

**SIDDAGANGA INSTITUTE OF TECHNOLOGY, TUMAKURU**  
**LECTURE PLAN FOR THE ACADEMIC YEAR 2023 – 2024**

Teacher	Dr. Jagadanba G	Dept.	Information Science and Engineering
Class	B.E. VII Semester	Course	RISE22: Big Data and Analytics

**Course Objectives:**

The objectives of this course are:

1. To provide a basic understanding of the types of digital data, the characteristics of big data, the challenges confronting the enterprises embracing big data.
2. To teach the fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.
3. To introduce programming tools PIG & HIVE in Hadoop ecosystem for storage, analysis and manipulation of data.
4. To enable students to have skills that will help them to solve complex real-world problems in for decision support.

Sl. No.	Date	Topic	Remarks
<b>UNIT – I</b>			
		<b>Getting An Overview Of Big Data</b>	
1	05/10/2023	What is Big Data? History of Data Management-Evolution of Big Data, Structuring Big Data-Types of Data	
2	07/10/2023	Elements of Big Data, Big Data Analytics: Advantages of Big Data Analytics	
		<b>Introducing Technologies For Handling Big Data</b>	
3	09/10/2023	Distributed and Parallel Computing for Big Data	
4	12/10/2023	Introducing Hadoop, Cloud Computing and Big Data	
5	14/10/2023	Features of Cloud Computing, Cloud Deployment Models	
6	16/10/2023	Cloud Delivery Models	
7	19/10/2023	Cloud Services for Big Data, Cloud Providers in Big Data Market	
8	21/10/2023	In-Memory Computing Technology for Big Data	
<b>UNIT - II</b>			
		<b>Understanding Hadoop Ecosystem</b>	
9	26/10/2023	Hadoop Ecosystem, Hadoop Distributed File System HDFS Architecture	
10	28/10/2023	MapReduce, and Hadoop YARN	
11	02/11/2023	Introducing HBase- HBase Architecture, Regions	
12	04/11/2023	Storing Big Data with HBase, Interacting with the Hadoop Ecosystem	Assign ment-1
13	06/11/2023	HBase in Operation-Programming with HBase	
14	09/11/2023	Combining HBase with HDFS	

15	11/11/2023	Pig and Pig Latin, Sqoop, Zookeeper	
16	13/11/2023	Flume, and Oozie	
<b>UNIT – III</b>			
		<b>Understanding MapReduce Fundamentals and HBase</b>	
17	20/11/2023	The MapReduce Framework: Exploring the Features of Map Reduce, Working of MapReduce, Exploring Map and Reduce Functions	
18	23/11/2023	Techniques to Optimize MapReduce Jobs: Hardware/ Network Topology	
19	02/12/2023	Synchronization, File System, Uses of MapReduce	
20	04/12/2023	Role of HBase in Big Data Processing: Characteristics of HBase, Installation of HBase	
		<b>Exploring Hive</b>	
21	07/12/2023	Introducing Hive, Getting Started with Hive: Hive Variables, Hive Properties	
22	09/12/2023	Hive Queries, Data Types in Hive, Built-in Functions in Hive	
23	11/12/2023	Hive DDL, Data Manipulation in Hive	
24	14/12/2023	Data Retrieval Queries, Using JOINS in Hive	
<b>UNIT – IV</b>			
		<b>Analyzing Data with Pig</b>	
25	16/12/2023	Introducing Pig: The Pig Architecture, Benefits of Pig: Properties of Pig, Running Pig	
26	18/12/2023	Getting Started with Pig Latin, Working with Operators in Pig	
27	23/12/2023	Debugging Pig, Working with Functions in Pig	
28	28/12/2023	Error Handling in Pig	
		<b>Using Oozie</b>	
29	29/12/2023	Introducing Oozie: Main Functional Components of Oozie	
30	30/12/2023	Benefits of Oozie, Installing and Configuring Oozie	
31	01/01/2024	Understanding the Oozie Workflow, Oozie Coordinator	
32	01/01/2024	Oozie Bundle, Oozie Parameterization with EL, Oozie Job Execution Model, Accessing Oozie, Oozie SLA	Assignme nt-2
<b>UNIT – V</b>			
		<b>NoSQL Data Management</b>	
33	06/01/2024	Introducing to NoSQL, Types of NoSQL Data Models, Characteristics of NoSQL	
34	07/01/2024	Schema-less Databases, Materialized Views, Distribution Models	
35	08/01/2024	CAP theorem, Sharding	
		<b>Cassandra</b>	
36	11/01/2024	Apache Cassandra - An Introduction and Features of Cassandra	
37	12/01/2024	SQL Data types, Import and Export commands and examples	
38	13/01/2024	Querying System Tables, CQLSH, Keyspaces	
39	13/01/2024	CRUD (Create, Read, Update and Delete) Operations	

**Text Books:**

1. Big Data: Black Book, DT Editorial Services, Publication: Dream Tech Press, Edition 2016
2. Big Data and Analytics, Seema Acharya, Subhashini Chellappan, Infosys Limited, Publication: Wiley India Private Limited, 1<sup>st</sup> Edition 2015

**Reference Books:**

1. Hadoop in Practice, Alex Holmes, Manning Publications Co. September 2014, Second Edition.
2. Programming Pig, Alan Gates, O'Reilly, Kindle Publication.
3. Programming Hive, Dean Wampler, O'Reilly, Kindle Publication.

**Course Outcomes:**

Upon completion of this course the student will be able to:

- CO1:** Identify the different types of digital data, sources, challenges, elements and technologies for handling Big Data.
- CO2:** Demonstrate the Hadoop Ecosystem and have broad comprehension of HDFS, MapReduce Fundamentals and HBase.
- CO3:** Apply Pig and Hive with Hadoop Distributed File System to analyze stored Big Data.
- CO4:** Describe managing Hadoop jobs using Oozie and basic concepts of NoSQL data management.
- CO5:** Create NoSQL Databases and explore Cassandra.

**Mapping of Course Outcomes (COs) to Program Outcomes (POs) & Program Specific Outcomes (PSOs)**

		POs												PSOs		
		1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
COs	CO1	2														2
	CO2	2														2
	CO3	1	2	2												2
	CO4	2														2
	CO5	2	2	2												2
ISPE23		2	2	2												2

Assessment Tools	COs				
Direct AT	CO1	CO2	CO3	CO4	CO5
CIE (Individual)	√	√	√	√	√
SEE (Individual)	√	√	√	√	√
Assignments (Individual/Group)	√	√	√	√	√
Micro Projects (Group)					
Topic seminar (Individual)					
Case studies (Individual/Group)					

Online courses (Individual)					
Indirect AT					
Course end survey (Students)	√	√	√	√	√
Student profile (Faculty)					

**Course delivery methods, assessment tools and sample questions:**

CO1	Identify the different types of digital data, sources, challenges, elements and technologies for handling Big Data.
Delivery Methods	Black Board Teaching, Power Point presentation, Online Video Tutorials
Assessment Tools	Quiz 1, TEST 1, SEE, Assignment-1
Sample Questions	1. What are the different characteristics of Structured, unstructured and semi-structured data? Give an example for each. (CO1, PO1, L1) 7
	2. Give the characteristics of Data which are not definitional traits of BigData. (CO1, PO1, L2) 5
	3. Explain a typical data warehouse and Hadoop environment w.r.t BigData. (CO1, PO1, L2) 5
	4. How was Big Data evolved? (CO1, PO1, L1) 4

CO2	Demonstrate the Hadoop Ecosystem and have broad comprehension of HDFS, MapReduce Fundamentals and HBase.
Delivery Methods	Black Board Teaching, Power Point presentation, Online Video Tutorials
Assessment Tools	Quiz 2, TEST 1, SEE, Assignment-1
Sample Questions	1. What is the use of Hadoop? (CO2, PO2, L1) 7
	2. Give the key consideration for huge popularity of Hadoop. (CO2, PO2, L1) 5
	3. Consider that there is a data analysis project in which 20 terabytes of data needs to be analyzed on 20 different Map Reduce server nodes. Write the informal steps to perform the map and reduce function. (CO2, PO2, L2) 4
	4. Sketch with a neat diagram Hadoop high-level architecture. (CO2, PO2, L2) 5
	5. How does the data is processed in a Hadoop Environment? (CO2, PO2, L1) 4

CO3	Apply Pig and Hive scripts with Hadoop Distributed File System to analyze stored Big Data.
Delivery Methods	Black Board Teaching, Power Point presentation, Online Video Tutorials
Assessment Tools	Quiz 3, TEST 2, SEE
Sample Questions	1. What is the main use of RC File Implementation, SERDE and UDF in HIVE? (CO3, PO2, L1) 5
	2. Write Hive command to execute the following: (CO3, PO2, L2) 6 i) Create a table with four columns, first name, last name, age and income. ii) To copy the table structure into a new table.



	<p>3. Create a data file for below schemes:</p> <p>Order: CustomerId, ItemId, ItemName, OrderDate, DeliveryDate  Customer: CustomerId, CustomerName, Address, City, State, Country.</p> <ol style="list-style-type: none"> <li>Create a table for order and customer data</li> <li>Write a HiveQL(HQL) to find number of items bought by each customer.</li> </ol> <p>4. Write HIVE commands for the following :</p> <ol style="list-style-type: none"> <li>Create a database with any two database properties.</li> <li>Create an external table.</li> <li>Copy the book-title column from table Lib-info to table list-titles.</li> <li>Display the Cartesian product of two tables.</li> </ol>	6	8
--	---	---	---

<b>CO4</b>	<b>Describe managing Hadoop jobs using Oozie and basic concepts of NoSQL data management.</b>
<b>Delivery Methods</b>	Black Board Teaching, Power Point presentation, Online Video Tutorials
<b>Assessment Tools</b>	Quiz 3, TEST 2, SEE, Assignment- 2
<b>Sample Questions</b>	<ol style="list-style-type: none"> <li>List and explain the key concepts of the Oozie coordinator. (CO4, PO1, L1) 5</li> <li>Discuss how Oozie as a Hadoop workflow engine is different from other workflow applications (CO4, PO1, L1). 4</li> <li>Consider that there is a data analysis project in which 20 terabytes of data needs to be analyzed on 20 different maps reduce server nodes. Write the informal steps to perform the map and reduce function. (CO4, PO1, L2) 6</li> <li>Why NoSQL is important in Big Data Analytics. (CO4, PO1, L2) 4</li> <li>Explain the ways in which data can be distributes (CO4, PO1, L1) 4</li> </ol>

<b>CO5</b>	<b>Create NoSQL Databases and explore Cassandra.</b>
<b>Delivery Methods</b>	Black Board Teaching, Power Point presentation, Online Video Tutorials
<b>Assessment Tools</b>	Quiz 4, SEE
<b>Sample Questions</b>	<ol style="list-style-type: none"> <li>What is MongoDB? Mention the different datatypes in MongoDB. (CO5, PO1, L1) 5</li> <li>Give the use of MapReduce Function in MongoDB with a suitable example. (CO5, PO2, L1) 4</li> <li>Mention the different features of HIVE. (CO5, PO1, L1) 4</li> <li>What are the different collections available in Cassandra? (CO5, PO1, L1) 5</li> <li>What are the different IMPORT and EXPORT views used in Cassandra to help the query language with suitable examples? (CO5, PO2, L2) 4</li> </ol>

Signature of Faculty

Signature of HOD

Signature of Principal