**Data engineering**

Data : Raw fact

Information : processed data or meaning full data.

Data :

1. Structured format : we can convert into table format and store in database.
2. Semi structure format : json or xml etc.
3. Un structure format : audio, picture, clips etc.

We store the data in

1. file system
   1. text format
   2. byte format
   3. steam format
2. Database system

What is data engineering:

Data engineering is the discipline of designing, building and maintaining the system that collect data from different source like file system, database system and process and generate the report as well as store in other format base upon client requirements.

Source data -🡪 file system, database system, rest api call, networking, cloud etc

Data transformation🡪 cleaning, normalizing, and structuring data for analysis.

Data storage -🡪 storing processed data in data warehouse or data lakes etc.

Dave delivery -> providing data pipelines that feed BI dashboard, MI models or generate the reports.

Database :

RDBMS 🡪 MySQL, Oracle, db2, sql server or postgres SQL Database.

NonSQL 🡪 Mongo Database: Json format

Mongo DB is an no SQL type of database which help to store the data using document concept with json format.

Mongo db is schema less database.

In mongo db we use collection concept to store the data

MySQL Mongo DB

Database database

Table collection

Records document

Data in table format data in json format key:value format

PK no PK \_id is like PK

Join join not required.

show databases;

use neha\_db;

show collection

or

show tables

mongo db provided pre defined object id **db**

**db.createCollection("Sample");**

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

to view document from collection in mongo

by default mongo db created \_id as pre defined attribute like PK.

db.Employees.insertMany([

{\_id:1,name:"Ravi",age:21,salary:45000,city:"Bangalore"},

{\_id:2,name:"Ramesh",age:24,salary:42000,city:"Mumbai"},

{\_id:3,name:"Rajesh",age:28,salary:49000,city:"Bangalore"},

{\_id:4,name:"Lokesh",age:29,salary:41000,city:"Pune"},

{\_id:5,name:"Mahesh",age:30,salary:46000,city:"Bangalore"},

{\_id:6,name:"Ajay",age:31,salary:47000,city:"Pune"}

]);

**filter document in mongo db**

db.Employees.find({\_id:1});

db.Employees.find({salary:45000});

db.Employees.find({city:"Bangalore"});

db.Employees.find({salary:{$gt:45000}});

db.Employees.find({salary:{$lt:45000}});

db.Employees.find({$and:[{\_id:1},{salary:45000}]});

db.Employees.find({$or:[{\_id:1},{city:"Bangalore"}]});

**retrieve specific field information from mongo db**

**it retrieve name and \_id**

db.Employees.find({},{name:1});

**it retrieve name,\_id and salary**

db.Employees.find({},{name:1,salary:1});

**it retrieve name**

db.Employees.find({},{name:1,salary:1,\_id:0});

Python --🡪

Basic programming

Looping

List, set, tuple, dictionary

Numpy

Pandas