

**Detailed Report on Big Data tools and technologies -Lab Oriented Approach**

**Date:** 24th and 25th March 2025  
**Venue:** Lab-405  
**Organized by:** Department of AI&DS   
**Target Audience:** 6th Semester Students

**Introduction**

The **Big Data tools and technologies -Lab Oriented Approach** workshop for 6th-semester students of aims to provide hands-on experience with big data technologies. This workshop enhances students’ understanding of data processing, storage, and analytics by working with industry-standard tools and real-world datasets. The resource Person was ***Mr. Raghavendra Swamy Kambhapathi – Senior Trainer, Techfortune Technologies.***

**Objectives**

* To introduce students to big data concepts and challenges.
* To provide practical exposure to Hadoop, Spark, and NoSQL databases.
* To enable students to implement data processing workflows using Apache tools.
* To apply machine learning algorithms on large datasets.
* To develop mini-projects using big data technologies.

**Workshop Schedule & Highlights**

**Day 1(Morning): Introduction to Big Data and Hadoop**

* Overview of Big Data and its importance.
* Hadoop ecosystem and its components.
* Installation and configuration of Hadoop.
* Hands-on session on Hadoop Distributed File System (HDFS).

**Day 1(Afternoon): Working with MapReduce and NoSQL , Databases Data Processing with Pig and Hive**

* Introduction to MapReduce and its working mechanism.
* Writing and executing MapReduce programs.
* Introduction to NoSQL databases (MongoDB/Cassandra).
* Performing CRUD operations in MongoDB.
* Understanding Apache Pig and its applications.
* Writing Pig Latin scripts for data transformation.
* Introduction to Apache Hive and HiveQL.
* Running queries and managing large datasets in Hive.

**Day 2(Morning): Apache Spark and Real-Time Analytics**

* Introduction to Apache Spark and its components.
* Writing Spark applications using RDDs and DataFrames.
* Performing transformations and actions in Spark.
* Introduction to Spark Streaming and real-time data processing.

**Day 2(Afternoon): Machine Learning with Big Data**

* Introduction to Spark MLlib for machine learning.
* Implementing classification, regression, and clustering models.
* Hands-on session on predictive analytics.
* Hadoop Ecosystem (HDFS, MapReduce, Pig, Hive)
* NoSQL Databases (MongoDB, Cassandra)
* Apache Spark (RDD, DataFrames, MLlib, Streaming)
* Programming Languages: Java, Python, and Scala
* Data Visualization Tools (Tableau, Matplotlib, Seaborn)

**Workshop Outcomes**

* Understand the fundamentals of big data and its ecosystem.
* Work with Hadoop and NoSQL databases for data storage and retrieval.
* Develop and execute MapReduce, Pig, and Hive queries for data processing.
* Utilize Apache Spark for large-scale data analytics and machine learning.
* Implement real-time analytics solutions using Spark Streaming.
* Complete a mini-project showcasing their skills in big data analytics.

**Feedback & Conclusion**

The Big Data Analytics workshop equips students with industry-relevant skills in handling and analyzing large-scale data. By participating in this workshop, 6th semester students gain practical exposure to essential big data technologies, improving their readiness for careers in data engineering, machine learning, and analytics

The successful completion of this workshop has contributed to the skill development of students, aligning with industry requirements and VTU curriculum standards.

**Pictures**

|  |  |
| --- | --- |
|  |  |
|  |  |

##### **Faculty Coordinator's HOD Sign**

##### Ms. Sushma M & Ms. Sindhu.G

##### Asst. Prof.

##### Dept. of AI&DS

##### BGSCET