

# **SRM VALLIAMMAI ENGINEERING COLLEGE**

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203.

## **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

### **QUESTION BANK**



### **VI SEMESTER**

### **AD3661 - BIG DATA ANALYTICS**

**Common to CSE, IT and AI &DS**

**Regulation – 2023**

**Academic Year 2025 – 2026**

**(Even Semester)**

*Prepared by*

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

**Dr.A. SAMYDURAI, Professor /CSE**



# SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur-603203.



## DEPARTMENT OF INFORMATION TECHNOLOGY

### QUESTION BANK

Year & Semester : III /VI  
 Subject : AD3661 - BIG DATA ANALYTICS  
 Degree & Branch : B. E - CSE

UNIT I - UNDERSTANDING BIG DATA			
Introduction to Big data – Convergence of key trends – Unstructured data – Industry examples of Big data – web analytics – Big data applications – Big data technologies –Introduction to Hadoop - Open source technologies – Cloud and big data – Mobile Business Intelligence – Crowd sourcing Analytics – Inter and trans firewall analytics			
PART A			
Q.No	Questions	BT Level	Competence
1.	What is big data?	BTL 2	Understand
2.	Name the four V's of Big Data.	BTL 1	Remember
3.	How does unstructured data differ from structured data?	BTL 2	Understand
4.	List two industries that heavily rely on big data analytics?	BTL 1	Remember
5.	What is the primary purpose of web analytics?	BTL 2	Understand
6.	Name a popular framework for distributed data processing in big data.	BTL 1	Remember
7.	What is Hadoop's role in the big data ecosystem?	BTL 2	Understand
8.	Give an example of a NoSQL database.	BTL 2	Understand
9.	How does cloud computing relate to big data?	BTL 2	Understand
10.	What is mobile business intelligence?	BTL 2	Understand
11.	How does crowdsourcing analytics work?	BTL 2	Understand
12.	What is the significance of inter-firewall analytics?	BTL 2	Understand
13.	What does trans-firewall analytics focus on?	BTL 2	Understand
14.	What are the key trends that led to the emergence of big data?	BTL 2	Understand

15.	How is big data different from traditional data analysis?		<b>BTL 2</b>	<b>Understand</b>
16.	Name a few open-source technologies commonly used in big data.		<b>BTL 1</b>	<b>Remember</b>
17.	How does big data benefit decision-making in healthcare?		<b>BTL 2</b>	<b>Understand</b>
18.	What is the purpose of data visualization in big data applications?		<b>BTL 2</b>	<b>Understand</b>
19.	Why is real-time data processing important in some big data scenarios?		<b>BTL 2</b>	<b>Understand</b>
<b>PART-B</b>				
1.	Explain Big Data and its characteristics. Discuss convergence of key trends with industry examples.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
2.	Describe unstructured data and challenges in Big Data analytics.	(16)	<b>BTL 3</b>	<b>Apply</b>
3.	Explain Big Data applications in web analytics and business intelligence.	(16)	<b>BTL 4</b>	<b>Analyze</b>
4.	Discuss Big Data technologies and open-source ecosystems.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
5.	Explain Hadoop framework and its role in Big Data analytics.	(16)	<b>BTL 4</b>	<b>Analyze</b>
6.	Discuss cloud computing and its integration with Big Data.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
7.	Explain mobile business intelligence and its significance.	(16)	<b>BTL 4</b>	<b>Analyze</b>
8.	Describe crowd sourcing analytics with examples.	(16)	<b>BTL 3</b>	<b>Apply</b>
9.	Explain inter- and trans-firewall analytics.	(16)	<b>BTL 6</b>	<b>Create</b>
10.	Analyze Big Data use cases in healthcare.	(16)	<b>BTL 4</b>	<b>Analyze</b>
11.	Analyze Big Data use cases in finance.	(16)	<b>BTL 4</b>	<b>Analyze</b>
12.	Explain challenges in Big Data management.	(16)	<b>BTL 4</b>	<b>Analyze</b>
13.	Discuss ethical and privacy issues in Big Data.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
14.	Compare traditional data processing and Big Data processing.	(16)	<b>BTL4</b>	<b>Analyze</b>
15.	Explain Big Data lifecycle in detail.	(16)	<b>BTL 4</b>	<b>Analyze</b>
16.	Discuss industry adoption of Big Data analytics.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
17.	Explain future trends in Big Data analytics.	(16)	<b>BTL 4</b>	<b>Analyze</b>

**UNIT II- NOSQL DATA MANAGEMENT**

**Introduction to NOSQL – Aggregate data models – Key-value and document data models– relationships – Graph databases – Schema less databases – Materialized views -distribution Models – Master-Slave replication – Consistency – Cassandra – Cassandra data model - Cassandra examples – Cassandra clients.**

<b>PART-A</b>				
<b>Q. No</b>	<b>Questions</b>	<b>BT Level</b>		<b>Competence</b>
1	Define NOSQL.	<b>BTL 3</b>		<b>Remember</b>
2	What are Schemaless databases?	<b>BTL 4</b>		<b>Understand</b>
3	What is Cassandra?	<b>BTL 3</b>		<b>Understand</b>
4	List out the components of Cassandra.	<b>BTL 1</b>		<b>Remember</b>
5	Write an advantages of NOSQL.	<b>BTL 1</b>		<b>Understand</b>
6	List out classification of NOSQL database.	<b>BTL 2</b>		<b>Remember</b>
7	Name any two examples of Key-value database.	<b>BTL 3</b>		<b>Remember</b>
8	Write syntax for materialized view.	<b>BTL 4</b>		<b>Understand</b>
9	Define Graph databases.	<b>BTL 2</b>		<b>Remember</b>
10	List out any four Cassandra example.	<b>BTL 6</b>		<b>Remember</b>
11	Describe Cassandra clients.	<b>BTL 1</b>		<b>Understand</b>
12	List out Key Features of NoSQL.	<b>BTL 6</b>		<b>Remember</b>
13	What is session consistency?	<b>BTL 5</b>		<b>Understand</b>
14	What is the primary purpose of master-slave replication in NoSQL databases?	<b>BTL 1</b>		<b>Understand</b>
15	What is shredding and replication?	<b>BTL 2</b>		<b>Understand</b>
16	What is write -write and read-write conflict?	<b>BTL 1</b>		<b>Understand</b>
<b>PART-B</b>				
1	Explain NoSQL databases and their need.	(16)	<b>BTL 4</b>	<b>Analyze</b>
2	Discuss aggregate data models in NoSQL.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
3	Explain key-value data model with examples.	(16)	<b>BTL 3</b>	<b>Apply</b>
4	Explain document-oriented data model.	(16)	<b>BTL 6</b>	<b>Create</b>
5	Explain graph databases and their applications.	(16)	<b>BTL 4</b>	<b>Analyze</b>
6	Discuss schema-less databases.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
7	Explain materialized views in NoSQL systems.	(16)	<b>BTL 4</b>	<b>Analyze</b>

8	Describe distribution models in NoSQL in detail.	(16)	<b>BTL 3</b>	<b>Apply</b>
9	Explain master-slave replication with examples.	(16)	<b>BTL 4</b>	<b>Analyze</b>
10	Discuss consistency models in NoSQL databases.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
11	Explain Cassandra architecture with a neat diagram	(16)	<b>BTL 4</b>	<b>Analyze</b>
12	Discuss Cassandra data model.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
13	Explain Cassandra read and write operations.	(16)	<b>BTL 4</b>	<b>Analyze</b>
14	Explain Cassandra replication and consistency.	(16)	<b>BTL 4</b>	<b>Analyze</b>
15	Discuss Cassandra query examples.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
16	Explain Cassandra clients.	(16)	<b>BTL 4</b>	<b>Analyze</b>
17	Compare RDBMS and NoSQL databases.	(16)	<b>BTL 4</b>	<b>Analyze</b>

### UNIT III - BASICS OF HADOOP

**Data format – Analyzing data with Hadoop – Scaling out – Hadoop streaming – Hadoop Pipes – Design of Hadoop Distributed File System (HDFS) – HDFS concepts – Java Interface – Data flow – Hadoop I/O – Data integrity – Compression – Serialization – Avro - File-based data structures – Cassandra – Hadoop integration**

#### PART-A

<b>Q.No</b>	<b>Questions</b>	<b>BT Level</b>	<b>Competence</b>
1	What if writable were not in MapReduce?	<b>BTL 2</b>	<b>Understand</b>
2	List out the failures in classic map reduce.	<b>BTL 1</b>	<b>Remember</b>
3	Define the term MR unit.	<b>BTL 1</b>	<b>Remember</b>
4	List out different types of job schedulers in Hadoop.	<b>BTL 1</b>	<b>Remember</b>
5	List the types of MapReduce.	<b>BTL 1</b>	<b>Remember</b>
6	Identify the features of YARN.	<b>BTL 2</b>	<b>Understand</b>
7	Mention the main components of YARN architecture.	<b>BTL 1</b>	<b>Remember</b>
8	What is MapFile?	<b>BTL 2</b>	<b>Understand</b>
9	Write the advantages and disadvantages of map reduce.	<b>BTL 2</b>	<b>Understand</b>
10	Mention an applications of Map Reduce.	<b>BTL 1</b>	<b>Remember</b>
11	What are the different stages involved in a MapReduce job run?	<b>BTL 1</b>	<b>Understand</b>
12	How would you optimize the shuffle and sort phase of a MapReduce job?	<b>BTL 2</b>	<b>Understand</b>
13	What happens if a client detects an error when reading a block in Hadoop?	<b>BTL 2</b>	<b>Understand</b>
14	State writable and its importance.	<b>BTL 1</b>	<b>Remember</b>

15	Write the heartbeat mechanism in HDFS.		<b>BTL 2</b>	<b>Understand</b>
16	List the design issues in HDFS.		<b>BTL 1</b>	<b>Remember</b>
17	Write a note on Job Tracker and Task Tracker.		<b>BTL 2</b>	<b>Understand</b>
<b>PART-B</b>				
1	Explain Hadoop ecosystem and architecture.	(16)	<b>BTL 4</b>	<b>Analyze</b>
2	Explain HDFS design and architecture.	(16)	<b>BTL 4</b>	<b>Analyze</b>
3	Discuss HDFS components and concepts.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
4	Explain HDFS data flow.	(16)	<b>BTL 6</b>	<b>Create</b>
5	Explain Hadoop I/O mechanisms.	(16)	<b>BTL 4</b>	<b>Analyze</b>
6	Discuss data integrity in Hadoop.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
7	Explain compression techniques in Hadoop in detail.	(16)	<b>BTL 4</b>	<b>Analyze</b>
8	Discuss serialization techniques.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
9	Explain Avro file format.	(16)	<b>BTL 3</b>	<b>Apply</b>
10	Explain file-based data structures.	(16)	<b>BTL 4</b>	<b>Analyze</b>
11	Discuss Hadoop streaming.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
12	Explain Hadoop pipes.	(16)	<b>BTL 4</b>	<b>Analyze</b>
13	Explain scaling out in Hadoop.	(16)	<b>BTL 4</b>	<b>Analyze</b>
14	Analyze Hadoop performance issues.	(16)	<b>BTL 4</b>	<b>Analyze</b>
15	Explain Cassandra-Hadoop integration.	(16)	<b>BTL 3</b>	<b>Apply</b>
16	Explain Hadoop Java interface.	(16)	<b>BTL 4</b>	<b>Analyze</b>
17	Explain data formats used in Hadoop with example.	(16)	<b>BTL 4</b>	<b>Analyze</b>

**UNIT IV- MAP REDUCE APPLICATIONS**

**MapReduce workflows – Unit tests with MRUnit – Test data and Local tests – Anatomy of MapReduce job run – Classic Map-reduce – YARN – Failures in classic Map-reduce and YARN – Job scheduling – Shuffle and Sort – Task execution – MapReduce types – Input Formats – Output formats.**

**PART-A**

Q.No	Questions	BT Level	Competence
1	Define the term scaling out.	<b>BTL 1</b>	<b>Remember</b>
2	State the term name node and data node.	<b>BTL 1</b>	<b>Remember</b>
3	Classify the advantages of Hadoop.	<b>BTL 2</b>	<b>Understand</b>
4	List out types of Hadoop file formats.	<b>BTL 1</b>	<b>Remember</b>
5	State HDFS.	<b>BTL 1</b>	<b>Remember</b>
6	Justify the disadvantages of Hadoop file system.	<b>BTL 2</b>	<b>Understand</b>
7	List the types of Hadoop data formats.	<b>BTL 1</b>	<b>Remember</b>
8	Define serialization.	<b>BTL 1</b>	<b>Remember</b>
9	What is Cassandra and its uses?	<b>BTL 2</b>	<b>Understand</b>
10	Write the Java interface used for interacting with HDFS.	<b>BTL 2</b>	<b>Understand</b>
11	How would you ensure data integrity in Hadoop?	<b>BTL 2</b>	<b>Understand</b>
12	How do Transformations Improve Linearity in Regression?	<b>BTL 2</b>	<b>Understand</b>
13	What are the uses of Hadoop Pipes?	<b>BTL 2</b>	<b>Understand</b>
14	Why is ensuring data integrity crucial in Hadoop Distributed System?	<b>BTL 1</b>	<b>Understand</b>
15	Define MapReduce.	<b>BTL 1</b>	<b>Remember</b>
16	List the Characteristics of MapReduce.	<b>BTL 1</b>	<b>Remember</b>
17	What are the major responsibilities of YARN?	<b>BTL 2</b>	<b>Understand</b>
18	Why is YARN used?	<b>BTL 2</b>	<b>Understand</b>
19	What is fair Scheduler ?	<b>BTL 2</b>	<b>Understand</b>
20	Why Hadoop works better with a small number of files?	<b>BTL 2</b>	<b>Understand</b>

**PART-B**

1	Explain MapReduce programming model.	(16)	<b>BTL 3</b>	<b>Apply</b>
2	Explain MapReduce workflow.	(16)	<b>BTL 4</b>	<b>Analyze</b>
3	Discuss anatomy of MapReduce job run.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
4.	Explain shuffle and sort phase.	(16)	<b>BTL 4</b>	<b>Analyze</b>
5	Discuss task execution in MapReduce.	(16)	<b>BTL 5</b>	<b>Evaluate</b>

6	Explain MapReduce input formats with example in detail .	(16)	<b>BTL 4</b>	<b>Analyze</b>
7	Explain MapReduce output formats with example in detail.	(16)	<b>BTL 3</b>	<b>Apply</b>
8	Discuss MapReduce job scheduling.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
9	Explain failures in classic MapReduce.	(16)	<b>BTL 6</b>	<b>Create</b>
10	Explain YARN architecture with neat diagram.	(16)	<b>BTL 4</b>	<b>Analyze</b>
11	Compare classic MapReduce and YARN.	(16)	<b>BTL 4</b>	<b>Analyze</b>
12	Explain MapReduce workflows.	(16)	<b>BTL 4</b>	<b>Analyze</b>
13	Explain MRUnit testing.	(16)	<b>BTL 4</b>	<b>Analyze</b>
14	Explain local testing of MapReduce jobs.	(16)	<b>BTL 3</b>	<b>Apply</b>
15	Explain different MapReduce job types.	(16)	<b>BTL 4</b>	<b>Analyze</b>
16	Explain fault tolerance in YARN.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
17	Explain MapReduce performance tuning.	(16)	<b>BTL 5</b>	<b>Evaluate</b>

#### UNIT V- : HADOOP RELATED TOOLS

**Hbase – Data model and implementations – Hbase clients – Hbase examples – praxis. 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

<b>Q.No</b>	<b>Questions</b>	<b>BT Level</b>	<b>Competence</b>
1	Classify the difference between RDBMS and HIVE.	<b>BTL 1</b>	<b>Remember</b>
2	State HBase.	<b>BTL 1</b>	<b>Remember</b>
3	What is HiveQL?	<b>BTL 1</b>	<b>Remember</b>
4	List the applications of pig grunt in big data.	<b>BTL 1</b>	<b>Remember</b>
5	Compare RDBMS and HIVE.	<b>BTL1</b>	<b>Remember</b>
6	What are the uses of HBase.	<b>BTL1</b>	<b>Remember</b>
7	Define HiveQL.	<b>BTL 1</b>	<b>Remember</b>
8	How would you use HiveQL to perform data aggregation and filtering?	<b>BTL 2</b>	<b>Understand</b>
9	Write the difference between HiveQL and SQL.	<b>BTL 2</b>	<b>Understand</b>
10	Write the syntax and semantics of the Pig Latin language.	<b>BTL 2</b>	<b>Understand</b>
11	What are the features of Hive?	<b>BTL 2</b>	<b>Understand</b>
12	What is Pig storage?	<b>BTL 1</b>	<b>Remember</b>

13	What is ZooKeeper?		<b>BTL 1</b>	<b>Remember</b>
14	What are the responsibilities of HMaster?		<b>BTL 2</b>	<b>Understand</b>
15	What is the Primary purpose of HiveQL queries in Hive Environment?		<b>BTL 2</b>	<b>Understand</b>
16	How does Hbase differ from traditional database in terms of Data Storage and Access Patterns?		<b>BTL 2</b>	<b>Understand</b>
<b>PART-B</b>				
1	Explain HBase architecture with a neat diagram.	(16)	<b>BTL 4</b>	<b>Analyze</b>
2	Evaluate HBase data model.	(16)	<b>BTL5</b>	<b>Evaluate</b>
3	Explain HBase read and write operations.	(16)	<b>BTL 2</b>	<b>Understand</b>
4.	Discuss HBase clients with suitable examples.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
5	Create a HBase for a suitable use cases.	(16)	<b>BTL 6</b>	<b>Create</b>
6	Explain Pig architecture with a neat diagram .	(16)	<b>BTL 4</b>	<b>Analyze</b>
7	Discuss Pig data model with examples.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
8	Explain Pig Latin statements in detail.	(16)	<b>BTL 4</b>	<b>Analyze</b>
9	Explain Pig script execution in detail with examples.	(16)	<b>BTL 6</b>	<b>Create</b>
10	Explain Hive architecture with a neat sketch and an example.	(16)	<b>BTL 4</b>	<b>Analyze</b>
11	Discuss Hive data types and file formats.	(16)	<b>BTL 5</b>	<b>Evaluate</b>
12	Explain HiveQL DDL commands in detail with suitable examples.	(16)	<b>BTL 4</b>	<b>Analyze</b>
13	Explain HiveQL DML commands in detail with suitable examples.	(16)	<b>BTL 4</b>	<b>Analyze</b>
14	Explain HiveQL queries with examples.	(16)	<b>BTL 4</b>	<b>Analyze</b>
15	Compare Pig and Hive in detail with suitable examples.	(16)	<b>BTL 4</b>	<b>Analyze</b>
16	Apply integration of Hive with Hadoop examples.	(16)	<b>BTL 3</b>	<b>Apply</b>
17	Discuss real-world applications of Hadoop tools.	(16)	<b>BTL 4</b>	<b>Analyze</b>