VAAGDEVI COLLEGE OF ENGINEERING
(AUTONOMOUS)

BIG DATA ANALYTICS

B.TECH-V Semester **L/T/P/C**
3/0 /0 /3

Pre-requisites: Programming Knowledge(C, Java), DBMS, Data Mining

Course Objectives:

- To understand the basic concepts of Big Data Analytics
- Explore the different tools for working with Big Data
- Understand the fundamentals of Hadoop and Map Reduce.
- Working on different command of HIVE & PIG.

UNIT-I: Introduction to Big Data

Types of Digital Data: Classification of Digital Data: Structured data, Semi-structured data and Unstructured. Introduction to Big Data: Characteristics of Data, Evolution of Big Data, Definition of Big Data, Challenges with Big Data, What is Big Data?, Why Big Data?, Traditional Business Intelligence (BI) versus Big Data, A Typical Data Warehouse Environment, A Typical Hadoop Environment, What is New Today?, What is Changing in the Realms of Big Data?

UNIT-II: Big Data Analytics

What is Big Data Analytics?, Sudden Hype Around Big Data Analytics?, Classification of Analytics, Challenges that Prevent Businesses from Capitalizing on Big Data, Top Challenges Facing Big Data, Why is Big Data Analytics Important?, Data Science, Data Scientist, Terminologies Used in Big Data Environments, Top Analytics Tools. The Big Data Technology Landscape: NoSQL: Types of NoSQL databases, advantages and comparison.

UNIT-III: Introduction to Hadoop

Features and advantages and versions of Hadoop. Hadoop Ecosystems and distributions. Hadoop versus SQL. Introducing Hadoop, RDBMS versus Hadoop, Distributed Computing Challenges, History of Hadoop, Hadoop Overview, HDFS (Hadoop Distributed File System), Processing Data with Hadoop, Managing Resources and Applications with Hadoop YARN (Yet Another Resource Negotiator), Interacting with Hadoop Ecosystem: PIG, HIVE & HBase.

UNIT-IV: Understanding Map Reduce

Introduction to Map Reduce,The Map Reduce framework, Techniques to optimize Map Reduce jobs, uses of Map Reduce. Mapper, Reducer, Combiner, Partitioner, Searching, Sorting, Compression.

UNIT-V: HIVE and PIG

Introduction to HIVE, Hive Architecture, Hive Data Types, Hive File Format, Hive Query Language (HQL), Hive Operations. Introduction to PIG, Pig Latin Overview: statements, keywords, identifiers, operators. Data Types in Pig: simple, complex. Running Pig, Execution Modes of Pig: local, Map Reduce.

Course Outcomes: After completing this course the students will be able to

CO-1: Understand the importance of Big Data.
CO-2: Learn Big Data Analytics and NoSQL databases.
CO-3: Understand the Hadoop Ecosystems and Hadoop Distributed File System
CO-4: Apply Map Reduce analytics using Hadoop
CO-5: Learn the commands and operations of HIVE & PIG.

TEXT BOOK