Business

Know

Decision making

OLTP ⇒ Real time data transaction processing

Run --Application

Operational source

College ; admin lib canteen hostel lab

Student :

Organisation : no due :

Java --oracle

Vb.net - sqlserver

Excel , txt , csv (flat file) .xml /json

SAP , pega, salesforce

Mainframes /c,c++ -legacy

OLTP : Dw

App /processing oriented subject oriented

Isolated integrated

Current historical

Volatile non-volatile

Puzzle : big ---31 pieces

Languages ---> diff lang ===> translate -english

OLTP ---E T L --DW =OLAP ==analysis =reporting

(oracle/teradata/netezza/DB2/Sqlserver)

DW

Business Unit : Data

Business domain : retail ⇒

Sales ⇒ 100 App ⇒

==================================

Cust\_name, customer name , cname , customer\_name

dd/mm/yy, mm/dd/yy , mm/dd/yyyy , dd-mon-yy

M,f , male ,female , 0,1

Inches, feet,cm

111 aaa 8000

Flag version

1000 111 aaa 6000 0 1001

1001 111 aaa 8000 0 1002

1002 111 aaa 9000 1

SCD → type 1 - current

Type 2 → history → version ,flag and date

Type 3 → previous and current data

CDC -- change Data capture ⇒

Sk empno ename sal begindate enddate

1000 111 aaa 6000 19/7/21 20/7/21

1001 111 aaa 8000 20/7/21 21/7/21

1002 111 aaa 9000 21/7/21 -

Type 3;

Sk empno ename sal prev\_sal

1 111 aaa 9000 8000

======================================

Dimension fact

Cleansing -scrubbing transform

OLTP:

LDM ---PDM --->

Dw; dim tables /fact tables

PDM → LDM ===> Dimensional modelling

OLAP: cubes :

hierarchy /level /attributes

========================================

Student performance :

Fees

Attendance

Univ

College

Degree

Subjects

Faculty

**Marks**

**Cgpa**

Semester

Measure /metric /fact / KPI =key performance Indicator

Point of view --dimension --perspective / context of the measure

**Sales analysis:**

Store

Product

Customer

Employee

Color

Size

**Price**

**Qty sold**

**Revenue**

**Profit**

Ordermethod

Star schema / snowflake

=====================================

Frequently book => semester , book , fetch

=======================================

OLTP --ETL --DW -- **OLAP → analysis -Reporting --info**

Knowledge discovery process ---decision making

Business Intelligence : BIDW -->

Raw data → ename ,sal ==product\_name

=======================================

Dashboard =car --multisource interactive reports

Scorecard --align strategy

Ad hoc -

Time series

Story

Olap operations : Data visualizations / DV =BI tools

Qlik sense /powerbi /tableau / Mosaic decision

Year

Qtr

Mon

Day

Drill down /drill up

Slice /dice

Drill across /drill thru

=======================================

Modern Data warehouse :

Data COE :

========================================

On-premises = > cloud

Cloud computing :

Anywhere

Scalability -- vertical scalability / horizontal scalability

Concurrency :

IAAS PAAS SAAS

==========================================

Low latency → fast queries →

Snowflake ---DW on cloud ⇒

Datawarehouse

Data engineering

Data lake -

Data science -

Analytical engineering ⇒

=======================================

Storage : Any volume / variety

Structured / semi-structured / unstructured ;

Human Machine

RDB /file /APP :

=================

Client ==server ===> scalability :

CPU ⇒ GPU ==TPU ⇒ tensor processing

Elastic ⇒

Vertical /horizontal scalability ⇒

ODC ⇒

Distributed storage :

Oracle --10 tb ---> 20 column ⇒ 2 min → 16 hrs

========================================

1--1

10---1 ⇒

Data centers ---> Cluster ===Rack ==Node

=========================================

Backup ===>

Oracle ==restore ===> ?

=================================

Distributed storage ;

Post office

C1 stamp/SP/RD c2 Speed/stamp c3 RD/stamp/SP

Fault tolerance : SPOF single point of failure

Massively parallel processing (MPP):

==========================================

Data availability :

Fast queries ===> low latency / high th

=========================================

Data storage :

Database ----->

Schema ===>

tables

Database : metadata /

Schema ⇒ structure /skeleton

Oracle sqlserver /mysql

===================================

HR /scott ==schema database

Schema /user user

===================================

Tables tables

=====================================

Dw --hierarchy --schema

=================================

Ltidb

⇒ ltischema

=> emp

Data engineering :

Write once ,Read many - worm

-immutable ===>

Hadoop / Hive ==spark >

ACID → CAP theorem /BASE theorem

Single Account

Sync -->check -in --check out

Core bank ATM 10k Mobile Internet

====================================

Data storage / data processing

Oracle db →

=======================================

Data storage ⇒ shared disk arch

Central repository

Distributed :

Sales.txt --->300 mb → blocks ---> 128 mb

Data processing : MPP

M1 java m2 -data

Namenode java

D! B1 D2 b2 d3 d4 b3

Human < -- food

Tiger --> food

Sharding

Sales.txt 300 mb --b1 128 mb

B2 128 mb

B3 44 mb

Snowflake ==AWS ===DAta storage ---> AWS s3

=======================================

Query processing /compute ⇒ warehouse

AWS EC2 ⇒ 10

====================================

10gb table ⇒ size ⇒ snowflake data storage

Warehouse → 10

1 1 1 1 1 1 1 1 1 1