**Assessment**

Which Hadoop Ecosystem project is best for joining, sorting, grouping etc. operations? (More than one can be correct)

* Mahuot
* **Hive**
* Sqoop
* **Impala**

Which type of data, you will prefer Hive/Impala?

* **Structured**
* **Semi-structured**
* Unstructured
* Messy data
* None of the above

Which of the below Hadoop ecosystem projects provide Fault tolerance by default? (More than one can be correct)

* Flume
* **Kafka**
* **HDFS**
* Impala
* **Map Reduce**
* All of the above

Which of the below Data format has limited Interoperability?

* Avro
* **Sequence File**
* Parquet
* ORC
* None of the above
* All of the above

Which of the below Data format(s) is/are self describing? (More than one can be correct)

* **Avro**
* **Parquet**
* Snappy
* Text
* None of the above
* All of the above

Which of the below projects provide High Throughput Inserts?

* **HBase**
* HDFS
* Cassandra
* Map Reduce
* Spark
* Impala
* None of the above
* All of the above

Which of the below projects provide High Throughput Ingestion?

* **Kafka**
* Flume
* Sqoop JDBC
* Spark
* Impala
* None of the above
* All of the above

What is standalone mode in Impala?

* Impala on one machine cluster
* Impala in Pseudo distributed Mode
* Impala on YARN
* **Impala running on its own Daemons**
* None of the above
* All of the above

Which of the following are columnar? (More than one can be correct)

* Spark SQL
* **HBase**
* **Impala**
* **Parquet**
* All of the above

Which is not a Spark Project/Library?

* MLLib
* Graphx
* Streaming
* **Drill**
* Spork
* None of the above

Which is purpose of Cloudera Manager?

* Installation
* Configuration
* Data Governance
* Maintenance
* Option 1, 3 and 4 above
* Option 1, 2 and 3 above
* **Option 1, 2 and 4 above**

Why do we need Impala?

* For long running ETL processes
* For Log File Analysis
* **For Interactive Data Exploration**
* None of the above
* All of the above

What are limitations of HDFS?

* It is not optimised for Sequence access
* **It is not optimised for Random Read/Write**
* It doesn't support append
* None of the above
* All of the above

Why MapReduce is slower than Spark?

* **It is Disk Based Framework**
* It runs on cheap machines
* None of the above
* All of the above

What happens when we drop an external table?

* It deletes Metadata from HDFS and Data from HDFS
* It deletes Metadata from Metastore and Data from HDFS
* It deletes Data from HDFS
* **It deletes Metadata from Metastore**
* It deletes Metadata from HDFS

Which is the fastest way to transfer RDBMS data into HDFS using Sqoop?

* Using Direct Option
* Using JDBC Connection
* Using ODBC Connection
* Using Drivers
* **Using Connectors**

What is default field delimiter in Hive?

* Tab Character
* Comma Character
* **Control A**
* Control B
* Control C

Which are salient features of AVRO datatype? (More than one is correct)

* **It is splittable format**
* **It supports Schema Evolution**
* **It is binary**
* It is columnar

Which of the below project is not based on Master worker architecture?

* Mahuot
* Map Reduce
* Hive
* Sqoop
* **Impala**

Which of the below project is not based on Map Reduce architecture?

* Mahuot
* Giraph
* Hive
* Pig
* Sqoop
* **None of them**

What kind of metadata Impala caches? (More than one can be correct)

* **Block Locations**
* **Hive Metadata**
* File vs Block Mapping
* File Level Attributes
* All of above

Which of the following are not Impala Standalone Daemons?

* **llamma**
* Catalog Daemon
* State Store Daemon
* Impala Daemon
* None of above

Which of the following is one of roles of Zookeeper in HiveServer2 Architecture?

* **Concurrency Management**
* Executing Queries
* Retrieving Metadata
* Providing access to JDBC clients
* None of above

Which is the best stack for performing Sentiment analysis of Data?

* Flume, HDFS, Impala
* **Flume, HDFS, Hive**
* Flume, HDFS, Pig
* Flume, HBase, Impala
* Flume, HBase, Hive
* Flume, HBase, Pig
* None of the above

Which of the below components are not part of Flume Architecture?

* Source
* Sink
* Channel
* **None of the above**
* All of the above

Which is the right option to transfer data from Sqoop to HDFS in Avro format?

* —as-avro
* **—as-avrodatafile**
* —as-avrodata
* —as-avsc
* None of the above
* All of above

How to view few records in parquet data format?

* Parquet-tools tojsonfile.parquet
* **Parquet-tools head file.parquet**
* Parquet-tools schema file.parquet
* Parquet-tools jsonfile.parquet
* None of the above
* All of above

What considerations to take while storing celebrity data in Hadoop Cluster?

* Partition data uniformly to achieve best write throughput
* Partition data uniformly to achieve best read throughput
* **Store data together to achieve best read throughput**
* Store data together to achieve best write throughput
* None of the above
* All of above

When you will use Wide table in HBase? (More than one option can be correct)

* **When we need to retrieve all attributes in one request**
* **When we need transaction while storing a row**
* When we have limited number of attributes in each row
* When we don't know start key and end key
* None of the above
* All of above

Which are recommended open source projects for Integrating Hadoop with BI Tools?

* ORC, Hive
* Parquet, Hive
* ORC, Impala
* **Parquet, Impala**
* None of the above
* All of above

What is performed during HBase Major compaction phase? (More than one option can be correct)

* Combining 3 files together
* Combining all WAL files together
* **Removing expired data**
* **Removing additional versions of data**
* **Physical removal of data**
* None of the above
* All of above

What is purpose of Bucketing in Hive? (More than one option can be correct)

* Filtering data faster
* **Sampling Data**
* **Faster Join of dataset**
* None of the above
* All of above

What is format of HFile in HBase?

* Avro
* Parquet
* **Sequence File**
* Custom Format
* None of the above
* All of above

What is difference between Block and Input Split?

* Block is logical and Input Split is Physical
* **Block is Physical and Input Split is Logical**
* Block is same as Input split always
* None of above

Where does Shuffle & Sort phase occur in MR Job?

* Mapper Local Machine
* Reducer Local Machine
* **At Mapper Local as well as Reducer Local Machine**
* Neither on Mapper Local nor on Reducer Local Machine

How do we determine number of reducers in MR job? (Multiple options can be correct)

* Automatically Determined By Framework
* No. of reducers are determined based on number of blocks of input data
* **Complexity of reducer algorithm**
* No. of reducers are equal to number of Mapper
* **Volume of intermediate data**

What is default field delimiter in Hive?

* Tab Character
* Comma Character
* **Control A**
* Control B
* Control C

Which of the following is not an example of Spark Action?

/

0

● Count

● Collect

● Take(n)

● saveAsTextFile

● Map

Add individual feedback

Which of the following is an example of Spark Transformation? (More than one can be correct)

/

0

● Zip

● reduceByKey

● flatMap

● Count

● Map

Option 6

Add individual feedback

How to perform filtering in Spark? (More than one can be correct)

/

0

● rdd.filter(lambda a: “keyword” in a)

● rdd.filter(a =>a.contains(“keyword”))

● rdd.filter(lambda a: a.contains(“keyword”))

● rdd.filter(a => “keyword” in a)

Option 5

Add individual feedback

Which of the following expects 1 to many function?

/

0

● Map

● Zip

● FlatMap

● reduceByKey

● Filter

Option 6

Add individual feedback

How to create Pair RDD? (More than one can be correct)

/

0

● sc.wholeTextFiles

● rdd.map function

● rdd.filter function

● rdd.zip function

Add individual feedback

When to use groupByKey?

/

0

● When operation is commutative

● When operation is associative

● When operation is commutative and associative

● When operation is not commutative and not associative

Add individual feedback

Which of the following expects 1 to many function?

/

0

● Map

● Zip

● FlatMap

● reduceByKey

● Filter

Consider 2 Tables Table1 and Table2 with data as follows -

Table1

ID Name

1 ABC

2 XYZ

3 LKJ

4 UIK

Table2

ID Name2

1 Hi

1 There

2 Bye

3 See

3 You

Question 1 -

Select count(\*) from Table1 Left Outer Join Table2 on Table1.ID = Table2.ID

Solutions Options

15

3

5

6 - Answer

Question 2 -

Select count(\*) from Table1 Inner Join Table2 on Table1.ID = Table2.ID

Solutions Options

15

5 - Answer

3

6

Question 3 -

Select count(\*) from Table1 Right Outer Join Table2 on Table1.ID = Table2.ID

Solutions Options

15

5 - Answer

3

6

Question 4 -

Select Table1.ID from Table1 Union ALL Select Table2.ID

Solutions Options

9 - Answer

5

3

4

Question 5 -

Select COUNT(DISTINCT id) FROM TABLE2

Solutions Options

3 - Answer

5

6

4

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

Consider a Table with data as follows -

Table

ID Name

1 ABC

2 XYZ

3 LKJ

3 UIK

Null UIK

Question 6 - Select count(ID) from Table ;

Solutions Options

4

5

3 - Answer

6

x`x`

Question 7 -

Select count(\*) from Table ;

Solutions Options

4

5 - Answer

3

6

Question 8 -

In order to select the Null record, what will be your SQL query ;

Solutions Options

select \* from Table where Name = 'UIK'

select \* from Table where ID = Null

select \* from Table where ID is Null ---Answer

select \* from Table where ID = 'Null'

Question 9 -

Which is the command to move the data from HDFS to MySQL database

Solutions Options

sqoop import

sqoop export - Answer

sqoop import all

sqoop export all

Question 10 -

How can we connect to the Hive shell ?

Solutions Options

hive -- Answer

hive -f

beeline -u ............ -- Answer

impala -shell

Question 11 -

What are the default number of mappers in a sqoop command

Solutions Options

1

2

4 ---- Answer

6

Question 12 -

How to create a new Hive table structure similar to a existing table

Solutions Options

Use "Like" Command -- Answer

Use "Create Table as Select" Command

Use "Create table as Old\_TableName"

Use "Alter Old\_tableName rename to New\_Table"

Question 13 -

How to extract the substring "Jammy" from the following field Line -

Line = "This is Jammy on way to Jam"

Solutions Options

select substr(Line,8,5) -- Answer

select substr(Line,9,4)

select substr(Line,24,5)

select substr(Line,7,5)