

PFB Pointers and prepare accordingly.

Sparks architecture

HBase Architecture and explanation,

Ubuntu Command,

Databases Join concepts,

Hive related Working questions,

Sqoop questions,

Projects related questions

Hive :

What are the optimization techniques in hive?

Diff b/W partition and bucketing?

File formats?

B/W avro and parquet files?

Sqoop:

Load Data from rdbms to HDFS without using primary key?

-- You can load data using the mapper 1 but ideally if you have big data this would take much time for that instance -

Sqoop split by?

-- you have to provide --split-by argument with some column name.

Pig :

Complex data types in Pig?

-- Bag,tuple,Map this are the complex datatype in Pig.

what is Vectorization in hive?

--vectorization is process of block of 1024 rows instead of single row at a time and within the block each column is stored as a vector (an array primitive datatype)

what is internal table in hive?

--Internal table in hive is used for storing data into hdfs and in this table data are bounded to the table schema like incase if you delete the internal table your data also get deleted, in internal table you have to load data.

what is external table in hive?

--External table in hive is used for analysis purpose here data is not bounded to the table like incase if you delete the table only schema(metadata of tables) get deleted data will be untouched or safely in hdfs , in external table you don't have to load data

difference between internal tables and external tables in hive?

--Main difference is internal table use for storing data and external table used for analysing the data. in external table you don't have to load data you can access the data by giving LOCATION of hive default directory path.

what is managed table in hive?

--Managed table and internal table is same.

what is the command to set number of mappers in sqoop ?

--

what is compile time safety in spark?

what is compilation time execution in spark?

what is mean by loading in pig?

what is sqoop import & export?

--- import mean migration of data from RDBMS to hdfs and export means migration of data from hdfs to RDBMS.

how to enable partition table in hive?

how to enable dynamic partition table in hive?

In sqoop export password how to secure?

what is sqoop import & export and explain sqoop import and export commands?

what is partitioning in hive?

what is bucketing in hive?

what is data frame in spark?

Can we import unstructured data into HDFS using sqoop?

If hive is there why do we need to import table to HDFS?

Given a figure find out right angles?

What is singleton in Python?

What are the data processing tools used in your project?

What is the use of Kafka, Flume and advantage over sqoop?

Spark basic questions?

How many mappers are used by default ?

Brief knowledge of your project should be known.

Spark:

What is meant by RDD?

How can we create Rdds?

How many default partitions are there, while creating Rdds?

Diff b/W Rdds, Data frames and Datasets?

1) what is Vectorization in hive?

- 2) difference between internal tables and external tables in hive?
- 3)what is the command to set number of mappers in sqoop ?
- 4)what is compile time safety in spark?
- 5) what is sqoop import & export and explain sqoop import and export commands?
- 6)what is partitioning in hive?
- 7)what is bucketing in hive?
- 8)what is data frame in spark?
- 9)what is rdd in spark?

Can we import unstructured data into HDFS using sqoop?

--- obvious No , We import data from relational databases so in RDBMS is contained only structured data.

If hive is there why do we need to import table to HDFS?

Given a figure find out right angles?

What is singleton in Python?

What are the data processing tools used in your project?

What is the use of Kafka,Flume and advantage over sqoop?

Spark basic questions?

How many mappers are used by default ?

Brief knowledge of your project should be known.

