

DSCI-272



Predicting with MLOps on Cloudera AI

Course Overview

Course Type

Instructor-Led Training

Level

Intermediate

Duration

4 days

Platform

Cloudera on premises
Cloudera on cloud

Topics Covered

- Projects and Sessions
- ML Runtimes
- Jobs and Pipelines
- Machine Learning Workflow
- Applications and AMPs
- Experiments and Models
- Model deployment
- Data Visualization
- CI/CD workflows
- Cloudera AI API
- Metrics and Monitoring
- Performance and GPUs

About This Training

Enterprise data science teams need collaborative access to business data, tools, and computing resources required to develop and deploy machine learning workflows. Cloudera AI, part of the Cloudera platform, provides the solution, giving data science teams the required resources.

This four-day course covers machine learning workflows and operations using Cloudera AI. Participants explore, visualize, and analyze data. You will also train, evaluate, and deploy machine learning models.

The course walks through an end-to-end data science and machine learning workflow based on realistic scenarios and datasets from a fictitious technology company. The demonstrations and exercises are conducted in Python (with PySpark) using Cloudera AI.

What Skills You Will Gain

Through lecture and hands-on exercises, you will learn how to:

- Utilize Cloudera SDX and other components of the Cloudera platform to locate data for machine learning experiments
- Use an Cloudera Accelerators for Machine Learning Projects (AMPs)
- Manage machine learning experiments
- Connect to various data sources and explore data
- Deploy an ML model as a REST API
- Manage and monitor deployed ML models

Who Should Take This Course?

The course is designed for data scientists who need to understand how to utilize Cloudera AI and the Cloudera platform to achieve faster model development and deliver production machine learning at scale. Data engineers, developers, and solution architects who collaborate with data scientists will also find this course valuable.

Other Training that Might Interest You

- Introducing Python
- Introducing Git

DSCI-272

Predicting with MLOps on Cloudera AI

Foundations

Git

- Introduction to Version Control: Understanding the importance of version control in collaborative environments
- Git Basics: Initialization, cloning, committing, pushing, and pulling
- Branching and Merging Strategies: Efficient collaboration techniques
- Hands-on: Creating and managing repositories

CI/CD

- Introduction to CI/CD Concepts: Continuous integration and deployment fundamentals
- Tools Overview: GitHub Actions
- Hands-on: Working with GitHub Actions
- Hands-on: Building a CI/CD pipeline with GitHub Actions

Docker

- Introduction to Containerization: Understanding container technology
- Docker Architecture and Components: Key elements of Docker
- Creating and Managing Docker Images and Containers: Practical usage
- Dockerfile Basics: Writing Dockerfiles
- Hands-on: Containerizing a simple application

Kubernetes

- Introduction to Container Orchestration: Kubernetes basics
- Kubernetes Architecture and Components: Core concepts
- Hands-on: Deploying Applications on Kubernetes: Practical deployment

Cloudera AI and MLflow

Introduction to Cloudera AI

- Overview of Cloudera AI: Introduction to key features and capabilities
- Navigating Cloudera AI Environment
- Hands-on: Creating and managing projects in Cloudera AI

Experiments in Cloudera AI

- Overview of MLflow: Key concepts and integration within Cloudera AI
- Experiments in Cloudera AI
- Hands-on: MLOps with MLflow

AI Registry

- Introduction: Overview of AI registry concepts
- Onboarding Walkthrough: Step-by-step guide to onboarding models
- Architecture Overview: Understanding the AI registry architecture

Working with Cloudera AI API

- Cloudera AI API Overview: Programmatically interacting with the Cloudera AI platform
- Using the Cloudera AI API: Managing projects, jobs, models, and applications via API
- Hands-on: Working with the Cloudera AI API Python client

DSCI-272

Predicting with MLOps on Cloudera AI

Advanced MLOps in Cloudera

MLOps in Cloudera AI

- Introduction to MLOps: Key concepts and principles
- MLOps Workflow: From development to production
- Challenges and Best Practices
- Hands-on:
 - Getting Connected and Set Up
 - Data Ingest, Exploration, and Model Training
 - Model Deployment and Model Operations
 - Model Registry and Model APIs
 - Model Management with Model Metric Store.

Monitoring ML Systems

- Continuous Model Monitoring with Evidently AI: Tracking model performance and detecting data drift
- Why Monitor Models?: Importance of model monitoring
- Fundamentals of Monitoring ML Systems: Core principles and best practices
- A Blueprint with Evidently & Cloudera AI
- Hands-on: Continuous model monitoring with Evidently AI

Configuring and Managing AI Workbenches

- Provisioning a Cloudera AI Workbench
- Cloudera AI Workbench Administration
- Cloudera AI Auto-Scaling
- Hands-on: Using Grafana dashboards for operational oversight

Advanced Topics in MLOps and Cloudera AI

Data Access and Lineage

- SDX Overview
- Data Catalog
- Authorization
- Lineage
- Hands-on: Data Access

Data Visualization in Cloudera AI

- Data Visualization Overview
- Cloudera Data Visualization Concepts
- Using Data Visualization in Cloudera AI
- Hands-on: Build a Visualization Application

Introduction to AMPs and the Workbench

- Editors and IDE
- Git
- Embedded Web Applications
- AMPs
- Hands-on: Streamlit

Autoscaling, Performance, and GPU Settings

- Autoscaling Workloads
- Working with GPUs
- Hands-on: Deep Learning with GPUs