Big Data and Hadoop Spark Developer

About **the Course**

The world is getting increasingly digital, and this means big data is here to stay. In fact, the importance of big data and data analytics is going to continue growing in the coming years, as an increasing number of businesses go digital. The need for specialists in big data and analytics will keep increasing correspondingly. Close to 80% of data scientists say there is currently a shortage of professionals working in the field. Choosing a career in the field of big data and analytics might just be the type of role that you have been trying to find to meet your career expectations.

The Big Data Hadoop training course lets you master the concepts of the Hadoop framework and prepares you for Cloudera's CCA175 Big data certification. With our online Hadoop training, you'll learn how the components of the Hadoop ecosystem, such as Hadoop 2.7, Yarn, MapReduce, HDFS, Pig, Impala, HBase, Flume, Apache Spark, etc. fit in with the Big Data processing lifecycle. Implement real life projects in banking, telecommunication, social media, insurance, and e-commerce on CloudLab.

Big data professionals have impressive earning potential, with the median salaries hovering around the $116,000 range, and even entry level positions can fetch an average annual salary of $92,000.

Key Features

Interactive Training

Hands on Projects with Real-world Applications

Exam Preparation Alignment

simpli·learn

Course **Advisor**

Ronald van Loon: Named by Onalytica as one of the three most influential people in Big Data, Ronald is also an author for a number of leading Big Data and Data Science websites, including Datafloq, Data Science Central, and The Guardian. He also regularly speaks at renowned events.

Course **Outline**

- Introduction: Course Information
- Lesson One: Introduction to Big Data and Hadoop Ecosystem
- Lesson Two: HDFS and YARN
- Lesson Three: MapReduce and Sqoop
- Lesson Four: Basics of Hive and Impala
- Lesson Five: Working with Hive and Impala
- Lesson Six: Types of Data Formats
- Lesson Seven: Advanced Hive Concept and Date File Partitioning
- Lesson Eight: Apache Flume and HBase

- Lesson Nine: Pig
- Lesson Ten: Basics of Apache Spark
- Lesson Eleven: RDDs in Spark
- Lesson Twelve: Implementation of Spark Applications
- Lesson Thirteen: Spark Parallel Processing
- Lesson Fourteen: Spark RDD Optimization Techniques
- Lesson Fifteen: Spark Algorithm
- Lesson 16: Spark SQL

---

simpl**learn** | simplilearn.com

Founded in 2009, Simplilearn is one of the world's leading providers of online training for Digital Marketing, Cloud Computing, Project Management, Data Science, IT Service Management, Software Development and many other emerging technologies. Based in Bangalore, India, San Francisco, California, and Raleigh, North Carolina, Simplilearn partners with companies and individuals to address their unique needs, providing training and coaching to help working professionals meet their career goals. Simplilearn has enabled over 1 million professionals and companies across 150+ countries train, certify and upskill their employees.

Simplilearn's 400+ training courses are designed and updated by world-class industry experts. Their blended learning approach combines e-learning classes, instructor-led live virtual classrooms, applied learning projects, and 24/7 teaching assistance. More than 40 global training organizations have recognized Simplilearn as an official provider of certification training. The company has been named the 8th most influential education brand in the world by LinkedIn.

India – United States – Singapore