

**Class Notes:**

Subject: ....E.D.A.....

Faculty: ..D.-Jyoti:....

Topic: Introduction to DB

Unit No. : 11

Lecture No.: 1

Link to Session

Planner (SP) S.No. ....of SP

Date Conducted:

Page No. 34

## Introduction to Database.

Data is recorded in a table which is collection of rows, columns and it is indexed so that to find relevant information becomes an easier task. A new information is added, data gets updated, expanded and deleted.

different types of databases.

- Flat
- Hierarchical
- Network.
- Relational.

Flat Database characteristics

- simple
- long and dominant
- useful for very scale and simple application

Relational Database characteristics

- organizes data such that it appears to the user to be stored in a series of interrelated tables
- used for high - performance applications
- Efficient

## Class Notes:

Subject: ...E.D.A.....

Faculty: ...D..J.yoti.....

Topic: ...Datawarehouse

Unit No.: II

Lecture No.: 2

Link to Session

Planner (SP) S.No. ....of SP

Date Conducted:

Page No. 27

### Datawarehouse:

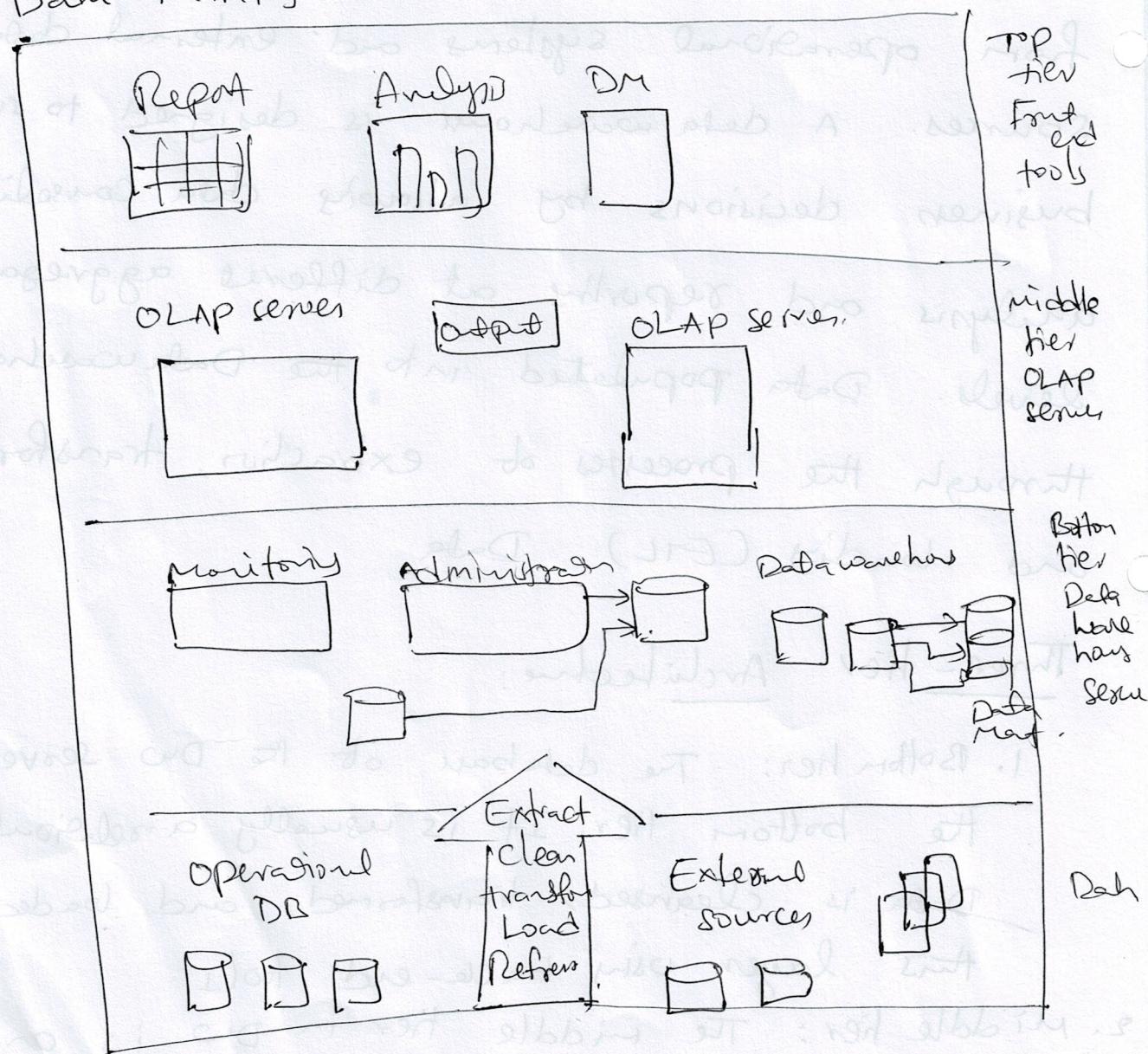
A datawarehouse (DW) is a collection of corporate information and data derived from operational systems and external data sources. A datawarehouse is designed to support business decisions by allowing data consolidation, analysis and reporting at different aggregate levels. Data populated into the Datawarehouse through the processes of extraction, transformation and loading (ETL) Data

### Three-tier Architecture:

1. Bottom tier: The database of the DW serves as the bottom tier. It is usually a relational DB. Data is cleansed, transformed, and loaded. into this layer using back-end tools.
2. Middle tier: The middle tier in DW is an OLA which is implemented using either ROLAP

top-tier:

The top tier is a front-end client layer. Top tier is the tools and API that you connect and get data out of datawarehouse. It could be query tools, reporting tools, merged query tools, Analysis tools and Data mining tools.



**Class Notes:**

Subject: ...EPA.....

Faculty: ...D. Jayalalitha...

Topic: Three tier architecture &amp; Datacube.

Unit No.: 11Lecture No.: 3 & 4

Link to Session

Planner (SP) S.No. ....of SP

Date Conducted:

Page No. 29Data cube

A Datacube refer to a multi dimensional data structure. That is, data within the data cube is explained by specific dimension values.

Example:

Consider a retail company based on the US that offers 4 types of clothing, t-shirts, shirts, jeans, and jackets.

product	→	T-shirts	Shirts	Jeans	Jackets
T-shirts					
\$12,000		\$10,500		\$16,300	\$15,200

## • 3-D Data cube

Location		Chicago			
		NY	7000	818	812
Quarter	Q1	605	828	74	400
	Q2	680	952	31	512
	Q3	812	1023	30	511
	Q4	921	1112	225	511

**Class Notes:**

Subject: ...EDA.....

Faculty: ...D. Jyoti.....

Topic: Storing data in cloud & Retrieval.Unit No. : IILecture No.: 6 & 5.8.7.

Link to Session

Planner (SP) S.No. ....of SP

Date Conducted:

Page No. 30Storing data in cloud.

On-premise storage networks, cloud storage uses servers to save data. However the data is sent to servers at an off-site location. Most of the servers you use are virtual machines. Hosted on a physical server. As your storage needs increase, the provider creates new virtual servers to meet demand.

public storage cloud: In this model, you connect over the internet to a storage cloud that's maintained by a cloud provider and used by other companies. Providers typically make services accessible from just about any device, including smartphones and desktops and let you scale up and down as needed.

private cloud storage: private cloud storage setups typically replicate the cloud model, but they reside within your network, leveraging cloud power to create instances of

to take full control of an on-premises private cloud or engage a cloud storage provider to build a dedicated private cloud that you can access with a private connection.

### Hybrid cloud storage:

This model combines elements of private and public clouds, giving organizations a choice of which data to store in which cloud. For instance, highly regulated data subject to strict archiving and replication requirements is usually more suited to a private cloud environment, whereas less sensitive data can be stored in the public cloud.

## Class Notes:

Subject: ....EDA.....

Faculty: ....D. Jyoti.....

Topic: Retrieving Data from cloud

Unit No.: 1

Lecture No.: 8 & 9

Link to Session

Planner (SP) S.No. ....of SP

Date Conducted:

Page No. 37

### Types of cloud storage.

#### File storage:

File storage saves data in the hierarchical file & folder structure with which most of us are familiar. The data retains its form, whether resides in the storage system or in the client where it originates, and the hierarchy makes it easier and more intuitive to find and retrieve file when needed. File storage is commonly used for development platforms, home directories, and repositories for video, audio and other files.

#### Block storage:

Traditionally employed in SANs, block storage is also common in cloud storage environments. In this storage model, data is organized into large volumes called "blocks". Each block represents a separate hard drive. Cloud storage providers use blocks to split large amounts of data among multiple storage nodes. Block storage resources provide better performance over a network to

## object storage:

object storage differs from files and block storage in that it manages data as objects. Each object includes the data in a file, its associated metadata, and an identifier, objects are stored data in the format it arrives in and makes it possible to customize metadata in ways that make the data easier to access and analyze. Instead of files and folder hierarchies, objects are kept in repositories that deliver virtually unlimited scalability.

## Class Notes:

Subject: ....EDA.....

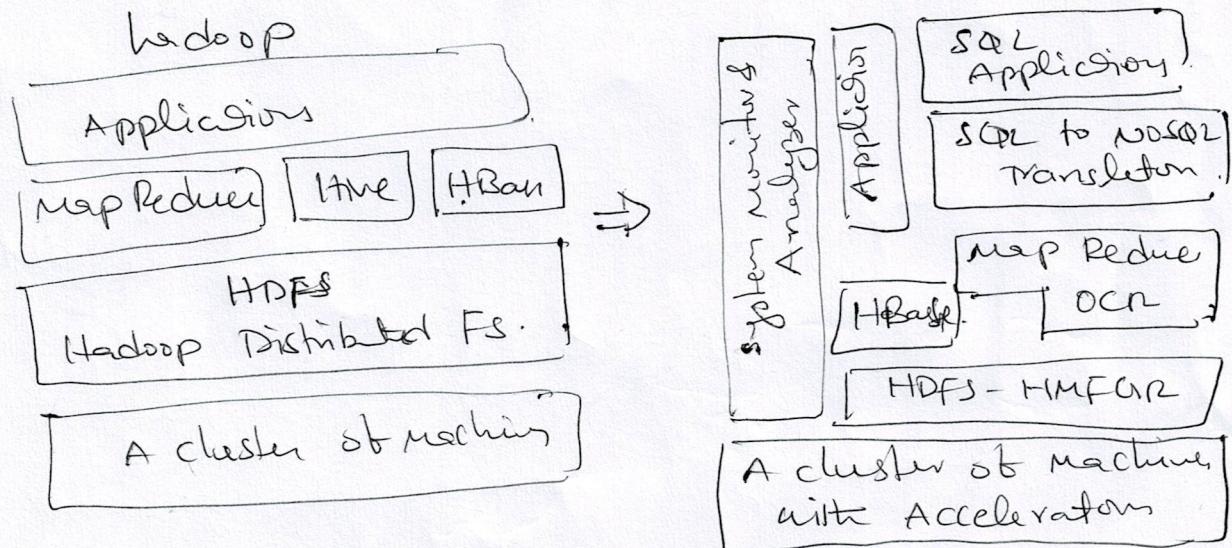
Faculty: ....D. Jyoti.....

Topic: Storing bigdata into HDFS

Unit No.: II  
Lecture No.: 10 & 11  
Link to Session  
Planner (SP) S.No. ....of SP  
Date Conducted:  
Page No. 33

Storing bigdata into HDFS.

HDFS (Hadoop Distributed Filesystem), distributed file system designed to store and process big data. It is a core component of Apache Hadoop ecosystem and allows for storing and processing large datasets across multiple commodity servers. It provides high throughput access to data and is optimized for large file transfers making it a suitable solution for bigdata.



## Class Notes:

Subject: .....EDA.....

Faculty: ....D. Jyoti.....

Topic: .....Storing BigData into NoSQL

Unit No. : 11

Lecture No.: 12

Link to Session

Planner (SP) S.No. ....of SP

Date Conducted:

Page No. 35

### Hadoop NoSQL Databases

The need to analyze data in large volumes, from different sources and formats, has given rise to NoSQL (not only SQL) technology.

They are not relational and not based on schemas.

NoSQL databases can grow and focus more on performance, allowing replication of data across multiple network nodes, reading, writing, and processing data at incredible speed, using distributed parallel processing paradigms.

### NoSQL Types:

1. Column Database (column-oriented)

2. Key-value Database (key/value oriented)

3. Document Database (document-oriented)

### Column Database:

A NoSQL database that stores data in tables by columns instead of rows.

It converts columns into database

26

### Apache HBase

Is a NoSQL-oriented columns. Developed to run on top of Hadoop with HDFS.

### 2- Key value Database:

Key value oriented NoSQL storage data in collections of key / value pairs. For example, a student id number may be the key, and the student's name may be the value.

### Apache Cassandra

Cassandra is a powerful NoSQL based key / value model

Cassandra runs without Hadoop, but it becomes powerful when connected to Hadoop and HDFS.

### Document Database (document-oriented)

Document-oriented NoSQL are like key / value documents. NoSQL organizes documents into collections analogous to relational tables.

Mongo DB: It is document oriented NoSQL,