



R21 Regulations

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR
(Established by Govt. of A.P., ACT No.30 of 2008)
ANANTHAPURAMU – 515 002 (A.P) INDIA

MASTER OF COMPUTER APPLICATIONS

| Course Code | BIG DATA TECHNOLOGIES | L | T | P | C |
|---|-----------------------|---------------------|---|---|---|
| 21F00307 | | 4 | 0 | 0 | 4 |
| Semester | | III | | | |
| Course Objectives: | | | | | |
| <ul style="list-style-type: none"> To understand the specialized aspects of big data including big data application, and big data analytics. To study different types Case studies on the current research and applications of the Hadoop and big data in industry. | | | | | |
| Course Outcomes (CO): Student will be able to | | | | | |
| <ul style="list-style-type: none"> Discuss the challenges and their solutions in Big Data Understand and work on Hadoop Framework and eco systems. Explain and Analyze the Big Data using Map-reduce programming in Both Hadoop and Spark framework. Demonstrate spark programming with different programming languages. Demonstrate the graph algorithms and live streaming data in Spark | | | | | |
| UNIT – I | | Lecture Hrs: | | | |
| What is big data, why big data, 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 | | Lecture Hrs: | | | |
| Introduction to NoSQL, aggregate data models, aggregates, key-value and document data models, relationships, graph databases, schemaless 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 mapreduce calculations | | | | | |
| UNIT – III | | Lecture Hrs: | | | |
| Data format, analysing 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 | | Lecture Hrs: | | | |
| 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 | | Lecture Hrs: | | | |
| 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. | | | | | |
| Text Books: | | | | | |
| <ol style="list-style-type: none"> Big Data Analytics, Introduction to Hadoop, Spark, and Machine-Learning, Raj kamal, PreetiSaxena, McGraw Hill, 2018. Big Data, Big Analytics: Emerging Business intelligence and Analytic trends for Today's Business, Michael Minelli, Michelle Chambers, and AmbigaDhiraj, John Wiley & Sons, 2013 | | | | | |