

AI COMPUTE

Total points 58/60

60 QUESTIONS. ALL QUESTIONS ARE COMPULSORY

Email *

patsonu63@gmail.com

NAME & PRN

Sumit Suresh patil 220980728007

COURSE

- DAC
- DAI
- DBDA



CENTRE

- 1. BENGALURU
- 2. CHENNAI
- 3. GUWAHATI
- 4. HYDERABAD
- 5. KARAD
- 6. KOCHI
- 7. KOLKATA
- 8. MUMBAI
- 9. NASHIK
- 10. NOIDA
- 11. PATNA
- 12. PUNE
- 13. SILCHAR
- 14. THIRUVANANTHAPURAM



✓ 1. Spark was initially started by _____ at UC Berkeley AMPLab in 2009.

- a) Mahek Zaharia
- b) Matei Zaharia
- c) Doug Cutting
- d) Stonebraker

✓ 2. Point out the correct statement. *

- a) RSS abstraction provides distributed task dispatching, scheduling, and basic I/O functionalities
- b) For cluster manager, Spark supports standalone Hadoop YARN
- c) Hive SQL is a component on top of Spark Core
- d) None of the mentioned

✓ 3. _____ is a component on top of Spark Core. *

- a) Spark Streaming
- b) Spark SQL
- c) RDDs
- d) All of the mentioned

✓ 4. Spark SQL provides a domain-specific language to manipulate _____ in Scala, Java, or Python.

- a) Spark Streaming
- b) Spark SQL
- c) RDDs
- d) All of the mentioned

✓ 5. Point out the wrong statement. *

- a) For distributed storage, Spark can interface with a wide variety, including Hadoop Distributed File System (HDFS)
- b) Spark also supports a pseudo-distributed mode, usually used only for development or testing purposes
- c) Spark has over 465 contributors in 2014
- d) All of the mentioned

✓ 6. _____ leverages Spark Core fast scheduling capability to perform streaming analytics.

- a) MLlib
- b) Spark Streaming
- c) GraphX
- d) RDDs



✓ 7. _____ is a distributed machine learning framework on top of Spark.

- a) MLlib
- b) Spark Streaming
- c) GraphX
- d) RDDs

✓ 8. _____ is a distributed graph processing framework on top of Spark. *

- a) MLlib
- b) Spark Streaming
- c) GraphX
- d) All of the mentioned

✓ 9. GraphX provides an API for expressing graph computation that can model the _____ abstraction.

- a) GaAdt
- b) Spark Core
- c) Pregel
- d) None of the mentioned

✓ 10. Spark architecture is _____ times as fast as Hadoop disk-based Apache Mahout and even scales better than Vowpal Wabbit.

- a) 10
- b) 20
- c) 50
- d) 100

✓ 11. Users can easily run Spark on top of Amazon's _____ *

- a) Infosphere
- b) EC2
- c) EMR
- d) None of the mentioned

✓ 12. Point out the correct statement. *

- a) Spark enables Apache Hive users to run their unmodified queries much faster
- b) Spark interoperates only with Hadoop
- c) Spark is a popular data warehouse solution running on top of Hadoop
- d) None of the mentioned



✓ 13. Spark runs on top of _____ a cluster manager system which provides efficient resource isolation across distributed applications.

- a) Mesjs
- b) Mesos
- c) Mesus
- d) All of the mentioned

✓ 14. Which of the following can be used to launch Spark jobs inside MapReduce?

- a) SIM
- b) SIMR
- c) SIR
- d) RIS

✓ 15. Point out the wrong statement. *

- a) Spark is intended to replace, the Hadoop stack
- b) Spark was designed to read and write data from and to HDFS, as well as other storage systems
- c) Hadoop users who have already deployed or are planning to deploy Hadoop YARN can simply run Spark on YARN
- d) None of the mentioned

✓ 16. Which of the following language is not supported by Spark? *

- a) Java
- b) Pascal
- c) Scala
- d) Python

✓ 17. Spark is packaged with higher level libraries, including support for _____ queries.

- a) SQL
- b) C
- c) C++
- d) None of the mentioned

✓ 18. Spark includes a collection over _____ operators for transforming data and familiar data frame APIs for manipulating semi-structured data.

- a) 50
- b) 60
- c) 70
- d) 80



✓ 19. Spark is engineered from the bottom-up for performance, running _____ faster than Hadoop by exploiting in memory computing and other optimizations.

- a) 100x
- b) 150x
- c) 200x
- d) None of the mentioned

✓ 20. Spark powers a stack of high-level tools including Spark SQL, MLlib for _____

- a) regression models
- b) statistics
- c) machine learning
- d) reproductive research



✓ 21. Apache Flume 1.3.0 is the fourth release under the auspices of Apache of the so-called _____ codeline.

- a) NG
- b) ND
- c) NF
- d) NR

✓ 22. Point out the correct statement.

- a) Flume is a distributed, reliable, and available service
- b) Version 1.5.2 is the eighth Flume release as an Apache top-level project
- c) Flume 1.5.2 is production-ready software for integration with hadoop
- d) All of the mentioned

✓ 23. _____ was created to allow you to flow data from a source into your Hadoop environment.

- a) Impala
- b) Oozie
- c) Flume
- d) All of the mentioned

✓ 24. A _____ is an operation on the stream that can transform the stream.

- a) Decorator
- b) Source
- c) Sinks
- d) All of the mentioned

✓ 25. Point out the wrong statement.

- a) Version 1.4.0 is the fourth Flume release as an Apache top-level project
- b) Apache Flume 1.5.2 is a security and maintenance release that disables SSLv3 and TLS 1.0 on all components in Flume that support SSL/TLS
- c) Flume is backwards-compatible with previous versions of the Flume 1.x codebase
- d) None of the mentioned

✓ 26. A number of _____ source adapters give you the granular control to grab a specific file.

- a) multimedia file
- b) text file
- c) image file
- d) none of the mentioned



✓ 27. _____ is used when you want the sink to be the input source for another operation.

- a) Collector Tier Event
- b) Agent Tier Event
- c) Basic
- d) All of the mentioned

✓ 28. _____ is where you would land a flow (or possibly multiple flows joined together) into an HDFS-formatted file system.

- a) Collector Tier Event
- b) Agent Tier Event
- c) Basic
- d) All of the mentioned



✓ 29. _____ sink can be a text file, the console display, a simple HDFS path, or a null bucket where the data is simply deleted.

- a) Collector Tier Event
- b) Agent Tier Event
- c) Basic
- d) None of the mentioned

✓ 30. Flume deploys as one or more agents, each contained within its own instance of_____

- a) JVM
- b) Channels
- c) Chunks
- d) None of the mentioned

✓ 31. _____ provides Java-based indexing and search technology. *

- a) Solr
- b) Lucene Core
- c) Lucy
- d) All of the mentioned



✓ 32. Point out the correct statement. *

- a) Building PyLucene requires GNU Make, a recent version of Ant capable of building Java Lucene and a C++ compiler
- b) PyLucene is supported on Mac OS X, Linux, Solaris and Windows
- c) Use of setuptools is recommended for Lucene
- d) All of the mentioned

✓ 33. _____ is a high performance search server built using Lucene Core.

- a) Solr
- b) Lucene Core
- c) Lucy
- d) PyLucene

✓ 34. _____ is a subproject with the aim of collecting and distributing free materials.

- a) OSR
- b) OPR
- c) ORP
- d) ORS



✓ 35. Point out the wrong statement. *

- a) PyLucene is a Lucene port
- b) PyLucene embeds a Java VM with Lucene into a Python process
- c) The PyLucene Python extension, a Python module called lucene is machine-generated by JCC
- d) PyLucene is built with JCC

✓ 36. _____ is a Python port of the Core project. *

- a) Solr
- b) Lucene Core
- c) Lucy
- d) PyLucene

✓ 37. The Lucene _____ is pleased to announce the availability of Apache Lucene 5.0.0 and Apache Solr 5.0.0.

- a) PMC
- b) RPC
- c) CPM
- d) All of the mentioned



✓ 38. _____ is a technology suitable for nearly any application that requires full-text search, especially cross-platform.

- a) Lucene
- b) Oozie
- c) Lucy
- d) All of the mentioned

✓ 39. Lucene provides scalable, high-Performance indexing over _____ per hour on modern hardware.

- a) 1 TB
- b) 150GB
- c) 10 GB
- d) None of the mentioned

✓ 40. Lucene index size is roughly _____ the size of text indexed. *

- a) 10%
- b) 20%
- c) 50%
- d) 70%



✓ 41. _____ is the architectural center of Hadoop that allows multiple data processing engines.

- a) YARN
- b) Hive
- c) Incubator
- d) Chuckwa

✓ 42. Point out the correct statement. *

- a) YARN also extends the power of Hadoop to incumbent and new technologies found within the data center
- b) YARN is the central point of investment for Hortonworks within the Apache community
- c) YARN enhances a Hadoop compute cluster in many ways
- d) All of the mentioned

✓ 43. YARN's dynamic allocation of cluster resources improves utilization over more static _____ rules used in early versions of Hadoop.

- a) Hive
- b) MapReduce
- c) Impala
- d) All of the mentioned

✓ 44. The _____ is a framework-specific entity that negotiates resources from the Resource Manager.

- a) NodeManager
- b) ResourceManager
- c) ApplicationMaster
- d) All of the mentioned

✓ 45. Point out the wrong statement. *

- a) From the system perspective, the ApplicationMaster runs as a normal container.
- b) The ResourceManager is the per-machine slave, which is responsible for launching the applications' containers
- c) The NodeManager is the per-machine slave, which is responsible for launching the applications' containers, monitoring their resource usage
- d) None of the mentioned

✓ 46. Apache Hadoop YARN stands for _____ *

- a) Yet Another Reserve Negotiator
- b) Yet Another Resource Network
- c) Yet Another Resource Negotiator
- d) All of the mentioned

✓ 47. MapReduce has undergone a complete overhaul in hadoop is _____

- a) 0.21
- b) 0.23
- c) 0.24
- d) 0.26

✓ 48. The _____ is the ultimate authority that arbitrates resources among all the applications in the system.

- a) NodeManager
- b) ResourceManager
- c) ApplicationMaster
- d) All of the mentioned

✓ 49. The _____ is responsible for allocating resources to the various running applications subject to familiar constraints of capacities, queues etc.

- a) Manager
- b) Master
- c) Scheduler
- d) None of the mentioned

✓ 50. The CapacityScheduler supports _____ queues to allow for more predictable sharing of cluster resources.

- a) Networked
- b) Hierarchical
- c) Partition
- d) None of the mentioned

✓ 51. Yarn commands are invoked by the _____ script. *

- a) hive
- b) bin
- c) hadoop
- d) home

✓ 52. Point out the correct statement.

- a) Each queue has strict ACLs which controls which users can submit applications to individual queues individual queues
- b) Hierarchy of queues is supported to ensure resources are shared among the same queues of an organization
- c) Queues are allocated a fraction of the capacity of the grid in the sense that a certain capacity of resources will be at their disposal
- d) All of the mentioned

✗ 53. The queue definitions and properties such as _____ ACLs can be changed, at runtime.

- a) tolerant
- b) capacity
- c) speed
- d) all of the mentioned

Correct answer

- b) capacity

✓ 54. The CapacityScheduler has a predefined queue called _____ *

- a) domain
- b) root
- c) rear
- d) all of the mentioned

✖ 55. Point out the wrong statement.

- a) The multiple of the queue capacity which can be configured to allow a single user to acquire more resources
- b) Changing queue properties and adding new queues is very simple
- c) Queues cannot be deleted, only addition of new queues is supported
- d) None of the mentioned

Correct answer

- d) None of the mentioned

✓ 56. The updated queue configuration should be a valid one i.e. queue-capacity at each level should be equal to _____

- a) 50%
- b) 75%
- c) 100%
- d) 0%



✓ 57. Users can bundle their Yarn code in a _____ file and execute it using jar command.

- a) java
- b) jar
- c) C code
- d) xml

✓ 58. Which of the following command is used to dump the log container? *

- a) logs
- b) log
- c) dump
- d) all of the mentioned

✓ 59. _____ will clear the RMStateStore and is useful if past applications are no longer needed.

- a) -format-state
- b) -form-state-store
- c) -format-state-store
- d) none of the mentioned

✓ 60. Which of the following command runs ResourceManager admin client?

- a) proxyserver
- b) run
- c) admin
- d) rmadmin

This content is neither created nor endorsed by Google. - [Terms of Service](#) - [Privacy Policy](#)

Google Forms





