



CSXX0249: Visual Big Data Analytics

L-T-P-Cr: 3-0-0-3

Objectives:

- To understand the fundamentals of big data and its characteristics.
- To realize storage and processing of big data
- To explore various tools and platforms used for big data visualization.
- To design and implement visual analytics solutions using big data platforms.

Course Outcomes:

At the end of the course, a student should have:

Sl. No.	Outcome	Mapping to POs
1.	Gaining Insight about Big Data concepts, tools and applications.	PO1, PO2
2.	Understanding Hadoop framework for the huge datasets.	PO1, PO2, PO5
3.	Understanding features of NoSQL, HIVE and PIG	PO1, PO2, PO5
4.	Application and understanding of key techniques and theory behind data visualization	PO1, PO2, PO5

Unit 1

Lecture 3

Evolution of Big data, What is Big data, Characteristics of Big Data, Types of Big Data, Traditional vs. Big Data, Big Data Applications, Big Data Architecture.

Unit 2

Lecture 5

RDBMS Vs Hadoop, Hadoop Architecture ,Hadoop Distributed File System (HDFS), Hadoop Distributed File System, Overview of Yarn, Features of YARN Rack Awareness, Application Workflow of Hadoop, HDFS Federation, Map Reduce: Developing a map-reduce application – Map-reduce working procedure: The Map Task, The Reduce Task, Grouping by Key, Partitioner and Combiners, Detail of Map Reduce Execution, Spark, Hadoop Ecosystem.

Unit 3

Lecture 9

Introduction to NoSQL, No SQL Business drivers, HIVE and PIG, HIVE: background, architecture, warehouse directory and meta-store, HIVE query language, loading data into table, HIVE built-in functions, joins in HIVE, Partitioning, PIG: background, architecture, PIG Latin Basics, PIG execution modes, PIG processing – loading and transforming data, PIG built-in functions, filtering, grouping, sorting data Installation of PIG and PIG Latin commands.

Unit 4

Lecture

15

Challenges of big data visualization, Categories of Data Visualization: Numerical Data, Categorical Data, Explanation of data visualization, Bar Charts, Histograms, Pie Charts, Scatter Plots, Line Plots, Regression, Time-series and Geospatial data visualization, Network and hierarchical data visualization, Networks and Trees - Heat Map – Tree Map - Map Color and Other Channels Manipulate View - Visual Attributes, Multivariate data visualization – Geometric projection techniques - Icon-based techniques -Pixel-oriented techniques - Hierarchical techniques - Scatterplot matrix - Hyper box – Trellis display - Parallel coordinates

10

Approaches to big data visualization: Data compression, Statistical methods, Information theory for big data visualization, High performance algorithms for visualization, Outlier detection using visualization, Visualization based data analysis techniques, Tools: Tableau, Power BI, D3.js

Text Books:

- Visualization Analysis and Design by Tamara Munzner, A K Peters Visualization Series, CRC Press.
- Thomas Erl ,Big Data Fundamentals-Concepts, Drivers and Techniques, Pearson publication, 2016.

Reference books:

- Ben Fry, Visualizing Data, 1st edition, O'Reilly Media, United States, 2008.
- The Visualization Handbook edited by Charles D. Hansen and Chris R. Johnson.