Code No: **RT4105B** **R13** **Set No. 1**

IV B.Tech I Semester Regular/Supplementary Examinations, October/November - 2017

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) Define Set? List out the various set implementations in Java. [3]
   b) What is the role of Data node and Name node in HDFS? [4]
   c) What is the role of Combiner and Partitioner in map reduce application? [4]
   d) Why key type need to be both writable and comparable in Map Reduce Programs? [4]
   e) What is a PIG? Specify its Role in Hadoop? [4]
   f) How to create and manage data bases in HIVE? [3]

PART–B *(3x16 = 48 Marks)*

2. a) Define Wrapper Class? Explain in brief about writable wrappers for java primitives. [8]
   b) Differentiate between Array List and class linked list functionalities. [8]

3. a) What are the modes that a Hadoop can run? [8]
   b) Discuss in brief about the building blocks of Hadoop? [8]

4. a) Describe in brief about API for Map reduce framework. [8]
   b) Discuss in brief about the implementation of map reduce concept with suitable example. [8]

5. a) What are object writable and generic writable? [8]
   b) Explain with an example, how Hadoop uses Scale out feature to improve the performance? [8]

6. a) Discuss in brief about running a pig script in local and distributed mode. [8]
   b) Describe in brief about PIG Commands. [8]

7. a) How can you create and manage data bases in Hadoop? [8]
   b) Explain in brief about Data manipulation in HIVE. [8]

Code No: **RT4105B** **R13** **Set No. 2**

IV B.Tech I Semester Regular/Supplementary Examinations, October/November - 2017

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 data structures in Java?	[4]
	b)	List out the reasons why hadoop is not using Java Serialization.	[3]
	c)	What is the role of mapper code and reducer code in map reduce application?	[3]
	d)	Define Byte, Object and Generic writable wrappers.	[4]
	e)	Write a PIG script for Word Count.	[4]
	f)	How to create a table by using HIVE QL?	[4]

PART–B *(3x16 = 48 Marks)*

2.	a)	Write a Java program to implement generic single linked list.	[8]
	b)	Explain about the conversion from primitive type to wrapper class and vice versa with suitable example.	[8]
3.	a)	Differentiate between HDFS and GFS.	[8]
	b)	Discuss in brief about the operational modes in Hadoop cluster configuration.	[8]
4.	a)	What are the real time industry applications of Hadoop?	[8]
	b)	Explain in brief about Name node, Data Node and Secondary Name node in HDFS.	[8]
5.	a)	Explain about the implementation of raw comparator and custom raw comparator with suitable examples.	[8]
	b)	Describe in brief about the implementation of a raw comparator for speed.	[8]
6.	a)	Discuss about the operators supported by pig.	[8]
	b)	Describe in brief about the PIG Architecture.	[8]
7.	a)	Explain in brief about the data types and schemas in HIVE.	[8]
	b)	How can you write user defined functions in HIVE?	[8]

Code No: **RT4105B** **R13** **Set No. 3**

IV B.Tech I Semester Regular/Supplementary Examinations, October/November - 2017

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) Define Wrapper class? List out the wrapper classes in Java? [4]
   b) What is the role of Job tracker and Task tracker in HDFS? [3]
   c) Define structured, semi structured and un structured data with examples? [4]
   d) List out the writable wrapper classes for Java Primitives? [4]
   e) What are the three key design principles PIG Latin? [3]
   f) Describe the various File formats supported by HIVE. [4]

PART–B *(3x16 = 48 Marks)*

2. [8]
   a) Discuss in brief about Linked list class functionalities with examples. [8]
   b) Explain in brief about various map implementations in Java with suitable examples.
3. a) What are the advantages and disadvantages of Hadoop? [8]
   b) Define Data node? How does name node tackle data node failures? [8]

4. a) Discuss in brief about the Name node, Check point name node and back up node? [8]
   b) What are the different modes in which hadoop can be installed and what is the use of each mode from application and developer point of view? [8]

5. a) Explain the significance of writable interface along with writable comparable and comparators with respect to serialization. [8]
   b) Describe in brief about writable Class hierarchy with suitable examples. [8]

6. a) Consider the student data file (st.txt) Data in the following format Name, District, Age, gender
      (i) Write a PIG Script to display Names of all Male Students.
      (ii) Write a PIG Script to find the number of students from Vizianagaram district.
      (iii) Write a PIG Script to display district wise count of all female students. [8]
   b) Explain about the various data types supported by pig in its data model with an example. [8]

7. a) Discuss in brief about the Architecture of HIVE. [6]
   b) What is Hive meta store? Which classes are used by the Hive to Read and Write HDFS Files? [10]

Code No: **RT4105B** **R13** **Set No. 4**

IV B.Tech I Semester Regular/Supplementary Examinations, October/November - 2017

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 difference between wild card (?) argument and Normal type argument Generic programming in Java. [4]
   b) Define Hadoop Cluster? How can you configure Hadoop cluster? [4]
   c) List out the components of map reduce application that we can develop? [3]
   d) Define Serialization? Write about RPC Serialization Format? [4]
   e) Write about any three PIG commands? [3]
   f) What is a HIVE? Specify its Role in Hadoop. [4]

PART–B *(3x16 = 48 Marks)*

2. a) Explain in brief about the operations performed on linked list and stack with suitable examples. [8]
   b) What are the advantages of object serialization in Java? Discuss in brief about serializing and de serializing an object with suitable examples. [8]

3. a) Discuss in brief about the basic building blocks in Hadoop. [8]
   b) Explain in brief about the Architecture of GFS. [8]

4. a) 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]
   b) What is a Data Node? How many instances of Data Node run on a Hadoop Cluster? [8]

5. a) Explain in brief about I/o primitives in Hadoop. [8]
   b) Discuss in brief about the writable wrappers for Java primitives. [8]

6. a) List the relational operators in Pig? [8]
   b) What are the components of Pig Execution Environment? [8]

7. a) What are views in HIVE? What is the difference between internal and external tables in HIVE? [8]
   b) Discuss in brief about the procedure for installation of Hive. [8]