edX

Course Learning Objectives

By the end of this course you should be able to:

- Understand the origins of Apache Hadoop and its ecosystem (data trends and the rise of Big Data in the enterprise IT; how the Apache Hadoop ecosystem is being developed and governed today, and how you can contribute to it; the importance of Open Enterprise Hadoop)

- Discuss the core building blocks of Hadoop: HDFS and YARN

- Leverage the three cornerstone parallel data processing frameworks: MapReduce, Apache Spark, and Apache Hive

- Gain hands-on experience with running small scale, but fully functional Hadoop environments, right from the comfort of your own desktop

- Manage Hadoop and parallel data processing frameworks from the command line

- Deploy Hadoop in a fully-clustered environment as part of a modern enterprise IT

- Build data lake management architectures around Hadoop

- Secure and scale your data lakes in multi-tenant enterprise environments.

Learn About Verified Certificates