Objectives:

- To explore the fundamental concepts of big data analytics
- To develop in-depth knowledge and understanding of the big data analytic domain.
- To learn to analyze the big data using intelligent techniques.
- To understand the various search methods and visualization techniques.
- To learn to use various techniques for mining data stream.
- To understand the applications using Map Reduce, pig, hive, spark, and MongoDB Concepts
- To get a in depth understanding of machine learning, Deep Learning, and NLP techniques.

Big Data Technologies

Introduction to Linux

- The Architecture and Structure of Linux
- Introduction to Linux File system
- File and Text processing commands,
- Basic of I/O commands
- Introduction to Users and Groups
- Essentials of Effective User
- Group, and Password Management

Introduction to Big data

- Introduction to big data platform
- Big data challenges
- Big Data Applications
- Types of Big Data Technologies
- Limitations and Solution of Big data Architecture
- Introduction to different big data Architectures

Hadoop Environment

- Introduction to Hadoop and Hadoop Architecture
- What is Hadoop?
- Brief History and Evolution of Hadoop
- Hadoop Distributions and Vendors
- Hadoop Architecture
- Core components of Hadoop

Hadoop Distributed File System

- What is HDFS?
- Core components of HDFS
- Hadoop Server Roles: Name Node, Secondary Name Node, and Data Node.
- HDFS Architecture overview,
- The HDFS command line and web interfaces,
- Analyzing the Data with Hadoop
- Demonstration to Cloudera
- Quickstart virtual machine
- How to set up Hadoop cluster and Install on Virtual Machine
- Hadoop Configuration
- Security in Hadoop
- Administering Hadoop,
- Security in a Cloudera cluster (HDFS, Hive)

Big data analytics with Map Reduce Framework

- Hadoop Map Reduce paradigm
- Map Reduce Execution Framework
- Anatomy of a Map Reduce Job
- Partitioners and Combiners
- Input Formats (Input Splits and Records, Text Input, Binary Input, Multiple Inputs)
- Output Formats (Text Output, Binary Output, Multiple Outputs)

Big data analytics with PIG

- Introduction to PIG
- Pig Execution Modes
- Basics of PIG Latin Programming Conventions
- Data Types
- Arithmetic and Relational Operators
- UDF Statements
- PIG Latin Scripting
- PIG Built-In Functions
- Eval Functions
- Load/Store Functions
- Math Functions
- String Functions
- Date Time Functions
- Writing a PIG UDF
- Piggy Bank and PIG Macros,
- Real-Time Data Analytics using PIG

Big data analytics with Hive

- The Hive Data-ware House
- Basics of Hive Query Language
- Working with Hive QL
- Operators and Functions
- Importing Data, Querying Data & Managing Outputs
- Hive Tables (Managed Tables and Extended Tables)
- Partitions and Buckets, Aggregating
- Joins Views
- Data manipulation with Hive
- User Defined Functions
- Writing HQL scripts.

Big data analytics with Spark

- Initializing Spark
- Spark Components and Architecture
- Resilient Distributed Datasets (RDDs)
- RDD Operations
- Passing Functions to Spark

- Working with Key-Value Pairs
- Shuffle operations
- RDD Persistence
- Shared Variables

Big data analytics with Spark

- Working with Spark with Hadoop
- Spark SQL
- Dataframes and Datasets
- Spark Streaming.

Big data analytics with MongoDB

- Overview of SQL (DDL, DML, TCL)
- Introduction to NoSQL
- Difference between SQL and NoSQL
- working with MongoDB (Installation, CRUD operations, Aggregation pipeline, Indexing, Data Modeling)

Python Programming

- Installing Python
- Introduction to Python Basic Syntax
- Data Types
- Variables
- Operators, Input/output, Strings
- Python data structure
- Lists, Tuples, Dictionaries, Sets.
- If, If- else, Nested if-else
- Looping, for, while and nested loops
- Control structure, uses of break & continue
- Functions and methods and Exception Handling
- OOPs Concepts
- Python classes and objects
- Introduction and Installation of Machine learning packages like PANDAS, NUMPY
- SKLearn, Matplotlib, Seaborn.
- Mathematical Computing with NumPy

- Data Manipulation with Pandas
- Machine Learning with Scikit–Learn.
- Introduction to Data Visualization in Python (i.e. matplotlib, Seaborn)

Machine learning

- Introduction to Machine Learning and data preprocessing
- What is machine learning?
- Types of learning
- Applications of Machine learning
- Evaluating ML techniques.
- Data cleaning
- Scaling of continuous features
- Encoding of categorical features,
- Train and test split
- Machine learning algorithms
- Linear Regression
- Decision Trees, Decision Trees case study
- Naive bayes classifier, assigning probabilities and calculating results, Naïve Bayes case study
- K-Nearest Neighbors, Algorithm and case study
- Ensemble Learning: Concept of model ensembling
- Random forest
- Gradient boosting Machines
- Model Stacking
- Support Vector Machines
- Different type of Unsupervised Machine Learning Algorithms
- Clustering, K-mean
- agglomerative clustering
- Association rule mining
- Apriori Algorithm

Introduction to Deep Learning

- Neural Network and its applications
- Single layer neural Network
- Constructing Neural Networks model
- Overview of Feed Forward Neural Network
- Back propagation
- Activation Functions: Sigmoid, Hyperbolic Tangent
- Introduction to deep Learning
- Why is Deep Learning taking off?
- Deep Learning Architecture
- Introduction to Tensorflow
- Introduction to Keras
- Building blocks of deep neural networks
- Activation Functions
- Why non-linear activation functions?
- Computer Vision:
- Introduction to Convolutional Neural Network.
- Sequence Modeling:
 - Recurrent Neural Network
- Real world case studies for CNN and RNN Model

Introduction to NLP

- Overview of NLP
- Pre-processing
- Need of Pre-processing Data
- Introduction to NLTK
- Using Python Scripts
- Shallow Parsing
- Deep Parsing
- Text featurization technique
- NLP with Machine Learning and Deep Learning
- Word2Vec models
- Building NLP Application.

Project Implementation