

Cloudera Data Science Workbench Training

Code: DATASWB-ON
URL: [View Online](#)

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

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

Who Can Benefit

- 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

- Collaboration in CDSW
 - Teams in CDSW
 - Using Git for Collaboration
 - Conclusion
-

Download Whitepaper: Accelerate Your Modernization Efforts with a Cloud-Native Strategy

Get Your Free Copy Now