what is big data

difference between mapreduce and spark

what is DAG

What are transformations? What are flatmap and map? Tell the difference with syntax.

Tasks performed in spark Project 1

Attributes of Spark.read function

What is AWS?

Real time services of cloud

sql joins

query for 2nd highest salary

difference between inline query and subquery

Rank and the dense rank function difference in SQL with syntax.

How to calculate commission percentage if salary and commission are given in an emp table

Difference between joins and union

difference between self join and subquery

normalisation in dbms : https://notepad.pw/ji1m21jm

python

given a list of names, find all the names starting with letter "H"

how to read csv files in python

Python anonymous functions and lambda function.

Python data types and data structures.

merging two or more csv files

dataframe operations

1.What all concepts you studies in training

2.What training are you doing rn

3.How to load data to snowflake

4.put and copy commands

5.what is the format of data stored in snowflake

6.hierarchy of data storage

7.concat in sql

8.wordcount program in pyspark

1. Partitioning vs Bucketing

2. Why spark does lazy evaluation

3. Internal Table vs External Table in Hive

4. Why do we perform transformations in Spark

Answers

**Normalization** in DBMS: 1NF, 2NF, 3NF and BCNF in Database

Normalization is a process of organizing the data in database to avoid data redundancy, insertion anomaly, update anomaly & deletion anomaly.

Here are the most commonly used normal forms:

First normal form(1NF) : As per the rule of first normal form, an attribute (column) of a table cannot hold multiple values. It should hold only atomic values.

Second normal form(2NF) : A table is said to be in 2NF if both the following conditions hold: 1) Table is in 1NF (First normal form)

2) No non-prime attribute is dependent on the proper subset of any candidate key of table.

Third normal form(3NF) : A table design is said to be in 3NF if both the following conditions hold: 1) Table must be in 2NF

2) Transitive functional dependency of non-prime attribute on any super key should be removed.

Boyce & Codd normal form (BCNF) : It is an advance version of 3NF that’s why it is also referred as 3.5NF. BCNF is stricter than 3NF. A table complies with BCNF if it is in 3NF and for every functional dependency X->Y, X should be the super key of the table.

for examples : https://beginnersbook.com/2015/05/normalization-in-dbms/

-------------------------------------------------------------------------------------------------------------------------------

Difference between

Mapreduce vs spark

Hive Managed table vs External

Partitioning vs Bucketing

row oriented vs column oriented

spark

map vs flatmap

RDD vs Dataframe vs Datasets

on-premises vs cloud

RDB vs nosql

ACID vs CAP /BASE

snowflake vs Redshift