- **Cloud Native Development (4 Credits)**
- Introduction to Cloud Computing
- Cloud Platforms and related Services
- Enterprise Applications
- Evolution of Deployment: Inside the Enterprise, the Web, the Data Center / Cloud
- Characteristics and Structure of a Cloud-Native Application
- Services and Microservices
- Microservices Architecture
- Composability and Decomposing solutions into Microservices
- Building Microservices
- Orchestration / Choreography of Microservices
- Shared data and communication
- Patterns and Best Practices for Cloud-Native Development
- Unit Testing Microservices
- Porting Monolithic applications to Microservices
- Overview of DevOps
- Continuous Integration and Continuous Delivery / Deployment
- Deploying Microservices via Containers
- CI/CD Pipeline for microservices and automation
- Deployment platform: automated operation/management of microservices
- Exercises on building applications using microservices
- Exercises on Containerization using Dockers
- Exercises on Kubernetes for deploying and managing microservices
- CI/CD exercises

Learning Objectives

- Understand basics of Cloud Computing and related Services
- Understand how to develop Applications on Cloud
- Understand Microservices Architecture
- Understand best practices in CI/CD pipelining
- Understand how to use Container Technologies
- Learn how to deploy on Cloud

- **Minor Project - Software Development (2 Credits)**

Design an application using UML, JUnit and Mockito and use TDD methodology to drive its development.

Learning Objectives

- Design an application using UML
- Implement the application using JUnit and Mockito
- Implement additional features using TDD methodology

Outline for Semester IV Courses

- **Data Engineering (4 Credits)**
- Concepts retailed to distributed computing
- Hadoop Distributed File System
- MapReduce Programming in Python
- Enterprise Data Management
- Relational Database Modelling
- Normal Forms and ER Diagrams
- Concepts of NoSQL Databases
- Introduction to Apache HBase

- HBase Python API
- Comparison of NoSQL Databases
- Spark Architecture
- RDD, DataFrame API,SparkSQL
- Exploratory Data Analysis with PySpark
- Predictive Analysis with Spark MLlib

Learning Objectives

- Understand the concept of distributed data processing
- Understand the methods of distributed storage
- Understand how Hadoop achieves distributed computing and storage
- Write MapReduce jobs in Python
- Understand the concepts of Data Management
- Execute Data Modeling from a Relational Database
- Understand concepts of NoSQL databases
- Understand the working of Apache HBase
- Learn about Apache Spark and how it achieves data processing
- Execute EDA using PySpark
- Execute Predictive Analysis using Spark's ML Library

- **Data Mining and Warehousing (4 Credits)**
- Understanding Data: Data and Attributes – Nominal, Ordinal, Interval, and Ratio. Measures of Similarity and Dissimilarity
- Understanding Datasets: Types of Datasets, Quality of Data in Datasets, Dimensions of data, Data Pre-processing including dimensionality reduction; Visualization: Review of basic techniques (from Probability & Statistics), Visualizing Spatial Temporal Data, Visualizing Higher Dimensional Data
- Association Rule Mining: Frequent Itemset Mining- The Apriori Principle and the Apriori Algorithm; Rule Generation in the Apriori Algorithm; Compact Representation of the Frequent Itemsets; FP-Growth algorithm and Frequent Itemset generation
- Clustering: Different types of clusters (resulting out of clustering); K-means algorithm, Agglomerative Hierarchical Clustering, Density-based Clustering, , Subspace Clustering, Graph-based Clustering, Self-Organizing Maps, Evaluation of Clusters
- Anomaly Detection: Causes of and approaches to detection
- Text Mining: Similarity Computation for Text data; Clustering Methods for text
- 4Vs of Big Data
- Big Data: Industry Case Studies
- Introduction to Data Warehouse and Data Lakes
- Designing Data Warehousing for an ETL Data Pipeline
- Designing Data Lake for an ETL Data Pipeline
- Fundamentals of Apache Hive
- Writing HQL for Data Analysis
- Partitioning and Bucketing with Hive
- Data warehousing with Redshift
- Analyze data with RedShift
- Running Spark on Multi Node cluster
- Spark Memory & Disk optimisation
- Optimising Spark Cluster environment
- Introduction to Apache Flink
- Batch Data Processing with Flink
- Stream Processing with Apache Flink
- SQL API
- Intro to real-time data processing architectures
- Fundamentals of Apache Kafka
- Setting up Kafka Producer and Consumer
- Kafka Connect API & Kafka Streams
- Spark Streaming Architecture
- Spark Streaming APIs
- Building Stream Processing Application with Spark