

Apache Pig and Hive

Description

This 4 day training course is designed for developers who need to create applications to analyze Big Data stored in Apache Hadoop using Pig and Hive. Topics include: Hadoop, YARN, HDFS, MapReduce, data ingestion, workflow definition, using Pig and Hive to perform data analytics on Big Data and an introduction to Spark Core and Spark SQL

Delegates will learn

- Hadoop Distributed File System
- Apache Pig
- Apache Hive
- Spark Core, Spark SQL and Oozie

Audience

Software developers who need to understand and develop applications for Hadoop.

Outline

An Introduction to the Hadoop Distributed File System

- Understanding Hadoop
- The Hadoop Distributed File System
- Ingesting Data into HDFS
- The MapReduce Framework

An Introduction to Apache Pig

- Introduction to Apache Pig
- Advanced Apache Pig Programming

An Introduction to Apache Hive

- Apache Hive Programming
- Using HCatalog
- Advanced Apache Hive Programming

Working with Spark Core, Spark SQL and Oozie

- Advanced Apache Hive Programming (Continued)
- Hadoop 2 and YARN
- Introduction to Spark Core and Spark SQL
- Defining Workflow with Oozie

Prerequisites

Students should be familiar with programming principles and have experience in software development. SQL knowledge is also helpful. No prior Hadoop knowledge is required.