Course Code: 6XL536G

Course Title: IBM Cloud Pak for Data (V4.x) Foundations

Description:

This learning offering tells a holistic story of Cloud Pak for Data and how you can extend the functions with services and integrations. You will explore some of the services and see how they enable effective collaboration across an organization. In this course, you will use Watson Knowledge Catalog, Data Virtualization, and Watson Studio (including Data Refinery and AutoAI). You will also examine some of the external data sets and industry accelerators that are available on the platform.

Objectives:

By the end of this course, you will be able to:

- Describe the Cloud Pak for Data implementation stack
- Summarize the Cloud Pak for Data workflow that implements the ModelOps process
- Construct a simple predictive model that reflects a typical Data Fabric solution
- Examine external data sets and industry accelerators that promote trustworthy AI
- Select services that align to the goals of a data-driven organization

Prerequisites:

Before you start this course, you should be able to complete the following tasks:

- Explain the purpose of Cloud Pak for Data and the value it brings to the business
- Describe its basic architecture
- State its deployment options
- Differentiate between Cloud Pak for Data and Red Hat OpenShift Container Platform
- Define the AI Ladder and its associated roles and services
- Identify the types of projects and how to collaborate on the platform
- Log in to Cloud Pak for Data and create an analytics project

You can review these skills in the *Solution Architect – Associate* learning path.

Duration:

6.4 Hrs

Topics:

Create an analytics project

- Summarize the ModelOps process
- Relate a process to a workflow
- Identify the predefined roles in Cloud Pak for Data

- Define analytics project
- Create an analytics project (data scientist)
- Request data (data scientist)

Add data to the project

- Respond to a data request
- Evaluate adding data from an integrated database versus data virtualization
- Differentiate between platform and service level connections
- Access an integrated database (data engineer)
- Create a catalog (data engineer)
- Connect to a data source (data engineer)
- Construct a virtualized table from a single data source (data engineer)

Organize the data

- Describe catalogs and their uses
- Summarize what you can do with the Watson Knowledge Catalog service
- List the types of governance artifacts
- Identify how to manage risk and regulatory challenges
- Profile data assets (data steward)
- Define a data protection rule (data steward)

Prepare the data

- List the ways to prepare data for use in projects
- Describe what you can do with Data Refinery
- Prepare data for modeling (data quality analyst)
- Validate data (data quality analyst)
- Visualize data (data quality analyst)
- Develop a Data Refinery flow (data quality analyst)
- Create a data set for modeling (data quality analyst)

Analyze the data and build a model

- Name the steps in the data analysis process
- List the criteria for choosing a modeling tool in analytics projects
- Summarize the AutoAI requirements
- Outline the AutoAI process
- Articulate the deployment process
- Describe how to use notebooks
- Build an AutoAI model (data scientist)
- Save an AutoAI pipeline model (data scientist)
- Deploy a model (data scientist)
- Save an experiment as a notebook

Expand to other scenarios

- Indicate how to monitor models
- List the aspects of trustworthy AI
- Identify how to collaborate with external stakeholders
- Describe how to extend Cloud Pak for Data functions
- Define scaling services
- Classify services
- List the most popular services from each category
- Associate Cloud Pak for Data use cases with the services that support them
- Explore solutions (solutions, services, external data sets and industry accelerators)

Audience:

Anyone who wants to gain foundational knowledge of IBM Cloud Pak for Data