

# **Processing Big Data with Hadoop in Azure HDInsight**

**Modality: On Demand**

**Duration: 25 Hours**

This course sets you up for the Exam of 70-775 prompting the Certification of 70-775. This program do not comprise the Official Examination Coupon but you can apply separately to buy the Official Exam Voucher.

## **About this course:**

An ever-increasing number of companies are taking on the challenge of examining large information. This course shows you how to utilize the technologies of Hadoop in MS Azure HDInsight to construct the solutions of batch processing that reshape and cleanse information for the analysis. In this course of five-week, you'll figure out how to utilize innovations like Pig, Hive, Sqoop, and Oozie with Hadoop in HDInsight; and the way to perform with the clusters of HDInsight from Linux, Windows, and Mac OSX customer PCs.

NOTE: To finish the hands-on components in this program, student will need a subscription of Azure and a Linux, Windows, or Mac OS X customer PC. You can pursue free trial subscription of Azure (required a valid credit card for authentication, but you won't be charged for the services of Azure). Note that the free trial isn't accessible in all areas. It is conceivable to finish the course and gain a credential without finishing the hands-on activities. Also, this course assists the understudies in the groundwork for the MS 70-775: certification exam of Perform Data Engineering on MS Azure HDInsight.

The normal pay for Big Data Analysts is **\$135,000** every year.

## **Course Objective:**

- Provision of the cluster of HDInsight.
- Connect to the cluster of HDInsight, upload data, and run jobs of MapReduce.
- Utilize Hive to process and store data.
- Processing big data with Pig
- Processing big data with Hive
- Utilize custom user-defined functions of Python from Pig and Hive.
- Describe and run workflows for the processing of data using Oozie.
- Move data between databases and HDInsight utilizing Sqoop.
- Orchestrating Big Data Workflows

## **Audience:**

This course is planned for

- Computer Programmers

- Big Data Professionals

## Prerequisites:

- Understanding with basic SQL query syntax and database concepts.
- Understanding with the fundamentals of programming
- Readiness for the study actively and persevere.

## Suggested prerequisites courses:

Programming for Absolute Beginners

SQL Basics

## Career path:

Cloud Computing Career Path.

## Course Outline:

### Module 1: Getting Started with HDInsight

- Lesson 1: Big Data, Hadoop, and HDInsight
- Lesson 2: Working with HDInsight
- Lab: Getting Started with HDInsight
- Module 1 Review

### Module 2: Processing Big Data with Hive

- Lesson 1: Working with Hive Tables
- Lesson 2: Developing Hive Applications
- Lab: Processing Big Data with Hive
- Module 2 Review

### Module 3: Going Beyond Hive with Pig and Python

- Lesson 1: Processing Data with Pig
- Lesson 2: Extending Pig and Hive with UDFs
- Lab: Beyond Hive - Pig and Python
- Module 3 Review

### Module 4: Building a Big Data Workflow

- Lesson 1: Implementing Workflows with Oozie
- Lesson 2: Transferring Data with Sqoop
- Lab: Orchestrating Big Data Workflows
- Module 4 Review

## Module 5: Course Exam

- Course Wrap-up
- Exam?