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

COMPUTER SCIENCE AND ENGINEERING

TUTORIAL QUESTION BANK

Course Title	**BIG DATA AND BUSINESS ANALYTICS**				
Course Code	ACS012				
Programme	B.Tech				
Semester	VII	CSE | IT			
Course Type	Core				
Regulation	**IARE - R16**				
Course Structure		**Theory**			**Practical**	
	Lectures	**Tutorials**	**Credits**	**Laboratory**	**Credits**
	3	1	4	3	2
Chief Coordinator	Ms. G Sulakshana, Assistant Professor				
Course Faculty	Mr. D Rahul, Assistant Professor				

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 (CLOs):

ACS012.01	Understand Define 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.

	UNIT-I			
	INTRODUCTION TO BIG DATA			
	Part - A (Short Answer Questions)			
S No	**QUESTIONS**	**Blooms Taxonomy Level**	**Course Outcomes**	**Course Learning Outcomes (CLOs)**
1	Define data?	Remember	CO1	ACS012.03
2	Define the term information?	Understand	CO1	ACS012.03
3	Define big data?	Remember	CO1	ACS012.03
4	List out various data formats?	Remember	CO1	ACS012.03
5	Explain the life cycle of data?	Remember	CO1	ACS012.04
6	Explain briefly the challenges of big data?	Remember	CO1	ACS012.04
7	Illustrate the importance of big data?	Understand	CO1	ACS012.04
8	List out the characteristics of big data?	Understand	CO1	ACS012.04
9	Differentiate between analysis and analytics?	Understand	CO1	ACS012.04
10	Define big data analytics?	Remember	CO1	ACS012.05
11	List out various big data technologies?	Understand	CO1	ACS012.01
12	What are the various applications of big data analytics?	Understand	CO1	ACS012.01
13	Define streaming data?	Remember	CO1	ACS012.01
14	Define batch data?	Understand	CO1	ACS012.01
15	Define Business Intelligence(BI)	Understand	CO1	ACS012.01
16	List out various terminologies in Big Data environments?	Remember	CO1	ACS012.03
17	Enumerate the terms a. OLAP b. OLTP c. RTAP	Understand	CO1	ACS012.03
18	Illustrate the term Data center?	Understand	CO1	ACS012.04
	Part - B (Long Answer Questions)			
1	Explain in detail about the various formats of data?	Remember	CO1	ACS012.03
2	Describe in detail about the life cycle of data?	Understand	CO1	ACS012.03
3	Explain the importance of big data?	Remember	CO1	ACS012.03
4	List out the various challenges facing big data in detail?	Remember	CO1	ACS012.04
5	List out and discuss the characteristics of big data?	Understand	CO1	ACS012.04
6	Explain in detail about the difference between Traditional Vs Big data business approach?	Understand	CO1	ACS012.04
7	Define what is the difference between analysis and analytics. Discuss the importance of big data analytics?	Remember	CO1	ACS012.05
8	Explain the current big data technologies in detail?	Understand	CO1	ACS012.01
9	Describe the various applications of big data analytics?	Understand	CO1	ACS012.01
10	Explain in detail about the ETL (Extract, Transform and Load) system?	Remember	CO1	ACS012.01
11	State the various drawbacks of using Traditional system approach?	Understand	CO1	ACS012.01
12	Classify the Difference between Traditional Business Intelligence BI) versus Big Data?	Understand	CO1	ACS012.01
13	Explain in detail the classification of Big Data Analytics from Analysis?	Remember	CO1	ACS012.03
14	Elaborate the different terminologies in Big Data environments?	Understand	CO1	ACS012.03
15	Describe the Big Data Technology Landscape?	Understand	CO1	ACS012.04
	Part - C (Problem Solving and Critical Thinking Questions)			
1	Explain about the challenges facing in processing big data now a days?	Understand	CO1	ACS012.04
2	Explain in detail about the case study of big data solutions?	Remember	CO1	ACS012.05
3	Describe Traditional Vs Big data business approach with its drawbacks?	Understand	CO1	ACS012.01
4	Express the various formats of data and illustrate with an real time examples?	Understand	CO1	ACS012.01

5	Extrapolate big data analytics and explain various applications in the real world scenario?	Remember	CO1	ACS012.01
6	List and discuss the elements of big data?	Understand	CO1	ACS012.01
	UNIT-II			
	INTRODUCTION TO HADOOP			
	Part – A (Short Answer Questions)			
1	List out various big data processing technologies?	Understand	CO 2	ACS012.08
2	Justify Why Big data is a problem?	Understand	CO 2	ACS012.07
3	Express the term bucketing data?	Understand	CO 2	ACS012.07
4	Discuss Why Hadoop came into an existence in processing big data?	Understand	CO 2	ACS012.08
5	Define Hadoop?	Remember	CO 2	ACS012.07
6	Classify the term Hadoop Cluster?	Understand	CO 2	ACS012.08
7	Demonstrate divide and conquer philosophy in Hadoop Cluster?	Remember	CO 2	ACS012.07
8	State the core components of Hadoop?	Understand	CO 2	ACS012.08
9	Define RDBMS(Relational Data Base Management System)?	Understand	CO 2	ACS012.08
10	Explain the Distributed Computing challenges?	Understand	CO 2	ACS012.07
11	List out the various use cases of Hadoop?	Remember	CO 2	ACS012.07
12	Illustrate Hadoop Distributors for processing Big Data?	Remember	CO 2	ACS012.07
13	Distinguish the Hadoop Ecosystem?	Understand	CO 2	ACS012.07
	Part - B (Long Answer Questions)			
1	Explain various big data processing technologies at present?	Understand	CO 2	ACS012.07
2	Explain why do companies store data? why we need data to gather?	Understand	CO 2	ACS012.07
3	Explain the various benefits of data analysis and analytics?	Understand	CO 2	ACS012.07
4	Illustrate the overview of Hadoop in detail?	Understand	CO 2	ACS012.07
5	Describe the overall history of Hadoop technology?	Understand	CO 2	ACS012.07
6	Explain in detail about the frame work of Hadoop?	Understand	CO 2	ACS012.07
7	Explain Define are the core components of Hadoop environment?	Understand	CO 2	ACS012.08
8	Describe the concept of Distributed and parallel computing challenges?	Understand	CO 2	ACS012.08
9	Explain the use cases of Hadoop in detail?	Understand	CO 2	ACS012.07
10	Explain the various big data Hadoop Distributors at present?	Understand	CO 2	ACS012.08
11	Implement the processing data with Hadoop?	Understand	CO 2	ACS012.07
12	Explain in detail interacting process with Hadoop Ecosystem?	Understand	CO 2	ACS012.07
	Part - C (Problem Solving and Critical Thinking Questions)			
1	Why to choose Hadoop for processing big data in detail?	Understand	CO 2	ACS012.07
2	Justify how hadoop technology satisfies the business insights now -a –days?	Understand	CO 2	ACS012.07
3	Explain divide and conquer philosophy in processing big data – Hadoop technology?	Understand	CO 2	ACS012.07
4	Illustrate the difference between the RDBMS versus Hadoop in detail?	Understand	CO 2	ACS012.07
5	Describe the architecture of Hadoop Technology?	Understand	CO 2	ACS012.07
	UNIT-III			
	THE HADOOP DISTRIBUTED FILE SYSTEM			
	Part - A (Short Answer Questions)			
1	Discuss the term HDFS in Hadoop environment?	Remember	CO 3	ACS012.09
2	Define Filesystem?	Remember	CO 3	ACS012.10
3	List out the basic Filesystem Operations?	Understand	CO 3	ACS012.10
4	Implement the Master-Slave architecture?	Remember	CO 3	ACS012.12
5	Illustrate what is commodity hardware?	Remember	CO 3	ACS012.13

6	Define what is block in HDFS?	Understand	CO 3	ACS012.12
7	Why is a block in HDFS so large?	Understand	CO 3	ACS012.13
8	Identify block replication in HDFS?	Remember	CO 3	ACS012.12
9	Express Throughput?	Understand	CO 3	ACS012.12
10	State how HDFS stores massive data in Hadoop Cluster?	Understand	CO 3	ACS012.13
11	Define racks in Hadoop Cluster?	Understand	CO 3	ACS012.10
12	List out the real time examples for Hadoop Cluster?	Remember	CO 3	ACS012.12
13	Extrapolate the Master components: Name node, Secondary Node and JobTracker?	Remember	CO 3	ACS012.13
14	Extrapolate the Slave components :Data Node and Task Tracker?	Understand	CO 3	ACS012.12
15	Define what is Daemons in Hadoop Cluster ?	Understand	CO 3	ACS012.13
Part – B (Long Answer Questions)				
1	Explain the design of HDFS?	Understand	CO 3	ACS012.12
2	Elaborate the basic concepts of HDFS architecture in detail?	Understand	CO 3	ACS012.13
3	Explain the Hadoop Cluster in processing big data environment?	Understand	CO 3	ACS012.12
4	Describe when not to use Hadoop setup in recent scenario?	Understand	CO 3	ACS012.13
5	Explain Hadoop Clusters are arranged in several racks with an real time example?	Understand	CO 3	ACS012.12
11	Ellaborate the Hadoop Cluster architecture?	Understand	CO 3	ACS012.12
12	Derive the core components of Hadoop Cluster?	Understand	CO 3	ACS012.13
13	Explain in detail the Master and Slave components of Hadoop Cluster?	Understand	CO 3	ACS012.12
14	Create a file in HDFS, Explain the Anatomy of a File Read and Write?	Understand	CO 3	ACS012.13
15	Explain in detail the Hadoop Filesystem?	Understand	CO 3	ACS012.12
Part – C (Problem Solving and Critical Thinking)				
1	Illuatrate the Hadoop cluster is a special type of computational cluster designed for storing and analyzing vast amount of unstructured data in a distributed computing?	Understand	CO 3	ACS012.12
2	Explain the Hadoop - Typical Workflow in HDFS?	Understand	CO 3	ACS012.13
3	Elaborate the Hadoop Daemon properties with real time scenario?	Understand	CO 3	ACS012.12
4	Explain the Hadoop processing of data in Cloud computing and AmazonEC2 with an examples?	Understand	CO 3	ACS012.13
UNIT-IV				
UNDERSTANDING MAP REDUCE FUNDAMENTALS				
Part – A (Short Answer Questions)				
1	Define what is MapReduce?	Understand	CO4	ACS012.14
2	Explain about Data Locality in MapReduce?	Understand	CO4	ACS012.17
3	How to overcome the Faults and handling of Errors?	Remember	CO4	ACS012.14
4	How to explore the Scale-out architecture?	Remember	CO4	ACS012.14
5	Discuss the term List Processing?	Understand	CO4	ACS012.14
6	List out the real time examples of MapReduce?	Understand	CO4	ACS012.14
7	Define what happens if the mapper output does not match the reducer input?	Understand	CO4	ACS012.14
8	Define Input Split?	Understand	CO4	ACS012.14
9	Generalise the term Record Reader/Writer?	Remember	CO4	ACS012.15
10	Design Map Phase?	Understand	CO4	ACS012.17
11	Design Reducer Phase?	Remember	CO4	ACS012.14
12	Define Partitioners?	Remember	CO4	ACS012.15
13	Explain the term YARN?	Remember	CO4	ACS012.14

14	Define Container?	Remember	CO4	ACS012.14
15	Define Combiner?	Understand	CO4	ACS012.14
	Part – B (Long Answer Questions)			
1	Explain in detail the framework of MapReduce?	Remember	CO4	ACS012.15
2	Explain the working principle of MapReduce?	Remember	CO4	ACS012.15
3	Estimate the entire process of data analysis conducted in the MapReduce programming model?	Understand	CO4	ACS012.14
4	Explain the description of MapReduce process for a specific case?	Remember	CO4	ACS012.14
5	Describe the uses of MapReduce?	Remember	CO4	ACS012.14
6	Define what conditions must be met to implement MapReduce application?	Understand	CO4	ACS012.16
7	Can MapReduce be used to solve any kind of computational problems? if not, explain the cases where MapReduce is not applicable?	Understand	CO4	ACS012.16
8	Implement the Input Format for Compute-Intensive applications?	Remember	CO4	ACS012.14
9	Write a short note on Input Split?	Understand	CO4	ACS012.16
10	Explain the types of MapReduce applications?	Understand	CO4	ACS012.14
11	Write a short on the FileInputFormat class?	Understand	CO4	ACS012.14
	Part – C (Problem Solving and Critical Thinking)			
1	List the main features of MapReduce?	Remember	CO4	ACS012.14
2	Describe the working of the MapReduce algorithm?	Understand	CO4	ACS012.15
3	Discuss some techniques to optimize MapReduce jobs?	Remember	CO4	ACS012.14
4	Discuss the points you need to consider while designing a file system in MapReduce?	Understand	CO4	ACS012.14
5	Write a short note on Input Format?	Understand	CO4	ACS012.16
	UNIT-V			
	INTRODUCTION TO PIG AND HIVE			
	Part - A (Short Answer Questions)			
1.	Breifly explain the architecture of Pig?	Understand	CO5	ACS012.14
2.	List out the benefits of Pig?	Remember	CO5	ACS012.16
3.	Define what are the properties of Pig?	Understand	CO5	ACS012.16
4.	Explain about schema?	Remember	CO5	ACS012.15
5.	Classify Pig Latin commands in Pig?	Understand	CO5	ACS012.16
6.	Explain the Pig Latin application flow?	Understand	CO5	ACS012.15
7.	Define UDF?	Understand	CO5	ACS012.16
8.	Discuss the modes of Pig scripts?	Understand	CO5	ACS012.17
9.	Define Hive?	Remember	CO5	ACS012.18
10.	Express DDL concepts in detail?	Understand	CO5	ACS012.18
11.	In Hive, explain the term 'aggregation' and its uses?	Understand	CO5	ACS012.18
12.	List out the Data types in Hive?	Remember	CO5	ACS012.19
13.	Interpret joins with an examples?	Understand	CO5	ACS012.17
	Part - B (Long Answer Questions)			
1	Discuss the two modes used for running the Pig scripts?	Remember	C05	ACS012.16
2	Define what are the main reasons for developing Pig Latin?	Remember	C05	ACS012.17
3	Define what do you understand by Pig Latin application flow?	Understand	C05	ACS012.18
4	Discuss the use of the FOREACH and ASSERT operator in Pig Latin?	Understand	C05	ACS012.16
5	Define the various Statements used in flow of data processing in Pig Latin?	Understand	C05	ACS012.16
6	Discuss the use of the FILTER and DISTINCT operator in Pig Latin	Understand	C05	ACS012.17
7	Define what do you understand by introducing Hive?	Understand	C05	ACS012.18
8	Discuss the various Hive services with an examples?	Understand	C05	ACS012.19
9	Describe the various Hive Data types?	Understand	C05	ACS012.20

10.	Explain the Built-in Functions in Hive?	Understand	CO5	ACS012.19
	Part – C (Problem Solving and Critical Thinking)			
1	Write a short note on the following operators: a. GROUP b. ORDER BY	Remember	C05	ACS012.17
2	Define are joins? How many types of joins are there in Pig Latin with examples?	Understand	C05	ACS012.19
3	Write the Hive command to create a table with four columns: First name, last name, age, and income?	Understand	C05	ACS012.18
4	A start-up company wants to use Hive for storing its data. List the collection types provided by Hive for this purpose? Write a shell command in Hive to list all the files in the current directory?	Understand	C05	ACS012.20
5	Write a shell command in Hive to list all the files in the current directory?	Understand	C05	ACS012.20

Prepared by:

Ms. G Sulakshana, Assistant Professor   **HOD, IT**