Creating a dataflow

A **dataflow** is a collection of tables that are created and managed in workspaces in the Power BI service. A **table** is a set of columns that are used to store data, much like a table within a database. You can add and edit tables in your dataflow, and manage data refresh schedules, directly from the workspace in which your dataflow was created.

A very common reason to use **dataflows** is the reusability of data among team members. Consistency and having a single source of truth is the main goal for many analysts, and a great application of this is a **Date** table. For this exercise, a Pro license with no additional premium capacity is necessary. In this section, we will explore the Power BI service to see where dataflows are created and then use simple code to produce a dataflow that will work in many different data models.

Dataflows are reusable **extract, transform, and load** (**ETL**) packages created and managed in the Power BI service, while the data you bring into a Power BI dataflow is stored as entities—basically flat tables—in the Common Data Model folders in Azure Data Lake Storage Gen2. These files are stored as CSVs with JSON files containing all the metadata and rules, which unify and standardize your self-service data warehouse. Once created, these dataflows then serve as a data source for Power BI reports, and can also be used with other Azure data services if you bring your own data lake.

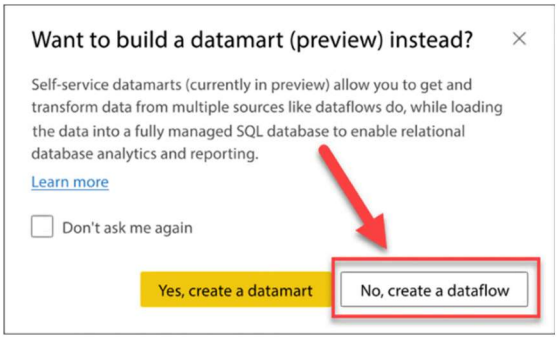
In order to create a Power BI dataflow, you need a new **workspace** experience. This has been the default for some time now—you cannot use Power BI dataflows in much older workspaces. When creating the workspace, there are advanced options to set up dedicated dataflow storage capacity—in other words, to connect your own data lake for additional storage capacity and options within the Azure ecosystem. Be sure to read the extensive admin documentation if your goal is to use your own data lake as the backend storage, rather than leveraging the default experience.

One key point is that you will want to ensure your capacity is set to the same region as your other Azure resources. Using My Workspace will never work, regardless of the region and capacity, because My Workspace is a personal folder that's not designed for sharing data organizationally.

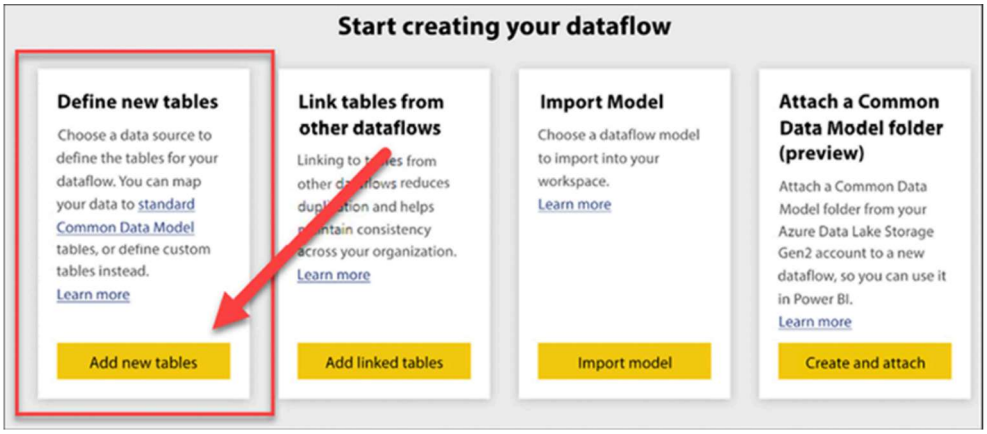
First, **create a new workspace** in your version of Power BI. A key detail is that the creator of the dataflow becomes the owner by default. Previously, ownership could not be transferred, but now there is an option to pass dataflow ownership to another Power BI user. The dataflow owner is the primary admin for dataflow functionality and permissions.

Once the workspace is created, there is an option for **Dataflow** under **+ New** toward the upper-left corner of your workspace.

When this option is selected, you will see a popup related to the new datamarts feature.



Confirm that you want to create a dataflow and not a datamart, to get started. After selecting **No, create a dataflow**, more options appear:



It's worth describing the various options:

- **Define new tables**: Once a dataflow contains data, you can create **computed tables**. Computed tables perform in-storage computations and they may be custom, or they can map to the **Common Data Model** standards.
- **Link tables from other dataflows**: Link to tables that have branched off from existing tables in the same or different workspaces. They are branches that allow transformation, but one key point is that they do not store data. If linked tables exist in the same workspace as the original tables, refreshes may be coordinated between them. However, different workspaces lead to different refresh schedules.
- **Import Model**: Import an existing dataflow from its locally stored JSON file.
- **Attach a Common Data Model folder (preview)**: This is the "bring your own data lake" option.

Using dataflows as a data source in the desktop

◄ Lesson 8 Quiz

Siirry...

Exercise 32 - Create and use a dataflow ►