



# Business Analytics with Power BI

Microsoft Services



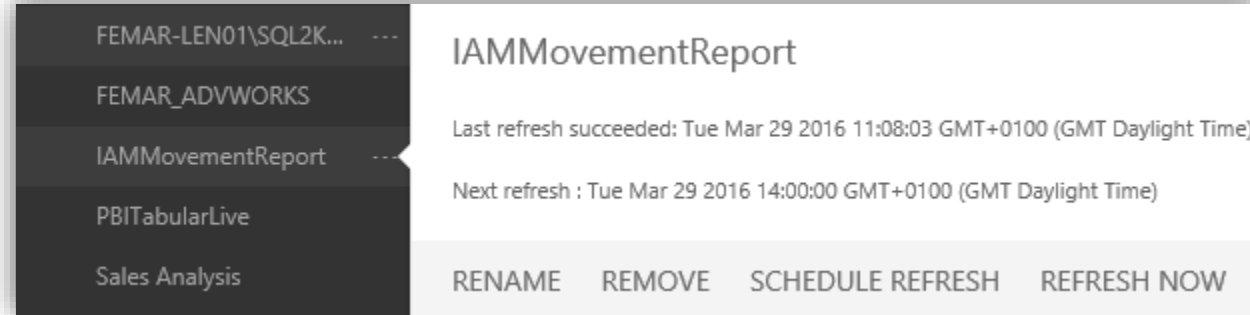
## Module 2: Power BI Service

### Lesson 3: Keeping Data Current

# Keeping Data Current

## Data Refresh

- When a **dataset** is created in Power BI it might have several **data sources**



- When you refresh data in Power BI, you're updating data in the dataset with new data from data sources
- Some data sources might be **live** or **self-updateable**
- Refresh Now** performs the operation immediately based on the definition of the data sources

# Keeping Data Current

## Data Refresh

- The **Scheduled Refresh** option allows configuration of **credentials** and schedule

▲ Data source credentials

ppr_applicationDB- <del>qmf2cd</del> .database.windows.net,1433	<a href="#">Edit credentials</a>
AzureBlob	<a href="#">Edit credentials</a>
OneDrive - Personal	<a href="#">Edit credentials</a>

Configure IAMMovementReport<sup>x</sup>

Server

Database

Authentication method

User name

Password

[Sign in](#) [Cancel](#)

# Keeping Data Current

## Data Refresh

- The **Scheduled Refresh** option allows configuration of credentials and schedule

Schedule Refresh

Keep your data up to date  
Yes ☒

Refresh frequency  
Daily

Time zone  
(UTC) Coordinated Universal Time

Time  
1 00 PM X

9 00 PM X

[Add another time](#)

☒ Send refresh failure notification email to me

Apply Discard

Schedule Refresh

Keep your data up to date  
Yes ☒

Refresh frequency  
Weekly

Time zone  
(UTC) Coordinated Universal Time

☒ Sunday ☒ Monday ☒ Tuesday ☒ Wednesday  
☒ Thursday ☒ Friday ☒ Saturday

Time  
1 00 PM X

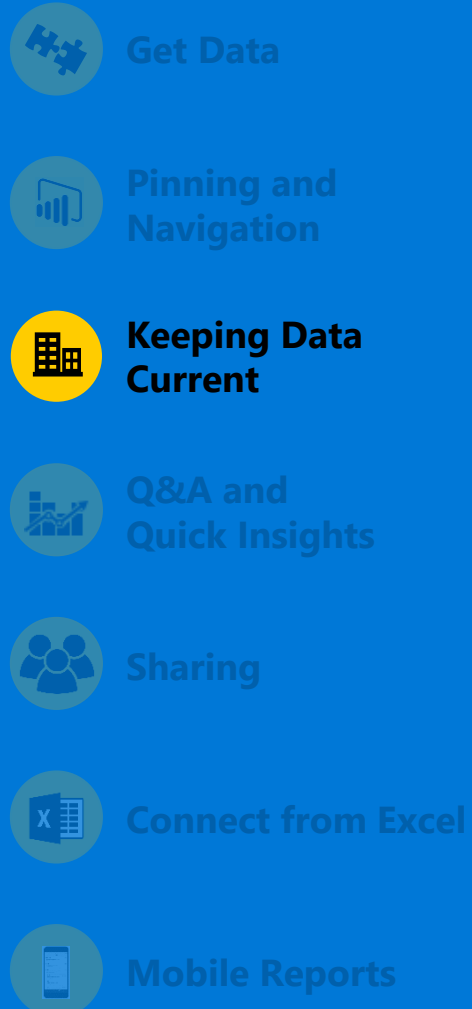
9 00 PM X

[Add another time](#)

☒ Send refresh failure notification email to me

Apply Discard

# Keeping Data Current



## Data Refresh – Data Source Types

- **Automatic Refresh** – no user configuration is required
  - Content packs from SaaS providers – Typically, once a day
  - Files loaded from Microsoft OneDrive/ Microsoft SharePoint Online – refreshes every hour (can be overridden)
- **Live with Direct Query** – Data is always in sync with the data source. No schedule can be defined
  - Microsoft Azure SQL Database
  - Microsoft Azure SQL Data Warehouse
  - Spark on Microsoft Azure HDInsight
  - SSAS (Multidimensional and Tabular) – requires a Gateway and allows row-level-security
- **Real-time data coming from Power BI API or Microsoft Stream Analytics or PubNub**
- **Refreshable Data Sources** – a schedule can be defined in Power BI service
  - SQL Azure Database
  - SQL Server On-Premises – requires a Gateway
  - File stored in OneDrive
  - ...

# Keeping Data Current



Get Data



Pinning and Navigation



Keeping Data Current



Q&A and Quick Insights



Sharing



Connect from Excel



Mobile Reports

## Data Refresh – Using Power BI REST APIs

- Programmatically refresh datasets in Power BI Service using APIs.
- Retrieve refresh history for any dataset that you own.
- Easily **automate** and **scale out** Power BI data refresh management.
- **Integrate data refresh** into your existing ETL or deployment processes:  
For e.g. trigger Power BI data refresh as the last step in your Azure Data Factory ETL pipeline

# Keeping Data Current



Get Data



Pinning and Navigation



Keeping Data Current



Q&A and Quick Insights



Sharing



Connect from Excel

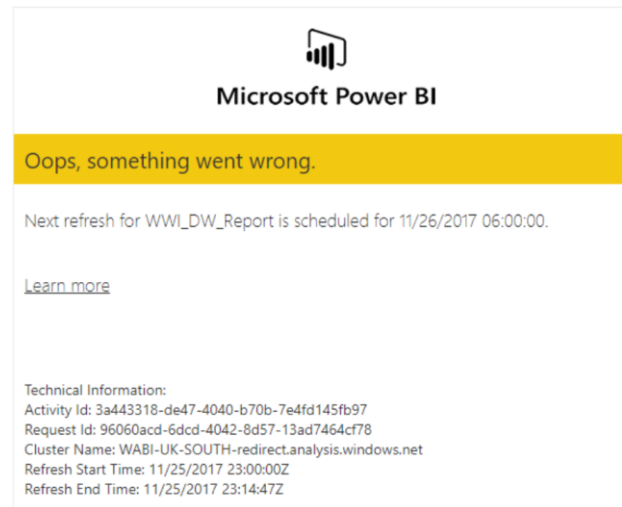


Mobile Reports

## Managing Power BI Dataset Refresh Failures

- Not for DQ, but when at least one report that relies on an on-premises database or file source and you need to schedule it to refresh regularly to get up to date data

## Built-in Failure Notifications



Send refresh failure notification email to me

Settings for WWI\_DW\_Report

This dataset has been configured by [cporteous@sqlglasgow.co.uk](mailto:cporteous@sqlglasgow.co.uk).

Next refresh: Sat Nov 25 2017 23:44:24 GMT+0000 (GMT Standard Time)

[Refresh history](#)

**BUT**, what if you have the same dataset across several workspaces, maybe in a DEV, QA, PROD scenario?

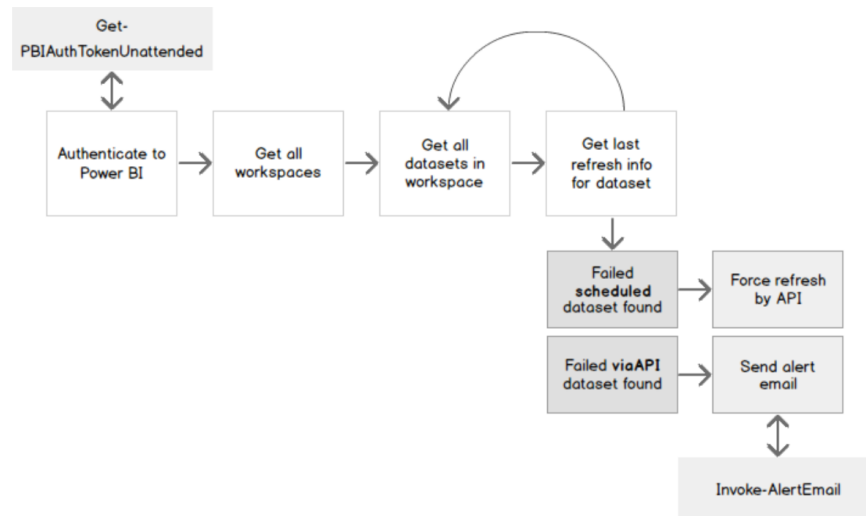


# Keeping Data Current



## Managing Power BI Dataset Refresh Failures PowerShell Alerting

### Custom Script on GitHub



### Sample Schedule For the Script:

- Run every 15 minutes.
- This means, at the longest interval, it would take up to **14 minutes** to detect a **failed dataset**, force a **manual refresh**, then another 15 minutes before it would **email out** to warn of the failure.

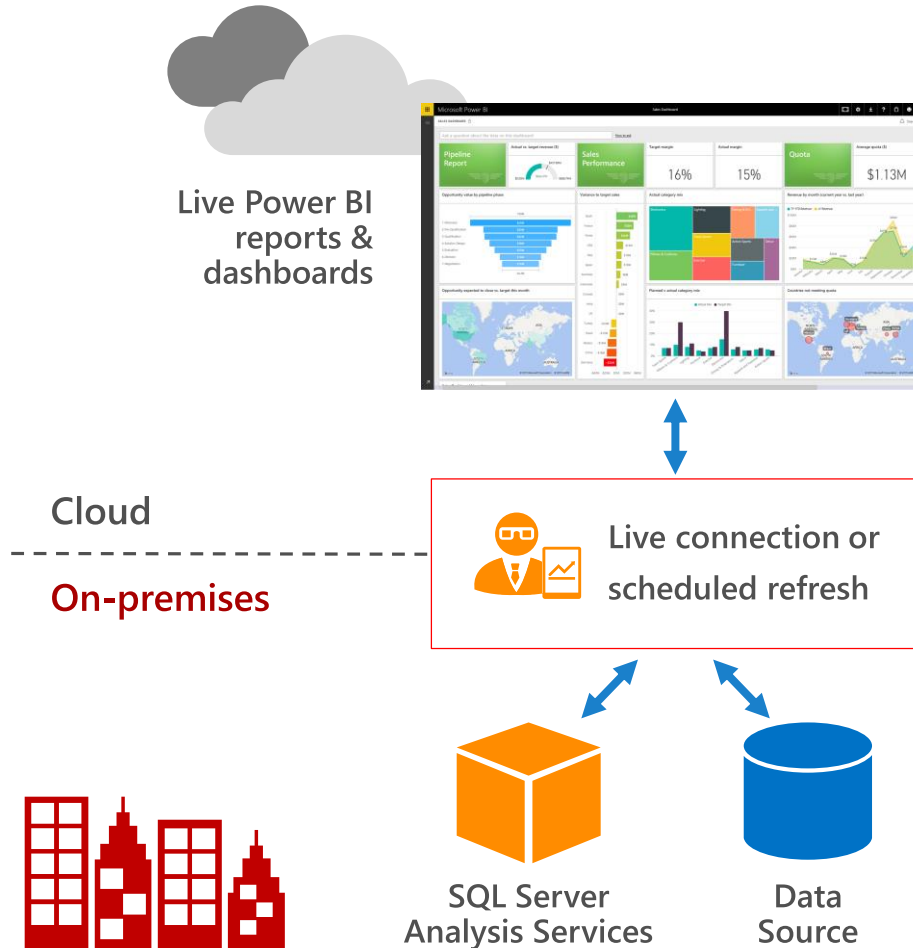
### What the Email Contains:

- A relevant **title**, (in this case, the name/purpose of the script)
- The name of the **parent workspace** & a **hyperlink** to that workspace
- The **name** & a **link** to the **failing dataset's settings page** so you can diagnose the failure easily

# Keeping Data Current

- Get Data
- Pinning and Navigation
- Keeping Data Current**
- Q&A and Quick Insights
- Sharing
- Connect from Excel
- Mobile Reports

## Gateways



Access on-premises data any way you like

- ➔ Scheduled refresh
- ➔ Live connection

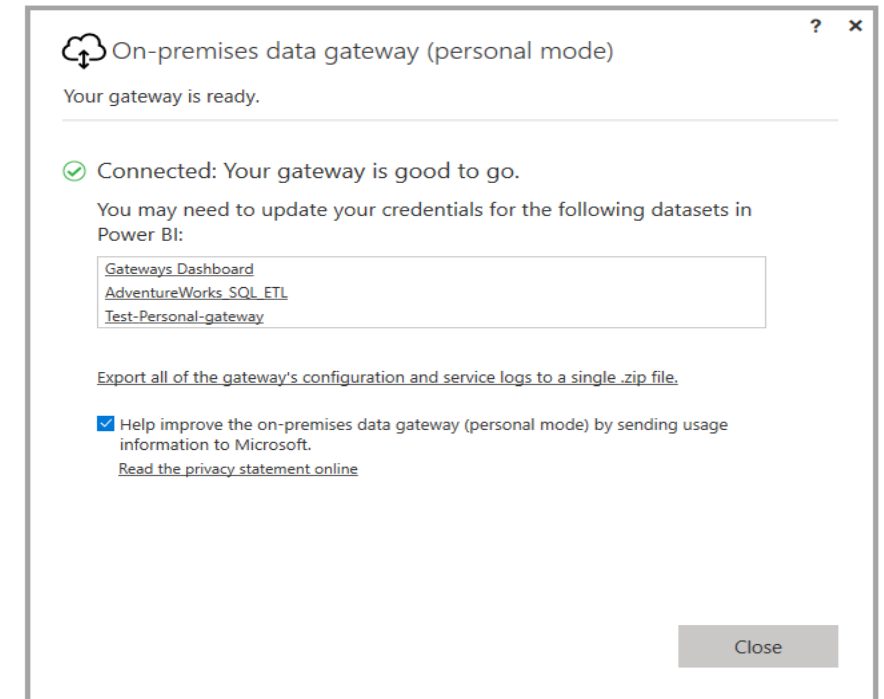
Connect to on-premises SQL Server and Analysis Services models

- ➔ Tabular and multidimensional

# Keeping Data Current

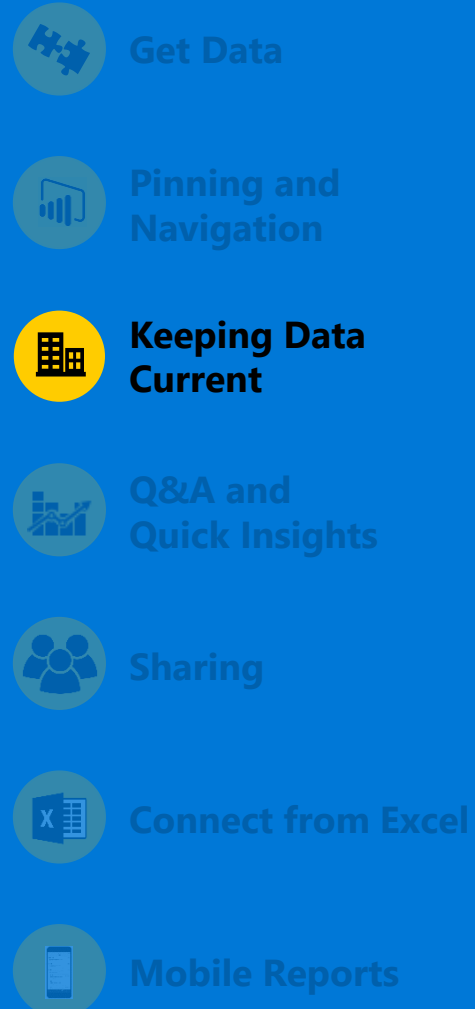
## Gateways – On-premises data gateway (personal mode)

- Allows datasets on the service to access data on-premises on a scheduled refresh
- Runs as a **service or application** on a **computer**, typically from a **business user**, and can serve **multiple sources**
- No need for IT intervention but computer has to be online
- When a gateway is installed, a **status** is shown in Scheduled Refresh



# Keeping Data Current

## Gateways – On-premises data gateway

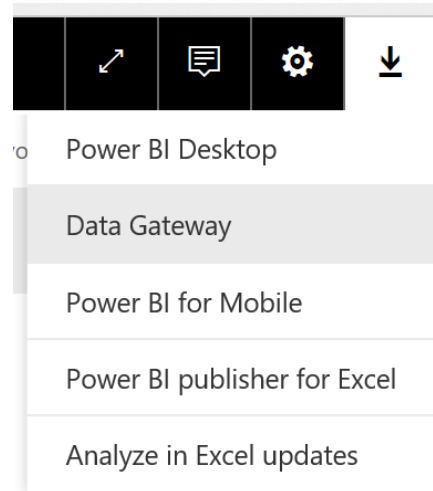


- Allows datasets on the service to **access data on-premises** on a **scheduled refresh** or **live**
- Runs as a **service** on a **server** and can serve multiple users and sources
- Typically set up by IT allows **access segregation** and **monitoring**
- **Supported data source types** (via Excel Data Model or PBI Desktop)
  - SQL Server (Live also)
  - Analysis Services (Live also and with Row-Level-Security)
  - SAP HANA (Live also)
  - File
  - Folder
  - Oracle (Live also) / DB2 / MySQL / Sybase
  - Teradata (Live also)
  - SharePoint list (on-premises)
  - Web/Odata
  - ...

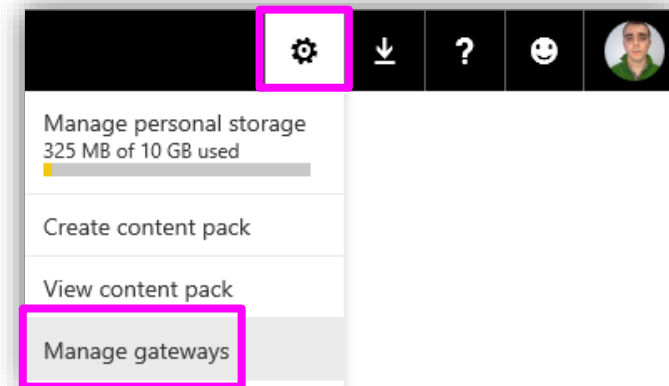
# Keeping Data Current

## Gateways

- Can be **downloaded** from powerbi.com and **installed** on-prem



- The data gateway is managed from the service



# Keeping Data Current

## Gateways – Manage Data Gateway

- **Data Sources** are added here
- **Security** is also defined centrally
- Some data sources are supported **Single Sign-on**
- **Users with access** can use this connection in their Power BI Desktop **reports**
- Users can create **schedules** to update their data
- For **SSAS, UPN Mapping** can be defined in the service, per data source.

The screenshot shows the 'ADD DATA SOURCE' configuration page. On the left, under 'GATEWAY CLUSTERS', the 'US Sales Gateway' is expanded, showing a 'US Sales Data' data source with a 'Test all connections' button. The main panel is titled 'Data Source Settings' and shows a 'Connection Successful' status. The configuration details include: Data Source Name: 'US Sales Data'; Data Source Type: 'SQL Server'; Server: '[REDACTED]\sql2016'; Database: 'TailspinToys-US'; Authentication Method: 'Windows'. Below these fields, a note states: 'The credentials are encrypted using the key stored on-premises on the gateway server. [Learn more](#)'. There are input fields for 'Username' and 'Password', both masked with dots. Under 'Advanced settings', the checkbox 'Use SSO via Kerberos for DirectQuery queries' is checked, with a note: 'This will only be applied for DirectQuery queries. Import will use the Username and Password specified in the data source details [Learn more](#)'. At the bottom, the 'Privacy Level setting for this data source' is set to 'Organizational'.