AZURE PURVIEW – FIRST LOOK

ADVANCING ANALYTICS

Microsoft Partner
Microsoft
Gold Cloud Platform
Gold Data Platform
Gold Data Analytics
Silver Application Integration

databricks

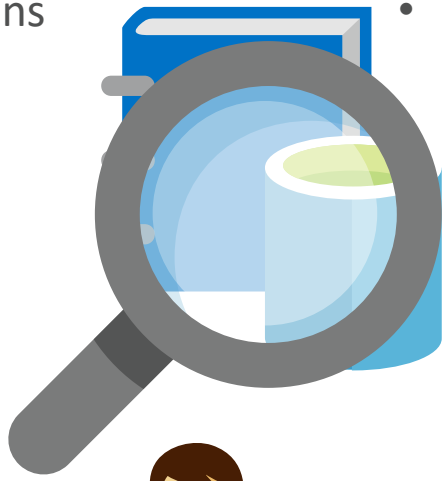
MVP Microsoft® Most Valuable Professional

DATA GOVERNANCE

WHAT IS DATA GOVERNANCE?

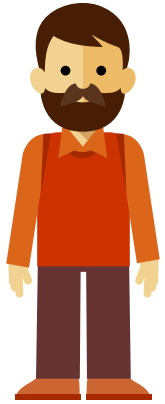
Information Security
- GDPR / PII
- Access Management
- Data Classifications
- Data Security

Data Owners
- Data Duplication
- Data Development Standards
- Data Dictionary
- Business Glossary



Data Consumers
- Data Discovery
- Data Connectivity
- Documentation

Data Stewards
- Data Quality Metrics
- Data Lineage
- Data Profiling



Data Engineering
- Data Processing Rules
- Dependency Chains
- Engineering Metadata

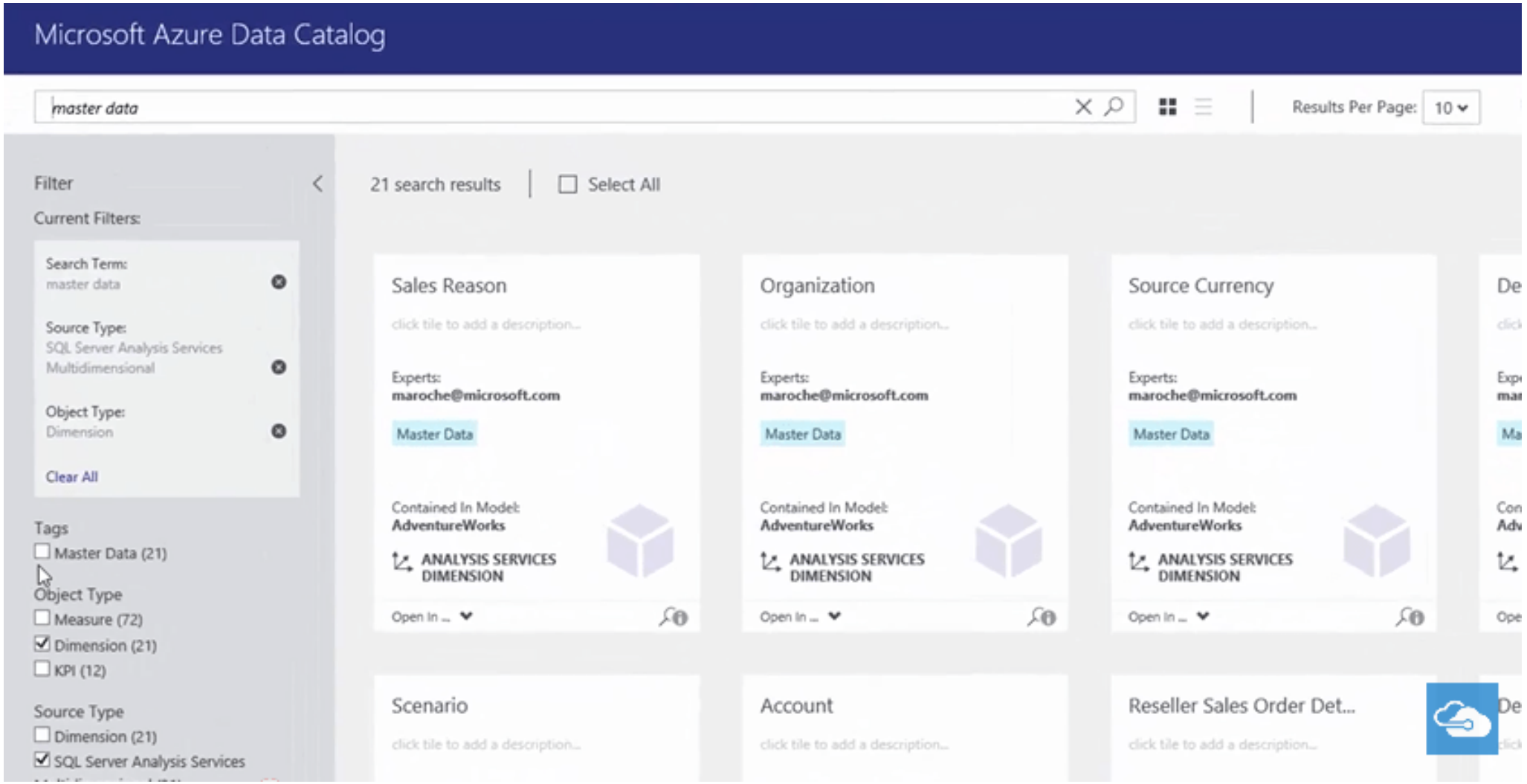
- *Master Data Management?*

DATA GOVERNANCE QUESTIONS

- *Who has access to my data?*
- *What would be impacted if I changed my data?*
- *How did they calculate this figure?*
- *Where do I go for data about X*
- *How much should I trust this data?*
- *Who is the specialist expert for this data set?*
- *Which process should this data follow?*
- *How secure does this data need to be?*

IN THE BEFORE-TIMES...

AZURE DATA CATALOG



SO LONG ADC…

Azure Data Catalog was good at…

- Data Documentation
- Data Dictionary
- Business Glossary
- Data Discovery
- Data Profiling

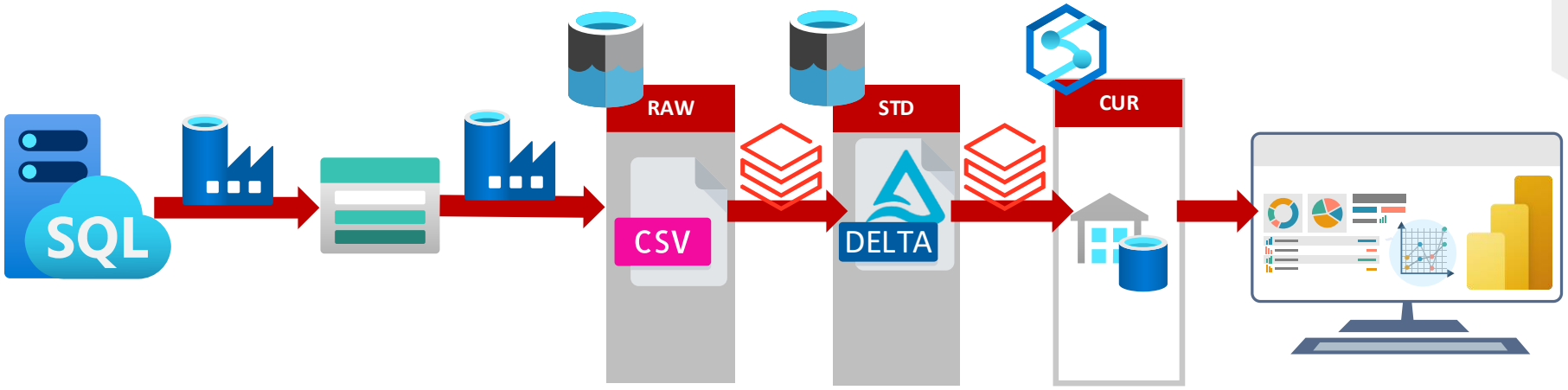
…for Relational Sources

But it lacked…

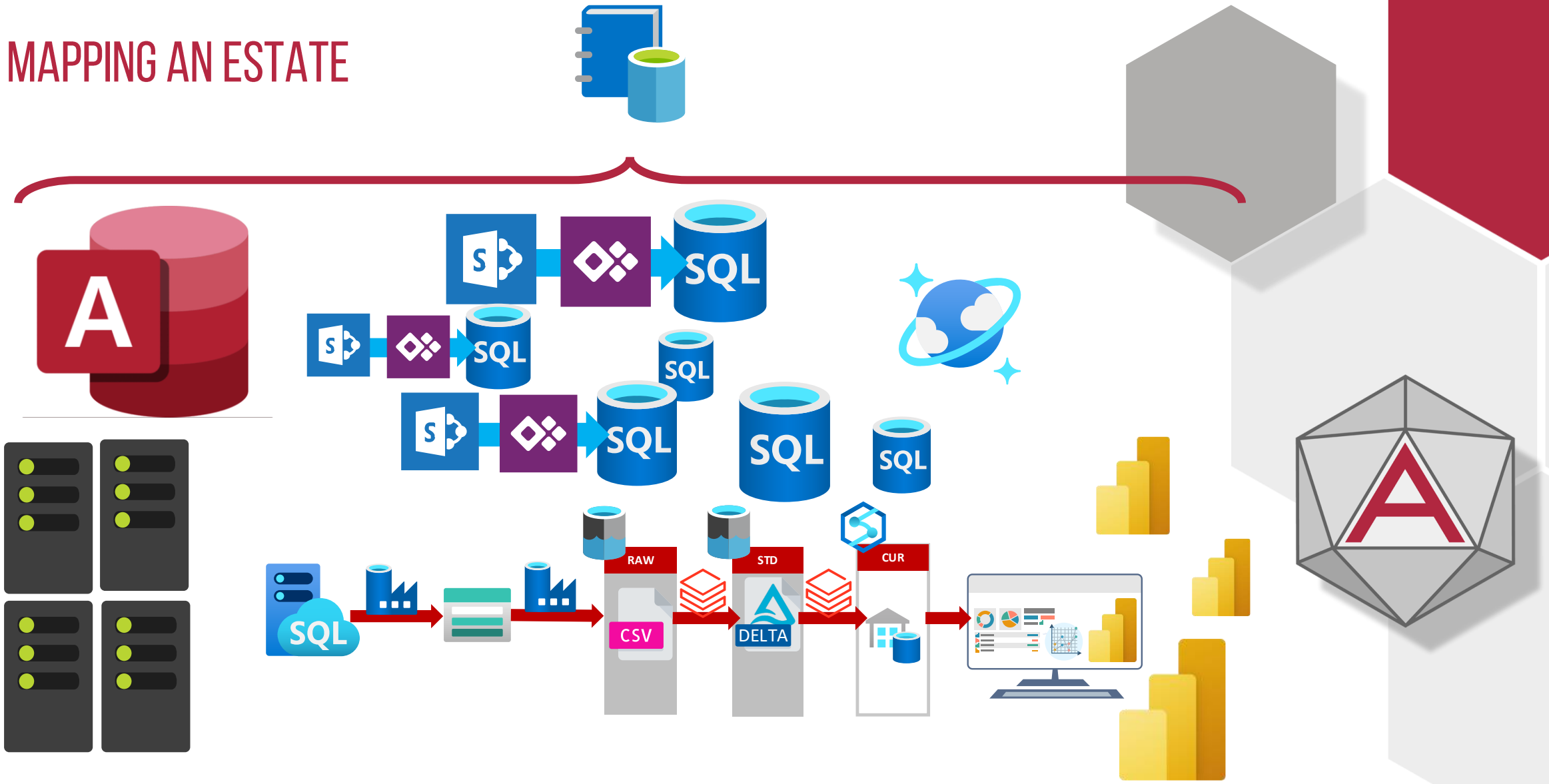
- Data Access Control
- More elaborate data concepts
- Lineage & Data Flow management
- **Programmability**

MAPPING AN ESTATE

This is the NICE use case



MAPPING AN ESTATE



AZURE PURVIEW

AZURE PURVIEW

(Azure Data Catalog Gen 2 + some bits)

5 sources 205 assets 1 glossary term

Knowledge center

Discover learning materials and tutorials.

Register sources

Register sources to your Purview.

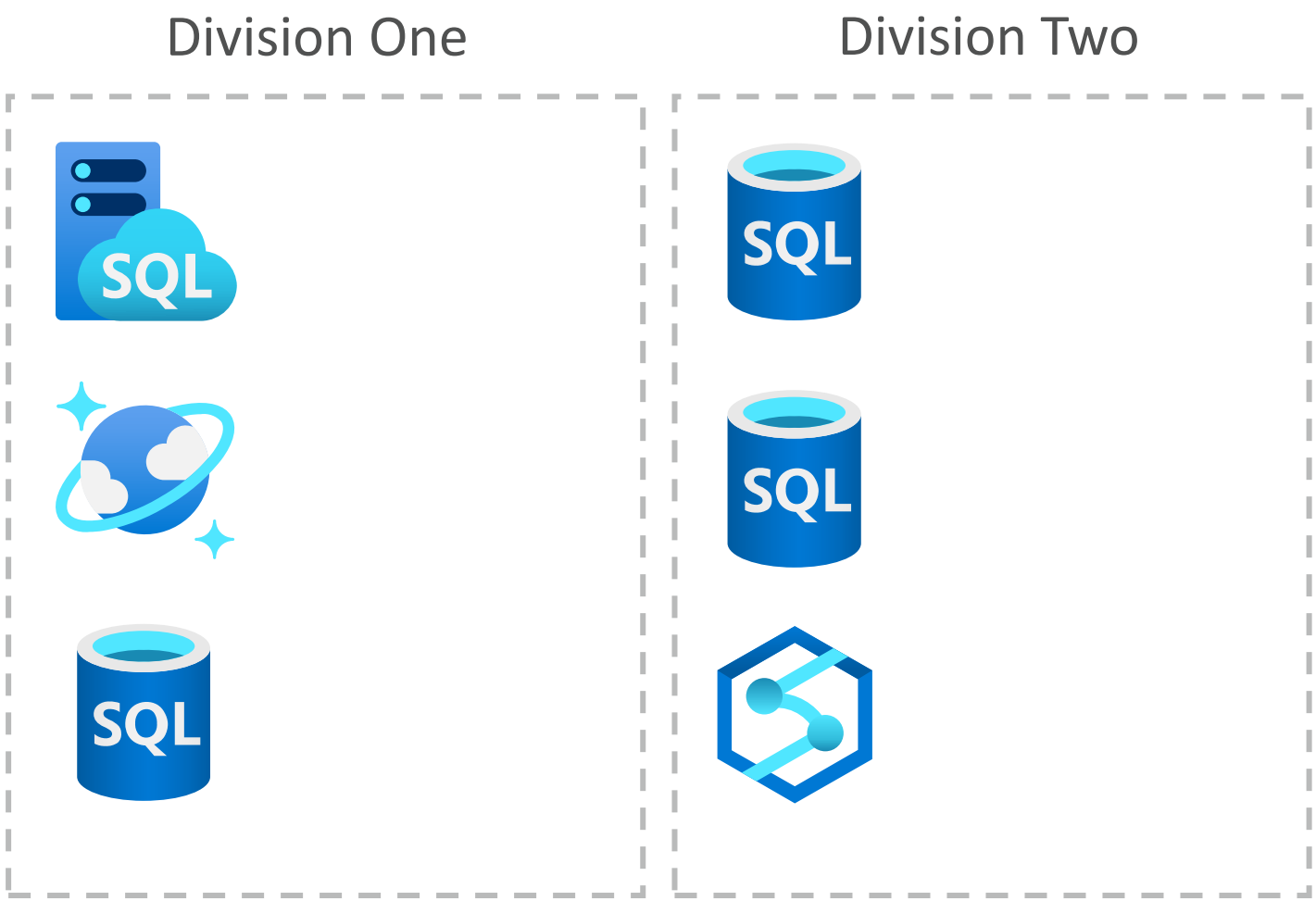
Browse assets

Explore assets by source type.

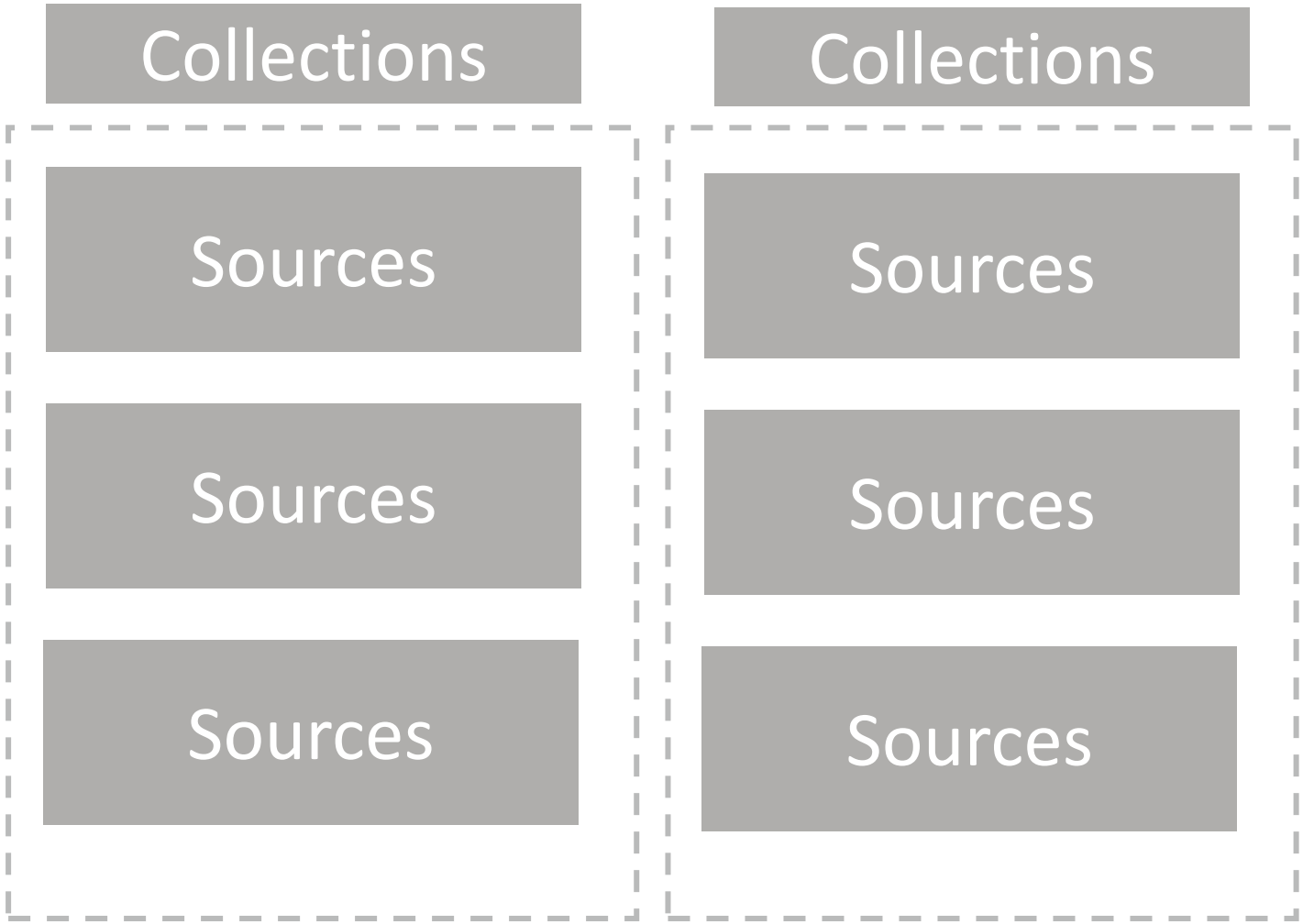
Manage glossary

Create and edit terms with custom term templates.

PURVIEW CONCEPTS



PURVIEW CONCEPTS



ASSET HIERARCHIES

Collection

Source

Scans

- Assets
- Assets
- Assets
- Assets
- Assets

- Description
- Classifications
- Properties
- Glossary Terms
- Lineage
- Contacts
- Related

ASSET HIERARCHIES

Collection

Source

Assets

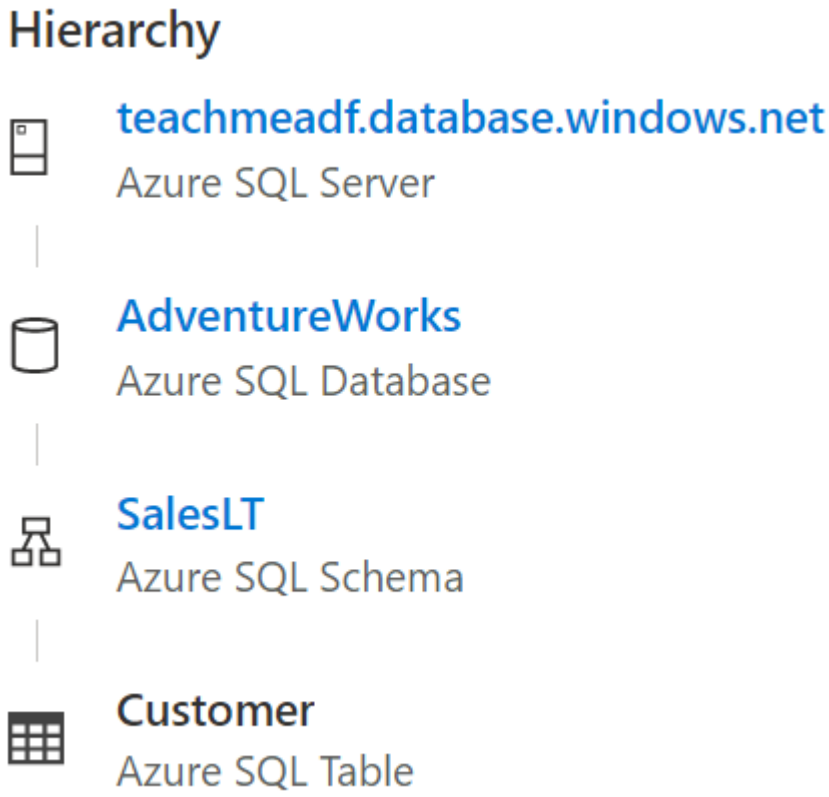
Assets

Assets

Assets

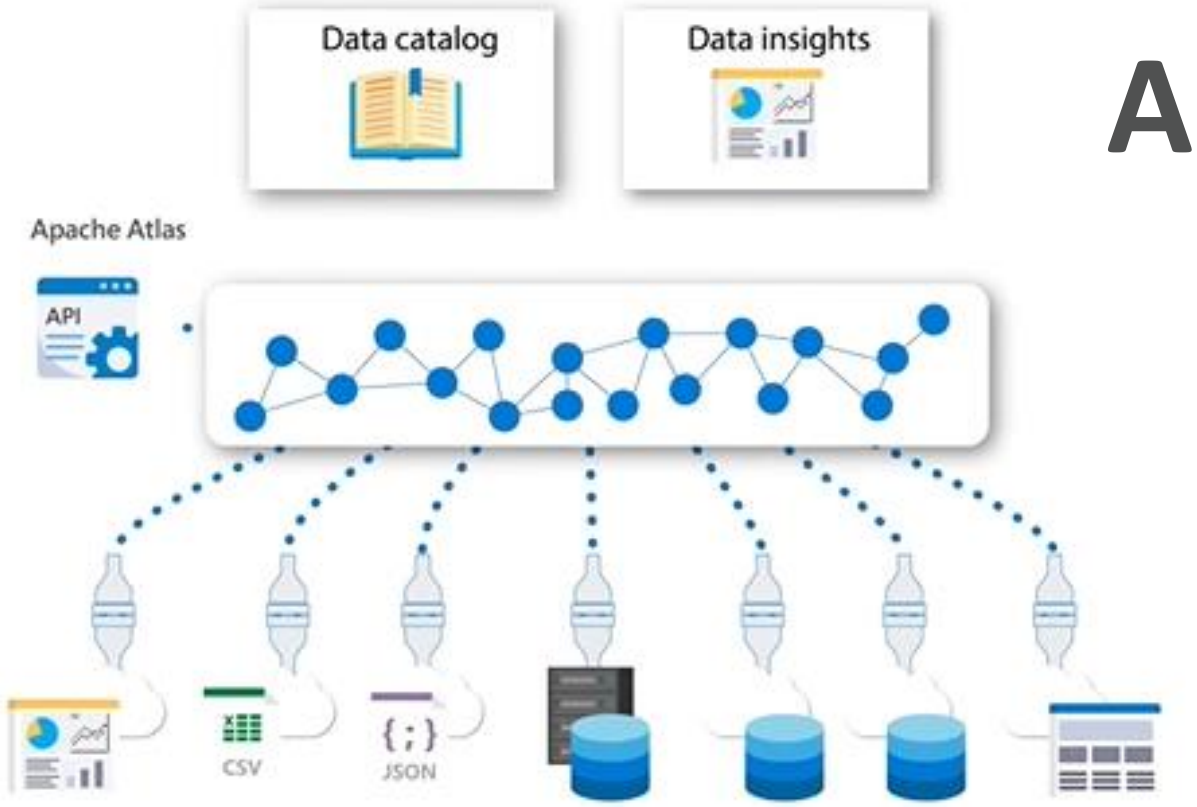
Assets

Assets



WHY THIS STRUCTURE?

Apache Atlas



API

GETTING STARTED

REMEMBER TO REGISTER PROVIDERS

Your subscription needs to have the following "resource providers" enabled:

- Purview
- EventHub
- Storage

SETUP CREDENTIALS

Azure Purview needs to know how to connect to your datasources:

- Basic Username/Password
- Service Principal
- SQL
- Blob Account

All MUST be Key-Vault Backed

GETTING STARTED WITH PURVIEW

- Creating a Purview Catalog
- Registering Sources
- Credentials
- Browsing Assets

ASSET TAGGING

WHAT IS THIS THING

Persons Name

- Restrict Access
- Wipe after business use complete
- Include in PII Checks

Customer Billing Name

- Used to uniquely identify purchasers
- Should be used on formal invoice
- Not to be used for single customer view – use master account holder name instead
- Speak to Peter in Shipping for further details

ASSET TAGGING

Classifications

- Identified by Scans
- Use a predefined set of classification rules
- Draw upon existing set of Microsoft classification rules
- Help determine the security treatment & source of the data

What is it?

Glossary Terms

- Manually tagged
- Less rule based, more descriptive
- Used for user searching & related term lookups
- Put the asset in business context

What does it mean & how should it be used?

ASSET TAGGING

Column name	Classifications	Glossary terms	Data type
CompanyName			nvarchar
CustomerID			int
EmailAddress	Email Address		nvarchar
FirstName	Person's Name	Customer Name	nvarchar
LastName	Person's Name	Customer Name	nvarchar

ASSET TAGGING

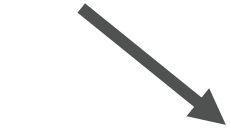
- Scan Results
- Curating Glossary Terms
- Creating Classifications

LINEAGE

ASSET LINEAGE

Customer

Azure SQL Table



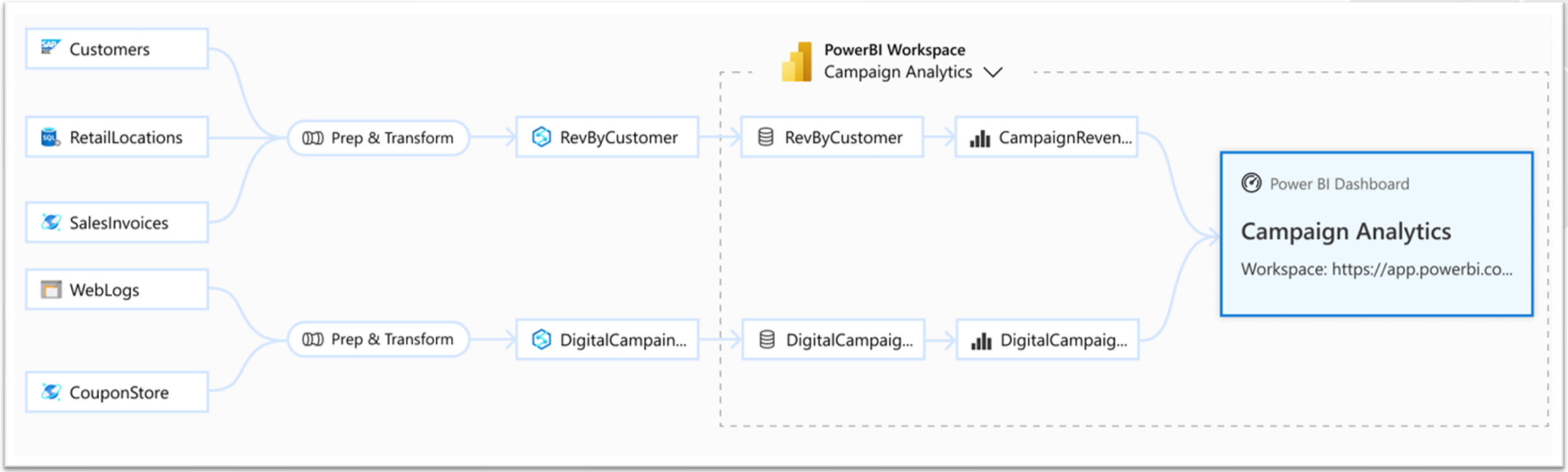
Copy Table to Lake

Azure Data Factory Copy Activity

SalesLT.Customer.csv

Azure Data Lake Storage Gen2 | File

ASSET LINEAGE



PURVIEW LINEAGE

- Setting up Data Factory connectivity
- Tracking Lineage

ASSET INSIGHTS

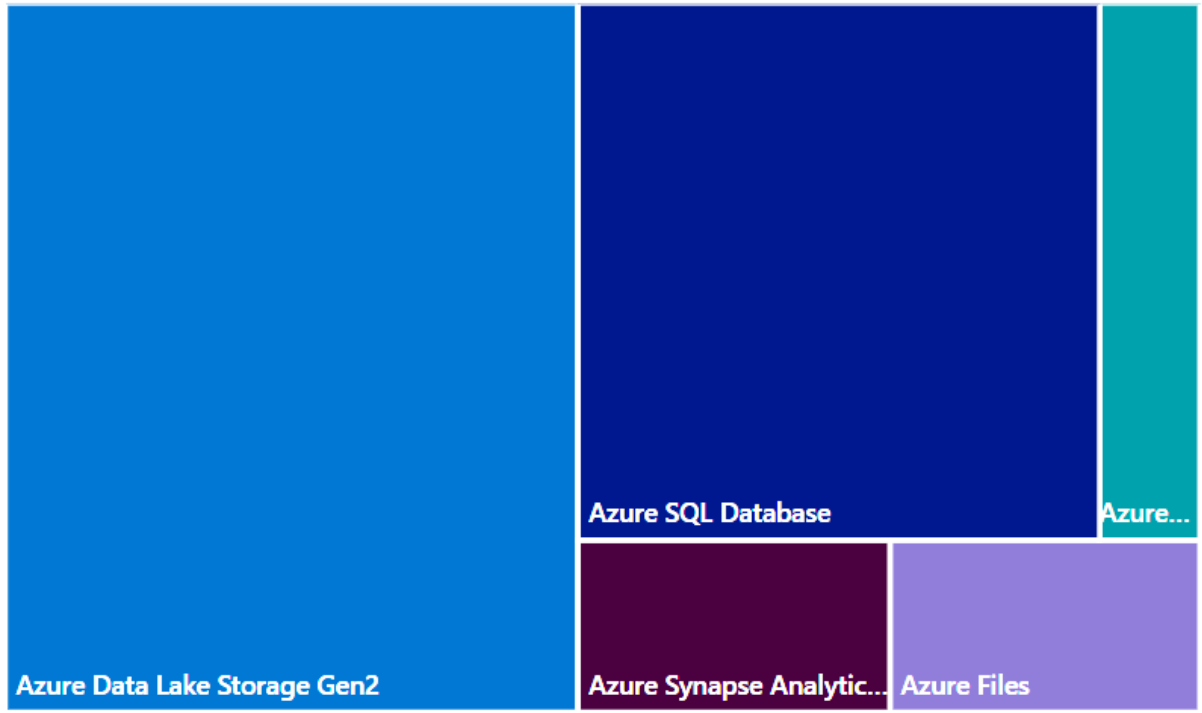
ASSET INSIGHT REPORTS

- Assets
- Scans
- Glossary
- Classification
- Sensitivity labels
- File extensions

Source types	Discovered assets	Classified assets
5	239	10

Asset count per source type

Classification category : **All** | Classification : **All**

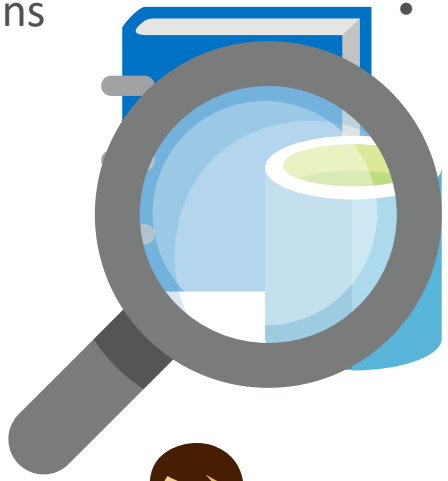


ASSET INSIGHTS

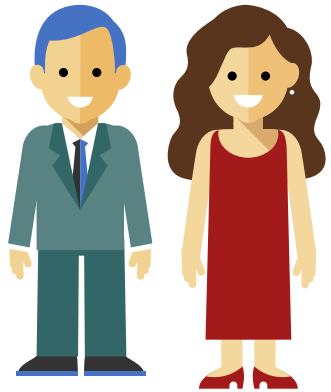
- Default Reports
- Digging Further

HOW WELL ARE WE COVERED?

Information Security
- GDPR / PII
- Access Management
- Data Classifications
- Data Security



Data Owners
- Data Duplication
- Data Development Standards
- Data Dictionary
- Business Glossary



Data Consumers
- Data Discovery
- Data Connectivity
- Documentation

Data Stewards
- Data Quality Metrics
- Data Lineage
- Data Profiling

Data Engineering
- Data Processing Rules
- Dependency Chains
- Engineering Metadata

THANKS FOR LISTENING

Twitter: @MrSiWhiteley

youtube.com/c/AdvancingAnalytics

AdvancingAnalytics.co.uk