Apache Flink

Duration: 2days (8hrs/day)

Prerequisites:

- Basic knowledge of Java
- Basic knowledge of Apache Hadoop or Apache Spark

Course Objective: This course aims to equip students with the knowledge and skills necessary to develop and implement real-time data processing applications using Apache Flink. You will learn about DataSet an DataStream API, navigate event time complexities, explore windowing techniques, and delve into stateful processing and fault tolerance. You'll learn to analyze data streams with Flink SQL and the Table API

Apache Flink Version: Latest

Lab Requirement: Koenig DC (CentOS)

Module 1 - Introduction to Stream Processing and Apache Flink

Batch Processing

What is Stream Processing

Stream Processing Architecture

Why is Stream Processing Important

Big Data

Hadoop and its Architecture

Apache Spark and its Architecture

Why is there a need for Apache Flink

Module 2 - Runtime Architecture

What is Apache Flink

Features of Flink

Flink Architecture

Flink Data Flow

Lab: Download and Install Apache Flink

Lab: Create a Maven Project for Apache Flink Program

Module 3 - Foundations of the DataStream API

DataSet Transformations

Types of DataSet Transformations

Lab: DataSet Transformations

DataStream in Flink

DataStream Transformations

Lab: DataStream Transformations

Module 4 - Data Pipelines and Stateful Stream Processing

Data Sources For DataStream API

Data Sinks For DataStream API

Lab: Reduce Operations on DataStream

Lab: Split Operations on DataStream

Stateful Stream Processing

Module 5 - Event Time and Watermarks

Event Time and its Importance

Watermarks

Late Elements & Allowed Lateness

Lab: Watermark creation on tumbling Windows

Module 6 - Windows and Streaming Analytics

Window in Apache Flink

Tumbling Window,Sliding Window,Session Window and Global Window

Window - Time Notions

Lab: Tumbling Processing Time Windows

Lab: Tumbling Event Time windows

Lab: Sliding Windows

Module 7 - State Backends and Fault Tolerance

State and its types

State Persistence

Lab: Value State

Lab: List State

CheckPointing

Fault Tolerance

Barriers and Barrier Snapshotting

Lab: CheckPointing

State Backend

Module 8 - Connector Ecosystem

Process Function

Side Output

Connector Ecosystem and its Components

Module 9 - Intro to Flink SQL and the Table API

Table & SQL API in Flink

Similarities and Differences

Lab: Table & SQL API

Module 10 - Use Cases and Application Patterns

Use Cases for Flink

Event-driven Applications

Data Analytics Applications

Data Pipeline Applications