

Table of Contents

- The demand for Big Data analysts and Hadoop administrators 4
- About DBA University, Inc..... 4
- Srini Ramineni : Big Data with Hadoop Faculty, Founder DBA University 5
- Pricing..... 6
- Cloud lab Access..... 6
- Course Topics 6
 - 1. Introduction to Big Data and the Hadoop Framework. 7
 - 1.1 What is Big Data and what are the 3 characteristics for Big Data classification..... 7
 - 1.2 Introduction to Apache Hadoop. 7
 - 1.3 History and current popular distributions of Hadoop. 7
 - 1.4 Big Data with Hadoop job market and current trends and future predictions. 7
 - 1.5 What are the use cases of Hadoop and learn about the Hadoop ecosystem. 7
 - 1.6 Lab Practice: Connect to the DBA University single node Hadoop server and browse its setup. Fully-distributed Hadoop cluster lab work will follow after a few topics are completed. 7
 - 1.7 Lab Practice: How to setup a single node Hadoop server on your own PC..... 7
 - 2. The Hadoop File System (HDFS)..... 7
 - 2.1 Introduction to the Hadoop Distributed File System (HDFS)..... 7
 - 2.2 What is replication factor in HDFS and learn about best practices in HDFS design. 7
 - 2.3 What is the Name Node, Secondary Name Node and what are Data Nodes in HDFS. 7
 - 2.4 Browse the HDFS using the web interface..... 7
 - 2.5 Identify configuration parameters for the Namenode and Datanode. 7
 - 2.6 High Availability of Data and Metadata (Name Node) in HDFS. 7
 - 2.7 Practice lab exercises working with HDFS using City of Chicago data sets..... 7
 - 3. Map Reduce computation paradigm. 8
 - 3.1 Introduction the Map Reduce computation paradigm for Big Data processing..... 8

- 3.2 What are mappers and reducers? 8
- 3.3 Learn about the distributed data processing in Map Reduce..... 8
- 3.4 Understand the differences between Map Reduce 1.0 and the latest Map Reduce with Yarn version..... 8
- 3.5 Learn about the different components of the Map Reduce computation framework. 8
- 4. Apache Sqoop and Hadoop..... 8
 - 4.1 Introduction to Apache Sqoop tool..... 8
 - 4.2 Prerequisites for the Sqoop data connector for Oracle and Hadoop. 8
 - 4.3 How to import data from a relational database to Hadoop using Sqoop. 8
 - 4.4 How to export data from Hadoop to a relational database using Sqoop..... 8
 - 4.5 Practice lab exercises with Apache Sqoop and an Oracle database. 8
- 4. Data warehousing in Hadoop (Apache Hive) 8
 - 5.1 Introduction to Apache Hive..... 8
 - 5.2 Understand the components and architecture of Apache Hive. 8
 - 5.3 The command line interfaces for running HiveQL: hive and beeline. 8
 - 5.4 Learn about Hive Partitions and Buckets..... 8
 - 5.5 Learn and practice HiveQL statements. 8
 - 5.6 How to work with the Twitter API to download tweets data..... 8
 - 5.7 Practice lab exercises working with real time data sets in Hive. 8
- 6. Apache Pig..... 9
 - 6.1 What is Apache Pig?..... 9
 - 6.2 Learn about the Pig Data Model. 9
 - 6.3 What are the rules and syntax of the Pig Latin language? 9
 - 6.4 What is a JSON data object and how to load and analyze JSON data sets using Pig? 9
 - 6.5 Practice lab exercises working with real time JSON data sets in Pig. 9
- 7. Install Cloudera Hadoop cluster in the cloud..... 9
 - 7.1 Choosing the hardware and compute resources for the servers (nodes) in a Hadoop cluster..... 9
 - 7.2 Software installation prerequisites of the Cloudera Hadoop cluster (CDH)..... 9
 - 7.3 Understand Cloudera Director and Cloudera Manager software components. 9
 - 7.4 Learn how to perform Cloud Computing using Amazon Web Services (AWS)..... 9
 - 7.5 Lab practice: How to install a 3 node Cloudera Hadoop cluster (CDH) in the cloud (AWS). 9

- 7.6 Lab practice: How to administer, manage and monitor the Hadoop cluster nodes using Cloudera Manager..... 9
- 7.5 Lab practice: Learn about Apache Hue web interface..... 9
- 7.6 Lab practice: Use Hue web interface to input Hive, Sqoop and Pig commands..... 9
- 8. Apache Spark 9
 - 8.1 Introduction to Apache Spark..... 10
 - 8.2 Compare Apache Spark and the Map Reduce computational framework..... 10
 - 8.3 Learn about Spark SQL and DataFrames..... 10
 - 8.4 How to store and analyze JSON documents using Apache Spark software framework..... 10
 - 8.5 Practice lab exercises using Apache Spark..... 10
- 9. Cloudera Impala..... 10
 - 9.1 Introduction to Cloudera Impala..... 10
 - 9.2 Key Features of Cloudera Impala..... 10
 - 9.3 Cloudera Impala vs Map Reduce computational framework..... 10
 - 9.4 Comparison among Apache Hive, Pig and Impala..... 10
 - 9.5 Practice lab exercises using Cloudera Impala..... 10
- 10. Apache Flume..... 10
 - 10.1 Apache Flume and real-world use cases..... 10
 - 10.2 What are the various components of Apache Flume..... 10
 - 10.3 Flume agent configuration..... 10
 - 10.4 Practice lab exercises using Apache Flume..... 10
- Unique Teaching Methodology..... 10
- More Information 11

The demand for Big Data analysts and Hadoop administrators

Welcome to a new world. Welcome to the world of Big Data. As per a recent [McKinsey Global Institute](#) report, there are almost 200,000 Big Data analytical talent positions available and 1.5 million more data-savvy managers needed to take full advantage of Big Data in the United States. The transformational potential of Big Data is in the below five domains.

1. Health Care (United States).
2. Public sector administration. (European Union).
3. Retail (United States).
4. Manufacturing (Global).
5. Personal location data (Global).

About DBA University, Inc.

DBA University, Inc. is a professionally run organization based in Chicago, USA that specializes in training and consulting services in the field of computer databases.

In our training division, we are currently offering courses in Big Data with Hadoop, Oracle database development and administration, Microsoft SQL Server database administration and AWS Cloud Computing using expert instructors and excellent REMOTE LAB access through [affordable prices](#).

We are a niche company and very dedicated and committed towards the success of our students. Our courses are very hands-on and aimed at preparing our students to face the real world with self-confidence and expert level technical skills. Our faculty and student video site are at <http://video.dbauniversity.com>

You are welcome to register for our courses. Registration for our upcoming courses can be done at <https://www.DBAuniversity.com>

Srini Ramineni : Big Data with Hadoop Faculty, Founder DBA University



The Big Data with Hadoop training is provided by Srini Ramineni (Srini).

Srini Ramineni is the founder of DBA University and has rich experience in the IT industry in USA managing both relational databases as well as Big Data with Hadoop environments. He also has completed a graduate degree credit hour program in Big Data using Hadoop in DePaul University, Chicago, USA in 2015. He provides training in an interactive method with lots of LAB practice work in his classes.

He can be reached at srini@dbauniversity.com

Pricing

We offer world class Big Data with Hadoop training at an affordable price of **\$649**. With \$649, you get 12 months of access to our pre-recorded video lectures (20 videos) on a 24*7 basis and 6 months of access to our cloud lab server (with an option to extend the remote lab access duration to 1 year during checkout). Each video recording is about 1 hour to 1.5 hours in duration.

- 1) 12 months of on Demand access to our pre-recorded live training videos on a 24*7 basis.
- 2) Even though these are video recordings, you still get a similar experience as a live training. Also, the courses are taught using a White Board to provide a classroom like experience.
- 3) Cloud based remote lab access for 6 months. (with an option to extend the remote lab access duration to 1 year.)
- 4) Training material PDF book with lab exercises through Dropbox.com email download.
- 5) There are 20 video recordings in this Big Data with Hadoop training course.
- 6) The all-inclusive course fee is **\$649**.

Please purchase online at https://dbauniversity.com/course/big_data_with_hadoop_training/

We accept online DEBIT CARD, CREDIT CARD or PayPal.

Cloud lab Access

- Each student gets access to a separate remote lab server.
- Students get an opportunity to work with Hadoop in both a single node cluster as well as a fully-distributed (multi-node) cluster in the Amazon cloud (AWS). There will be an additional charge of \$30 approximately by AWS for using their cloud infrastructure to setup the 3-node cluster. The Hadoop distribution that will be used for the lab work is the Cloudera distribution.

Course Topics

1. Introduction to Big Data and the Hadoop Framework.

- 1.1 What is Big Data and what are the 3 characteristics for Big Data classification.
- 1.2 Introduction to Apache Hadoop.
- 1.3 History and current popular distributions of Hadoop.
- 1.4 Big Data with Hadoop job market and current trends and future predictions.
- 1.5 What are the use cases of Hadoop and learn about the Hadoop ecosystem.
- 1.6 Lab Practice: Connect to the DBA University single node Hadoop server and browse its setup. Fully-distributed Hadoop cluster lab work will follow after a few topics are completed.
- 1.7 Lab Practice: How to setup a single node Hadoop server on your own PC.

2. The Hadoop File System (HDFS).

- 2.1 Introduction to the Hadoop Distributed File System (HDFS).
- 2.2 What is replication factor in HDFS and learn about best practices in HDFS design.
- 2.3 What is the Name Node, Secondary Name Node and what are Data Nodes in HDFS.
- 2.4 Browse the HDFS using the web interface.
- 2.5 Identify configuration parameters for the Namenode and Datanode.
- 2.6 High Availability of Data and Metadata (Name Node) in HDFS.
- 2.7 Practice lab exercises working with HDFS using City of Chicago data sets.

3. Map Reduce computation paradigm.

- 3.1 Introduction the Map Reduce computation paradigm for Big Data processing.
- 3.2 What are mappers and reducers?
- 3.3 Learn about the distributed data processing in Map Reduce.
- 3.4 Understand the differences between Map Reduce 1.0 and the latest Map Reduce with Yarn version.
- 3.5 Learn about the different components of the Map Reduce computation framework.

4. Apache Sqoop and Hadoop.

- 4.1 Introduction to Apache Sqoop tool.
- 4.2 Prerequisites for the Sqoop data connector for Oracle and Hadoop.
- 4.3 How to import data from a relational database to Hadoop using Sqoop.
- 4.4 How to export data from Hadoop to a relational database using Sqoop.
- 4.5 Practice lab exercises with Apache Sqoop and an Oracle database.

4. Data warehousing in Hadoop (Apache Hive)

- 5.1 Introduction to Apache Hive.
- 5.2 Understand the components and architecture of Apache Hive.
- 5.3 The command line interfaces for running HiveQL: hive and beeline.
- 5.4 Learn about Hive Partitions and Buckets.
- 5.5 Learn and practice HiveQL statements.
- 5.6 How to work with the Twitter API to download tweets data.
- 5.7 Practice lab exercises working with real time data sets in Hive.

6. Apache Pig

6.1 What is Apache Pig?

6.2 Learn about the Pig Data Model.

6.3 What are the rules and syntax of the Pig Latin language?

6.4 What is a JSON data object and how to load and analyze JSON data sets using Pig?

6.5 Practice lab exercises working with real time JSON data sets in Pig.

7. Install Cloudera Hadoop cluster in the cloud.

7.1 Choosing the hardware and compute resources for the servers (nodes) in a Hadoop cluster.

7.2 Software installation prerequisites of the Cloudera Hadoop cluster (CDH).

7.3 Understand Cloudera Director and Cloudera Manager software components.

7.4 Learn how to perform Cloud Computing using Amazon Web Services (AWS).

7.5 Lab practice: How to install a 3 node Cloudera Hadoop cluster (CDH) in the cloud (AWS).

7.6 Lab practice: How to administer, manage and monitor the Hadoop cluster nodes using Cloudera Manager.

7.5 Lab practice: Learn about Apache Hue web interface.

7.6 Lab practice: Use Hue web interface to input Hive, Sqoop and Pig commands.

8. Apache Spark

8.1 Introduction to Apache Spark.

8.2 Compare Apache Spark and the Map Reduce computational framework.

8.3 Learn about Spark SQL and DataFrames.

8.4 How to store and analyze JSON documents using Apache Spark software framework.

8.5 Practice lab exercises using Apache Spark.

9. Cloudera Impala

9.1 Introduction to Cloudera Impala.

9.2 Key Features of Cloudera Impala.

9.3 Cloudera Impala vs Map Reduce computational framework.

9.4 Comparison among Apache Hive, Pig and Impala.

9.5 Practice lab exercises using Cloudera Impala.

10. Apache Flume

10.1 Apache Flume and real-world use cases.

10.2 What are the various components of Apache Flume.

10.3 Flume agent configuration.

10.4 Practice lab exercises using Apache Flume.

Unique Teaching Methodology

We use a *Digital White Board* for teaching our online courses. Teaching with a DIGITAL WHITE BOARD gives our students a traditional *classroom like experience*. The instructor writes and draws by free hand on the DIGITAL WHITE BOARD and all our students can see it in real time in the online classroom.

More Information

1. Please purchase online at https://dbauniversity.com/course/big_data_with_hadoop_training
2. For any additional information email srini@DBAuniversity.com