

## To introduce DWH concepts and its importance in Analytics

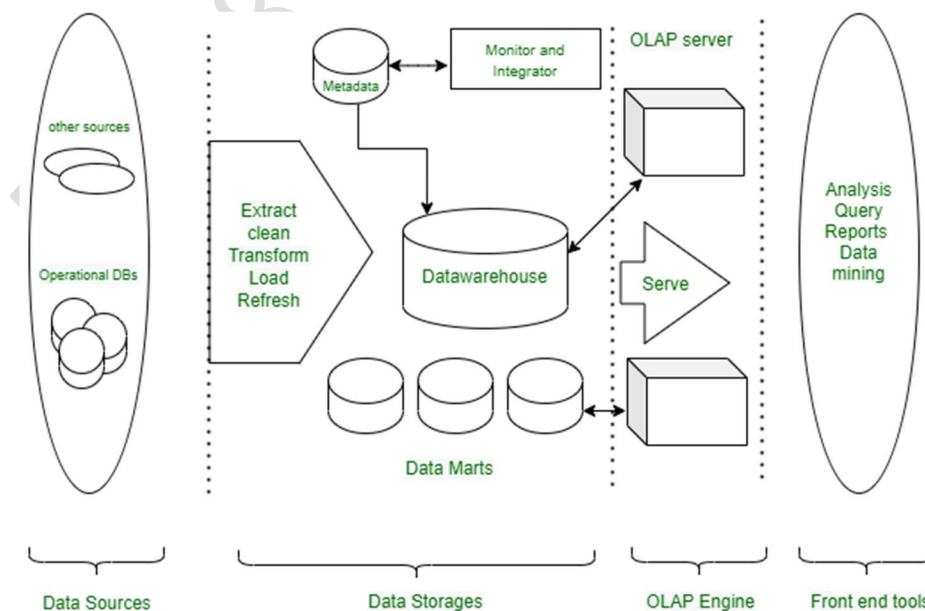
### An introduction to Data Warehousing

Data Warehousing is a system for storing and analyzing large volumes of data from various sources to support business intelligence and decision-making processes. It involves collecting, cleaning, transforming, and storing data in a centralized repository. Key components include ETL (Extract, Transform, Load) processes, which prepare data for storage; a central data warehouse that holds the structured data; and tools for querying and analyzing the data. Data warehousing enables organizations to gain insights, identify trends, and make data-driven decisions efficiently.

### Purpose of Data Warehouse

The purpose of a Data Warehouse is to provide a centralized, reliable source of data for analysis and decision-making. It consolidates data from various operational systems, transforming it into a format suitable for querying and reporting. This enables organizations to gain a comprehensive view of their data, support historical analysis, identify trends, and improve strategic planning. Data Warehouses are optimized for complex queries, ensuring quick access to large datasets and facilitating business intelligence, reporting, and analytics.

### Data Warehouse Architecture



## **Operational Data Store**

An Operational Data Store (ODS) is a type of database that serves as an intermediary between operational systems and a data warehouse. It consolidates and integrates real-time or near-real-time data from multiple sources, providing a centralized location for current, up-to-date information. Unlike a data warehouse, which is optimized for historical analysis, an ODS focuses on current operational data, making it suitable for quick, short-term queries that support day-to-day activities. It is often used in situations that require immediate, accurate information, such as customer service or transactional processing.

## **OLTP Vs Warehouse Applications**

OLTP (Online Transaction Processing) systems are optimized for real-time, day-to-day transactional tasks like order processing, involving frequent reads and writes of detailed, current data. In contrast, Data Warehouses are designed for data analysis and reporting, storing large volumes of historical, summarized data for complex, read-only queries. OLTP systems prioritize speed and accuracy in transaction handling, while Data Warehouses focus on efficient data retrieval for strategic decision-making.

## **Data Marts**

A Data Mart is a subset of a data warehouse focused on a specific business area, department, or subject. It contains summarized or detailed data tailored to the needs of a particular group, such as sales, finance, or marketing, enabling faster access to relevant information and insights. Data marts support decision-making by providing specialized data, often structured for quick query performance. They can be categorized into dependent data marts, which derive data from a centralized data warehouse, and independent data marts, which are standalone and gather data directly from sources.

## **Data marts Vs Data Warehouses**

Data Marts and Data Warehouses both store organizational data, but they serve different purposes:

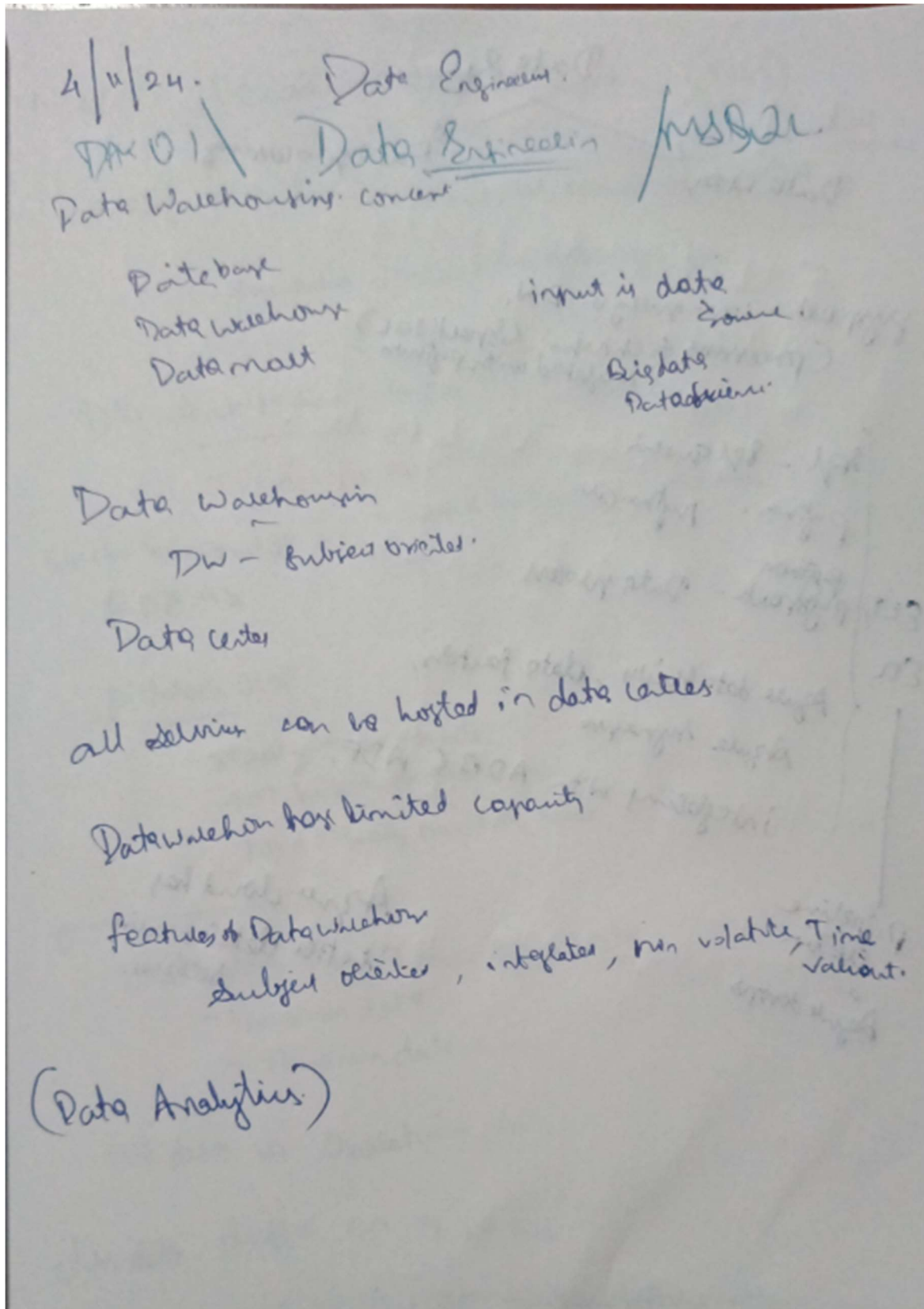
- **Data Warehouse:** A large, centralized repository that consolidates data from multiple sources across the organization. It stores vast amounts of historical data, serving as a single source of truth and supporting enterprise-wide analysis and reporting.

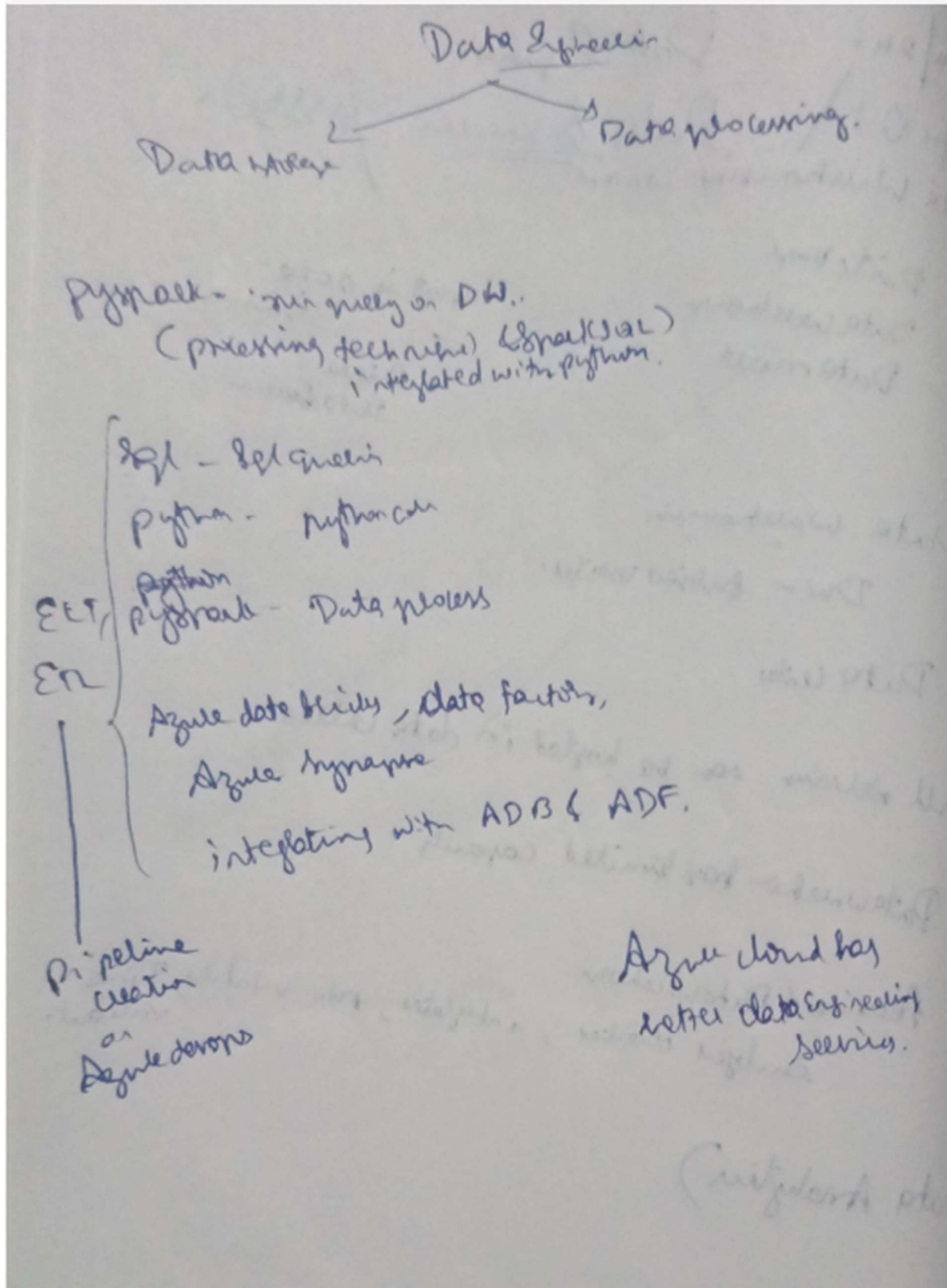
- **Data Mart:** A smaller, specialized subset of a data warehouse, tailored for specific business functions or departments (e.g., sales, marketing). It focuses on specific data needs, allowing faster and more targeted data access.

### Data Warehouse Life cycle

The Data Warehouse Life Cycle includes these key stages:

1. **Business Requirements Analysis:** Define objectives and scope based on business needs.
2. **Data Modeling:** Design data structures (conceptual, logical, and physical models).
3. **ETL (Extract, Transform, Load):** Gather, clean, and load data into the warehouse.
4. **Design and Development:** Build warehouse architecture and storage.
5. **Testing:** Validate data accuracy, performance, and reliability.
6. **Deployment:** Move to production and train users.
7. **Maintenance:** Update data, enhance performance, and address issues.
8. **Optimization:** Monitor and improve query and storage efficiency.





Need for Decision Support System (DSS).

every task has a structured component as well as <sup>unstructured component</sup>  
 unstructured - readable, processed in

DSS Architectural Styles

OLTP & OLAP

used by traditional

RDBMS

Used by Data Warehouse

pitfalls of OLTP

require instant update.  
 not suitable for data analysis.  
 need to query multiple table.

→

data store has 2 main types of data

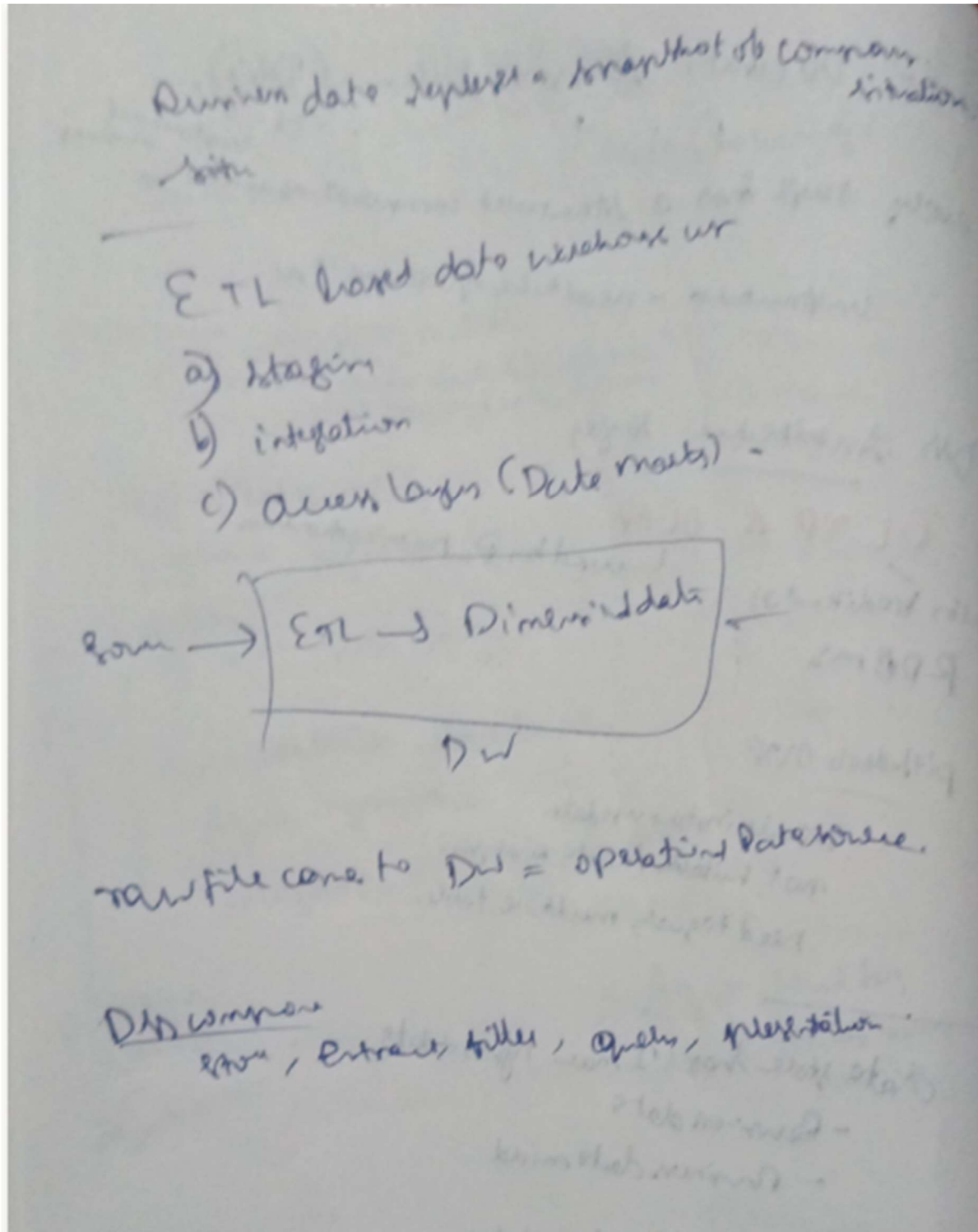
- Current data
- Historical data mod.

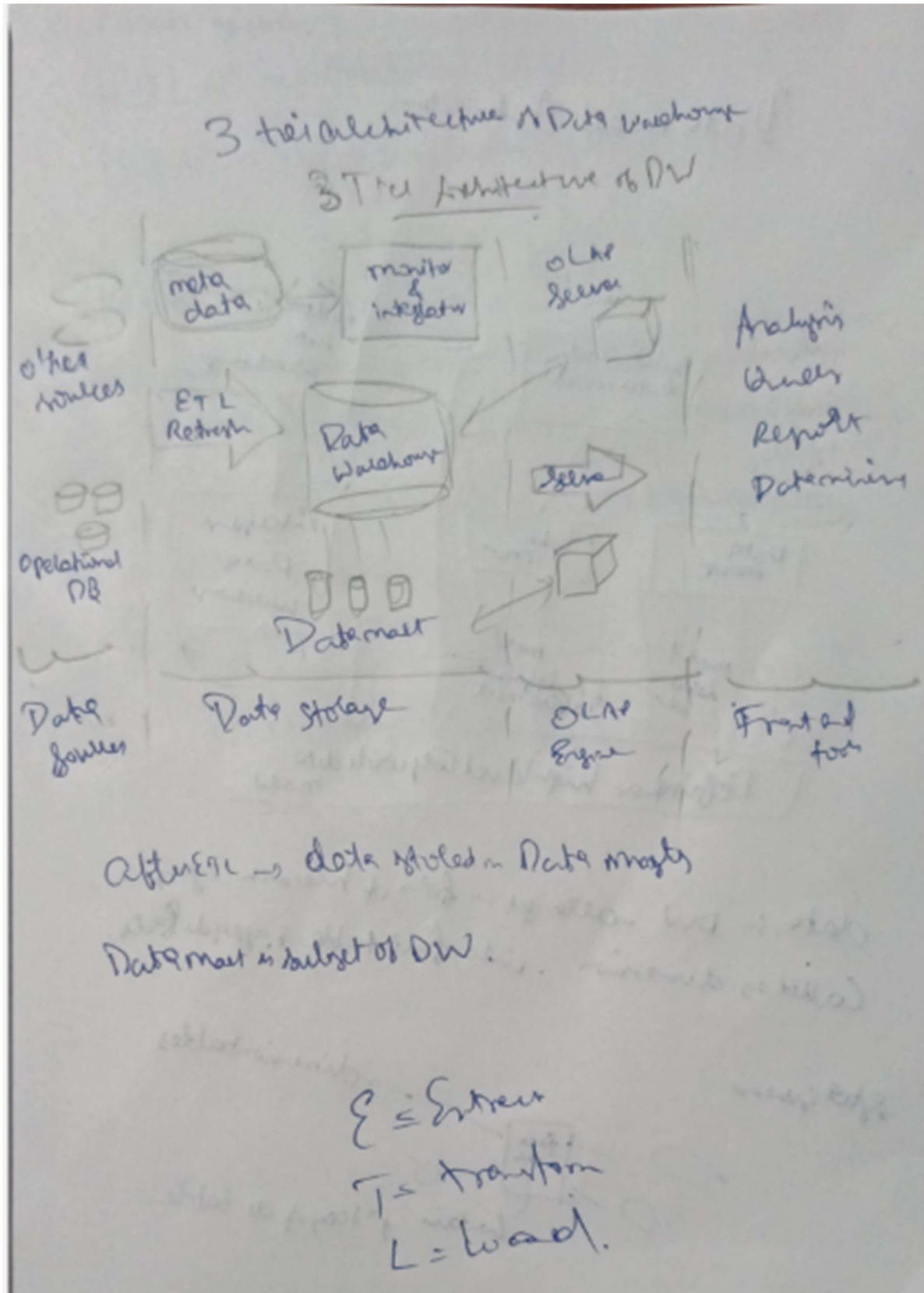
DSS data vs Operational data.

DSS data differ in 3 areas

time span, granularity, dimensionality.

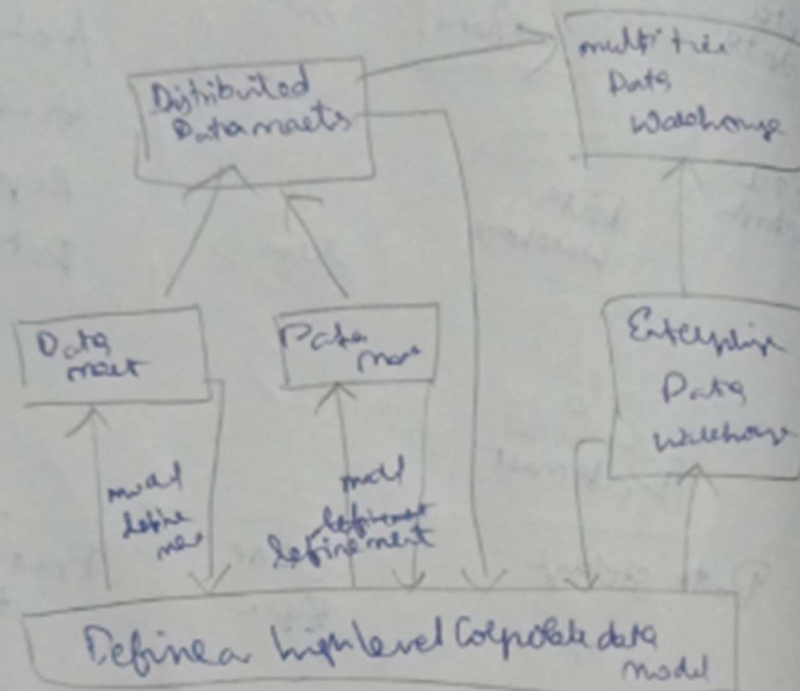




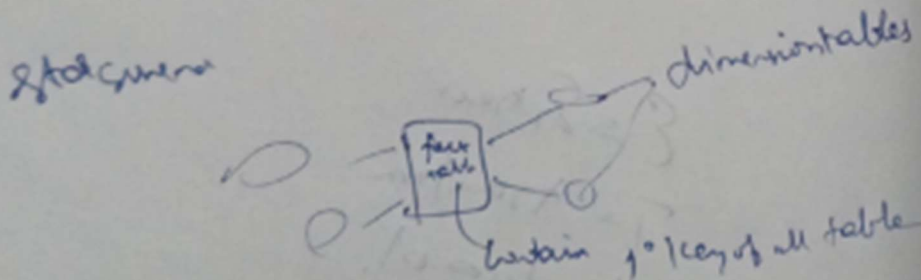




## Data mart Architecture



data in DW always in form of hierarchical glass  
Called as dimension . into fact table & aggregate table



LP server quality.

ROLAP = Relational OLAP.

MOLAP = Multidimensional OLAP.  
(olap dice intro)

3 Kinds of DW applications

- info processing - support querying, basic statistic analysis, known using cost table, table
- Analytical processing (support OLAP operations like slice dice drill, pivot)
- Data Mining - Knowledge discovery for hidden pattern

Big Data

Date of large qty

DE138