# BCSE0157: INTRODUCTION TO BIG DATA ANALYTICS

**Credits: 03** **L-T–P: 3–0–0**

| Module No. | Content | Teaching Hours |
|---|---|---|
| I | **Big Data technology Landscape**: Types of Digital Data (Structured, Semi-Structured, Unstructured),Concept, importance and characteristics of data, Challenges with big data, Big data stack, Big Data 1.0, 2.0 and 3.0, Traditional BI vs. Big Data Environment, NoSQL Databases, NoSQL Vs. RDBMS, New SQL, Introduction to Data Science/Scientist<br><br>**HADOOP 1.0:** Introducing Hadoop 1.0, Limitations of RDBMS, Hadoop Components, High Level Architecture of Hadoop, History of Hadoop, Special Features of Hadoop, Introduction to HDFS 1.0, Architecture, Daemons, Working with HDFS Command, Introduction to Map-Reduce 1.0, Architecture, Daemons<br><br>**HADOOP 2.0:** Introducing Hadoop 2.0, Limitations of 1.0, Introduction to HDFS 2.0, Architecture, Daemons, Introduction to Map-Reduce 2.0, YARN, Architecture, Daemons, Word Count Example using Java, Introduction to Hadoop 3.0, Difference among Hadoop1.0, Hadoop2.0, Hadoop3.0 | 13 |
| II | **Introduction to Mongo DB**: RDBMS vs. MongoDB, JSON, Unique Key, Dynamic Queries, Sharding, Replication, MongoDB QL: Create, Drop Database and Collections, CRUD: Create, Insert, Find, Update, Delete, Map Reduce Programming, Aggregations<br>**Introduction to Cassandra DB**: Features of Cassandra, CQL Data Types, CQLSH: CRUD, Counter, TTL, List, Set, Map, Tracing, Import Export csv files | 14 |
| III | **HADOOP Ecosystem and Flume:** Introduction to Hadoop Ecosystem, Sqoop, Zookeeper, Plug-in Components: Impala, Hue, Flume: Introduction, Application, Advantage, Features.<br>**Introduction to HIVE:** Hive Architecture, Hive Data types, Hive Collection Types, Hive File Formats, Hive Query Language, Hive Partitions, Bucketing, Views, RCFile Implementation, Hive User Defined Function, SerDe, UDF<br>**Introduction to Pig:** History and Anatomy of Pig, Pig on Hadoop, Use Case for Pig, Pig Primitive Data Types, Pig Latin Overview, Execution Modes of Pig, Field, Tuple, Bag, User Defined Function, Parameters in Pig, Piggy Bank, Word count example using Pig, Pig vs Hive, When to use Pig. | 13 |

**Text Book**:

- Seema Acharya and Subhashini Chellappan, “Big Data and Analytics”, 1st Edition, 2015, Wiley, India.
- Jure Leskovec, Anand Rajaraman, Jeff Ullman, “Mining of Massive Datasets”, 2nd Edition, 2014, Cambridge University Press.

**Reference Books:**

- Chuck Lam, “Hadoop in Action”, 2nd Edition, 2014, Manning Publications.

**Outcome:** At the end of the course, student will be able to

CO 1. Understand the concept and challenges of big data

CO 2. Work with existing technology to collect, manage, store, query, and analyze the various form of big data;

CO 3. Perform job scheduling of various applications and resource management in the cluster using Hadoop and Yarn

CO 4. Do the data summarization, query, and analysis over the big data with the help of pig and hive

CO 5. Prepare the regression model, cluster and decision tree over the real big data

CO 6. Gain hands-on experience in large-scale analytics tools to solve some open big data problems;