Program: BE ----Computer----- Engineering

Curriculum Scheme: Revised 2016

Examination: Final Year Semester VII

Course Code: \_\_ CSDLO7032 and Course Name: \_\_ Big Data & Analytics

Time: 1hour Max. Marks: 50

====================================================================

Note to the students:- All the Questions are compulsory and carry equal marks .

|  |  |
| --- | --- |
| Q1.  | Following is not HIVE function. |
| Option A: | Summarization |
| Option B: | Query |
| Option C: | Analysis |
| Option D:  | Optimization |
|  |  |
| Q2. | Hive also support custom extensions written in  |
| Option A: | C# |
| Option B: | Java |
| Option C: | PHP |
| Option D: | C++ |
|  |  |
| Q3. | \_\_\_\_\_\_\_ jobs are optimized for scalability but not latency |
| Option A: | Mapreduce |
| Option B: | Drill |
| Option C: | Oozie |
| Option D: | Hive |
|  |  |
| Q4. | Point out the wrong statement. |
| Option A: | Replication Factor can be configured at a cluster level (Default is set to 3) and also at a file level |
| Option B: | Block Report from each DataNode contains a list of all the blocks that are stored on that DataNode |
| Option C: | User data is stored on the local file system of DataNodes |
| Option D: | DataNode is aware of the files to which the blocks stored on it belong to |
|  |  |
| Q5. | \_\_\_\_\_\_\_\_ is the slave/worker node and holds the user data in the form of Data Blocks. |
| Option A: | DataNode |
| Option B: | NameNode |
| Option C: | Data block |
| Option D:  | Replication |
|  |  |
| Q6. | HDFS is implemented in \_\_\_\_\_\_\_\_\_\_\_\_\_ programming language |
| Option A: | C++ |
| Option B: | Java |
| Option C: | Scala |
| Option D:  | PHP |
|  |  |
| Q7.  | For YARN, the \_\_\_\_\_\_\_\_\_\_\_ Manager UI provides host and port information |
| Option A: | Data Node |
| Option B: | NameNode |
| Option C: | Resource |
| Option D:  | Replication |
|  |  |
| Q8.  | During start up, the \_\_\_\_\_\_\_\_\_\_\_ loads the file system state from the fsimage and the edits log file. |
| Option A: | DataNode |
| Option B: | NameNode |
| Option C: | ActionNode |
| Option D:  | Secondary Node |
|  |  |
|  |  |
| Q9.  | Shared Nothing Architecture shares |
| Option A: | Shares RAM |
| Option B: | shares Nothing |
| Option C: | shares disk |
| Option D:  | shares Processor |
|  |  |
| Q10.  | In master-slave distribution model,cluster can be configured with a standby master  |
| Option A: | that’s continually updated from the slave node. |
| Option B: | that’s continually updated from the master node. |
| Option C: | that’s continually updated from the both master and slaves node. |
| Option D:  | It automatically updates itself. |
|  |  |
| Q11.  | A hash ring technique in NoSQL used  |
| Option A: |  to create hash values. |
| Option B: |  to distribute big data loads over many servers . |
| Option C: |  to create hash tables. |
| Option D: | to create hash map. |
|  |  |
| Q12. | Following is not example of sources of stream data. |
| Option A: | Sensor data |
| Option B: | Image data |
| Option C: | Internet and web traffic |
| Option D:  | Microdata |
|  |  |
| Q13.  | For counting the number of ones in a long stream of bits approximately,following algorithm is used. |
| Option A: | DSMS |
| Option B: | DGIM |
| Option C: | MGIM |
| Option D:  | PGIM |
|  |  |
| Q14. | In DGIM,the right end of a bucket is always a position with a .......... |
| Option A: | 1 |
| Option B: | 0 |
| Option C: | Not Fixed |
| Option D:  | Any number |
|  |  |
| Q15.  | Select correct statement. |
| Option A: | Overlaping of buckets in DGIM is allowed. |
| Option B: | Overlaping of buckets in DGIM is not allowed. |
| Option C: | Overlaping of buckets in DGIM is conditionally allowed. |
| Option D:  | No buckets are considered. |
|  |  |
| Q16. | The following is common technique sampling stream |
| Option A: | Basic Sampling |
| Option B: | Random Sampling |
| Option C: | Mixed Sampling |
| Option D: | Hetro sampling |
|  |  |
| Q17. | For building a one-pass synopsis of a data set in a streaming environment,following is used. |
| Option A: | Passing |
| Option B: | Modeling |
| Option C: | Mixing |
| Option D:  | Sampling |
|  |  |
| Q18.  | Point out the wrong statement. |
| Option A: |  k-means clustering is a method of vector quantization |
| Option B: |  k-means clustering aims to partition n observations into k clusters |
| Option C: |  k-nearest neighbor is same as k-means |
| Option D:  |  K-Means is heirarchical clustering algorithm. |
|  |  |
| Q19. | Which of the following function is used for k-means clustering? |
| Option A: |  k-means |
| Option B: |  k-mean |
| Option C: |  Heatmap |
| Option D: | Kinetic Means |
|  |  |
| Q20. | Cosine similarity is often used in information retrieval within the  |
| Option A: | 3D Model |
| Option B: | Scalar Space Model |
| Option C: | Vector Space Model |
| Option D:  | Homogenous model |
|  |  |
| Q21.  | Which of the following methods can be used to solve the edit distance problem? |
| Option A: |  Recursion |
| Option B: |  Dynamic programming |
| Option C: |  Both dynamic programming and recursion |
| Option D:  |  Greedy Algorithm |
|  |  |
| Q22. | It is a model or representation of a social network |
| Option A: | Social network tree |
| Option B: | Social network graph |
| Option C: | social network media |
| Option D:  | bar graph |
|  |  |
| Q23.  | One of Analysis of social network’s tasks is  |
| Option A: | detecting communities |
| Option B: | finding number of nodes |
| Option C: | finding scale of graph |
| Option D:  | Joining many graphs |
|  |  |
| Q24. | Page Rank is  |
| Option A: | iterative graph processing algorithm |
| Option B: | recursive graph processing algorithm |
| Option C: | functional algorithm |
| Option D: | Procedural algorithm |
|  |  |
| Q25. | Two K-Clique is adjacent if |
| Option A: | they share K+1 nodes |
| Option B: | they share K nodes |
| Option C: | they share K-1 nodes |
| Option D:  | they share any number of nodes. |