Data Movement: Data Lake Store to Azure SQL DW

Updated on: 1/13/2017

Introduction

You will learn how to setup a recurring job to run and how to copy the output of that job in a recurring format from the Data Lake Store to SQL DW. This is a common pattern employed to move transformed data to a database for reporting/analytics on aggregated data scenarios.

Prerequisites

For this lab, you will need:

- Access to a Data Lake Store account that you can write to
- Access to a Data Lake Analytics account that you can submit jobs to
- Access to the **adltrainingsampledata** Data Lake Store account that you can read from
    - This means be a member of the ADLTrainingUsers security group.
- Access to a Data Factory account where you can author
- Access to an Azure SQL DW (or Azure SQL DB if you prefer)

Create Table in SQL DW

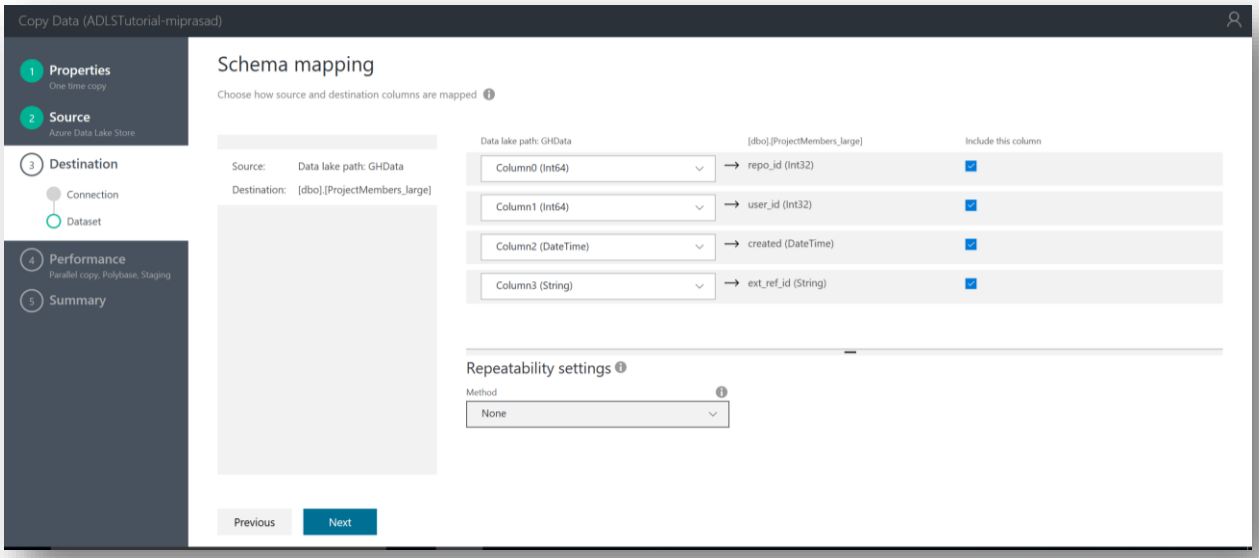
- Connect to your SQL DW using Visual Studio
- Run this query to create a table called **ProjectMembers_large**

```
CREATE TABLE ProjectMembers_large
(
         repo_id int,
         user_id int,
         created datetime,
         ext_ref_id varchar(255)
);
```

Create a copy pipeline in ADF

- Open your ADF account
- click on **Copy data (PREVIEW)**
- Setup properties for the ADF Task
    - **Name** for the Task(Activity) -> call it any thing you want
    - Set **Recurrence** to Daily
- Select the source data store to be Data Lake Store

- Setup a Linked Service for the Data Lake Store
    - Type in a name (or use the default)
    - Select the Data Lake Store account
    - Browse to the folder containing your data
- Select the output data store to be Azure SQL DW
    - Fill in the connection info for the Azure SQL DW you want to use
    - Select the table created in Step #1 and the columns will be mapped automatically



- Create the pipeline and Deploy it

Verify that data was copied

- Run this query in the Azure SQL DW

```
SELECT TOP 3 * FROM [dbo].[ProjectMembers_large]
```