

# Cloudera Data Scientist Training

**Duration:** 32.00 hours (4 days)

**26.0 CPD Hours**

**Rating:** ★ 4.6 (5,878 reviews)

## Course Information

**Delivery Format:** Instructor Led - Online

## Course Overview

This workshop covers data science and machine learning workflows at scale using Apache Spark 2 and other key components of the Hadoop ecosystem. The workshop emphasizes the use of data science and machine learning methods to address real-world business challenges. Using scenarios and datasets from a fictional technology company, students discover insights to support critical business decisions and develop data products to transform the business. The material is presented through a sequence of brief lectures, interactive demonstrations, extensive hands-on exercises, and discussions. The Apache Spark demonstrations and exercises are conducted in Python (with PySpark) and R (with sparklyr) using the Cloudera Data Science Workbench (CDSW) environment. The workshop is designed for data scientists who currently use Python or R to work with smaller datasets on a single machine and who need to scale up their analyses and machine learning models to large datasets on distributed clusters. Data engineers and developers with some knowledge of data science and machine learning may also find this workshop useful.

## About This Course

This workshop covers data science and machine learning workflows at scale using Apache Spark 2 and other key components of the Hadoop ecosystem. The workshop emphasizes the use of data science and machine learning methods to address real-world business challenges. Using scenarios and datasets from a fictional technology company, students discover insights to support critical business decisions and develop data products to transform the business. The material is presented through a sequence of brief lectures, interactive demonstrations, extensive hands-on exercises, and discussions. The Apache Spark demonstrations and exercises are conducted in Python (with PySpark) and R (with sparklyr) using the Cloudera Data Science Workbench (CDSW) environment. The workshop is designed for data scientists who currently use Python or R to work with smaller datasets on a single machine and who need to scale up their analyses and machine learning models to large datasets on distributed clusters. Data engineers and developers with some knowledge of data science and machine learning may also find this workshop useful.

## Who Should Attend

The workshop is designed for data scientists who currently use Python or R to work with smaller datasets on a single machine and who need to scale up their analyses and machine learning models to large datasets on distributed clusters. Data engineers and developers with some knowledge of data science and machine learning may also find this workshop useful.

# Learning Outcomes

---

**Upon successful completion of this course, participants will be able to:**

Overview of data science and machine learning at scale

Overview of the Hadoop ecosystem

Working with HDFS data and Hive tables using Hue

Introduction to Cloudera Data Science Workbench

Overview of Apache Spark 2

Reading and writing data

Inspecting data quality

Cleansing and transforming data

Summarizing and grouping data

Combining, splitting, and reshaping data

Exploring data

Configuring, monitoring, and troubleshooting Spark applications

Overview of machine learning in Spark MLlib

Extracting, transforming, and selecting features

Building and evaluating regression models

Building and evaluating classification models

Building and evaluating clustering models

Cross-validating models and tuning hyperparameters

Building machine learning pipelines

Deploying machine learning models

Spark, Spark SQL, and Spark MLlib

PySpark and sparklyr

Cloudera Data Science Workbench (CDSW)

Hue

## Additional Course Details

---

Nexus Humans Cloudera Data Scientist Training training program is a workshop that presents an invigorating mix of sessions, lessons, and masterclasses meticulously crafted to propel your learning expedition forward.

This immersive bootcamp-style experience boasts interactive lectures, hands-on labs, and collaborative hackathons, all strategically designed to fortify fundamental concepts.

Guided by seasoned coaches, each session offers priceless insights and practical skills crucial for honing your expertise. Whether you're stepping into the realm of professional skills or a seasoned professional, this comprehensive course ensures you're equipped with the knowledge and prowess necessary for success.

While we feel this is the best course for the Cloudera Data Scientist Training course and one of our Top 10 we encourage you to read the course outline to make sure it is the right content for you.

Additionally, private sessions, closed classes or dedicated events are available both live online and at our training centres in Dublin and London, as well as at your offices anywhere in the UK, Ireland or across EMEA.

# Frequently Asked Questions

---

## **Q: What delivery options are available for Cloudera Data Scientist Training?**

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
  - Traditional Instructor-Led Classroom Training (ILT)
  - On-site delivery at your offices anywhere in United Kingdom
  - Private dedicated courses customized for your team
- 

## **Q: How many CPD hours does this course provide?**

The 4-day Cloudera Data Scientist Training course provides up to 26.0 CPD hours of structured learning. CPD certificates can be provided upon request.

---

## **Q: What is the duration of the Cloudera Data Scientist Training training?**

The training takes place over 4 day(s), with each day lasting approximately 32.00 hours including breaks for lunch and refreshments.

---

## **Q: Do you provide corporate training for Cloudera Data Scientist Training?**

Yes, we provide corporate training, dedicated training, and closed classes for Cloudera Data Scientist Training. Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

---

## Q: Why choose Nexus Human for Cloudera Data Scientist Training?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

---

## Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your Cloudera Data Scientist Training training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

# Nexus Human

## Professional Training & Development

 Email: [info@nexushuman.com](mailto:info@nexushuman.com)

 Website: [www.nexushuman.com](http://www.nexushuman.com)

 Phone: +353 1 XXX XXXX (Ireland) | +44 20 XXXX XXXX (UK)