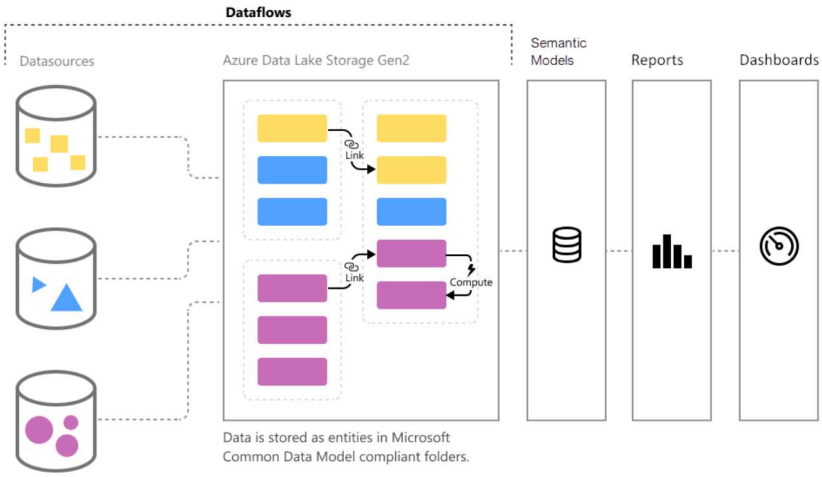
Takaisin välilehdelle

Tee: Käy oppitunti läpi loppuun asti

Getting started with dataflows

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.



Power BI dataflows can be simple or complex depending on your needs—and licensing. Licensing and workspace capacity are key factors in how you can leverage Power BI dataflows. It is important to research your organization's governance and overall architecture to conform to best practices. Below is a table from the Microsoft documentation that outlines the features and type of Power BI license needed.

Feature	Power BI
Store data in Dataverse tables (standard dataflow)	N/A
Store data in Azure Data Lake Storage (analytical dataflow)	Power BI Pro / Power BI Premium
Store data in customer-provided Azure Data Lake Storage (analytical dataflow; bring your own Azure Data Lake Storage)	Power BI Pro / Power BI Premium
The enhanced compute engine (running on Power BI Premium capacity / parallel execution of transforms)	Power BI Premium
DirectQuery connection to dataflow	Power BI Premium
AI capabilities in Power BI	Power BI Premium

Linked entities	Power BI Premium
Computed entities (in-storage transformations using M)	Power BI Premium
Schedule refresh	Power BI Pro / Power BI Premium
Dataflow authoring with Power Query Online	Power BI Pro / Power BI Premium
Dataflow management	Power BI
New connectors	Power BI Pro / Power BI Premium
Standardized schema, built-in support for Common Data Model	Power BI Pro / Power BI Premium
Dataflow data connector in Power BI Desktop	Power BI Pro / Power BI Premium
Dataflow incremental refresh	Power BI Premium

Dataflows are designed to support the following scenarios:

- Create reusable transformation logic that can be shared by many datasets and reports inside Power BI. Dataflows promote reusability of the underlying data elements, preventing the need to create separate connections with your cloud or on-premises data sources.
- Expose the data in your own Azure Data Lake Gen 2 storage, enabling you to connect other Azure services to the raw underlying data.
- Create a single source of the truth by forcing analysts to connect to the dataflows, rather than connecting to the underlying systems. This single source provides you with control over which data is accessed and how data is exposed to report creators. You can also map the data to industry standard definitions, enabling you to create tidy curated views, which can work with other services and products in the Power Platform.
- If you want to work with large data volumes and perform ETL at scale, dataflows with Power BI Premium scales more efficiently and gives you more flexibility. Dataflows supports a wide range of cloud and on-premises sources.
- Prevent analysts from having direct access to the underlying data source. Since report creators can build on top of dataflows, it might be more convenient for you to allow access to underlying data sources only to a few individuals, and then provide access to the dataflows for analysts to build on top of. This approach reduces the load to the underlying systems, and gives administrators finer control of when the systems get loaded from refreshes.

For the sake of simplicity, in this course we are giving the basic features with the Power BI Pro license; however, it may be worth a deeper dive into Microsoft's documentation to truly uncover all the dataflow features and administrative requirements needed to use all the Power BI Premium features (find out more at https://learn.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-introduction-self-service).

In the next section, we will create a practical, reusable dataflow. But before moving to that step, let's review the key differences between a dataflow and a datamart, a new preview feature that will soon be generally available.

Exploring datamarts

Datamarts help bridge the gap between business users and IT. Datamarts are self-service analytics solutions, enabling users to store and explore data that is loaded in a fully managed database. Datamarts provide a simple and optionally no-code experience to ingest data from different data sources, extract transform and load (ETL) the data using Power Query, then load it into an Azure SQL database that's fully managed and requires no tuning or optimization.

Once data is loaded into a datamart, you can additionally define relationships and policies for business intelligence and analysis. Datamarts automatically generate a dataset or semantic model, which can be used to create Power BI reports and dashboards. You can also query a datamart using a T-SQL endpoint or using a visual experience.

Datamarts offer the following benefits:

- Self-service users can easily perform relational database analytics, without the need for a database administrator
- Datamarts provide end-to-end data ingestion, preparation and exploration with SQL, including no-code experiences
- Enable building semantic models and reports within one holistic experience

Datamart features:

- 100% web-based, no other software required
- A no-code experience resulting in a fully managed datamart
- Automated performance tuning
- Built-in visual and SQL Query editor for ad-hoc analysis
- Support for SQL and other popular client tools
- Native integration with Power BI, Microsoft Office and other Microsoft analytics offerings
- Included with Power BI Premium capacities and Premium Per User.

Datamarts are targeted toward interactive data workloads for self-service scenarios. For example, if you're working in accounting or finance, you can build your own data models and collections, which you can then use to self-serve business questions and answers through T-SQL and visual query experiences. In addition, you can still use those data collections for more traditional Power BI reporting experiences. Datamarts are recommended for customers who need domain oriented, decentralized data ownership and architecture, such as users who need data as a product or a self-service data platform.

Datamarts are designed to support the following scenarios:

- **Departmental self-service data:** Centralize small to moderate data volume (approximately 100 GB) in a self-service fully managed SQL database. Datamarts enable you to designate a single store for self-service departmental downstream reporting needs (such as Excel, Power BI reports, others), thereby reducing the infrastructure in self-service solutions.

- **Relational database analytics with Power BI:** Access a datamart's data using external SQL clients. Azure Synapse and other services/tools that use T-SQL can also use datamarts in Power BI.

- **End-to-end semantic models:** Enable Power BI creators to build end-to-end solutions without dependencies on other tooling or IT teams. Datamarts gets rid of managing orchestration between dataflows and datasets through auto-generated datasets, while providing visual experiences for querying data and ad-hoc analysis, all backed by Azure SQL DB.

Datamarts and dataflows integration

In some cases it can be useful to incorporate both dataflows and datamarts in the same solution. The following situations could find incorporating both dataflows and datamarts advantageous:

- For solutions with existing dataflows:
    - Easily consume the data with datamarts to apply any additional transformations or enable ad-hoc analysis and querying using SQL queries
    - Easily integrate a no-code data warehousing solution with no management of datasets
- For solutions with existing datamarts:
    - Perform reusable extract, transform and load (ETL) at scale for large data volumes

- Bring your own data lake and use dataflows as a pipeline for datamarts

Comparing dataflows to datamarts

Dataflows provide reusable extract, transform and load (ETL). Tables can't be browsed, queried, or explored without a dataset, but can be defined for reuse. The data is exposed in Power BI or CDM format if you bring your own data lake. Dataflows are used by Power BI to ingest data into your datamarts. You should use dataflows whenever you want to reuse your ETL logic.

Use **dataflows** when you need to:

- Build reusable and shareable data prep for items in Power BI.

Datamarts are a fully managed database that enables you to store and explore your data in a relational and fully managed Azure SQL DB. Datamarts provide SQL support, a no-code visual query designer, Row Level Security (RLS), and auto-generation of a dataset for each datamart. You can perform ad-hoc analysis and create reports, all on the web.

Use **datamarts** when you need to:

- Sort, filter, do simple aggregation visually or through expressions defined in SQL
- For outputs that are results, sets, tables, and filtered tables of data
- Provide accessible data through a SQL endpoint
- Enable users who don’t have access to Power BI Desktop

Creating a dataflow

◄ Lesson 8 Quiz

Siirry...

Exercise 32 - Create and use a dataflow ►