SRM VALLIAMMAI ENGINEERING COLLEGE
(An Autonomous Institution)
SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF INFORMATION TECHNOLOGY

QUESTION BANK

M.Tech DATA SCIENCE -I SEMESTER

1924101-INTRODUCTION TO BIG DATA ANALYSTICS

Regulation – 2019

Academic Year 2021 – 2022 (ODD SEMESTER)

Prepared by

Dr. D. Sridevi, Assistant Professor (Sr. G) / IT

SUBJECT   : 1924101-INTRODUCTION TO BIG DATA ANALYSTICS

SEM / YEAR: I Sem / I Year- M. Tech Data Science

UNIT I -INTRODUCTION
Introduction to Digital Data-types - Definition of Big Data - Challenges with Big Data -Evolution of Big data – Best Practices for Big data Analytics – Big data characteristics – Validating – The Promotion of the Value of Big Data -Big Data Applications – Perception and Quantification of Value - Understanding Big Data Storage –Challenges with Big Data - 3Vs of Big Data - Non Definitional traits of Big Data - Business Intelligence vs. Big Data - Data warehouse and Hadoop environment - Coexistence.

PART – A			
Q.No	Questions	BT Level	Competence
1.	List the types of Digital Data.	BTL1	Remembering
2.	What is Unstructured Data?	BTL1	Remembering
3.	How to store Unstructured Data?	BTL2	Understanding
4.	Differentiate between Structured and Unstructured Data	BTL3	Applying
5.	What is Big Data?	BTL1	Remembering
6.	List out the best practices of Big Data Analytics	BTL1	Remembering
7.	Compare BI and Data Science.	BTL3	Applying
8.	Discuss Business Acumen(expertise) Skills of Data scientist.	BTL2	Understanding
9.	List down the four computing resources of Big Data Storage.	BTL4	Analyzing
10.	What is HDFS?	BTL1	Remembering
11.	Compare Big Data vs Data Warehouse	BTL3	Applying
12.	What is Semi Structured data?	BTL1	Remembering
13.	Difference between Traditional Business Intelligence (Bi) and Big Data	BTL2	Understanding
14.	Discuss the three characteristics of big data.	BTL2	Understanding
15.	Where does Semi-structured Data Come from?	BTL4	Analyzing
16.	Mention the tasks of Big Data Analytics.	BTL6	Creating
17.	List any four Top Analytical Tools.	BTL5	Evaluating
18.	Point out the 3V's of Big Data.	BTL5	Evaluating
19.	Draw Typical Analytical Architecture.	BTL6	Creating
20.	Analyze Big Data Analytics Challenges.	BTL4	Analyzing
PART – B			
1.	(i) Explain in detail various types of Digital data. (6) (ii) Discuss the characteristics of Big Data. (7)	BTL1	Remembering
2.	(i)Write about the importance of Big Data in detail. (6) (ii)Write about Business intelligence in detail. (7)	BTL1	Remembering
3.	Extrapolate big data analytics and develop a summary of various applications in the real-world scenario. (13)	BTL6	Creating

4.	Explain in detail about 3V's of Big Data (13)	BTL1	Remembering
5.	Discuss in detail about Big Data and Data Warehouse with an example. (13)	BTL2	Understanding
6.	Explain in detail about what Big Data Analytics challenges? (13)	BTL2	Understanding
7.	Discuss in detail various Top Analytics Tools . (13)	BTL2	Understanding
8.	Explain in detail Data Analytics Life Cycle. (13)	BTL1	Remembering
9.	Illustrate big Data and Business Intelligence with an example. (13)	BTL3	Applying
10	Analyze how Big Data solution differs in many aspects of BI to use. (13)	BTL4	Analyzing
11.	(i) What are the best practices in Big Data analytics? (6) (ii) Explain the techniques used in Big Data Analytics. (7)	BTL4	Analyzing
12.	Explain the roles and stages in data science project (13)	BTL4	Analyzing
13.	Assess in detail (i) Hadoop Environment. (6) (ii) Hadoop Community Package (7)	BTL5	Evaluating
14.	(i) Generalize the list of tools related to Hadoop. (6) (ii) How does Hadoop work? (7)	BTL3	Applying
PART C			
1.	Explain in detail about Big data framework (15)	BTL6	Creating
2.	Write in detail about Four Big Data strategies (15)	BTL6	Creating
3.	Assess various techniques used to find patterns in or interpret unstructured data. (15)	BTL5	Evaluating
4.	Assess the Popular Big Data Techniques and Vendors. (15)	BTL5	Evaluating

UNIT II – CLASSIFICATION OF ANALYTICS

Big Data Analytics: Classification of analytics - Data Science - Terminologies in Big Data - CAP Theorem - BASE Concept. NoSQL: Types of Databases – Advantages – NewSQL - SQL vs. NOSQL vs NewSQL. Introduction to Hadoop: Features – Advantages – Versions - Overview of Hadoop Eco systems - Hadoop distributions - Hadoop vs. SQL – RDBMS vs. Hadoop - Hadoop Components – Architecture – HDFS. Hadoop 2 (YARN): Architecture - Interacting with Hadoop Eco systems.

PART – A

Q.No	Questions	BT Level	Competence
1	Give the classification of analytics.	BTL1	Remembering
2	State Descriptive analytics.	BTL1	Remembering
3	List the Two techniques used in descriptive analytics.	BTL1	Remembering
4	Define Predictive analytics.	BTL1	Remembering
5	Compare Descriptive Analytics with Predictive.	BTL3	Applying
6	List out the Terminologies used in Big Data Environments.	BTL3	Applying
7	Difference between Descriptive and Predictive.	BTL2	Understanding
8	What do you mean by NoSQL?	BTL2	Understanding
9	Distinguish RDBMS Versus Hadoop	BTL2	Understanding
10	Compare SQL and NoSQL	BTL2	Understanding
11	What is SQL?	BTL1	Remembering
12	Define CAP Theorem.	BTL1	Remembering
13	Draw Hadoop Architecture Diagram.	BTL3	Applying
14	What do you mean by HDFS?	BTL4	Analyzing
15	Analyze Advantages of Hadoop	BTL4	Analyzing
16	What do you mean by YARN?	BTL4	Analyzing
17	Assess what is Scheduler?	BTL5	Evaluating
18	What are the Components of Hadoop Ecosystem?	BTL5	Evaluating
19	Investigate HBase.	BTL6	Creating
20	Construct Hadoop 2 (YARN): Architecture.	BTL6	Creating

PART – B

1	Describe in detail about the Classification of analytics.	(13)	BTL1	Remembering
2	Illustrate CAP Theorem with an example.	(13)	BTL3	Applying
3	Discuss in detail various Terminologies used in Big Data environment.	(13)	BTL1	Remembering
4	List various Types of Databases and explain in detail.	(13)	BTL1	Remembering
5	Compare NewSQL - SQL vs. NoSQL vs NewSQL.	(13)	BTL2	Understanding
6	Explain in detail: (i) Hadoop features (6) (ii) Advantages of Hadoop. (7)		BTL2	Understanding

7	Compare SQL and NoSQL	(13)	BTL2	Understanding
8	(i) How Does Hadoop Work? (ii) Illustrate with an example.	(4) (9)	BTL3	Applying
9	Illustrate with an example Hadoop Distributed File System.	(13)	BTL3	Applying
10	Explain in detail various Components of Hadoop Ecosystem and how data stored in HDFS?	(13)	BTL4	Analyzing
11	Explain Hadoop 2 (YARN) Architecture in detail.	(13)	BTL4	Analyzing
12	Explain the core tasks that Hadoop performs.	(13)	BTL4	Analyzing
13	(i)Evaluate Processing Data with Hadoop and Managing Resources (ii)Applications with Hadoop YARN.	(9) (4)	BTL5	Evaluating
14	Construct the Hadoop Architecture and explain in detail.	(13)	BTL6	Creating

PART C				
1.	Investigate the differences between RDBMS and NoSQL databases	(15)	BTL6	Creating
2.	Write in detail about: (i) R Language (ii) Apache Spark (iii) MongoDB	(5) (5) (5)	BTL6	Creating
3.	Evaluate: (i) Data Mining (ii) Data Warehousing (iii) Data Science	(5) (5) (5)	BTL5	Evaluating
4.	Evaluate various challenges of Distributed Computing.	(15)	BTL5	Evaluating

UNIT III - MONGO DB
MongoDB: Introduction – Features - Data types - MongoDB Query language -. Cassandra: Introduction – Features -Data types – CQLSH - Key spaces - CRUD operations – Collections – Counter – TTL - Alter commands - Import and Export - Querying System tables. Map Reduce: Mapper – Reducer – Combiner – Partitioner – Searching – Sorting – Compression.

PART – A		
Q.No	Questions	BT Level	Competence
---	---	---	---
1.	What is MongoDB?	BTL1	Remembering
2.	Define Database.	BTL1	Remembering
3.	What is Collection?	BTL1	Remembering
4.	What do you mean by Document?	BTL1	Remembering
5.	What is Cassandra?	BTL1	Remembering
6.	List Advantages of MongoDB over RDBMS	BTL1	Remembering
7.	Why Use MongoDB?	BTL2	Understanding
8.	Where to Use MongoDB?	BTL2	Understanding
9.	What are data types supported by MongoDB?	BTL2	Understanding
10.	Describe find method.	BTL2	Understanding
11.	Analyze the NOR in MongoDB.	BTL4	Analyzing
12.	Examine relationship of RDBMS terminology with MongoDB.	BTL4	Analyzing

13.	What CQL Cassandra Query Language	BTL3	Applying
14.	What is keyspaces in Cassandra?	BTL3	Applying
15.	Analyze the data types provided by CQL.	BTL4	Analyzing
16.	Classify the CRUD operations.	BTL3	Applying
17.	Assess Time To Live (TTL) for a column in Cassandra	BTL5	Evaluating
18.	Assess tasks of The MapReduce algorithm.	BTL5	Evaluating
19.	Write the actions performed by MapReduce algorithm.	BTL6	Creating
20.	What is the use of MapReduce?	BTL6	Creating

PART – B

1	Describe in detail about MongoDB datatypes. (13)	BTL1	Remembering
2	(i)List the advantages of MongoDB over RDBMS (7) (ii) Why use MongoDB? (6)	BTL1	Remembering
3	(i)Discuss in detail Why CRUD is so important (10) (ii)List the benefits of CURD (3)	BTL1	Remembering
4	(i)What do you mean by mongosh? (3) (ii)List two options of editor mode and Write about Mongo Shell vs Legacy mongo shell. (10)	BTL1	Remembering
5	Describe in detail various User-defined datatypes in Cqlsh (13)	BTL2	Understanding
6	Discuss in detail about: (i) $eq operator (7) (ii) $gt operator (6)	BTL2	Understanding
7	Explain $gte operator with an example program. (13)	BTL2	Understanding
8	Illustrate Time To Live (TTL) for a column in Cassandra for delete with an example. (13)	BTL3	Applying
9	Illustrate with your own example to retrieve the document(s) whose first name is not "X" and last name is not "Y". (13)	BTL3	Applying
10	Illustrate Time To Live (TTL) for a column in Cassandra for Insert with an example. (13)	BTL3	Applying
11	Explain in detail with example about Importing and Exporting Data in MongoDB. (13)	BTL4	Analyzing
12	Explain sorting MapReduce algorithm (13)	BTL4	Analyzing
13	Evaluate how MapReduce employs Searching algorithm to find out the details of the employee who draws the highest salary in a given employee dataset. (13)	BTL5	Evaluating
14	Write about Aggregation Pipeline Stages in MongoDB. (13)	BTL6	Applying

1.	Write a program to retrieve the document with title MongoDB Overview (15)	BTL6	Creating
2.	Investigate various built-in data types available in CQL. (15)	BTL6	Creating
3.	Evaluate CRUD operations with an example. (15)	BTL5	Evaluating
4.	How MapReduce Works? Explain with an example. (15)	BTL5	Evaluating

UNIT IV - HADOOP ECO SYSTEMS

Hive – Architecture - data type - File format – HQL – SerDe - User defined functions - Pig: Features – Anatomy - Pig on Hadoop - Pig Philosophy - Pig Latin overview - Data types - Running pig - Execution modes of Pig - HDFS commands - Relational operators - Eval Functions - Complex data type - Piggy Bank -

	User defined Functions - Parameter substitution - Diagnostic operator.			

PART – A

Q.No	Questions	BT Level	Competence
1.	Define is Hive in Big Data.	BTL1	Remembering
2.	List two modules of Hadoop.	BTL1	Remembering
3.	What is MapReduce?	BTL1	Remembering
4.	Define HQL.	BTL1	Remembering
5.	List sub projects of Hadoop Eco system	BTL1	Remembering
6.	What do you mean by sqoop?	BTL4	Analyzing
7.	What is Pig?	BTL2	Understanding
8.	Illustrate Hive with an example.	BTL3	Applying
9.	What are the Features of Hive?	BTL2	Understanding
10.	Describe SerDe in Hive.	BTL2	Understanding
11.	Illustrate Apache Pi with an example.	BTL3	Applying
12.	Differences between Apache MapReduce and PIG	BTL2	Understanding
13.	List the uses of Pig technology.	BTL1	Remembering
14.	Analyze Pig Latin.	BTL4	Analyzing
15.	Analyze various Latin Data Types.	BTL4	Analyzing
16.	Illustrate Pig Data Types with an example.	BTL3	Applying
17.	Assess four different types of diagnostic operators of Pig Latin.	BTL5	Evaluating
18.	Assess Types of UDF's in Java	BTL5	Evaluating
19.	Write the six programming languages which UDF supports.	BTL6	Creating
20.	State the Use of Piggy Bank.	BTL6	Creating

PART – B

1	Describe in detail Architecture of Hive. (13)	BTL1	Remembering
2	Discuss in detail about: (i)Working of Hive (7) (ii)How Hive interacts with Hadoop framework (6)	BTL1	Remembering
3	List the various File Formats in Hive and explain in detail. (13)	BTL1	Remembering
4	Describe about Hive DDL Commands in detail. (13)	BTL1	Remembering
5	Discuss in detail about Hive DML Commands (13)	BTL2	Understanding
6	(i)What is Hive Query Language? (4) (ii)Discuss in detail with an example. (9)	BTL2	Understanding
7	Discuss in detail various Latin Data Types with an example. (13)	BTL2	Understanding
8	Evaluate Apache Pig scripts executed in: (i)Interactive mode (4) (ii)Batch mode (4) (iii)Embedded mode. (5)	BTL5	Evaluating
9	Illustrate Apache Pig Execution Modes with an example. (13)	BTL3	Applying
10	Explain Pig Latin Relational Operations with an example. (13)	BTL4	Analyzing

SRM

11	Examine list of eval functions provided by Apache Pig with an example.(13)	BTL4	Analyzing
12	Analyze Pig Versus Hive with an example. (13)	BTL4	Analyzing
13	Write in detail about the following Eval Functions: (i) AVG() and BagToString() (7) (ii) CONCAT() and COUNT() (6)	BTL6	Creating
14	Illustrate how to write a sample UDF using Eclipse. (13)	BTL3	Applying
PART C			
1	Evaluate Bucketing and give its advantages. (15)	BTL5	Evaluating
2	Write a Word Count Example Using Pig Script. (15)	BTL6	Evaluating
3.	Assess the following with an example, (i)Dump Operator and Describe Operator (8) (ii) Explanation Operator and Illustration Operator (7)	BTL5	Creating
4.	Write in detail Why we go for Hive When Pig is There? (15)	BTL6	Creating

UNIT V - CASE STUDIES
Big Data Case Studies – Retail sector, public sector, banking sector, small business, scientific research, health care sector.

PART – A

Q.No	Questions	BT Level	Competence
1.	Why retail companies using Big Data?	BTL1	Remembering
2.	What is the use of Big Data in Retail Industry	BTL6	Creating
3.	What are the Big Data use cases in Retail Industry?	BTL1	Remembering
4.	Analyze Why public sectors using Big Data?	BTL4	Analyzing
5.	State use of Big Data in public sectors	BTL1	Remembering
6.	Big Data use cases in public sectors	BTL1	Remembering
7.	Why banking sector using Big Data?	BTL2	Understanding
8.	Analyze the use of Big Data in banking sector	BTL4	Analyzing
9.	Assess use cases in banking sector	BTL5	Evaluating
10.	Why small business using Big Data?	BTL2	Understanding
11.	Illustrate Big Data in small business with an example.	BTL3	Applying
12.	Show the use cases in small business	BTL3	Applying
13.	Apply big Data in scientific research	BTL3	Applying
14.	Describe the use of Big Data in scientific research	BTL2	Understanding
15.	List Big Data use cases in scientific research	BTL1	Remembering
16.	Why health care sector using Big Data?	BTL4	Analyzing
17.	Describe use of Big Data in health care sector	BTL2	Understanding
18.	What are the Big Data use cases in health care sector?	BTL1	Remembering
19.	Assess various applications of Big Data in Banking Sector	BTL5	Evaluating

20.	Investigate the use of Data Mining by Walmart.	BTL6	Creating
1	Discuss in detail: (i)Personalizing customer experience (7) (ii)Predicting demands (6)	BTL1	Remembering
2	Describe in detail: (i)Operational efficiency (6) (ii)Customer journey analytics (7)	BTL1	Remembering
3	Discuss the opportunities and challenges of Big Data in Public Sector. (13)	BTL2	Understanding
4	Explain in detail: (i)Risk Management (4) (ii)Fraud Detection (4) (iii)Customer Contentment (5)	BTL1	Remembering
5	Explain in detail the use of Big Data by Walmart in achieving their goals.(13)	BTL2	Understanding
6	Explain in detail Netflix Case study. (13)	BTL1	Remembering
7	What do you know about the Big Data implementation by eBay? (13)	BTL2	Understanding
8	Explain case study on Scientific Research using Big Data. (13)	BTL4	Analyzing
9	Illustrate big data case study in healthcare challenges and opportunities (13)	BTL3	Applying
10	Explain BiClustering with an example. (13)	BTL4	Analyzing
11	Explain with suitable example the concept of Internal and External Data sources for performing data analysis in the business environment. (13)	BTL6	Creating
12	Explain with a suitable example the various tasks for a business analyst and the required skills for data analysis in a business environment. (13)	BTL4	Analyzing
13	Evaluate Facebook has been one of the most successful companies in the world at gathering our data and turning it into profit – and why some think its business practices sometimes overstep the mark. (13)	BTL5	Evaluating
14	Illustrate with a suitable example the concept of Internal and External Data sources for performing data analysis in the business environment. (13)	BTL3	Applying
PART C			
1.	Investigate How Big Data Analytics implementation helped Uber to reach greater heights. (15)	BTL5	Evaluating
2.	What are IOT Devices and how they are related to Big Data and Cloud Technologies? (15)	BTL6	Creating
3.	Evaluate Big Data in predicting the uncertainties. (15)	BTL5	Evaluating
4.	How is the emergence of Cloud Technologies related to the growth in Big Data? (15)	BTL6	Creating