Business Intelligence
An Overview

Zahra Mansoori

Contents

1. Preferences

Preference

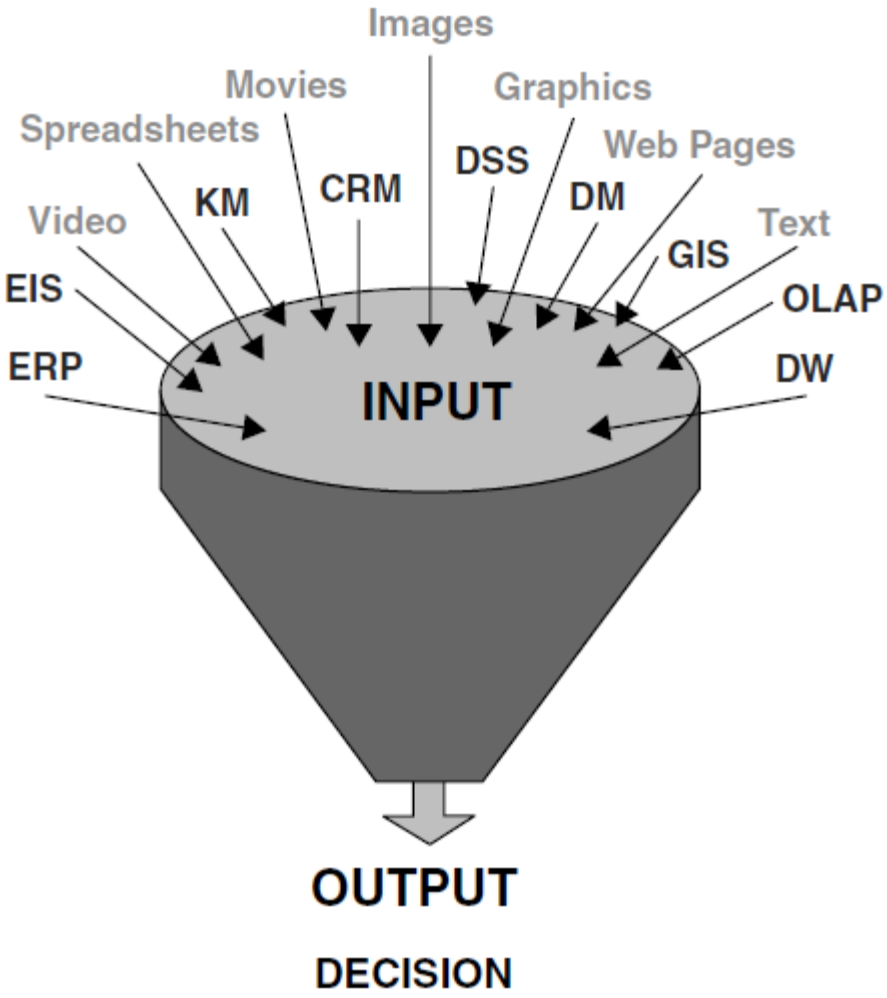
Over the past two decades

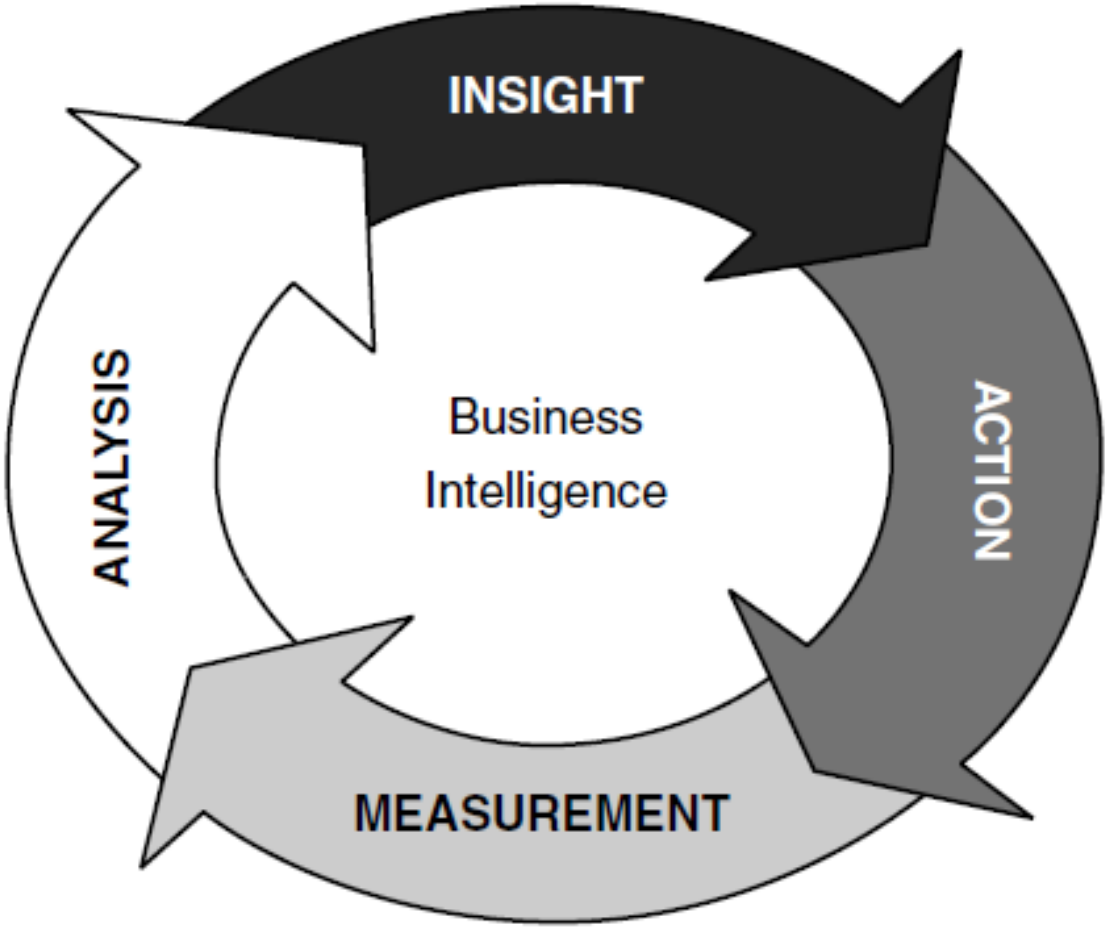
- companies have gathered tons and tons of data about their operation
- Information is said to double every 18 months

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

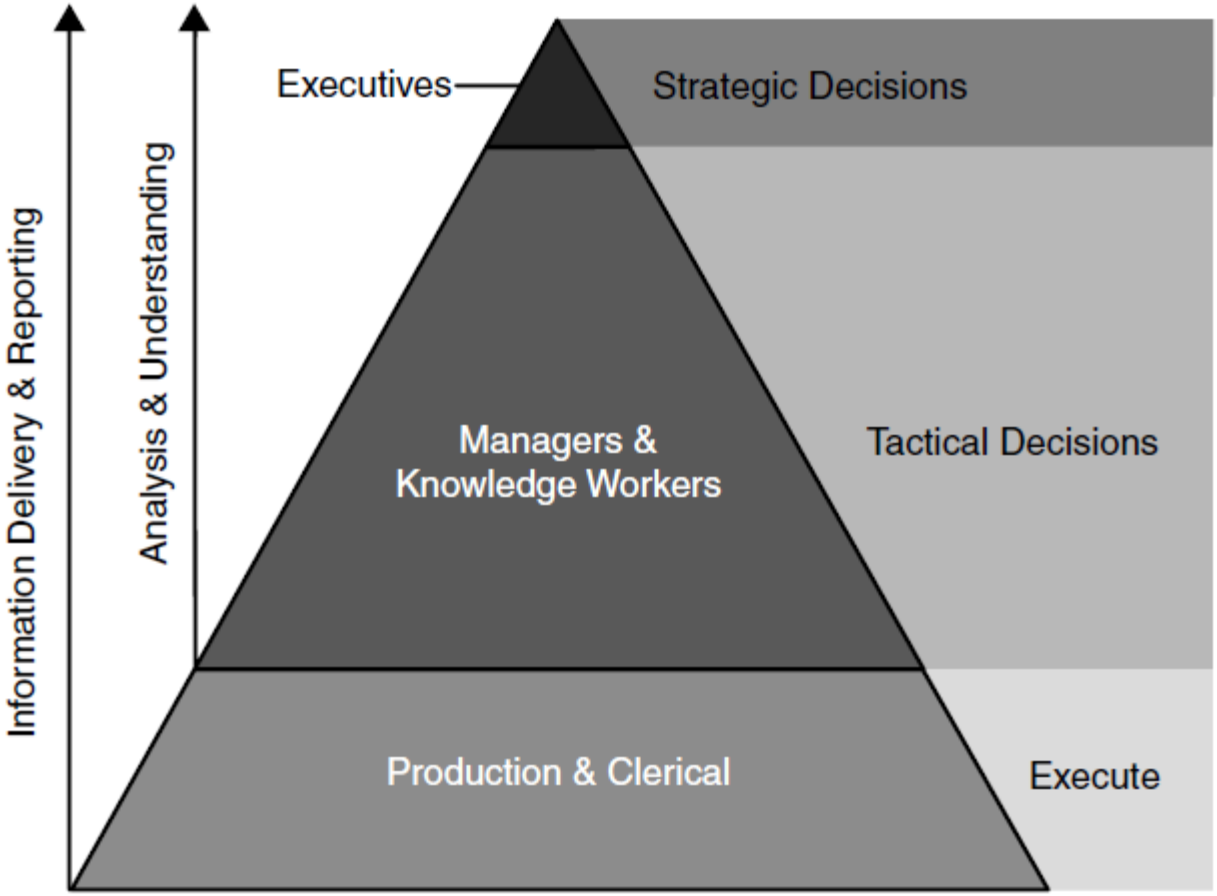




Business Intelligence Lifecycle

Decision making

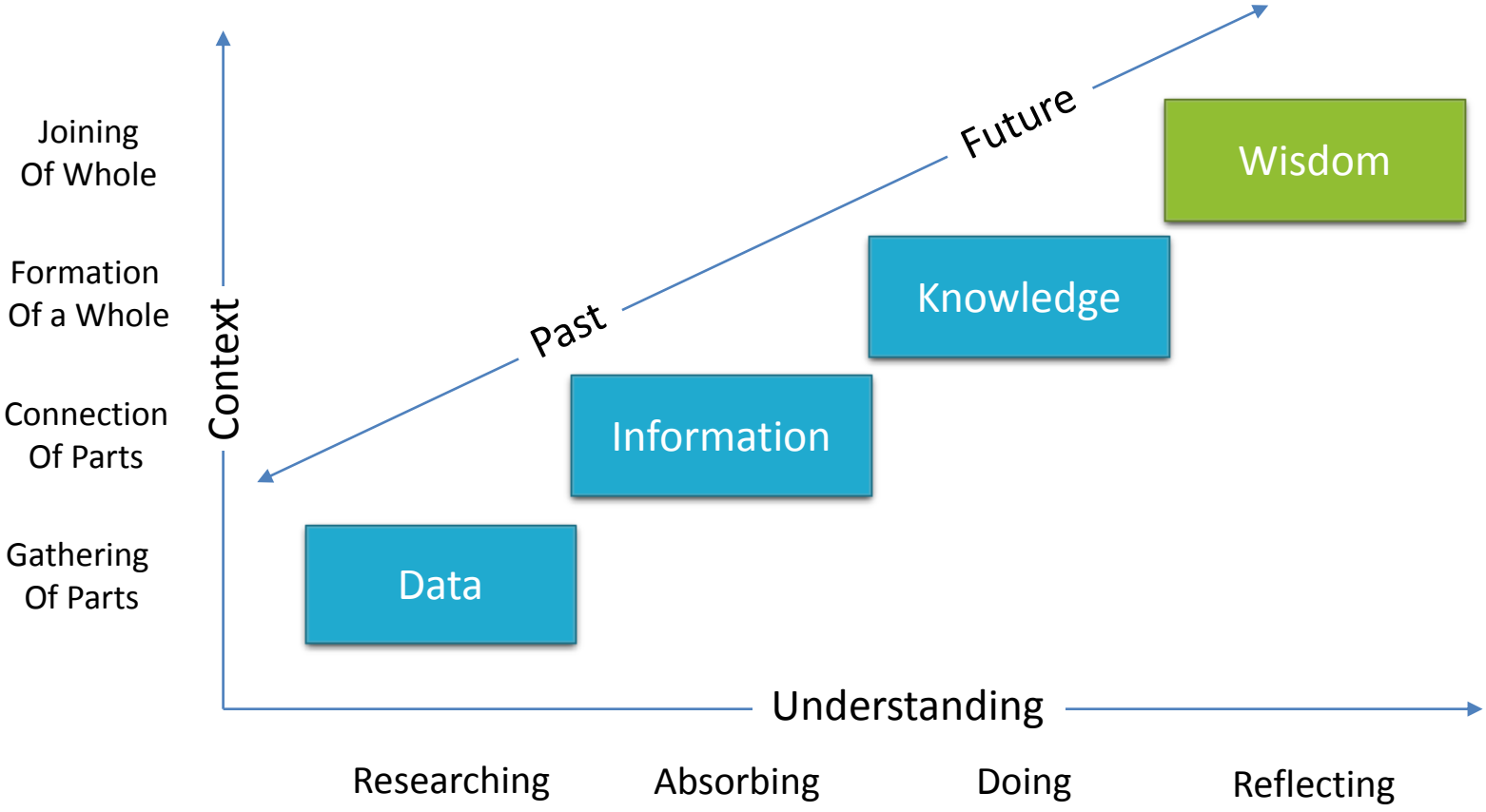
- Operational Decision making
  - Operational Systems
- Tactical Decision making
  - Meeting certain business objectives within a specific time frame
- Strategic Decision making
  - Long Term Goals
  - Far-reaching impact on the organization



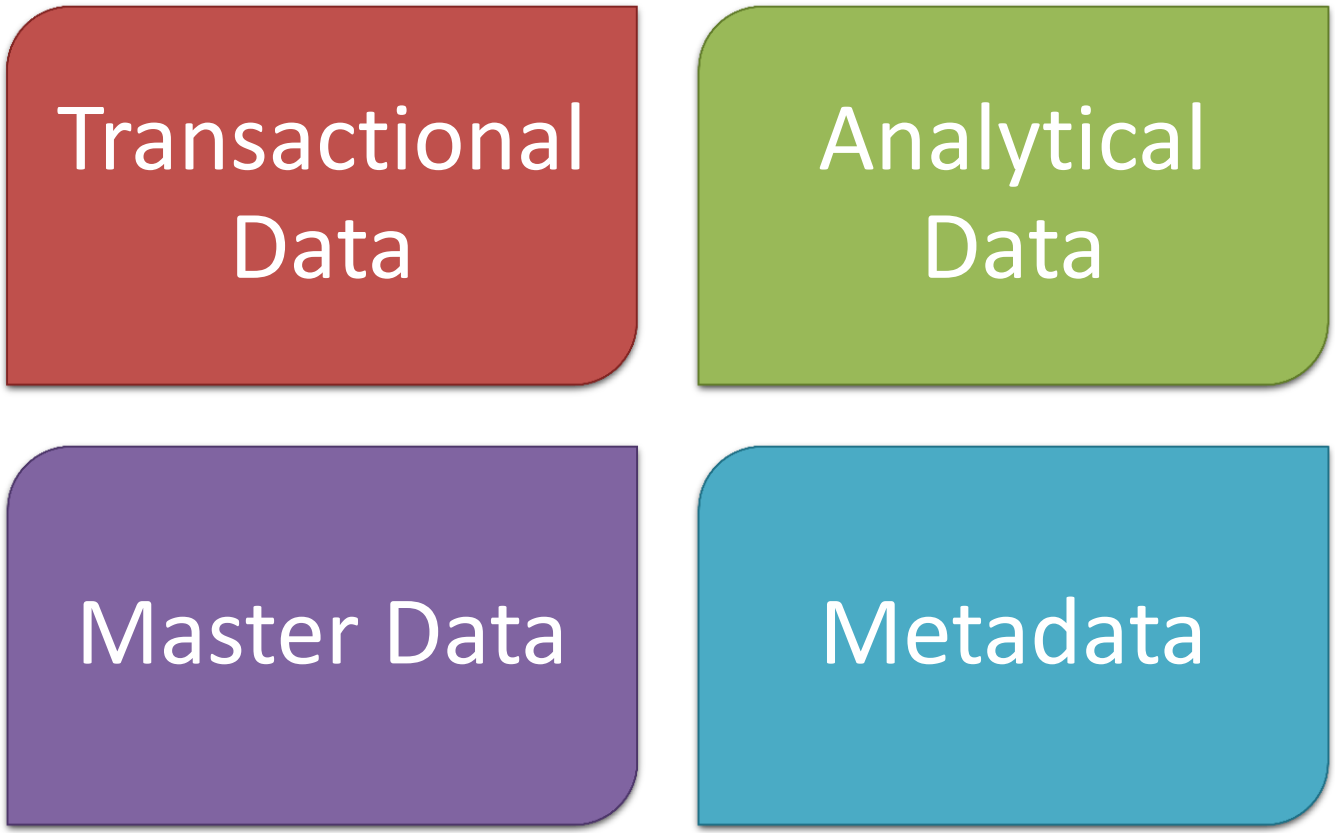
Decision Making Pyramid

Data Evolution (DIKW Pyramid)

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



Enterprise Data



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.

Definition: KPI

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.).

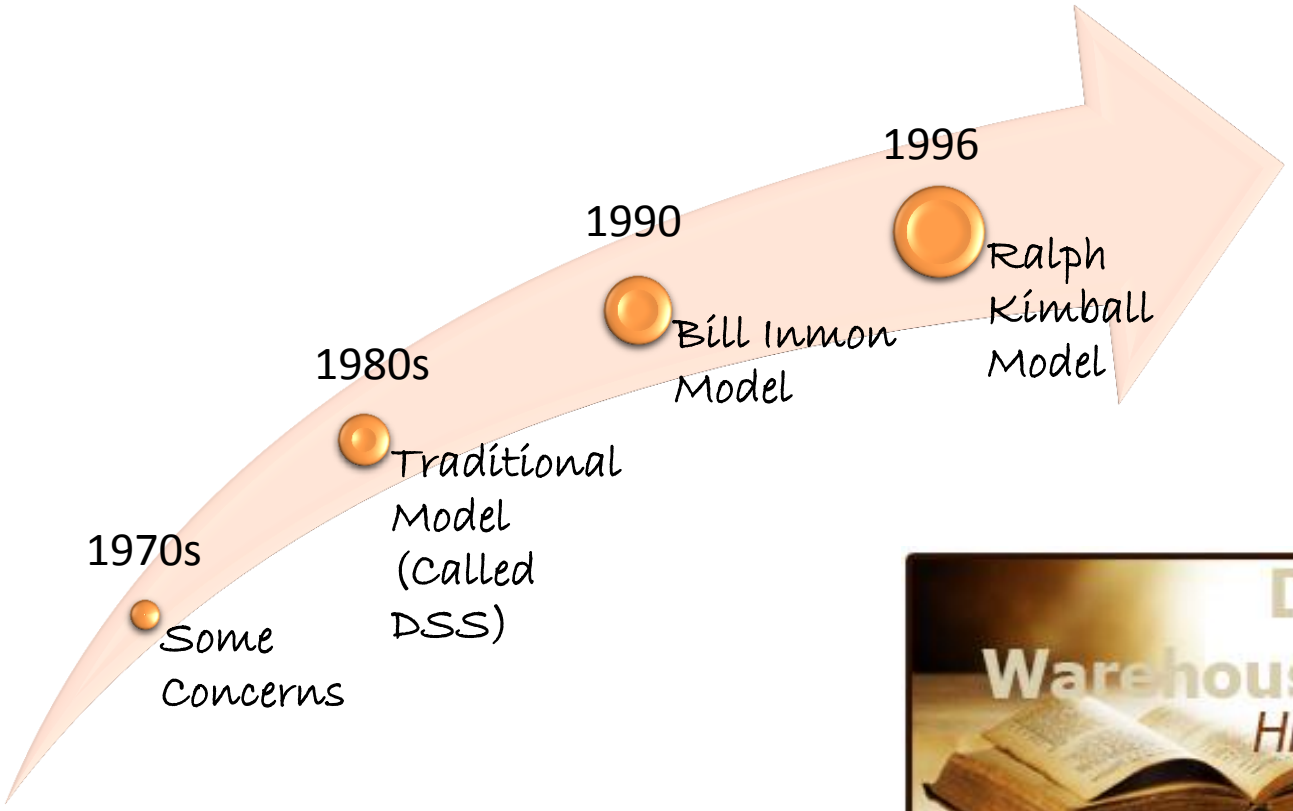
Nature of Data Warehouse

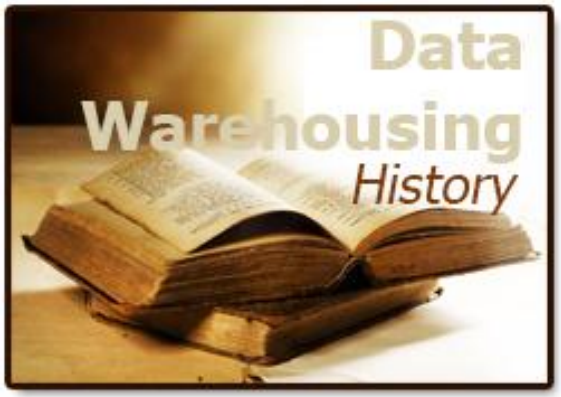
- Historical Data
- Easy to query
- Show the relationship between unrelated data
- Time-stamped data
- User-friendly access tools
- Reasonable response time

Business Intelligence Concerns

2. History

Evolution by Time





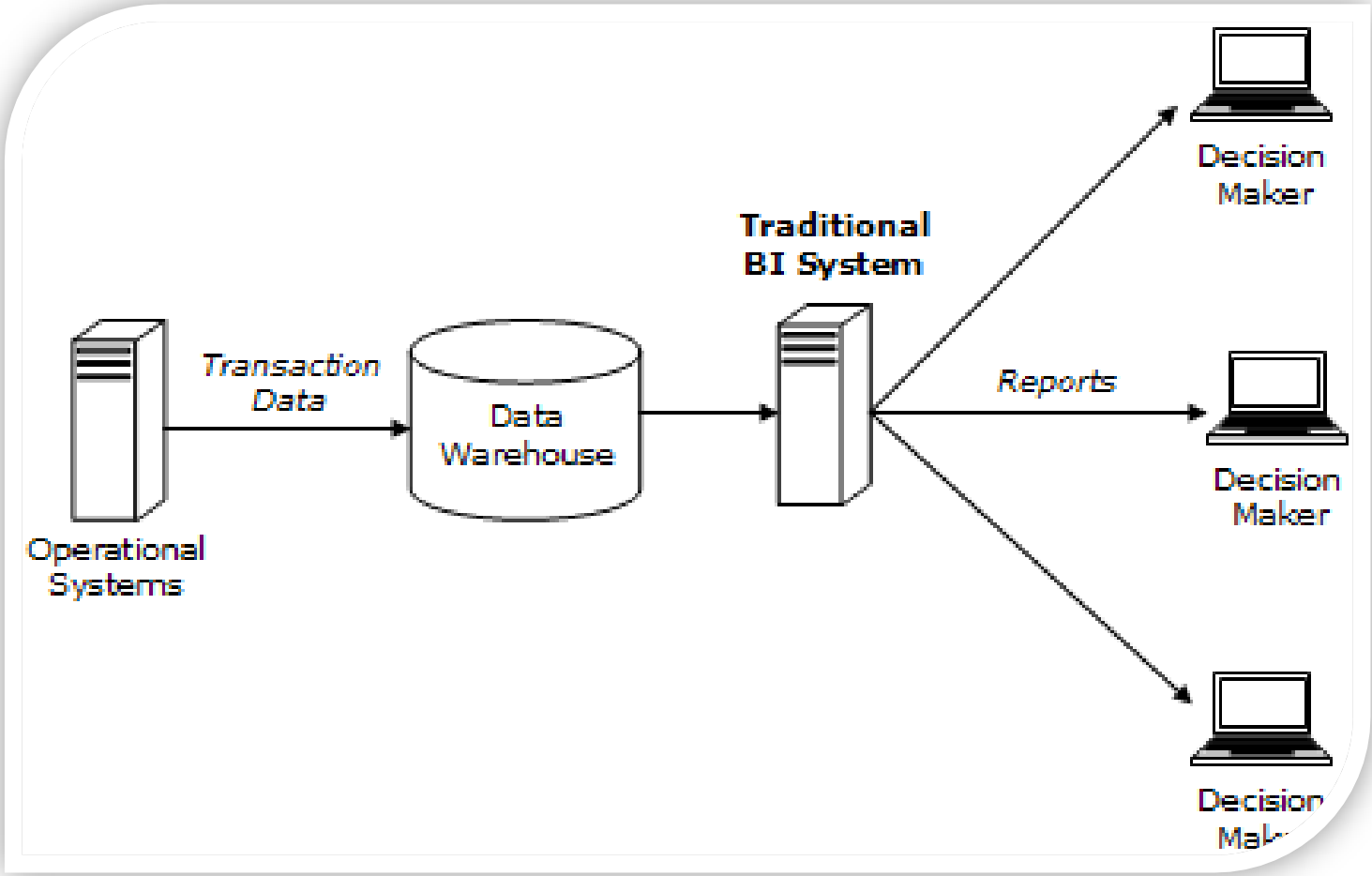
Traditional DSS Models

An IBM Systems Journal article published in 1988, "***An architecture for a business information system***", coined the term "**Business Data Warehouse**".

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

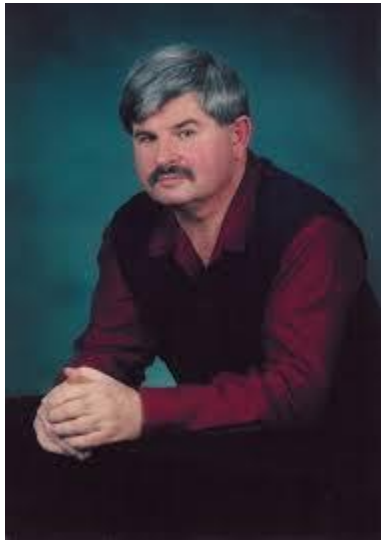
With a Traditional BI system:

- 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

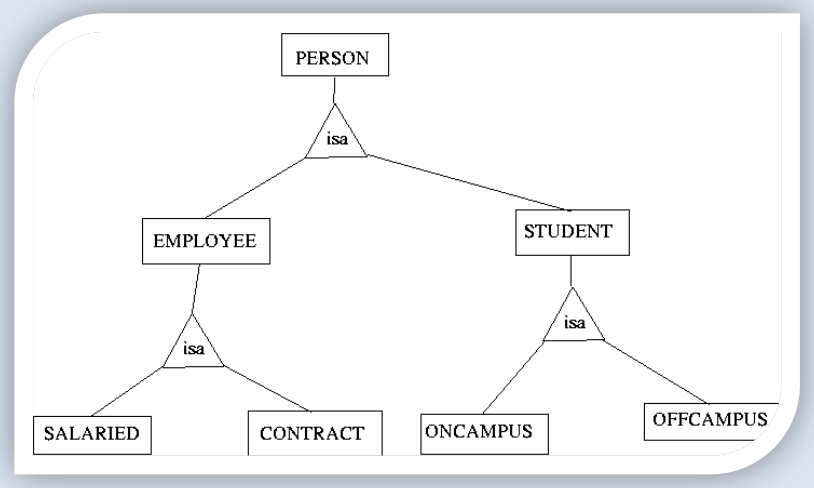
1990 - Bill Inmon Model

- 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



1996 - Ralph Kimball Model

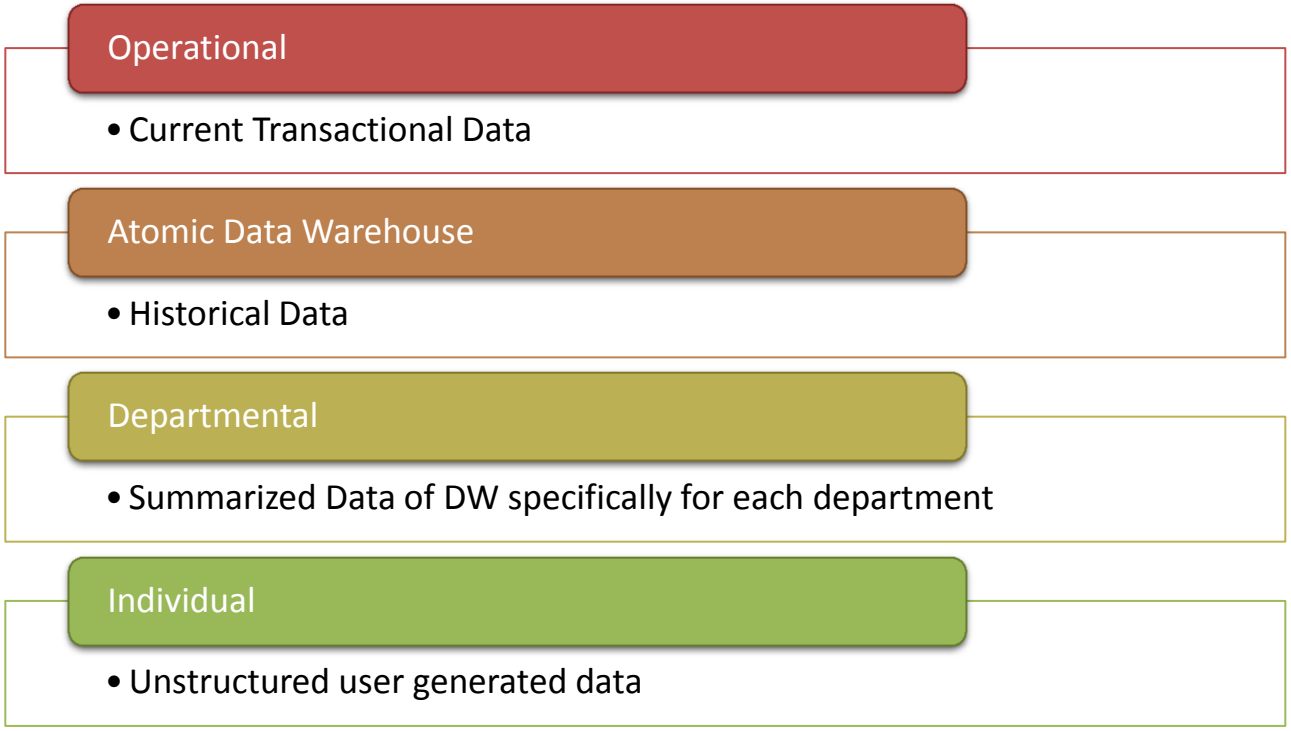
- In 1996, Kimball, a scholar-practitioner developed a model that compete Inmon's
- In 2002 he complete his model
- Recommends an architecture multiple databases, called **Data Marts**, organized by business processes
- The sum of Data Marts comprises the **Data Warehouse**
- **A BOTTOM-UP** approach that must adhere to an enterprise-wide standard "**Data Bus**"



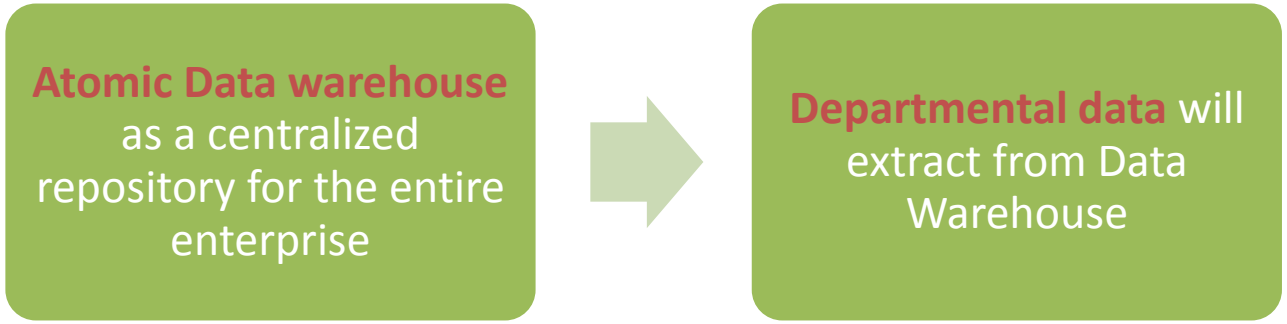
3. Inmon Model - Inmonities

Definition

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

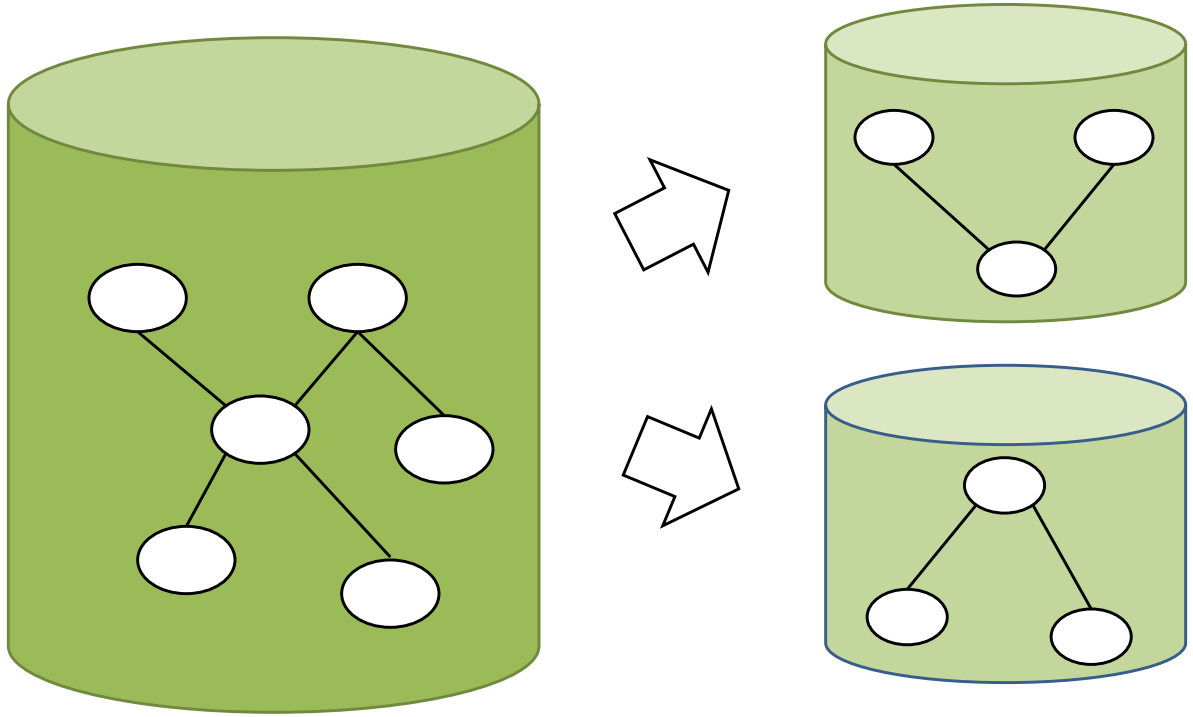


Inmon's Top-down design



Inmon Top-Down Schema

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



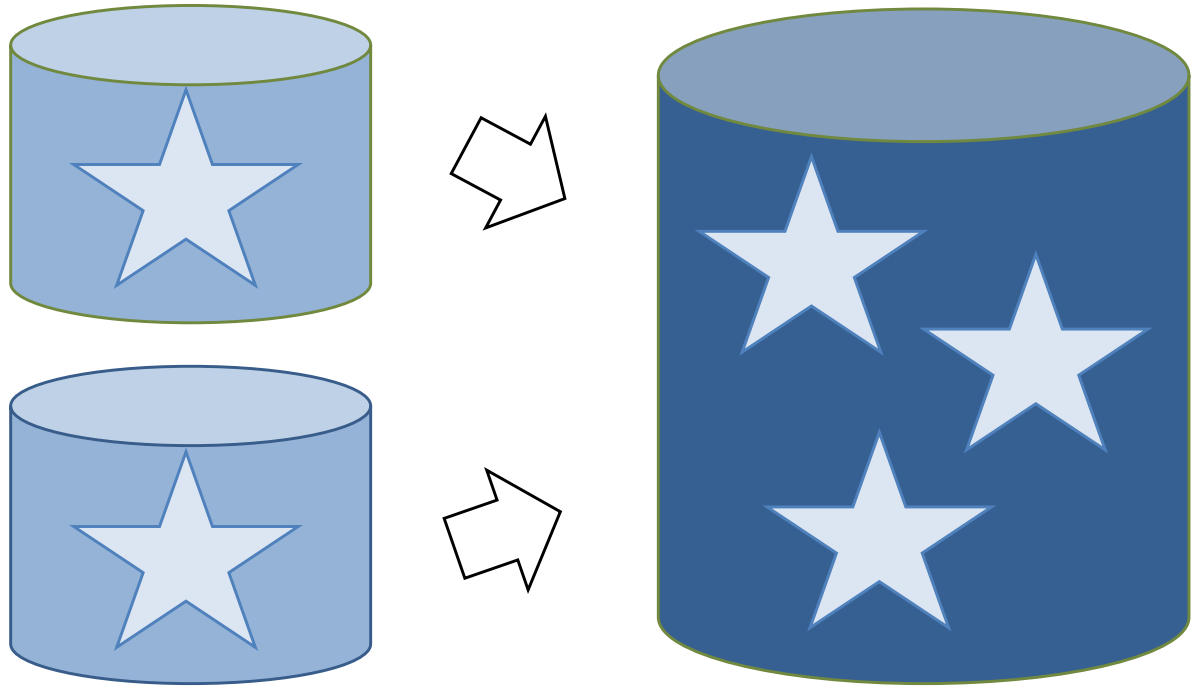
4. Kimball Model – Kimballities

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



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

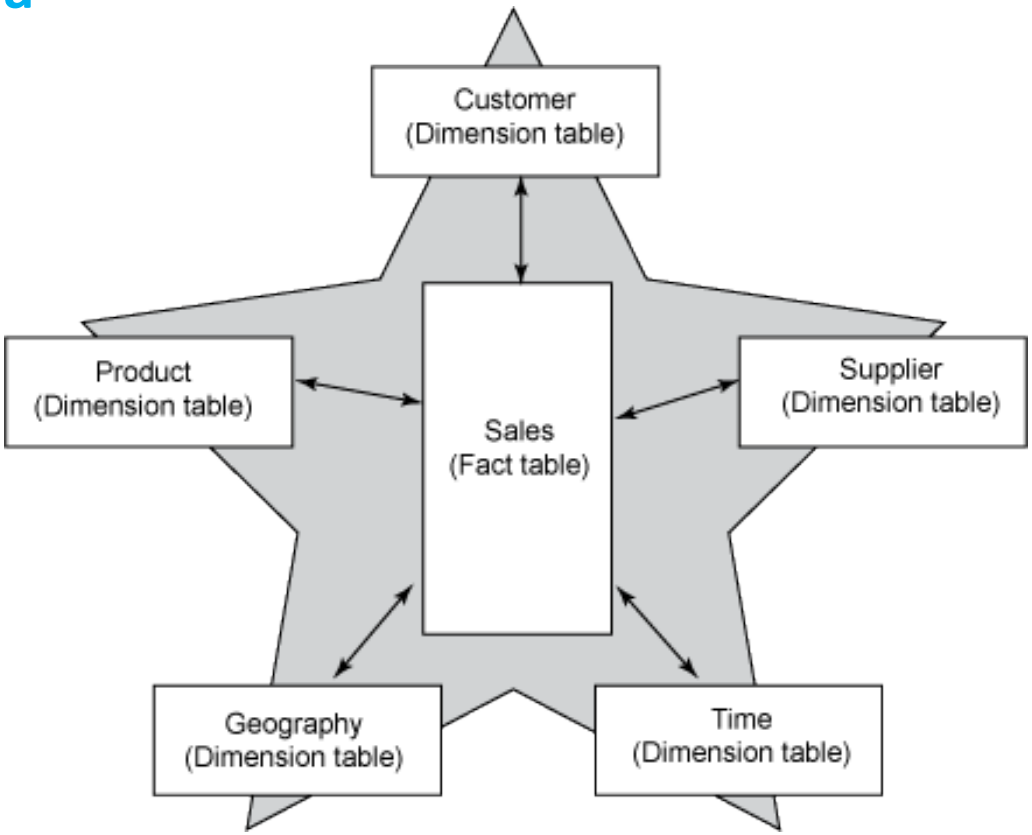
- Contain metrics
- Contain many rows and relatively few columns (for query performance)

Dimension Table

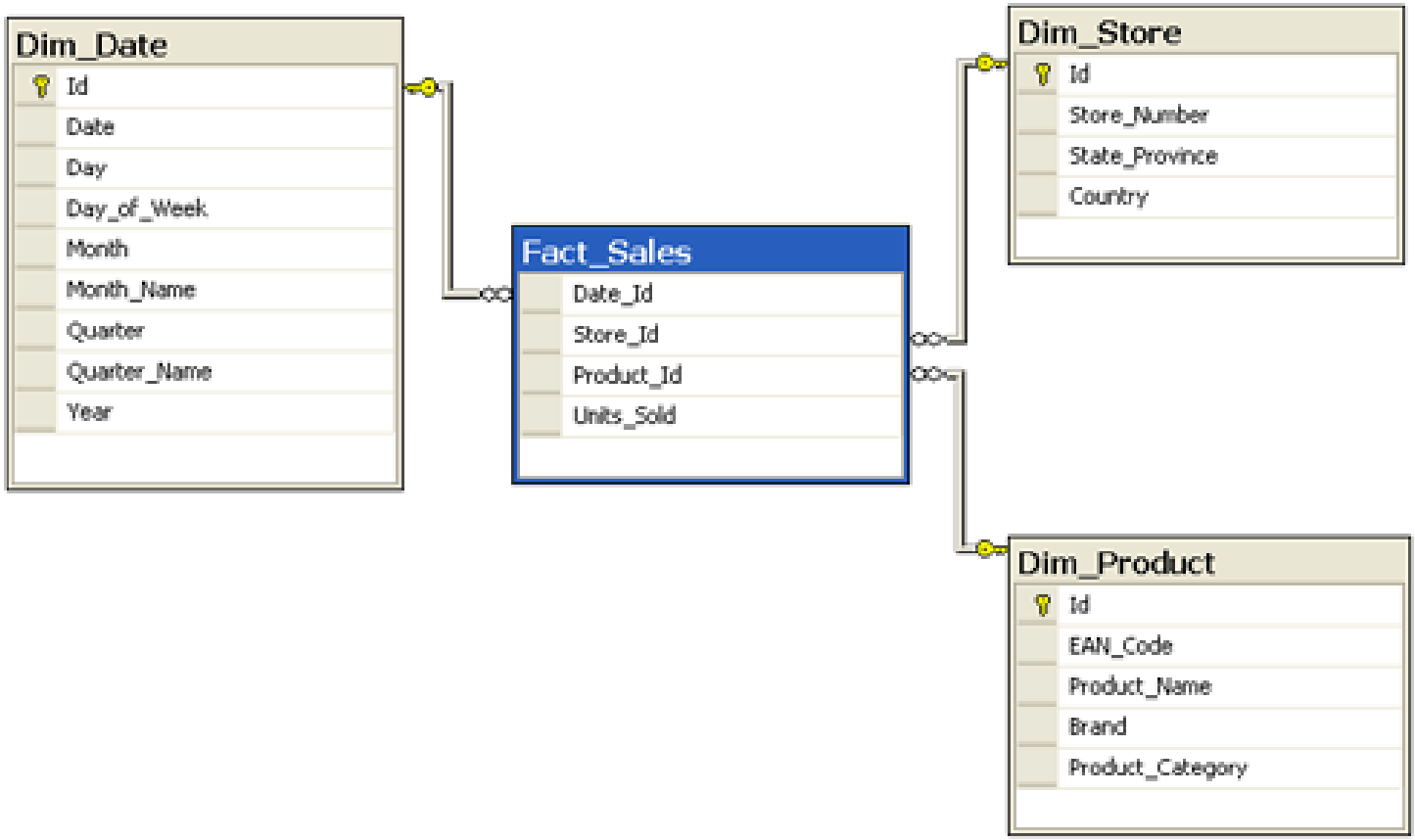
- Contain attributes of the metrics of fact table
- Have only hundreds or thousands of rows
- Hundred columns or more

Star Schema

- Relationship between Fact and Dimension Tables are in **Star Schema**



Fact and Dimension Tables in Star Schema



Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Time Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Fact Table - Transactions

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

1 0..* 1 0..* 0..* 1 1

Sale point Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

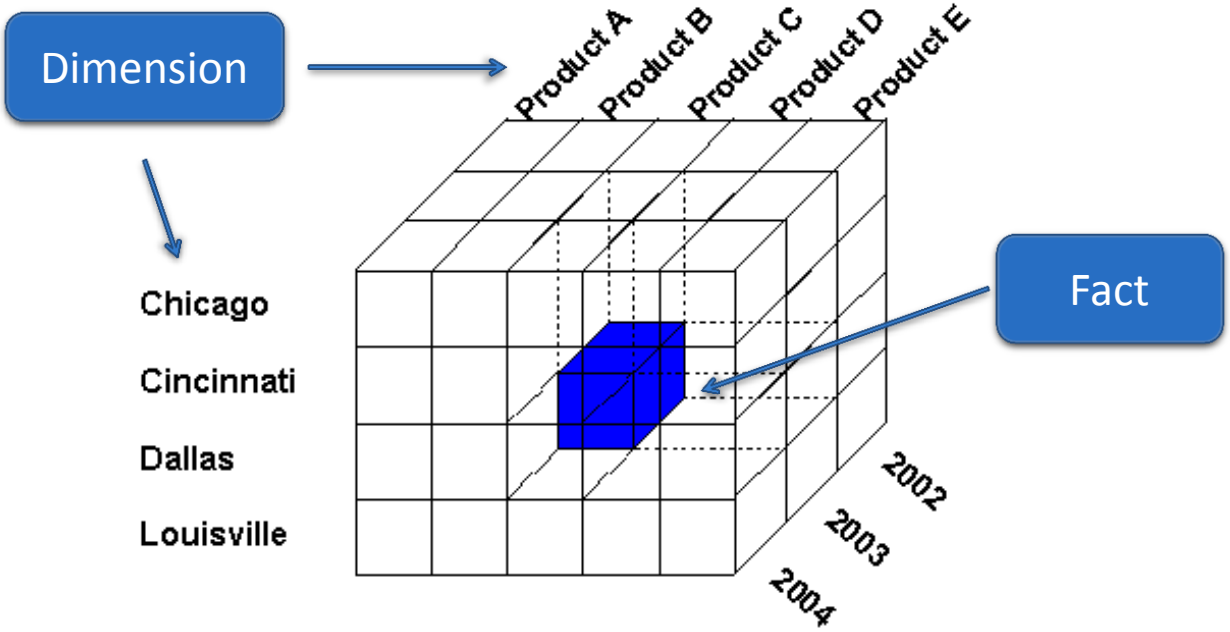
Example of Fact and Dimension table

4 dimensions: Service, Time, Sales Point, Customer

1 Fact: Transactions

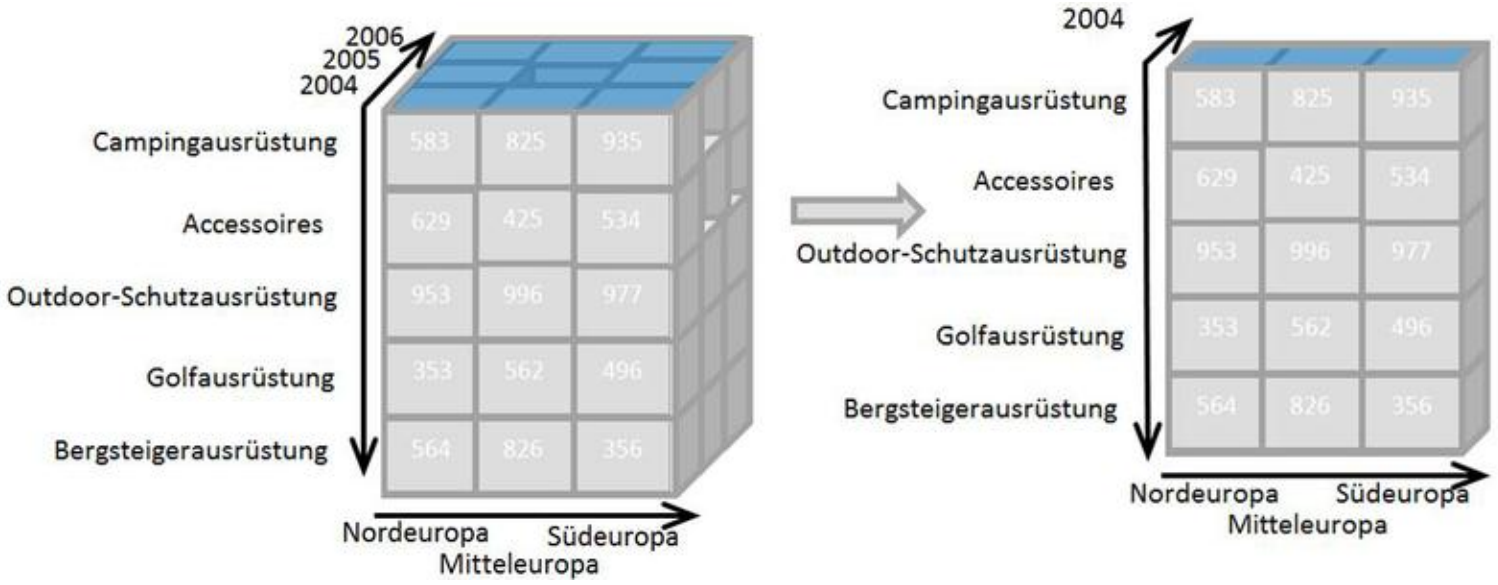
OLAP Fact-Dimension Cube

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



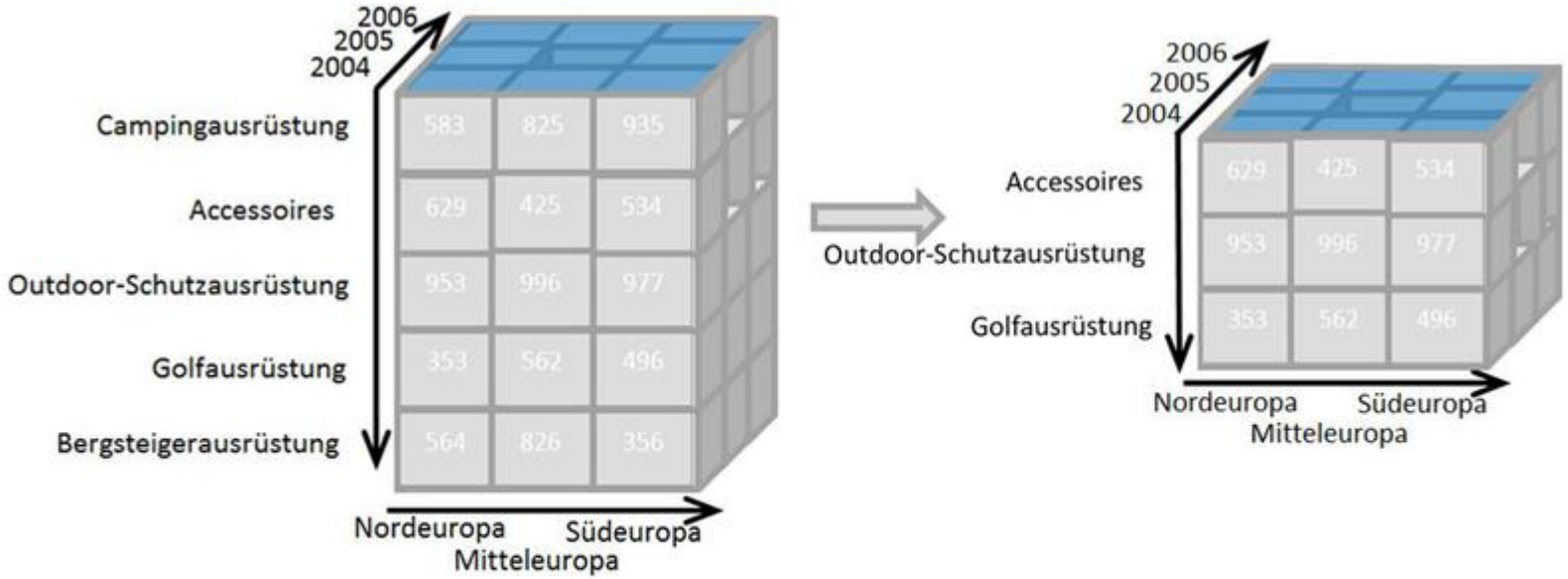
Operation: Slice

- **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**



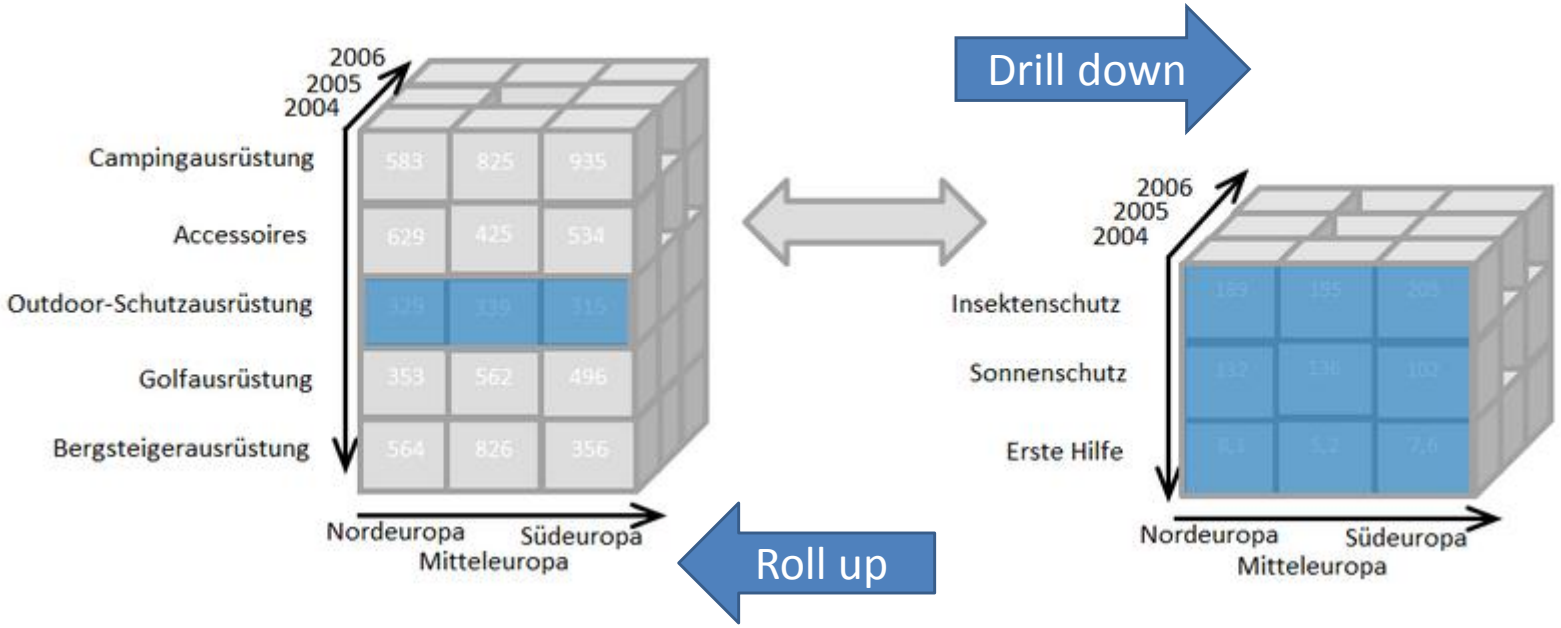
Operation: Dice

- **Dice** operation produces a **Subcube** by allowing the analyst to pick specific values of multiple dimensions

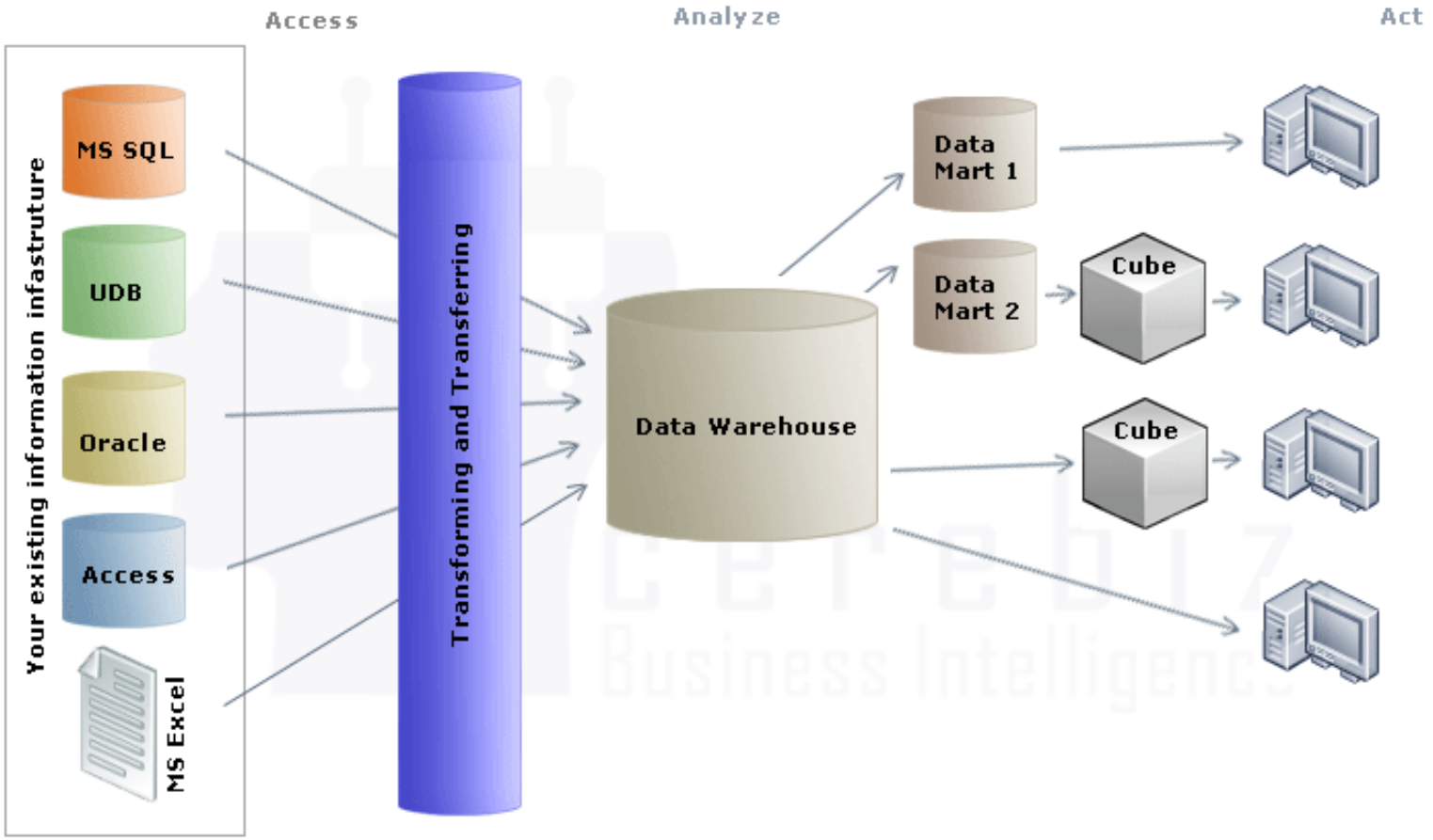


Operation: Drill Down/ Roll Up

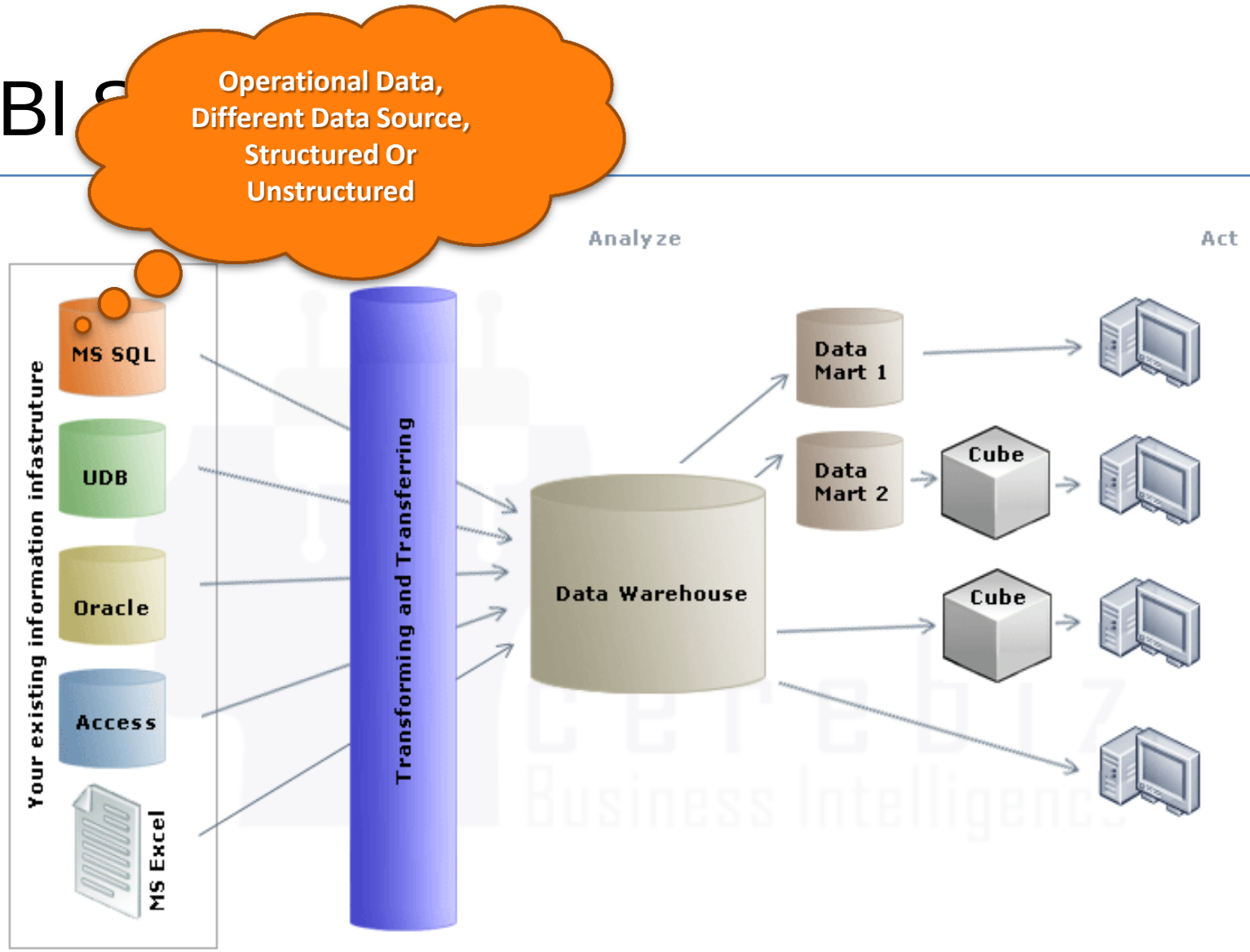
- **Drill Down/Roll Up** allows the user to **Navigate Among Levels Of Data** ranging from the **Most Summarized** (Roll Up) to the **Most Detailed** (Drill Down).



BI Structure

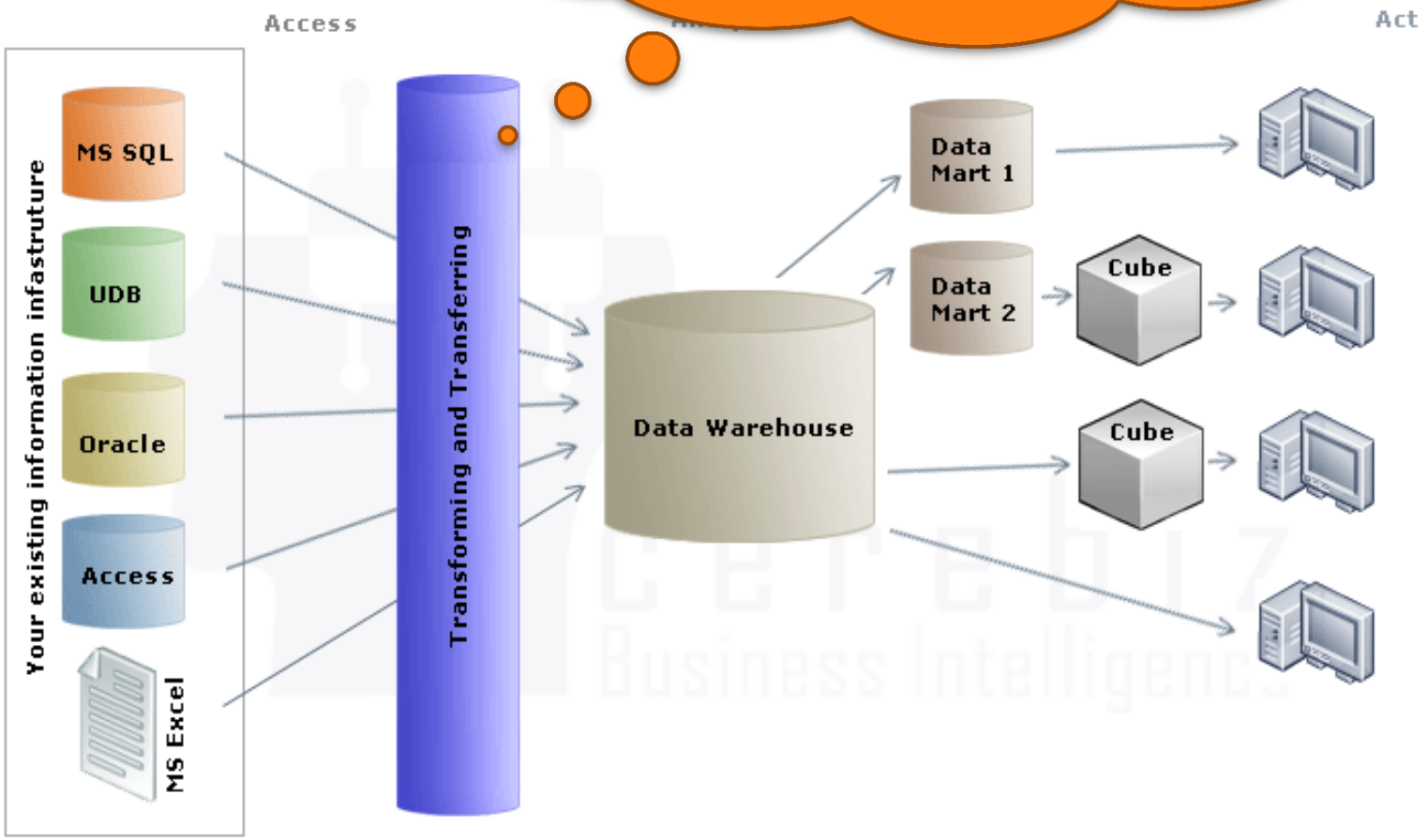


BI S

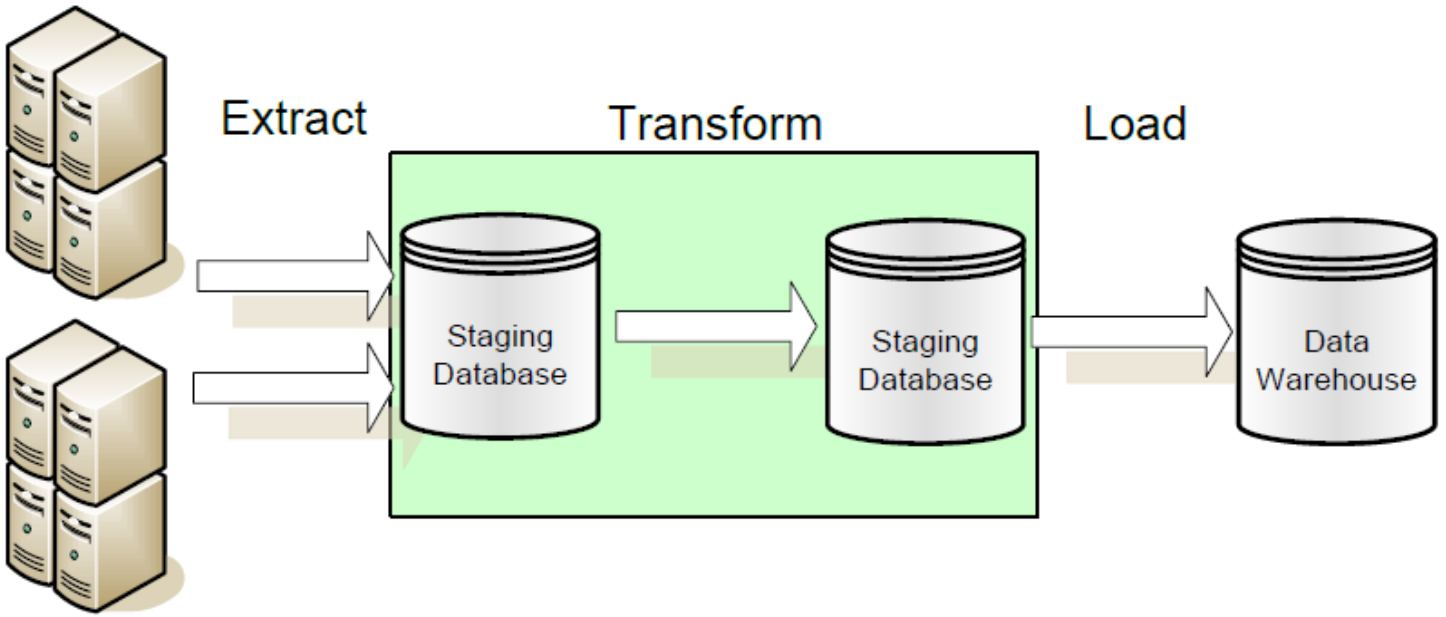

Operational Data,
Different Data Source,
Structured Or
Unstructured
Analyze
Act
Your existing information infastruture
MS SQL
UDB
Oracle
Access
MS Excel
Transforming and Transferring
Data Warehouse
Data Mart 1
Data Mart 2
Cube
Cube

BI Structure

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

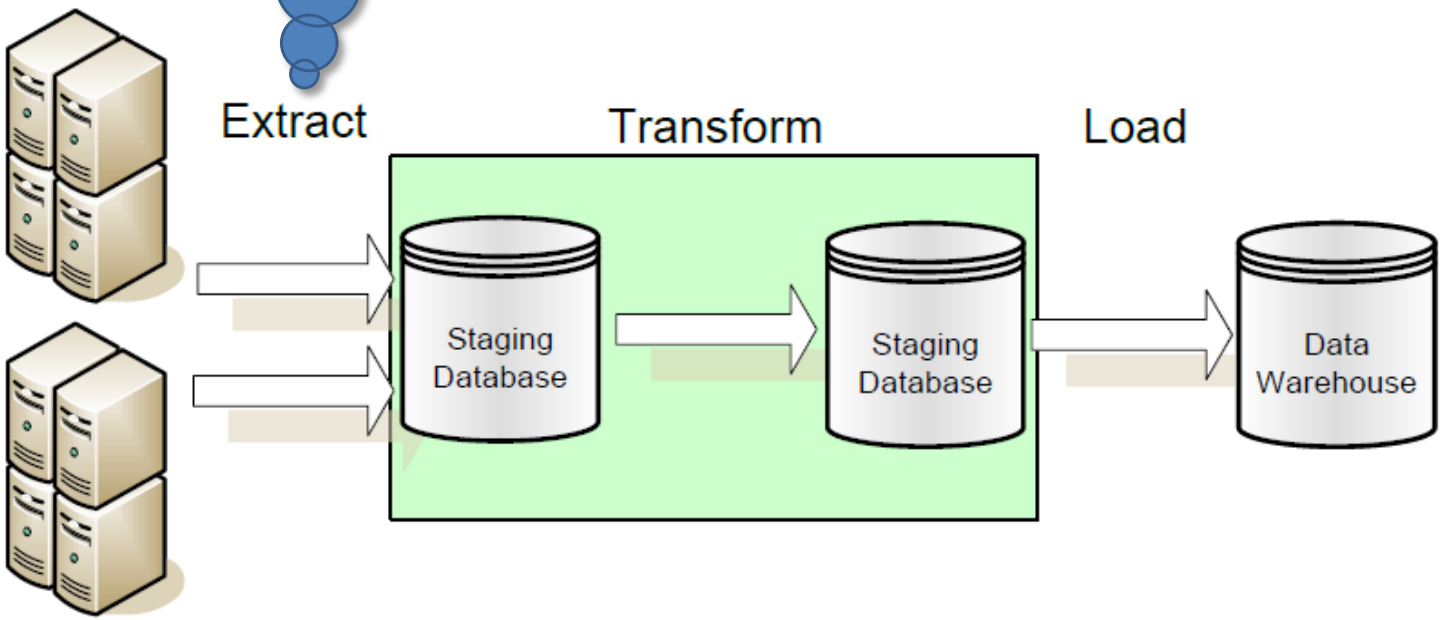


ETL

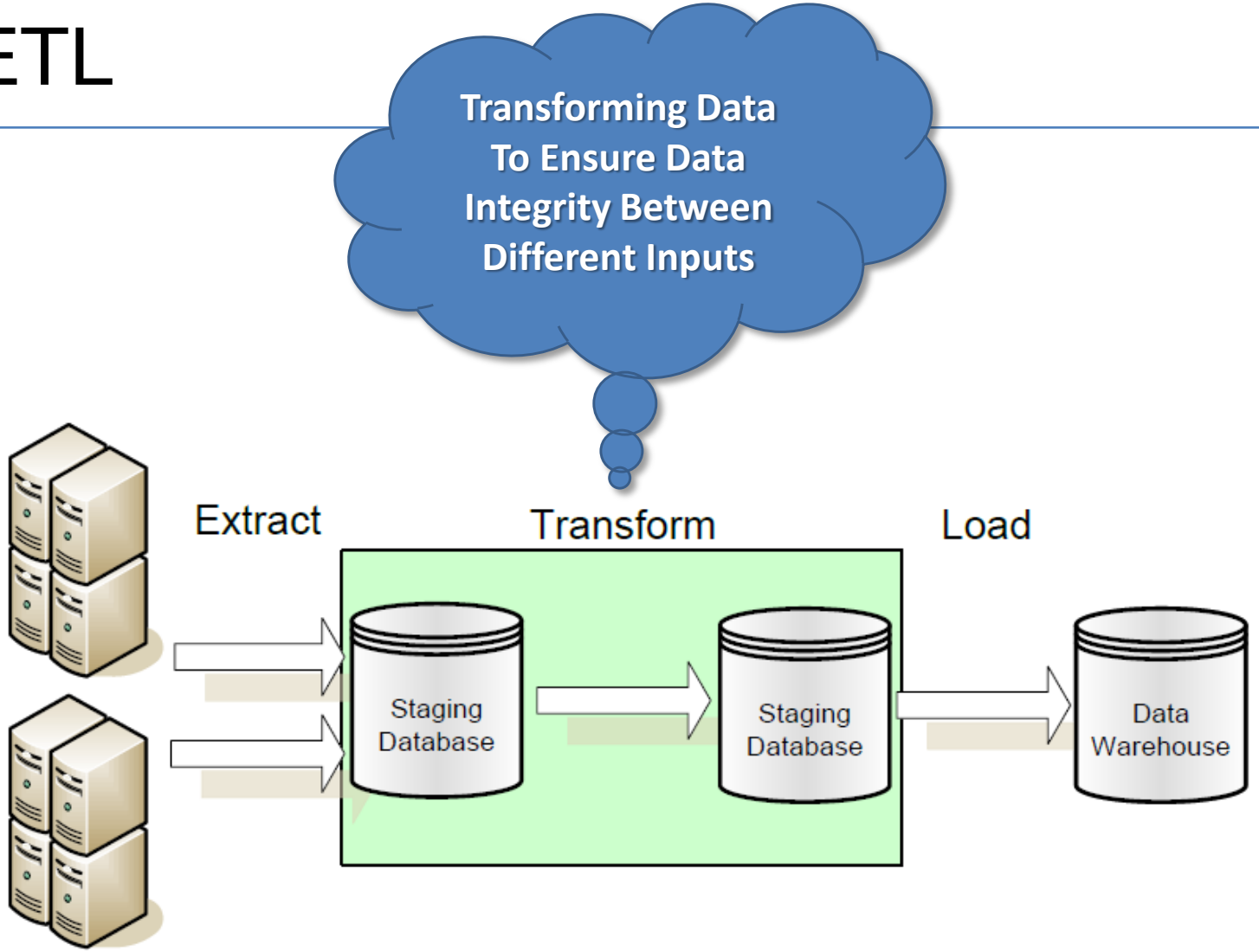

Extract
Transform
Load
Staging Database
Staging Database
Data Warehouse

ETL

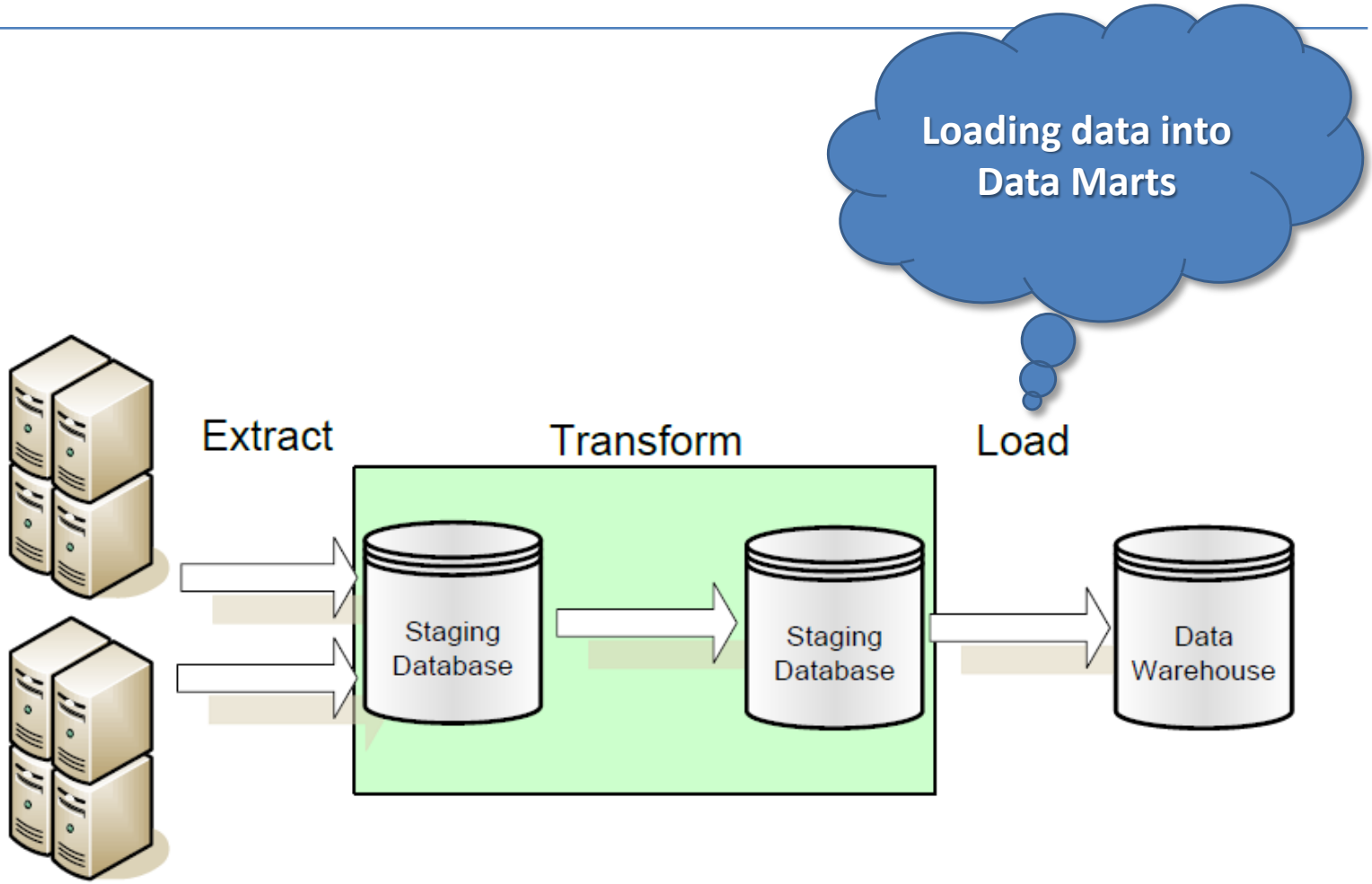
Moving data from operational systems to a persistent staging area



ETL

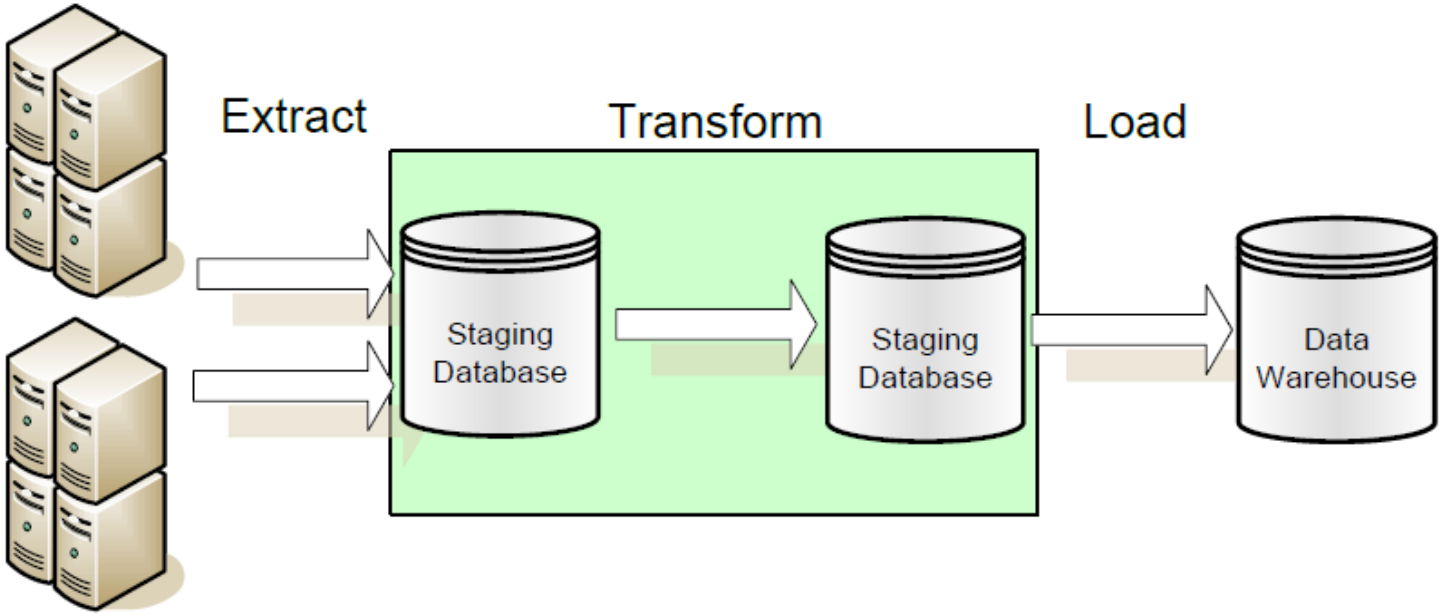

Transforming Data To Ensure Data Integrity Between Different Inputs
Extract
Transform
Load
Staging Database
Staging Database
Data Warehouse

ETL

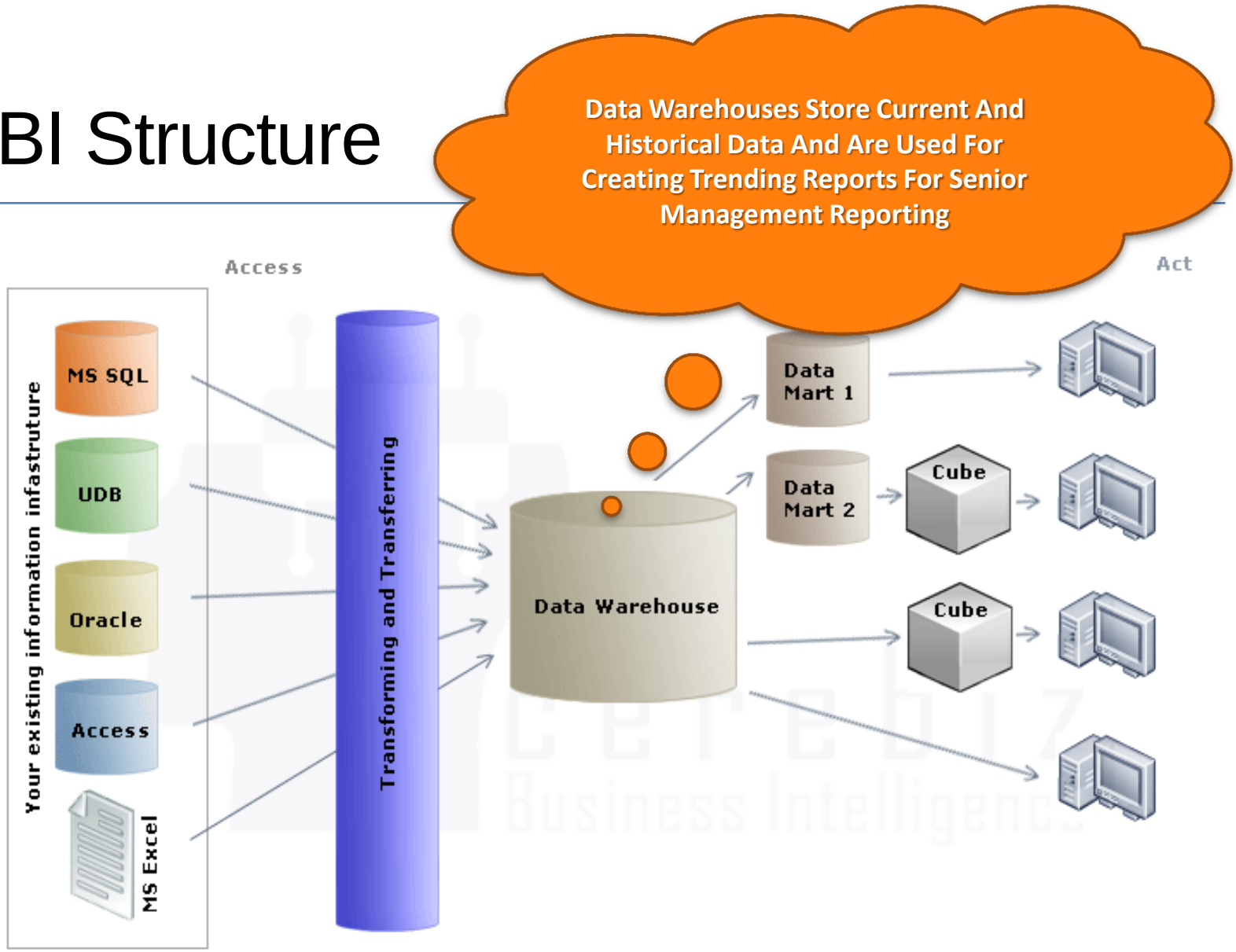

Loading data into
Data Marts
Extract
Transform
Load
Staging
Database
Staging
Database
Data
Warehouse

ETL

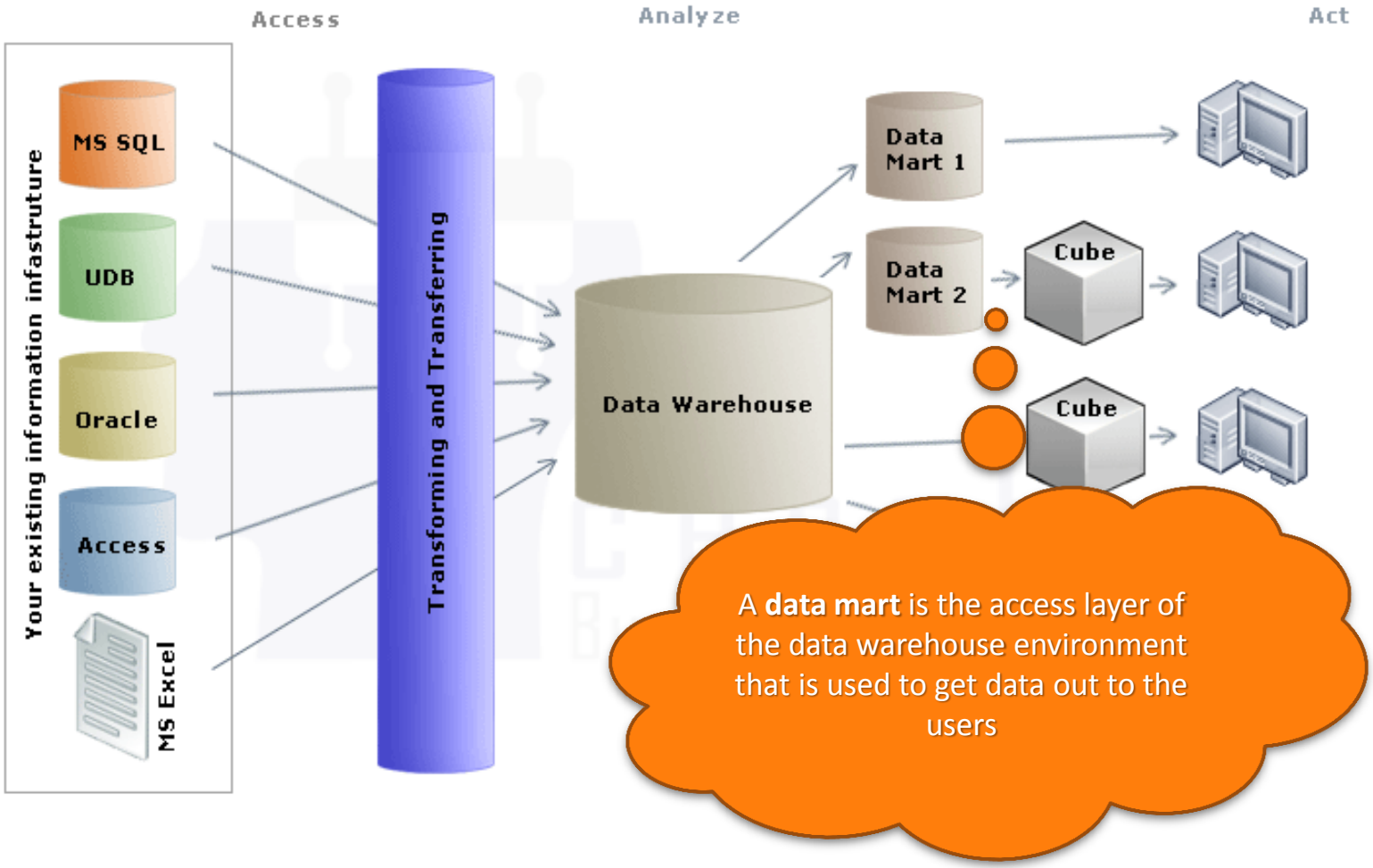
Garbage In–Garbage Out!



BI Structure



BI Structure



BI Structure

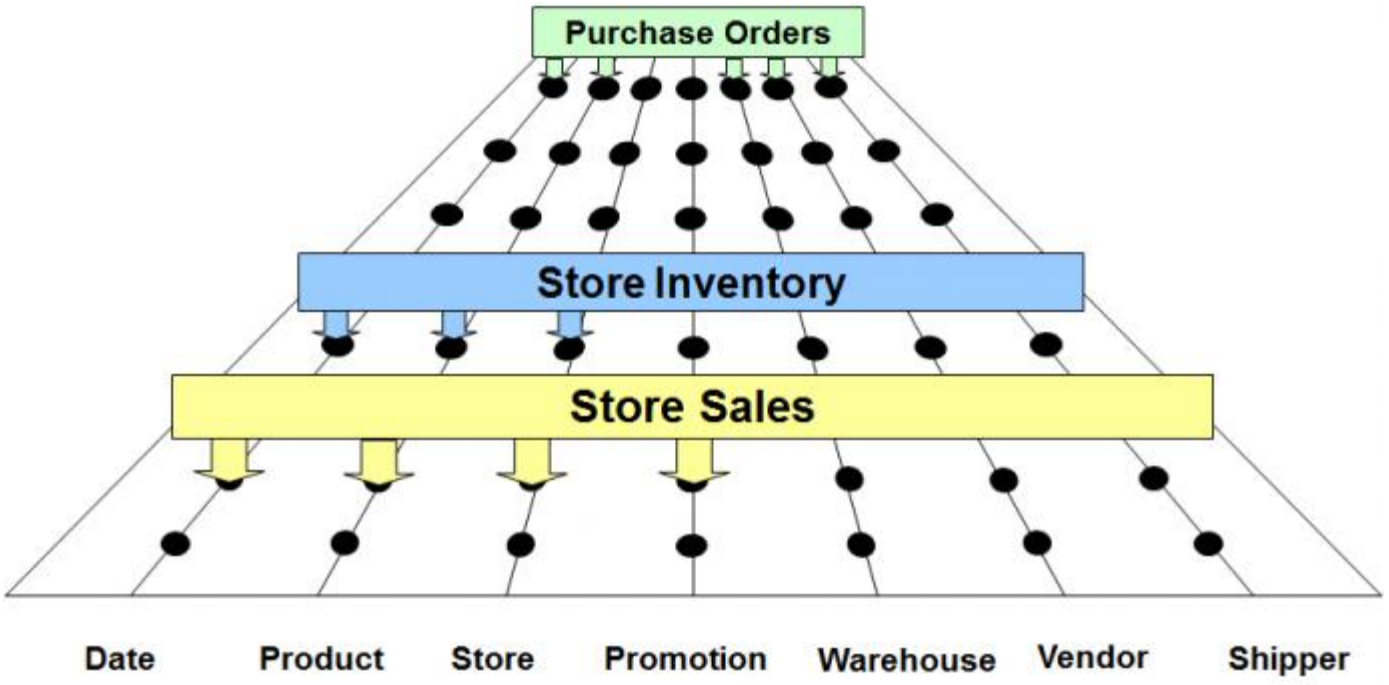
EDW Bus Architecture

Database-independent Bus Architecture	
Decomposes The DW/BI Process	By Focusing On The Organization's Core Business Processes By Using Conformed Dimensions
Conformed Dimensions:	Master Common Standardized Dimensions
	Created Once In The ETL
	Reused By Multiple Fact Tables

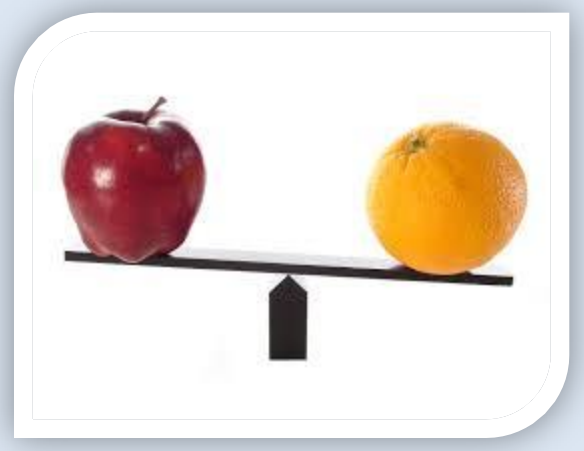
Bus Architecture – Example

BUSINESS PROCESSES	COMMON DIMENSIONS						
	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

Bus Architecture – Example.



Data warehouse



5. Inmon vs. Kimball

Inmon vs. Kimball

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


Amount by Time
Revenue
Predicted Sales
Amount by Country
Top Stories
Next Meeting Location

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

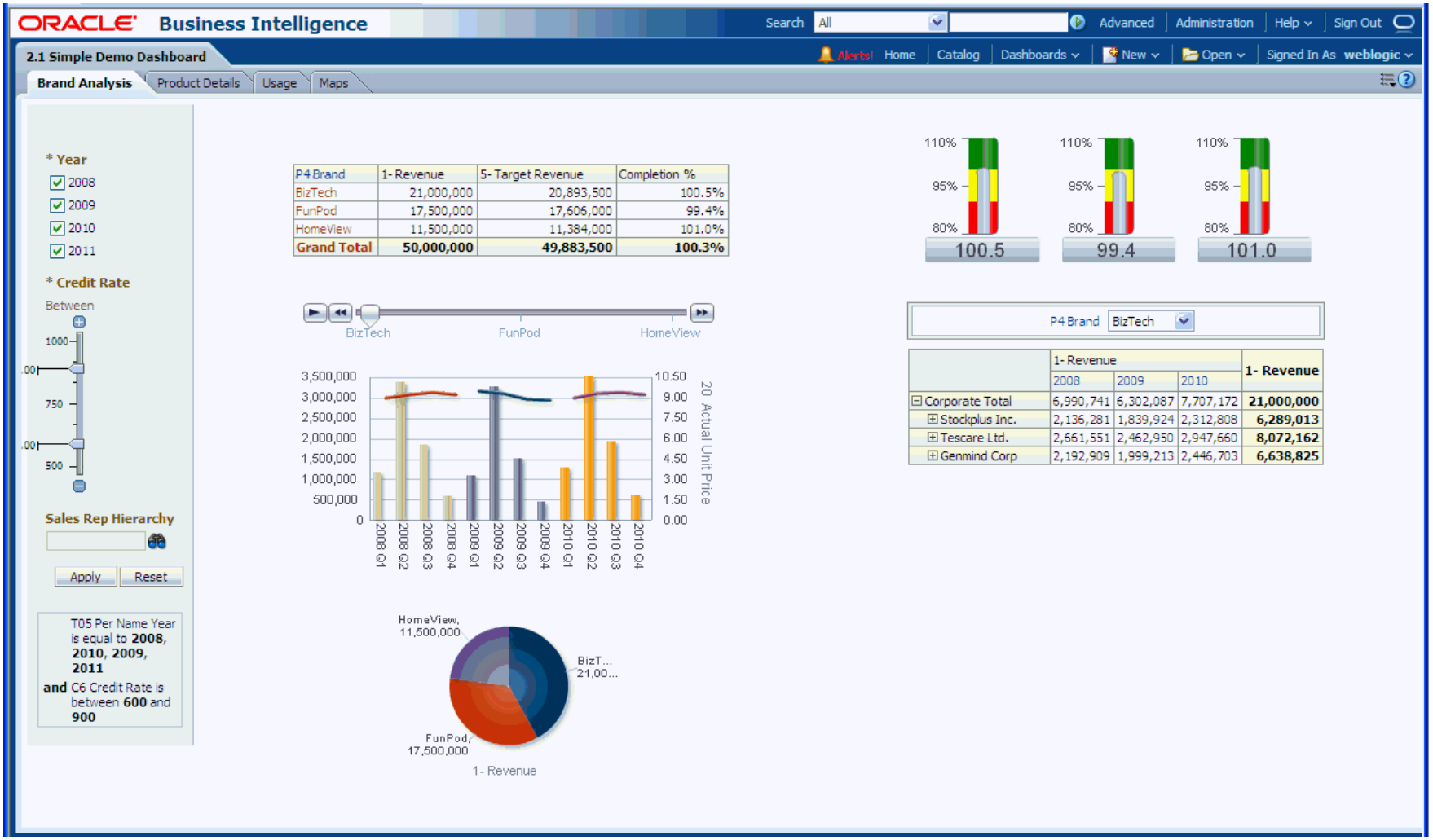
- Contain high-level, aggregated company strategic data with comparisons and performance indicators
- Include both static and interactive reports
- Lots of graphics, charts and illustrations

Write-back reports

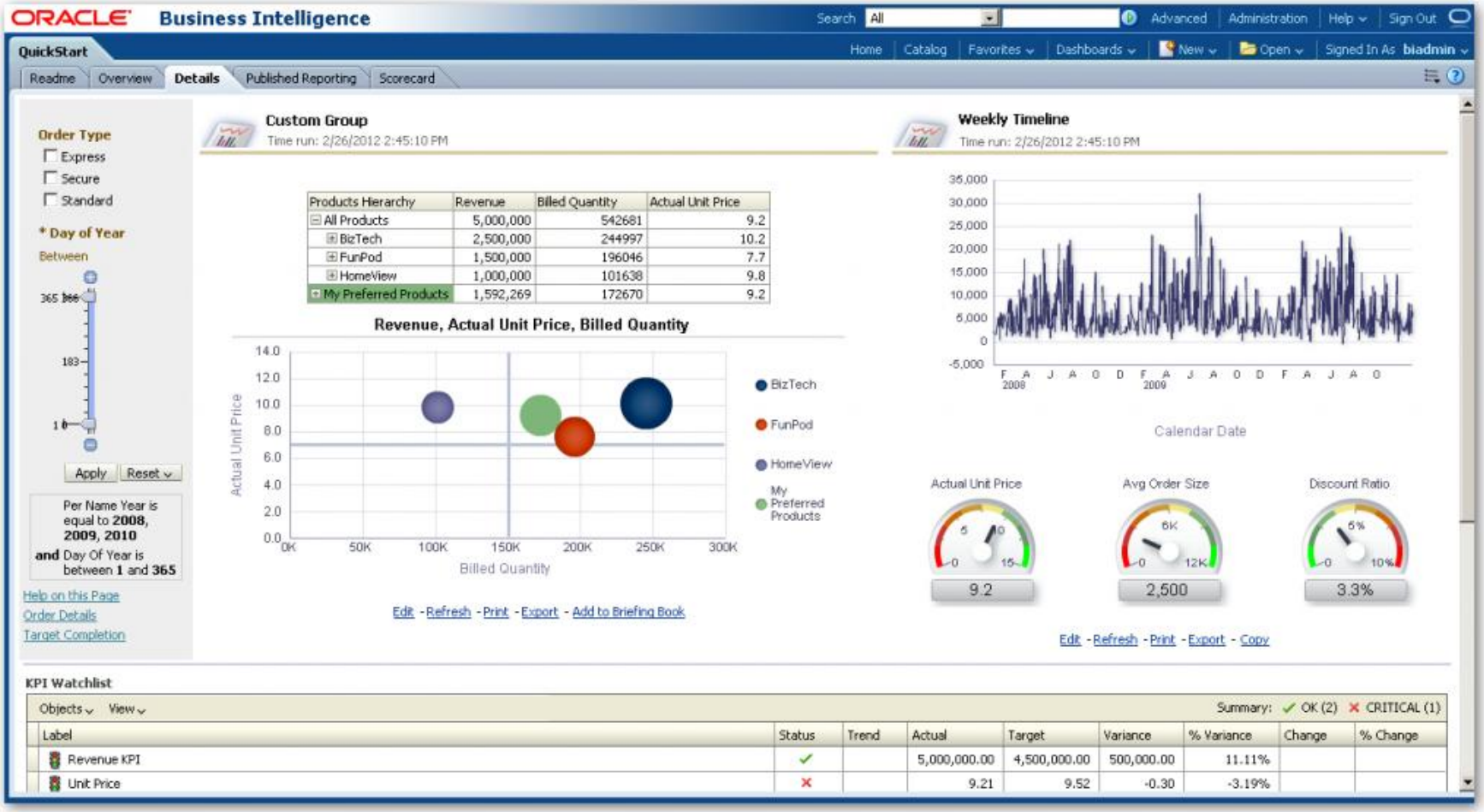
- Those are interactive reports directly linked to the Data Warehouse which allow modification of the data warehouse data.
  By far the most often use of this kind of reports is:
  - Editing and customizing products and customers grouping
  - Entering budget figures, forecasts
  - Setting sales targets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



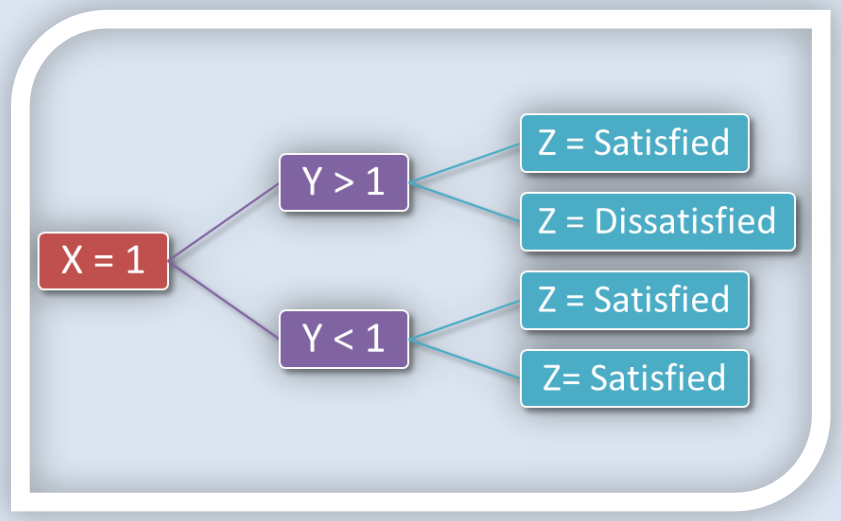
Sample BI dashboard



Sample BI dashboard

Most widely used BI Systems:

IBM Cognos
SAP Business Objects and Crystal Reports
Oracle Hyperion and Siebel Analytics
Microstrategy
Microsoft Business Intelligence (SQL Server Reporting Services)
SAS
Pentahoo Reporting and Analysis
BIRT - Open Source Business Intelligence and Reporting Tools
JasperReports
Qlickview



7. BI Algorithms

Popular algorithms used by BI software

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

Regression Analysis

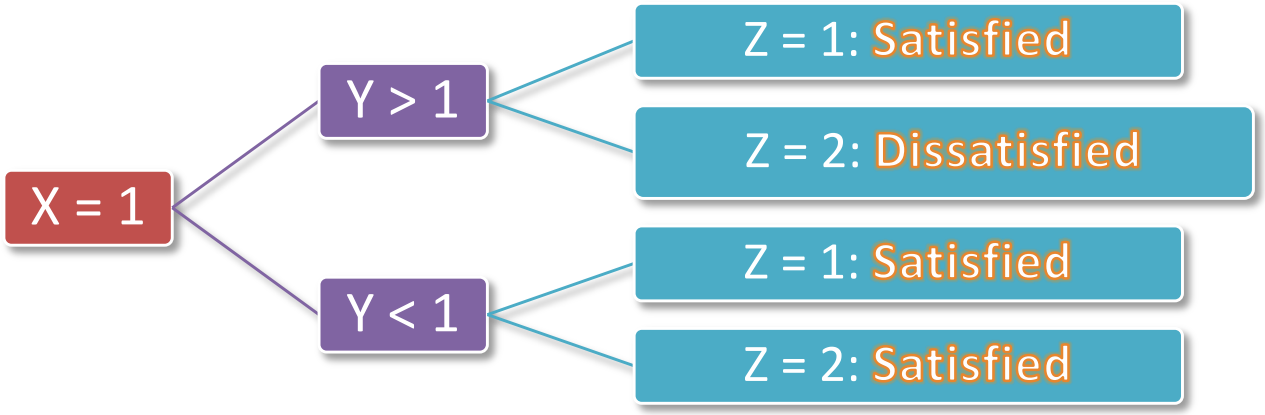
- Y = aX + b
- Example: Profit is Linear/Non-linear function of Revenue, so we forecast future Profit by assessing historical information

	2012	2013	2014
Revenue	1,000	2,000	?
Profit	200	300	?

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



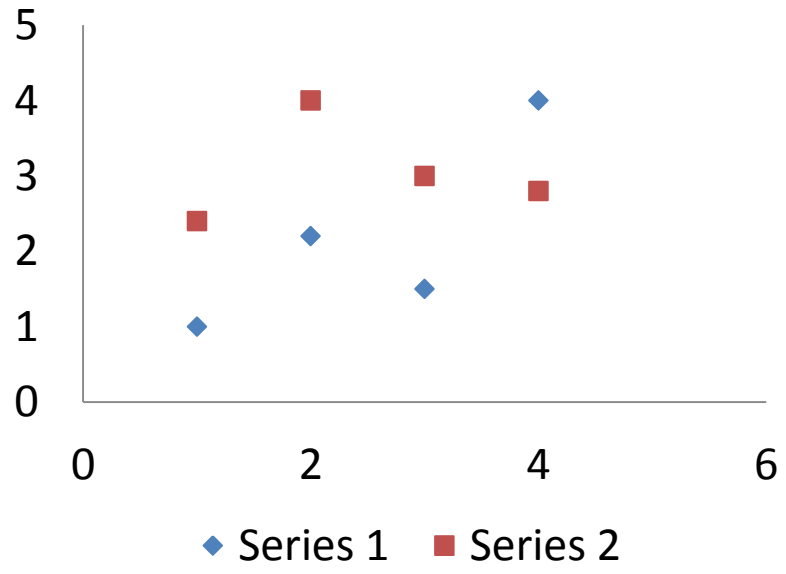
Association Analysis

- 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 : $I = \{I_1, I_2, \ldots, I_m\}$, $T \subseteq I$ (T *is a Transaction*), $X \subseteq T$
- Define: $X \Rightarrow Y \quad \Leftrightarrow \quad Y \subseteq T \;\&\; X \cap Y = \emptyset$

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

- David Butler, Bob Stackowiak, "*Master Data Management, An Oracle White Paper",* June 2009
- Wikipedia.com
- businessintelligence.com/dictionary/
- www.vanguardsw.com/products/vanguard-system/business-intelligence.htm
- swiki.net/reliable-business-intelligence-datawarehouse-datamart-and-datamining.html
- www.oracle.com/technetwork/articles/madison-models-086845.html

9. References

- www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/kimball-data-warehouse-bus-architecture/
- 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
- http://help.sap.com/
- Chih-Fong Tsai, “Mao-Yuan Chen, Variable selection by association rules for customer churn prediction of multimedia on demand”, Elsevier, 2009

Thanks for your attention