Handbook ADF as ETL for SC Navigator

This handbook describes how to initially set up and configure your Azure Data Factory instance for usage as an ETL for inputting data for usage in the SC Navigator application.

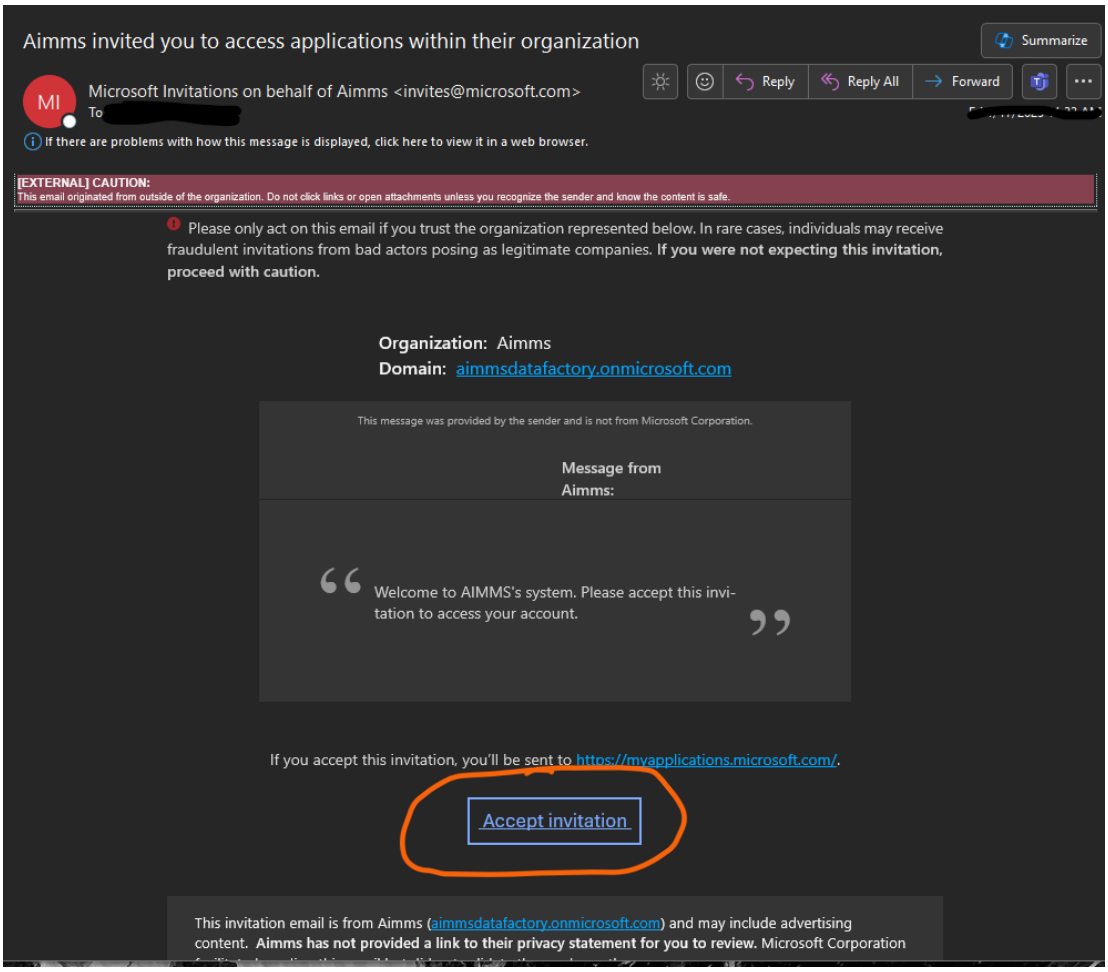
Make sure you have the following ready:

- *An email address you want to use to login to and configure the Azure Data Factory instance;*
- *A mobile phone which you can use for the Multi Factor Authentication (MFA). This will require you to install the Microsoft Authenticator App OR to use another MFA-app of your choice.*

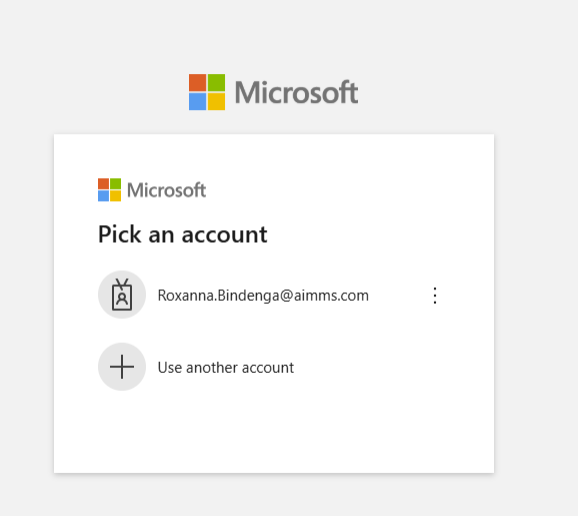
1. Initial setup 2
2. Transforming your data 8
3. Ingesting data to SC Navigator 9
   3.1 Linked Service within your ADF instance 9
   3.2 Copying data to the linked service 11

1. Initial setup

You will receive an invitation via email from Microsoft Invitations on behalf of AIMMS which you will need to accept by clicking the ‘Accept invitation’ button:

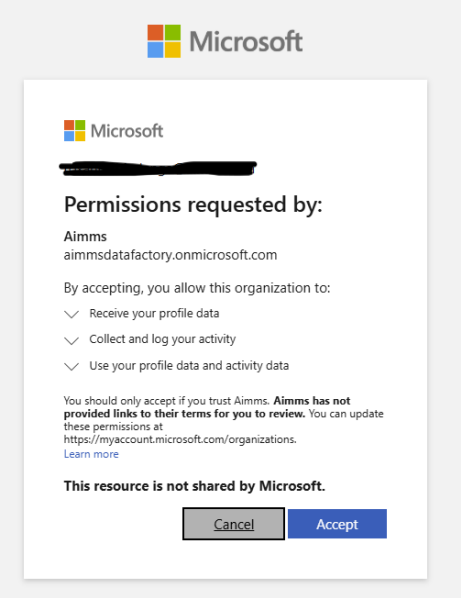


If you already have a Microsoft account, you are now requested to login. If you don't have a Microsoft account, or you want to use another account, you can choose to do so here too:

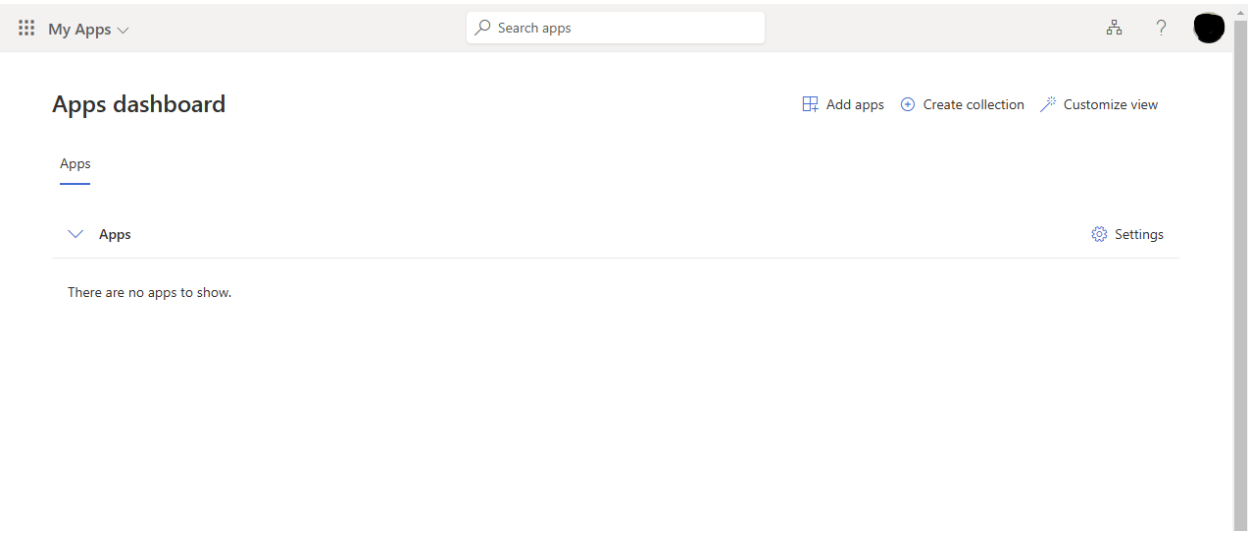


Please note that the Azure Data Factory instance is on a separate tenant and will in no way connect to any of your other linked services on Azure.

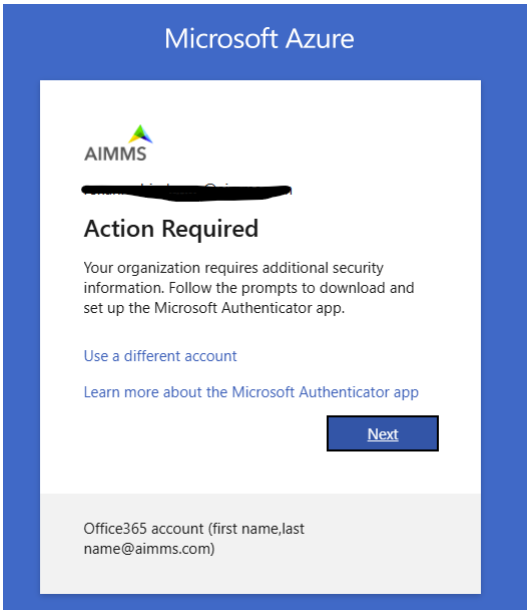
Microsoft will then request you to accept the following:



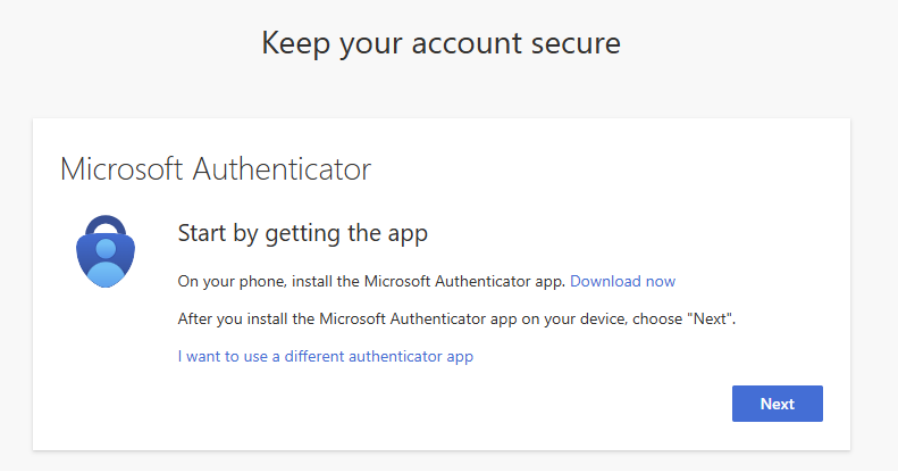
Choose Accept. It will lead you to a screen like this:



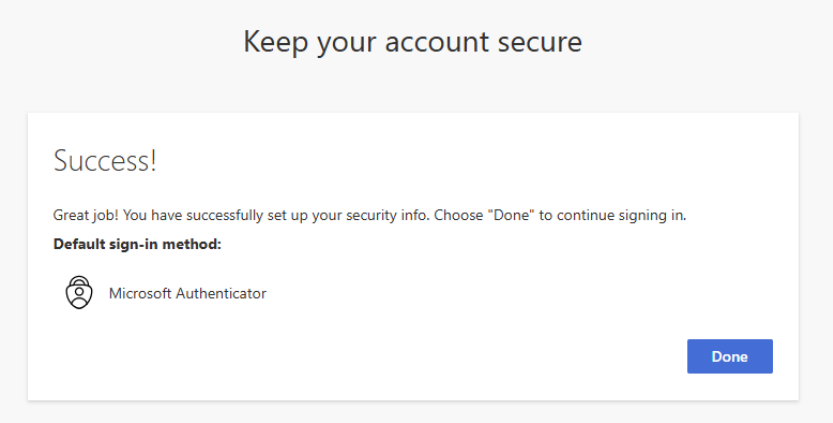
There's nothing to do here, but now that you're logged in you can go to the URL that was sent to you in the introduction email. You can also go to https://portal.azure.com. If this is your first time logging in, Microsoft Azure will request you to configure MFA.



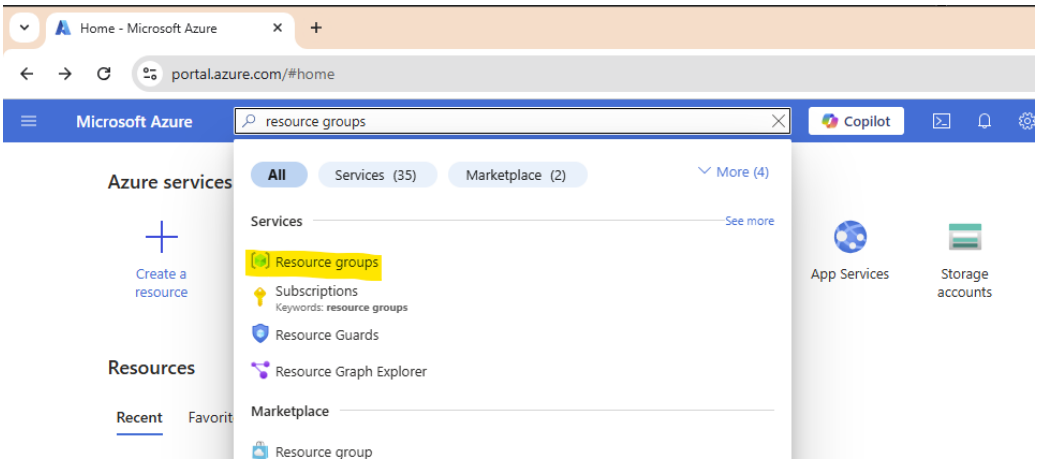
Pressing ‘Next’ will lead you to this screen:



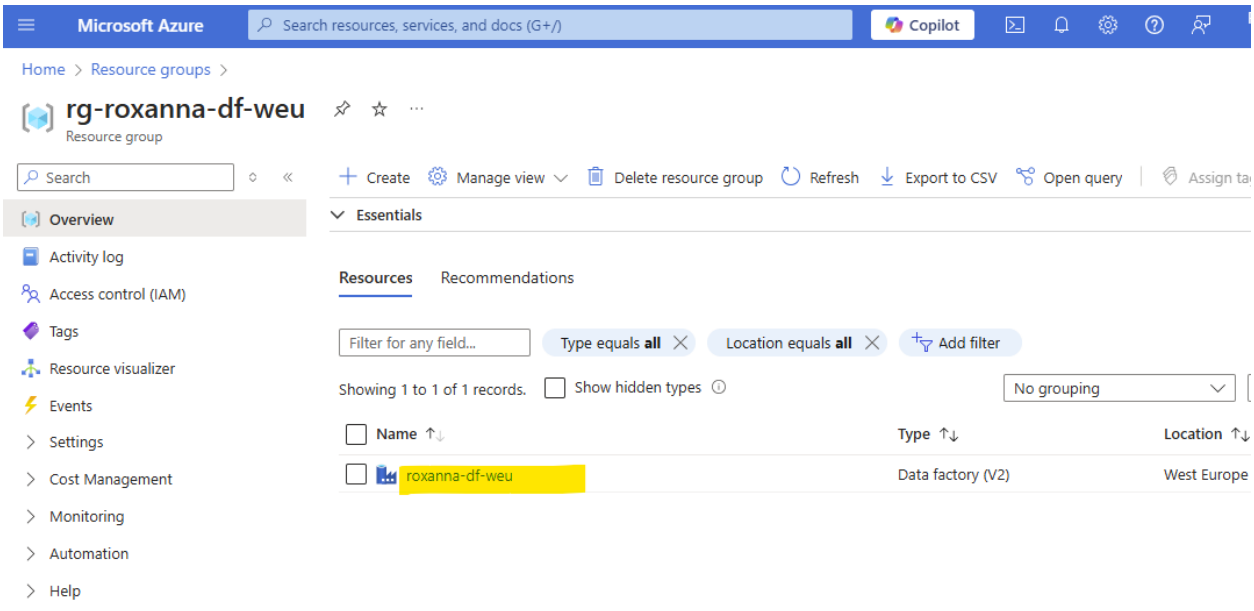
You can choose here to either download the Microsoft Authenticator App on your mobile device or choose another authenticator app to set up the MFA. Follow the instructions on the screen. If all went successfully, you will see a success screen:



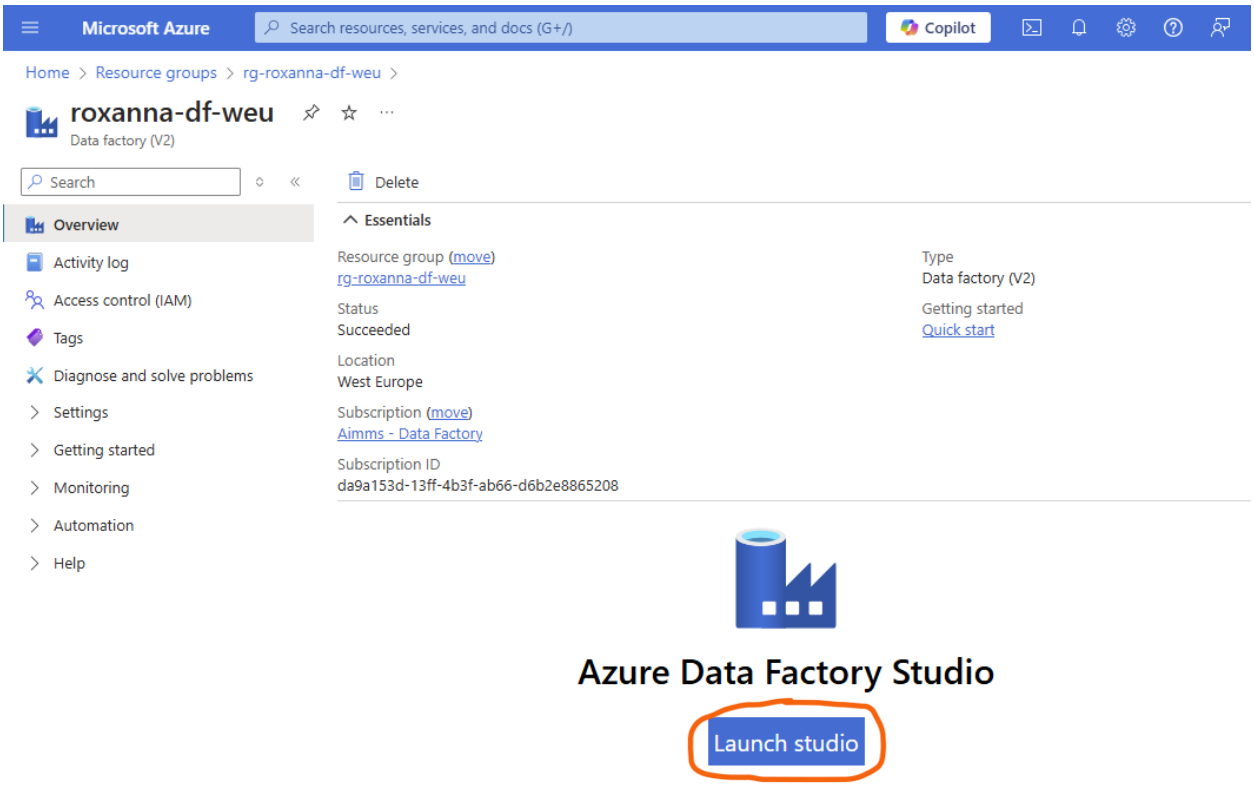
Now you can go to https://portal.azure.com. From here, go to Resource groups. If this is not visible, use the search bar and type ‘Resource groups’ and it will show up in the search results:



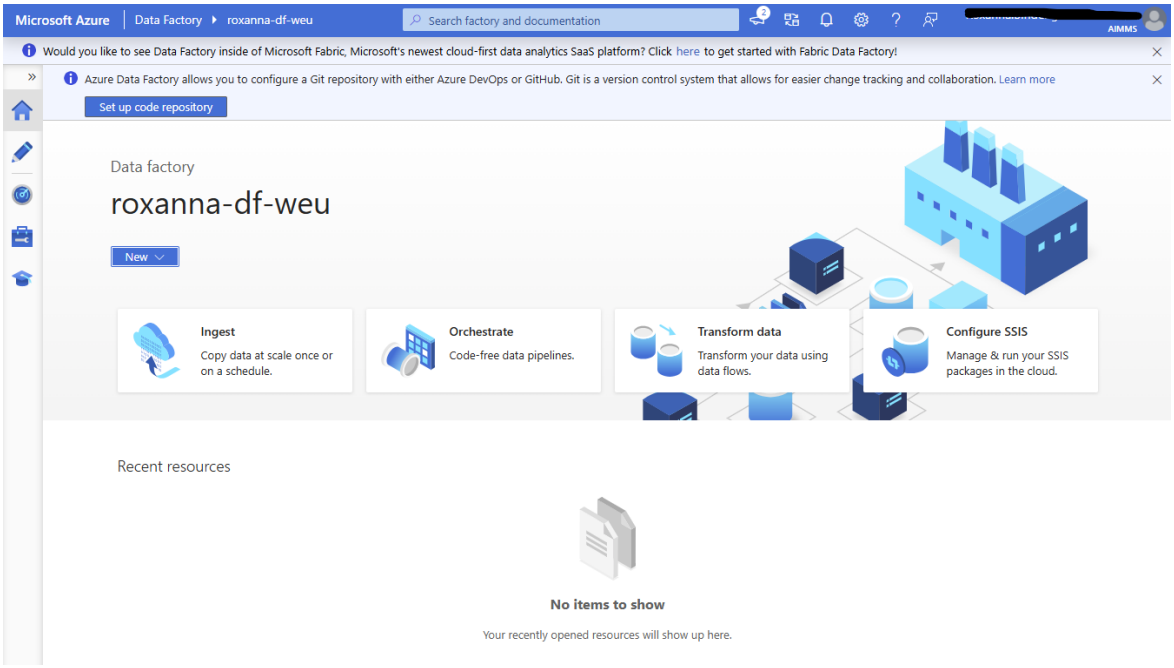
There will only be one resource group available. Click it, it will lead you to a screen where you can access the ADF, click it:



And then click ‘Launch studio’:



You are now in the Azure Data Factory instance and can use it in any way you want.


Microsoft Azure
Data Factory
roxanna-df-weu
Search factory and documentation
AIMMS
Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click here to get started with Fabric Data Factory!
Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration. Learn more
Set up code repository
Data factory
roxanna-df-weu
New
Ingest
Copy data at scale once or on a schedule.
Orchestrate
Code-free data pipelines.
Transform data
Transform your data using data flows.
Configure SSIS
Manage & run your SSIS packages in the cloud.
Recent resources
No items to show
Your recently opened resources will show up here.

2. Transforming your data

How you want and/or should transform your data fully depends on what your input data looks like and what should be done with it before merging it into the SC Navigator template.

Here is a list of documentation that might help you in this process:

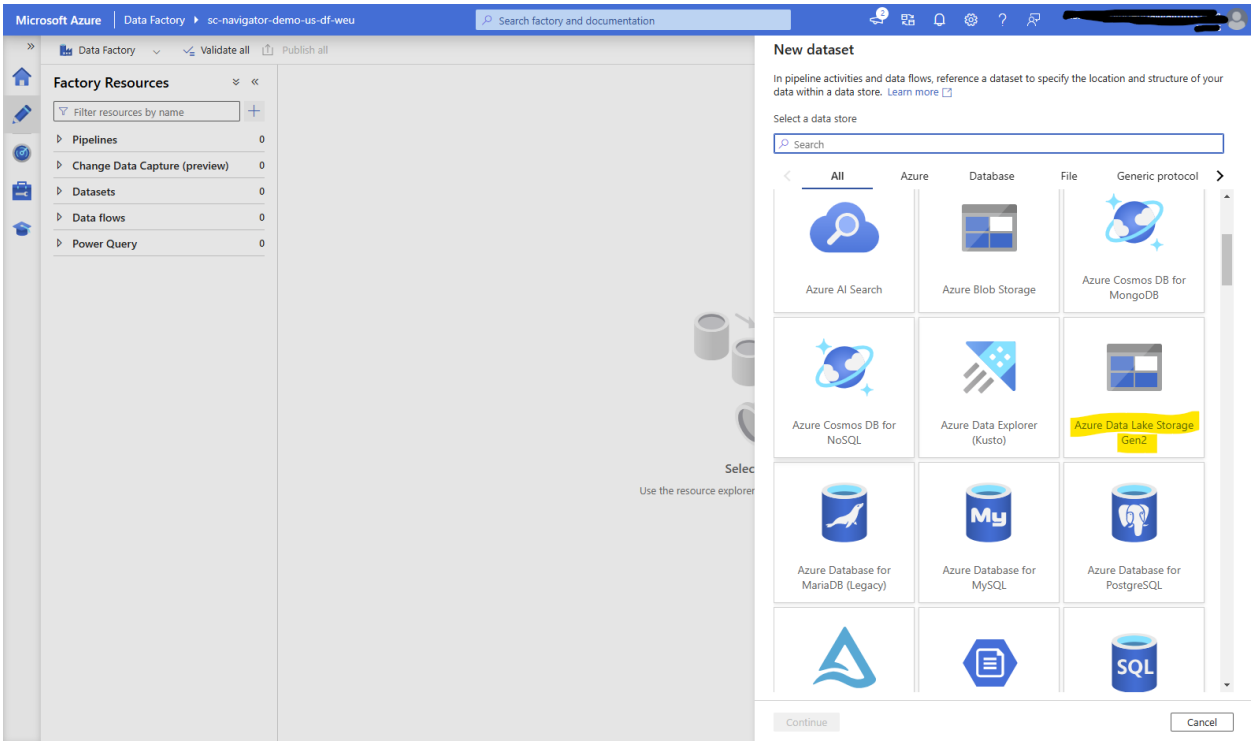
- Microsoft offers a lot of documentation related to Azure Data Factory. This would be the starting point for that.
- To learn how to copy/ingest data into Azure Data Factory, here is a list of tutorials.
- Microsoft's YouTube page on Azure Data Factory offers a lot of videos about data wrangling and data flows.
- This guide will help you understand more about 'exporting' your result data; more specifically this guide explains more in depth about copying data to Azure blob storage (which is part of the SC Navigator cloud).

3. Ingesting data to SC Navigator

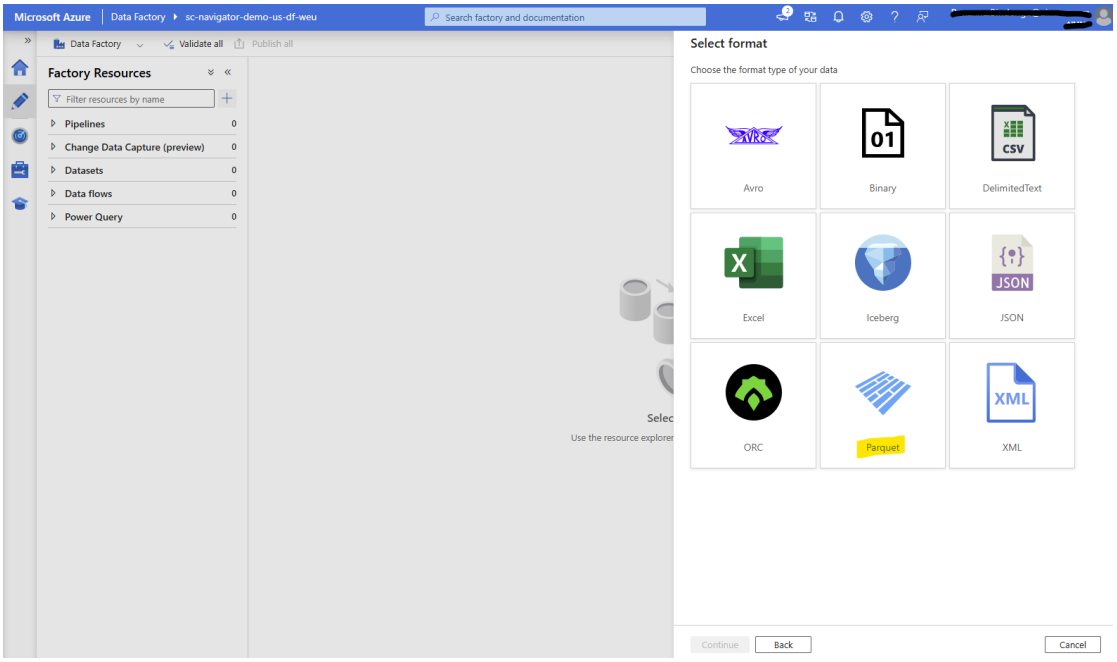
When you've transformed your data as preferred, you will need to push it as a dataset to SC Navigator so you can start using it in the application. In your ADF, we have automatically set up a link (through Storage Account Key) to your Azure Data Lake Storage Gen2 attached to your SCN Cloud-account.

3.1 Linked Service within your ADF instance

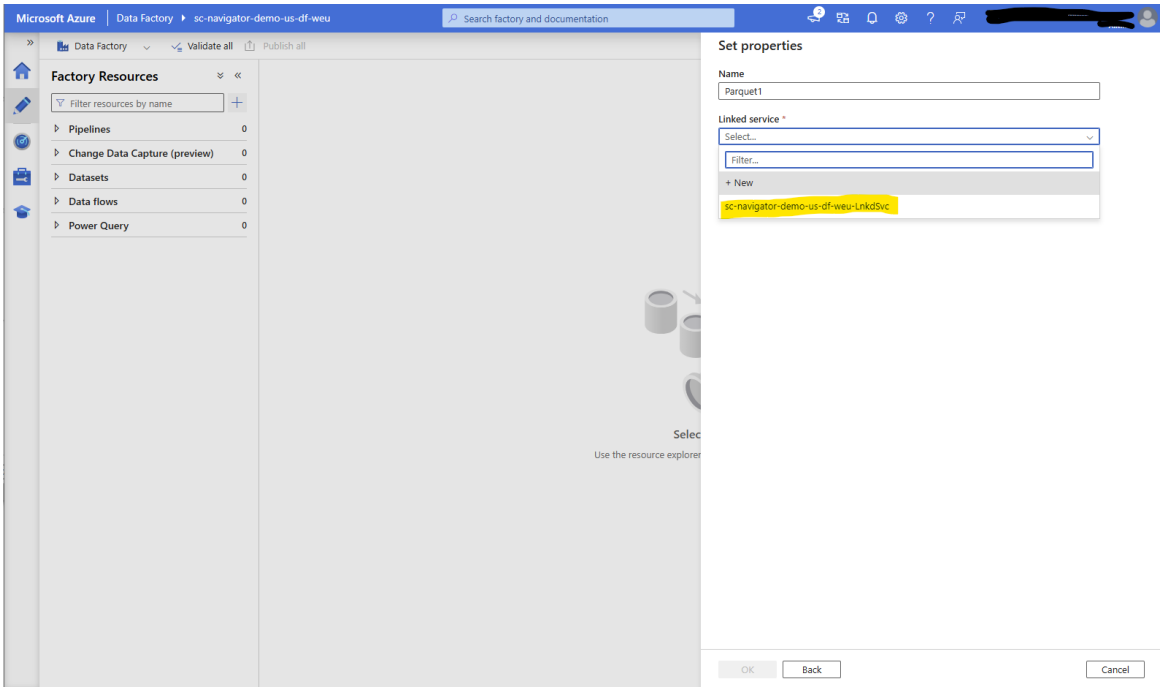
By default, there will be a connection preconfigured with your Azure Data Lake Storage Gen2 (ADLS) on the AIMMS Cloud, connected to your SC Navigator application. This is the location where you should push your input files to if you want to use them as a dataset in the SC Navigator environment. Whenever you are configuring a data-involved step, you get the possibility to choose from predefined sources. Here you should choose the Azure Data Lake Storage Gen2:



In the following step, you can choose a format. In the context of exporting data from your pipeline to the SC Navigator environment, the data should always be exported in the pre-defined parquet-file format:



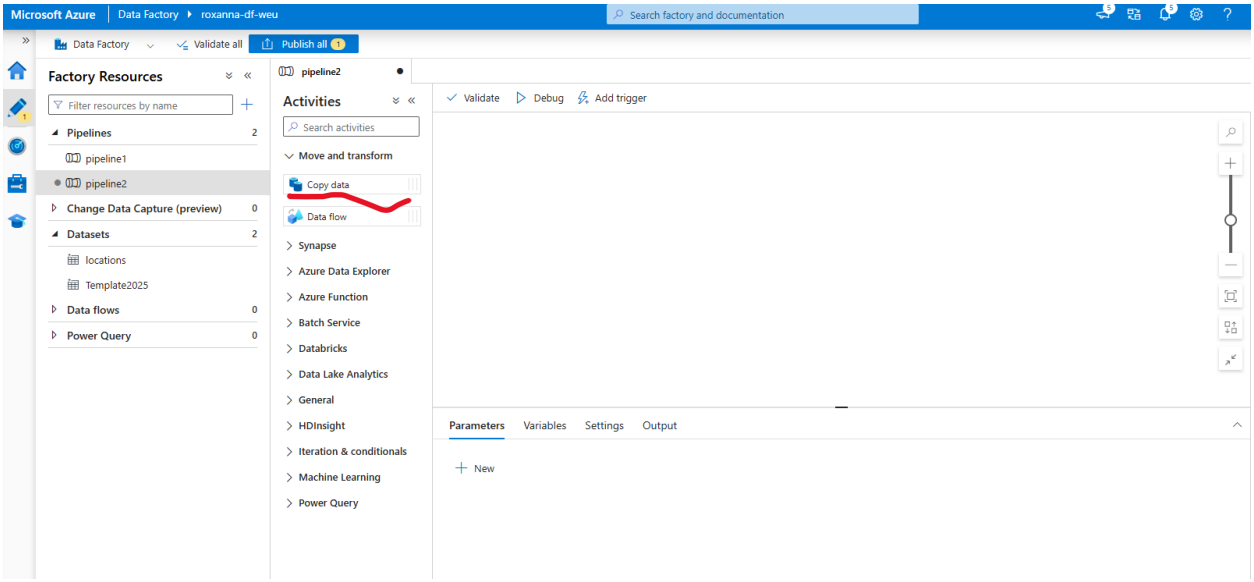
In the next step, you should by default be able to select a Linked service which looks something like this:



This is the pre-configured connector to your related Azure Data Lake Storage on the SC Navigator cloud. The folder you should be exporting to is always called '**sc-navigator-import**'.

3.2 Copying data to the linked service

To export your data to the Azure Data Lake Storage related to your SC Navigator cloud, in a pipeline you need to use the “Copy data” activity which can be found under activities in the category “Move and transform”. See the image below:



In this activity you can configure a “source” and a “sink”. The source refers to the input dataset (which you loaded or transformed with ADF) and the sink refers to the final result (a parquet file for SC Navigator).

In the source tab of the copy data activity select the source dataset (which is the input dataset). Next in the sink tab, you need to click on new then on Azure Blob Storage and continue. In this screen you need to select the file type which is “Parquet”. Provide a name (which is the file name!) and select your linked service. Please use a file name which corresponds to the template of SC Navigator (for instance “locations”).

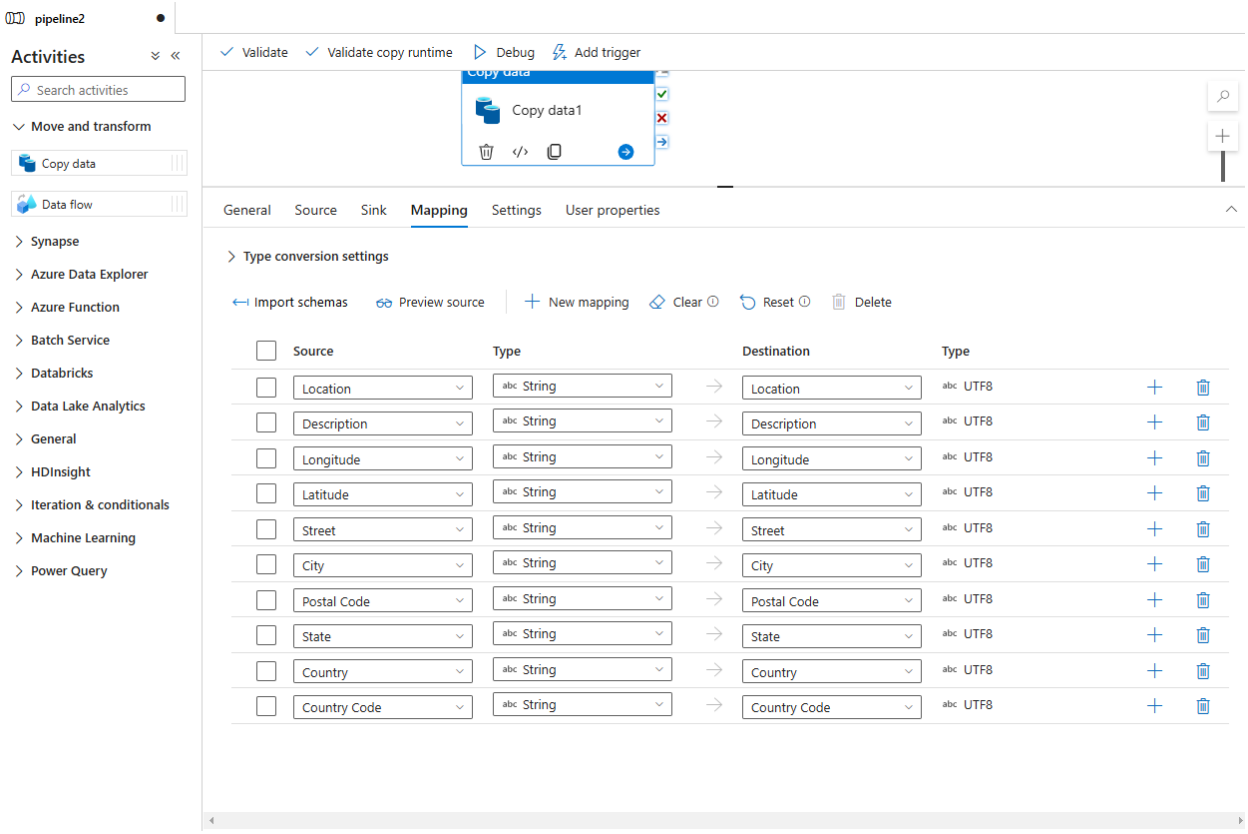
Your linked service can be a data source you've added manually, or you can make use of the preconfigured link to your Azure Data Lake Storage related to your SC Navigator Cloud account.

In the next step you can select a place to store the parquet file. You need to select a container, a folder, and a filename. Make sure you store the file in the “sc-navigator-import” container. This is the container where SC Navigator looks when you open the load dataset dialogue. The dataset will appear in this dialogue when there is a folder in the “sc-navigator-import” container. The name of the folder is used as the dataset name.

It’s good practice to make a new folder/dataset for every import run. You can come up with a dataset name yourself or you could dynamically let ADF make a dataset name for example based on the date.

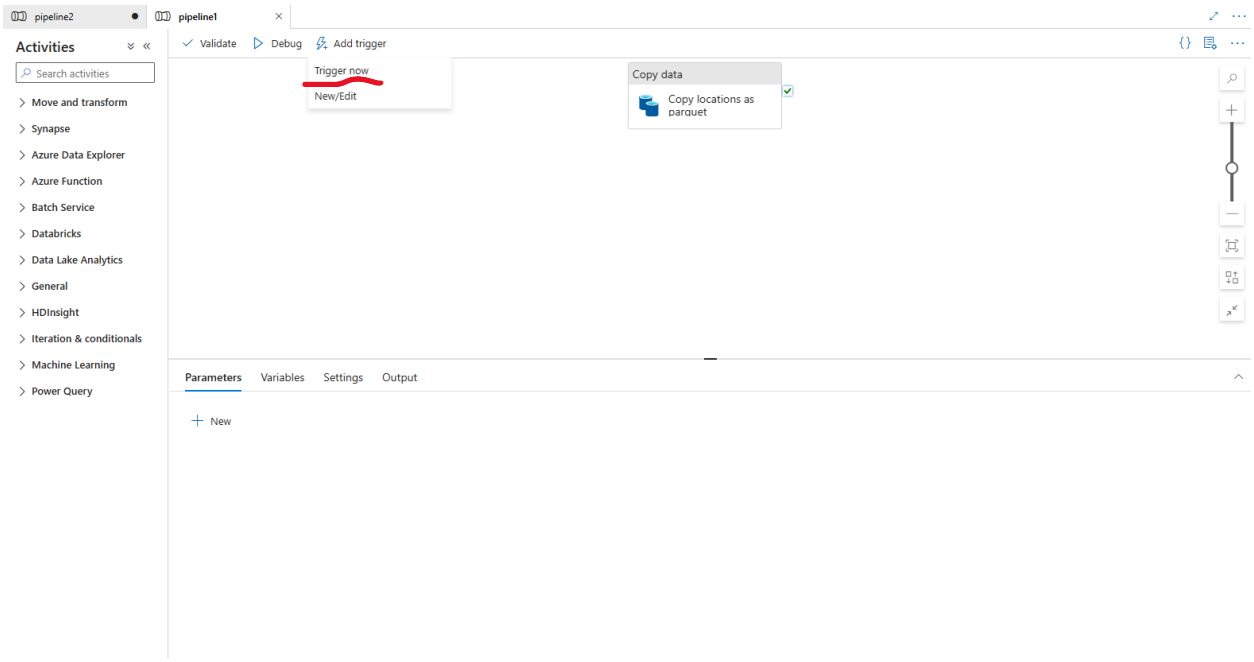
At this point you don't need to provide a file name. You already did this with the name property. If you provide a file name you will get an error message that the blob does not exist.

The next step is to define the mapping from source to sink. In the mapping tab you can click on the import schemas button to create schema mappings. In the example below we have source data with locations which we would like to write to the locations.parquet. In the screenshot you will see the generated mappings by ADF:



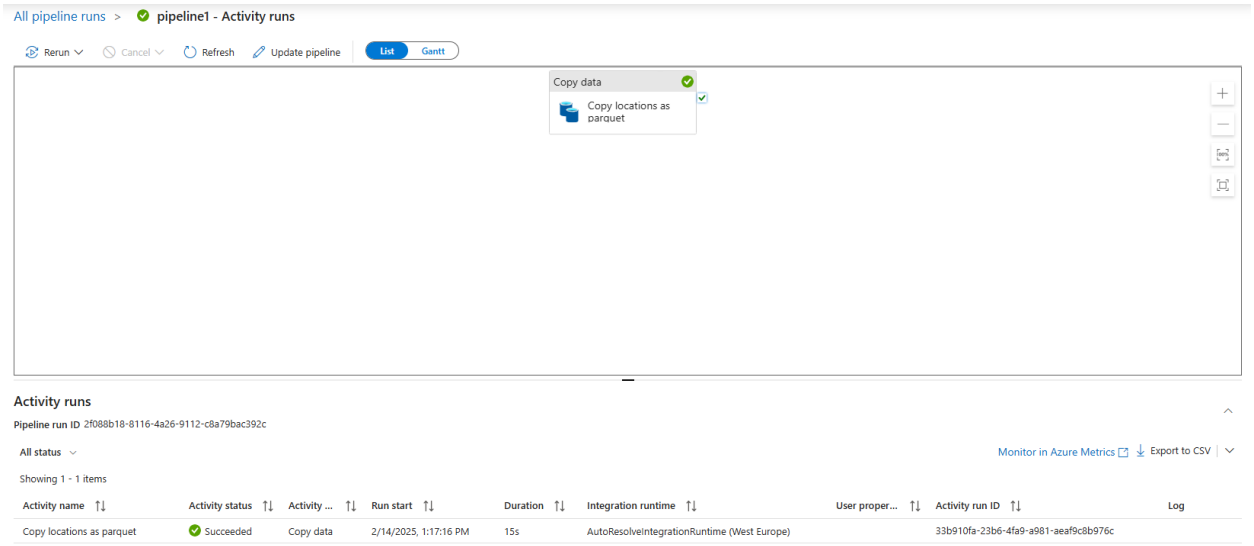
The generated mapping looks correct, however we need to change the column names in the destination since Paquet does not support whitespace in the column names. When copying the data into a parquet file, make sure that you removed spaces and other special characters (like "[]{}/-"). The Parquet file format is strict on these conventions and any deviations from it will result in an error. Please refer to the SC Navigator template for Parquet files which file names and column names are correct (you can download this file in the data section of SC Navigator).

Once we changed "Postal Code" and "Country Code" to "Postal_Code" and "Country_Code", we need to publish the pipeline using the blue "Publish all" button at the top of the screen. Now we are ready to run the pipeline!



You can run the pipeline using the “Add trigger” dropdown and click on the “Trigger now” button. A pop-up message will appear where you need to confirm that you would like to trigger this pipeline. Once you confirm a message will appear that the pipeline is running, and you can view this pipeline run. If you click on this message, it will bring you to “Monitor” screen where you can see the status of the pipeline execution.

If all went well, you will see a successful run which means that the file is successfully imported in the SC Navigator environment.



Once all the required files for a dataset are in the SC Navigator environment in the corresponding dataset folder. You can go to SC Navigator and open the import dataset dialogue. In this dialogue you will see the dataset you created with ADF, and you can use it to create your scenario’s and do your analysis.