Building a Simple Data Pipeline with Azure

SHARES KNOWLEDGE ON
DATACRAFT
DATA ENGINEERING

Introduction

- **Objective**: Move data from a local SQL Server (AdventureWorksLT2019) to Azure Blob Storage using Azure Data Factory (ADF) and create a table in Azure SQL Database.
- **Focus**: Business-specific example with a video demo.

Business Scenario

- **Context**: Retail company migrating customer data to the cloud for scalability and accessibility.
- **Objective**: Efficiently transfer the Customer table to enable advanced analytics and reporting.

Solution Overview

- **Local SQL Server**: Source database with the Customer table.
- **Azure Blob Storage**: Intermediate storage for data transfer.
- **Azure Data Factory (ADF)**: Orchestrates data movement using a Self-hosted Integration Runtime (IR).

Prerequisites for the Demo

- MS SQL 2019 Developer Edition.
- AdventureWorksLT2017.bak.
- Azure Account

Files
dev
Go to file t
1 What is Data Engineering
2 Key Concepts of Data Engineeri...
3 Essential Skills for Data Engineers
4 Data Engineering vs Data Science
5 Building a Simple Data Pipeline...
notes
steps_for_demo.md
slides
src
FolderTemplate
README.md

datacraftconnect addingdemosteps c4fa1d4 · 4 days ago History

Data Pipeline Demo

Overview

This subfolder contains scripts and configurations for the data pipeline demo, which moves data from a local SQL Server (AdventureWorksLT2019) to Azure Blob Storage using Azure Data Factory (ADF) and then creates a table in Azure SQL Database.

Steps for Demo

Step 1: Create Azure Blob Storage

1. **Sign in to the Azure portal.**
2. **Create a storage account**:
   - Navigate to "Storage accounts" and click "Create".
   - Fill in the required details (Subscription, Resource group, Storage account name, etc.).
   - Click "Review + create" and then "Create".
3. **Create a container**:
   - Go to your storage account.
   - Under "Data storage", select "Containers".
   - Click "+ Container", name it (e.g., `mycontainer`), and set the public access level to "Private".

Step 2: Set Up Azure Data Factory (ADF)

1. **Create a Data Factory**:
   - In the Azure portal, search for "Data factories" and click "Create".

Key Steps

- **Set Up Azure Blob Storage:**
  Create a storage account and container.
- **Configure Azure Data Factory:**
  Create a Data Factory instance.
  Install Self-hosted IR on the local machine.
- **Create Linked Services:**
  Connect to local SQL Server and Azure Blob Storage in ADF.
- **Build and Run Pipeline:**
  Design a pipeline to copy data from SQL Server to Blob Storage.

Business Benefits

- **Scalability**: Scale data storage and processing in the cloud.
- **Accessibility**: Enable remote data access for global teams.
- **Advanced Analytics**: Utilize Azure's analytics tools for better insights.
- **Cost Efficiency**: Reduce on-premises infrastructure costs.

Conclusion

- Demonstrated building a data pipeline using ADF, to migrate data from on premise database server to Azure Blob Storage.