

TRAINING OFFERING | DEV-331

HDP Developer: APACHE HIVE & ADVANCED SQL

SUBJECT MATTER EXPERT

This training course is designed for analysts and developers who need to create and analyze Big Data stored in Apache Hadoop using Hive. Topics include: Understanding of HDP, HDF and its integration with Hive. Hive on Tez, LLAP and Druid OLAP query analysis. Hive data ingestion using HDF and Spark. Also, a focus on EDW offload capabilities in HDP using Hive.

PREREQUISITES

Students should be familiar with programming principles and have experience in software development. SQL, data modeling and light scripting knowledge is also helpful. No prior Hadoop Knowledge is needed.

TARGET AUDIENCE

Analysts, developers and data engineers who need to understand, do analysis and develop applications for Hive on HDP.

FORMAT

50% Lecture/Discussion | 50% Hands-on Labs

AGENDA SUMMARY

Information Architecture for Big Data, An introduction to Hadoop, An Introduction to Hive, Hive Programming, Hive File Formats and Compression, Hive Performance Tuning, Hive Security and Data Governance, Pig, HBase and Phoenix, Druid, Streaming Hive Data Ingest, and an EDW Offload to Hive Workshop

INFORMATION ARCHITECTURE FOR BIG DATA

- Intro to NoSQL & NewSQL Data Access Methods
- Comparison and Contrast of Hive with RDBMS
- Enterprise Data Warehouse Introduction
- Federated Data Warehouse Architecture
- Entity Types and Modeling for NoSQL versus EDW

INTRODUCTION TO HIVE

- About Hive
- Overall HDP 3.0 ecosystem and Hive version 3.0.0
- Comparing Hive to other RDBMS
- Hive Architecture
 - Overall Hive Architecture
 - Hive on Tez Architecture
 - LLAP Architecture
 - Hive Metastore
- Hive 3.0 enhancements

About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

INTRODUCTION TO HIVE (CONTINUED)

- Materialized Views
- Query Results Cache
- Workload Management in LLAP
- Better Statistics
- ACID by default
- Big Data Technology integrations(Phoenix and Druid)
- Security and Governance updates

HIVE PROGRAMMING

- Submitting Hive Queries
 - Apache Superset with Ambari
 - Beeline
 - Hive Client (JDBC,ODBC, Thrift)
 - HiveServer2
- Introduction to Hive SQL
 - Managed and External Tables
 - Storage Formats
 - Row Formats & SerDe
 - Partitions and Dynamic Partitions
 - Buckets for Only Specific Use Cases
 - Skew Tables
 - Temporary Tables
 - Constraints (PL and FK are for the Optimizer and BI Tools)
 - DDL (Create/drop (External and Managed tables)/alter/truncate)
 - DDL Indexes and Archiving

About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405



About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

HIVE PROGRAMMING (CONTINUED)

- DML (Select/Insert/Update/delete/merge)
- DML (Import/Export)
- DML Explain Plans
- DML (Introduction to Joins)
- Advanced Hive SQL
 - Complex Joins
 - Operators and UDFs
 - Vectorized Query Execution
 - CDC
 - Windowing
 - HIVE ACID transactions

HIVE FILE FORMATS AND COMPRESSION

- ORC
- ORC and Compression
- Avro SerDe
- Parquet SerDe
- CSV SerDe
- JSON SerDe
- RCFile
- Parquet
- Hive SerDes

About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

HIVE PERFORMANCE TUNING

- Dynamic Partitions with Tez
- More on Explain Plans Analysis
- Introduction to Query Tuning
- Hive Table Statistics
- Hive Cost-Based Optimization (CBO)
- Hive Major and Minor Compaction

HIVE SECURITY AND DATA GOVERNANCE

- Hive Ranger Policies
- Hive Integration with Atlas

About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

HBASE AND PHOENIX

- About HBase and Phoenix
- HBase Architecture
- Storage Handlers
- Mapping Hive to HBase Column Families

Druid

- About Druid
- Hive Storage Handlers - (Hive tables stored in Druid)
- Executing Queries in Druid from Hive
- OLAP Queries in Hive
- OLAP Indexes

About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

STREAMING HIVE DATA INGEST

- Spark
 - Introduction to Spark 2.x
 - Spark Session and SparkSQL
 - Spark connections to Hive
 - Spark ETL with Hive
 - Spark and Hive Catalog
 - Spark Structured Streaming
- HDF(NiFi)
 - Introduction to HDF and NiFi
 - Hive Processors
 - Hive Data Flow ETL with NiFi

WORKSHOP(OPTIONAL)

- EDW Offload to Hive

About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

HANDS-ON LABS AND DEMOS

- Setting up the Environment
- Demo: Setting up Hive LLAP
- Hive Beeline client from command line
- Hive, Zeppelin and Apache Superset
- Introduction to Hive SQL
- Hive Transactions (Hive ACID)
- Hive SQL and File Formats
- Hive Partitions
- Hive Materialized Views
- Hive joins
- Hive integration with Ranger
- Hive Integration with Atlas
- Hive and Druid
- Hive and Spark
- HDF and Hive
- Optional Lab: Map HBase Table to Hive Table

Revised 09/25/2018

About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405