Oracle Cloud Infrastructure Data Science Professional (2025): Hands-on Workshop

Oracle Cloud Infrastructure

DURATION	MODULES	COURSE CODE
3 Days	**6 Lectures**	**—**

Course Overview

The Oracle Cloud Infrastructure (OCI) Data Science Professional course is designed to equip Data Scientists, Machine Learning/AI Engineers with the skills and knowledge to effectively utilize OCI's comprehensive suite of data science tools. Spanning the entire machine learning lifecycle, from establishing a data science workspace to deploying and overseeing machine learning models in real-world environments, this course enables participants to harness OCI's robust capabilities. They will learn to construct, train, and deploy machine learning models, apply MLOps best practices, and integrate other OCI services to optimize their data science workflows.After completing this course, you should be able to: Identify OCI services to implement an ML solution for a business use case. Configure and manage OCI Data Science workspaces and projects. Utilize Accelerated Data Science (ADS) SDK for streamlined data science processes. Implement end-to-end machine learning lifecycle, including data preparation, model training, evaluation, and deployment. Apply MLOps practices to automate and monitor machine learning workflows. Integrate with other OCI services such as OCI Vault, OCI Object Storage, OCI Generative AI, OCI Data Flow and OCI Data Labeling. Be prepared to take OCI Data Science Professional 2025 certification.

What You Will Learn

Module 1: Welcome to Data Science

- Lesson 1: Data Science Professional Course Overview
- Course Speakers
- Course Contributors
- Target Audience
- Prerequisites
- What the Data Science Exam Validates
- Course Outline
- Data Science Demos and Labs
- Getting the Most Out of This Course
- Ratings and Feedback

Module 2: Introduction and Configuration

- Lesson 1: Data Science: Introduction
- Data Science and Machine Learning (ML) in History
- Data Science and ML Today
- Importance of Data Science and AI
- What is Oracle AI?
- OCI Services That Support AI and ML
- Oracle Cloud Infrastructure Data Science Overview
- Core Principles of OCI Data Science
- OCI Data Science Details
- Data Science Features and Terminology
- Ways to Access OCI Data Science
- Where to Find Data Science
- Lesson 2: ADS SDK Overview
- Accelerated Data Science (ADS) SDK Overview
- Ways to Access ADS SDK
- ADS SDK Features
- Data Visualization
- Feature Engineering
- Model Training
- Model Evaluation
- Model Interpretation and Explainability
- Model Deployment
- Lesson 3: Tenancy Configuration Basics
- Tenancy Configuration Concepts
- How Data Science Components Work Together
- Compartments & Creating Compartments
- User Groups & Dynamic Groups
- Policies (Required, Optional)
- Demo
- Lesson 4: Configure a Tenancy with OCI Resource Manager
- Automatic Configuration and Data Science Service Template
- What the Data Science Service Template Creates
- Running Oracle Resource Manager (ORM) Stack
- Accessing Terraform Script
- Demo
- Lesson 5: Networking for Data Science
- Cloud Networking Components Overview
- Data Science and Networking Connectivity
- Default Networking
- Custom Networking
- Demo
- Lesson 6: Authenticate to OCI APIs
- Importance of Authentication
- Authenticating Different Interfaces

- Resource Principals & Data Science Service Authentication
- OCI Configuration File & Format
- Demo

Module 3: Workspace Design and Setup

- Lesson 1: Projects
- What Are Projects?
- Create, View, Edit, Delete Projects
- Demo
- Lesson 2: Notebook Sessions
- What Are Notebook Sessions?
- Create, View, Edit, Delete Notebook Sessions
- Activate/Deactivate Notebook Sessions
- Notebook Session Metrics
- Demo
- Lesson 3: How to Work with JupyterLab
- JupyterLab Overview and Interfaces
- Features
- Demo
- Lesson 4: Conda Environments: Overview
- What is a Conda Environment?
- Benefits
- Environment Explorer
- Data Science Conda Environments (Published & Installed)
- Demo
- Lesson 5: Data Science Conda Environments
- Types and Families of Conda Environments
- Naming Conventions
- Special Use Cases: Computer Vision, ML, NLP, ONNX, PyTorch, PySpark, TensorFlow
- Modify & Publish
- Lesson 6: Manage Conda Environments
- Recap & Functionality
- Browse, Search, Install, Clone, Delete
- Create from YAML
- Demo
- Lesson 7: OCI Vault: Introduction
- Importance of OCI Vault in Data Science
- Vaults, Keys, Master and Data Encryption Keys
- Rotating Keys
- Secrets
- Lesson 8: Using OCI Vault in OCI Data Science
- Oracle-Managed and Customer-Managed Key Encryption
- Encrypt, Encode, Store, Retrieve Secrets
- Using OCI Vault with ADS
- Demo

- Lesson 9: Code Repositories (Git)
- Version Control Overview
- Centralized vs Distributed VCS
- Using Git in Data Science Workflows
- OCI Git Extension & Terminology
- Working with Remote Repo: GitHub
- Git Commands
- Demo

Module 4: Machine Learning Lifecycle

- Lesson 1: ML Lifecycle Overview

Machine Learning Lifecycle Stages: Data Access, Exploration, Feature Engineering, Modeling, Validation, Deployment, Monitoring

- Lesson 2: Access Data

Data Sources: Object Storage, Local Storage, Autonomous DB, MySQL, Amazon S3, HTTP/HTTPS

- DatasetBrowser, PyArrow
- Supported Sources & Formats
- Demo
- Lesson 3: Data Preprocessing
- Data Transformations, Cleaning, Imputation, Dummy Variables
- Outlier Detection, Feature Scaling, Dimensionality, Text Data
- ADS Automated Transformations
- Demo
- Lesson 4: Introduction to Feature Types
- Feature Types, Exploratory Data Analysis (EDA), Correlation Tables & Plots
- Lesson 5: Custom Feature Types
- Feature Statistics, Plots, Warnings, Validators, Creating Custom Feature Types
- Lesson 6: Data Visualization
- ADS Smart Visualization Tool, AutoML Pipeline
- Methods: corr(), show_in_notebook(), plot(), Seaborn, Matplotlib, GIS
- Data Profiling
- Lesson 7: Model Training
- Model Training Process, Libraries, Ways to Train
- Lesson 8: Oracle AutoML: Introduction
- AutoML Overview, Approaches, Benefits, Workflow
- Model Explanations
- Lesson 9: Hyperparameter Tuning: ADSTuner
- Search Spaces, Tuning Process, Custom Search Spaces
- Lesson 10: Model Evaluation
- ADS Evaluators: Binary, Multiclass, Regression
- Metrics, Charts, and Plots
- Lesson 11: Model Explanations: Global Explainer

Feature Permutation Importance, Feature Dependence Explanations, Accumulated Local Effects (ALE)

- Lesson 12: Model Explanations: Local and WhatIf Explainers
- Lesson 13: Model Catalog: Overview

Artifacts, Metadata, Input/Output Schema, Model Provenance, Introspection, Taxonomy

- Lesson 14: Model Serialization
- Saving Models via ADS SDK, Model Catalog Operations
- Lesson 15: Model Deployment
- Deployment Architecture, Create/Invoke Models, Monitoring, Managing Deployments
- Demo
- Lesson 16: LLM Training & LangChain Integration
- Fine-Tuning Pre-Trained Models, LangChain Integration
- Demo
- Lesson 17: OCI Data Science Operators
- Types, Demo
- Lesson 18: OCI Data Science AI Quick Actions
- AI Quick Actions Overview
- Demo

Module 5: MLOps Practices

- Lesson 1: MLOps Architecture
- Continuous Practices, Importance, Automation
- OCI MLOps Architecture
- Lesson 2: OCI Data Science Jobs
- Jobs Service, Life Cycle, Runs, Scaling, Batch/Distributed Inference
- Oracle Cloud Access for Jobs
- Lesson 3: Jobs Monitoring and Logging
- Job Metrics, Logs, Events, Rules, and Actions
- Lesson 4: Data Science Pipeline
- Pipeline Steps, Life Cycle, Demo Scenario
- Lesson 5: Model Deployment: Autoscaling
- Data Science Model Scaling Issues, Metric-Based Autoscaling

Module 6: Related OCI Services

- Lesson 1: Spark Applications, Data Flow, and Data Science
- OCI Data Flow Overview, Components, Capabilities, Security
- Spark Applications Configuration and Integration with Data Science
- Build and Train ML Models with Data Flow
- Best Practices
- Learn More About ADS
- Lesson 2: Oracle Open Data
- Overview, Benefits, Access
- Lesson 3: OCI Data Labeling
- What, Who Uses, Industry Use Cases
- Data Labeling in AI/ML Life Cycle
- Scenario: Data Labeling for AI
- Data Label Types and Integrations