



SRI VENKATESWARA UNIVERSITY

4-Year UG Honours in B.Sc. (Data Mining) as Major in consonance with Curriculum framework (W.E.F. AY-2026-27)

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
II	IV	8	Data Warehousing & OLAP	3	3
			Data Warehousing & OLAP-Practical	2	1
		9	Big Data Tools & Technologies	3	3
			Big Data Tools & Technologies-Practical	2	1
		10	Data Communications and Computer Networks	3	3
			Data Communications and Computer Networks-Practical	2	1



SRI VENKATESWARA UNIVERSITY

II Year Syllabus for 4-Year UG Honours in B.Sc. (Data Mining) as Major in consonance with the Curriculum framework w.e.f. AY 2026-27

II Year IV Semester

Sl. No.	Category	Course No.	Course Name	No. of Hours		Total No. of Hours	No. of Credits		Total No. of Credits
				Theory	Practical		Theory	Practical	
1	Major-Core	8	Data Warehousing & OLAP	3	2	5	3	1	4
2	Major-Core	9	Big Data Tools & Techniques	3	2	5	3	1	4
3	Major-Core	10	Data Communications and Computer Networks	3	2	5	3	1	4
4	Minor	2		3	2	5	3	1	4
5	MDC	3		2	0	2	2	0	2
6	SEC	4	SEC: Design Thinking/Problem solving/ others	2	0	2	2	0	2
				16	8	24	16	4	20

**SEMESTER
-IV**

**COURSE 9: BIG DATA TOOLS AND
TECHNOLOGIES**

**Theory
hrs/week**

Credits: 3

3

Course Objectives:

1. To introduce the fundamentals of Big Data, its ecosystem, and modern challenges in data management.
2. To familiarize students with distributed storage and processing frameworks such as Hadoop and Spark.
3. To provide practical skills in using data ingestion, cleaning, and processing tools for big data.
4. To enable students to apply big data analytics for real-world problems across domains.
5. To expose learners to NoSQL databases and big data storage models.
6. To develop the ability to select appropriate big data tools and technologies for given applications.

Course Outcome:

On successful completion of this course, students will be able to:

1. Demonstrate the use of Hadoop Distributed File System (HDFS) and MapReduce.
2. Apply Apache Spark for large-scale data analytics.
3. Work with NoSQL databases (MongoDB, Cassandra) for unstructured data.
4. Use data ingestion and workflow tools such as Apache Kafka, Flume, and Sqoop.

Syllabus: UNIT 1:

Big Data overview: definition, characteristics (5Vs), Traditional data vs. Big Data, Big Data ecosystem and architecture, Challenges in Big Data storage and processing, Distributed computing basics, Introduction to Hadoop ecosystem, Use cases of Big Data in industry.

UNIT 2:

Hadoop architecture, Hadoop Distributed File System (HDFS) – design and working, MapReduce programming model, YARN resource management, Data ingestion with Apache Flume or Sqoop, Hands-on with HDFS commands.

UNIT 3:

Limitations of Hadoop MapReduce, Apache Spark architecture and RDDs, Spark DataFrames and Datasets, Spark SQL, Spark lib for machine learning, PySpark basics.

UNIT 4:

Introduction to NoSQL databases, Types of NoSQL databases – key-value, column-based, document-oriented, graph, MongoDB basics: CRUD operations, HBase and columnar storage, Comparison of NoSQL and RDBMS.

UNIT 5:

Apache Kafka – real-time data streaming, Apache Storm and Spark Streaming, Workflow management tools – Apache Oozie, Data visualization tools for Big Data - Tableau, Power BI, Big Data in cloud platforms - AWS, Azure, GCP, Big Data security and governance, Applications of Big Data in business, healthcare, IoT, social

networks, Future trends: AI integration with Big Data.

Activities:

1. Identify and analyze Big Data use cases across industries (retail, banking, healthcare).
2. Compare small dataset vs. large dataset processing performance in Python.
3. Store and retrieve files in HDFS.
4. Write a simple MapReduce program for word count.
5. Perform data transformation using Spark RDDs and DataFrames.
6. Create and query collections in MongoDB
7. Build a simple real-time data pipeline using Kafka.

Prescribed Textbooks:

- Tom White – *Hadoop: The Definitive Guide*, O'Reilly.
- Raj Kamal – *Internet of Things: Architecture and Design Principles*, McGraw Hill.
- Vignesh Prajapati – *Big Data Analytics with R and Hadoop*, Packt Publishing.
- Chuck Lam – *Hadoop in Action*, Manning Publications.

Reference Books:

- Alan Gates – *Programming Pig*, O'Reilly.
- George Coulouris et al. – *Distributed Systems: Concepts and Design*, Pearson.
- Sam R. Alapati – *Hadoop Administration*, Packt.
- Shashank Tiwari – *Professional NoSQL*, Wiley.

SEMESTER-IV

COURSE 9: BIG DATA TOOLS AND TECHNOLOGIES

Practical

Credits: 1

2 hrs/week

List of experiments:

1. Explore a real-world dataset (e.g., sales, weather, tweets) and identify how it satisfies the 5Vs of Big Data.
2. Upload a file into HDFS, list the contents of the directory, and retrieve the file back.
3. Run a MapReduce program to count word frequency in a text dataset.
4. Register a Spark DataFrame as a temporary table and run SQL queries (e.g., GROUP BY, AVG).
5. Create a MongoDB collection, insert multiple documents, and query them with conditions.
6. Set up Kafka and demonstrate message exchange between a producer and consumer.
7. Build a mini big data pipeline: ingest a dataset into HDFS, process it using Spark, store results in MongoDB, and visualize the output.

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Mining Honours
II year IV Semester
COURSE 9: Big Data Tools & Technologies

MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 3 marks.

(5X3=15)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer **FIVE** of the following. Each Question Carries 12 marks.

(5X12=60)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V