

HDP Apache Hive Training

Course Outline

Information Architecture and Big Data

- Enterprise Data Warehouse Optimization

Introduction to Apache Hive

- About Apache Hive
- About Apache Zeppelin and Apache Superset (incubating)

Apache Hive Architecture

- Apache Hive Architecture

Apache Hive Programming

- Apache Hive Basics
- Apache Hive Transactions (Hive ACID)

File Formats

- SerDes and File Formats

Partitions and Bucketing

- Partitions
- Bucketing
- Skew and Temporary Tables

Advanced Apache Hive Programming

- Data Sorting
- Apache Hive User Defined Functions (UDFs)
- Subqueries and Views
- Joins
- Windowing and Grouping
- Other Topics

Apache Hive Performance Tuning

- Cost-Based Optimization and Statistics
- Bloom Filters

- Execution and Resource Plans

Live Long and Process (LLAP) Deep Dive

- Live Long and Process Overview
- Apache Hive and LLAP Performance
- Apache Hive and LLAP Installation

Security and Data Governance

- Apache Ranger
- Apache Ranger and Hive
- Apache Atlas
- Apache Atlas and Hive Integration

Apache HBase and Phoenix Integration with Hive

- Apache HBase Overview
- Apache HBase Integration with Apache Hive
- Apache Phoenix Overview

Apache Druid (incubating) with Apache Hive

- Apache Druid (incubating) Overview
- Apache Druid (incubating) Queries
- Apache Druid (incubating) and Hive Integration

Apache Sqoop and Integration with Apache Hive

- Overview of Apache Sqoop

Apache Spark and Integration with Apache Hive

- Introduction to Apache Spark
- Apache Hive and Spark

Introduction to HDF (Apache NiFi) and Integration with Apache Hive

- Introduction to Apache NiFi
- Apache NiFi and Apache Hive

Appendix: EDW Offload Workshop