

Implementing Real-Time Analytics with Hadoop in Azure HDInsight

Modality: On Demand

Duration: 10 Hours

About this course:

You'll learn how to implement low-latency and streaming Big Data solutions using Hadoop technologies like HBase, Storm, and Spark on Microsoft Azure HDInsight.

Note: *To complete the hands-on elements in this course, you will require an Azure subscription and a Windows, Linux, or Mac OS X client computer. You can sign up for a free Azure trial subscription (a valid credit card is required for verification, but you will not be charged for Azure services). Note that the free trial is not available in all regions. It is possible to complete the course and earn a certificate without completing the hands-on practices.*

Course Objective:

In this course, you'll learn how to use:

- HBase to implement low-latency NoSQL data stores.
- Storm to implement real-time streaming analytics solutions.
- Spark for high-performance interactive data analysis.

Audience:

- Developers

Prerequisite:

- Familiarity with Hadoop clusters and Hive in HDInsight
- Familiarity with database concepts and basic SQL query syntax
- Familiarity with basic programming constructs (for example, variables, loops, conditional logic). Experience with Java or C# is useful but not essential
- A willingness to learn actively and persevere when troubleshooting technical problems is essential

Course Outline:

Using HBase for NoSQL Data

- Introduction HBase
- Accessing Data in HBase
- Additional Resources
- Lab

- Review

Using Srm for Streaming Data

- Introduction Srm
- Implementing Srm Topologies
- Additional Resources
- Lab
- Review

Using Spark for Interactive Analysis

- Introduction Spark
- Exploring Data with Spark
- Additional Resources
- Lab
- Review

Introducing Kafka

- Introduction Kafka

Final Exam

- Course Wrap-Up
- Exam