INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad - 500 043

MODEL QUESTION PAPER - I

B.Tech VII Semester End Examinations (Regular), November – 2019
Regulations: IARE-R16
BIG DATA AND BUSINESS ANALYTICS
(Information Technology)

Time: 3 hours   **Max. Marks: 70**

Answer ONE Question from each Unit
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

UNIT – I

1. a) What is Big Data? Explain how big data processing differs from distributed processing. **[7M]**

   b) List various application of big data. How it can be used to improve business for a superstore. **[7M]**

2. a) Explain characteristics of Big Data in detail and illustrate in which condition data is called by "Big Data". **[7M]**

   b) How Big Data Analytics can be useful in the development of smart cities and explain the landscape of Big Data Technology? **[7M]**

UNIT – II

3. a) What are the advantages of Hadoop? Explain Hadoop architecture and its components with proper diagram. **[7M]**

   b) Write Short note on Hadoop Ecosystem also explain various usecases involved in Hadoop. **[7M]**

4. a) Why to choose Hadoop for processing Big Data in detail and explain the concept of Distributed and parallel computing challenges? **[7M]**

   b) Explain in detail the interacting process with Hadoop Ecosystem? List out various big data processing technologies? **[7M]**

UNIT – III

5. a) Define HDFS? Discuss the HDFS Architecture and HDFS Commands in brief. Write down the goals of HDFS. **[7M]**

   b) How does HDFS ensure data Integrity in a Hadoop Cluster? **[7M]**

6.	a)	Define racks in Hadoop Cluster? Explain how Hadoop Clusters are arranged in several racks with an real time example?	**[7M]**
	b)	Create a file in HDFS, Explain the Anatomy of a File Read and Write?	**[7M]**

UNIT – IV

7.	a)	Explain Map-reduce framework in brief and Draw the architectural diagram for Physical Organization of Compute Nodes	**[7M]**
	b)	Explain working of following phases of Map Reduce with one common example. (i) Map Phase (ii) Combiner Phase (iii) Shuffle and Sort Phase (iv)Reducer Phase	**[7M]**
8.	a)	Describe the working of the MapReduce algorithm?	**[7M]**
	b)	Write Map Reduce code for counting occurrences of specific words in the input text file(s). Also write the commands to compile and run the code.	**[7M]**

UNIT – V

9.	a)	What is Apache Pig and why do we need it and draw the architecture of Apache Pig and explain in breif?	**[7M]**
	b)	Explain Pig data Model in detail and Discuss how it will help for effective data flow.	**[7M]**
10.	a)	Draw and explain Architecture of APACHE HIVE. Explain various data insertion techniques in HIVE with example.	**[7M]**
	b)	What do you mean by HiveQL Data Definition Language?	**[7M]**

COURSE OBJECTIVES:

The course should enable the students to:

I	Optimize business decisions and create competitive advantage with Big data analytics.
II	Understand several key big data technologies used for storage, analysis and manipulation of data.
III	Recognize the key concepts of Hadoop framework, map reduce.
IV	Demonstrate the concepts in Hadoop for application development.

COURSE OUTCOMES:

CO 1	Understand the key issues in big data analytics and its associated applications in business analytics.
CO 2	Illustrate different types of big data technologies in Hadoop parallel world.
CO 3	Interpret disparate data storing in Hadoop Distributed File Systems (HDFS).
CO 4	Explore map reduce framework and optimize its jobs.
CO 5	Explain the basic methodologies of pig and hive.

COURSE LEARNING OUTCOMES:

Students, who complete the course, will have demonstrated the asking to do the following:

ACS012.01	Understand what Big Data, importance and various sources of data.
ACS012.02	Describe the elements of big data-volume, variety, velocity and veracity.
ACS012.03	Understand the importance and challenges of big data.
ACS012.04	Define big data analytics advantages and its applications.
ACS012.05	Define distributed and parallel computing for big data.
ACS012.06	Analyze the core components of hadoop with basic commands.
ACS012.07	Explain the key features of hadoop in processing big data.
ACS012.08	Understand hadoop ecosystem with its animal planet.
ACS012.09	Explain the basic terminology of Hadoop Distributed File Systems (HDFS).
ACS012.10	Describe in detail about Distributed file system.
ACS012.11	Understand the concept of Hadoop cluster architecture.
ACS012.12	Explain a file in HDFS and represent the anatomy of file read and write.
ACS012.13	Understand Map Reduce and its qualities and retain advanced Map Reduce thoughts.
ACS012.14	Understand the architecture of Map Reduce framework.
ACS012.15	Demonstrate the techniques to optimize Map Reduce jobs.
ACS012.16	Understand the typical use occasions of input and output forms of Map Reduce.
ACS012.17	Demonstrate an ability to use frameworks like pig and hive to process Big Data and Analytics.
ACS012.18	Design the architecture of pig with its data types and operations.
ACS012.19	Explain the architecture of hive with different operations.
ACS012.20	Design and implement different technologies for processing big data in pig and hive.

MAPPING OF SEMESTER END EXAMINATION TO COURSE LEARNING OUTCOMES:

SEE Question Number			COURSE LEARNING OUTCOME	Course Outcomes	Blooms Taxonomy Level
1	a	ACS012.01	Understand what Big Data, importance and various sources of data.	CO 1	Understand
	b	ACS012.02	Describe the elements of big data-volume, variety, velocity and veracity.	CO 1	Understand
2	a	ACS012.03	Understand the importance and challenges of big data.	CO 1	Remember
	b	ACS012.04	Define big data analytics advantages and its applications.	CO 1	Remember
3	a	ACS012.05	Define distributed and parallel computing for big data.	CO 2	Understand
	b	ACS012.06	Analyze the core components of hadoop with basic commands.	CO 2	Understand
4	a	ACS012.07	Explain the key features of hadoop in processing big data.	CO 2	Understand
	b	ACS012.08	Understand hadoop ecosystem with its animal planet.	CO 2	Understand
5	a	ACS012.09	Explain the basic terminology of Hadoop Distributed File Systems (HDFS).	CO 3	Understand
	b	ACS012.10	Describe in detail about Distributed file system.	CO 3	Understand
6	a	ACS012.11	Understand the concept of Hadoop cluster architecture	CO 3	Remember
	b	ACS012.12	Explain a file in HDFS and represent the anatomy of file read and write	CO 3	Remember
7	a	ACS012.13	Understand Map Reduce and its qualities and retain advanced Map Reduce thoughts.	CO 4	Understand
	b	ACS012.14	Understand the architecture of Map Reduce framework.	CO 4	Understand
8	a	ACS012.15	Demonstrate the techniques to optimize Map Reduce jobs.	CO 4	Remember
	b	ACS012.16	Understand the typical use occasions of input and output forms of Map Reduce.	CO 4	Understand
9	a	ACS012.17	Demonstrate an ability to use frameworks like pig and hive to process Big Data and Analytics.	CO 5	Remember
	b	ACS012.18	Design the architecture of pig with its data types and operations.	CO 5	Understand
10	a	ACS012.19	Explain the architecture of hive with different	CO 5	Understand
	b	ACS012.20	Design and implement different technologies for processing big data in pig and hive.	CO 5	Remember

HOD, IT