Data prep using Dataflows in Power BI

Prashant Kumar Mishra
Engineering Architect

Instructions:

Please keep your microphones on mute during the entire session
Type in your questions in chat window, it will be answered towards the end of our session

Data prep using Dataflows in Power BI

Welcome

Course Navigation

Introduction to Power BI
Section 1

Dataflows in Power BI
Section 2

Section 1

- Introduction to Power BI
- Power BI products
- Power BI pricing
- Transform, shape and model data
- Create your own measures
- Create calculated columns
- Publish dashboard
- Schedule refresh

Section 2

- Introduction to Dataflows
- Connect to data sources with dataflows
- Use dataflow with on-premises data sources
- Use incremental refresh with dataflows
- Connect to dataflows in Power BI desktop
- Connect data lake gen2 for dataflow storage

PlatformAzureData.com

Introduction to Power BI

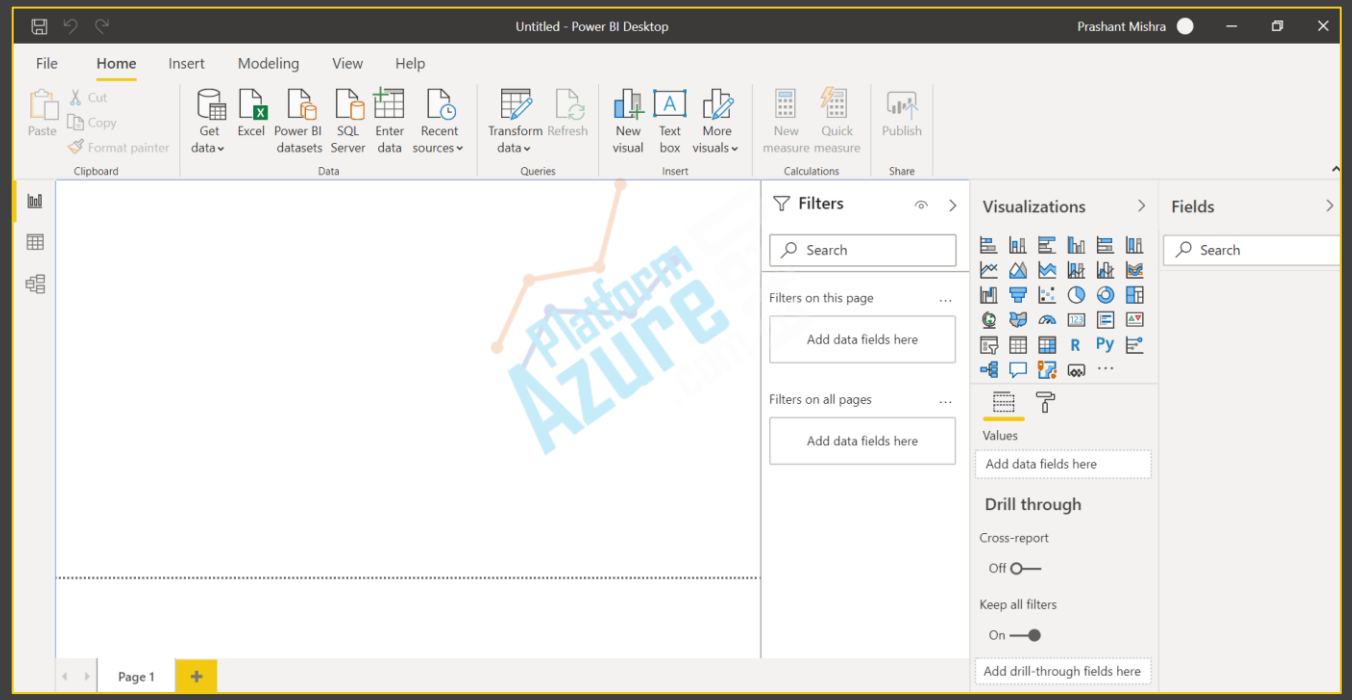
Introduction to Power BI

Introduction to Power BI
Section 1

- Introduction to Power BI
- Power BI products
- Power BI pricing
- Transform, shape and model data
- Create your own measures
- Create calculated columns
- Publish dashboard
- Schedule refresh

Dataflows in Power BI
Section 2

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data may be an Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses. Power BI lets you easily connect to your data sources, visualize and discover what's important, and share that with anyone or everyone you want.



Introduction to Power BI

Power BI products

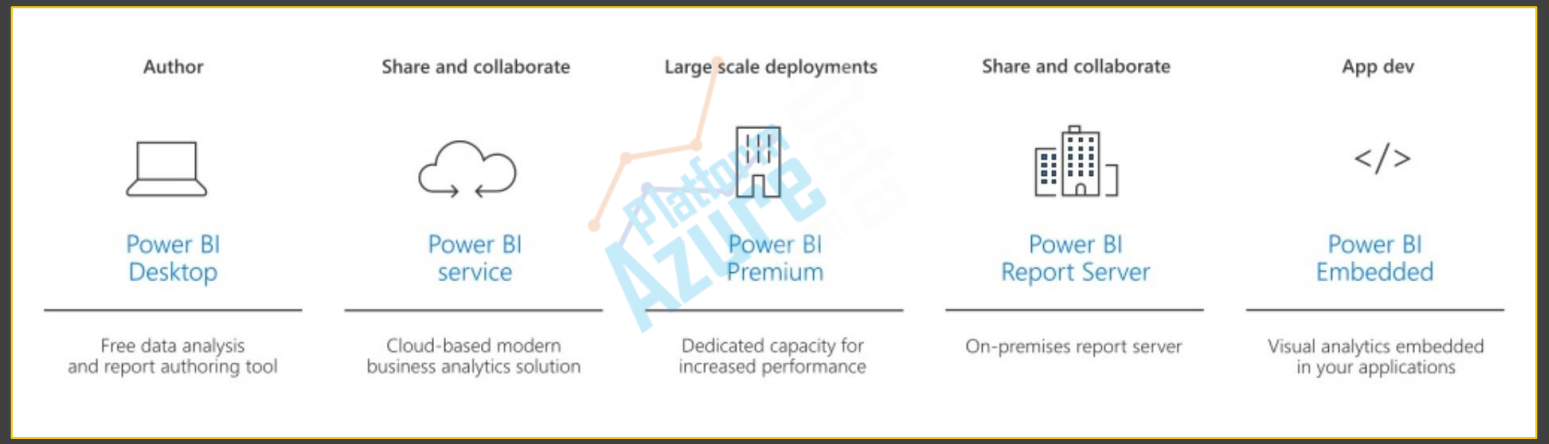
Introduction to Power BI
Section 1

Introduction to Power BI
Power BI products
Power BI pricing
Transform, shape and model data
Create your own measures
Create calculated columns
Publish dashboard
Schedule refresh

Dataflows in Power BI
Section 2

Power BI Desktop, the service, and the mobile apps—are designed to let you create, share, and consume business insights in the way that serves you and your role most effectively.

A fourth element, Power BI Report Server, allows you to publish Power BI reports to an on-premises report server, after creating them in Power BI Desktop.



Introduction to Power BI

Power BI pricing

Introduction to Power BI
Section 1

Introduction to Power BI
Power BI products
Power BI pricing
Transform, shape and model data
Create your own measures
Create calculated columns
Publish dashboard
Schedule refresh

Dataflows in Power BI
Section 2

Use the free Power BI Desktop tool to author reports and prepare, model, and create data visualizations.

Choose Power BI Pro

- Self-service and modern BI in the cloud
- Collaboration, publishing, sharing, and ad-hoc analysis
- Fully managed by Microsoft

Power BI Pro

$9.99

Monthly price per user

Add Power BI Premium

- Enterprise BI, big data analytics, cloud and on-premises reporting
- Advanced administration and deployment controls
- Dedicated cloud compute and storage resources
- Allows any user to consume Power BI content

Power BI Premium

$4,995

Monthly price per dedicated cloud compute and storage resource with annual subscription

PlatformAzureData.com

Introduction to Power BI

Transform, shape and model data

Introduction to Power BI
Section 1

- Introduction to Power BI
- Power BI products
- Power BI pricing
- Transform, shape and model data
- Create your own measures
- Create calculated columns
- Publish dashboard
- Schedule refresh

Dataflows in Power BI
Section 2

- Connect to data
- Shape and combine data
- Group rows
- Pivot columns
- Create custom columns
- Query formulas

You can use quick measures to quickly and easily perform common, powerful calculations. A quick measure runs a set of Data Analysis Expressions (DAX) commands behind the scenes, then presents the results for you to use in your report. You don't have to write the DAX, it's done for you based on input you provide in a dialog box.

PlatformAzure.com Data

Introduction to Power BI

Create your own measures

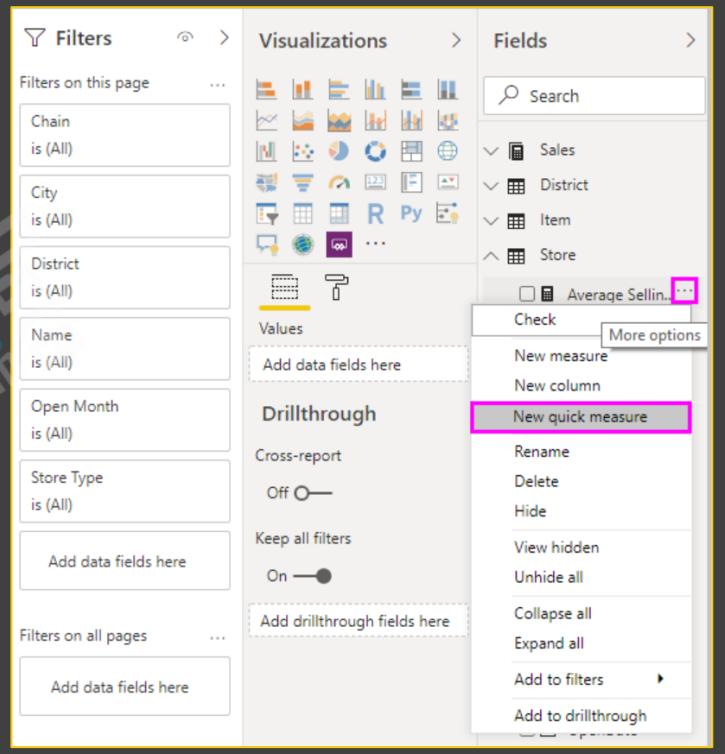
Introduction to Power BI
Section 1

- Introduction to Power BI
- Power BI products
- Power BI pricing
- Transform, shape and model data
- Create your own measures
- Create calculated columns
- Publish dashboard
- Schedule refresh

Dataflows in Power BI
Section 2

In Power BI Desktop, measures are created and displayed in Report View or Data View. Measures you create yourself appear in the Fields list with a calculator icon. You can name measures whatever you want and add them to a new or existing visualization just like any other field.

Measures calculate a result from an expression formula. When you create your own measures, you'll use the Data Analysis Expressions (DAX) formula language. DAX includes a library of over 200 functions, operators, and constructs. Its library provides immense flexibility in creating measures to calculate results for just about any data analysis need.



Introduction to Power BI

Create your own measures

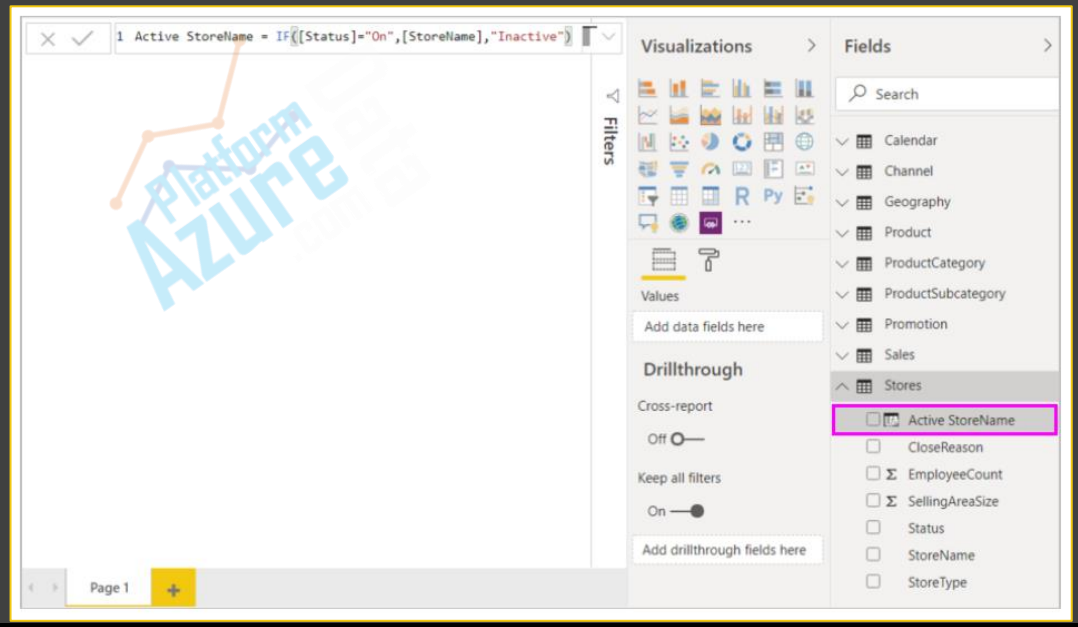
Introduction to Power BI
Section 1

- Introduction to Power BI
- Power BI products
- Power BI pricing
- Transform, shape and model data
- Create your own measures
- Create calculated columns
- Publish dashboard
- Schedule refresh

Dataflows in Power BI
Section 2

Sometimes the data you're analyzing doesn't contain a particular field you need to get the results you're after. This is where calculated columns come in. Calculated columns use Data Analysis Expressions (DAX) formulas to define a column's values, anything from putting together text values from a couple of different columns to calculating a numeric value from other values.

Calculated columns are similar to measures in that both are based on DAX formulas, but they differ in how they are used. You often use measures in a visualization's Values area, to calculate results based on other fields. You use calculated columns as new Fields in the rows, axes, legends, and group areas of visualizations.



Introduction to Power BI

Publish Dashboard

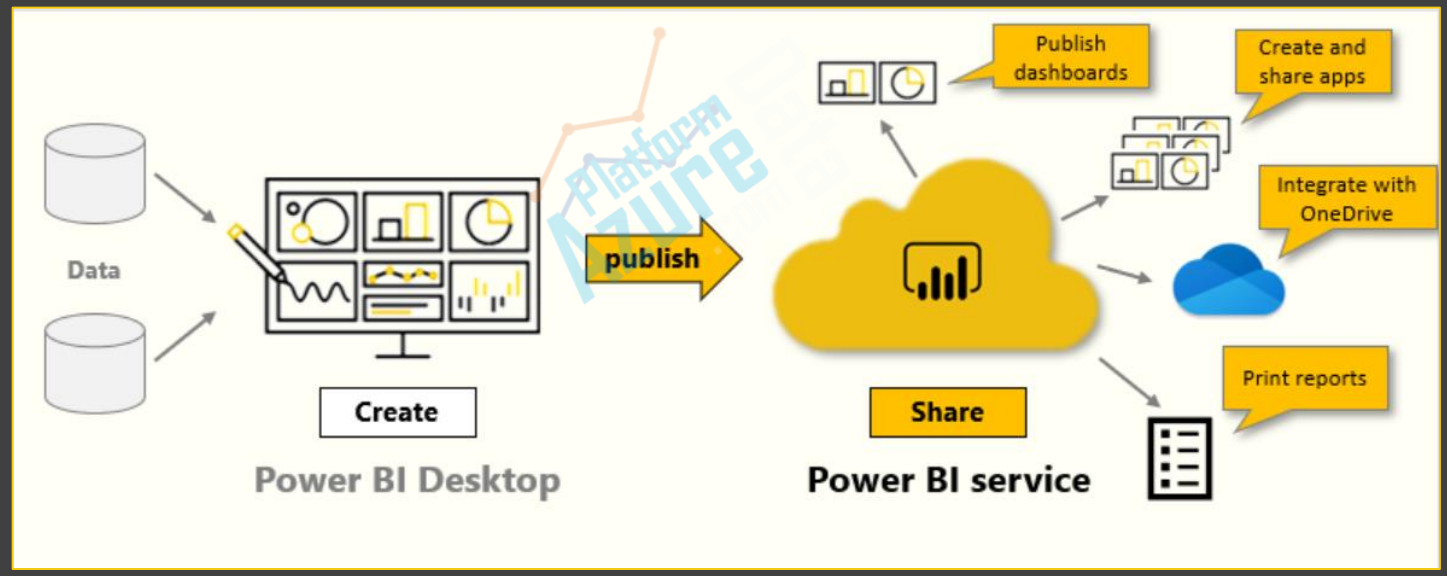
Introduction to Power BI
Section 1

- Introduction to Power BI
- Power BI products
- Power BI pricing
- Transform, shape and model data
- Create your own measures
- Create calculated columns
- Publish dashboard
- Schedule refresh

Dataflows in Power BI
Section 2

Power BI offers various ways of sharing and collaborating with colleagues on your dashboards, reports, and data including:

Publish your reports from Power BI Desktop to the Power BI service.
Create Apps that package a dashboard, report, and dataset to send to your colleagues, who can use the content pack as a starting point and further enhance it.
Create Groups, which you can use as a security model to identify a subset of users who have access to dashboards, reports, and datasets that you create.
Publish to the web so that live reports can be embedded in a webpage.



Introduction to Power BI

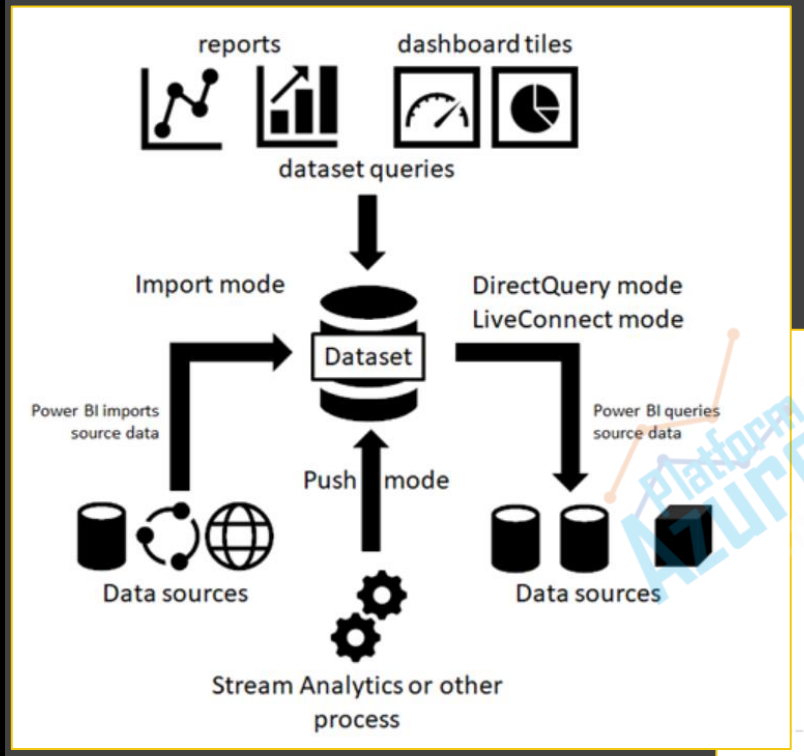
Schedule Refresh

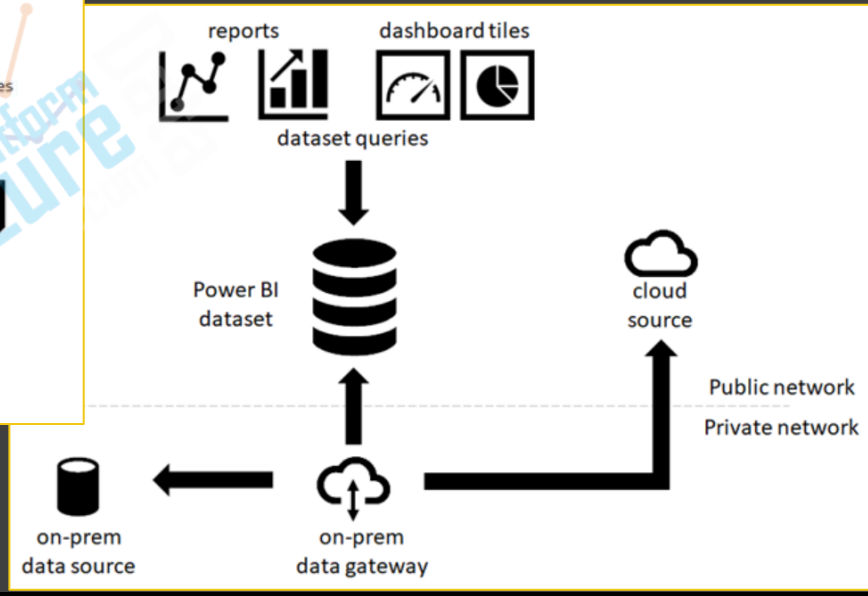
Introduction to Power BI
Section 1

- Introduction to Power BI
- Power BI products
- Power BI pricing
- Transform, shape and model data
- Create your own measures
- Create calculated columns
- Publish dashboard
- Schedule refresh

Dataflows in Power BI
Section 2

Power BI offers various ways of sharing and collaborating with colleagues on your dashboards, reports, and data including:





Dataflows in Power BI

Introduction to Dataflows

Introduction to Power BI
Section 1

Dataflows in Power BI
Section 2

Introduction to Dataflows

Connect to data sources with dataflows

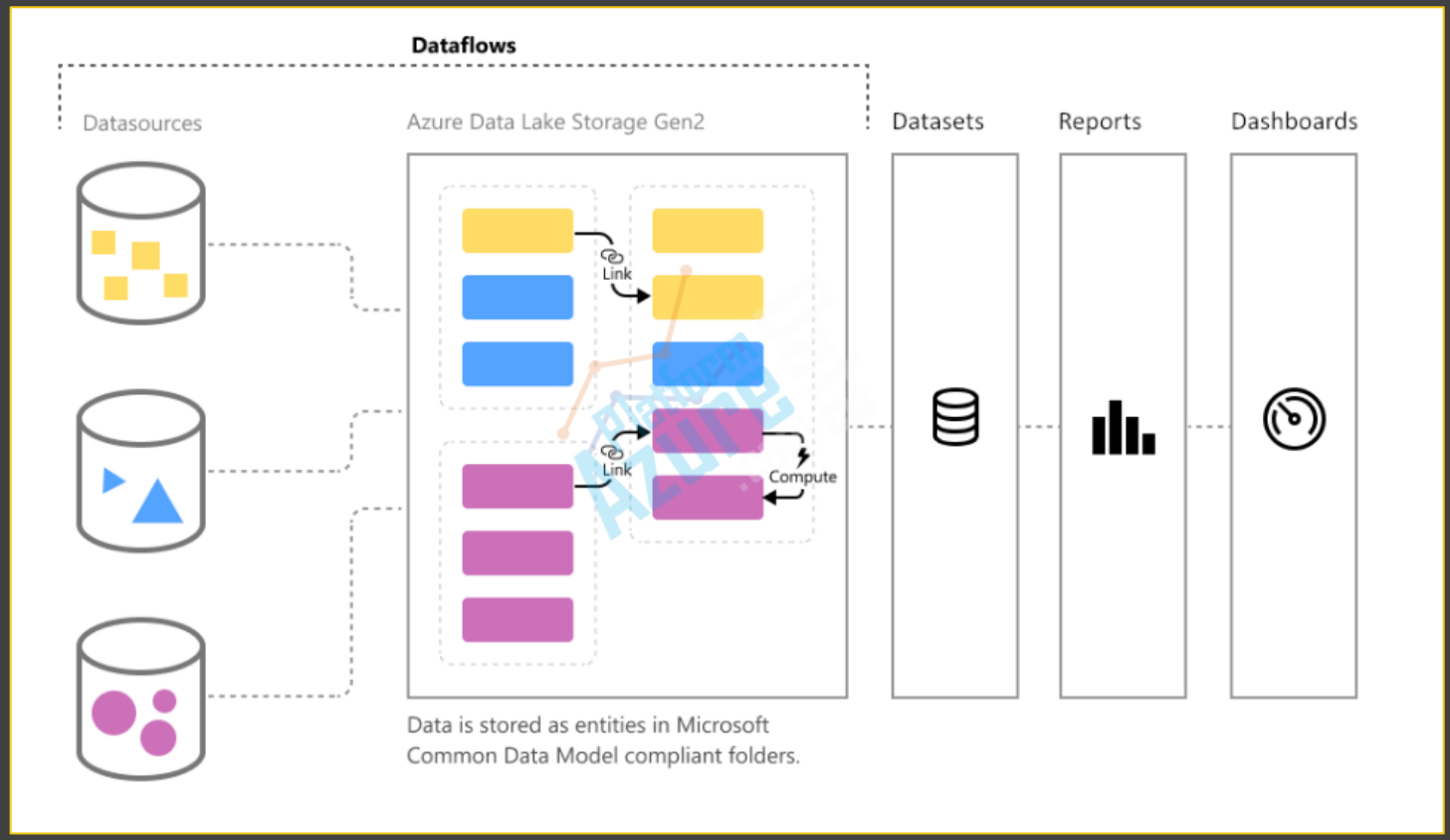
Use dataflow with on-premises data sources

Use incremental refresh with dataflows

Connect to dataflows in Power BI desktop

Connect data lake gen2 for dataflow storage

Power BI introduces dataflows to help organizations unify data from disparate sources and prepare it for modeling. Analysts can easily create dataflows, using familiar, self-service tools. Dataflows are used to ingest, transform, integrate, and enrich big data by defining data source connections, ETL logic, refresh schedules, and more.



PlatformAzure.com Data

Dataflows in Power BI

Connect to data sources with dataflows

Introduction to Power BI
Section 1

Dataflows in Power BI
Section 2

- Introduction to Dataflows
- Connect to data sources with dataflows
- Use dataflow with on-premises data sources
- Use incremental refresh with dataflows
- Connect to dataflows in Power BI desktop
- Connect data lake gen2 for dataflow storage

With Power BI dataflows, you can connect to many different data sources to create new dataflows or add new entities to an existing dataflow.

Dataflows in Power BI

Use dataflow with on-premises data sources

Introduction to Power BI
Section 1

Dataflows in Power BI
Section 2

- Introduction to Dataflows
- Connect to data sources with dataflows
- Use dataflow with on-premises data sources
- Use incremental refresh with dataflows
- Connect to dataflows in Power BI desktop
- Connect data lake gen2 for dataflow storage

To create a dataflow using a gateway, the user must be the Enterprise Gateway's administrator, or the administrator must have shared the data source they are planning to use with the user.

When creating a dataflow, select an on-premises data source from the data sources list

Once you make your selection, you're prompted to provide the connection details for the Enterprise Gateway that will be used to access the on-premises data. You must select the gateway itself and provide credentials for the selected gateway.

Limitations:

- Each dataflow may use only one gateway. As such, all queries should be configured using the same gateway.
- Changing the gateway impact the entire dataflow.
- If several gateways are needed, the best practice is to build several dataflows (one for each gateway) and use the compute or entity reference capabilities to unify the data.
- Dataflows are only supported using enterprise gateways. Personal gateways will not be available for selection in the dropdown lists and settings screens.

Platform Azure Data .com

Dataflows in Power BI

Use incremental refresh with dataflows

Introduction to Power BI
Section 1

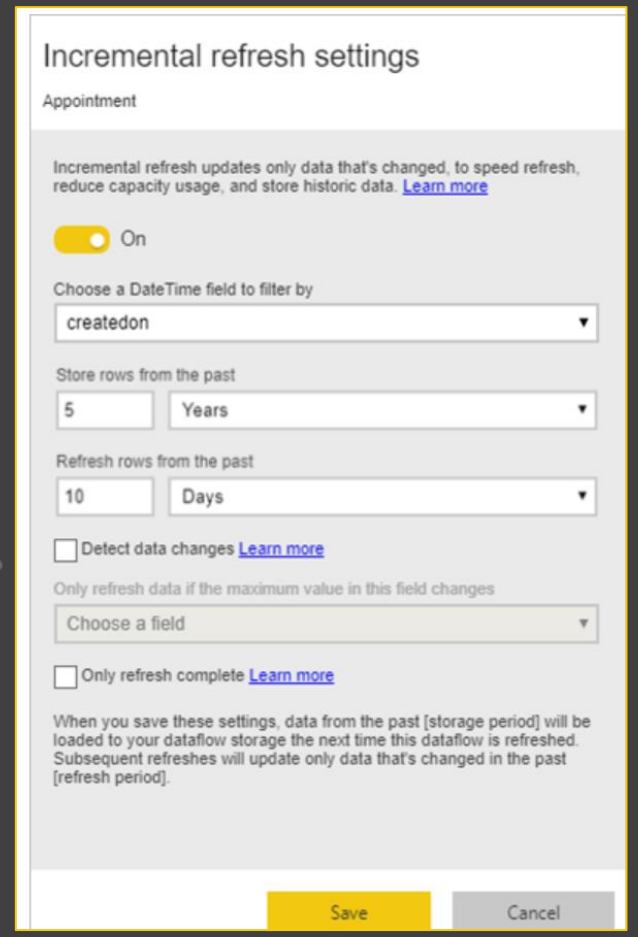
Dataflows in Power BI
Section 2

- Introduction to Dataflows
- Connect to data sources with dataflows
- Use dataflow with on-premises data sources
- Use incremental refresh with dataflows
- Connect to dataflows in Power BI desktop
- Connect data lake gen2 for dataflow storage

Benefits

- Refresh occurs faster
- Refresh is more reliable
- Resource consumption is reduced

Time zone support in incremental refresh
Dataflow incremental refresh is dependent on the time in which it is run. The filtering of the query is dependent on the day it runs.



Dataflows in Power BI

Connect to dataflows in Power BI desktop

Introduction to Power BI
Section 1

Dataflows in Power BI
Section 2

- Introduction to Dataflows
- Connect to data sources with dataflows
- Use dataflow with on-premises data sources
- Use incremental refresh with dataflows
- Connect to dataflows in Power BI desktop
- Connect data lake gen2 for dataflow storage

The Power Platform dataflows connector lets you connect to entities created by dataflows in the Power BI service.

To use the Power Platform dataflows connector, you must be running a recent version of Power BI Desktop.

Power BI Desktop runs locally on the computer on which it is installed. Ingestion performance of dataflows is determined by a variety of factors. Those factors include the size of the data, your computer's CPU and RAM, network bandwidth, distance form the data center, and other factors.

If the ingested data size is too large for Power BI Desktop to manage on your computer, you can use linked and computed entities in dataflows to aggregate the data (within dataflows) and ingest only the pre-prepared, aggregated data.

**Most dataflows reside in the Power BI service tenant. However, Power BI Desktop users cannot access dataflows that are stored in Azure Data Lake Storage Gen2 account, unless they are the owner of the dataflow, or they have been explicitly authorized to the dataflow's CDM folder.

AzurePlatformData.com

Introduction to Power BI
Section 1

Dataflows in Power BI
Section 2

Introduction to Dataflows

Connect to data sources with dataflows

Use dataflow with on-premises data sources

Use incremental refresh with dataflows

Connect to dataflows in Power BI desktop

Connect data lake gen2 for dataflow storage

Dataflows in Power BI

Connect data lake gen2 for dataflow storage

To use Azure Data Lake Storage Gen2 for dataflows, you need the following:

- Power BI tenant - at least one account in your Azure Active Directory (AAD) tenant must have signed up for Power BI
- A Global Administrator account - this account is required to connect and configure Power BI to store the dataflow definition, and data, in your Azure Data Lake Storage Gen2 account
- An Azure subscription - you need an Azure subscription to use Azure Data Lake Storage Gen2
- Resource group - use a resource group you already have, or you can create a new one
- An Azure Storage account with Data Lake Storage Gen2 feature enabled

PlatformAzureData.com