Business Intelligence
An Overview

Zahra Mansoori
Contents

1. Preference
2. History
3. Inmon Model - Inmonities
4. Kimball Model - Kimballities
5. Inmon vs. Kimball
6. Reporting
7. BI Algorithms
8. Summary
9. References
1. Preferences
Preference

Over the past two decades

- companies have gathered tons and tons of data about their operation
- Information is said to double every 18 months

The theory behind BI systems:

- you cannot improve what you do not measure
- Without some sort of feedback mechanism, you are essentially **driving blind**
Structured and Unstructured data input
Business Intelligence Lifecycle
Decision making

• Operational Decision making
  • Operational Systems

• Tactical Decision making
  • Meeting certain business objectives within a specific time frame

• Strategic Decision making
  • Long Term Goals
  • Far-reaching impact on the organization
Decision Making Pyramid
Data Evolution (DIKW Pyramid)

- **Data** is the foundation of **Information**, **Knowledge** and ultimately, **Wisdom**

**Context**
- Joining Of Whole
- Formation Of a Whole
- Connection Of Parts
- Gathering Of Parts

**Understanding**
- Researching
- Absorbing
- Doing
- Reflecting

**Past**
- Information

**Future**
- Knowledge
- Wisdom
Enterprise Data

- Transactional Data
- Analytical Data
- Master Data
- Metadata
Definition: OLAP vs. OLTP

**OLAP**
- **Online Analytical Processing**, or **OLAP**, is an approach to answering multi-dimensional analytical queries.
- OLAP tools enable users to analyze multidimensional data interactively from multiple perspectives.
- Databases configured for OLAP use a multidimensional data model, allowing for complex analytical and ad hoc queries with a rapid execution time. They borrow aspects of navigational databases, hierarchical databases and relational databases.

**OLTP**
- **Online transaction processing**, or **OLTP**, is a class of information systems that manage transaction-oriented applications, typically for data entry and retrieval transaction processing. OLTP has also been used to refer to processing in which the system responds immediately to user requests.
KPI

- A Performance Indicator or Key performance indicator (KPI) is a type of performance measurement. An organization may use KPIs to evaluate its success, or to evaluate the success of a particular activity in which it is engaged. Sometimes success is defined in terms of making progress toward strategic goals, but often success is simply the repeated, periodic achievement of some level of operational goal (e.g. zero defects, 10/10 customer satisfaction, etc.).
Nature of Data Warehouse

- Historical Data
- Easy to query
- Show the relationship between unrelated data
- Time-stamped data
- User-friendly access tools
- Reasonable response time
Business Intelligence Concerns

- Fraud Analysis
- Churn Analysis
- Traffic Analysis
- Product Bundling
2. History
Evolution by Time

- 1970s: Some Concerns
- 1980s: Traditional Model (Called DSS)
- 1990: Bill Inmon Model
- 1996: Ralph Kimball Model
Traditional DSS Models


Traditional DSS systems consist of:

- **One**: A Central Data Warehouse that contains company Transaction Data
- **Two**: A Reporting Mechanism that allows users to access the data in several summary and ad hoc formats
- **Three**: A Common Interface is a Dashboard that reports how the company is doing on Key Performance Indicators (KPIs)
Traditional BI – cont.
Merits and Demerits of Traditional Model

With a Traditional BI system:

• You are no longer driving blind, but,
• Because all information is historical, your only view of the world is through your rear-view mirror
• If the road on which you are driving is long, featureless, and straight, you can stay on course by making small corrections and watching how the road drifts behind you
• However, if there is a fork in the road ahead (an opportunity) you won't see it until it passes
• And, if there is a sharp curve, you crash!

What you need is a system that gives you a forward view
1990 - Bill Inmon Model

• The term Business Intelligence is a popularized introduced by Gartner Group in 1989

• In 1990, Bill Inmon Became “Father of Data warehousing”

• The Industry soon began to implement Inmon’s vision

• In 2002 Inmon introduced new concept to his model

• Data stored into single database called Data Warehouse

• Data extract from this database to smaller Departmental Databases

• Decision support users query and create reports from departmental databases – a TOP-DOWN approach
In 1996, Kimball, a scholar-practitioner developed a model that compete Inmon’s

In 2002 he complete his model

Recommends an architecture multiple databases, called Data Marts, organized by business processes

The sum of Data Marts comprises the Data Warehouse

A BOTTOM-UP approach that must adhere to an enterprise-wide standard “Data Bus”
3. Inmon Model - Inmonities
Definition

• All Data of an Organization: **Corporate Information Factory (CIF) contains**

- **Operational**
  - Current Transactional Data

- **Atomic Data Warehouse**
  - Historical Data

- **Departmental**
  - Summarized Data of DW specifically for each department

- **Individual**
  - Unstructured user generated data
Inmon’s Top-down design

Atomic Data warehouse as a centralized repository for the entire enterprise

Departmental data will extract from Data Warehouse
Inmon Top-Down Schema

- Data stores in ERD
- Summarized Data From Data warehouse to Data marts

Data Warehouse

Departmental Data (Data Marts)
4. Kimball Model – Kimballities
Kimball Model

• Uses a data modeling method unique to the Data Warehouse

• Known as “Dimensional Data Modeling”

• Multiple databases as Data Marts consolidate to each other – highly interoperable

• Data Bus – another invention
Kimball Bottom-Up Schema

- Data stores in Fact-Dimension Model

Data Marts ➔ Data Warehouse
Definition: Fact

Fact

• If the business process is SALES, then the corresponding fact table will typically contain columns representing both raw facts and aggregations in rows such as:
  - $12,000, being "sales for New York store for 15-Jan-2005"
  - $34,000, being "sales for Los Angeles store for 15-Jan-2005"
  - $22,000, being "sales for New York store for 16-Jan-2005"
  - $50,000, being "sales for Los Angeles store for 16-Jan-2005"
  - $21,000, being "average daily sales for Los Angeles Store for Jan-2005"
  - $65,000, being "average daily sales for Los Angeles Store for Feb-2005"
  - $33,000, being "average daily sales for Los Angeles Store for year 2005"
Definitions: Dimension

**Dimension**

- The dimension is a data set composed of individual, non-overlapping data elements. The primary functions of dimensions are threefold: to provide filtering, grouping and labeling.
- Typically dimensions in a data warehouse are organized internally into one or more hierarchies. "Date" is a common dimension, with several possible hierarchies:
  - "Days (are grouped into) Months (which are grouped into) Years",
  - "Days (are grouped into) Weeks (which are grouped into) Years"
  - "Days (are grouped into) Months (which are grouped into) Quarters (which are grouped into) Years"
  - etc.
Fact vs. Dimension Table

**Fact Table**
- Contain metrics
- Contain many rows and relatively few columns (for query performance)

**Dimension Table**
- Contain attributes of the metrics of fact table
- Have only hundreds or thousands of rows
- Hundred columns or more
Star Schema

- Relationship between Fact and Dimension Tables are in Star Schema
Fact and Dimension Tables in Star Schema
Example of Fact and Dimension table

4 dimensions: Service, Time, Sales Point, Customer

1 Fact: Transactions
OLAP Fact-Dimension Cube

- **Fact Table** is the Cartesian Product of **Dimension Tables**
- Operation On Dimension Cubes: *Slice, Dice, Drill down, Roll Up*
Operation: Slice

- **Slice** is the act of picking a rectangular **Subset** of a cube by choosing a **Single Value** for one of its dimensions, creating a new cube with **One Fewer Dimension**.
Operation: Dice

- **Dice** operation produces a Subcube by allowing the analyst to pick specific values of multiple dimensions.
Operation: Drill Down/ Roll Up

- **Drill Down/Roll Up** allows the user to **Navigate Among Levels Of Data** ranging from the **Most Summarized** (Roll Up) to the **Most Detailed** (Drill Down).
BI Structure
BI Structure:
Operational Data, Different Data Source, Structured Or Unstructured
BI Structure

- Extracts data from outside sources
- Transforms it to fit operational needs
- Loads it into the end target (data mart, or data warehouse)
ETL

**Extract**
- Staging Database

**Transform**
- Staging Database

**Load**
- Data Warehouse
ETL

Moving data from operational systems to a persistent staging area
ETL

Transforming Data To Ensure Data Integrity Between Different Inputs
ETL

Loading data into Data Marts
ETL

Garbage In—Garbage Out!
Data Warehouses Store Current And Historical Data And Are Used For Creating Trending Reports For Senior Management Reporting.
A data mart is the access layer of the data warehouse environment that is used to get data out to the users.
BI Structure

Dashboard which is used by business users by getting query from data marts
EDW Bus Architecture

Database-independent Bus Architecture

<table>
<thead>
<tr>
<th>Decomposes The DW/BI Process</th>
<th>By Focusing On The Organization’s Core Business Processes By Using Conformed Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conformed Dimensions:</td>
<td>Master Common Standardized Dimensions</td>
</tr>
<tr>
<td></td>
<td>Created Once In The ETL</td>
</tr>
<tr>
<td></td>
<td>Reused By Multiple Fact Tables</td>
</tr>
</tbody>
</table>
Bus Architecture – Example

![Table showing common dimensions and business processes]

<table>
<thead>
<tr>
<th>BUSINESS PROCESSES</th>
<th>Date</th>
<th>Product</th>
<th>Warehouse</th>
<th>Store</th>
<th>Promotion</th>
<th>Customer</th>
<th>Employee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue Purchase Orders</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receive Warehouse Deliveries</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Warehouse Inventory</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receive Store Deliveries</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Store Inventory</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Retail Sales</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Retail Sales Forecast</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Retail Promotion Tracking</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Returns</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Returns to Vendor</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Frequent Shopper Sign-Ups</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Bus Architecture – Example.

Data warehouse
5. Inmon vs. Kimball
## Inmon vs. Kimball

<table>
<thead>
<tr>
<th></th>
<th>Inmon</th>
<th>Kimball</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methodology And Architecture</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Approach</td>
<td>Top - Down</td>
<td>Bottom - Up</td>
</tr>
<tr>
<td>Architectural Structure</td>
<td>Enterprisewide (atomic) data warehouse “feeds” departmental databases</td>
<td>Data Marts model a single business process, enterprise consistency achieved through data bus and conformed dimensions</td>
</tr>
<tr>
<td>Complexity</td>
<td>Quite complex</td>
<td>Fairly simple</td>
</tr>
<tr>
<td><strong>Data Modeling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Orientation</td>
<td>Subject or data – driven</td>
<td>Process Oriented</td>
</tr>
<tr>
<td>Tools</td>
<td>Traditional (ERD, DIS)</td>
<td>Dimensional modeling</td>
</tr>
<tr>
<td>End-user Accessibility</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>
6. Reporting
Types of reporting

- Standard, static reports
- Ad-hoc reports
- Interactive, multidimensional OLAP reports
- Dashboards
- Write-back reports
- Technical reports
Standard, static reports

• Subject oriented, reported data defined precisely before creation

• **Reports with fixed layout** defined by a report designer when the report is created

• Very often the static reports contain sub-reports and perform calculations or implement advanced functions

• Generated either on request by an end user or refreshed periodically from a scheduler

• Usually are made available on the web server or a shared drive
Ad-Hoc Reports

• Simple reports created by the end users on demand

• Designed from scratch or using a standard report as a template
Interactive, multidimensional OLAP reports

- Usually provide more general information - using dynamic drill-down, slicing, dicing and filtering users can get the information they need

- Reports with fixed design defined by a report designer

- Generated either on request by an end user or refreshed periodically from a scheduler

- Usually are made available on the web server or a shared drive
Dashboards

- Contain **high-level, aggregated** company strategic data with comparisons and performance indicators
- Include both **static** and **interactive reports**
- Lots of **graphics, charts and illustrations**
Write-back reports

• Those are interactive reports directly linked to the Data Warehouse which allow modification of the data warehouse data.
By far the most often use of this kind of reports is:
  • Editing and customizing products and customers grouping
  • Entering budget figures, forecasts
  • Setting sales targets
  • Refining business relevant data
Technical reports

This group of reports is usually generated to fulfill the needs of the following areas:

- IT technical reports for monitoring the BI system, generate execution performance statistics, data volumes, system workload, user activity etc.
- Data quality reports - which are an input for business analysts to the data cleansing process
- Metadata reports - for system analysts and data modelers
Sample BI dashboard
Sample BI dashboard
Most widely used BI Systems:

<table>
<thead>
<tr>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM Cognos</td>
</tr>
<tr>
<td>SAP Business Objects and Crystal Reports</td>
</tr>
<tr>
<td>Oracle Hyperion and Siebel Analytics</td>
</tr>
<tr>
<td>Microstrategy</td>
</tr>
<tr>
<td>Microsoft Business Intelligence (SQL Server Reporting Services)</td>
</tr>
<tr>
<td>SAS</td>
</tr>
<tr>
<td>Pentaho Reporting and Analysis</td>
</tr>
<tr>
<td>BIRT - Open Source Business Intelligence and Reporting Tools</td>
</tr>
<tr>
<td>JasperReports</td>
</tr>
<tr>
<td>Qlickview</td>
</tr>
</tbody>
</table>
7. BI Algorithms
Popular algorithms used by BI software

- Regression Analysis
- Decision Tree
- Association Analysis
- Cluster Analysis
Regression Analysis

• $Y = aX + b$

• Example: Profit is Linear/Non-linear function of Revenue, so we forecast future Profit by assessing historical information

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>1,000</td>
<td>2,000</td>
<td>?</td>
</tr>
<tr>
<td>Profit</td>
<td>200</td>
<td>300</td>
<td>?</td>
</tr>
</tbody>
</table>

Profit = 0.1Revenue + 100
Decision Tree

• Decision trees are used to learn from historic data and to make predictions about the future

• Example: Customer Satisfactory

\[ X = 1 \]

\[ Y > 1 \]
\[ Z = 1: \text{Satisfied} \]
\[ Z = 2: \text{Dissatisfied} \]

\[ Y < 1 \]
\[ Z = 1: \text{Satisfied} \]
\[ Z = 2: \text{Satisfied} \]
Association Analysis

• Helps you to identify cross-selling opportunities, for example. You can use the rules resulting from the analysis to place associated products together in a catalog

• Let: \( I = \{I_1, I_2, \ldots, I_m\} , T \subset I \) (\( T \) is a Transaction), \( X \subset T \)

• Define: \( X \Rightarrow Y \iff Y \subset T \) & \( X \cap Y = \emptyset \)

if a customer purchases an airline ticket, then he is likely to rent a car and make a hotel reservation
Cluster Analysis

• Example:
  1. Gathers attributes about Customers with the same purchases
  2. Predicts which product should be chosen by a specific customer with specific attribute
8. Summary

• BI Objectives
• Traditional BI
• Inmon Departmental model
• Kimball fact-dimension model and bus architecture
• BI Dashboard and Reporting
• BI Algorithms
9. References

- Wikipedia.com
- businessintelligence.com/dictionary/
- swiki.net/reliable-business-intelligence-datawarehouse-datamart-and-datamining.html
- www.oracle.com/technetwork/articles/madison-models-086845.html
9. References

- Robert J Davenport, ETL vs ELT, A Subjective View, June 2008
- Erik Perjons, DSV, SU/KTH, DW Architecture and Lifecycle
- Mary Breslin. “Data warehousing, Battle of Giants”, 2004
Thanks for your attention