

Mini Project Report

Project Title:

Data Export from Azure SQL to CSV, Parquet, and Avro using Azure Data Factory

Author:

Samiksha Kharche

Abstract:

This mini-project demonstrates how to extract data from an Azure SQL Database and export it into three different file formats - CSV, Parquet, and Avro - using Azure Data Factory. This approach supports various downstream use cases such as data warehousing, reporting, and analytics.

Technologies & Services Used:

- Azure SQL Database
- Azure Blob Storage
- Azure Data Factory (ADF V2)
- File formats: CSV, Parquet, Avro

Step-by-Step Implementation:

1. Created Azure SQL Database with a sample 'Students' table.
2. Created a Blob Storage container (export-data) with required directories.
3. Set up Azure Data Factory (V2) in the same region.
4. Created Linked Services:
 - Azure SQL Database (source)
 - Azure Blob Storage (sink)
5. Defined Datasets:
 - SQL_Students (source dataset)
 - CSV_Students, Parquet_Students, Avro_Students (sink datasets)
6. Built a pipeline (ExportToAllFormats) with three parallel Copy Data activities:

Mini Project Report

- Copy SQL data to CSV
- Copy SQL data to Parquet
- Copy SQL data to Avro

7. Published the pipeline and triggered it manually.

8. Verified successful file creation in the Blob Storage container.

Conclusion:

This project provides a hands-on understanding of integrating Azure services for efficient data movement and format conversion. It showcases the flexibility of Azure Data Factory in handling multiple file outputs from a relational source, which is highly useful in real-world data workflows.