Consume Data from REST API into Azure Data Lake

Table of Contents

1. Introduction

This document outlines how to invoke a REST API using Azure PaaS components (Azure Data Factory i.e. ADF) and then ingest the data into ADLS or Azure SQL DB into tabular structure for further use.

2. Solution Components

A **RESTful API** is an application program interface (API) that uses HTTP requests to GET, PUT, POST and DELETE data and is also referred to as a RESTful web service -- is based on representational state transfer (REST) technology, an architectural style and approach to communications often used in web services development.

Azure Data Factory is the platform for these kinds of scenarios. It is a cloud-based data integration service that allows you to create data-driven workflows in the cloud that orchestrate and automate data movement and data transformation.

Microsoft **Azure Data Lake** is a highly scalable public cloud service that allows developers, scientists, business professionals and other Microsoft customers to gain insight from large, complex data sets.

Microsoft **Azure SQL Database** is a managed cloud database provided as part of Microsoft Azure. A cloud database is a database that runs on a cloud computing platform, and access to it is provided as a service. Managed database services take care of scalability, backup, and high availability of the database.

3. Solution Diagram

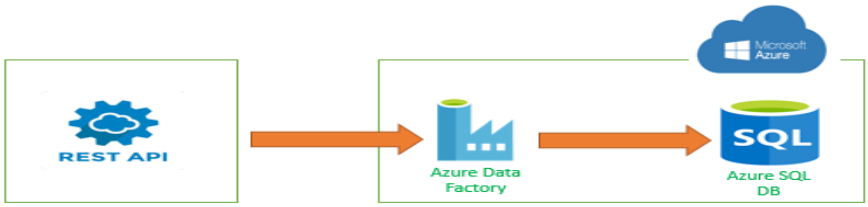


Fig-1

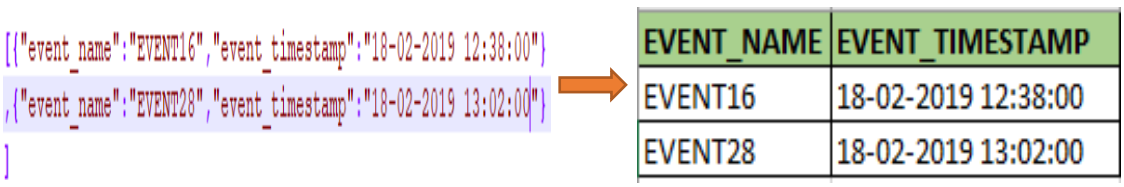


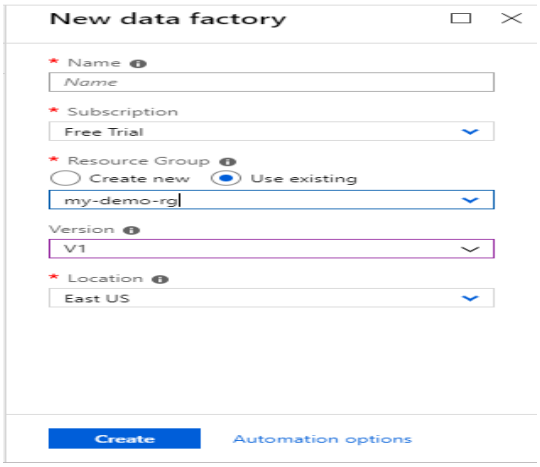
Fig-2

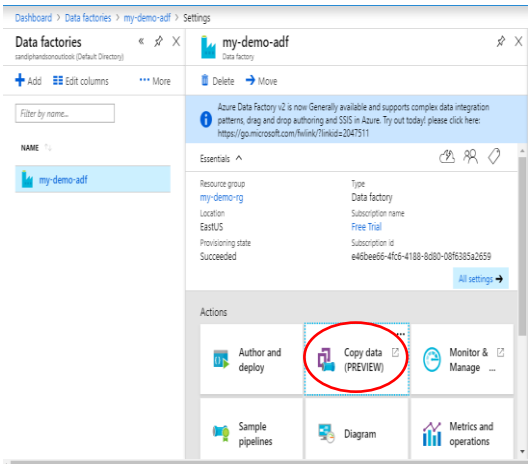
4. Detailed Steps

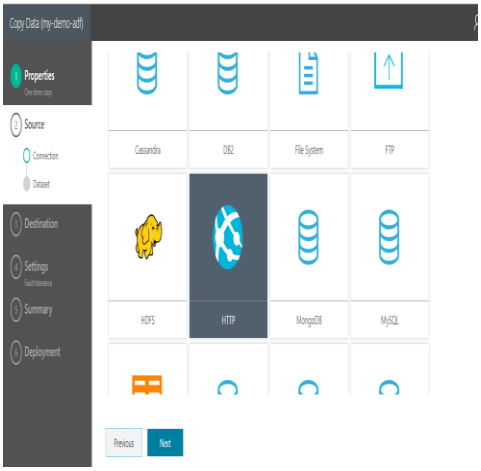
Step 1 - Identify the REST API you need to consume along with the pre-requisites (credentials, api keys or auth tokens etc.) handy.

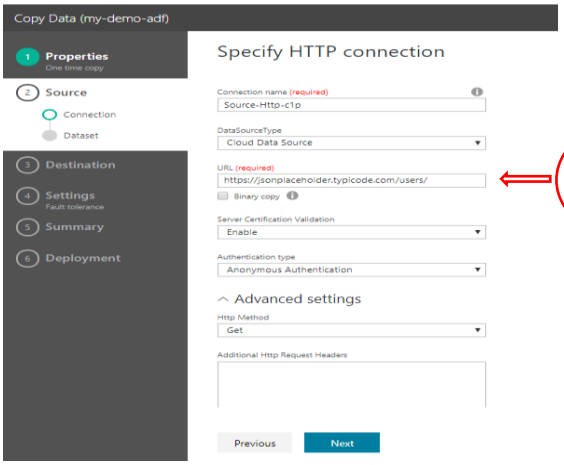
E.g - You can register yourself in https://developer.nytimes.com/apis and generate the required API key/tokens to access the REST API hosted there but here I will demonstrate the simple case (no authentication needed) with anonymous user.

Step 2 – Create the ADF pipeline with HTTP as source and SQL DB as destination.

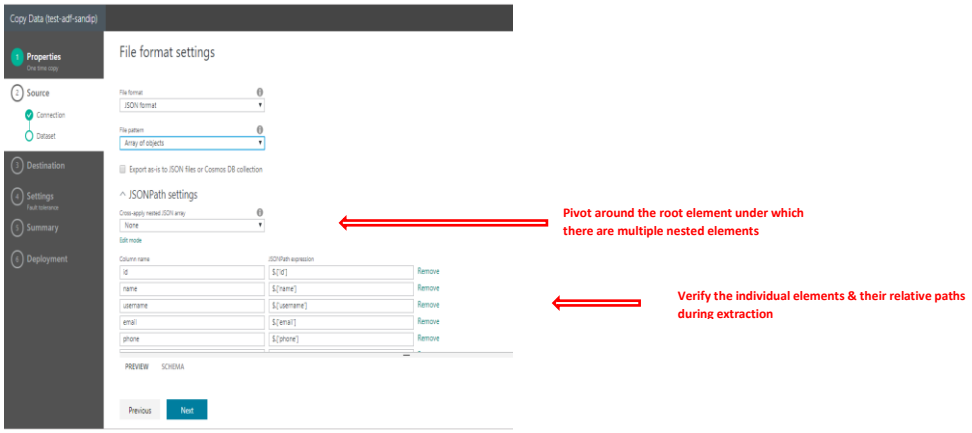




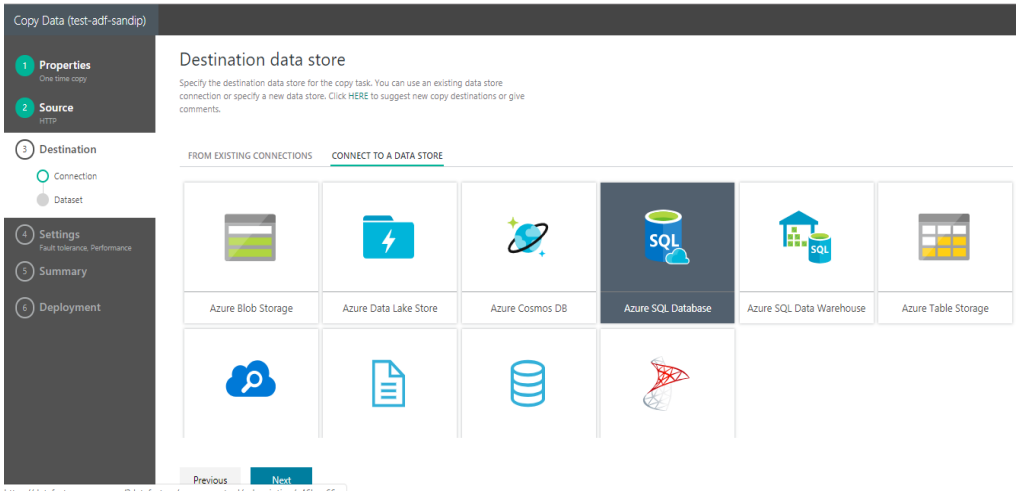


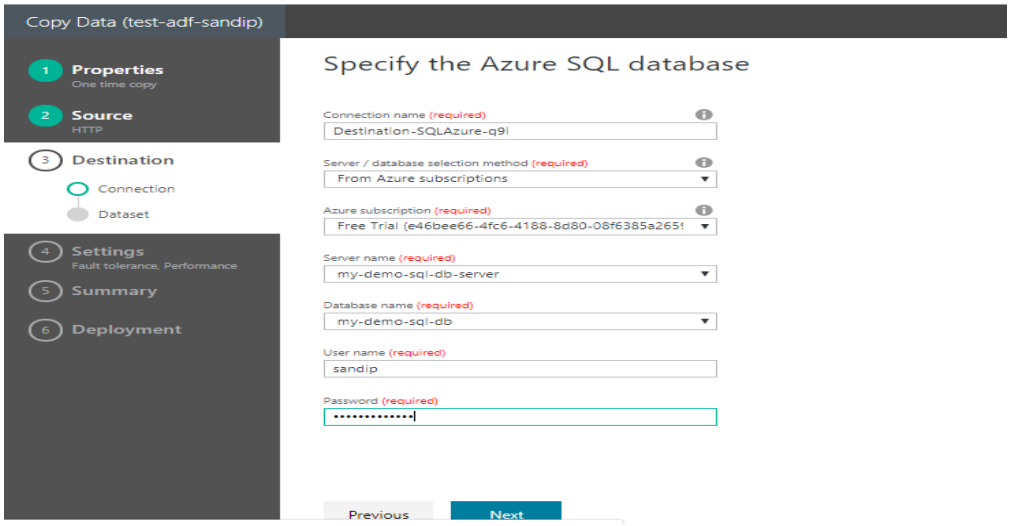


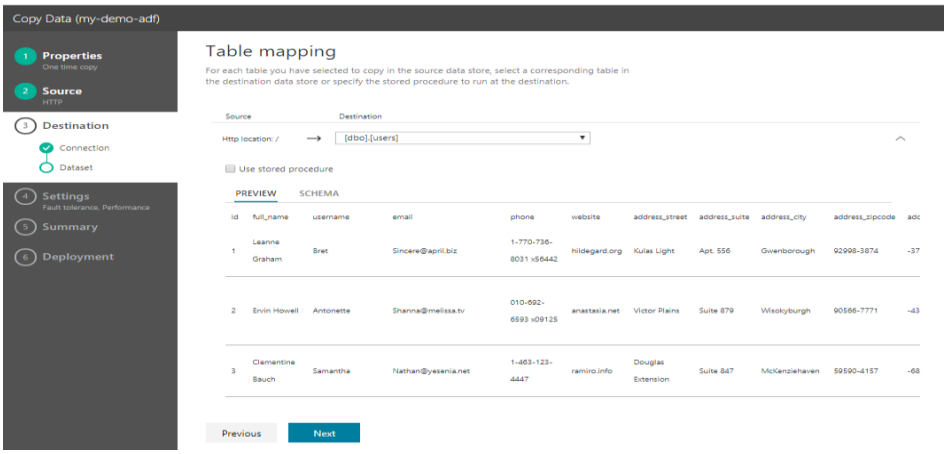
Pass access tokens/API keys as URL parameter (if required)



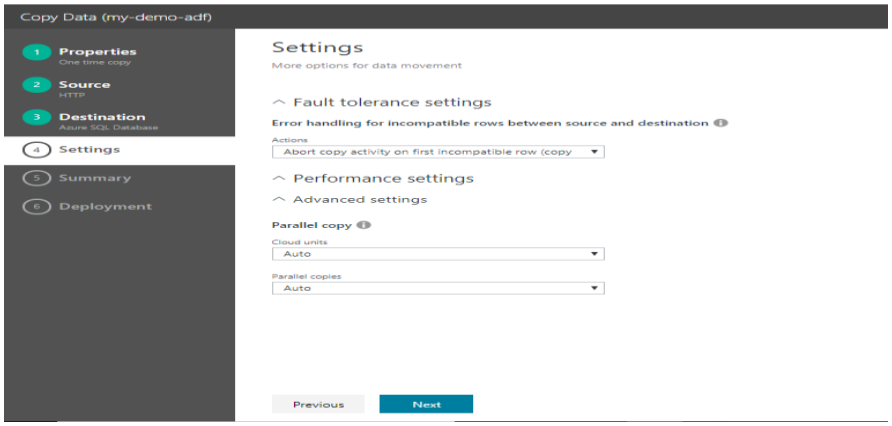
Step 3 – Select SQL DB as destination of the ADF pipeline. If you just want to dump the data into storage, then you can choose either Blob or ADLS as destination instead of SQL DB.



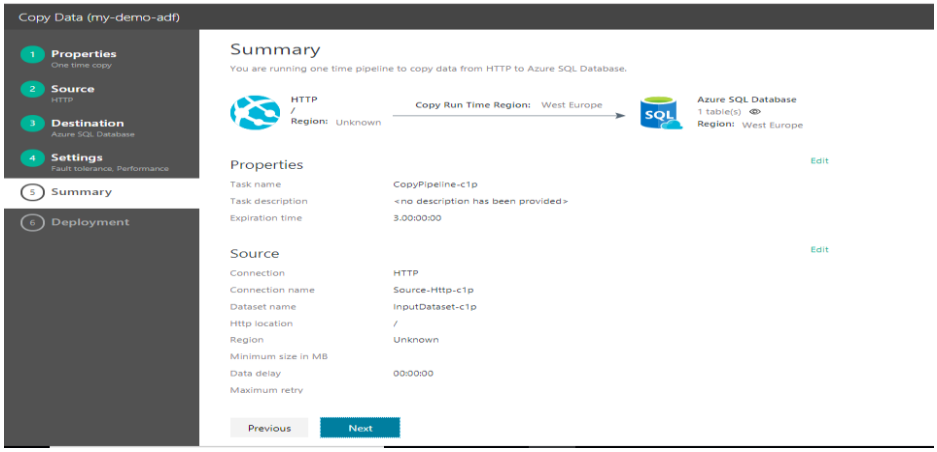


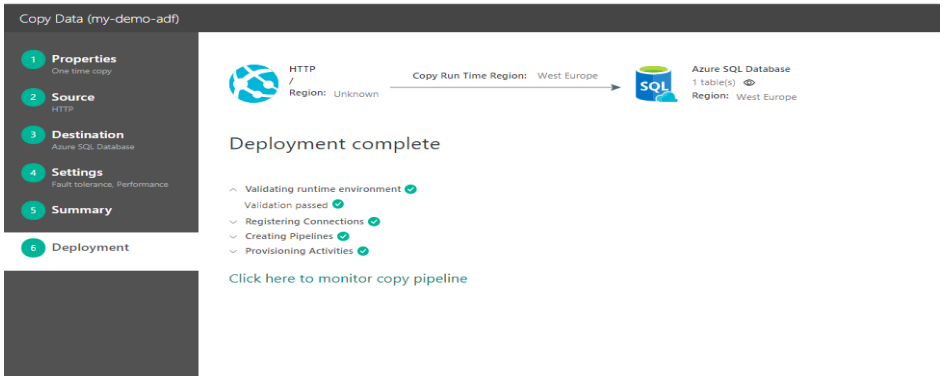


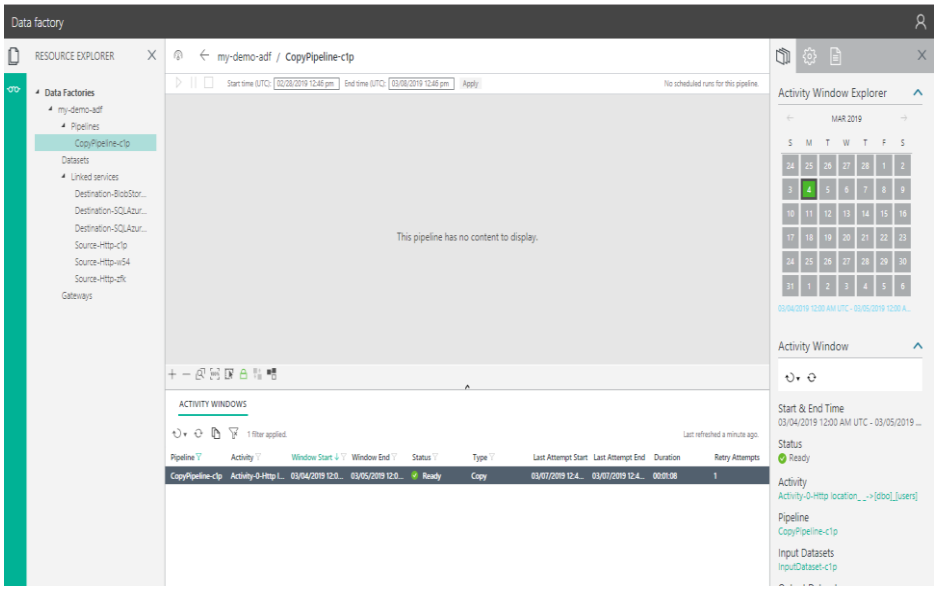
Step 4 – Specify the settings



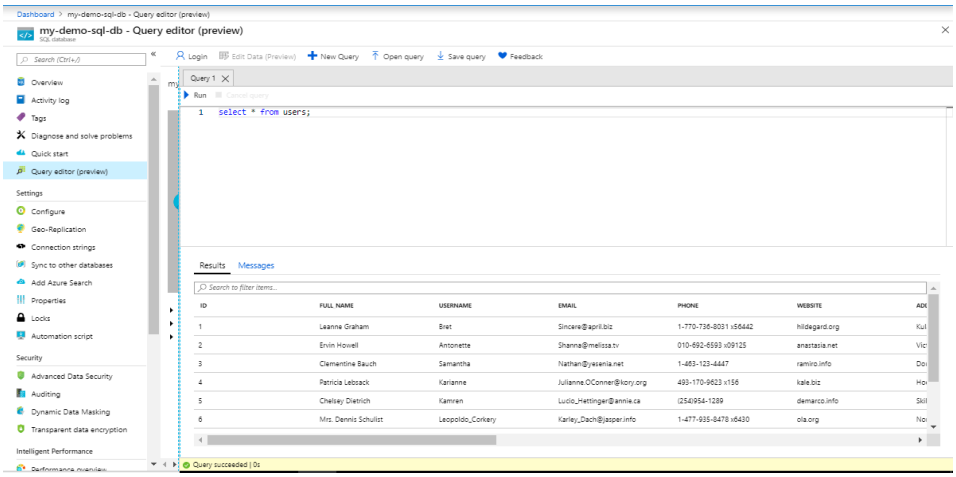
Step 5 – Review the Summary and initiate Deployment.







Step 6 – Once the deployment status is ready, go to SQL DB window and check whether records from REST API response have been captured there or not.



5. References

https://docs.microsoft.com/en-us/azure/data-factory/connector-http

https://docs.microsoft.com/en-us/azure/data-factory/tutorial-copy-data-tool

https://docs.microsoft.com/en-us/azure/sql-database/