databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Databricks Academy Course Catalog

FINAL UPDATE: June 17, 2024

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Welcome to the Databricks Academy

About the Databricks Academy

The Databricks Academy is the training arm of Databricks - our goal is to help our users achieve efficient and effective use of the Databricks Lakehouse Platform to reach their big data and AI goals.

Via the Databricks Academy, you can access self-paced e-learning and instructor-led courses that help you prepare for Databricks certification exams and focus on how to use the Databricks Lakehouse Platform for:

- Data analytics
- Data engineering
- Data science/machine learning
- Generative AI engineering
- Platform administration

Training Offerings

Self-paced e-learning is virtual training available 24/7 to individuals signed up for the Databricks Academy. Databricks customers and partners are granted access to self-paced e-learning for free. Non-Databricks customers and partners are able to purchase a subset of content available. Training currently includes mostly lectures and demos on how to use the Databricks Lakehouse Platform.

Instructor-led training is virtual training available to everyone (Databricks customers and partners and the general public) for a fee. Instructor-led training courses include roughly 8 to 16 hours of training and include lectures, demos, and hands-on labs.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Blended learning is a cohort-based, multi-week, virtual training available to everyone (Databricks customers and partners and the general public). It combines self-paced, self-directed study with once a week meetings with instructors.

Accreditations are 30-minute quizzes available via the Databricks Academy after completing a selection of Databricks courses or learning plans. Upon successful completion of an accreditation, badges are issued that can be shared on social media sites and professional networks.

Certifications are 1.5 to two hours exams available via our certification platform. Upon successful completion of an exam, badges are issued that can be shared on social media sites and professional networks, and validate your data and AI skills in the Databricks Lakehouse Platform.

Learning paths

Learning paths are designed to help guide users to the courses most relevant to them.

Current pathways are available for:

- Databricks fundamentals
- Generative AI Fundamentals
- Data Analysts
- Data Engineers
- Machine Learning Practitioners
- Generative AI Engineers
- Platform Administrators
- Platform Architects (AWS, Azure, Google Cloud)

Please log-in to the Databricks Academy to access these learning plans. You'll need to click on your active catalogs and navigate to catalog DB001 [FREE Self-Paced] Databricks Academy - Role-Based/Specialty Learning Path Catalog.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Databricks Academy Offerings

Certification exam/accreditations

For a full list of available certification exams/accreditations, along with their descriptions, please click here.

Instructor-led courses

For a full list of available instructor-led courses, along with their descriptions, please click here.

Self-paced courses

Note: All self-paced courses are free for Databricks customers and partners.
Non-customers can purchase some courses through role-based learning plans available via the Databricks Academy.

Access Shared Data Externally with Delta Sharing

Click here for the customer enrollment link.

Duration: 35 Minutes

Delta Sharing is an open protocol for secure data sharing with other organizations regardless of which computing platforms they use. Databricks provides both open source and managed options for Delta Sharing. Databricks-managed Delta Sharing allows data providers to share data and data recipients to access the shared data. With Delta Sharing, users can share collections of tables in a Unity Catalog metastore in real time without copying them, so that data recipients can immediately begin working with the latest version of the shared data.

This course will focus on sharing data externally and accessing shared data from external tools such as PowerBI and Apache Spark. First, we will demonstrate how to

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

share data externally. Then, we will show how to access the shared data from external tools.

Prerequisites:

- Beginning-level knowledge of the Databricks Lakehouse platform (high-level knowledge the structure and benefits of the Lakehouse platform)
- Basic understanding of how Unity Catalog is integrated into the Databricks Lakehouse Platform
- Beginning-level knowledge of python, PySpark API and pandas API
- Admin account privileges or have been granted required privileges for creating and managing shares

Learning objectives:

- Describe the external data sharing process with Databricks-managed Delta Sharing
- Share data externally with Databricks-managed Delta Sharing
- Access shared data in PowerBI
- Access shared data using PySpark
- Access shared data using pandas

Advanced Data Engineering with Databricks

Click here for the customer enrollment link ***- Updated January 2024***

Duration: 12 hours

Course description:In this course, participants will build upon their existing knowledge of Apache Spark, Delta Lake, and Delta Live Tables to unlock the full potential of the data lakehouse by utilizing the suite of tools provided by Databricks. This course places a heavy emphasis on designs favoring incremental data processing, enabling systems optimized to continuously ingest and analyze ever-growing data. By designing workloads that leverage built-in platform optimizations, data engineers can reduce the burden of code maintenance and on-call emergencies, and quickly adapt production code to new demands with

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

minimal refactoring or downtime. The topics in this course should be mastered prior to attempting the Databricks Certified Data Engineering Professional exam.

Prerequisites:

- Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc)
- Intermediate programming experience with PySpark
- Extract data from a variety of file formats and data sources
- Apply a number of common transformations to clean data
- Reshape and manipulate complex data using advanced built-in functions
- Intermediate programming experience with Delta Lake (create tables, perform complete and incremental updates, compact files, restore previous versions etc.)
- Beginner experience configuring and scheduling data pipelines using the Delta Live Tables (DLT) UI
- Beginner experience defining Delta Live Tables pipelines using PySpark
- Ingest and process data using Auto Loader and PySpark syntax
- Process Change Data Capture feeds with APPLY CHANGES INTO syntax
- Review pipeline event logs and results to troubleshoot DLT syntax

Learning objectives:

- Design databases and pipelines optimized for the Databricks Lakehouse Platform.
- Implement efficient incremental data processing to validate and enrich data driving business decisions and applications.
- Leverage Databricks-native features for managing access to sensitive data and fulfilling right-to-be-forgotten requests.
- Manage code promotion, task orchestration, and production job monitoring using Databricks tools.

Apache Spark Programming with Databricks

Click here for the customer enrollment link.

Duration: 12 hours

NOTE: This is an e-learning version of the Apache Spark Programming with

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Databricks instructor-led course. It is an on-demand recording available via the Databricks Academy and covers the same content as the instructor-led course. For more information about what's in the course itself, please visit this link.

Automate Production Workflows

Click here for the customer enrollment link.

Duration: 2 hours

Course description: In this course, you'll learn how to programmatically deploy batch and streaming workloads using workflow patterns for common use cases.

Prerequisites:

- Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc).
- Intermediate programming experience with PySpark:
    - Extract data from a variety of file formats and data sources
    - Apply a number of common transformations to clean data
    - Reshape and manipulate complex data using advanced built-in functions
- Intermediate programming experience with Delta Lake (create tables, perform complete and incremental updates, compact files, restore previous versions etc.)

Note: The data engineering skills required for this course can be learned by taking the Data

Engineering with Databricks course by Databricks Academy.

Learning objectives:

- Describe common workflow patterns to deploy batch and streaming workloads Configure task dependencies and task values for a multi-task workflow job Programmatically interact with workflow jobs using the Databricks API and CLIVersion control the configuration settings for a DLT pipeline using Files in Repos and the Databricks CLIDescribe how Databricks workspace/resource provisioning can be automated using Terraform to manage infrastructure as code

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

AWS Databricks Cloud Integrations

Click here for the customer enrollment link.

Duration: 2 hours

Course description: This 2 hour video series will walk you through three common integrations to help build out the capabilities of your AWS Databricks applications.

Prerequisites:

- Beginner-level knowledge of AWS (EC2, IAM, Kinesis, Redshift, S3)
- Access to your AWS console, with the ability to create buckets, Kinesis data streams, Redshift clusters, and IAM roles and policies
- Account administrator capabilities in your Databricks account

Learning objectives:

- Create your own S3 bucket and access data from Databricks
- Set up a data stream in Amazon Kinesis and stream records from Databricks
- Set up a data warehouse in Amazon Redshift, connect it to Databricks, and exchange data.

AWS Databricks Networking and Security Fundamentals

Click here for the customer enrollment link.

Duration: 1 hour

Course description: This 1 hour video series will provide you with the background needed to customize the structure of your environment and reinforce security at the infrastructural level.

Prerequisites:

- Beginner-level knowledge of AWS (IAM, KMS, S3, VPC)
- Knowledge of networking concepts (IPv4 addressing, DNS)

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Access to your AWS console, with the ability to create buckets, VPCs and their associated resources, and IAM roles and policies
- Account administrator capabilities in your Databricks account

Learning objectives:

- Create your own VPCs
- Deploy workspaces into your own managed VPCs
- Create your own customer-managed keys
- Apply customer-managed keys to achieve different levels of encryption in your Databricks workspaces

AWS Databricks Platform Administration Fundamentals

Click here for the customer enrollment link.

Duration: 1 hour

Course description: This 1 hour video series will provide you with a high-level overview of the AWS environment as it relates to Databricks, and it will guide you through how to perform some fundamental tasks related to deploying and managing Databricks workspaces in your organization.

Prerequisites:

- Beginner-level knowledge of AWS (IAM and S3)
- Access to your AWS console, with the ability to create buckets and IAM roles
- Account administrator capabilities in your Databricks account

Learning objectives:

- Describe and identify elements of the AWS Databricks architecture
- Create and manage workspaces and metastores, and supporting resources
- Automate administration operations

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Azure Databricks Cloud Integrations

Click here for the customer enrollment link.

Duration: 40 Minutes.

Course description: This course is designed to provide additional information about individual topics of integration in the Azure Cloud with Azure Databricks. These videos are stand-alone integration videos that build on the foundational concepts covered in the associated courses, Azure Databricks Workspace Administration Fundamentals and Azure Databricks Networking and Security Fundamentals.

Prerequisites:

- The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require the following privileges:-
- Azure Databricks Account administration
- Azure Cloud administration

Learning objectives:

- Describe the basics of establishing cloud integrations with an Azure Databricks workspace.
- Identify common integration methods used in an Azure Databricks workspace for specific cloud services integrations.
- Explain the importance of integrating a storage account with Azure Databricks.
- Connect an Azure storage account to an Azure Databricks workspace.
- Explain why you would use Azure Data Factory with Azure Databricks.
- Connect Azure Data Factory to an Azure Databricks workspace.
- Explain why you would use Power BI with Azure Databricks.
- Connect Power BI to an Azure Databricks workspace.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Azure Databricks Networking and Security Fundamentals

Click here for the customer enrollment link.

Duration: 1 hour 10 minutes

Course description: This course focuses on providing you with a foundation for the networking and security needs for an Azure Databricks workspace in your Azure Cloud ecosystem. You'll be able to explore how identity and access is managed through Azure Active Directory and Unity Catalog. Additionally, you'll be able to review some foundational networking concepts and how they are applicable to the Azure Databricks environment, such as Azure Software Defined Networks, CIDR ranges, subnets, and VNet peering. You'll also explore how Azure Databricks workspaces can be secured through IP Access List, User Defined Routes, private and service endpoints, and private DNS zones to support Data Loss Prevention strategies.

Prerequisites:

- "The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require the following privileges:
- Account administration
- Cloud administration
- Additional prerequisites include a solid understanding of IPv4 addresses, subnets, CIDR ranges, and general networking concepts."

Learning objectives:

- Describe components of the Azure Databricks platform architecture and deployment model.
- Explain network security features including no public IP address, Bring Your Own VNET, VNET peering, and IP access lists.
- Describe identity provider and Azure Active Directory integrations and access control configurations for an Azure Databricks workspace.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Explain encryptions and permissions available for data protection, such as identity provider authentication, secrets, and table access control.
- Describe security standards and configurations for compliance, including cluster policies, Bring Your Own Key, and audit logs.

Azure Databricks Workspace Administration Fundamentals

Click here for the customer enrollment link.

Duration: 1 hour 10 minutes

Course description: This course is designed to introduce you to the fundamentals of of Azure Databricks Workspace Administration including the reference architecture and deployment options for your workspace. You'll also be introduced to some of the resources necessary to deploy an Azure Databricks Workspace in your environment.

Prerequisites:

- The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require the following privileges:
- Account administration
- Cloud administration
- Additional prerequisites include a solid understanding of IPv4 addresses, subnets, CIDR ranges, and general networking concepts."

Learning objectives:

- "Explain the first-party service relationship Databricks has with Microsoft.
- Identify the responsibilities of the Platform Administrator/Platform Architect with an Azure Databricks implementation.
- Describe foundational concepts of the Azure cloud ecosystem.
- Identify how Azure Databricks is part of the Azure ecosystem.
- Describe additional resources that may be included with Azure Databricks in an Azure based architecture.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Recognize the impact of Azure Databricks on the platform's cost management and planning.
- Review the decisions necessary to implement Azure Databricks for your architecture.
- Identify the resources needed to implement an Azure Databricks workspace.
- Create necessary backing resources for Azure Databricks.
- Differentiate between the available options for deploying an Azure Databricks workspace.
- Determine the impact of networking for Azure on your workspace.
- Deploy an Azure Databricks workspace using the default method.
- Deploy an Azure Databricks workspace with VNet Injection.
- Describe how Terraform can automate the deployment process of Azure Databricks.

Build Data Pipelines with Delta Live Tables

Click here for the customer enrollment link.

Duration: 3 hours

Course description: Use Delta Live Tables with Spark SQL and Python to define and schedule pipelines that incrementally process new data from a variety of data sources into the Lakehouse. Note: These lessons are a subset of the Data Engineering with Databricks course.

Prerequisites:

- Beginner familiarity with cloud computing concepts (virtual machines, object storage, etc.)
- Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc)
- Beginning programming experience with Delta Lake
- Use Delta Lake DDL to create tables, compact files, restore previous table versions, and perform garbage collection of tables in the Lakehouse.
- Use CTAS to store data derived from a query in a Delta Lake table

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Use SQL to perform complete and incremental updates to existing tables.
- Beginning programming experience with Spark SQL or PySpark
- Extract data from a variety of file formats and data sources
- Apply a number of common transformations to clean data
- Reshape and manipulate complex data using advanced built-in functions
- Production experience working with data warehouses and data lakes

Learning objectives:

- Describe how Delta Live Tables tracks data dependencies in data pipelines.
- Configure and run data pipelines using the Delta Live Tables UI.
- Use Python or Spark SQL to define data pipelines that ingest and process data through multiple tables in the lakehouse using Auto Loader and Delta Live Tables.
- Use APPLY CHANGES INTO syntax to process Change Data Capture feeds.
- Review event logs and data artifacts created by pipelines and troubleshoot DLT syntax

CI/CD Administration in Databricks

Click here for the customer enrollment link.

Duration: 45 Minutes

Course description: This Continuous Integration and Delivery Administration in Databricks course will talk about the various elements provided in the Databricks environment that can help fulfill the requirements of your organization's CI/CD processes.

Prerequisites:

- The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require account administration, workspace administration, and/or metastore ownership.

Learning objectives:

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Describe roles and responsibilities related to administering the Databricks platform
- Administer metastores, workspaces and supporting resources
- Automate administration operations
- Describe Databricks identities and how they apply across the platform
- Configure groups, users and service principals
- Automate user and group provisioning
- Configure workspace settings
- Secure access to workspace assets
- Use Databricks secrets to implement security best practices in your organization
- Describe compute options and features available in the Databricks platform
- Secure access to compute resources
- Secure access to Databricks from a BI tool
- Describe data access patterns in Databricks
- Define data access rules and manage data ownership
- Secure access to external storage
- Upgrade legacy data assets to Unity Catalog
- Describe features offered by the Databricks platform to support continuous integration and deployment
- Implement revision control in the workspace
- Schedule execution of a Data Science and Engineering workloads using Databricks Workflows
- Automate actions in response to code updates

Common Applications with Large Language Models

Click here for the customer enrollment link.

Duration: 50 Minutes

Course description: Welcome to Common Applications with Large Language Models! Please note: This module is a subset of the larger Large Language Models with Databricks course available in the Databricks Academy.

databricks Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Prerequisites:

- Experience/familiarity with NLP
- Working knowledge of machine learning and deep learning is helpful.

Learning objectives:

- Describe the components of common LLM applications.
- List popular tools in the NLP ecosystem.
- Explain the importance of Hugging Face in the NLP environment.
- Describe how to select a model for your application.
- Use a variety of existing models for a variety of common applications.

Compute Resources and Unity Catalog

Click here for the customer enrollment link.

Duration: 30 Minutes

Course description: Unity Catalog is a central hub for administering and securing your data which enables granular access control and built-in auditing across the Databricks platform. This course guides learners through creating compute resources capable of accessing Unity Catalog.

Prerequisites:

- This course has no specific course prerequisites.

Learning objectives:

- Describe how to access Unity Catalog through Databricks compute resources
- Create a Unity Catalog enabled cluster
- Access Unity Catalog through Databricks SQL

Data Access Control in Unity Catalog

Click here for the customer enrollment link.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Duration: 1 hour

Course description: Unity Catalog is a central hub for administering and securing your data which enables granular access control and built-in auditing across the Databricks platform. This course guides learners through techniques for creating and governing data objects in Unity Catalog.

Prerequisites:

- This course has no specific course prerequisites

Learning objectives:

- Describe the security model for governing data objects in Unity Catalog
- Define data access rules and manage data ownership
- Secure access to external storage

Data Administration in Databricks

Click here for the customer enrollment link.

Duration: 1.50 hours

Course description: This Data Administration in Databricks course will provide you with a basis of how data is managed in the Databricks platform with a feature known as Unity Catalog.

Prerequisites:

- The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require account administration, workspace administration, and/or metastore ownership.

Learning objectives:

- Describe data access patterns in Databricks
- Define data access rules and manage data ownership
- Secure access to external storage

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Upgrade legacy data assets to Unity Catalog

~~Data Analysis with Databricks SQL~~

~~Click here for the customer enrollment link - *Updated January 2024*~~

~~Duration: 6 hours~~

~~NOTE: This is an e-learning version of the Data Analysis with Databricks instructor-led course. It is an on-demand recording available via the Databricks Academy and covers the same content as the instructor-led course. For more information about what's in the course itself, please visit this link.~~

Data Engineering with Databricks

Click here for the customer enrollment link.

Duration: 12 hours

NOTE: This is an e-learning version of the Data Engineering with Databricks instructor-led course. It is an on-demand recording available via the Databricks Academy and covers the same content as the instructor-led course. For more information about what's in the course itself, please visit this link.

Data Preparation for Machine Learning

Click here for the customer enrollment link for the free course (no labs). If you would like to purchase this course with labs, please speak to your Databricks representative about a Databricks Academy lab subscription.

Duration: 4 Hours

Course description: This course focuses on the fundamentals of preparing data for machine learning using Databricks. Participants will learn essential skills for exploring, cleaning, and organizing data tailored for traditional machine learning

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

applications. Key topics include data visualization, feature engineering, and optimal feature storage strategies. Through practical exercises, participants will gain hands-on experience in efficiently preparing data sets for machine learning within the Databricks. This course is designed for associate-level data scientists and machine learning practitioners. and individuals seeking to enhance their proficiency in data preparation, ensuring a solid foundation for successful machine learning model deployment.

Prerequisites:

- Familiarity with Databricks workspace and notebooks
- Familiarity with Delta Lake and Lakehouse
- Intermediate level knowledge of Python

Learning objectives:

- Describe Databricks Data Intelligence Platform and its features for machine learning.
- Explain data storage and governance features on Databricks.
- Perform exploratory data analysis and feature engineering using Spark and integrated visualization tools.
- Perform data pre-processing for missing data handling, data encoding and data standardization.
- Utilize Feature Store for storing and retrieving features.

Databricks Compute Resource Administration

Click here for the customer enrollment link.

Duration: 1 Hour

Course description: This Databricks Compute Resource Administration course, along with its accompanying labs, will show you how to create clusters and SQL warehouses, going through some of the important parameters and their meanings.

Prerequisites:

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require account administration, workspace administration, and/or metastore ownership.

Learning objectives:

- Create and configure clusters and SQL warehouses
- Control access to clusters and SQL warehouses
- Manage costs associated with compute resources
- Optimize clusters for performance and cost

Databricks, Generative AI, and Natural Language Processing

Click here for the customer enrollment link.

Duration: 35 Minutes

Course description: Welcome to Databricks, Generative AI, and Natural Language Processing! Please note: This module is a subset of the larger Large Language Models with Databricks course available in the Databricks Academy (just search for the title).

Generative AI is a type of artificial intelligence focused on the ability of computers to use models to create content like images, text, code, and synthetic data. In this module, we'll give you a gentle introduction to generative AI concepts, including a dive into the Databricks generative AI and LLM vision. Then, we'll focus the majority of the course on natural language processing (NLP), of which LLMs often provide a basis for.

Prerequisites:

- Intermediate-level experience with Python
- Working knowledge of machine learning and deep learning is helpful

Learning objectives:

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Explain fundamental concepts about generative AI.
- Give examples of LLM business use cases.
- Explain fundamental concepts about natural language processing.
- Summarize the Databricks generative AI vision.

Databricks Identity Administration

Click here for the customer enrollment link.

Duration: 40 Minutes

Course description: This Databricks Identity Administration course will provide you with a basis of how identity and access management is done in the Databricks platform.

Prerequisites:

- The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require account administration, workspace administration, and/or metastore ownership.

Learning objectives:

- Describe Databricks identities and how they apply across the platform
- Configure groups, users and service principals
- Automate user and group provisioning

Databricks Platform Administrator Fundamentals

Click here for the customer enrollment link.

Duration: 1.25 hours

Course description:

The Databricks Platform Administration Fundamentals course will provide you with a high-level overview of the Databricks platform, targeting the specific needs of platform administrators.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Prerequisites:

The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require account administration capabilities.

Learning objectives:

- Describe and identify the roles and responsibilities related to administering the Databricks platform
- Automate interactions with the Databricks platform

Databricks Workspace Administration and Security

Click here for the customer enrollment link.

Duration: 45 Minutes

Course description: This Databricks Workspace Administration and Security course, along with its accompanying labs, will show you how to configure the workspace using the Workspace Admin console and the SQL admin console.

Prerequisites & Requirements

- Prerequisites
    - The course is primarily demo-based, but learners can follow along assuming they have sufficient privileges. Some exercises require account administration, workspace administration, and/or metastore ownership.

Learning objectives

- Configure workspace settings
- Configure workspace groups, users and service principals
- Employ access control to secure workspace assets
- Implement and advocate security best practices in your notebooks

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Data Privacy and Governance Patterns

Click here for the customer enrollment link.

Duration: 2 hours

Course description: In this course, you'll learn how to apply patterns to securely store and delete personal information for data governance and compliance in the Lakehouse.

Prerequisites & Requirements

- Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc).
- Intermediate programming experience with PySpark:
- Extract data from a variety of file formats and data sources
- Apply a number of common transformations to clean data
- Reshape and manipulate complex data using advanced built-in functions
- Intermediate programming experience with Delta Lake (create tables, perform complete and incremental updates, compact files, restore previous versions etc.)
- Note: The data engineering skills required for this course can be learned by taking the Data Engineering with Databricks course by Databricks Academy.

Learning objectives

- Store sensitive data appropriately to simplify granting access and processing deletes.
- Process deletes to ensure compliance with the right to be forgotten.
- Perform data masking and configure fine-grained access control to configure appropriate privileges to sensitive data.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Delta Sharing Best Practices

Click here for the customer enrollment link.

Duration: 2 hours

Course description: Delta Sharing is an open protocol for secure data sharing with other organizations regardless of which computing platforms they use. Databricks provides both open source and managed options for Delta Sharing. Databricks-managed Delta Sharing allows data providers to share data and data recipients to access the shared data.

While data sharing brings a wide range of opportunities for data practitioners, the performance and security concerns become apparent. Thus, it is crucial to ensure the security and performance of the shared. This course will focus on the most important best practices for data sharing with Databricks-managed Delta Sharing.

Prerequisites:

- Beginning-level knowledge of the Databricks Lakehouse platform (high-level knowledge the structure and benefits of the Lakehouse platform)
- Basic understanding of how Unity Catalog is integrated into the Databricks Lakehouse Platform
- Admin account privileges or have been granted required privileges for creating and managing shares
- Databricks CLI (only for configuring IP access list)

Learning objectives:

- Describe data sharing best practices with Delta Sharing
- Configure and manage data access permissions
- Configure and manage recipient tokens
- Share and access partial data
- Enable audit logging for shared data

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

~~Deep Learning with Databricks~~

~~Click here for the customer enrollment link.~~

~~Duration: 12 hours~~

~~NOTE: This is an e-learning version of the Deep Learning with Databricks instructor-led course. It is an on-demand recording available via the Databricks Academy and covers the same content as the instructor-led course. For more information about what's in the course itself, please visit this link.~~

Deploy Workloads with Databricks Workflows

Click here for the customer enrollment link.

Duration: 3 hours

Course description: Moving a data pipeline to production means more than just confirming that code and data are working as expected. By scheduling tasks with Databricks Jobs, applications can be run automatically to keep tables in the Lakehouse fresh. Using Databricks SQL to schedule updates to queries and dashboards allows quick insights using the newest data. In this course, students will be introduced to task orchestration using the Databricks Workflow Jobs UI. Optionally, they will configure and schedule dashboards and alerts to reflect updates to production data pipelines. Note: These lessons are a subset of the Data Engineering with Databricks course.

Prerequisites:

- Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc)
- Ability to configure and run data pipelines using the Delta Live Tables UI
- Beginner experience defining Delta Live Tables (DLT) pipelines using PySpark
- Ingest and process data using Auto Loader and PySpark syntax
- Process Change Data Capture feeds with APPLY CHANGES INTO syntax
- Review pipeline event logs and results to troubleshoot DLT syntax

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Production experience working with data warehouses and data lakes.

Learning objectives:

- Orchestrate tasks with Databricks Workflow Jobs.
- Use Databricks SQL for on-demand queries.
- Configure and schedule dashboards and alerts to reflect updates to production data pipelines.

Evaluating Large Language Models

Click here for the customer enrollment link.

Duration: 45 Minutes

Course description: Welcome to Evaluating Large Language Models! Please note: Please note: This module is a subset of the larger Large Language Models with Databricks course available in the Databricks Academy (just search for the title).

Once you've trained and fine-tuned your LLMs, it's time to evaluate how they're doing. In this module, we'll discuss methods for evaluating traditional ML models.

Prerequisites:

- Intermediate-level experience with Python
- Working knowledge of machine learning and deep learning is helpful

Learning objectives:

- Compare and contrast the evaluation of traditional ML models and LLMs.
- Describe how LLMs are generally evaluated, using a variety of metrics.

Fine-Tuning Large Language Models

Click here for the customer enrollment link.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Duration: 45 Minutes

Course description: Welcome to Fine-Tuning Large Language Models! Please note: This module is a subset of the larger Large Language Models with Databricks course available in the Databricks Academy (just search for the title).

Fine-tuning LLMs is all about how we can improve the behavior of our models. In this module, we'll highlight different methods for fine-tuning models, dive into the pros and cons of each, discuss the importance of using your own data in the model-tuning process, and explore training/fine-tuning tools.

Prerequisites:

- Intermediate-level experience with Python
- Working knowledge of machine learning and deep learning is helpful

Learning objectives:

- Explain when and how to fine-tune models.
- Discuss the advantages and disadvantages of each method for fine-tuning models.
- Discuss the advantages of building models using your own data
- Describe common tools for training and fine-tuning, such as those from Hugging Face and DeepSpeed.

Get Started with Databricks for Data Engineering

Click here for the customer enrollment link.

Duration: 3 hours

Course description: In this course, you will learn basic skills that will allow you to use the Databricks Lakehouse Platform to perform a simple data engineering workflow. You will be given a tour of the workspace, and you will be shown how to work with notebooks. You will create a basic data engineering workflow while you perform tasks like creating and using compute resources, working with repositories, and creating and managing basic workflow jobs. The course will also introduce you to

databricks
Academy

Please note – with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Databricks SQL. Finally, you will see how data is stored, managed, governed, and secured within the lakehouse.

Prerequisites & Requirements

- Prerequisites
    - Basic knowledge of data engineering topics such as extraction, cleaning (and other transformations), and loading

Learning objectives

- Explain fundamental concepts about using the Databricks Lakehouse. Platform for new users responsible for data engineering workflows.
- Perform basic notebook tasks using the Databricks Lakehouse Platform.
- Manage Delta tables using the Databricks Lakehouse Platform.
- Describe features available through the Databricks Lakehouse Platform to secure and govern data.
- Use Workflow Jobs within the Databricks Lakehouse Platform to automate a basic data engineering workflow.
- Use Databricks to complete a simple data engineering workflow..

Get Started with Databricks for Business Leaders

Click here for the customer enrollment link.

Duration: 1.5 hours

Course description: The Databricks Data Science and Engineering Workspace (Workspace) provides a collaborative analytics platform to help data practitioners get the most out of Databricks when it comes to data science and engineering tasks. This course guides practitioners through fundamental Workspace concepts and components necessary to achieve a basic development workflow.

Prerequisites & Requirements

- Prerequisites

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Beginning-level knowledge of the Databricks Lakehouse platform (high-level knowledge the structure and benefits of the Lakehouse platform)
- Intermediate-level knowledge of Python (good understanding of the language as well as ability to read and write code)
- Beginning-level knowledge of SQL (ability to understand and construct basic queries)

Learning objectives

- Describe the Databricks architecture and the services it provides.
- Navigate the Databricks Data Science and Engineering Workspace.
- Create and manage Databricks clusters for running code.
- Manage data using the Databricks File System and Delta Lake.
- Create and run Databricks Notebooks.
- Schedule non-interactive execution of Databricks Notebooks using Databricks Jobs.
- Integrate a hosted Git service for revision control using Databricks Repos.

Get Started with Databricks for Machine Learning

Click here for the customer enrollment link.

Duration: 4 hours

Course description:

In this course, you will learn basic skills that will allow you to use the Databricks Data Intelligence Platform to perform a simple data science and machine learning workflow. You will be given a tour of the workspace, and you will be shown how to work with notebooks. You will train a baseline model with AutoML and transition the best model to production. Finally, the course will also introduce you to MLflow, feature store and workflows and demonstrate how to train and manage an end-to-end machine learning lifecycle.

Prerequisites & Requirements

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Basic knowledge of data science and machine learning topics such as regression/classification models, model evaluation metrics and machine learning libraries such as scikit-learn.

Learning objectives

- Explain fundamental concepts about using the Databricks Lakehouse Platform for machine learning.
- Perform basic notebook tasks using the Databricks Lakehouse Platform.
- Store and manage data in the Lakehouse for machine learning tasks.
- Describe features available through the Databricks Lakehouse Platform for end-to-end machine learning development.
- Create and use a baseline model using AutoML.
- Create and use a feature store table for model training.
- Track, register and manage the stage of a model with MLflow.
- Use a registered model for batch and real-time inference.

Getting Started with Data Analysis on Databricks

Click here for the customer enrollment link.

Duration: 1 hour

Course description: This course is designed to introduce Business Leaders to Databricks and the Databricks Lakehouse Platform. They will learn about the benefits the lakehouse provides to their businesses through this introductory content. This content will cover high-level, business impacting topics of value to a business leader and will not go into technical depth on Databricks products.

Prerequisites & Requirements

- Prerequisites
    - None as this is an introductory course.

Learning objectives

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Describe Databricks and the Databricks Lakehouse Platform, it's services, the availability of ISV partnership, and how to begin migrating to Databricks.
- Identify members of the Databricks customer account team that will be interacting with you throughout your customer journey.
- Locate where to find further information and training on the Databricks Lakehouse Platform.
- Describe how Databricks promotes a secure data environment that can be easily governed and scaled.
- Explain how, as an organization using Databricks, your company will be able to reduce their total cost of ownership of data management solutions by using Databricks.
- Define common data science terms used by Databricks when discussing the Databricks Lakehouse Platform.

GCP Databricks Platform Administration Fundamentals

Click here for the customer enrollment link.

Duration: 1.30 hours

Course description: This video series will provide you with a high-level overview of the Google Cloud Platform (GCP) environment as it relates to Databricks, and it will guide you through how to perform some fundamental tasks related to deploying and managing Databricks workspaces in your organization through GCP.

Prerequisites:

- Beginner-level knowledge of GCP
- Beginner-level knowledge of Terraform is helpful
- Access to a GCP project is required, with the ability to add principals and service accounts
- Account administrator capabilities in your Databricks account

Learning objectives:

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Describe and identify elements of the GCP Databricks architecture
- Create and manage workspaces and metastores, and supporting resources
- Automate administration operations

GCP Databricks Networking and Security Fundamentals

Click here for the customer enrollment link.

Duration: 1 hour

Course description: This video series will provide you with the background needed to customize the structure of your environment and reinforce security at the infrastructural level.

Prerequisites:

- Beginner-level knowledge of GCP
- Knowledge of networking concepts (IPv4 addressing, DNS).
- Account administrator capabilities in your Databricks account
- Access to your a GCP project, with the ability to enable APIs and create service accounts and VPCs

Learning objectives:

- Create your own VPCs.
- Deploy workspaces into your own managed VPCs.
- Create your own customer-managed keys.
- Apply customer-managed keys to achieve different levels of encryption in
- your Databricks workspaces.

GCP Databricks Cloud Integrations

Click here for the customer enrollment link.

Duration: 1 hour

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Course description: This video series will provide you with the background needed to customize the structure of your environment and reinforce security at the infrastructural level.

Prerequisites:

- Beginner-level knowledge of GCP
- Account administrator capabilities in your Databricks account
- Access to your a GCP project, with the ability to enable APIs and create service accounts and buckets

Learning objectives:

- Create your own external GCP bucket and access data from Databricks.
- Set up a topic and subscription in Google Pub/Sub (Lite) and stream messages from Databricks.
- Set up a data warehouse in Google BigQuery, connect it to Databricks, and exchange data.

Generative AI Fundamentals

Click here for the customer enrollment link.

Duration: 2 hours

Course description: Welcome to Generative AI Fundamentals. This course provides an introduction to how organizations can understand and utilize generative artificial intelligence (AI) models. First, we'll start off with a quick introduction to generative AI - we'll discuss what it is and pay special attention to large language models, also known as LLMs. Then, we'll move into how organizations can find success with generative AI - we'll take a deeper dive into what LLM applications are, discuss how Lakehouse AI can help you succeed, and discuss essential considerations for adopting AI in general. Finally, we'll tackle important aspects to consider when evaluating the potential risks and challenges associated with using/adopting generative AI.

databricks Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Prerequisites:

- None

Learning objectives:

- By the end of this course you will be able to:-
- Describe how generative artificial intelligence (AI) is being used to revolutionize practical AI applications.
- Describe how organizations can find success with generative AI applications.
- Recognize the potential legal and ethical considerations of using generative AI applications.

Generative AI Engineering with Databricks

Click here for the customer course enrollment link.

Duration: 16 hours

Course description: This course, updated June 2024, is the latest version of the Generative AI Engineering with Databricks course. It is aimed at data scientists, machine learning engineers, and other data practitioners looking to build generative AI applications with the latest and most popular frameworks and Databricks capabilities.

The four modules included in this course (with their learning objectives) are:

Generative AI Solution Development

- Describe RAG architecture.
- Use Mosaic AI Playground to explore the significance of contextual information.
- Prepare data for generative AI solutions.
- Connect data preparation for generative AI solutions to building a RAG architecture.
- Describe fundamental concepts about context embedding, vectors, vector databases, and the utilization of Mosaic AI Vector Search.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Generative AI Application Development

- Explain how to decompose a problem into its components and select the most suitable model for each step to enhance business use cases.
- Construct a multi-stage reasoning chain utilizing LangChain and HuggingFace transformers.
- Design an autonomous agent using generative models on Databricks.

Generative AI Application Evaluation and Governance

- Explain the meaning behind and motivation for building evaluation and governance/security systems.
- Explain Databricks Data Intelligence Platform features for LLM evaluation and governance.
- Describe evaluation techniques for specific components and types of applications.
- Analyze entire AI systems with respect to performance and cost.

Generative AI Application Deployment and Monitoring

- Explain best practices for deploying generative AI applications using tools like Model Serving.
- Explain how to operationalize generative AI applications following best practices and recommended architectures.
- Use Lakehouse Monitoring to monitor generative AI applications and their components.

Prerequisites:

- Familiarity with natural language processing concepts
- Familiarity with prompt engineering/prompt engineering best practices
- Familiarity with the Databricks Data Intelligence Platform
- Familiarity with RAG (preparing data, building a RAG architecture, concepts like embedding, vectors, vector databases, etc.)
- Experience with building LLM applications using multi-stage reasoning LLM chains and agents

databricks
Academy

Please note - with the launch of of our [learning catalog](), this course catalog will no longer be updated as of June 17, 2024.

- Familiarity with Databricks Data Intelligence Platform tools for evaluation and governance.

Incremental Processing with Spark Structured Streaming

Click here for the customer enrollment link.

Duration: 2 hours

Course description: In this course, you'll learn how to incrementally process data to power analytic insights with Structured Streaming and Auto Loader.

Prerequisites:

Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc).

Intermediate programming experience with PySpark:

Extract data from a variety of file formats and data sources

Apply a number of common transformations to clean data

Reshape and manipulate complex data using advanced built-in functions

Intermediate programming experience with Delta Lake (create tables, perform complete and incremental updates, compact files, restore previous versions, etc.)

Note: The data engineering skills required for this course can be learned by taking the Data Engineering with Databricks course by Databricks Academy.

Learning objectives:

- Describe the computation model used by Spark Structured Streaming.
- Configure required options to perform a streaming read on a source.
- Describe the requirements for end-to-end fault tolerance. Configure required options to perform a streaming write to a sink.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Perform aggregation on a streaming dataset and describe watermarking.

Introduction to Delta Sharing

Click here for the customer enrollment link.

Duration: 20 minutes

Course description: Delta Sharing is an open protocol for secure data sharing with other organizations regardless of which computing platforms they use. Databricks provides both open source and managed options for Delta Sharing. Databricks-managed Delta Sharing allows data providers to share data and data recipients to access the shared data. It can share collections of tables in a Unity Catalog metastore in real time without copying them, so that data recipients can immediately begin working with the latest version of the shared data.

This course will give an overview of Delta Sharing with an emphasis on the benefits of Delta Sharing over other methods of data sharing and how Delta Sharing fits into the Lakehouse architecture. Then, there will be a demonstration on how to configure Delta Sharing on a metastore and how to enable external data sharing.

Prerequisites:

- Beginning-level knowledge of the Databricks Lakehouse platform
- Basic understanding of how Unity Catalog is integrated into the Databricks Lakehouse Platform
- Metastore admin or account admin privileges (optional, but needed if following along with the configuration demo)

Learning objectives:

- Describe the benefits of Delta Sharing compared to traditional data sharing systems
- Describe features and use cases of Delta Sharing
- Define key components of Delta Sharing
- Configure Delta Sharing on a metastore

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Enable external data sharing for an account

~~Introduction to Photon~~

~~Click here for the customer enrollment link.~~

~~Duration: 30 minutes~~

~~Course description: In this course, you'll learn how Photon can be used to reduce Databricks total cost of ownership (TCO) and dramatically improve query performance. You'll also learn best practices for when to use and not use Photon Finally, the course will include a demonstration of a query run with and without Photon to show improvement in query performance.~~

~~Prerequisites:~~

- ~~Administrator privileges~~
- ~~Introductory knowledge about the Databricks Lakehouse Platform (what the Databricks Lakehouse Platform is, what it does, main components, etc.)~~

~~Learning objectives:~~

- ~~Explain fundamental concepts about Photon on Databricks.~~
- ~~Describe the benefits of enabling Photon on Databricks.~~
- ~~Identify queries that would benefit from using Photon~~
- ~~Describe the performance differences between a query run with and without Photon enabled~~

Introduction to Python for Data Science and Data Engineering

Click here for the customer enrollment link.

Duration: 12 hours

NOTE: This is an e-learning version of the Introduction to Python for Data Science and Data Engineering instructor-led course. It is an on-demand recording available

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

via the Databricks Academy and covers the same content as the instructor-led course. For more information about what's in the course itself, please visit this link.

Introduction to Unity Catalog

Click here for the customer course enrollment link.

Duration: 1 hour

Course description: Unity Catalog is a central hub for administering and securing your data which enables granular access control and built-in auditing across the Databricks platform. This course guides learners through fundamental Unity Catalog concepts and tasks.

Prerequisites:

- This course has no specific course prerequisites.

Learning objectives:

- Describe key concepts related to Unity Catalog
- Describe how Unity Catalog is integrated with the Databricks platform
- Perform a variety of tasks with Unity Catalog (managing users and groups, creating a Unity Catalog-enabled cluster, accessing Unity Catalog through Databricks SQL, create and govern table assets, limit table access with views, govern file access)

Large Language Models Operations

Click here for the customer course enrollment link.

Duration: 45 minutes

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Course description: Welcome to Large Language Models Operation Models! Please note: This module is a subset of the larger Large Language Models with Databricks course available in the Databricks Academy (just search for the title).

In this module we'll discuss LLMOps and as the Databricks product changes, notable developments in the product itself. We'll start with an overview of DevOps concepts, ensuring a comprehensive understanding, before diving into MLOps. Then, we'll discuss topics like cost performance, deployment options, monitoring, and more.

Prerequisites:

- Intermediate-level experience with Python
- Working knowledge of machine learning and deep learning is helpful

Learning objectives:

- Discuss how traditional MLOps can be adapted for LLMs
- Review an MLOps end-to end workflow and architecture
- Assess key concerns for LLMOps such as cost/performance tradeoffs, deployment options, monitoring, and feedback.
- Outline the development-to-production workflow for deploying a scalable LLM-powered data pipeline.

Machine Learning with Databricks

Click here for the customer enrollment link.

Duration: 16 hours

Course description: This course is your gateway to mastering machine learning workflows on Databricks. Dive into data preparation, model development, deployment, and operations, guided by expert instructors. Learn essential skills for data exploration, model training, and deployment strategies tailored for Databricks. By course end, you'll have the knowledge and confidence to navigate the entire machine learning lifecycle on the Databricks platform, empowering you to build and deploy robust machine learning solutions efficiently.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Prerequisites:

At a minimum, you should be familiar with the following before attempting to take this content:

- Familiarity with Databricks workspace and notebooks
- Familiarity with Delta Lake and Lakehouse
- Intermediate level knowledge of Python

Learning objectives:

After consuming this content, you should be able to:

- Describe the comprehensive features of the Databricks Data Intelligence Platform tailored for machine learning, including data storage, governance, and exploratory data analysis techniques with Spark and integrated visualization tools.
- Explain fundamental machine learning concepts, MLflow components for model development, and hyperparameter tuning methods, while performing practical skills in utilizing MLflow for model tracking and tuning, and leveraging Databricks AutoML for rapid model experimentation.
- Master deployment strategies for batch, pipeline, and real-time scenarios, understanding their advantages and limitations, and demonstrating proficiency in performing batch, pipeline, and real-time inference using Databricks features like DLT and Model Serving.
- Develop expertise in modern machine learning operations encompassing DevOps, DataOps, and ModelOps principles, and architect machine learning operations solutions based on Databricks-recommended best practices. Perform an end-to-end implementation of a machine learning project using MLOps Stacks and other Databricks capabilities.

Machine Learning in Production(V2)

Click here for the customer enrollment link.

Duration: 12 hours

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

NOTE: This is an e-learning version of the Machine Learning in Production instructor-led course. It is an on-demand recording available via the Databricks Academy and covers the same content as the instructor-led course. For more information about what's in the course itself, please visit this link.

Machine Learning Model Deployment

Click here for the customer enrollment link.

Duration: 4 hours

Course description: This comprehensive course provides a practical guide to developing traditional machine learning models on Databricks, emphasizing hands-on demonstrations and workflows using popular ML libraries. This course focuses on executing common tasks efficiently with AutoML and MLflow. Participants will delve into key topics, including regression and classification models, harnessing Databricks' capabilities to track model training, leveraging feature stores for model development, and implementing hyperparameter tuning. Additionally, the course covers AutoML for rapid and low-code model training, ensuring that participants gain practical, real-world skills for streamlined and effective machine learning model development in the Databricks environment.

Prerequisites:

- Knowledge of fundamental concepts of regression and classification methods
- Familiarity with Databricks workspace and notebooks
- Intermediate level knowledge of Python

Learning objectives:

- Describe fundamental concepts of machine learning.
- Describe the main components of MLflow for model development.
- Describe hyperparameter tuning and methods.
- Utilize MLflow for model tracking and model tuning with hyperopt.
- Explain benefits and features of Databricks AutoML

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Create an AutoML experiment, identify the best model, and modify the generated models.

Machine Learning Model Development

Click here for the customer enrollment link.

Duration: 4 hours

Course description: This comprehensive course provides a practical guide to developing traditional machine learning models on Databricks, emphasizing hands-on demonstrations and workflows using popular ML libraries. This course focuses on executing common tasks efficiently with AutoML and MLflow. Participants will delve into key topics, including regression and classification models, harnessing Databricks' capabilities to track model training, leveraging feature stores for model development, and implementing hyperparameter tuning. Additionally, the course covers AutoML for rapid and low-code model training, ensuring that participants gain practical, real-world skills for streamlined and effective machine learning model development in the Databricks environment.

Prerequisites:

- Knowledge of fundamental concepts of regression and classification methods
- Familiarity with Databricks workspace and notebooks
- Intermediate level knowledge of Python

Learning objectives:

- Describe fundamental concepts of machine learning.
- Describe the main components of MLflow for model development.
- Describe hyperparameter tuning and methods.
- Utilize MLflow for model tracking and model tuning with hyperopt.
- Explain benefits and features of Databricks AutoML
- Create an AutoML experiment, identify the best model, and modify the generated models.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Machine Learning Operations

Click here for the customer enrollment link.

Duration: 4 hours

Course description: This course will guide participants through an exploration of machine learning operations on Databricks. This will begin with an introduction to modern machine learning operations as a combination of DevOps, DataOps, and ModelOps, including an overview of each component. Next, the course will cover the Databricks-recommended architectures and solutions for machine learning operations. This will focus on capabilities like MLOps Stacks, MLflow, Model Serving, Unity Catalog, and Lakehouse Monitoring. Finally, the final section of the course will be a practical walkthrough of a simple machine learning operations project using these capabilities, including: an initial development phase, a staging phase, and a production stage. By the end of this course, learners will be familiar with the basic end-to-end machine learning operations workflow on Databricks.

Prerequisites:

At a minimum, you should be familiar with the following before attempting to take this content:

- Basic knowledge of traditional machine learning concepts
- Beginner experience with traditional machine learning development on Databricks
- Intermediate knowledge of Python for machine learning projects
- Recommended: Beginner experience with basic DevOps concepts like CI/CD

Learning objectives:

After consuming this content, you should be able to:

- Describe modern machine learning operations in the context of DevOps, DataOps, and ModelOps.
- Architect basic machine learning operations solutions for traditional machine learning applications based on Databricks-recommended best practices.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Describe the process of setting up and implementing a machine learning project on Databricks using MLOps Stacks and other Databricks capabilities.

~~Manage Data with Delta Lake~~

~~Click here for the customer enrollment link.~~

~~Duration: 3 hours~~

~~Course description: The Databricks Lakehouse Platform enables individuals throughout an organization to collaboratively develop, productionalize, and derive insights from data assets using a set of common tools and a unified collection of databases. This module presents an overview of the data lakehouse and provides an in-depth hands-on introduction to Delta Lake. Note: These lessons are a subset of the Data Engineering with Databricks course.~~

~~Prerequisites:~~

- ~~Beginner familiarity with cloud computing concepts (virtual machines, object storage, etc.)~~
- ~~Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc)~~
- ~~Beginning programming experience with Spark SQL~~
- ~~Extract data from a variety of file formats and data sources~~
- ~~Apply a number of common transformations to clean data~~
- ~~Reshape and manipulate complex data using advanced built-in functions~~

~~Learning objectives:~~

- ~~Describe how Delta Lake transactional guarantees provide the foundation for the data lakehouse architecture.~~
- ~~Use Delta Lake DDL to create tables, compact files, restore previous table versions, and perform garbage collection of tables in the Lakehouse.~~
- ~~Use CTAS to store data derived from a query in a Delta Lake table~~

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- ~~Use SQL to perform complete and incremental updates to existing tables.~~

Manage Data Access with Unity Catalog

Click here for the customer enrollment link.

Duration: 3 hours

Course description: Unity Catalog is a central hub for administering and securing your data which enables granular access control and built-in auditing across the Databricks platform. This course guides learners through fundamental Unity Catalog concepts and tasks necessary to satisfy data governance requirements.

Prerequisites:

- Beginning-level knowledge of the Databricks Lakehouse platform (high-level knowledge the structure and benefits of the Lakehouse platform)
- Beginning-level knowledge of SQL (ability to understand and construct basic queries)

Learning objectives:

- Describe Unity Catalog key concepts and how it integrates with the Databricks platform
- Access Unity Catalog through clusters and SQL warehouses
- Create and govern data assets in Unity Catalog
- Adopt Databricks recommendations into your organization's Unity Catalog based solutions
- Data Science Workspace

Multi-stage Reasoning with Large Language Models Chains

Click here for the customer enrollment link.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Duration: 45 minutes

Course description: Welcome to Multi-stage Reasoning with Large Language Models Chains! Please note: This module is a subset of the larger Large Language Models with Databricks course available in the Databricks Academy (just search for the title).

In previous modules in the Large Language Models with Databricks course, we learned how we can download LLMs from places like Hugging Face to solve various tasks in Natural Language Processing. We also saw how we can convert our data to vector format and provide vector-type searches using vector databases. In this module, we'll show you how you can combine these two features together to enhance applications that we can build as developers.

Prerequisites:

- Intermediate-level experience with Python
- Working knowledge of machine learning and deep learning is helpful

Learning objectives:

- Describe the flow of LLM pipelines with tools like LangChain.
- Apply LangChain to leverage multiple LLM providers such as Open AI and Hugging Face.
- Create complex logic flow with agents in LangChain to pass prompts and use logical reasoning to complete tasks.

New Capability Overview: bamboolib

Click here for the customer enrollment link.

Duration: 70 minutes

Course description: bamboolib delivers an extendable GUI that exports Python code for fast, simple data exploration and transformation without any coding required for users. The UI-based workflows help make Databricks accessible for both citizen data scientists and experts alike and reduces employee on-boarding and training

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

costs. These no-code use cases include data exploration, data transformation and data visualization.

There are many benefits of using bamboolib on Databricks for team members who have coding skills and team members who cannot code. Team members who have coding skills can speed up their data exploration and transformation process by avoiding repetitive tasks and use out-of-the box best-practice analyses provided by bamboolib. Others who cannot code can use bamboolib for all stages of the data analysis process without writing code or they may use bamboolib as a great entrypoint for getting started to learn coding.

This course will introduce you to bamboolib, discuss its use case, and demonstrate how to use it on Databricks platform. The demonstration will cover loading data from various sources, exploring data, transforming data and visualizing data.

Prerequisites & Requirements

- In this course, learners are expected to have;
- Intermediate level of data analysis concept and methods including data transformation, data visualization, and summary statistics
- Basic understanding of statistical methods

Learning objectives

- Describe features and use cases of the bamboolib on Databricks
- Import data from static files and database tables into the bamboolib UI
- Utilize bamboolib for data exploration
- Utilize bamboolib for data transformation
- Utilize bamboolib for data visualization

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

New Capability Overview: Databricks Assistant

Click here for the customer course enrollment link.

Duration: 40 minutes

Course Description:

This course provides an in-depth overview of the new capability, **Databricks Assistant**, introduced in the Databricks Data Intelligence Platform, revolutionizing the coding experience. Databricks Assistant is designed to streamline code and SQL authoring processes across various Databricks platforms. This course aims to introduce Databricks Assistant, covering fundamental concepts, competitive positioning, and practical demonstrations of its capabilities.

Prerequisites:

At a minimum, you should be familiar with the following before attempting to take this content:

- Very basic SQL or Python
- Opening and using Databricks Notebooks and clusters

Learning objectives:

- Explain fundamental concepts about the impact of Databricks Assistant on the Databricks Data Intelligence Platform.
- Follow along with a gold-standard demonstration of how Databricks Assistant can provide value to common/important use cases.
- Perform hands-on activities using Databricks Assistant, showcasing the most important features of the Databricks Data Intelligence Platform.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

New Capability Overview: Data Lineage with Unity Catalog

Click here for the customer enrollment link.

Duration: 1 hour 30 minutes

Course description: Databricks Unity Catalog, the unified governance solution for all data and AI assets on lakehouse, brings support for data lineage. Data lineage includes capturing all the relevant metadata and events associated with the data in its lifecycle, including the source of the data set, what other data sets were used to create it, who created it and when, what transformations were performed, what other data sets leverage it, and many other events and attributes. With a data lineage solution, data teams get an end-to-end view of how data is transformed and how it flows across their data estate. Lineage is supported for all languages and is captured down to the column level. Lineage data includes notebooks, workflows, and dashboards related to the query. Lineage can be visualized in Data Explorer in near real-time and retrieved with the Databricks REST API.

In this course, we are going to cover the fundamentals of data lineage, how to create lineage enabled clusters, and demonstrate how to capture and view lineage data.

Prerequisites:

- Basic knowledge of Databricks SQL

Learning objectives:

- Describe fundamental concepts of data lineage
- Explain common use cases and features of data lineage with Unity Catalog
- Describe the technical requirements for data lineage on Databricks
- Configure a cluster and SQL warehouse to enable data lineage
- Utilize Data Explorer UI to explore lineage information at table and column level
- Manage data lineage permissions at data and entity level

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

New Capability Overview: Data Profiles in Databricks Notebooks

Click here for the customer enrollment link.

Duration: 30 minutes

Course description: Exploratory data analysis is a key part of the repeating cycle of exploration, development, and validation that makes up data asset development, and establishing a baseline understanding of a data set is a crucial job that is done as part of EDA. Databricks simplifies this process of understanding data sets, making it possible to generate a robust profile or summary of the data set with the push of a button. These are available in notebooks in Databricks Machine Learning and Databricks Data Science and Engineering Workspaces.

Prerequisites:

- Basic Python skills (e.g. can declare variables, distinguish between methods and attributes).
- Ability to describe and utilize basic summary statistics (e.g mean, standard deviation, median).

Learning objectives:

- Explain fundamental concepts about creating Data Profiles to help the process of EDA.
- Perform basic tasks using Data Profiles in Databricks notebooks.

~~New Capability Overview: Databricks SQL Serverless~~

~~Click here for the customer enrollment link.~~

~~Duration: 10 minutes~~

~~Course description: In this course, you will learn about how Databricks SQL users can leverage the nearly instant -on capability of Serverless to make working with Databricks SQL~~

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

~~faster than ever. In addition, you will learn how customer data is protected and isolated. Finally, you will learn how customers can enable this feature.~~

~~Prerequisites & Requirements~~

- ~~Prerequisites~~
  - ~~Experience with Databricks SQL~~
  - ~~Familiarity with the basics of how cloud compute resources are provisioned~~
  - ~~Administrator access~~

~~Learning objectives~~

- ~~Explain how Databricks SQL Serverless fits into the lakehouse architecture~~
- ~~Describe how Databricks SQL Serverless works~~
- ~~Explain how Databricks SQL Serverless keeps customer data secure and isolated~~
- ~~Implement Databricks SQL Serverless on customer accounts~~

New Capability Overview: Global Search

Click here for the customer enrollment link.

Duration: 20 minutes

Course description: Delta Sharing is an open protocol for secure data sharing with other organizations regardless of which computing platforms they use. Databricks provides both open source and managed options for Delta Sharing. Databricks-managed Delta Sharing allows data providers to share data and data recipients to access the shared data. It can share collections of tables in a Unity Catalog metastore in real time without copying them, so that data recipients can immediately begin working with the latest version of the shared data.

This course will give an overview of Delta Sharing with an emphasis on the benefits of Delta Sharing over other methods of data sharing and how Delta Sharing fits into the Lakehouse architecture. Then, there will be a demonstration on how to configure Delta Sharing on a metastore and how to enable external data sharing.

databricks
Academy

Please note - with the launch of of our [learning catalog](), this course catalog will no longer be updated as of June 17, 2024.

Prerequisites:

- Beginning-level knowledge of the Databricks Lakehouse platform
- Basic understanding of how Unity Catalog is integrated into the Databricks Lakehouse Platform
- Metastore admin or account admin privileges (optional, but needed if following along with the configuration demo)

Learning objectives:

- Describe the benefits of Delta Sharing compared to traditional data sharing systems
- Describe features and use cases of Delta Sharing
- Define key components of Delta Sharing
- Configure Delta Sharing on a metastore
- Enable external data sharing for an account

New Capability Overview: ipywidgets

Click here for the customer enrollment link.

Duration: 1:15 hour

Until recently, Databricks widgets have been the only option for users wanting to use widgets in Databricks notebooks. Now, with the integration of ipywidgets, Databricks users have an alternative method for building interactive notebooks.

Widgets can be embedded into the Databricks notebooks to provide a user-friendly interface that can be used to collect and use these data in the code without having to change the code. This method can be used to create interactive data apps in notebooks.

This course will introduce you to ipywidgets, discuss use cases for when to use them, and demonstrate how to embed various ipywidgets controls in Databricks notebooks.
Prerequisites & Requirements

- Prerequisites
    - In this course, learners are expected to have;
    - Intermediate level python coding skills

databricks Academy

Please note - with the launch of of our [learning catalog](), this course catalog will no longer be updated as of June 17, 2024.

- Basic knowledge of pyspark for reading and querying data
- Basic knowledge of CSS for styling the controls

Learning objectives

- Describe ipywidgets features and use cases
- Explain the difference between ipywidgets and Databricks widgets
- List supported widget types
- List supported and unsupported ipywidgets controls on DB
- Utilize ipywidgets controls to create interactive notebooks
- Use layout and styling controls to build custom layouts

New Capability Overview: Model Serving

Click here for the customer enrollment link.

Duration: 1 Hour

Course description: This course provides an in-depth overview of the new capability, Model Serving, introduced in the Databricks Data Intelligence Platform. It covers fundamental concepts, competitive positioning, and hands-on demonstrations to showcase its value in various use cases. The course includes detailed instruction on deploying models, querying endpoints, and monitoring performance, offering participants a comprehensive understanding of Model Serving's capabilities.

Prerequisites & Requirements

- Prerequisites
  - Write basic machine learning code
  - Able to train models within the Databricks platform
  - Ability or previous experience deploying models is helpful but not required

Learning objectives

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Explain fundamental concepts about the impact that Model Serving has on the Databricks Data Intelligence Platform.
- Perform a gold-standard demonstration of how Model Serving can be used to provide value to common/essential use cases.
- Perform hands-on activities using Model Serving that showcases the most important features of the Databricks Data Intelligence Platform.

New Capability Overview: Personal Compute

Click here for the customer course enrollment link.

Duration: 30 Minutes

Course description: Databricks is no longer a tool that is only suitable for massive, Spark-powered workloads. Personal compute is a new, Databricks-managed default compute policy that will appear in all customers' workspaces.

Prerequisites:

- Log in to Databricks
- Access to cluster creation

Learning objectives:

- Compare and contrast Personal Compute node and multi-node clusters
- Create a Personal Compute node

New Capability Overview: VS Code Extension for Databricks

Click here for the customer course enrollment link.

Duration: 00:45 hours

Course description: Integrated development environments (IDEs) are important for Databricks users because they provide a variety of features that can increase

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

productivity and streamline the development process. The new Databricks extension for VS Code allows users to develop the Databricks Lakehouse platform from VSCode. This extension allows Databricks users to connect their local VSCode environment to remote Databricks workspaces, synchronize local code in VS Code with remote workspaces, and run Python files and Python notebooks in remote workspaces. In this course, we are going to introduce you to this new extension and discuss its main features. In the demo section of the course, we are going to demonstrate how to install and configure the extension for VS Code and run locally developed Python applications on Databricks clusters in the cloud.

Prerequisites:

- Basic knowledge of Python programming.
- Basic knowledge of Databricks cluster and notebook concepts.
- Familiarity with VS Code interface.

Learning objectives:

- Explain the main features of VS Code extension for Databricks.
- Describe the local and workspace requirements for using the extension.
- Install and configure the extension to run local Python files.
- Run Python files and Python notebooks as standalone and workflow jobs.
- Run a unit test function using custom runtime configuration.

~~New Capability Overview: Workspace Browser for Databricks SQL~~

~~Click here for the customer course enrollment link.~~

~~Duration: 20 minutes~~

~~Course description: Creating, managing and sharing workspace entities on the Databricks platform is a very common task among data engineers, data scientists and data analysts. The new Workspace Browser for DBSQL unifies browsing for content across all three Databricks personas. With this feature, you are able to~~

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

~~create, browse, manage and share existing workspace content (notebooks, experiments, etc.) and DBSQL specific content (queries, dashboards, alerts, etc.).~~

~~This short, introductory course will introduce you to the new Workspace Browser within Databricks SQL and demonstrate how to use this feature effectively. The first section of the course will cover features and use cases of the new Workspace Browser and in the second section of the course we will demonstrate how to create, organize and share assets using this new feature.~~

~~Prerequisites:~~

- ~~There are no prerequisites for this course.~~

~~Learning objectives:~~

- ~~Describe the Workspace Browser features and use cases~~
- ~~Migrate existing objects into the Workspace Browser~~
- ~~Create, organize and share new objects using the Workspace Browser~~

~~Optimizing Apache Spark on Databricks~~

~~Click here for the customer course enrollment link.~~

~~Duration: 12 hours~~

~~Course description: In this course, students will explore five key problems that represent the vast majority of performance problems in an Apache Spark application: Skew, Spill, Shuffle, Storage, and Serialization. With each of these topics, we explore coding examples based on 100 GB to 1+ TB datasets that demonstrate how these problems are introduced, how to diagnose these problems with tools like the Spark UI, and conclude by discussing mitigation strategies for each of these problems.~~

~~We continue the conversation by looking at a series of key ingestion concepts that promote strategies for processing Tera Bytes of data including managing Spark Partition sizes, Disk Partitioning, Bucketing, Z-Ordering, and more. With each~~

databricks
Academy

Please note – with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

~~of these topics, we explore when and how each of these techniques should be implemented, new challenges that productionalizing these solutions might provide along with corresponding mitigation strategies.~~

~~Finally, we introduce a couple of other key topics such as issues with Data Locality, IO-Caching and Spark-Caching, Pitfalls of Broadcast Joins, and new features like Spark 3's Adaptive Query Execution and Dynamic Partition Pruning. We then conclude the course with discussions and exercises on designing and configuring clusters for optimal performance given specific use cases, personas, the divergent needs of various teams, and cross-team security concerns.~~

~~Prerequisites:~~

- ~~Intermediate to advanced programming experience in Python or Scala~~
- ~~Hands-on experience developing Apache Spark applications~~

~~Learning objectives:~~

- ~~Articulate how the five most common performance problems in a Spark application can be mitigated to achieve better application performance.~~
- ~~Summarize some of the most common performance problems associated with data ingestion and how to mitigate them.~~
- ~~Articulate how new features in Spark 3.0 can be employed to mitigate performance problems in your Spark applications.~~
- ~~Configure a Spark cluster for maximum performance given specific job requirements and while considering a multitude of other factors.~~

~~Performance Optimization with Spark and Delta Lake~~

~~Click here for the customer course enrollment link.~~

~~Duration: 12 hours~~

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

~~Course description:In this course, you'll learn how to optimize workloads and physical layout with Spark and Delta Lake and analyze the Spark UI to assess performance and debug applications.~~

~~Prerequisites:~~

- ~~Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc).~~
- ~~Intermediate programming experience with PySpark:~~
- ~~Extract data from a variety of file formats and data sources~~
- ~~Apply a number of common transformations to clean data~~
- ~~Reshape and manipulate complex data using advanced built-in functions~~
- ~~Intermediate programming experience with Delta Lake (create tables, perform complete and incremental updates, compact files, restore previous versions etc.)~~

~~Note: The data engineering skills required for this course can be learned by taking the Data Engineering with Databricks course by Databricks Academy.~~

~~Learning objectives:~~

- ~~Describe strategies and best practices for optimizing workloads on Databricks~~
- ~~Analyze information presented in the Spark UI and Cluster UI to assess performance and debug applications~~

Preparing for Databricks Certification Exams

Click here for the customer enrollment link.

Duration: 20 Minutes

Course description: This course outlines important details about Databricks Certification exams. While most of the information found within this course is

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

available on the Databricks Certification website, this training seeks to summarize that content to simplify your preparation.

Please note that this course will not teach any of the content on any particular exam.

Instead, this course will review the general information necessary for your test-taking experience to go as smoothly as possible, and point you to resources you can use to study for each exam.

Prerequisites:

- N/A

Learning objectives:

- Identify key concepts about the Databricks Certification Program.
- Explain best practices for taking an online proctored exam.
- Describe how to access and use study resources to prepare for your exam.

Retrieval-Augmented Generation with Vector Search and Storage

Click here for the customer course enrollment link.

Duration: 45 Minutes

Course description: Welcome to Retrieval-Augmented Generation with Vector Search and Storage! Please note: This module is a subset of the larger Large Language Models with Databricks course available in the Databricks Academy (just search for the title).

Retrieval Augmented Generation is gaining popularity for leveraging large language models, also known as LLMs, to extract information from documents or a knowledge base to answer questions and perform area-specific tasks. One of the important applications that we see at Databricks is when our customers use custom datasets and internal knowledge bases with LLMs to address domain-specific knowledge.

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Prerequisites:

- Intermediate-level experience with Python
- Working knowledge of machine learning and large language models is helpful

Learning objectives:

- Explain vector search strategies and how to evaluate search results.
- Define the utility of vector databases.
- Differentiate between vector databases, vector libraries, and vector plugins.
- Discuss best practices for when to use vector stores and how to improve search-retrieval performance.

~~Scalable Machine Learning with Apache Spark (V2)~~

~~Click here for the customer course enrollment link.~~

~~Duration: 12 hours~~

~~NOTE: This is an e-learning version of the Scalable Machine Learning with Apache Spark instructor-led course. It is an on-demand recording available via the Databricks Academy and covers the same content as the instructor-led course. For more information about what's in the course itself, please visit this link.~~

Share Data within Databricks Using Delta Sharing

Click here for the customer course enrollment link.

Duration: 45 Minutes

Course description: Delta Sharing is an open protocol for secure data sharing with other organizations regardless of which computing platforms they use. Databricks provides both open source and managed options for Delta Sharing. Databricks-managed Delta Sharing allows data providers to share data and data

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

recipients to access the shared data. With Delta Sharing, users can share collections of tables in a Unity Catalog metastore in real time without copying them, so that data recipients can immediately begin working with the latest version of the shared data.

This course will focus on the data sharing process with Delta Sharing. On Databricks users manage data sharing processes through UI or using Databricks SQL queries. In this course, we will start by demonstrating how to share data using UI and then we will use SQL for the same process. It is highly recommended that you are familiar with both methods.

Prerequisites:

- Beginning-level knowledge of the Databricks Lakehouse platform (high-level knowledge the structure and benefits of the Lakehouse platform)
- Basic understanding of how Unity Catalog is integrated into the Databricks Lakehouse Platform
- Beginning-level knowledge of SQL
- Admin account privileges or have been granted required privileges

Learning objectives:

- Describe the data sharing and data access process with Delta Sharing within Databricks
- Share data and access shared data within Databricks using the UI
- Share data and access shared data within Databricks using SQL queries
- Share and access partial data within Databricks

Society and Large Language Models

Click here for the customer course enrollment link.

Duration: 45 Minutes

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

Course description: Welcome to Society and Large Language Models! Please note: This module is a subset of the larger Large Language Models with Databricks course available in the Databricks Academy (just search for the title).

LLMs are changing history - but with great power, comes great responsibility. In this module, we'll talk about why LLMs are often considered a double-edged sword. While the benefits that LLMs are bringing to the world are widespread and powerful, it's also important to discuss concerns/problems that may arise from their use as we roll them out into production to the world and to our organizations.

DISCLAIMER: In this module we'll be discussing things like what may happen when models output potentially offensive, inaccurate, or biased information or instructions.

Prerequisites:

- Intermediate-level experience with Python
- Working knowledge of machine learning and large language models is helpful

Learning objectives:

- Discuss the merits and risks of LLM usage.
- Examine datasets used to train LLMs and assess their inherent bias.
- Identify the underlying causes and consequences of hallucination, and discuss evaluation and mitigation strategies.
- Discuss ethical and responsible usage and governance of LLMs.

Streaming ETL Patterns with Delta Live Tables

Click here for the customer enrollment link.

Duration: 2 hours30min

Course description: In this course, you'll learn how to apply design patterns for designing workloads to perform ETL in the Lakehouse with Delta Live Tables.

Prerequisites:

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc).
- Intermediate programming experience with PySpark:
    - Extract data from a variety of file formats and data sources
    - Apply a number of common transformations to clean data
    - Reshape and manipulate complex data using advanced built-in functions
- Intermediate programming experience with Delta Lake (create tables, perform complete and incremental updates, compact files, restore previous versions etc.)

Note: The data engineering skills required for this course can be learned by taking the Data

Engineering with Databricks course by Databricks Academy.

Learning objectives:

- Ingest raw streaming data into a multiplex bronze table and apply metadata
- Enforce quality with expectations and quarantine tables
- Implement and update a Type 2 slowly changing dimension table
- Explore and tune state information using streaming joins

SWE Practices for Delta Live Tables

Click here for the customer enrollment link.

Duration: 2 hours 30 min

Course description: In this course, you'll learn how to implement software engineering best practices to develop, test and deploy DLT pipeline code into production.

Prerequisites:

- Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc).
- Intermediate programming experience with PySpark:
    - Extract data from a variety of file formats and data sources
    - Apply a number of common transformations to clean data
    - Reshape and manipulate complex data using advanced built-in functions

databricks
Academy

Please note - with the launch of of our [learning catalog](), this course catalog will no longer be updated as of June 17, 2024.

- Intermediate programming experience with Delta Lake (create tables, perform complete and incremental updates, compact files, restore previous versions etc.)

Note: The data engineering skills required for this course can be learned by taking the Data Engineering with Databricks course by Databricks Academy.

Learning objectives:

- Describe how to implement SWE best practices and CI/CD workflows for Delta Live Table pipelines
- Modularize code with libraries, parameterization, metaprogramming, and portable expectations
- Configure unit testing and deployment environments for DLT pipelines
- Monitor data pipelines to ensure continued operation and data quality by querying and visualizing metrics from event logs

Transform Data with Spark

Click here for the customer course enrollment link.

Duration: 3 hours

Course description: While the data lakehouse combines the best aspects of the data warehouse and the data lake, users familiar with one or both of these environments may still encounter new concepts as they move to Databricks. By the end of these lessons, students will feel comfortable defining databases, tables, and views in the Lakehouse, ingesting arbitrary data from a variety of sources, and writing simple applications to drive ETL pipelines. Note: This course covers both PySpark and SQL. Also, the lessons in this course are a subset of the Data Engineering with Databricks course.

Prerequisites:

- Beginner familiarity with basic cloud concepts (virtual machines, object storage, identity management)

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc)
- Intermediate familiarity with basic SQL concepts (select, filter, groupby, join, etc)
- Beginner programming experience with Python (syntax, conditions, loops, functions)
- Beginner programming experience with the Spark DataFrame API: Configure DataFrameReader and DataFrameWriter to read and write data, Express query transformations using DataFrame methods and Column expressions, Navigate the Spark documentation to identify built-in functions for various transformations and data types.

Learning objectives:

- Extract data from a variety of file formats and data sources.
- Apply a number of common transformations to clean data.
- Reshape and manipulate complex data using advanced built-in functions.
- Leverage UDFs for reusable code and apply best practices for performance.

Unity Catalog Patterns and Best Practices

Click here for the customer course enrollment link.

Duration: 30 Minutes

Course description: Unity Catalog is a central hub for administering and securing your data which enables granular access control and built-in auditing across the Databricks platform. This course guides learners through recommended practices and patterns for implementing data architectures centered on Unity Catalog.

Prerequisites:

- This course has no specific course prerequisites.

Learning objectives:

databricks
Academy

Please note - with the launch of of our learning catalog, this course catalog will no longer be updated as of June 17, 2024.

- Adopt Databricks recommendations into your organization's Unity Catalog based solutions

What is Big Data?

Click here for the customer course enrollment link.

Duration: 1 hour

Course description: This course was created for individuals who are new to the big data landscape and want to become conversant with big data terminology. It will cover foundational concepts related to the big data landscape including: characteristics of big data; the relationship between big data, artificial intelligence, and data science; how individuals on data science teams work with big data; and how organizations can use big data to enable better business decisions.

Prerequisites:

- Experience using a web browser

Learning objectives:

- Explain foundational concepts used to define big data.
- Explain how the characteristics of big data have changed traditional organizational workflows for working with data.
- Summarize how individuals on data science teams work with big data on a daily basis to drive business outcomes.
- Articulate examples of real-world use-cases for big data in businesses across a variety of industries.