

Code No: RT4105B

R13**Set No. 1**

IV B.Tech I Semester Regular/Supplementary Examinations, Oct/Nov - 2018

HADOOP AND BIG DATA

(Common to Computer Science and Engineering and Information Technology)

Time: 3 hours

Max. Marks: 70

*Question paper consists of Part-A and Part-B**Answer ALL sub questions from Part-A**Answer any THREE questions from Part-B*

PART-A (22 Marks)

- 1 a) Write the various methods in Map interface. [4]
- b) List the characteristics of Big Data. [4]
- c) What are the functions of a combiner? [3]
- d) List the writable wrapper classes for java primitives. [4]
- e) Differentiate Apache pig with Map Reduce. [4]
- f) What do you mean by HiveQL Data Definition Language? [3]

PART-B (3x16 = 48 Marks)

- 2 a) Explain the Linked List data structure with sample example program. [8]
- b) What is data serialization? With proper examples discuss and differentiate structured, unstructured and semi-structured data. Make a note on how type of data affects data serialization. [8]
- 3 Explain in detail building blocks of Hadoop with neat sketch. [16]
- 4 a) Write Map Reduce steps for counting occurrences of specific numbers in the input text file(s). Also write the commands to compile and run the code. [8]
- b) Discuss on the different types and formats of Map-reduce with an example each one. [8]
- 5 a) What do you mean by a custom writable and explain the implementation of a custom writable with an example. [8]
- b) Explain the Writable class hierarchy with a neat sketch. [8]
- 6 a) Draw and explain architecture of APACHE PIG in detail. [8]
- b) Discuss how Pig data model will help in effective data flow. [8]
- 7 a) Draw and explain Architecture of APACHE HIVE. Explain various data insertion techniques in HIVE with example. [8]
- b) Explain any three Hive QL DDL command with its syntax and example. [8]

Code No: **RT4105B****R13****Set No. 2****IV B.Tech I Semester Regular/Supplementary Examinations, Oct/Nov - 2018****HADOOP AND BIG DATA****(Common to Computer Science and Engineering and Information Technology)****Time: 3 hours****Max. Marks: 70***Question paper consists of Part-A and Part-B**Answer ALL sub questions from Part-A**Answer any THREE questions from Part-B************PART-A (22 Marks)**

- 1 a) List various wrapper classes in Java. [4]
- b) Write the reasons for Why Hadoop won't be using JAVA serialization. [3]
- c) Define structured, semi structured and un structured data with examples. [4]
- d) Define Byte writable and Object writable wrappers. [4]
- e) Write a PIG script for Word Count. [4]
- f) Explain Metastore in Hive. [3]

PART-B (3x16 = 48 Marks)

- 2 a) Discuss about serialization concept in java. [8]
- b) Explain in brief about various map implementations in Java with suitable examples. [8]
- 3 a) Explain the procedure for Installing Hadoop in Pseudo Distributed Mode. [8]
- b) Discuss Big data in terms of three dimensions, volume, variety and velocity. [8]
- 4 a) Write Map Reduce steps for counting sum of numbers in the input text file(s). Also write the commands to compile and run the code. [8]
- b) What are core methods of a reducer? What happens if you try to run a Hadoop job with an output directory that is already present? [8]
- 5 a) Explain about the implementation of raw comparator and custom raw comparator with an example. [8]
- b) Explain the significance of Writable interface along with Writable Comparable and comparators w.r.to implementing the serialization. [8]
- 6 a) List any five commands of pig script. [8]
- b) Discuss Pig Latin Application Flow. [8]
- 7 a) Explain working of Hive with proper steps and diagram. [8]
- b) What are the different Hive data types? Explain them briefly. [8]

Code No: RT4105B

R13**Set No. 3**

IV B.Tech I Semester Regular/Supplementary Examinations, Oct/Nov - 2018

HADOOP AND BIG DATA

(Common to Computer Science and Engineering and Information Technology)

Time: 3 hours

Max. Marks: 70

*Question paper consists of Part-A and Part-B**Answer ALL sub questions from Part-A**Answer any THREE questions from Part-B*

PART-A (22 Marks)

- 1 a) What are the uses of serialization? [4]
- b) What are the various functions of name node? [4]
- c) Define the role of combiner and partitioner in a map reduce application. [3]
- d) How a custom raw comparator differs from the raw comparator? [4]
- e) List the components of Pig Execution Environment. [3]
- f) Why HIVE is relevant in Hadoop Eco system? [4]

PART-B (3x16 = 48 Marks)

- 2 a) Explain the Stack and Queue Data structure in java with sample program. [8]
- b) Discuss the use of wildcards in generic methods. [8]
- 3 a) Define HDFS. Discuss the HDFS Architecture and HDFS Commands in brief. [8]
- b) Explain in detail Hadoop distributed file system. [8]
- 4 a) Explain Map-reduce framework in detail. [8]
- b) Discuss the following [8]
 - (i) Mapper (ii) Reducer (iii) Combiner
- 5 a) Describe in brief about writable Class hierarchy with suitable examples. [8]
- b) Discuss about the writable collections. [8]
- 6 a) Discuss the various data types in Pig. [8]
- b) Write a word count program in Pig to count the occurrence of similar words in a file. [8]
- 7 a) Write the Hive commands to create a sample table of a student with roll number, name and address And insert two rows into that table. [8]
- b) Explain procedure to write user defined functions in HIVE. [8]

