

Hall Ticket No

Question Paper Code: ACS012



# INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad - 500 043

## MODEL QUESTION PAPER - II

B.Tech IV Semester End Examinations (Regular), November – 2019

Regulations: IARE-R16

### BIG DATA AND BUSINESS ANALYTICS

(Computer Science and Engineering)

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 are the benefits of Big Data? Discuss challenges under Big Data. How Big Data Analytics can be useful in the development of smart cities. [7M]
- b) Define the following terms for Big Data: i. Structured data [7M]  
ii. Semi structured data  
iii. Unstructured data.
2. a) What is Big data? Discuss it in terms of four dimensions, volume, velocity, variety and veracity. [7M]
- b) Explain Big Data Analytics and discuss the terminologies used in Big Data Environments? [7M]

#### UNIT – II

3. a) Explain core architecture of Hadoop with suitable block diagram. Discuss role of each component in detail. [7M]
- b) Illustrate the difference between the RDBMS versus Hadoop in detail and justify how Hadoop technology satisfies the business insights now-a-days? [7M]
4. a) Discuss the overview of Hadoop in detail and classify the term Hadoop Cluster with various daemons? [7M]
- b) Explain the various Big Data Hadoop Distributors at present? Describe the overall history of Hadoop technology? [7M]

#### UNIT – III

5. a) Draw and explain HDFS Architecture. Explain the function of NameNode and DataNode. What is a Secondary Namenode? Is it a substitute to the Namenode? [7M]
- b) Explain any two commands of HDFS from following commands with syntax and at least one example of each. (i) copyFromLocal (ii) setrep (iii) checksum. [7M]

6. a) Illustrate the Hadoop cluster is a special type of computational cluster designed for storing and analyzing vast amount of unstructured data in a distributed computing? [7M]  
b) Explain the concept of Blocks and Heartbeat Message in HDFS Architecture. What are the benefits of block transfer? [7M]

#### UNIT – IV

7. a) Explain the features and uses of Map Reduce? Discuss the techniques to optimize Map Reduce jobs? [7M]  
b) Explain “Shuffle & Sort” phase and “Reducer Phase” in MapReduce [7M]
8. a) Write down the differences between Apache Pig and MapReduce with an real time scenario. [7M]  
b) Explain how matrix multiplication is carried out by using the working principles of MapReduce algorithms? [7M]

#### UNIT – V

9. a) Define what are the main reasons for developing Pig Latin and Discuss the two modes used for running the Pig scripts? [7M]  
b) Define are joins? How many types of joins are there in Pig Latin with an examples? [7M]
10. a) Discuss the various Hive services with an examples and Explain the Built-in Functions in Hive? [7M]  
b) Write the Hive command to create a table with four columns: First name, last name, age, and income? [7M]



# INSTITUTE OF AERONAUTICAL ENGINEERING (Autonomous)

## 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 operations.	CO 5	Understand
	b	ACS012.20	Design and implement different technologies for processing big data in pig and hive.	CO 5	Remember

**HOD, CSE**