

# SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(COMMON TO INFORMATION TECHNOLOGY)

### QUESTION BANK



VII SEMESTER

**1904011 – BIG DATA ANALYTICS**

**Regulation – 2019**

**Academic Year 2025 – 2026 ODD Semester**

*Prepared by*

**Dr.V. Dhanakoti, Professor /CSE**

**Ms.A.Lalitha, Assistant Professor (Sel.G) /CSE**

## UNIT 1 INTRODUCTION TO BIG DATA

**Big Data – Definition, Characteristic Features – Big Data Applications - Big Data vs Traditional Data - Risks of Big Data - Structure of Big Data - Challenges of Conventional Systems - Web Data – Evolution of Analytic Scalability - Evolution of Analytic Processes, Tools and methods - Analysis vs Reporting - Modern Data Analytic Tools**

### PART – A

Q.No	Question	Competence	Level
1	What is Big Data?	Remember	BTL 1
2	Differentiate Big Data and Conventional Data.	Understand	BTL 2
3	List the various dimensions of growth of Big Data.	Remember	BTL 1
4	List the main characteristics of Big Data.	Remember	BTL 1
5	Write the risk of big data.	Remember	BTL 1
6	What is web data?	Remember	BTL 1
7	List the sources of big data.	Remember	BTL 1
8	What are the challenges in big data?	Remember	BTL 1
9	Why domain expertise is required for any type of Data Analytics?	Remember	BTL 1
10	Give reason: “Web Data is the most popular Big Data”.	Understand	BTL 2
11	Give reason “Accuracy in big data is beneficial”	Understand	BTL 2
12	Give the structure of big data.	Understand	BTL 2
13	Give the list of big data applications.	Understand	BTL 2
14	List the challenges of convectional system.	Remember	BTL 1
15	How to simplify the role of analytical scalability in big data?	Understand	BTL 2
16	Give the structure of big data.	Understand	BTL 2
17	How big data can be represented?	Understand	BTL 2
18	State the importance of analysis vs reporting.	Remember	BTL 1
19	What are various technologies used to handle big data?	Understand	BTL 2
20	What is a sand box?	Remember	BTL 1
21	Give the traditional analytical architecture.	Understand	BTL 2
22	List the types of Data.	Remember	BTL 1
23	Differentiate data warehouse architecture vs MPP Architecture.	Understand	BTL 2

24	What do you mean by the data privacy?	Understand	BTL 2
----	---------------------------------------	------------	-------

<b>PART-B</b>				
Q.No.	Question		Competence	Level
1	<b>Analyze</b> What is Big data? <b>Describe</b> the main features of a big data in detail.	13	Analyze	BTL 4
2	i. Explain the main characteristics of big data. ii. <b>Describe</b> big data architecture with a neat schematic diagram.	4 9	Analyze	BTL 4
3	Explain in detail various risks in handling big data.	13	Analyze	BTL 4
4	Illustrate the structure of big data representation with neat sketch.	13	Apply	BTL 3
5	i. <b>Point out</b> the features of Massive parallel processing system. ii. <b>Explain</b> the use of Massive Parallel Processing system in big data analytics	5 8	Analyze	BTL 4
6	<b>Analyze</b> the challenges faced by traditional system.	13	Analyze	BTL 4
7	Illustrate with an example the various analysis tools and reporting tools used in Big-data.	13	Apply	BTL 3
8	i. <b>What</b> is an analytical data set? ii. <b>Explain</b> the types of analytical data set.	3 10	Analyze	BTL 4
9	i. <b>Summarize</b> what is web data. ii. <b>Discuss</b> what the web data reveals.	6 7	Analyze	BTL 4
10	i. <b>Describe</b> how big data are effectively filtered. . ii. <b>Describe</b> how big data are mixed with traditional one.	6 7	Analyze	BTL 4
11	<b>Illustrate the Evolution</b> Tools and Method in big data with an example.	13	Apply	BTL 3
12	i. <b>Assess</b> the difficulties faced by conventional systems. ii. <b>Explain</b> the differences between big data from the traditional one.	5 8	Apply	BTL 3
13	<b>Illustrate</b> how the analytical scalability is handled in big data with an example.	13	Apply	BTL 3
14	i. <b>Pointout</b> some of the web data in current action today. ii. <b>Illustrate</b> the modern tools for big data analysis with an example	6 7	Analyze	BTL 4
15	<b>Illustrate</b> the Analysis and reporting methods and tools <b>with an example.</b>	13	Apply	BTL 3
16	<b>Illustrate the importance of</b> analytical sandbox in detail <b>with an example</b>	13	Apply	BTL 3
17	<b>Illustrate</b> the Evolution of Analytical Scalability.	13	Apply	BTL 3
<b>PART C</b>				
1.	<b>Generalize</b> in detail about the challenges of the Big data	15	Create	BTL 6
2.	<b>Justify</b> the Statement “Web Data is the Most Popular Big Data” with reference to data analytic professional.	15	Evaluate	BTL 5
3.	<b>Comment</b> on the statement “Is the “Big” Part or the “Data” Part More Important “ in the term big data.	15	Evaluate	BTL 5

4.	<b>Develop</b> the role of Analytic Sandbox and its benefits in the Analytic Process.	15	Create	BTL 6
5.	<b>Explain the</b> evolution of Analytical Scalability.	15	Evaluate	BTL 5

## UNIT II HADOOP FRAMEWORK

Distributed File Systems - Large-Scale Filesystem Organization – HDFS concepts - MapReduce Execution, Algorithms using MapReduce, Matrix-Vector Multiplication – Hadoop YARN

### PART – A

Q.No	Question	Competence	Level
1	<b>What</b> is Hadoop?	Remember	BTL 1
2	How Map-Reduce computation executes?	Remember	BTL 1
3	List the key advantages in Hadoop.	Remember	BTL 1
4	<b>What</b> is Hadoop YARN?	Understand	BTL 2
5	<b>List</b> the core concepts of HADOOP.	Remember	BTL 1
6	<b>Define</b> MAP REDUCE concepts.	Remember	BTL 1
7	How a key value pair is formed?	Understand	BTL 2
8	State the importance of DFS.	Understand	BTL 2
9	<b>What</b> is HDFS?	Remember	BTL 1
10	What are the features of HDFS?	Remember	BTL 1
11	What is name node?	Understand	BTL 2
12	<b>What are</b> the goals of HDFS?	Remember	BTL 1
13	What is data node?	Understand	BTL 2
14	<b>What is</b> shuffle and sort algorithm?	Understand	BTL 2
15	<b>What are</b> the Advantages of HDFS?	Understand	BTL 2
16	<b>List</b> out the Hadoop applications.	Remember	BTL 1
17	<b>What</b> is matrix multiplication?	Understand	BTL 2
18	Why the partitions are shuffled in map reduce?	Understand	BTL 2
19	List the steps in MapReduce algorithm.	Remember	BTL 1
20	List the role of application manager in Hadoop.	Remember	BTL 1
21	List the application of distributed system.	Remember	BTL 1
22	<b>State the importance of</b> resource manager in Hadoop.	Understand	BTL 2
23	<b>What</b> is YARN?	Understand	BTL 2
24	<b>What are</b> the advantages of Hadoop?	Understand	BTL 2

### PART-B

Q.No.	Question	Marks	Competence	Level
1	Explain the features of Hadoop and explain the functionalities of Hadoop?	13	Analyze	BTL 4
2	<b>Discuss</b> the various core components of the Hadoop.	13	Analyze	BTL 4
3	<b>Explain</b> about Hadoop distributed file system architecture with neat diagram.	13	Analyze	BTL 4
4	<b>Summarize</b> briefly on i. Algorithms using MapReduce. ii. Advantages of MapReduce.	8 5	Analyze	BTL 4
5	<b>Compare</b> and Contrast the Hadoop and MapReduce.	13	Analyze	BTL 4
6	<b>Analyze</b> the steps of Map Reduce Algorithms.	13	Analyze	BTL 4
7	<b>Describe</b> the concepts of HDFS.	13	Apply	BTL 3
8	i. <b>Identify the</b> map and reduce algorithm in detail. ii. <b>Illustrate</b> the map reduce algorithm with an example.	6 7	Apply	BTL 3
9	Explain the phases in map reduce with an example.	13	Analyze	BTL 4
10	<b>Illustrate</b> the Apache Hadoop YARN architecture with neat sketch.	13	Apply	BTL 3
11	i. <b>Describe</b> Map Reduce framework in detail. Draw the architectural diagram ii. <b>Define</b> HDFS. Explain HDFS in detail.	13	Analyze	BTL 4
12	<b>Illustrate the</b> Matrix Vector Multiplication in detail <b>with an example.</b>	13	Apply	BTL 3
13	<b>i. Explain</b> what YARN is. ii. <b>Illustrate</b> HADOOP YARN architecture with neat diagram	6 7	Apply	BTL 3
14	Use map reduce architecture to illustrate the concept of Hadoop for the following example Welcome to Hadoop class Hadoop is interesting Hadoop is useful Hadoop is useful in big data	13	Apply	BTL 3
15	Illustrate the distributed file system <b>with neat sketch.</b>	13	Analyze	BTL 4
16	<b>Discuss</b> the benefits of using HDFS. <b>Discuss</b> the benefits of HDFS in marketing application.	8 5	Analyze	BTL 4
17	<b>Illustrate the HDFS</b> data replication with an example.	13	Apply	BTL 3
<b>PART C</b>				
1	<b>Recommend</b> a procedure to find the number of Occurrence of a word in a document.	15	Create	BTL 5
2	<b>i. Generalize</b> with a neat sketch about processing of a job in Hadoop. ii. <b>List</b> the various operational modes of Hadoop cluster configuration and explain in detail about configuring/installing the Hadoop in local/standalone mode	15	Evaluate	BTL 6

3	<b>Summarize</b> how google file system differs from the Hadoop file system and explain the google file system architecture with a neat sketch.	15	Create	BTL 5
4	<b>Prepare</b> a Map-Reduce Algorithm to get the Dot Product of two Large Vectors. Assuming Only non-zero elements of those vectors are given in input files and output file should show only non-zero entries( assuming two vectors are same size) ex: v1=[ 5 4 0 1 2] v2=[ 4 2 1 0 6] file1:                      file2:                      output: (0,5)                      (0,4)                      (0,20) (1,4)                      (1,2)                      (1,8) (3,1)                      (2,1)                      (4,12) (4,2)                      (4,6)	15	Evaluate	BTL 6
5	<b>Consider</b> a collection of literature survey made by a researcher in the form of a text document with respect to cloud and big data analytics. Using Hadoop and Map Reduce, write a program to count the occurrence of pre dominant key words.	15	Evaluate	BTL 6

### UNIT III - DATA ANALYSIS

Statistical Methods: Regression modelling, Multivariate Analysis - Classification: SVM & Kernel Methods - Rule Mining - Cluster Analysis, Types of Data in Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density Based Methods, Grid Based Methods, Model Based Clustering Methods, Clustering High Dimensional Data - Predictive Analytics – Data analysis using R.

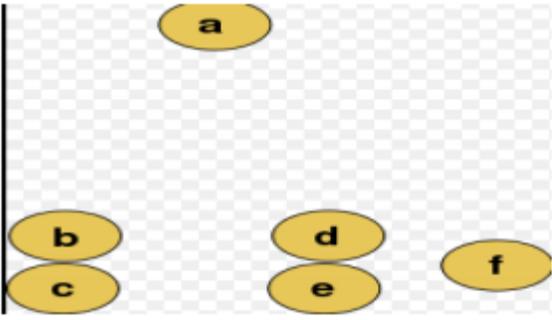
#### PART – A

Q.No.	Question	Competence	Level
1	<b>What</b> is data analysis?	Remember	BTL 1
2	<b>What</b> is classification?	Remember	BTL 1
3	<b>State</b> support-vector machines.	Remember	BTL 1
4	<b>What</b> is regression?	Understand	BTL 2
5	<b>List</b> out the different types of regression	Remember	BTL 1
6	<b>Define</b> multivariate analysis.	Remember	BTL 1
7	<b>Classify</b> different types of data analysis.	Understand	BTL 2
8	<b>State</b> the importance of classification in data analysis.	Remember	BTL 1
9	<b>What</b> is kernel?	Remember	BTL 1
10	<b>What</b> is rule mining?	Remember	BTL 1
11	<b>Give</b> the types of kernels.	Understand	BTL 2
12	<b>What</b> is Multiple Linear Regression?	Remember	BTL 1
13	<b>What</b> is predictive analysis?	Remember	BTL 1
14	<b>Differentiate</b> regression and correlation.	Understand	BTL 2
15	<b>What</b> is clustering?	Understand	BTL 2
16	<b>List</b> the types of clustering.	Remember	BTL 1
17	<b>What</b> is SVM?	Understand	BTL 2
18	<b>What do you mean by</b> classification and clustering?	Understand	BTL 2

19	State the importance of clustering in data analysis.	Understand	BTL 2
20	What is R Language?	Remember	BTL 1
21	Define density based clustering.	Understand	BTL 2
22	What is the grid based clustering?	Understand	BTL 2
23	List the types of hierarchical clustering.	Understand	BTL 2
24	Classify the partitioning methods in clustering.	Understand	BTL 2

### PART-B

Q.No	Question	Marks	Competence	Level
1	i. What is regression? List the types of regression.	5	Analyze	BTL 4
	ii. Explain the purpose of using Regression Modeling in Data Analysis.	8		
2	i. Assess when we use multivariate analysis.	5	Analyze	BTL 4
	ii. Explain in detail about the various Multivariate Analysis Techniques with examples	8		
3	i. Describe SVM in detail.	7	Analyze	BTL 4
	ii. List out and explain some of the applications of SVM in detail	6		
4	Explain about kernel methods in detail.	13	Analyze	BTL 4
5	List and Explain on types of data in clustering and its Importance.	13	Analyze	BTL 4
6	i. Discuss in detail about the rule mining.	6	Analyze	BTL 4
	ii. Explain in detail about association rule mining	7		
7	i. Examine clustering in data analysis.	3	Apply	BTL 3
	ii. Illustrate density based and Grid based clustering in detail.	10		
8	Describe how clustering is used in high dimensional data.	13	Analyze	BTL 4
9	Illustrate the approaches of clustering.	13	Apply	BTL 3
10	Discuss model based clustering and high dimensional clustering in detail.	13	Analyze	BTL 4
11	Describe partitioning method of clustering in detail.	13	Analyze	BTL 4
12	Illustrate k-means clustering algorithm with an example.	13	Apply	BTL 3
13	i. What is prediction? Depict how prediction helps in data analysis.	5	Apply	BTL 3
	ii. Explain in detail about predictive analysis.	8		
14	Illustrate the different hierarchical clustering techniques with an example.	13	Apply	BTL 3
15	Illustrate the model based clustering with an example.	13	Apply	BTL 3
16	Illustrate the density based clustering with a neat diagram.	13	Apply	BTL 3
17	Illustrate grid based clustering in detail with neat sketch	13	Apply	BTL 3
<b>PART – C</b>				
1	Comment the statement in detail: “Data Analysis is not a decision-making system, but a decision-supporting system”.	15	Analyze	BTL 6
2	Create a Regression Model for “happy people get many hours of sleep” using your own data and what kind of inferences it provides.	15	Create	BTL 6
3	Summarize hierarchical clustering in detail. Analyse the given	15	Evaluate	BTL 5

	<p>diagram and draw the dendrogram using hierarchical clustering algorithm .</p> 			
4	<p><b>Compose</b> the K-means partitioning algorithm using the given data. Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)</p>	15	Create	BTL 6
5	<p><b>Summarize</b> predictive analysis with some application.</p>	15	Evaluate	BTL 5

#### UNIT IV MINING DATA STREAMS

Streams: Concepts – Stream Data Model and Architecture - Sampling data in a stream - Mining Data Streams and Mining Time-series data - Real Time Analytics Platform (RTAP) Applications - Case Studies - Real Time Sentiment Analysis, Stock Market Predictions

#### PART – A

Q.No.	Question	Competence	Level
1	<b>List</b> the main characteristics of stream sources.	Remember	BTL 1
2	<b>What</b> is a data stream?	Remember	BTL 1
3	Why data stream management is relevant in data mining?	Understand	BTL 2
4	<b>Define</b> what is meant by data stream.	Remember	BTL 1
5	<b>What</b> is Sampling data in a stream?	Remember	BTL 1
6	<b>List</b> out the few challenges of data stream mining algorithms.	Remember	BTL 1
7	<b>Differentiate</b> between DBMS and DSMS.	Understand	BTL 2
8	State “Filtering a Data Stream”.	Remember	BTL 1
9	<b>Give</b> the applications of DSMS.	Understand	BTL 2
10	<b>What</b> is Real-Time Analysis?	Understand	BTL 2
11	How to deal with infinite streams?	Remember	BTL 1
12	<b>What</b> is a Data Stream Management System?	Remember	BTL 1
13	<b>List</b> examples you can find for stream sources.	Remember	BTL 1
14	<b>What</b> is called Data Stream Mining?	Understand	BTL 2
15	<b>Compare</b> and contrast RTAP (real time analytics platform) and RTSA (real time sentiment analysis)?	Understand	BTL 2
16	What is the need of RTAP?	Understand	BTL 2
17	What is Real Time Analytics Platform?	Understand	BTL 2

18	<b>Define</b> forecasting.	Understand	BTL 2
19	<b>What is</b> real time data?	Understand	BTL 2
20	What information is used to substitute the view of streams over databases?	Remember	BTL 1
21	<b>State the</b> importance of social media analytics.	Remember	BTL 1
22	<b>Give</b> the reasons why do we need Real Time Analytics Platform?	Remember	BTL 1
23	<b>Define</b> prediction and forecasting.	Understand	BTL 2
24	<b>Illustrate</b> what is time series data.	Understand	BTL 2

### PART-B

Q.No.	Question	Marks	Competence	Level
1	i. <b>Define</b> data stream.	3	Analyze	BTL 4
	ii. <b>Describe</b> the Data Stream model with a neat architecture diagram.	10		
2	<b>Illustrate</b> briefly about the sources of data stream.	13	Apply	BTL 3
3	Based on what you know, <b>Analyze</b> how would you partition the following bit stream into buckets 1001011011101? Find all of them?	13	Analyze	BTL 4
4	<b>Explain</b> the stream data model and its architecture.	13	Analyze	BTL 4
5	<b>Analyze</b> and write a short note on Aurora system model.	13	Analyze	BTL 4
6	i. <b>Explain</b> Sampling in Data Streams.	5	Analyze	BTL 4
	ii. <b>Explain</b> the sampling types in detail	8		
7	<b>Describe</b> about Aurora query model.	13	Analyze	BTL 4
8	<b>Illustrate</b> how mining is done with data streams with an example.	13	Apply	BTL 3
9	i. <b>Describe</b> briefly how to count the distinct elements in a stream.	9	Analyze	BTL 4
	ii. <b>What</b> do you meant by count–distinct problem?	4		
10	Explain in detail:	7	Analyze	BTL 4
	i. Sliding window concept ii. Land mark window concept	6		
11	<b>Illustrate</b> how you would describe the various windowing approach to data stream mining.	13	Apply	BTL 3
12	i. <b>List</b> the methods for analyzing time series data.	7	Analyze	BTL 4
	ii. Explain the several types of motivation and data analysis available for time series.	6		
13	i. <b>Illustrate</b> what approaches are used to estimate the moments.	8	Apply	BTL 3
	ii. <b>Examine</b> the function cost of exact counts.	5		
14	i. <b>Evaluate</b> what is real time sentiment analysis.	5	Apply	BTL 3
	ii. <b>Apply</b> the mining concept in real time sentiment analysis and explain in detail.	8		
15	<b>Illustrate</b> can you identify the following? Suppose our stream consists of the integers 3, 1, 4, 1, 5, 9, 2, 6, 5. Our hash functions will all be of the form $h(x) = ax+b \text{ mod } 32$ for some a and b. You should treat the result as a 5-bit binary integer. Determine the tail length for each stream element and the resulting estimate of the number of distinct elements if the	13	Apply	BTL 3

	hash function is: (a) $h(x) = 2x + 1 \pmod{32}$ . (b) $h(x) = 3x + 7 \pmod{32}$ . (c) $h(x) = 4x \pmod{32}$			
16	i. <b>Express</b> what bloom filters are. ii. Illustrate the relevance of bloom filters in data mining with an example.	3 10	Apply	BTL 3
17	Illustrate how data analysis is used in i. Stock market predictions. ii. Weather forecasting predictions.	7 6	Apply	BTL 3
<b>PART – C</b>				
1	<b>Evaluate</b> the process of Data Stream Mining with suitable examples.	15	Evaluate	BTL 5
2	Summarize data streaming algorithms in detail. <b>Evaluate</b> key stream mining problems and discuss the challenges associated with each problem.	15	Evaluate	BTL 5
3	<b>Generalize</b> data stream management systems in detail.	15	Create	BTL 6
4	<b>Prepare</b> a generic design for Real-time Analytics Platform (RTAP). Discuss your answer related to real time sentiment analysis.	15	Create	BTL 6
5	<b>Evaluate</b> the Bloom Filter in detail with an algorithm. Apply this bloom filter algorithm in Adhar card( Unique Identification number)	15	Evaluate	BTL 5

### UNIT V BIG DATA FRAMEWORKS

Introduction to NoSQL – Aggregate Data Models – Hbase: Data Model and Implementations – Hbase Clients – Examples – .Cassandra: Data Model – Examples – Cassandra Clients – Hadoop Integration. Pig – Grunt – Pig Data Model – Pig Latin – developing and testing Pig Latin scripts. Hive – Data Types and File Formats – HiveQL Data Definition – HiveQL Data Manipulation – HiveQL Queries

#### PART – A

Q.No.	Question	Competence	Level
1	<b>What</b> is NoSQL?	Remember	BTL 1
2	Why do we need NoSQL?	Understand	BTL 2
3	<b>List</b> the components of Hadoop framework.	Remember	BTL 1
4	<b>Classify</b> the categories of NoSQL.	Understand	BTL 2
5	<b>What</b> is the advantage of NoSQL?	Remember	BTL 1
6	<b>Give</b> the disadvantages of NoSQL.	Understand	BTL 2
7	<b>What</b> is HBase?	Remember	BTL 1
8	List the advantage of Cassandra.	Remember	BTL 1
9	Who is <b>generating</b> big data and what are the ecosystem projects used for processing?	Remember	BTL 1
10	Difference between HBase and Hive.	Understand	BTL 2
11	<b>List</b> the aggregate data models.	Remember	BTL 1
12	What is Pig in Hadoop?	Understand	BTL 2
13	<b>What</b> is Apache pig?	Remember	BTL 1
14	Difference between Pig and Hive.	Understand	BTL 2
15	<b>Classify</b> the usage of Pig, Hive and HBase.	Understand	BTL 2
16	<b>Give</b> the features of Hive.	Understand	BTL 2

17	<b>Define</b> Pig.	Remember	BTL 1
18	<b>What</b> is hive in Big Data?	Remember	BTL 1
19	What is Cassandra Client?	Understand	BTL 2
20	<b>List</b> out the types of built-in operator in HIVE.	Remember	BTL 1
21	Differentiate Row-Oriented vs. Column-Oriented data stores	Understand	BTL 2
22	<b>Define</b> The Role Of Hmaster in Hbase.	Understand	BTL 2
23	<b>State</b> the importance of Cassandra.	Remember	BTL 1
24	<b>Give</b> the types of data modeling.	Understand	BTL 2

<b>PART-B</b>				
Q.No.	Question	Marks	Competence	Level
1	i. <b>Describe</b> what is NoSQL.	7	Analyze	BTL 4
	i. <b>List</b> the advantages and disadvantages of NoSQL.	6		
2	i. <b>Illustrate</b> in detail about Hive data manipulation, queries, and data types.	8	Apply	BTL 3
	ii. <b>Illustrate</b> data definition in Hive.	5		
3	<b>Describe</b> the system architecture and components of Hive and Hadoop.	13	Analyze	BTL 4
4	<b>Explain</b> briefly on aggregate data models.	13	Analyze	BTL 4
5	<b>Describe</b> Pig in detail.	13	Analyze	BTL 4
6	i. <b>Describe</b> about HBase in detail.	7	Analyze	BTL 4
	ii. Explain Hbase clients in detail.	6		
7	i. <b>Analyze</b> how Cassandra is integrated with Hadoop.	6	Apply	BTL 3
	ii. <b>Explain</b> the tools related to Hadoop.	7		
8	<b>Illustrate</b> Hbase architecture with neat diagram.	13	Apply	BTL 3
9	Explain in detail:		Analyze	BTL 4
	i. Conceptual data modeling.	4		
	ii. Logical data modeling.	4		
	iii. Physical data modeling	5		
10	<b>Discuss</b> about Cassandra clients.	13	Analyze	BTL 4
11	<b>Compare</b> and Contrast the Hbase and Hive.	13	Analyze	BTL 4
12	i. <b>Explain</b> about Hive in detail	7	Analyze	BTL 4
	ii. <b>What</b> is invoking a Grunt shell?	6		
13	<b>Illustrate</b> Pig data model in detail with neat diagram.	13	Apply	BTL 3
14	i. <b>Assess</b> the difference between hive and map reduce.	4	Apply	BTL 3
	ii. <b>Explain</b> about Hive in detail.	9		
15	<b>Illustrate</b> the data modeling in detail with an example.	13	Apply	BTL 3
16	<b>Illustrate</b> hive datatypes and file formats with an example.	13	Apply	BTL 3
17	<b>Illustrate</b> in detail Hive QL	13	Apply	BTL 3
<b>PART-C</b>				
1	<b>Recommend</b> a procedure to find the number of occurrence of a word in a document using Hive.	15	Evaluate	BTL 5
2	<b>Evaluate</b> the use of Hive. How Does Hive Interact With Hadoop explain in detail?	15	Analyze	BTL 5
3	<b>Perform analysis on web server report</b>	15	Create	BTL 6

	<p><b>Sample Data:</b></p> <pre>teleman.pr.mcs.net,-,-,[01/Jul/2005:00:03:57,- 0400],"GET,/images/KSC-logosmall.gif,HTTP/1.0",304,0 teleman.pr.mcs.net,-,-,[01/Jul/2005:00:03:57,- 0400],"GET,/images/KSC-logosmall.gif,HTTP/1.0",304,0</pre> <p>The data is comma separated. It consists of the user IP address, time at which the request is received, timezone, request type, requested link, request details, response code and bytes transferred. Usually the scale of these datasets is quite huge and running queries in a conventional method is not possible.</p> <p>Hence use Pig programming on this dataset to retrieve the necessary statistics which helps us to understand the load and usage of the server, user visit frequency, webpage popularity and the total bytes transferred.</p> <p>.</p>			
4	<p><b>Formulate</b> a Hbase table from the following data Data_file.txt contains the below data</p> <pre>1,India,Bihar,Champaran,2009,April,P1,1,5 2,India, Bihar,Patna,2009,May,P1,2,10 3,India, Bihar,Bhagalpur,2010,June,P2,3,15 4,United States,California,Fresno,2009,April,P2,2,5 5,United States,California,Long Beach,2010,July,P2,4,10 6,United States,California,San Francisco,2011,August,P1,6,2 0</pre> <p>Column family region has three column qualifiers: country, state, city</p> <p>Column family Time has two column qualifiers: year, month</p>	15	Create	BTL 6
5	<p><b>Explain what is Apache Cassandra?</b> Cassandra has become so popular because of its outstanding technical features.Comment on this statement.</p>	15	Evaluate	BTL 5