DOPS-246

Streaming on Cloudera with Apache Flink

Course Overview

Course Type
Instructor-Led Training

Level
Intermediate

Duration
2 days

Platform
Cloduera Base

Topics Covered
- Apache Flink
- SQL Stream Builder
- Schema Registry
- Integration with Kafka

About This Training

During this two-day instructor-led training course, participants will learn development and operations for Cloudera Streaming Analytics, a framework for low-latency processing and analytics powered by Apache Flink and Cloudera's innovative SQL Stream Builder.

Through extensive hands-on exercises, students will gain experience deploying and managing a Flink cluster, developing and running Flink applications, and using SQL Stream Builder's continuous SQL to perform analytics on streaming data.

What Skills You Will Gain

During this course, you learn how to:

- Deploy a Flink cluster using Cloudera Manager
- Develop Flink batch and streaming applications
- Run and view Flink jobs
- Transform data streams
- Use watermarks and windows to analyze streaming data
- Analyze data with Cloudera SQL Stream Builder
- Monitor Flink application metrics

Who Should Take This Course?

This course is designed for those who have experience with administration and application development on the Cloudera platform. Students must have at least basic familiarity with Java and Linux. Our Apache Kafka on Cloudera course, or equivalent experience with Apache Kafka, is a recommended prerequisite.

Other Training that Might Interest You

- Apache Kafka on Cloudera
- Ingesting with Cloudera DataFlow and Apache NiFi
- Administering Cloudera on premises
- Preparing with Cloudera Data Engineering and Apache Spark

DOPS-246

Streaming on Cloudera with Apache Flink

Overview

- Introduction to Apache Flink and Stream Processing
- Typical Use Cases
- Related Products

Service Deployment

- Planning Requirements
- Installation
- Flink Dashboard
- Exercise: Running a Flink Program

Flink Basics

- Execution Environment
- Flink Application Structure
- Create a Flink Project
- Build a Flink Program
- Exercise: Building a Simple Flink Program

Architecture

- Logical
- Physical
- Parallelism
- Fault Tolerance
- Data Storage

DataStream API

- DataStream API Overview
- Data Types and Serialization
- Sources and Sinks
- Transformations
- Exercise: Batch Processing Using Flink
- Exercise: Creating a Flink Streaming Application

DataStream API (continued)

- Using Kafka as a Source and Sink
- Exercise: Creating a Streaming Application Using a Kafka Source

FlinkSQL and Table API

- Streaming Concepts
- Programming Options
- Integrations
- Exercise: Using FlinkSQL and Kafka

Stateful Stream Processing

- Stateful Streaming Applications
- Checkpoints
- Event Time Processing
- Watermarks
- Windows
- Exercise: Tumbling Windows with Event Time

Cloudera SQL Stream Builder

- Overview
- Streaming SQL Console
- Data Providers
- SQL Stream Jobs
- Exercise: Using SQL Stream Builder

Monitoring

- Metrics
- Logging
- Backpressure
- Resource Configuration
- Exercise: Monitoring