Code No.: IT17411

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (Autonomous)
B.E. (IT) IV/IV I Sem (Main) Examination Nov 2017

Big Data Analytics

Time: 3 Hours
Max Marks: 75

Note: Answer all questions from **Section-A** at one place in the same order
Answer any **five** questions from **Section-B**

Section - A (25 Marks)

1	What are the applications of big data?	(2)
2	What are the main components of a Hadoop Application?	(2)
3	What are the core methods of a Reducer?	(2)
4	Explain combiner function?	(3)
5	Explain mapreduce types.	(2)
6	What are the main components of Mapreduce Job	(3)
7	Differentiate SQL and NOSQL.	(3)
8	What for aggregation is used.	(3)
9	How to define Tables in Hive?	(3)
10	Compare Pig and Hive.	

Section - B (50 Marks)

11	(a) Explain the anatomy of file read operation in HDFS.	(5)
	(b) Discuss the basic HDFS commands with examples.	(5)
12	Write a mapper and reducer program to find an average temperature of season based on last ten years dataset.	(10)
13	Write short notes on application master failure, resource manager failure and node manager failure.	(10)
14	(a) Write a short note on indexing and aggregation operations in Mongo DB.	(5)
	(b) Discuss the advantages and disadvantages of Mongo DB.	(5)
15	(a) Write a short note on user-defined functions in Pig Latin.	(5)
	(b) Explain the architecture of Hive in detail.	(5)
16	Explain in detail about how to run a distributed Map Reduce job with an example program.	(10)
17	(a) Explain the architecture of Pig framework in detail.	(5)
	(b) Write about Data Processing Operators in PIG Latin.	(5)

******

Code No.: IT18411s

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (Autonomous)
B.E. (IT) IV/IV I Sem (Suppl) Examination Apr - May 2018

Big Data Analytics

Time: 3 Hours **Max Marks: 75**

Note: Answer all questions from **Section-A** at one place in the same order
Answer any **five** questions from **Section-B**

Section - A (25 Marks)

1	Define Big data.	(2)
2	Differentiate between Structured and Unstructured data	(2)
3	Explain about the partitioning, shuffle and sort phase	(3)
4	Explain the Task Execution process in Map Reduce.	(2)
5	What are the Key/Value Pairs in Map Reduce framework?	(2)
6	What is data serialization	(3)
7	Define MongoDB.	(3)
8	Define Indexing.	(2)
9	Define Data Processing Operators in Pig.	(3)
10	Explain what is a Hive variable. What do we use it for	(3)

Section - B (50 Marks)

11	Discuss in detail about HDFS architecture with neat sketch.	(10)
12 (a)	What is rack awareness in HDFS? Explain with a neat sketch.	(5)
(b)	Discuss in detail about the use of combiner function in map reduce.	(5)
13	Explain how matrix multiplication can be performed in Map Reduce framework with suitable code examples. Also explain the input format of the matrices to be processed by MR.	(10)
14	Explain the differences between SQL and NoSQL databases.	(10)
15 (a)	Illustrate with steps how to install and run Pig Latin.	(5)
(b)	Discuss different data types in Hive query language.	(5)
16 (a)	Explain how coherency model work in Hadoop distributed file system.	(5)
(b)	Discuss about any two applications where big data can be considered as a solution.	(5)
17	Explain in detail about Anatomy of a MapReduce Job Run with a neat sketch.	(10)

******