Welcome!   Module 2

- Data Warehouse – Business Intelligence Concepts

"A collection of integrated, subject-oriented databases designed to support the DSS function where each unit of data is relevant to some moment in time..."

Inmon, Imhoff and Sousa, The Corporate Information Factory

"A copy of transaction data specifically structured for query and analysis."

Ralph Kimball, The Data Warehouse Toolkit

Introduction

About Me

- **Parwaz Dalvi**

  Sr. Architect / Consultant DW-BI & Database

  TOGAF 8 Certified (The Open Group Architecture Framework)

My Session For you

- **Data Warehouse Concepts**

Session's Objective

- Understand what Data Warehousing means
- Realize the Need, Advantages & Challenges in implementation of a DW Solution
- Understand Data Warehouse Architecture and its components
- Understand IBM Reference DW-BI Architecture
- Understand IBM's IOD initiative and realize how DW-BI helps in achieving this objective
- Know the DW-BI Tools and Products, the trends in DW-BI
- Know your Growth Prospects in the DW-BI Arena within IBM

Course Content

Module	Content	Duration
1	Data Warehouse Evolution	
2	*Data Warehouse Concepts*	
3	Data Warehouse Architecture – Part 1 – GENERIC	
4	Data Modeling in DW-BI	
5	Data Warehouse Architecture – Part 2 – SPECIFIC	
6	DW-BI - IBM Reference Architecture & IOD	
7	DW-BI Tools and Products	
8	Trends in DW-BI	
9	Growth Path of DW-BI Professionals	

Course Title:

Module 2 : Data Warehouse Concepts

Module Objectives

- At the completion of this chapter you should be able to understand:

    – What is OLTP, OLAP & ODS & difference between them

    – Why do we need to build an OLAP for a Data Warehouse

    – Steps to transform OLTP to OLAP

    – Differentiating factors between OLTP, DW, DSS, BI & OLAP

    – What are the Data Warehouse system qualifiers

Module 2: Data Warehouse Concepts : Agenda

- Topic 1. OLTP, OLAP, ODS

- Topic 2. Why OLAP for Data Warehouse

- Topic 3. OLTP to OLAP transformation steps

- Topic 4. Differentiate between OLTP, OLAP, & DW

- Topic 5. Data Warehouse system qualifiers
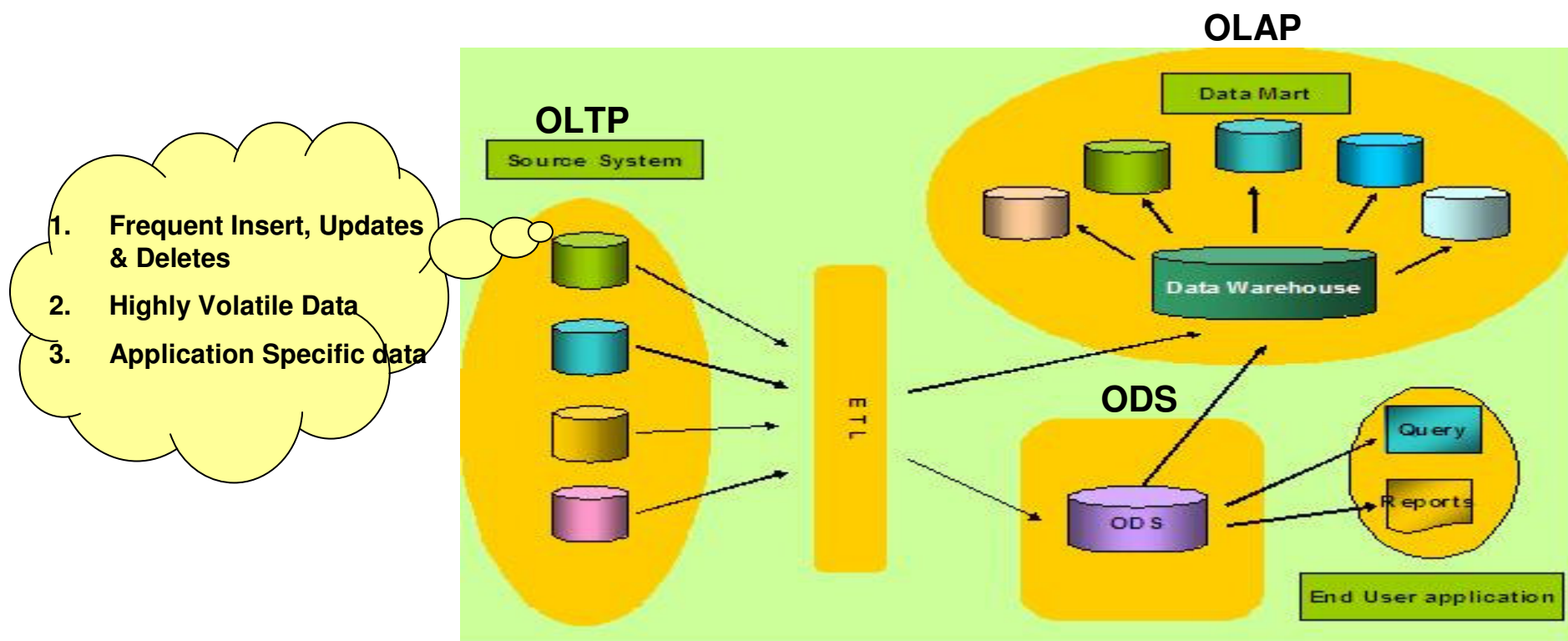
Module 2: > Topic 1: OLTP, OLAP, ODS

OLTP

- OLTP Definition

- OLTP Examples

- OLTP Description

- OLTP Characteristics

Module 2: > Topic 1: OLTP, OLAP, ODS

OLTP – On-line Transaction Processing - Definition

A system that keeps track of an establishment's daily transactions and updates the warehouse at periodic intervals

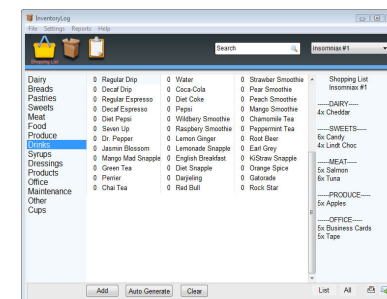1. **Frequent Insert, Updates & Deletes**
2. **Highly Volatile Data**
3. **Application Specific data**

Module 2: > Topic 1: OLTP, OLAP, ODS

OLTP – Examples

Cash Register at counter
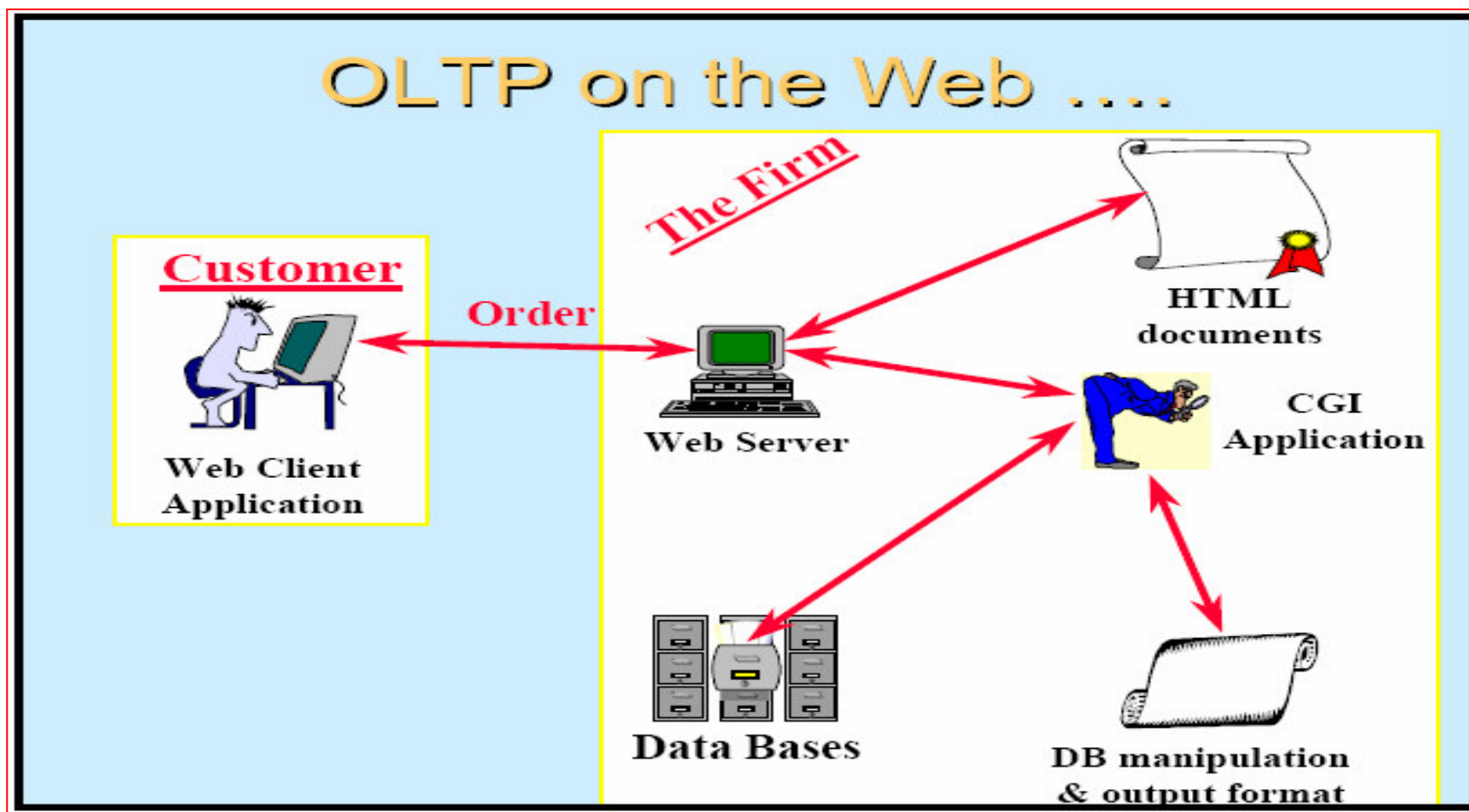
Stock update at back door

Inventory Management System

Module 2: > Topic 1: OLTP, OLAP, ODS

OLTP – Examples

Module 2: > Topic 1: OLTP, OLAP, ODS

OLTP – Description

By OLTP (On-Line Transaction Processing) systems, we mean a class of systems which facilitate and manage transaction-oriented applications, mainly data entry and retrieval transactions

Most of the systems that are used in the day to-day business are of OLTP type, such as:

- Order entry

- Inventory management

- Railway Reservation system

- Payroll or production tracking

Module 2: > Topic 1: OLTP, OLAP, ODS

OLTP – Characteristics

- Majority of transaction are record entry. In Some cases retrieval of records and then modify or delete.

- Millions of transactions per day

- The system is up to date with current state of business transaction.

- Non-Integrated, Application specific, Volatile data

- ER data model
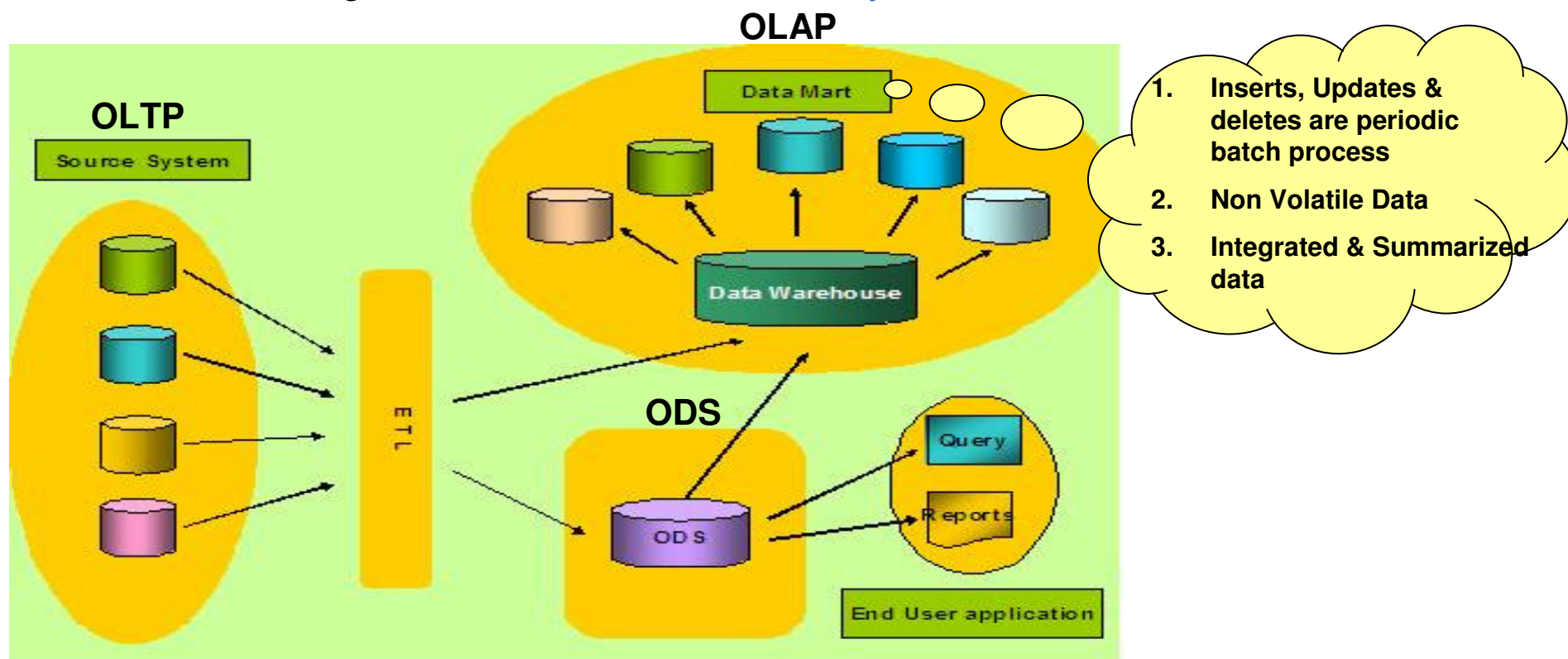
Module 2: > Topic 1: OLTP, OLAP, ODS

OLAP

- OLAP Definition

- OLAP Examples

- OLAP Description

- OLAP Characteristics

Module 2: > Topic 1: OLTP, OLAP, ODS

OLAP – On-line Analytical Processing - Definition

Is a technology that uses a <u>multidimensional view</u> of <u>aggregate data</u> to provide quick access to strategic information for further <u>analysis</u>

OLAP

Data Mart

OLTP

Source System

Data Warehouse

ODS

ODS

Query

Reports

ETL

End User application

1. **Inserts, Updates & deletes are periodic batch process**
2. **Non Volatile Data**
3. **Integrated & Summarized data**

Module 2: > Topic 1: OLTP, OLAP, ODS

OLAP – Examples

- Target Marketing

- Market Segmentation

- Market based Analysis – POS Analysis

- Churn Analysis

- Profitability Management

- Event Tracking

Module 2: > Topic 1: OLTP, OLAP, ODS

OLAP – Examples

Manufacturers

Customers

- Sales – Target vs. Achieved

- Customers

- Credit Rating Agencies

- Financial Reporting & Consolidation

- Material Management

- Product Lifecycle

Module 2: > Topic 1: OLTP, OLAP, ODS

OLAP – Description

OLAP (On-Line Analytical Processing) systems are those, which enables end-users to perform ad hoc analysis of data in multiple dimensions, thereby providing the insight and understanding they need for better decision making

The approach is to quickly provide answers to analytical queries

Typical OLAP applications are:

– Business Reporting for Sales, Marketing, Management

– Budgeting & Forecasting

– Financial Reporting

Module 2: > Topic 1: OLTP, OLAP, ODS

OLAP – Characteristics

- Global

- Summarized and Integrated data

- Non-Volatile data

- Not focused on detailed pieces of information

- Data Model and representation is business friendly

Module 2: > Topic 1: OLTP, OLAP, ODS
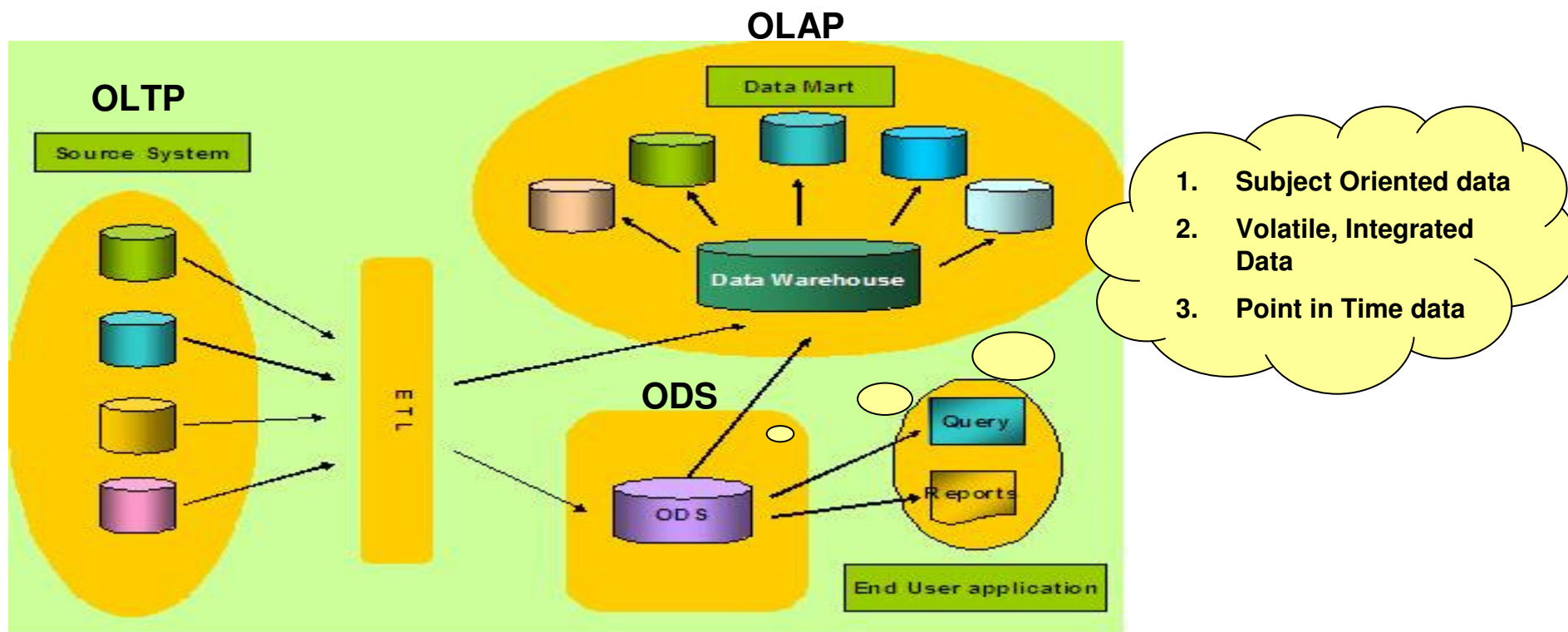
ODS

- ODS Definition

- ODS Examples

- ODS Description

- ODS Characteristics

Module 2: > Topic 1: OLTP, OLAP, ODS

ODS – Operational Data Store - Definition
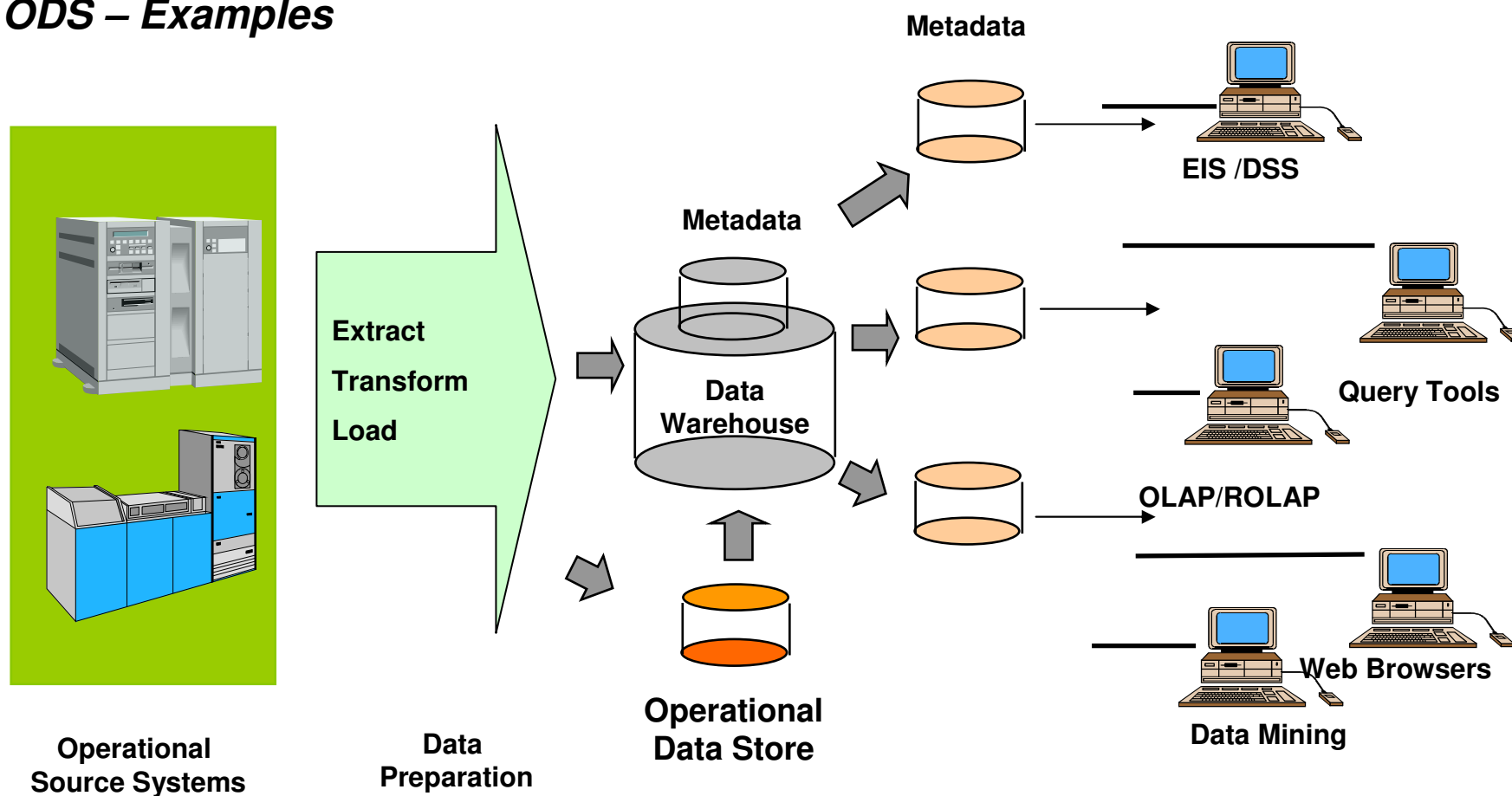
It is a subject oriented, integrated, volatile, current valued data store containing only corporate detailed data

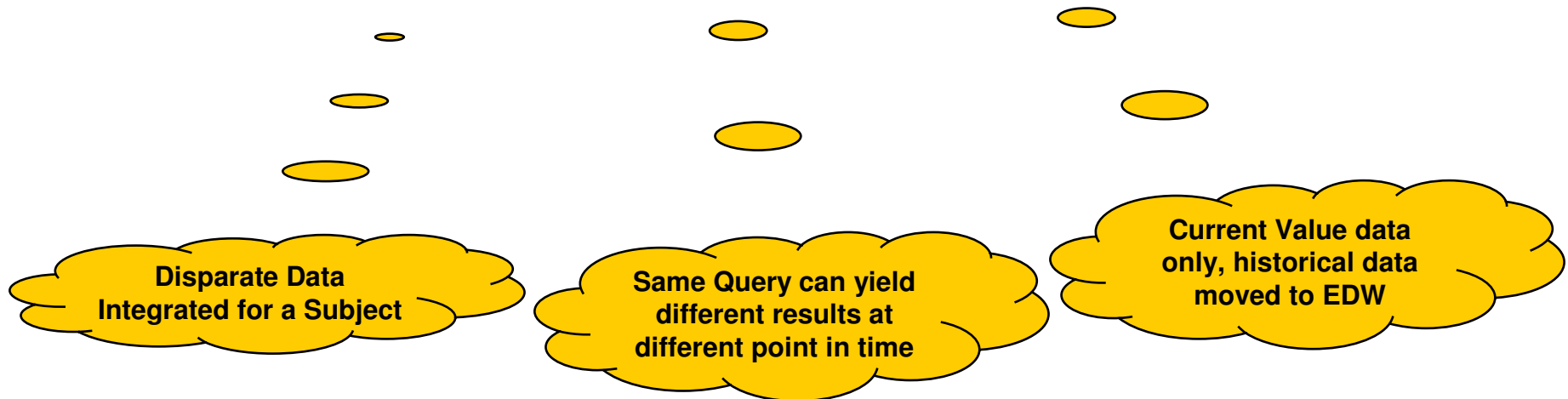# Module 2: > Topic 1: OLTP, OLAP, ODS

ODS – Examples

Metadata

Metadata

Extract

Transform

Load

Data Warehouse

EIS /DSS

Query Tools

OLAP/ROLAP

Web Browsers

Data Mining

Operational Source Systems

Data Preparation

Operational Data Store

Module 2: > Topic 1: OLTP, OLAP, ODS

ODS – Description

By ODS (Operational Data Store) systems, we mean a class of systems which is a <u>subject oriented</u>, <u>integrated</u>, <u>volatile</u>, <u>current valued</u> data store containing only corporate <u>detailed</u> data

Disparate Data Integrated for a Subject

Same Query can yield different results at different point in time

Current Value data only, historical data moved to EDW

Module 2: > Topic 1: OLTP, OLAP, ODS

ODS – Characteristics

- Is is meant only for Operational Systems

- Contains current value and near current values

- Contains detailed data

- Is meant for day to day decisions & operational activities

- Facilitates Operation Reports & Statistics by means of OPER-MART

Module 2: > Topic 1: OLTP, OLAP, ODS

OLTP vs. ODS vs. OLAP

Characteristics	OLTP	ODS	OLAP
Used for What	Day to Day Transaction	Operational activities & Day-to-day decisions	Information Managemnt in an Enterprise
Database Size	Moderate	Moderate	Very Large
Data Load	Field by Field	Field by Field	Batch Upload
Accesed By	Operational Users	Analyts & Operational Users	Executives, Managers & Analysts
Kind of Data	Individual records	Individual records	Set of Records
Type of Data	Transaction	Transaction & Analysis	Analysis

Module 2: > Topic 1: OLTP, OLAP, ODS

OLTP vs. ODS vs. OLAP

Characteristics	OLTP	ODS	OLAP
Methodology	Operational Requirements	Data Driven	Evolutionary
Structure of Data	Detailed	Detailed & Lightly Summarized	Higly Summarized
Organization of Data	Functional	Subject Oriented	Subject Oriented
Source of Data	Homogeneous, Application Centric	Homogenous	Heterogeneous
Redundancy of Data	Not Redundant	Redundant to some extent	Manged Redundancy
Structure of Data	Detailed	Detailed & Lightly Summarized	Higly Summarized

Module 2: > Topic 1: OLTP, OLAP, ODS   Summary

- Having completed this topic, you should be able to:

    – OLTP Definition, Examples, Description & Characteristics

    – OLAP Definition, Examples, Description & Characteristics

    – ODS Definition, Examples, Description & Characteristics

    – Differences between OLTP, ODS & OLAP

Module 2: > Topic 1: OLTP, OLAP, ODS   Review

Module 2: > Topic 2: Why OLAP for Data Warehouse

- OLAP Integrates Data

- OLAP Improve Performance

- OLAP models data specifically for DSS
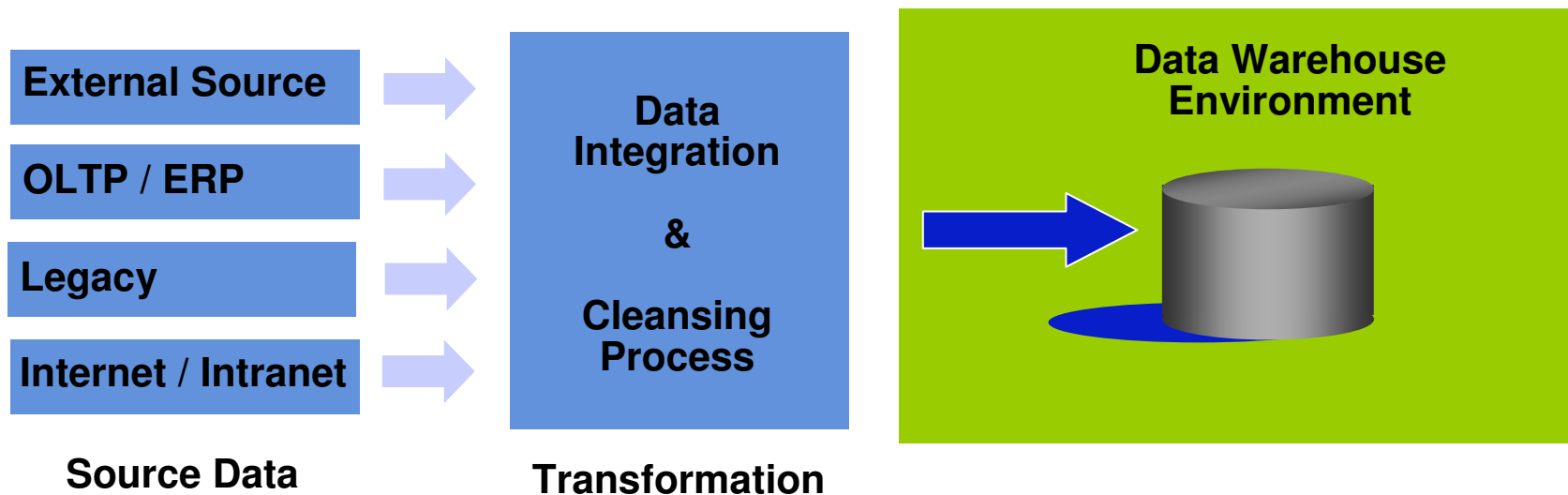
- OLAP takes advantage of DSS Tools

Module 2: > Topic 2: Why OLAP for Data Warehouse

- Integration of data:

  - History is not available in the OLTP system, but is required for trend analysis

  - Data from other operational systems will be needed for analysis

  - Data purchased from commercial data sources

  - Archived data

- Improve performance:

  - Analytical queries degrade OLTP system performance

- Model the data specifically for Decision Support:

  - provide faster performance and a friendlier interface for users

- Take advantage of specialized Decision Support tools:

  - for example, to provide Decision Support access via the Inter/Intranet or to establish a Data Mining capability

Module 2: > Topic 2: Why OLAP for Data Warehouse

OLAP Integrates Data

– History may not being converted in the OLTP system, but is required for trend analysis

– Data from other operational systems will be needed for analysis

– Data purchased from commercial data sources

– Archived data

External Source		Data Integration & Cleansing Process	Data Warehouse Environment
OLTP / ERP			
Legacy			
Internet / Intranet			

Source Data **Transformation**

Module 2: > Topic 2: Why OLAP for Data Warehouse

OLAP Improves Performance

OLAP based data model provides **high** **Response Time**

Performance Terms in Database and Data Warehouse -

Throughput - *It is the amount of work that a computer can do in a given time period*

 – "Cost per million instructions per second (**MIPS**)" provides a basis for comparing the cost of raw computing over time or by manufacturer, **throughput** theoretically tells you how much useful work the MIPS are producing"

Response Time - *The amount of time between a single interactive user request being entered and receiving the application's response is known as response time*

 – "The **speed** with which one or a set of batch programs run with a certain **workload** or how many interactive user requests are being handled with what responsiveness"

Module 2: > Topic 2: Why OLAP for Data Warehouse

OLAP models data specifically for DSS

– Add value to the data through aggregation, transformation and integration with external data

– Apply summary hierarchies not in the OLTP

– Apply business terminology references for users

– Provide the ability to re-state data

– Provide the ability to perform modeling on volume data

	Category	Hub	Month	Sales (INR '000)
Record 1	MS	Chennai	June95	11,342.868
Record 2	HSD	Bangalore	May95	4,020,220
Record 3	ATF	Hydrabad	June95	573,935
Record 4	MS	Bangalore	April95	1,751.316
Record 5	HSD	Chennai	June95	13,551,933
Record 6	ATF	Hydrabad	May95	597,987
Record 7	MS	Chennai	April95	8,963,856
Record 8	FO	Bangalore	May95	9,964,483
Record 9	HSD	Chennai	June95	612,817
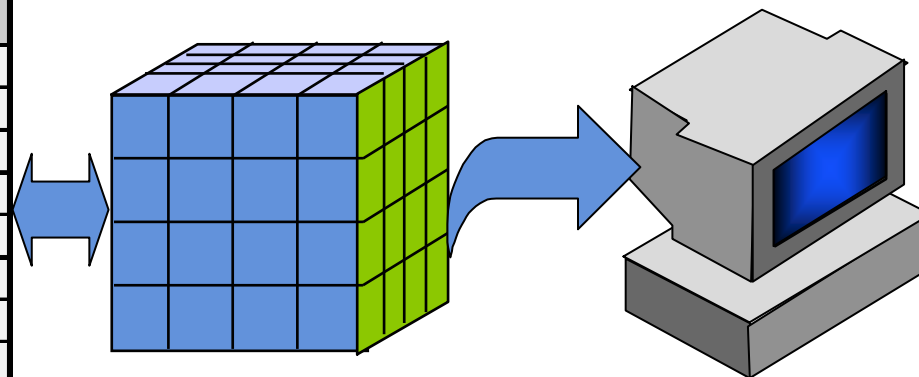
Module 2: > Topic 2: Why OLAP for Data Warehouse

OLAP takes advantage of DSS Tools

– Extract Knowledge using features of DSS tools -
  Drill Through, Slice and Dice, Drill Across

– Data Virtualization and Universes provide unlimited
  integration of data across applications and
  enterprises

– Provide the ability to re-state data

– Provide the ability to perform modeling on volume
  data

– Business friendly representation of data using
  Dimensional Modeling

– Facilitate data based decision making that would
  drive & improve business

– Facilitate Trend Analysis

Module 2: > Topic 2: Why OLAP for DW   Summary

- Having completed this topic, you should be able to:

    – Understand Integration of data

    – Know Response Time and Throughput for Performance

    – OLAP data model specifically for DSS

    – DSS tools utility advantages for OLAP

Module 2: > Topic 2: Why OLAP for DW   Review

Module 2: > Topic 3: Steps to transform OLTP to OLAP

- Data Merging

- Data Scrubbing

- Data Aggregation

- Data Storing

Module 2: > Topic 3: Steps to transform OLTP to OLAP

Data Merging

Data consists of many sources.

- We must be able to merge all the data related to specific items (products, customers, etc) from multiple OLTP systems into a single OLAP system. The merge process must resolve differences in encoding between the different OLTP systems.

- For example, one system may assign an ID to each customer, and the other systems have no customer IDs. The merge process must be able to match common customer data from both systems, perhaps by comparing customer names and addresses. The merge process must also be able to convert data stored using different data types in each OLTP system to a single data type used in the OLAP system. You must also select which columns in the OLTP system are not relevant to an OLAP system, and exclude these columns from the merge process.

Module 2: > Topic 3: Steps to transform OLTP to OLAP

Data Scrubbing

Merging the OLTP data into an OLAP system gives you an opportunity to scrub data.

For example, you may find that various OLTP systems spell items differently, or the merge process may uncover previously unknown spelling errors. You may find other inconsistencies, such as having different addresses for the same store, employee, or customer. These inconsistencies have to be addressed before the data can be loaded into the data warehouse for use by the OLAP system.

Module 2: > Topic 3: Steps to transform OLTP to OLAP

Data Aggregation

- OLTP data records all transaction details. OLAP queries typically need summary data, or data aggregated in some fashion.

  For example, a query to retrieve the monthly sales totals for each product over the last year runs much faster if the database only has summary rows showing the daily or hourly sales for each product, than if the query must scan every transaction detail record for the last year.

- The degree to which you aggregate the data in a OLAP system depends on a number of design factors, such as the speed requirements of your OLAP queries and the level of granularity required for your analysis.

  For example, if you aggregate sales details into daily summaries instead of hourly summaries, your OLAP queries would run faster, but you could only do this if you had no need to analyze sales on an hourly basis.

Module 2: > Topic 3: Steps to transform OLTP to OLAP

Data Storing

- Relational OLTP data is organized in a way that makes some analysis processing difficult and time-consuming. When OLTP data is moved into an OLAP system, it must be transformed into an organization that better supports decision support analysis. The process of building a data warehouse involves reorganizing OLTP data stored in relational tables into OLAP data stored in multidimensional cubes.

- We will talk about Cubes a little later !!

Module 2: > Topic 3: Steps - OLTP to OLAP   Summary

- Having completed this topic, you should be able to:

  – Understand Data Merging

  – Understand Data Scrubbing

  – Understand Data Aggregation

  – Understand Data Storing

Module 2: > Topic 3: Steps - OLTP to OLAP   Review

Module 2: > Topic 4: Differentiate OLTP, OLAP, DW, DSS, BI

- OLTP vs. OLAP

- Data Warehouse vs. OLAP

- DSS vs. DW vs. BI vs. OLAP

Module 2: > Topic 4: Differentiate OLTP, OLAP, DW, DSS, BI

OLTP vs. OLAP

Characteristics	OLTP	OLAP
Data Content	Current Values	Historical, Summarized, Calculated Data
Data Organization	Application by Application	Subject Area wise, Across an Enterprise
Data Updation	Field by Filed	Batch Loads & Updates
Nature of Data	Volatile	Non Volatile, Static between refresh
Data Model	Operational computaion	Business analysis
Usage	Operational	Analytical Analyis and Trends
Performance	Critical	Relaxed
In Other Words	Used for running the business	Used for managing the business

Module 2: > Topic 4: Differentiate OLTP, OLAP, DW, DSS, BI

Data Warehouse vs. OLAP

DATA WAREHOUSE (DW)

- Data from different data sources is stored in a relational database for end use analysis

- Data is organized in summarized, aggregated, subject oriented, non volatile patterns.

- Data is a data warehouse is consolidated, flexible collection of data. Supports analysis of data but does not support online analysis of data.

On-Line Analytical Processing (OLAP)

- A tool to evaluate and analyze the data in the data warehouse using analytical queries.

- A tool which helps organize data in the data warehouse using multidimensional models of data aggregation and summarization.

- Supports the data analyst in real time and enables online analysis of data with speed and flexibility

Module 2: > Topic 4: Differentiate OLTP, OLAP, DW, DSS, BI

DSS vs. DW vs. BI

- **Decision Support Systems (DSS)** are interactive computer-based systems intended to help decision makers utilize data and models to identify and solve problems and make decisions

- **Data Warehouse (DW)** is the foundation of DSS process. It is a Strategy and a Process for Staging Corporate Data

- **Business intelligence (BI)** refers to technologies, applications and practices for the collection, integration, analysis, and presentation of business information and sometimes to the information itself.

- **Online Analytical Processing (OLAP)** can be defined as a technology which allows the users to view the aggregate data across measurements (like Sales Amount, Interest Rate etc.) along with a set of related parameters called dimensions (like Customer, Product, Location, Time, etc.)

Module 2: > Topic 4: OLTP, OLAP, DW, DSS, BI Summary

- Having completed this topic, you should be able to:

    - Difference between OLTP and OLAP

    - Difference between Data Warehouse and OLAP

    - Difference between DSS, DW, BI, and OLAP

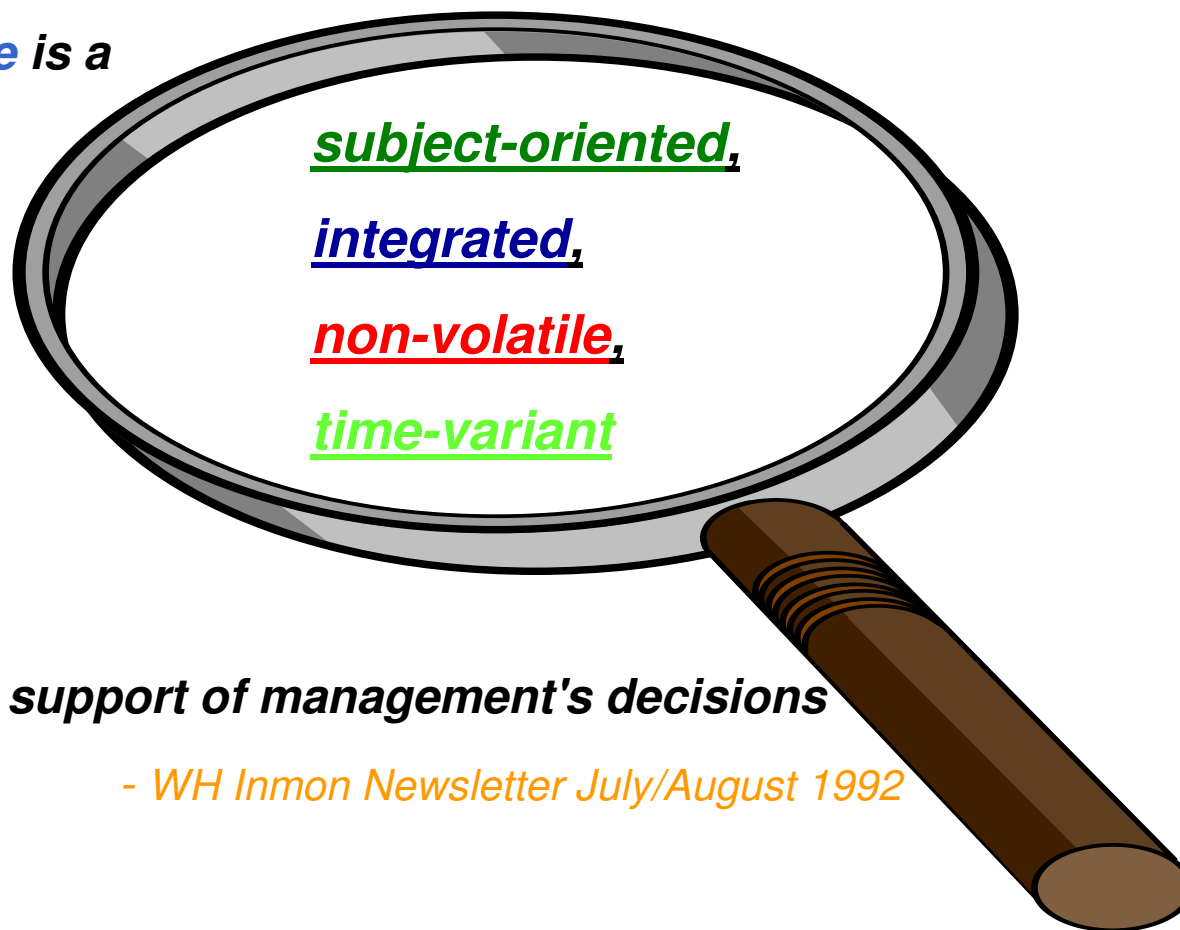# Module 2: > Topic 4: OLTP, OLAP, DW, DSS, BI Review

Module 2: > Topic 5: DW System Qualifiers

- DW Definition Recap

- Subject Oriented data

- Integrated data

- Non-Volatile data

- Time Variant data

Module 2: > Topic 5: DW System Qualifiers

DW Definition Recap

A *Data Warehouse* is a

subject-oriented,

integrated,

non-volatile,

time-variant

collection of data in support of management's decisions
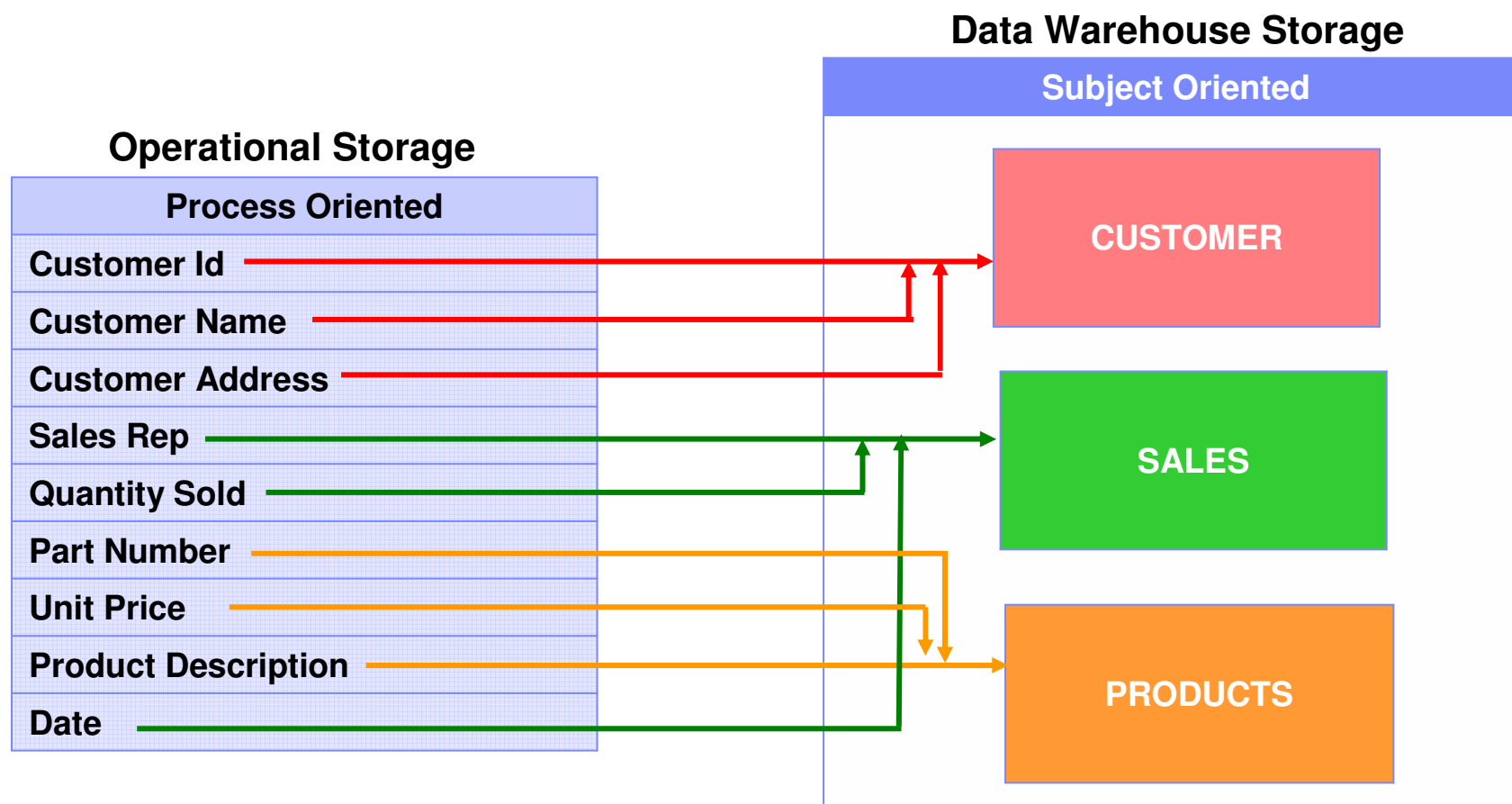
- WH Inmon Newsletter July/August 1992

Module 2: > Topic 5: DW System Qualifiers

- **Subject Oriented** - Data warehouses are designed to help you analyze your data. For example, you might want to learn more about your company's sales data. To do this, you could build a warehouse concentrating on sales. In this warehouse, you could answer questions like "Who was our best customer for this item last year?" This kind of focus on a topic, sales in this case, is what is meant by subject oriented.

- **Integrated** - Integration is closely related to subject orientation. Data warehouses need to have the data from disparate sources put into a consistent format. This means that naming conflicts have to be resolved and problems like data being in different units of measure must be resolved.

- **Nonvolatile** - Nonvolatile means that the data should not change once entered into the warehouse. This is logical because the purpose of a warehouse is to analyze what has occurred.

- **Time Variant -** Most business analysis requires analyzing trends. Because of this, analysts tend to need large amounts of data. This is very much in contrast to OLTP systems, where performance requirements demand that historical data be moved to an archive.
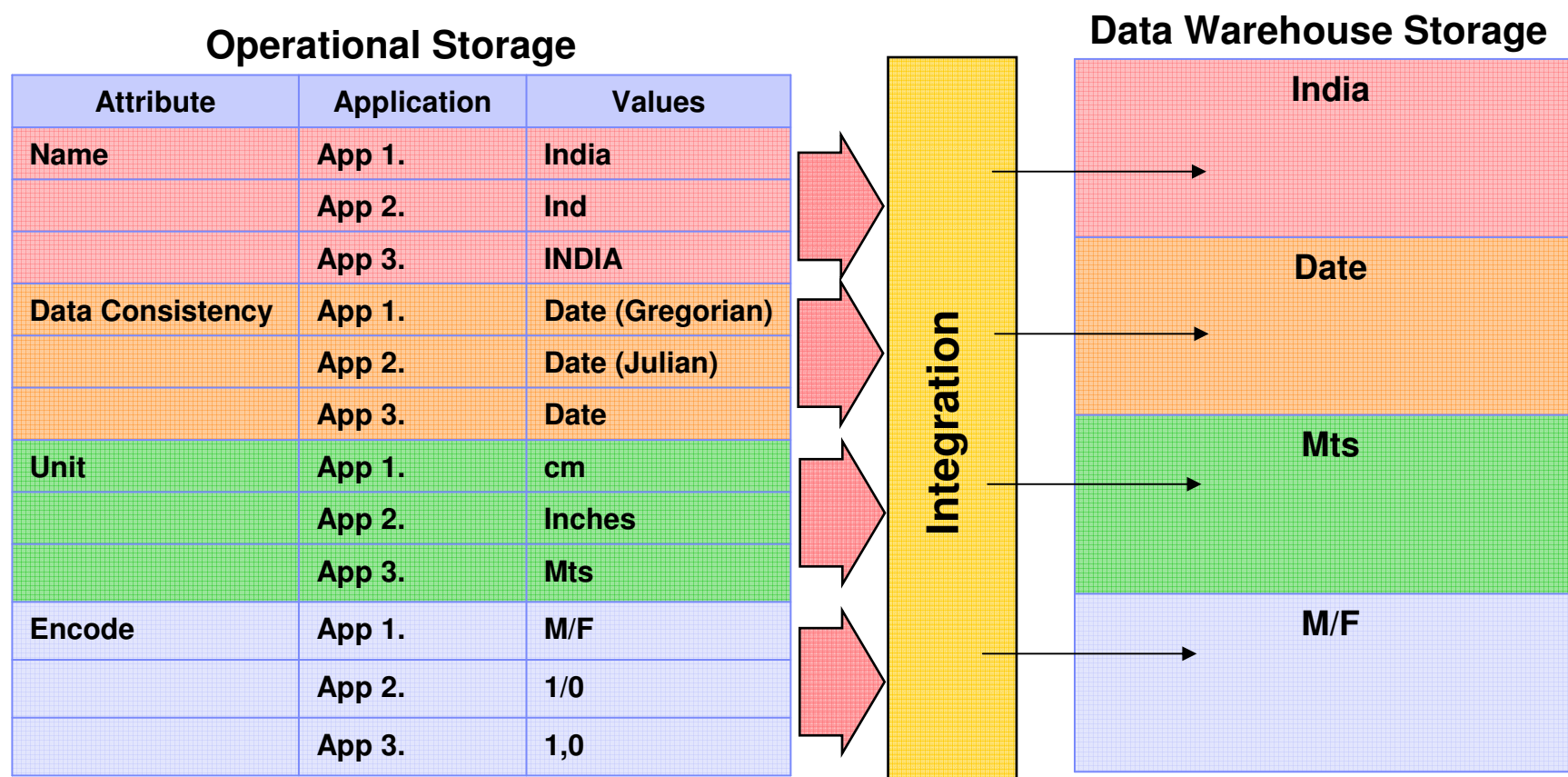
Module 2: > Topic 5: DW System Qualifiers

Subject Oriented Data

Data Warehouse Storage

Operational Storage

Subject Oriented
CUSTOMER
SALES
PRODUCTS

Process Oriented
Customer Id
Customer Name
Customer Address
Sales Rep
Quantity Sold
Part Number
Unit Price
Product Description
Date

Module 2: > Topic 5: DW System Qualifiers

Integrated Data

Operational Storage

Attribute	Application	Values
Name	App 1.	India
	App 2.	Ind
	App 3.	INDIA
Data Consistency	App 1.	Date (Gregorian)
	App 2.	Date (Julian)
	App 3.	Date
Unit	App 1.	cm
	App 2.	Inches
	App 3.	Mts
Encode	App 1.	M/F
	App 2.	1/0
	App 3.	1,0

Integration

Data Warehouse Storage

India

Date

Mts

M/F

Module 2: > Topic 5: DW System Qualifiers

Non-Volatile Data

Volatile Data

Insert
Delete
Update

→ **Operational System** → **Access**

Non-Volatile Data

Mass Load
Mass Load
Mass Load

→ **Data Warehouse System** →

Access
Access
Access
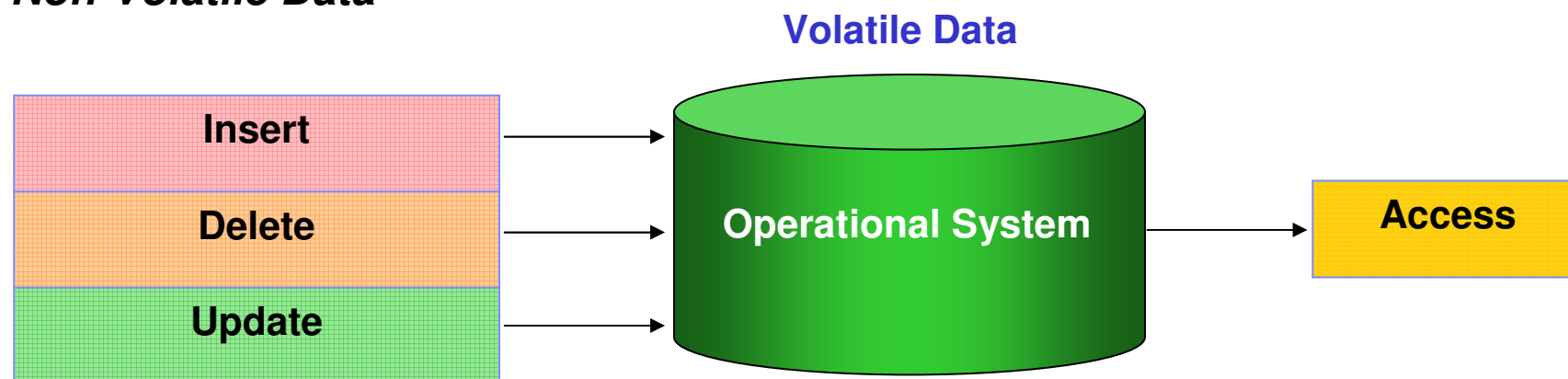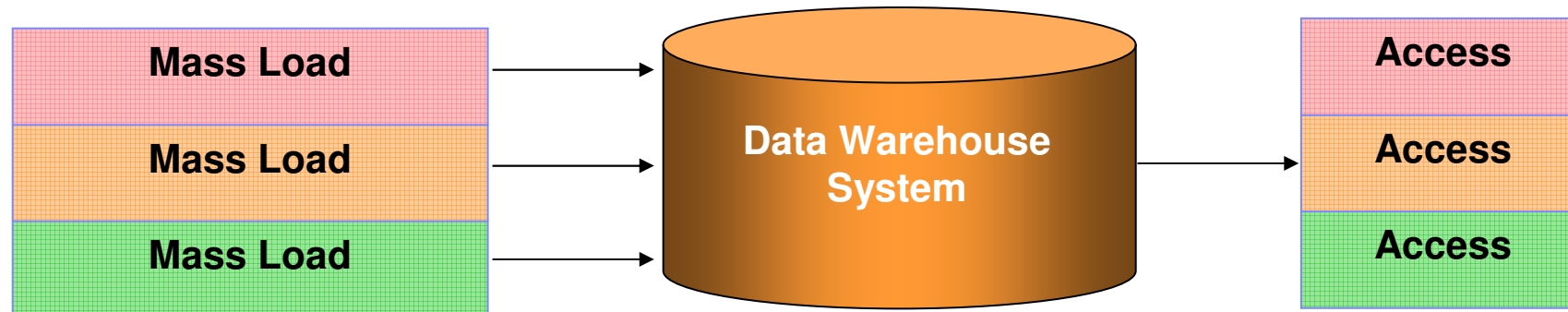
Module 2: > Topic 5: DW System Qualifiers

Time Variant Data

Product Sales Million $

Year 2007-2008

- North
- West
- South
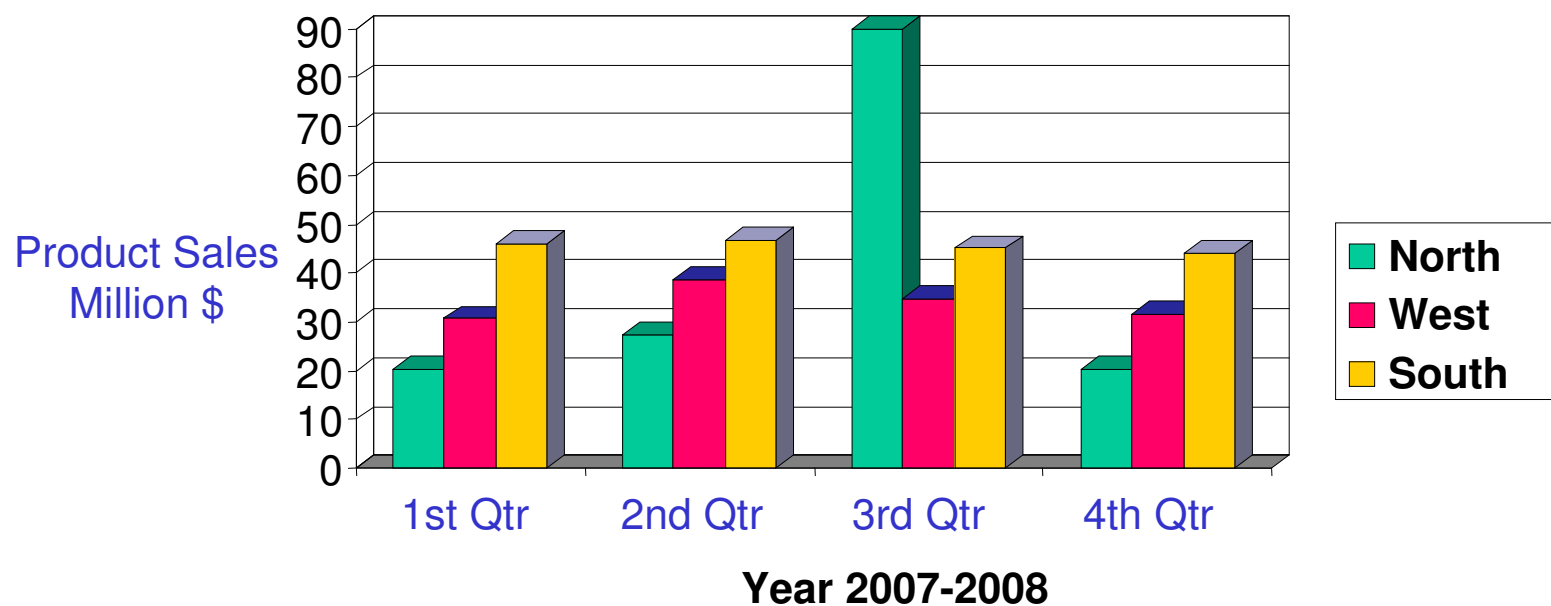
Module 2: > Topic 5: DW System Qualifiers   Summary

- Having completed this topic, you should be able to:

    – Understand what Subject Oriented Data means in DW context

    – Understand what Integrated Data means in DW context

    – Understand what Non-Volatile Data means in DW context

    – Understand what Time Variant Data means in DW context

Module 2: > Topic 5: DW System Qualifiers   Review

References

- TDWI … www.tdwi.org

- DM Review … www.dmreview.com

- Forrester Research … www.forrester.com