UNIT-01

INTRODUCTION TO DATA WAREHOUSE

Data Warehouse Introduction - Data Warehouse components - Operational database vs data Warehouse - Data Warehouse Architecture - Three tier Data Warehouse architecture - autonomous Data Warehouse - autonomous Data Warehouse vs snowflake - Modern Data Warehouse

Data Warehouse Introduction: TOPIC NAME:-

- DW is a type of data mgmt system designed to enable & support Business Intelligence (BI) & analytics
- Contain large amount of historical data
- derived from log files & transaction apps
- consolidate large amount of data from multiple source
- data to improve decision-making
- build data scientists & business analysts

Need for DW:- Complex business world used

→ Enhance turnaround time for analysis & reporting
↓ Allow critical data & retrive data from multiple sources save money, time

→ Improved Business Intelligence - Support organizational data

→ Benefit of historical data - day to day basis of historical data

→ standardization of data & Immense ROI (Return on Investment)
↳ Reduce expense

Benefits of Data Warehouse:-

→ Scalability - easily expand
→ access to Historical Insights — access past data
→ Works on premises & oncloud — depend existing system
→ Better Effeciency — Reliable, faster, improve performance
→ Improved Data Security — prevent misuse data
→ Increase Revenue & Returns — strengthen business
→ Faster and Accurate Data Analytics — generate accurate report

Characteristics of Data Warehouse:-

* distinguish data
→ subject oriented : specific organizational operations
emphasize existing activities
easy & accurate one.

DW is subject - oriented

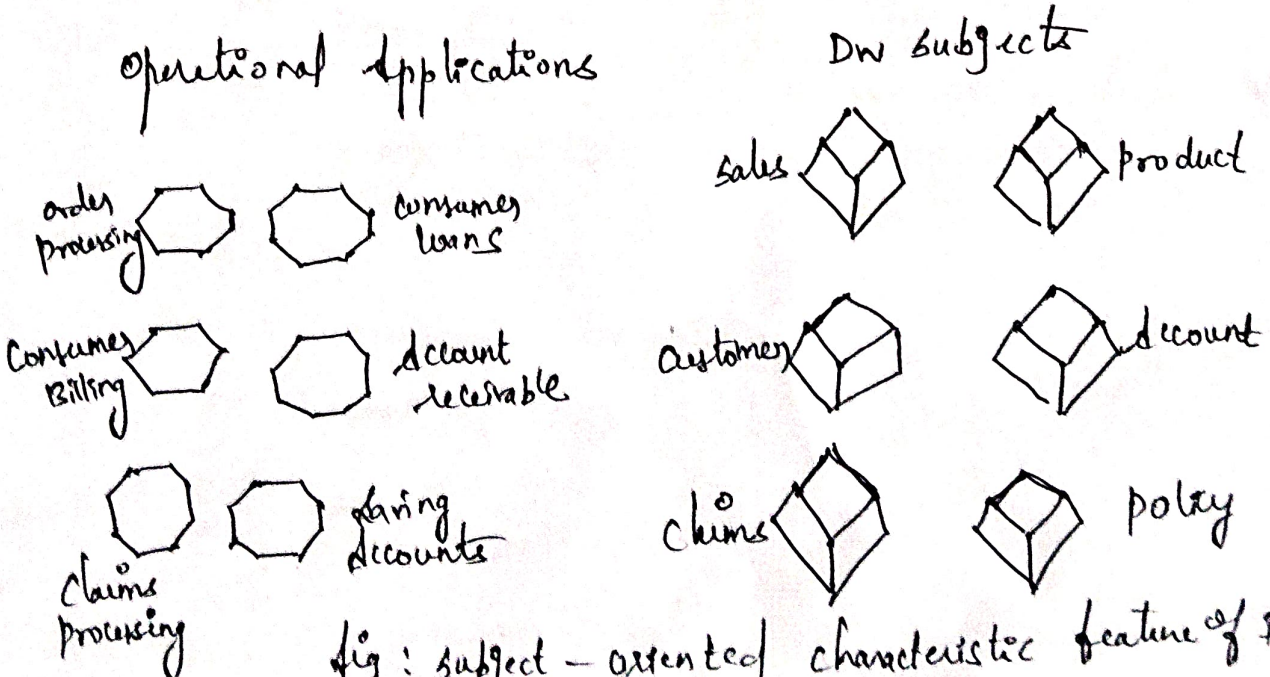


fig : subject - oriented characteristic feature of DW

→ Integrated :-

* measure all similar data from multiple systems
* Repositories database stored in secured manner
* combine relational database & consistent, readable, coded

DH is Integrated

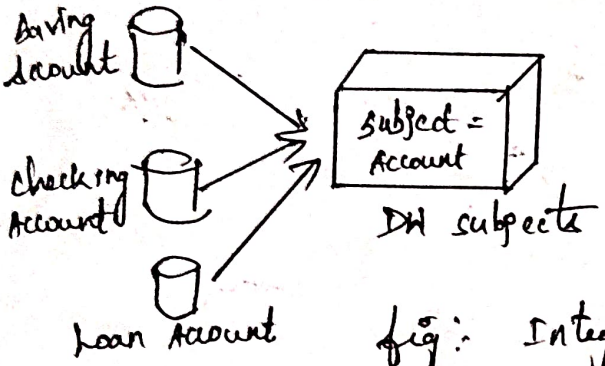


fig :- Integrated characteristic feature of DW

→ Time variant :-

provide online transaction
cover operational systems
predict history

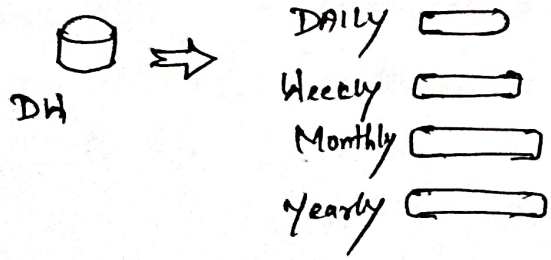


fig :- Time variant characteristic feature of DW

→ Non-volatile :-

dw is permanent / non-volatile
ensure amount of data to be analyzed

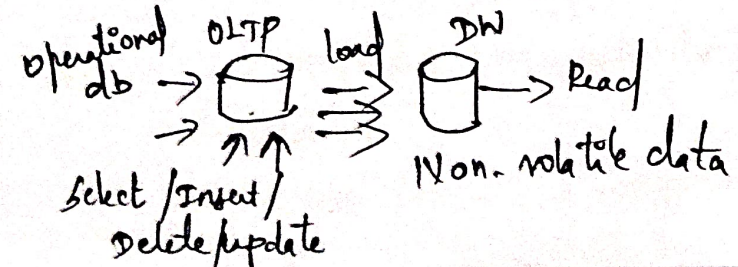


fig :- Non-volatile characteristic feature of DW

Data Warehouse Applications :-

→ e-commerce, telecommunication, transport & marketing

* Investment & Insurance – analyze market trend & stock markets
* Health care – backbone of healthcare system & relevant information
* Retail – Examine trends & incorporate for forecasting
* Social Media Website – Facebook, twitter, linked In etc
* Banking – use special offers, deals etc
* Government – analyze government theft detect
* Airlines – analyse promotions
* public sector – manage data & records

TOPIC NAME:- DW COMPONENTS:-

* provide users with current decisions
* Effective integration, Include relational db, client/server

Overall Architecture :-

- Based on relational db mgmt system
- designed functional & manageable source data's
- Involve conversion, summarization, filtering & condensation } of data
- Capable of holding & managing large volume of data

component :
* DW Db
* Sourcing, Acquisition Cleanup & transformation tools
* Metadata
* Access tools :- 8 tools

ACCESS TOOLS

Types:- 1) Query & Reporting tools
2) Appln
3) OLAP
4) Data Mining
5) Visualization
6) Data Mart
7) DW & mgmt
8) Information delivery system

1) DW Database:- Architecture

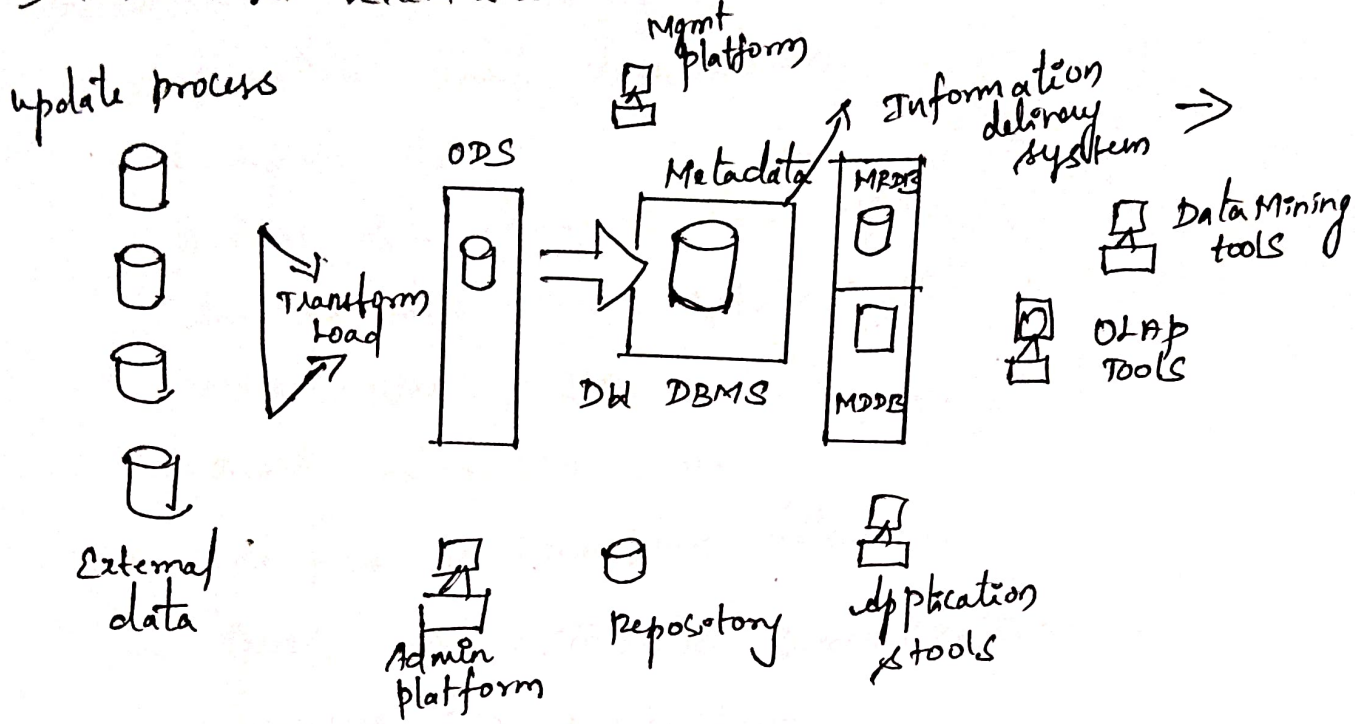


fig: DataWarehouse Architecture

→ DW Database :-

* Implement RDBMS
* Include parallel & symmetric procs
* tightly coupled OLAP [online analytical processing]
* categorize query, reporting analysis, mining tools

→ Sourcing acquisition, cleanup & transformation tools :-

Need to transform data
Maintain Meta data
Removing unwanted data
Summarize missing data
deal with A) DB heterogeneity
B) Data heterogeneity

→ Metadata:-

Information about data source
transform data
perform data cleanup & data mgmt
disclosure data access, history etc
extract usage of data
provide interactive access
Integrate & transform data
Specify conversion & summarization
accessible from any platform
support specified formats

→ ACCESS TOOLS:-

* Interact front end tools
* assess risk
* obtain access to data stored [o/p tools]
* employ multidimensional db
* divide into - data query & reporting tool
application development tool
Execution information sys tool
online analytical processing tool
Data mining tools

A) QUERY & REPORTING TOOLS:-

Reporting tool & Managed query tools
Subject oriented database
designed for easy to use

B) APPLICATIONS:-

find complex set of queries & reporting tools
develop graphical data access
designed client/server environment
Eg: visual Basic, Microsoft, etc

c) OLAP:-

analyze complex views
organize multidimensional model

D) DATA MINING:-

ability to use information effectively
define meaningful patterns
ability to build → discover knowledge ← segmentation, classification, association, preferencing
→ visual data
→ connected data

E) DATA VISUALIZATION:-

visible to domain experts
technique - 3D imaging, Analyze

F) DATA MARTS

Inexpensive
dedicate user group
analyze limited portion of database
Scalability & Integration

* DATA WAREHOUSE ADMINISTRATION & MANAGEMENT:-

Summarize update, quality check, Backup & recovery etc

* INFORMATION DELIVERY SYSTEM:-

→ enable user specified scheduling algorithm
→ Easy access to information
→ Inexpensive access to data
→ engage valid information.

TOPIC NAME:- OPERATIONAL DATA BASE VS DATA WAREHOUSE

* OLTP (or) operational db → manage dynamic data
* DW – organize various users, called as OLAP - online analytical processing

DIFFERENCE :-

Operational db	Data Warehouse
→ Transaction processing	→ analytical processing eg ; OLAP
→ current data	→ historical data
→ update regularly to need	→ new data added / Non-volatile
→ realtime dealing & processes	→ designed by subject area & attribute
→ simple	→ complex
→ validate data table	→ no realtime validation
→ concurrent client support	→ support OLTP
→ process oriented	→ subject oriented
→ small volume of data	→ high volume of data
→ Data In	→ Data out
→ less Number of data accessed	→ large Number of data accessed
→ OLTP [online transaction processing]	→ OLAP [online analytical processing]

TOPIC NAME:- DATA WAREHOUSE ARCHITECTURE:-

→ Method of defining overall architecture of data communication processing & presentation

→ designed for OLTP [online transaction processing]

→ gather data for day to day operations

→ designed OLAP [Online Analytical processing]

→ Applications — summary reporting / profiling / trend analysis

→ reflect summarize & combining data

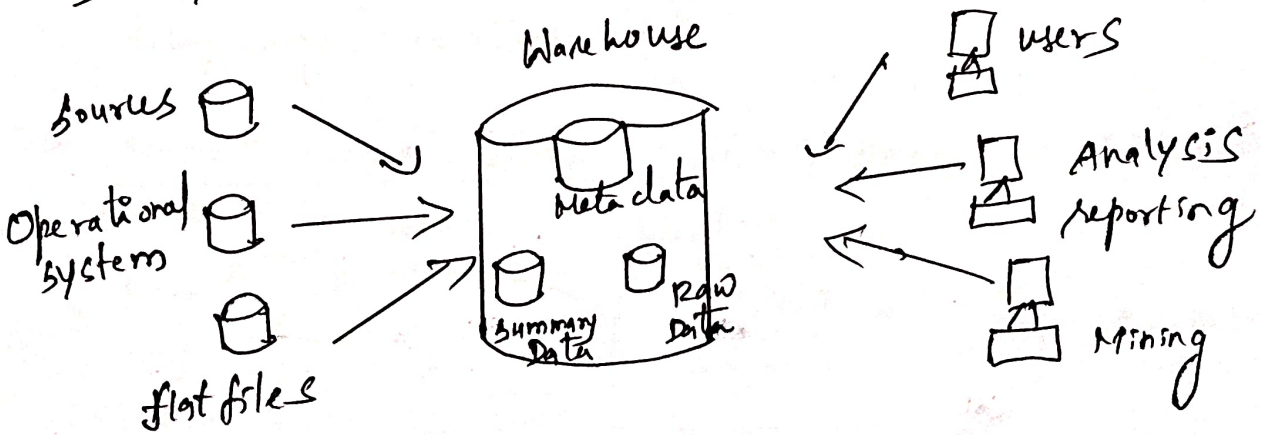
COMMON ARCHITECTURE:-

DW Architecture : Basic

DW Architecture With Staging Area

DW Architecture With Staging Area & Data Marts

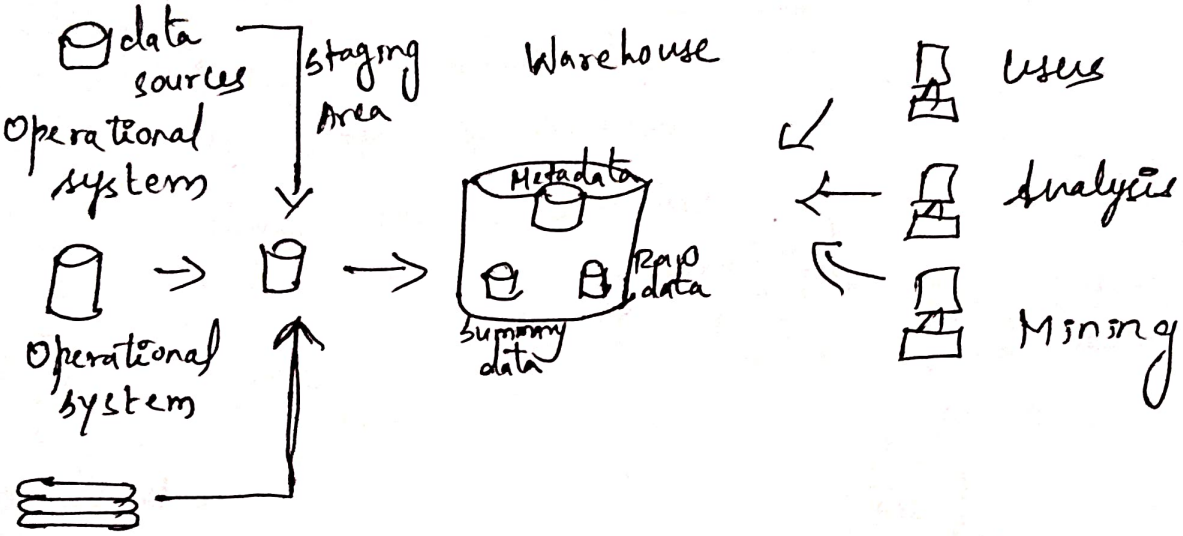
DW Architecture : BASIC



* Operational systems — process day to day transactions
* flat files — transactional data is stored
* Meta data — defines information about other data
* lightly & highly summarized data — speed up query performance

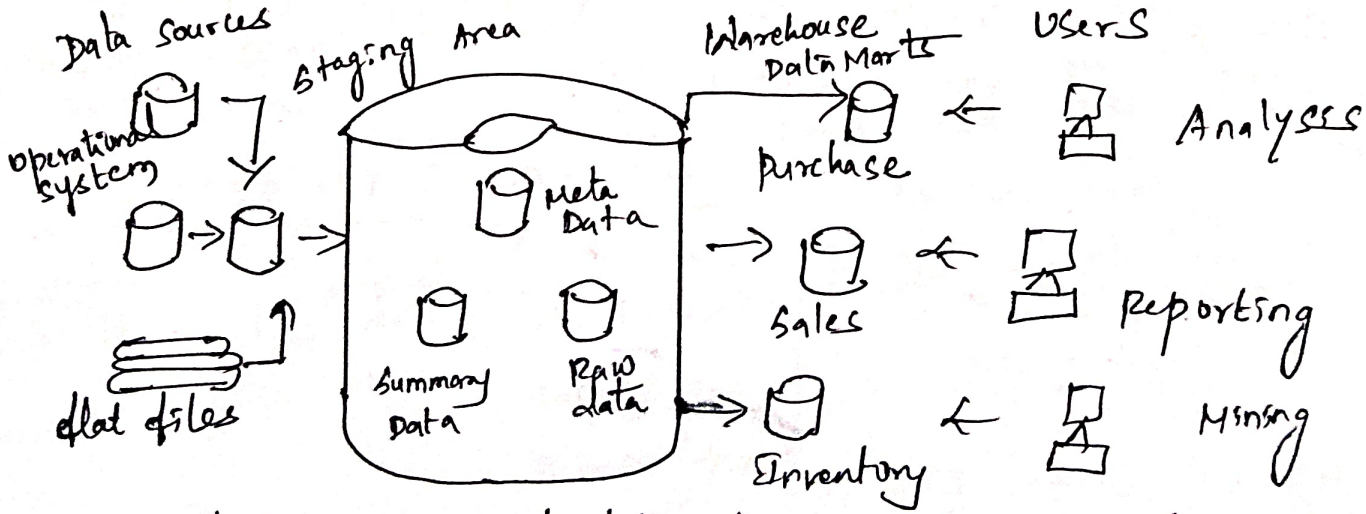
* End user access tools – Reporting & Query tools
Application development tools
Executive Information system tools
online analytical processing tools
Data Mining TOOLS.

DW ARCHITECTURE: WITH STAGING AREA



Simplifies data cleansing & consolidation
temporary location

DW ARCHITECTURE : WITH STAGING AREAS & DATA MARTS



analyze historical data for purchase & sales about customer behaviour Eg: payroll

Properties of Data Warehouse Architecture :-

properties of data Warehouse Architecture
- Separation / Analyze & transaction
- scalability / upgrade & manage
- extensibility / redesign
- Security / Strategic data stored
- Administerability / not complicated

Types of data Warehouse Architectures :-

Single - Tier Architecture
Two - Tier Architecture
Three - Tier Architecture

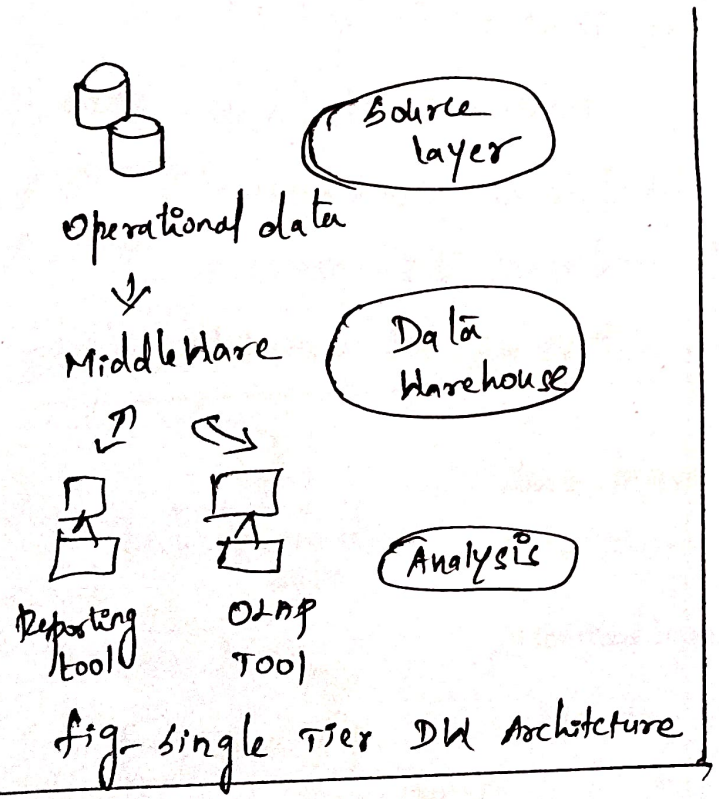


fig- Single Tier DW Architecture

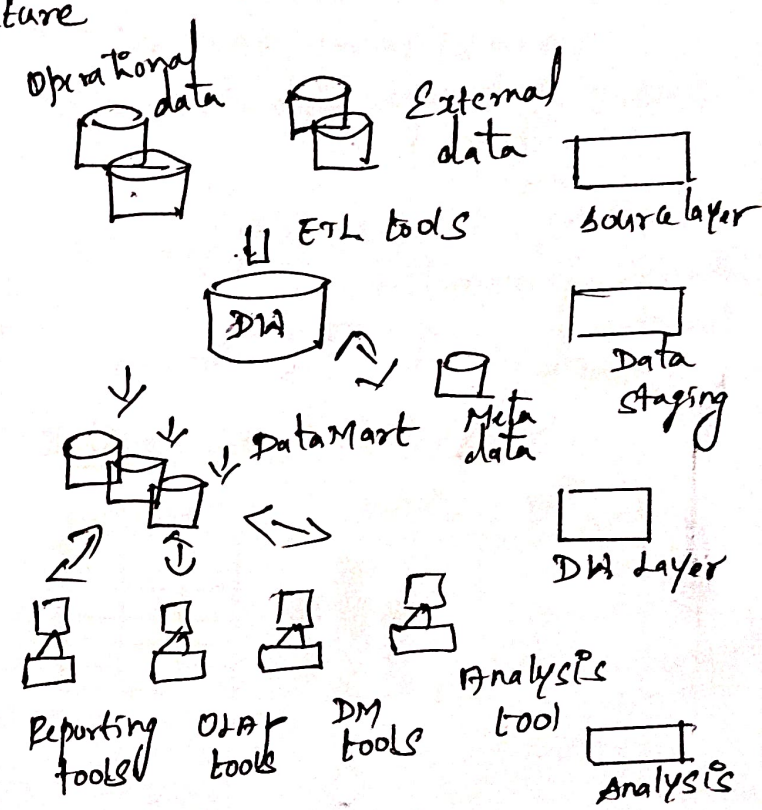
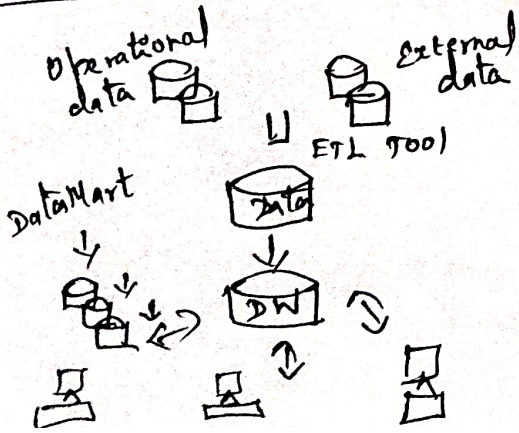


fig: Two-tier DW Architecture



Source layer
Data staging
Reconcile layer
Loading
DW layer
Analysis

→ fig - three tier DW Architecture

Single-tier architecture:- analytic & transaction processing

Two-tier architecture :- Source layer - outside analysis
Data staging - ETL / Extraction, Transformation & Loading (ETL)
DW layer - access users
Analysis - dynamic analysis & OLAP

Three-tier architecture:- reconciled layer

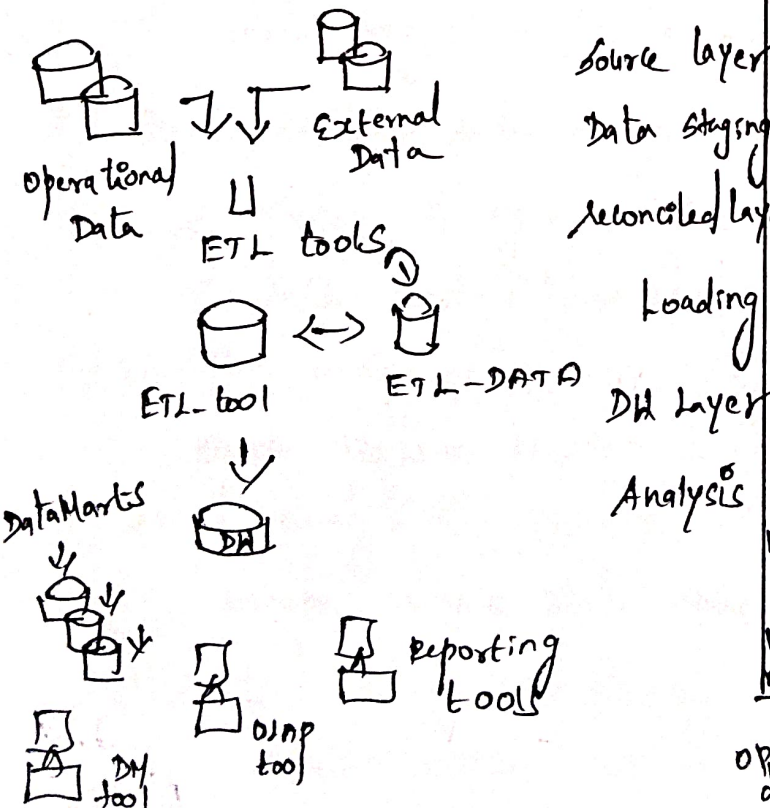
TOPIC NAME:- THREE TIER DW ARCHITECTURE:- 3 layers

A) BOTTOM TIER (DW Server) - Data mart & metadata
Eg: ODBC / OLE DB / JDBC
B) MIDDLE TIER (OLAP Server) - OLAP / MOLAP / ROLAP
C) Top TIER (Front end tools) - Meta data repository - RDBMS

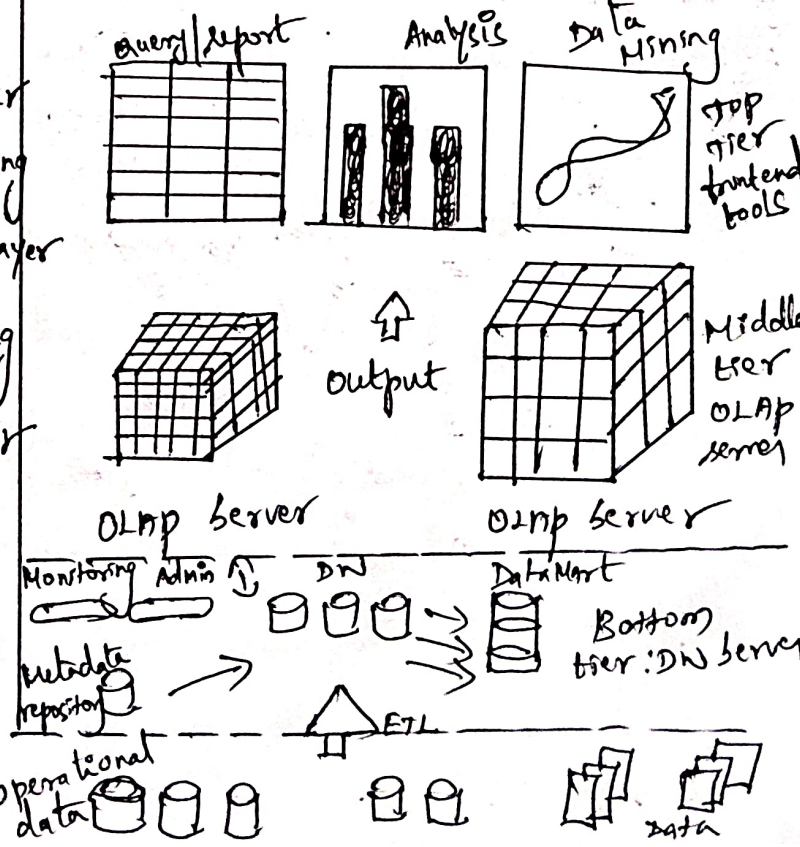
* ODBC - open db connection
* OLE-DB - Open linking & Embedding for DB
* JDBC - Java DB connection

* MOLAP - Multidimension OLAP
* ROLAP - Extended relational DBMS

3-tier Architecture for DW system



3-tier DW Architecture



TOPIC NAME:-

AUTONOMOUS DATA WAREHOUSE

* Oracle Autonomous DW (ADW) is fully managed cloud based DW solutions
* ADW handle large volume of data & decision making
* using data of any size and type

AUTONOMOUS DW FEATURES:-

A) Data lakehouse – modern data platform for business insights - sparks, python

B) Autonomous mgmt - simplified DW mgmt with built in automation features * Auto backups / patching / repairing / scaling / securing / tuning

C) High Availability - Improve the availability of DW & analytic workshops used in storage mgmt, data safe, clusters, data guard etc

D) Security - Reduce risk with DB security
* Identify sensitive data & masks / access data
* data encryption / data masking & redaction / audit & reporting

E) Data Studio - Autonomous DB data studio
* Easily integrate cloud services, db sources, dependencies
* data loading / transformation / analysis / sharing

F) Machine learning & AI
Oracle Machine learning Notebooks / python / SQL

G) Graph analytics - Uncover hidden relationship in data
manage complex data relationships & automated graph modelling

H) Spatial analytics - Making sense of location
spatial appln & location of tracking servers

I) oracle applns - E-business & Net suite
ready to use data model

J) developing & deploying multi cloud applications
Oracle & Azure services & run high performance
simplify monitoring & trouble shoot

Adv of oracle autonomous DW :- location & tracking server

K) applications - Business suite/fusion/Net suite
Built in service data & ETL

L) developing & deploying multicloud applications
Eg: Azure data center

Advantage of oracle autonomous DW:-
auto-pilot warehouse mgmt
productivity / Analytics / Availability
Industry

TOPIC NAME:- AUTONOMOUS DW VS SNOWFLAKE

Autonomous DW
- Require backup
- run on class capability
- 95% Availability
- upgrade data isolation
- not fixed blocks
- built in oracle - cloud integration

SNOWFLAKE
- require cluster set
- runs on Azure - informati
- 99% availability
- customer application two
- higher cost
- advanced analytics

TOPIC NAME:- MODERN DATA WAREHOUSE

⇒ cloud based solution to store information
→ Make smarter decision

COMPONENTS:- level 1: data Acquisition
level 2: Data Engineering
level 3: Data Mgmt and
level 4: Reporting & Business Intelligence
level 5: Data Science: Eg: AI, deep learning, ML

TOPIC NAME: DW/ARCHITECTURE- is modern ~~DW~~ Data used
A) Multiple partial/parallel processing Architecture (MPP)
uses share nothing
B) Multi structured Data:- define Big data
C) Lambda architecture: speed/batch/serving
D) Hybrid architecture:- cloud / utilize Data structure

feature of Modern DW:-
* support master data
* Improve speed
* Support self service data preparation
* Support Agile delivery

Importance of Modern DW
Data lakes / Data Divided Across organization / IOT streaming

Benefits of Modern data Warehouse:-
* lower cost
* less maintanence
* faster speeds
* More flexibility
* Easier to scale

TRADITION DW	MODERN DW
* location — onsite	-> cloud
* purpose - decision making	-> any form
* Data source - transaction	-> any data source Eg: blog/sensor
* Scope - Business Intelligence	-> Extract insights
* Architecture - ETL schema	-> No architecture set
* cost - higher	-> lower

Location — cloud Infrastructure
purpose — desirable Workload / decision making
Scope — OLAP
Data source — virtual environment
Architecture — Extract / load / transform
cost - expensive

— X —