

2 mark questions (Big Data)

→ Advantages and disadvantages of NOSQL

→ easily scalability of data model

→ simplicity and maintenance of data model

→ less code to write the data code

→ easy maintenance

disadvantages:-

→ less flexible

→ High Scaled data is very difficult to filter

→ Queries are not found directly because of

Row to Column maintenance

②

Mongo DB

RDBMS

① It is a non-Relational data Base

② Suitable for hierarchical data storage

③ It is document based searching algorithm

④ It is easy for maintaining Structured and unstructured data easily

① It is a Relational data base

② Not useful for hierarchical data storage

③ It is a row based searching algorithm

④ not used for unstructured data values as they don't contain relation b/w the points

→ Features of Spark:-

- ① It is easily used for accessing real-time streaming processes
- ② Speed of Retrieving data from HDFS is more
- ③ more accurate in data modification and less error production
- ④ It is easily memorise the system ram allowance and send the data according to that
- ⑤ virtual RAM data is not created

→ HDFS Working:-

→ HDFS is known as hadoop distributed file system

→ All the files are stored in the HDFS format as with the use of NoSQL (or) MongoDB Data Base units

→ It stores the values as Nodes in column-row architecture

→ It uses map reduce function in Retrieving the data from HDFS.

→ Hadoop ~~stream~~ streaming:-

Hadoop streaming is a part of hadoop distribution system. It facilitates ease of writing map reduce programs and codes.

Hadoop streaming support all type of programming languages to write the script for retrieving data.

→ By the use of streaming mechanism help in increasing the speed of data finding and modifying.

→ Hadoop Command and distribution:-

- Hadoop count command
- Hadoop append command
- Hadoop retrieve command
- Hadoop mkdir

→ use of apache oozie

oozie is a java web application scheduled apache Hadoop jobs.

→ oozie combines different jobs which are integrated in single unit of work space

→ It integrates Hadoop stack with Yarn as an architectural center and support Hadoop jobs for Apache map reduce, pig, Hive(etc).