Performing Data Engineering on Microsoft Cloud Services

Código del curso: 20776

Duración: 5 días

Acerca de este curso

This five-day instructor-led course describes how to process Big Data using Azure tools and services including Azure Stream Analytics, Azure Data Lake, Azure SQL Data Warehouse and Azure Data Factory. The course also explains how to include custom functions, and integrate Python and R.

Perfil del usuario objetivo

The primary audience for this course is data engineers (IT professionals, developers, and information workers) who plan to implement big data engineering workflows on Azure.

Al finalizar el curso

After completing this course, students will be able to:

- Describe common architectures for processing big data using Azure tools and services.
- Describe how to use Azure Stream Analytics to design and implement stream processing over large-scale data.
- Describe how to include custom functions and incorporate machine learning activities into an Azure Stream Analytics job.
- Describe how to use Azure Data Lake Store as a large-scale repository of data files.
- Describe how to use Azure Data Lake Analytics to examine and process data held in Azure Data Lake Store.
- Describe how to create and deploy custom functions and operations, integrate with Python and R, and protect and optimize jobs.
- Describe how to use Azure SQL Data Warehouse to create a repository that can support large-scale analytical processing over data at rest.
- Describe how to use Azure SQL Data Warehouse to perform analytical processing, how to maintain performance, and how to protect the data.
- Describe how to use Azure Data Factory to import, transform, and transfer data between repositories and services.

Requisitos previos:

In addition to their professional experience, students who attend this training should already have the following technical knowledge:

- A good understanding of Azure data services.
- A basic knowledge of the Microsoft Windows operating system and its core functionality.
- A good knowledge of relational databases.

Contenido del curso:

Module 1: Architectures for Big Data Engineering with Azure

This module describes common architectures for processing big data using Azure tools and services.

Lessons

- Understanding Big Data
- Architectures for Processing Big Data
- Considerations for designing Big Data solutions

Lab : Designing a Big Data Architecture

- Design a big data architecture

After completing this module, students will be able to:

- Explain the concept of Big Data.
- Describe the Lambda and Kappa architectures.
- Describe design considerations for building Big Data Solutions with Azure.

Module 2: Processing Event Streams using Azure Stream Analytics

This module describes how to use Azure Stream Analytics to design and implement stream processing over large-scale data.

Lessons

- Introduction to Azure Stream Analytics
- Configuring Azure Stream Analytics jobs

Lab : Processing Event Streams with Azure Stream Analytics

- Create an Azure Stream Analytics job
- Create another Azure Stream job
- Add an Input

- Edit the ASA job
- Determine the nearest Patrol Car

After completing this module, students will be able to:

- Describe the purpose and structure of Azure Stream Analytics.
- Configure Azure Stream Analytics jobs for scalability, reliability and security.

Module 3: Performing custom processing in Azure Stream Analytics

This module describes how to include custom functions and incorporate machine learning activities into an Azure Stream Analytics job.

Lessons

- Implementing Custom Functions
- Incorporating Machine Learning into an Azure Stream Analytics Job

Lab : Performing Custom Processing with Azure Stream Analytics

- Add logic to the analytics
- Detect consistent anomalies
- Determine consistencies using machine learning and ASA

After completing this module, students will be able to:

- Describe how to create and use custom functions in Azure Stream Analytics.
- Describe how to use Azure Machine Learning models in an Azure Stream Analytics job.

Module 4: Managing Big Data in Azure Data Lake Store

This module describes how to use Azure Data Lake Store as a large-scale repository of data files.

Lessons

- Using Azure Data Lake Store
- Monitoring and protecting data in Azure Data Lake Store

Lab : Managing Big Data in Azure Data Lake Store

- Update the ASA Job
- Upload details to ADLS

After completing this module, students will be able to:

- Describe how to create an Azure Data Lake Store, create folders, and upload data.
- Explain how to monitor an Azure Data Lake account, and protect the data that it contains.

Module 5: Processing Big Data using Azure Data Lake Analytics

This module describes how to use Azure Data Lake Analytics to examine and process data held in Azure Data Lake Store.

Lessons

- Introduction to Azure Data Lake Analytics
- Analyzing Data with U-SQL
- Sorting, grouping, and joining data

Lab : Processing Big Data using Azure Data Lake Analytics

- Add functionality
- Query against Database
- Calculate average speed

After completing this module, students will be able to:

- Describe the purpose of Azure Data Lake Analytics, and how to create and run jobs.

- Describe how to use USQL to process and analyse data.
- Describe how to use windowing to sort data and perform aggregated operations, and how to join data from multiple sources.

Module 6: Implementing custom operations and monitoring performance in Azure Data Lake Analytics

This module describes how to create and deploy custom functions and operations, integrate with Python and R, and protect and optimize jobs.

Lessons

- Incorporating custom functionality into Analytics jobs
- Managing and Optimizing jobs

Lab : Implementing custom operations and monitoring performance in Azure Data Lake Analytics

- Custom extractor
- Custom processor
- Integration with R/Python
- Monitor and optimize a job

After completing this module, students will be able to:

- Describe how to incorporate custom features and assemblies into USQL.
- Describe how to implement security to protect jobs, and how to monitor and optimize jobs to ensure efficient operations.

Module 7: Implementing Azure SQL Data Warehouse

This module describes how to use Azure SQL Data Warehouse to create a repository that can support large-scale analytical processing over data at rest.

Lessons

- Introduction to Azure SQL Data Warehouse
- Designing tables for efficient queries
- Importing Data into Azure SQL Data Warehouse

Lab : Implementing Azure SQL Data Warehouse

- Create a new data warehouse
- Design and create tables and indexes
- Import data into the warehouse.

After completing this module, students will be able to:

- Describe the purpose and structure of Azure SQL Data Warehouse.
- Describe how to design table to optimize the processing performed by the data warehouse.
- Describe tools and techniques for importing data into a warehouse at scale.

Module 8: Performing Analytics with Azure SQL Data Warehouse

This module describes how to import data in Azure SQL Data Warehouse, and how to protect this data.

Lessons

- Querying Data in Azure SQL Data Warehouse
- Maintaining Performance
- Protecting Data in Azure SQL Data Warehouse

Lab : Performing Analytics with Azure SQL Data Warehouse

- Performing queries and tuning performance

- Integrating with Power BI and Azure Machine Learning
- Configuring security and analysing threats

After completing this module, students will be able to:

- Describe how to perform queries and use the data held in a data warehouse to perform analytics and generate reports.
- Describe how to configure and monitor a data warehouse to maintain good performance.
- Describe how to protect data and manage security in a data warehouse.

Module 9: Automating the Data Flow with Azure Data Factory

This module describes how to use Azure Data Factory to import, transform, and transfer data between repositories and services.

Lessons

- Introduction to Azure Data Factory
- Transferring Data
- Transforming Data
- Monitoring Performance and Protecting Data

Lab : Automating the Data Flow with Azure Data Factory

- Automate the Data Flow with Azure Data Factory

After completing this module, students will be able to:

- Describe the purpose of Azure Data Factory, and explain how it works.
- Describe how to create Azure Data Factory pipelines that can transfer data efficiently.
- Describe how to perform transformations using an Azure Data Factory pipeline.
- Describe how to monitor Azure Data Factory pipelines, and how to protect the data flowing through these pipelines.