U-2 (BI)

The architecture of DW & BI.

DOMS | Page No.
Date / /

- Data Warehousing Concept-
  - process of constructing & using a data warehouse.
  - Integrating data from multiple heterogeneous sources → analytical reporting
- Data warehouse → store historical info from multiple sources to allow you to analyze & report on related data.

Goals →
- reporting / analysis
- maintain → organization's historical info.
- foundation → decision making.

* organizations → info → data warehouse.
  - Increasing customer focus
  - Repositioning products & managing products
  - Analysing operations & looking for sources of profit
  - managing customer relationships, making environmental corrections & managing cost of corporate assets.

* Charateristics of DW →
  - Subject oriented
  - Integrated
  - Non-Volatile
  - Time Varient
  - access & high speed query
  - end user → time sensitive
  - Large amt → historical data
  - Queries → large amt of data.

DOMS | Page No.
Date / /

Architecture of DW → Single tier architecture
→ two-tier "
→ three-tier " (multi-tier archi)

Need for DW →
- Business user
- store historical data
- Make strategic decisions
- For data consistency & quality
- High response time.
- Data integration
- BI
- Performance
- Data quality.

Benefits →
- Improved decision making
- Increased efficiency
- Better data quality
- Historical analysis (large data)
- cost saving
- competitive advantage.
- Improved data governance.

Data Warehouse Types → Enterprise warehouse
→ Data mart
→ Virtual warehouse.

ETL process → Extraction, Transformation & Loading.

DOMS | Page No.
Date / /

#	Business Intelligence (BI)	Data Warehouse (DW)
•	set of strategies & technologie to analyze & visualize data to make decisions	• central location → store consilidated data from multiple data sources
•	data → user's behaviour belongs to BI.	• customer consumption → data warehouse
•	set of technologies & strategies.	• storage
•	present data → reports, charts & graphs.	• presents data in tables
•	top executives & senior manager use BI.	• data engineers, data & business analysts use data warehouses.
•	example – Datapine.	• example – Amazon Redshift
•	create business insights.	• storing data from several sources.
•	provides user-friendly tools for data analysis & visulization.	• Typically accessed via SQL queries or reporting tools.
•		

DOMS | Page No.
Date / /

	ROLAP	MOLAP	HOLAP
implementation	• based on relational DBMS	• based on multidimensional DBMS;	• both relational & multidimensional techniques
Adv →	• handle large data • can leverage • Functionalities • inherit in the relational database	• excellent performance • perform complex calculations.	• HOLAP tools can utilize both pre-calculated cubes & relational sources.
Disadv →	• Performance → slow • Limited by SQL functionalities	• limited data handling • requires additional investment	• supports disadv of MOLAP.

	ROLAP	MOLAP
SCHEMA	• User Star Schema • Additional dimensions can be added dynamically.	• User Data cubes • Add" dimensions require restruction of data base.
Database size	• Medium to Large	• Small to medium.
Architecture	• client/server	• client/server.
Access - Speed →	• support ad-hoc requests • Good with small data sets. • Avg for medium to large data set.	• limited to pre-defined dimensions • Faster for small to medium data sets • Avg for large datasets.
Flexibility & Scalability	• High	• Low

DOMS | Page No.
Date / /

OLTP	OLAP.
• support long trans" → complex queries.	• short trans" → query & updates.
• trad" DBMS to accomodate a large volume of real-time transaction.	• has multidimensional schema
• Tables → OLTP database are normalized.	• tables → are not normalized.
• are designed for use for frontline workers.	• for data scientists, business analyst & knowledge.
• contains current data.	• contains historical data.
• appl" → current data.	• focuses on information out.
• DB size → 100GB to TB	• DBsize - 100MB to GiB
• used for day-to day operations.	• used for decision support systems
• flexibility → low	• high.
• priority → high performance	• ~~maybe low~~. high flexibility.
• performance high → few seconds per query	• performance → loco for complex query.
• Updating → continous & irregular	• updating → periodic & regular.

DOMS | Page No.
Date / /

* Data Models → 1) conceptual models
2) Logical model
3) Physical model.

Types of Business Intelligence Tools -
- spreadsheets.
- dashboards
- Data mining tools
- Adhoc data analysis
- Online analytical processing
- Mobile BI
- Realtime BI
- Operational BI
- Open some BI
- Collaborative BI
- Data visualization software
- Reporting & query software
- Data warehousing tools.

* Data Cube -
allows data to be modelled & viewed in multiple dimensions.

n-D base cube → base cuboid.
0-D cuboid → highest level of summarization → apex cuboid

Adv
- ease in aggregating & summarizing data.
- provide better visulization of data.
- huge amt of data → simplified way
- increases overall efficiency of data warehouse.
- aggregated data in data cube → analysing data fast & thereby reducing access time.

DOMS | Page No.
Date / /

Fact table	Dimension table
• contains measurements along attributes of dimension table.	• contains attributes alongwhich Fact table calculates metric
• contains less attributes & more records.	• contains more attributes & less records.
• Fact table grows vertically	• dimension table grows ~~less~~ horizontally
• Fact table → primary key which concatenation of primary keys → dimension table	• each dimensional table contains → primary key.
• schema contains less no. of fact tables	• schema contains more no. of dimension tables.
• Fact table can have data innumeric as well textual format.	• always contains attributes in textual format.

Star Schema	Snowflake Schema.
• simple & common modelling paradigm where data warehouse comprises of fact table.	• kind of star schema which includes hierarchial form of dimensional tables
• Star schema → not use normalization.	• uses normalization to elimination redunacy of data.
• contains fact & dimension tables.	• sub-dimension tables ↳ including fact & dimensional tables.
• simple to understand & easily.	• Hard to understand & design.
• high level of data redundancy.	• low level of data redundacy
• Cube processing is faster.	• cube processing might be slow because of complex join.
• uses more space.	• It uses less space