
CLOUDERA DATA SCIENCE WORKBENCH TRAINING

Accelerate data science in the enterprise

////////////////////////////////////
**Cloudera Data Science Workbench
enables fast, easy, and secure
self-service data science for the
enterprise.**
////////////////////////////////////

Cloudera has not only prepared us
for success today, but has also
trained us to face and prevail over
our big data challenges in the future.

Persado

Cloudera Data Science Workbench Training prepares learners to complete data science and machine learning projects using Cloudera Data Science Workbench (CDSW).

Get hands-on experience

Through narrated demonstrations and hands-on exercises, learners achieve proficiency in CDSW and develop the skills required to:

- _ Navigate CDSW's options and interfaces with confidence
- _ Create projects in CDSW and collaborate securely with other users and teams
- _ Develop and run reproducible Python and R code
- _ Customize projects by installing packages and setting environment variables
- _ Connect to a secure (Kerberized) Cloudera or Hortonworks cluster
- _ Work with large-scale data using Apache Spark 2 with PySpark and sparklyr
- _ Perform end-to-end machine learning workflows in CDSW using Python or R (read, inspect, transform, visualize, and model data)
- _ Measure, track, and compare machine learning models using CDSW's Experiments capability
- _ Deploy models as REST API endpoints serving predictions using CDSW's Models capability
- _ Work collaboratively using CDSW together with Git

What to Expect

This course is designed for learners at organizations using CDSW under an enterprise license or a trial license. The learner must have access to a CDSW environment on a Cloudera or Hortonworks cluster running Apache Spark 2. Some experience with data science using Python or R is helpful but not required. No prior knowledge of Spark or other Hadoop ecosystem tools is required.

Course Details

Overview of CDSW

- _Introduction to CDSW
- _Who Can Use CDSW
- _How to Access CDSW
- _Navigating around CDSW
- _User Settings
- _Hadoop Authentication

Projects in CDSW

- _Creating a New Project
- _Navigating around a Project
- _Project Settings

The CDSW Workbench Interface

- _Using the Workbench
- _Using the Sidebar
- _Using the Code Editor
- _Engines and Sessions

Running Python and R Code in CDSW

- _Running Code
- _Using the Session Prompt
- _Using the Terminal
- _Installing Packages
- _Using Markdown in Comments

Using Apache Spark 2 in CDSW

- _Scenario and Dataset
- _Copying Files to HDFS
- _Interfaces to Apache Spark 2
- _Connecting to Spark
- _Reading Data
- _Inspecting Data

Data Science and Machine Learning in CDSW

- _Transforming Data
- _Using SQL Queries
- _Visualizing Data from Spark
- _Machine Learning with MLlib
- _Session History

Experiments and Models in CDSW

- _Machine Learning Workflow
- _Running Experiments
- _Using Packages in Experiments
- _Deploying Models
- _Calling Models
- _Using Packages in Models

Teams and Collaboration in CDSW

- _Collaboration in CDSW
- _Teams in CDSW
- _Using Git for Collaboration
- _Conclusion

Visit our website and learn more at
university.cloudera.com