Code No: **R164105A** **R16** **Set No. 1**

IV B.Tech I Semester Regular/Supplementary Examinations, March - 2021

BIG DATA ANALYTICS

(Common to Computer Science & 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 FOUR questions from Part-B
****

PART–A *(14 Marks)*

1. a) Why primitive data types are not allowed in JAVA generics? [2]
   b) Which node takes the responsibility when the active NameNode fails? [2]
   c) What is the order of the three steps to MapReduce? [3]
   d) Mention the six writable collection types in Hadoop [2]
   e) Specify the role of Pig Latin in Hadoop. [2]
   f) What is partition and bucketing in hive? [3]

PART–B *(4x14 = 56 Marks)*

2. a) Explain the Characteristics of a Map Interface in JAVA. Give its hierarchy and write about the classes that implement Map interface. [7]
   b) What is meant by Serialization in JAVA? Why do we need Serialization in JAVA? Can we serialize a non serializable object in JAVA? Explain. [7]

3. a) Explain about Google File System. [6]
   b) What is Big Data? Explain any four significant characteristics of big data. [4]
   c) Explain the following [4]
      i) Job tracker ii)Task tracker

4. a) Explain the work flow of MapReduce process with a suitable example. [7]
   b) With suitable example, briefly describe usage of MapReduce with and without combiner. [7]

5 a) Why GenericWritable is required? With an example, illustrate the usage of GenericWritable. [7]
   b) What is ArrayWritable? Illustrate the usage of MapWritable and SortedMapWritable. [7]

6. a) Explain about Pig Latin data model and its data types. [6]
   b) Write about the three key design principles of Pig Latin [4]
   c) Write about Apache Pig execution modes and mechanism. [4]

7. a) Illustrate Create Database in HIVE. What is the importance of IF NOT EXISTS in the context of Create Database? [4]
   b) Explain any four functions on ALTER TABLE [3]
   b) When it is appropriate to go for Internal and External tables in Hive? Explain. [7]

||''|''|''|''||

Code No: **R164105A** **R16** **Set No. 2**

IV B.Tech I Semester Regular/Supplementary Examinations, March – 2021

BIG DATA ANALYTICS

(Common to Computer Science & 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 FOUR questions from Part-B

*****

PART–A *(14 Marks)*

1. a) Can we use primitive data types in generics? justify your answer. [2]
   b) What happens when a data node fails? [2]
   c) What decides number of mappers in a mapreduce job? [3]
   d) Write about GenericWritable class. [2]
   e) What is Grunt in Pig Latin? [3]
   f) Why Hive is used instead of Pig Latin? [2]

PART–B *(4x14 = 56 Marks)*

2. a) Write a JAVA program to implement various operations like adding, changing and removing elements using Map interface and HashMap class. [7]
   b) What is meant by Serialization and Deserialization in JAVA? Why static and transient variables are not serialized? Explain. [7]

3. a) Explain in detail about the key components of Hadoop architecture. [7]
   b) What is Big data? Why is big data analytics so important in today's digital era? Expain. [7]

4. a) Explain about the important Hadoop APIs for MapReduce framework. [7]
   b) Explain the order of execution of Mapper, Combiner and Partitioner in a MapReduce job with a suitable example. [7]

5. a) What is the significance of RawComparator and at what scenarios it is more appropriate? Give explanation. [7]
   b) Explain the sequence of steps for creating custom Key writable data types in Hadoop with an example program. [7]

6. a) Explain the architecture of Apache Pig with neat sketch. [7]
   b) Explain about the complex data types in Pig Latin. [7]

7. a) Demonstrate the work flow between Hive and Hadoop with suitable diagrams. [7]
   b) How does data distribution happens in Hive bucketing? Explain. [7]

||"|""|"|'|""||

Code No: **R164105A** **R16** **Set No. 3**

IV B.Tech I Semester Regular/Supplementary Examinations, March – 2021

BIG DATA ANALYTICS

(Common to Computer Science & 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 FOUR questions from Part-B
*****

PART–A *(14 Marks)*

1.	a)	How do you instantiate a generic array in Java?	[2]
	b)	How the name node detects that a particular data node is down?	[2]
	c)	What is Shuffle phase in map reduce jobs?	[3]
	d)	Write about Writable Comparable interface.	[2]
	e)	List the three key design principles of Pig Latin.	[3]
	f)	Is Hive a database? Justify your answer.	[2]

PART–B *(4x14 = 56 Marks)*

2.	a)	Write a JAVA program to perform various operations like adding, accessing and removing elements on SortedSet.	[7]
	b)	What are Generics in JAVA? What is the use of generics in Java? Discuss the advantages and limitations of Generics in JAVA.	[7]
3.	a)	What is big data? Explain the significant applications of big data.	[4]
	b)	With a neat sketch explain the typical architecture of Hadoop cluster.	[10]
4.	a)	Explain the following i) Drivercode ii) Mappercode iii) Reducercode	[6]
	b)	Write a MapReduce program in JAVA to count the number of words in a file.	[8]
5.	a)	Explain the following i) NullWritable ii) BytesWritable iii) ObjectWritable iv) GenericWritable	[7]
	b)	Explain the implementation of raw comparator and custom raw comparator with an example	[7]
6.	a)	Consider the Departmental Stores data file (stores.txt) in the following format customerName, deptName, purchaseAmount. i) Write a Pig script to list total sales per departmental store. ii) Write a Pig script to list total sales per customer.	[7]
	b)	Explain the following operators in Pig Latin. i) flatten operator ii) Relational operators	[7]
7.	a)	Explain about column types, literals and complex types supported by Hive.	[7]
	b)	Explain the following clauses with example HQL statements i) cluster by ii) distribute by	[7]

||''|'''|''|'|''''||

Code No: **R164105A** **R16** **Set No. 4**

IV B.Tech I Semester Regular/Supplementary Examinations, March – 2021

BIG DATA ANALYTICS

(Common to Computer Science & 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 FOUR questions from Part-B

*****

PART–A *(14 Marks)*

1. a) What is the use of generics in Java? [2]
   b) What kind of information is stored in NameNode? [2]
   c) Illustrate the reducer phase in a MapReduce job? [3]
   d) Write about Writable Interface. [2]
   e) Write about the key components of Apache Pig framework. [2]
   f) Which property of Hive enables users to work with different file formats? [3]

PART–B *(4x14 = 56 Marks)*

2. a) How can we implement a Stack using Queue in JAVA? Explain with an example program. [7]
   b) What is a Wrapper class in JAVA? How do you create a Wrapper class in JAVA? Why Wrapper classes are immutable in JAVA? Give explanation. [7]

3. a) What are the major sources of big data? Write about various technologies available to manage big data. [7]
   b) Explain the step by step procedure to install and setup a 5-Node Hadoop Cluster. [7]

4. a) Explain the role of combiner and partitioner in MapReduce application with a suitable example. [7]
   b) Write a JAVA program to implement matrix multiplication using Map-Reduce paradigm. [7]

5. a) How the implementation of RawComparator will speed up your Hadoop MapReduce jobs? explain. [7]
   b) Explain in detail about Writable class hierarchy with a neat sketch. [7]

6. a) Write the major differences between Apache Pig and SQL. [7]
   b) List and Explain various operators of Pig Latin. [7]

7. a) With a neat diagram explain the key components of Hive architecture. [7]
   b) Explain the available Hive operators to access the elements of Complex Types. [7]

||''|''|''|''||