



JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA
KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF INFORMATION TECHNOLOGY

III Year – II Semester		L	T	P	C
		3	0	0	3
BIG DATA ANALYTICS					

Course Objectives:

- To optimize business decisions and create competitive advantage with Big Data analytics
- To learn to analyze the big data using intelligent techniques
- To introduce programming tools PIG & HIVE in Hadoop ecosystem

Course Outcomes:

At the end of the course, the students will be able to

- Illustrate big data challenges in different domains including social media, transportation, finance and medicine
- Use various techniques for mining data stream
- Design and develop Hadoop
- Identify the characteristics of datasets and compare the trivial data and big data for various applications
- Explore the various search methods and visualization techniques

UNIT I:

Introduction: Introduction to big data: Introduction to Big Data Platform, Challenges of Conventional Systems, Intelligent data analysis, Nature of Data, Analytic Processes and Tools, Analysis vs Reporting.

UNIT II:

Stream Processing: Mining data streams: Introduction to Streams Concepts, Stream Data Model and Architecture, Stream Computing, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream, Estimating Moments, Counting Oneness in a Window, Decaying Window, Real time Analytics Platform (RTAP) Applications, Case Studies - Real Time Sentiment Analysis - Stock Market Predictions.

UNIT III:

Introduction to Hadoop: Hadoop: History of Hadoop, the Hadoop Distributed File System, Components of Hadoop Analysing the Data with Hadoop, Scaling Out, Hadoop Streaming, Design of HDFS, Java interfaces to HDFS Basics, Developing a Map Reduce Application, How Map Reduce Works, Anatomy of a Map Reduce Job run, Failures, Job Scheduling, Shuffle and Sort, Task execution, Map Reduce Types and Formats, Map Reduce Features Hadoop environment.

UNIT IV:

Frameworks and Applications: Frameworks: Applications on Big Data Using Pig and Hive, Data processing operators in Pig, Hive services, HiveQL, Querying Data in Hive, fundamentals of HBase and ZooKeeper.

UNIT V:

Predictive Analytics and Visualizations: Predictive Analytics, Simple linear regression, Multiple linear regression, Interpretation of regression coefficients, Visualizations, Visual data analysis techniques, interaction techniques, Systems and application



JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF INFORMATION TECHNOLOGY

Text Books:

1. Tom White, “Hadoop: The Definitive Guide”, Third Edition, O’reilly Media, Fourth Edition, 2015.
2. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, “Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data”, McGrawHill Publishing, 2012.
3. Anand Rajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, CUP, 2012

Reference Books:

1. Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, John Wiley & sons, 2012.
2. Paul Zikopoulos, DirkdeRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, “Harness the Power of Big Data: The IBM Big Data Platform”, Tata McGraw Hill Publications, 2012.
3. Arshdeep Bahga and Vijay Madisetti, “Big Data Science & Analytics: A Hands On Approach “, VPT, 2016.
4. Bart Baesens, “Analytics in a Big Data World: The Essential Guide to Data Science and its Applications (WILEY Big Data Series)”, John Wiley & Sons, 2014.

Software Links:

1. Hadoop: <http://hadoop.apache.org/>
2. Hive: <https://cwiki.apache.org/confluence/display/Hive/Home>
3. Piglatin: <http://pig.apache.org/docs/r0.7.0/tutorial.html>