

1) Which component of the apache ambari architecture stores the cluster configuration

Postgres RDBMS

2) Which of the following can sqoop import from a relational database? (two answers)

All rows of table.

Specific rows and columns using a query.

3) When creating a waston studio project , what do you need to specify?

sparkservice

4) Who can control waston studio project assests?

Editors

5) What is the primary purpose of apache Nifi ?

Collect and send data into a stream

6) What python statement is used to add a library to the current code cell?

Import

7) Which feature allows application developers to easily use the ambari interface to integrate Hadoop provisioning management a capabilities into their own application?

Rest APis

8) Which visualization library is developed by IBM as an add-on to python ...

PixieDust

9) Which feature allows the big sql user to securely access data in Hadoop on behalf ..

Impersonation

10) What is the name of the Scala programming feature that provides functions with no names?

Lambda functions

11) You need to create a table that is not managed by the BigSql database manager

External

12) In Big Sql, what is used for table definitions , location and storage format of input file

The hive Metastore

13) Which two options can be used to start and stop Big Sql?

Ambari web interface

Command line

14) Which Statement describes the purpose of Ambari?

Its is used for provisioning, managing and monitoring Hadoop clusters.

15) which two are the driving principles of MapReduce?

Spread data across a large cluster of computers

Run your programs on the nodes that have the data

16) Which component of the hdfs architecture manages the file system namespace and metadata

NameNode

17) What is one disadvantage to using CSV formatted data in a Hadoop data store?

It is difficult to represent complex data structures such as maps

18) which statement describes “BIGDATA” as it is used in the moderns?

non-conventional strategies and innovative technologies used by

businesses and organizations to capture, manage, process, and make sense of a large volume of data

19) Which two are features of the Hadoop distributed file system (HDFS)

3 replicas of each unit of data

files are split into blocks

20) Which statement is true about Hadoop Distributed file System (HDFS)

HDFS links the disks on multiple nodes into one large file system

21) Which Hortonworks Data Platform HDP component provides a common Web user interface for applications running on a Hadoop cluster?

Ambari

22) When sharing a notebook what will always point to the most recent version of the notebook?

The permalink

23) Which two of the following are column-based data encoding formats?

Parquet

ORC

24) Which two are use cases for deploying Zookeeper?

Configuration bootstrapping for new nodes

Simple data registry between nodes

25) what can be used to surround a multi-line string in python code

"""

26) Which two are examples of personally identifiable information (PII)?

E-mail address

IP address

27) Which component of the spark unified stack provides processing of data arriving at the

Spark Streaming

28) Which two of the following data sources are currently supported by big sql

Oracle

Teradata

29) which data type can cause significant performance degradation and should be avoided

String

30) What is the default web location for a local jupyter instance?

localhost8888

31) you need to define a server to act as the medium between an application and a data source

Create Server

32) under the MapReduce v1 architecture which element of MapReduce controls job execution

jobTracker

33) the big sql head node has a set of processes running. What is the name of

Bigsq1

34) what must be done before using sqoop to import from a relation database

Copy an appropriate jdbc drive jar

35) Which spark RDD operation returns values after performing the evaluation

Actions

36) What is the default data format sqoop parses to export to a database?

CSV

37) Under the Mapreduce architecture which element of the system manages the map and reduce

TaskTracker

38) You can import preinstalled libraries if you are using which languages?

Python

R

39) in a Hadoop cluster which two are the result of adding more nodes of the cluster?

It increases available processing power

It adds capacity to the file system

40) Which Big Sql authentication mode is designed to provide strong authentication secret-key

cryptography

Kerberos

41) Under the Mapreduce architecture which function is performed by the jobTracker

Accepts mapReduce jobs submitted by clients

42) Under the HDFS storage model , what is the default method of replication

3 replicas 2 on the same rack 1 on a different rack

43) What is hortonworks dataPlane services DPS used for

Manage Secure and govern data stored across all storage environments

44) which command would you run to make a remote table accessible using an alias

Create Wrapper

45) Which type of foundation does Big Sql build on

Apache Hive

46) What are three examples of big data

Messages tweeted on twitter , web server logs , bank records

47) Which the primary advantage of using columns-based data formats over record-based formats

Faster query execution

48) Which of the five V of big data describes the real purpose of deriving business insight from

bigdata

Value

49) Which Spark RDD operation creates a directed acyclic graph through lazy evaluations

Transformations

50) Which element of Hadoop is responsible of spreading data across the cluster

HDFS

51) how does Mapreduce use Zookeeper

Coordination between servers

52) What are three examples of “Data Exhaust”

Cookies

Logfiles

Browser cache

53) Which statement is true about spark’s resilient distributed dataset (RDD)

It’s a distributed collection of elements that are parallelized across the cluster