**BIG DATA ANALYTICS**

**Unit I**

Introduction to Big Data: What is big data, why big data, the convergence of key trends, unstructured data, industry examples of big data, web analytics, big data and marketing, fraud and big data, risk and big data, credit risk management, big data and algorithmic trading, big data and healthcare, big data in medicine, advertising and big data, big data technologies, introduction to Hadoop, open-source technologies, cloud and big data, mobile business intelligence, Crowd-sourcing analytics, inter and trans firewall analytics. 

**Unit II**

No SQL: Introduction to NoSQL, aggregate data models, aggregates, key-value and document data models, relationships, graph databases, schema less databases, materialized views, distribution models, sharding, master-slave replication, peer-peer replication, sharding and replication, consistency, relaxing consistency, version stamps, map-reduce, partitioning and combining, composing map-reduce calculations.  

**Unit III**

Hadoop: Data format, analyzing data with Hadoop, scaling out, Hadoop streaming, Hadoop pipes, design of Hadoop distributed file system (HDFS), HDFS concepts, Java interface, data flow, Hadoop I/O, data integrity, compression, serialization, Avro, file-based data structures. 

**Unit IV**

MapReduce: MapReduce workflows, unit tests with MRUnit, test data and local tests, anatomy of MapReduce job run, classic Map-reduce, YARN, failures in classic Map-reduce and YARN, job scheduling, shuffle and sort, task execution, MapReduce types, input formats, output formats. 

**Unit V**

Recent Trends in Big Data Analytics: HBase, data model and implementations, HBase clients, HBase examples, praxis. Cassandra, Cassandra data model, Cassandra examples, Cassandra clients, Hadoop integration, Hive, data types and file formats, HiveQL data definition, HiveQL data manipulation, HiveQL queries. 

**Textbooks / References:**

1. Big Data Analytics, Introduction to Hadoop, Spark, and Machine-Learning, Raj Kamal, Preeti Saxena, McGraw Hill, 2018.
2. Big Data, Big Analytics: Emerging Business intelligence and Analytic Trends for Today's Business, Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, John Wiley & Sons, 2013.
3. Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013
4. Hadoop: The Definitive Guide, Tom White, Third Edition, O’Reilly, 2012.
5. Hadoop Operations, Eric Sammer, O’Reilly, 2012.
6. Programming Hive, E. Capriolo, D. Wampler, and J. Rutherglen, O’Reilly, 2012.
7. HBase: The Definitive Guide, Lars George, O’Reilly, 2011.
8. Cassandra: The Definitive Guide, Eben Hewitt, O’Reilly, 2010.
9. Programming Pig, Alan Gates, O’Reilly, 2011.