Program: BE Engineering

Curriculum Scheme: Revised 2012

Examination: Semester VIII

Course Code: CPE8035 and Course Name: Big Data Analytics

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

|  |  |
| --- | --- |
| Q1.  | Big Data is processed a huge amount of data using  |
| Option A: | Multiprocessing |
| Option B: | Batch processing |
| Option C: | Sequentially |
| Option D:  | Variety |
|  |  |
| Q2. | The Facebook always generated Big data in the form of  |
| Option A: | Image |
| Option B: | Image & video |
| Option C: | Video |
| Option D: | Structured |
|  |  |
| Q3. | Any data that can be stored, accessed and processed in the form of fixed format is termed as a  |
| Option A: | Huge data |
| Option B: | Structured |
| Option C: | Unstructured  |
| Option D: | Semi Structured |
|  |  |
| Q4. | which is not the example of Big Data |
| Option A: | stock exchanges,  |
| Option B: | Semi Structured |
| Option C: | social media sites |
| Option D: | jet engines |
|  |  |
| Q5. | When the Name node goes down |
| Option A: | Job tracker will control |
| Option B: | Task tracker will control |
| Option C: | System restart |
| Option D:  | System failure |
|  |  |
| Q6. | Data Node creates a block (replica ) of minimum size is |
| Option A: | 64KB |
| Option B: | 64GB |
| Option C: | 64MB |
| Option D:  | 64TB |
|  |  |
| Q7.  | Map reduced is a  |
| Option A: | Programming Model |
| Option B: | Distributed Model |
| Option C: | HDFS parts |
| Option D:  | Blocks  |
|  |  |
| Q8.  | MongoDB managed Structured as well as unstructured data in the form of  |
| Option A: | Table |
| Option B: | Document |
| Option C: | Image |
| Option D:  | Collection |
|  |  |
| Q9. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_is multi-structural data. |
| Option A: | Document |
| Option B: | XML |
| Option C: | Dynamic DB |
| Option D:  | Collection |
|  |  |
| Q10.  | MongoDB uses  |
| Option A: | Dynamic schemas |
| Option B: | JSON |
| Option C: | BSON |
| Option D:  | XML |
|  |  |
| Q11.  | The following is the rules that must be followed when representing a stream by buckets. In DGIM |
| Option A: | All sizes must be a power of 3. |
| Option B: | There are one or two buckets of any given size, up to some maximum size. |
| Option C: | The right end of a bucket is always a position with a 0. |
| Option D:  | Buckets can decrease in size. |
|  |  |
| Q12.  | Euclidean distance is a measure of the true straight line distance between two points in |
| Option A: | Collection |
| Option B: | XY Space |
| Option C: | Documents |
| Option D: | Euclidean space |
|  |  |
| Q13. | Streaming algorithms are the method of choice if emerging data must be processed in a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ without completely storing it. |
| Option A: | Distributed |
| Option B: | Collection |
| Option C: | Streaming data |
| Option D:  | Real-time manner |
|  |  |
| Q14.  | Sliding-window Model is divide the data stream into\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Option A: | Blocks by time |
| Option B: | Blocks by size |
| Option C: | Document By Time |
| Option D:  | Small Chunks |
|  |  |
| Q15. | Time based Sliding window model maintain Frequents \_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Option A: | Documents |
| Option B: | Collection |
| Option C: | items set |
| Option D:  | Big data |
|  |  |
| Q16.  | **CURE(Clustering Using Representatives) is based on** |
| Option A: | Agglomerative Hierarchical Cluster |
| Option B: | Sampling Cluster |
| Option C: | Hierarchical Cluster |
| Option D:  | K- NN  |
|  |  |
| Q17. | A process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources called as  |
| Option A: | Random Filtering  |
| Option B: | Aglomattive Filtering |
| Option C: | Collaborative Filtering |
| Option D: | Viewpoints Filtering |
|  |  |
| Q18. | Facebook, Twitter, Google+ called as  |
| Option A: | Big Data Network |
| Option B: | Social Network |
| Option C: | Social Media |
| Option D:  | Unstructured network |
|  |  |
| Q19.  | Social networks are naturally modeled as graphs, which we sometimes refer to as a |
| Option A: | Directed Graph |
| Option B: | Undirected Graph |
| Option C: | Social graph |
| Option D:  | Structured Graph |
|  |  |
| Q20. | A view of the same data is as a graph in which the nodes are papers. Two papers are connected by an edge if they have at least one author in common. Now, we form communities that are collections of papers on the same topic called as : |
| Option A: | Collaboration Networks |
| Option B: | Email Networks |
| Option C: | Telephonic Networks |
| Option D: | Social media Network |
|  |  |
| Q21. | Finding maximal cliques is a  |
| Option A: | N |
| Option B: | NP |
| Option C: | NP- Complete |
| Option D:  | N- Complete |
|  |  |
| Q22.  | HDFS works by the |
| Option A: | Job Tracker |
| Option B: | Task Tracker |
| Option C: | Name Node |
| Option D:  | Secondary Name Node |
|  |  |
| Q23. | Split And merge task always done in  |
| Option A: | Reduced- Mapping Phase |
| Option B: | Mapping- Reduced phase |
| Option C: | Job Tracker |
| Option D:  | Data Node |
|  |  |
| Q24.  | Fault Tolerance is the Feature of  |
| Option A: | Hadoop  |
| Option B: | Big Data |
| Option C: | DGIT |
| Option D:  | Clustering |
|  |  |
| Q25. | The number of shortest paths going through each edge always describe in method called the  |
| Option A: | Clustering |
| Option B: | Girvan-Newman (GN) Algorithm |
| Option C: | Hadoop |
| Option D:  | DAG |