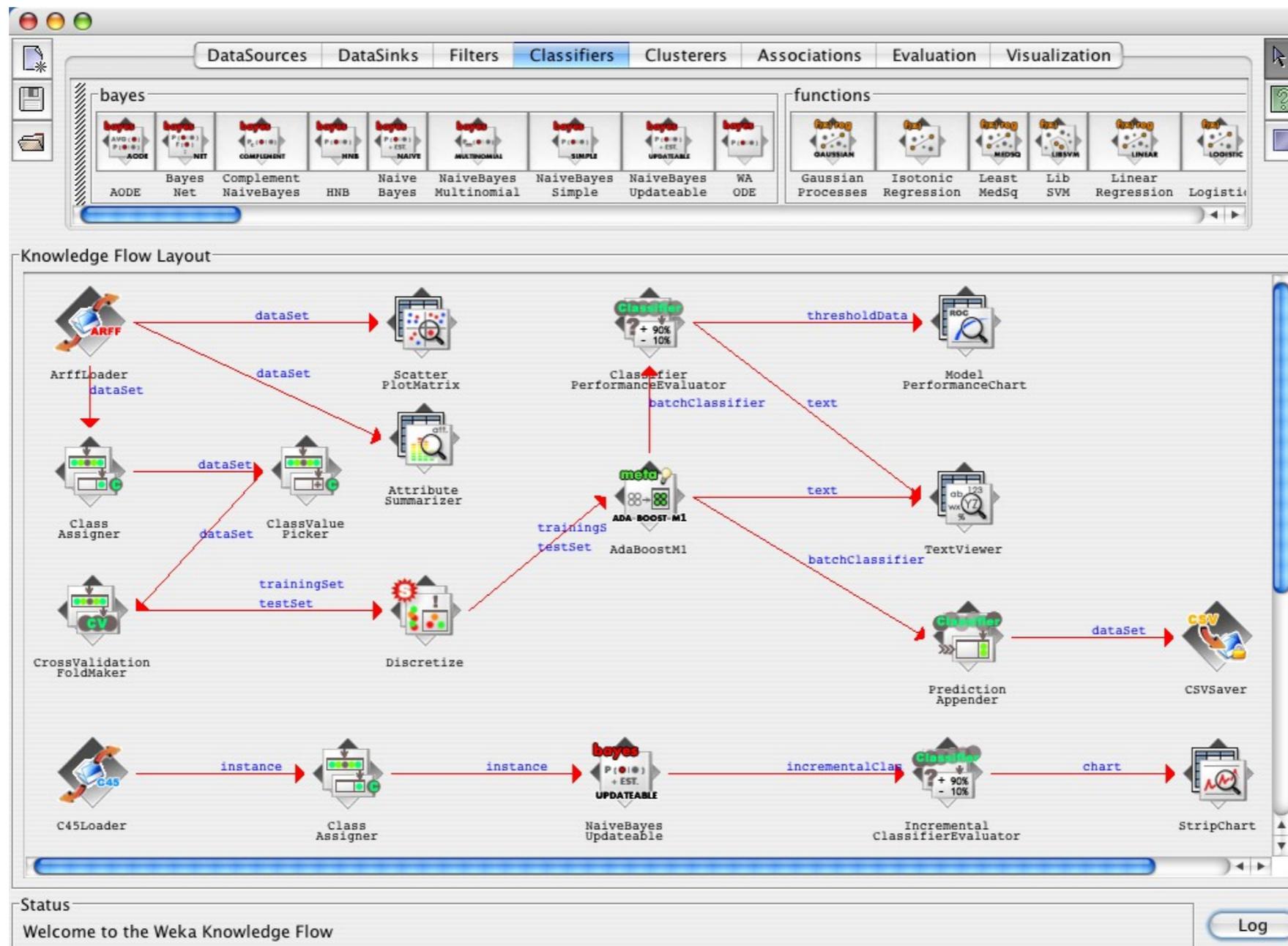
- CDF e CDE sono il modulo base e l'editor interattivo per realizzare Dashboards raffinate





Data Mining - Weka

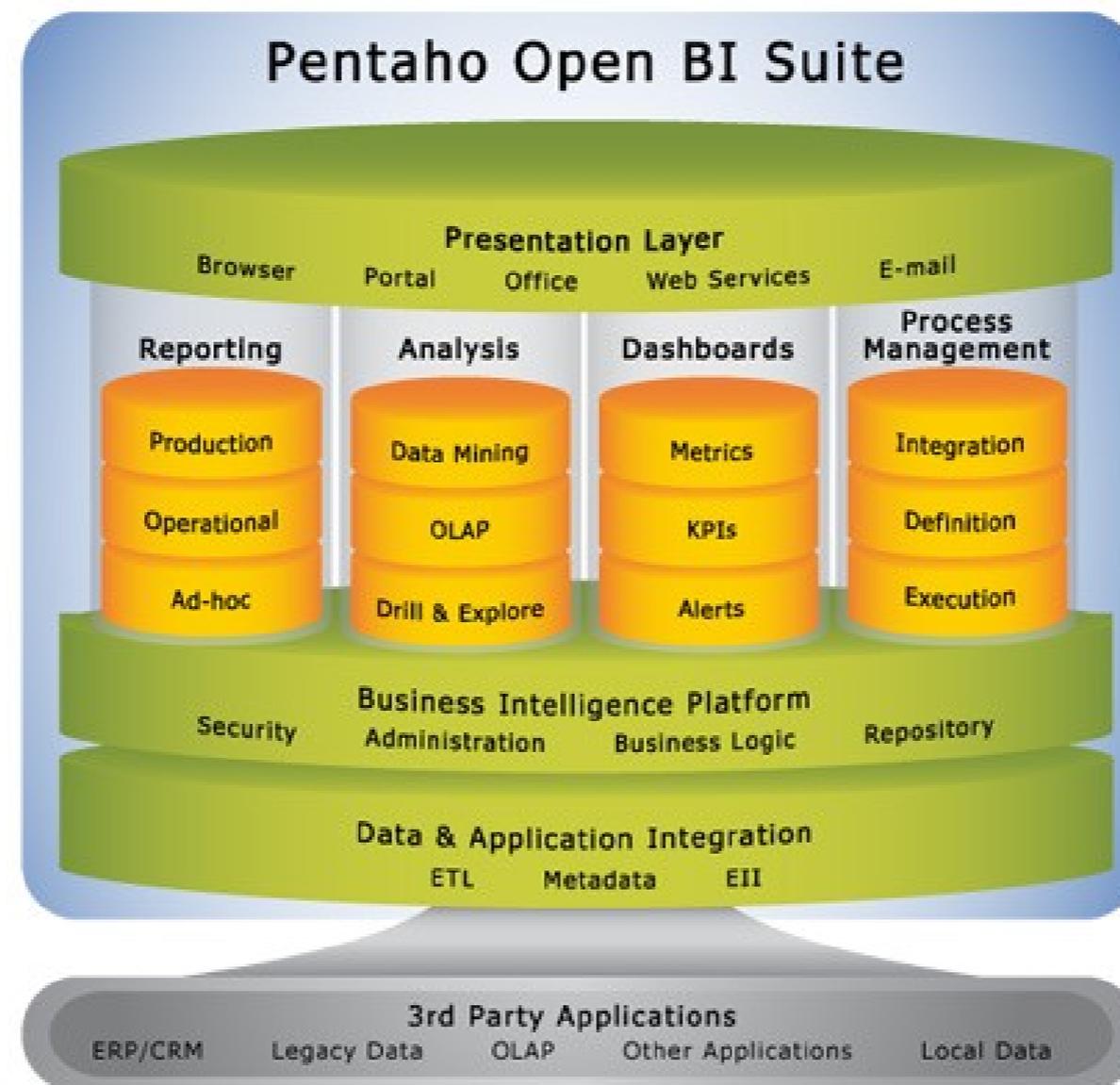
- Weka è il modulo per realizzare analisi ed esplorazione dati (data mining)





Community vs enterprise edition

- La versione community è liberamente disponibile per il download e l'uso. La versione enterprise comprende alcune versioni grafiche evolute dei tool per la generazione di Dashboards (ma ora esiste anche una CDE competitiva) e propone un contratto di assistenza.





Esempi d'uso



GRAZIE PER L'ATTENZIONE