**Day 1:**

**Data Engineering:**

Data : raw fact. Number format, string format, Boolean format, picture format, clips format.

Information : meaning full data or processed data.

Software : to develop the application to solve the problem.

Input a=10,b=20

Process sum = a+b;

Output write sum;

To store the data

1. File system

Limitation of file base system :

* 1. Data redundancy : duplicate records.
  2. Data inconsistency :

1. Format of the file. 100-raj-21000
   1. Security :for file we can provide security as read, write and execute etc.
   2. CRUD Operation : create/insert, delete, update and read more complex.
2. Database system

In Database we store data in proper format.

DBMS : Database Management system : it is a software which help to store in table format using row and columns.

Database model

Hierarchical model : table or files are connected top to bottom.

Network model : more than one table or files are connected with each others.

Relational model two table or files are connect with each others using relation. Logical connection using PK and FK.

RDBMS : Relational database management system.

MySQL, oracle, db2, sql server, etc.

ORDBMS

No SQL database :

Mongo db

Casandra

HBase

Neo4J

Date Engineer : it collect the data from different source like file system, rdbms or no sql database, transform and manage data to make it available for analysis and decision making.

Key responsibilities and skill

Tools and language

1. SQL, No SQL, python, big data, java, Hadoop, Scala, Apache spark etc,

Cloud provider AWS, Azure, Google cloud etc.

ETL extract transform load)

Data Engineering Vs Data science Vs Data Analytics :

Data engineering using BI (business intelligence ) and ML Machine learning and reporting tools to ensure data is clean, available, structure for analysis the data.

My SQL :

No SQL Database :

Python :

Java/ Scala

Spark

No SQL database : Mongo DB

Mongo DB is no sql category database which use collection concept to store the data using documents. Mongo db use to store the data in json format.

JSON : JavaScript object notation. {key:value}

My SQL Vs Mongo Db

Database database

Table collection

Records documents.

Schema base database schema less.

show databases;

Or

show dbs;

create database databasename;

use databasename; in mysql database.

use databasename; in mongo db if database present it will switch inside existing database else it will create and switch inside that database.

Mongo db provided pre defined object ie db. This object contains lot of pre defined functions which help to deal with mongo db database.

Created the collection

db.createCollection("Sample");

collection is like a table.

Inside a collection we store the documents. Documents is like a records.

Insert documents inside a mongo db database.

db.Sample.insertOne({name:"John"})

view all document from Sample collection

\_id is like a primary key in mongo db.

Inserting more than one document inside a collection without creating collection explicitly

db.Employees.insertMany(

[

{\_id:100,name:"Lex",age:21,salary:55000},

{\_id:101,name:"Raj",age:28,salary:59000},

{\_id:102,name:"John",age:21,salary:52000},

{\_id:103,name:"Neena",age:29,salary:59000},

{\_id:104,name:"Alice",age:22,salary:54000},

{\_id:105,name:"Bob",age:21,salary:51000},

{\_id:106,name:"Charles",age:25,salary:58000}

]

);

To view all document from a collection

**db.Employees.find();**

it is equal to select query in RDBMS database.

Select \* from Employees; Oracle

db.Empoyees.find(); mongo db

select id,name from Employees; In oracle /mysql database

db.Employees.find({},{name:1}) view name and \_id

db.Employees.find({},{name:1,age}) view name,\_id and age

db.Employees.find({},{name:1,age:1,\_id:0}) view name and age

where clause in Oracle or MySQL

select \* from employee where id=100;

select \* from employee where salary > 25000;

db.Employees.findOne({\_id:100})

db.Employees.find({salary:{$gt:55000}});

db.Employees.find({salary:{$lt:55000}});

db.Employees.find({salary:{$eq:55000}});

in Oracle or mysql

update employees set salary = 65000 where id=100;

db.Employees.updateOne({\_id:100},{$set:{salary:65000}})

db.Employees.updateOne({\_id:100},{$salary:65000})

db.Employees.updateMany({age:{$gt:25}},{$set:{salary:65000}});

in oracle

delete from Employee where id=100;

db.Employees.deleteOne({\_id:100});

Students

Sid,sname,age,skillset

100, Akash,40, Java, Python

Mongo DB relationship

1. Embedded style : we store huge data in one collection.
2. Linking style

Trainer

TID(PK) TName tech

Students

SID(PK) SName age TID(FK)