

Using the Dataflow Scanner API Solution

Introduction

The following is an updated solution to collect and store Power BI tenant metadata. To see version 1 (using Power Automate), see here: [LINK](#). This solution utilizes Dataflow Best Practices, and utilizes the "bronze / silver / gold" approach to dataflows to A) Send a single request for a new scan and B) use the silver/gold dataflow to collect the metadata from that scan.

Without the use of two dataflows, Power Query would attempt per query to post a new scan.

Note About V2 & V3

There is currently a v3 solution that will not only incorporate the Scanner API, but utilize the same approach for collecting Activity Data from the Power BI Audit Log using incremental refresh to avoid the 30 day retrieve limit. V3 will also incorporate Refresh History of Datasets in a single solution.

V2 is focused on improving the Power Automate method which has limitations and errors.

Pre-Requisites

Before starting, you must have already completed the following in order to use the Scanner API at all. Please see the documentation for each to set up:

- [Enable Service Principal Authentication for Read-Only Admin API's](#)
- [Create an Azure AD app](#)
 - Ensure there are no Power BI admin-consent required permissions on the application
 - Create a client secret / key and copy the value for later use
- Create a new Security Group in AD, add the app to the group
- [Enable Allow Service Principals to use read-only Power BI admin API's in the Power BI Tenant](#)

What's In the Box?

The solution consists of 2 dataflows & a Power BI Template file.

Dataflows

bronzeGetScannerResultID

This is the first dataflow needed to be imported into your tenant. This dataflow should be imported into a development or staging workspace.

This is essential as in order to retrieve the metadata, you must post and retrieve a scan ID. The only purpose of this dataflow is to post, and then output the Scan ID to be used in the silverGetScannerResults dataflow.

Important

Ensure you do not load this dataflow multiple times a day, or even go back to editing it. Everytime this dataflow is triggered it will post and return a new scan ID!

Workspace List

The list of workspaces in your tenant (exlcuding MyWorkspaces). This is needed as the collect metadata requires the workspace id.

scanResults

This is the latest returned Scan ID. This is essential when calling for data from a Scan.

Whenever the bronze dataflow is run, it will send a request to scan the Power BI tenant and metadata. This scanID is what will return the metadata in the dependent dataflow.

getScanIDWS

Information on the status of the dataflow