Servus!
Hi!
Welcome!



Wolfgang Strasser
Data Consultant @ ACP CUBIDO Digital Solutions
workingondata.wordpress.com
https://www.youtube.com/@WolfgangStrasser
Microsoft® MVP Most Valuable Professional

The Sky is not the Limit
Extending Microsoft Purview

Wolfgang Strasser





THANK YOU TO OUR SPONSORS

PASS DEUTSCHLAND e.V. Microsoft Data Platform Community

pmOne

SOLARWINDS

smart casual datadesign

Mainzer Datenfabrik



b.telligent
smart data. smart decisions.

HMS analytical software

ORAYLIS
turn your data into value

TIMEXTENDER



Microsoft

ITGAIN



SOLISYON
www.solisyon.de

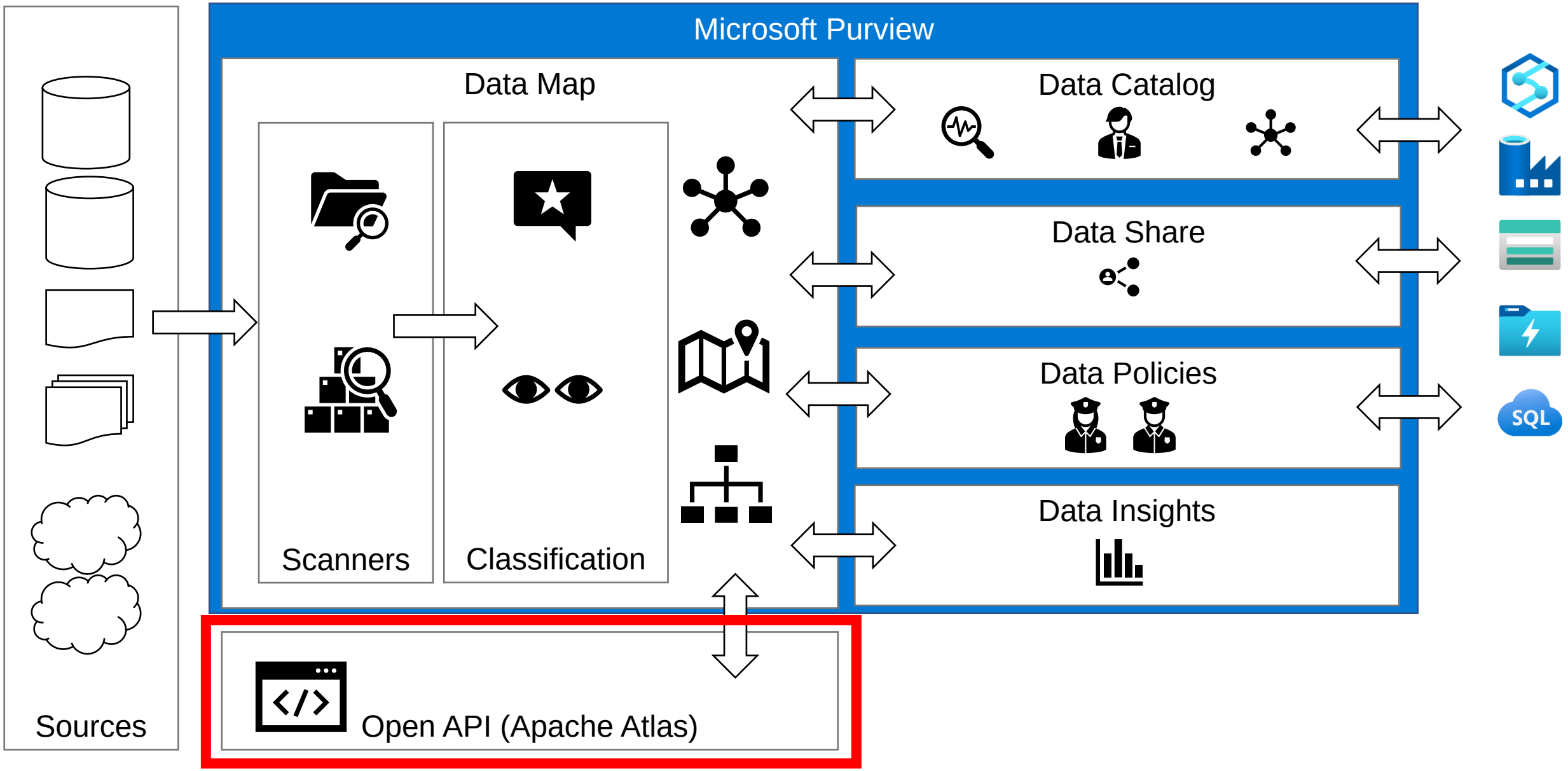


GDS BUSINESS INTELLIGENCE GmbH

Microsoft Purview

Data Governance




Microsoft Purview
Data Map
Scanners
Classification
Data Catalog
Data Share
Data Policies
Data Insights
Sources
Open API (Apache Atlas)
SQL

Home

Data Catalog
Overview
Browse
Source types
Collections
Data sharing

Apps

Related apps
Data Map
Data Estate Insights

Data Catalog

Browse, search, and discover data assets across your organization.

14 sources 890+ assets 17 glossary terms

Search catalog



Explore your data

Microsoft Azure

View Azure subscriptions and their contents.

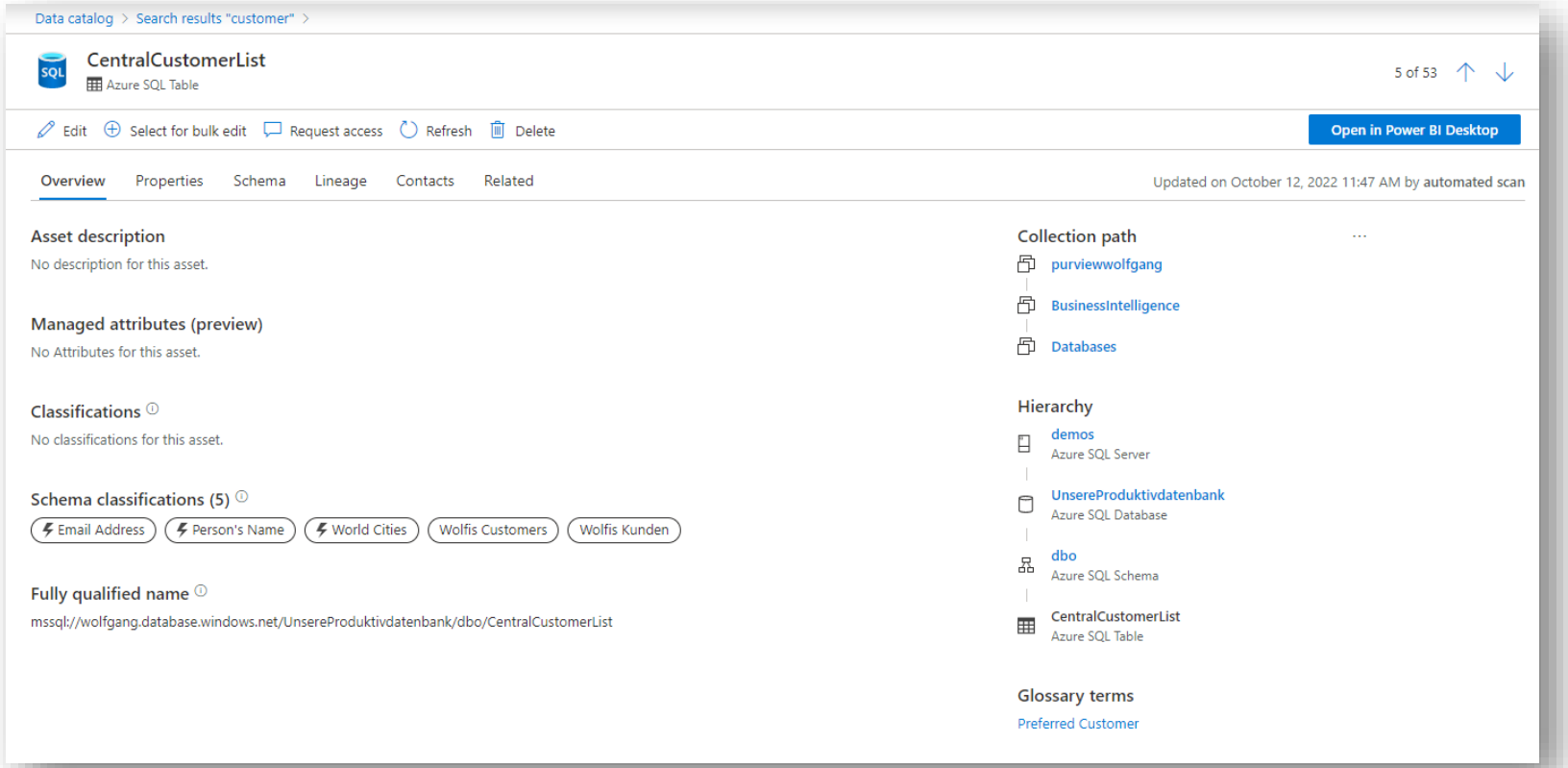
Quick access

Recent My items

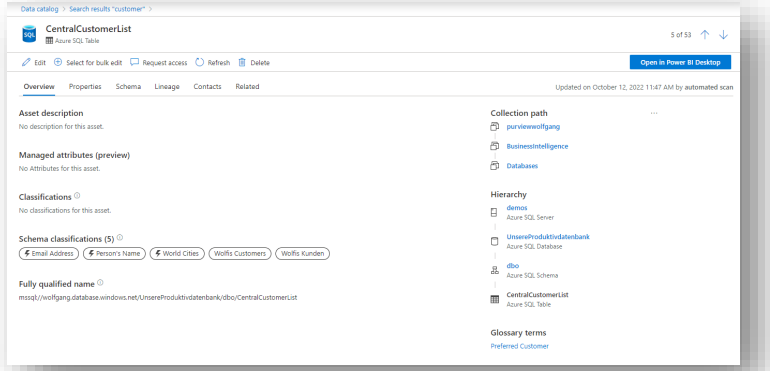
Name	Last update
CentralCustomerList	a month ago
BuildVersion	---
Articles	---
Sales	a month ago
Sales with Employees	2 months ago
Another_Lakehouse	a month ago

View all recently accessed

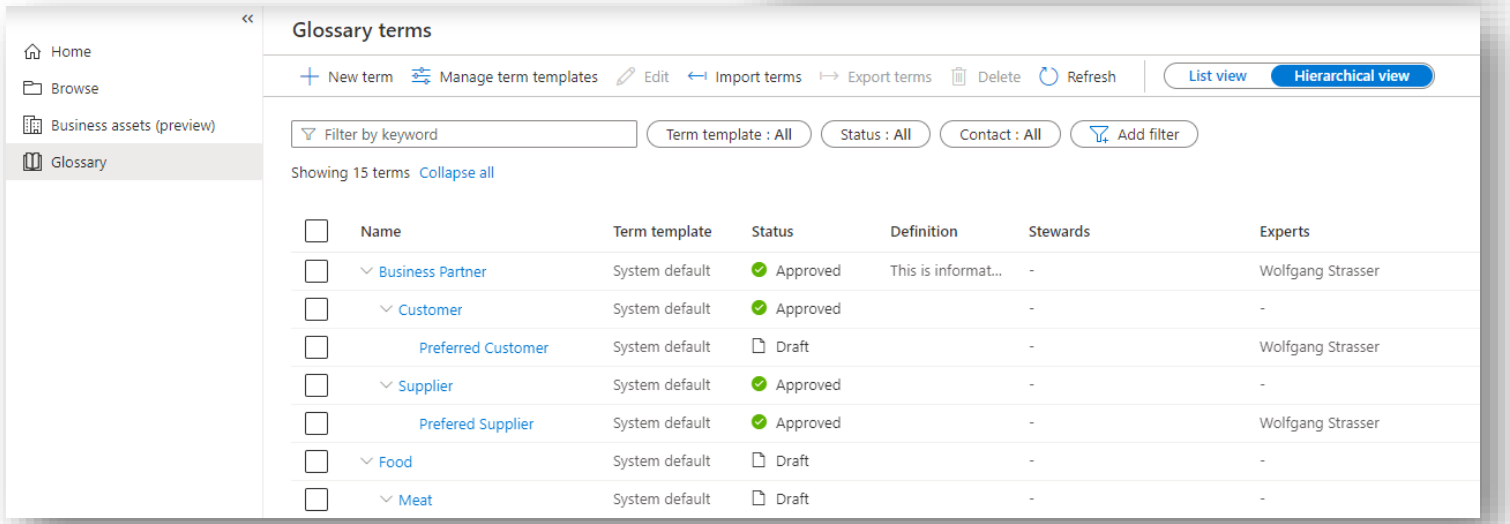
Data Assets


Data catalog > Search results "customer" >
CentralCustomerList
Azure SQL Table
5 of 53
Edit
Select for bulk edit
Request access
Refresh
Delete
Open in Power BI Desktop
Overview
Properties
Schema
Lineage
Contacts
Related
Updated on October 12, 2022 11:47 AM by automated scan
Asset description
No description for this asset.
Managed attributes (preview)
No Attributes for this asset.
Classifications
No classifications for this asset.
Schema classifications (5)
Email Address
Person's Name
World Cities
Wolfis Customers
Wolfis Kunden
Fully qualified name
mssql://wolfgang.database.windows.net/UnsereProduktivdatenbank/dbo/CentralCustomerList
Collection path
purviewwolfgang
BusinessIntelligence
Databases
Hierarchy
demos
Azure SQL Server
UnsereProduktivdatenbank
Azure SQL Database
dbo
Azure SQL Schema
CentralCustomerList
Azure SQL Table
Glossary terms
Preferred Customer

Data Assets



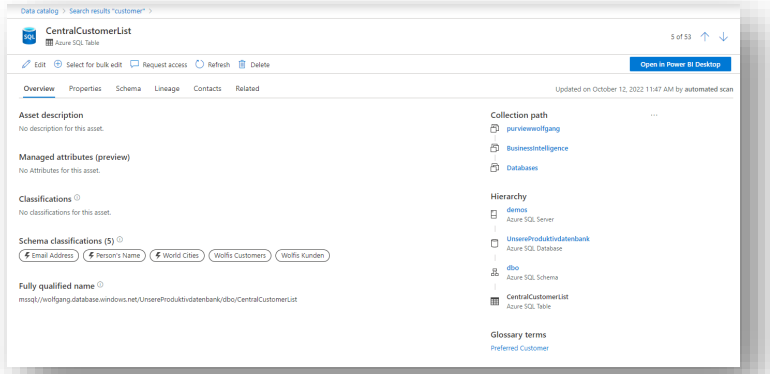
Glossary (Items)

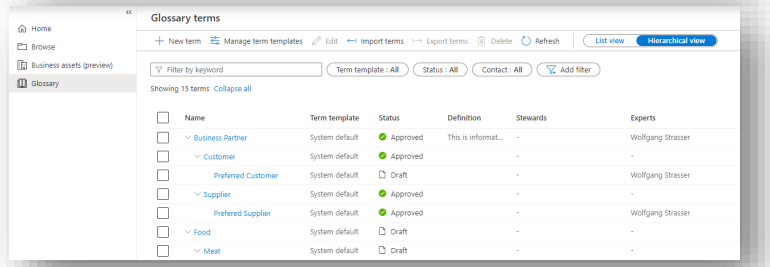


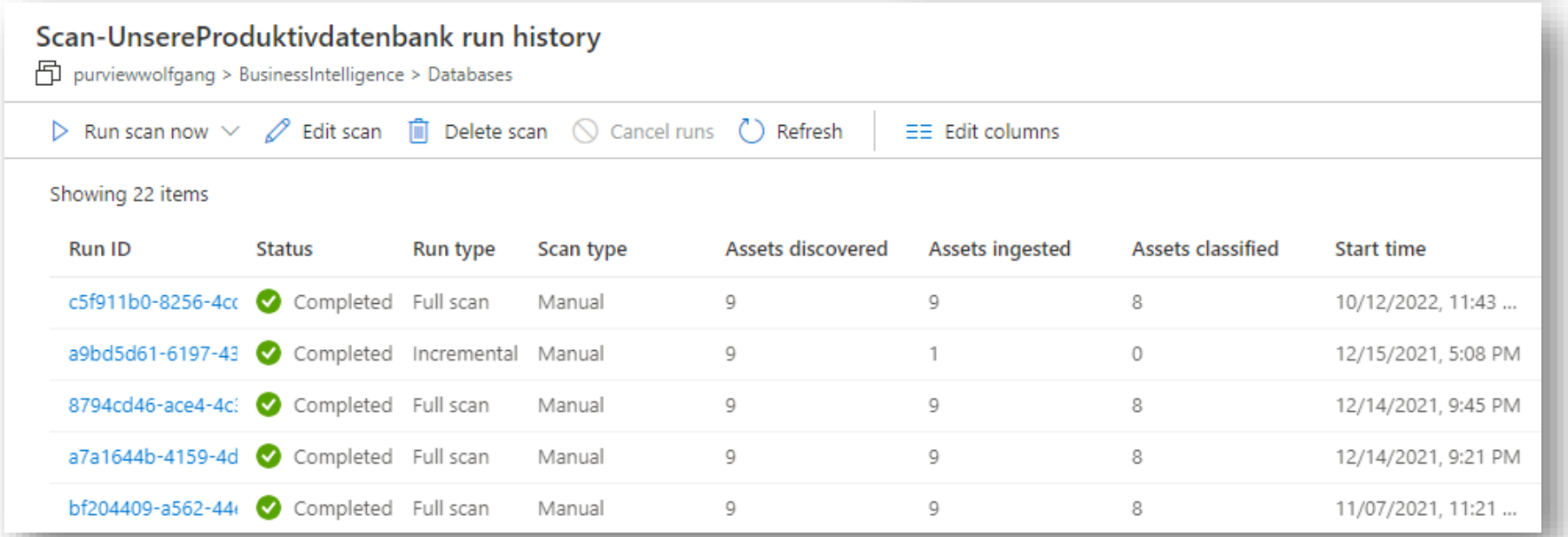
Data Assets

Glossary (Items)

Scans





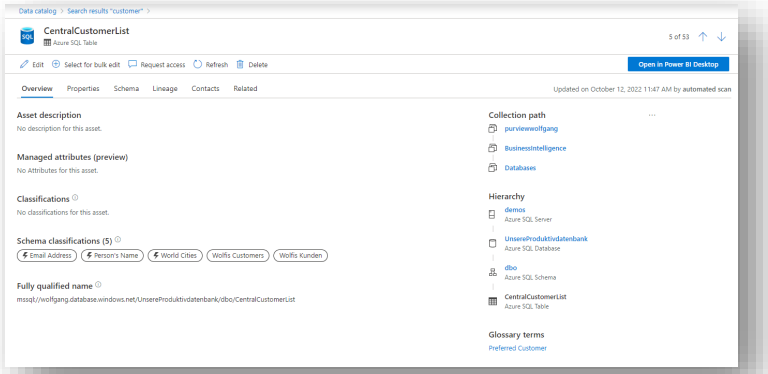


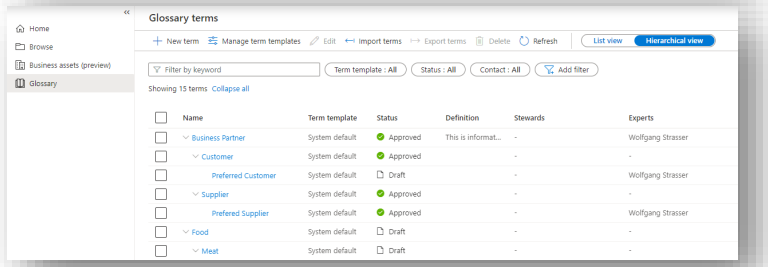
Data Assets

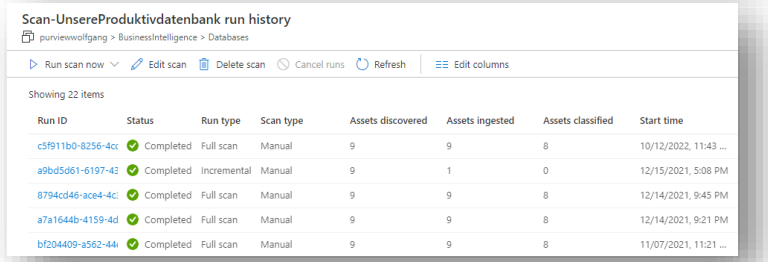
Glossary (Items)

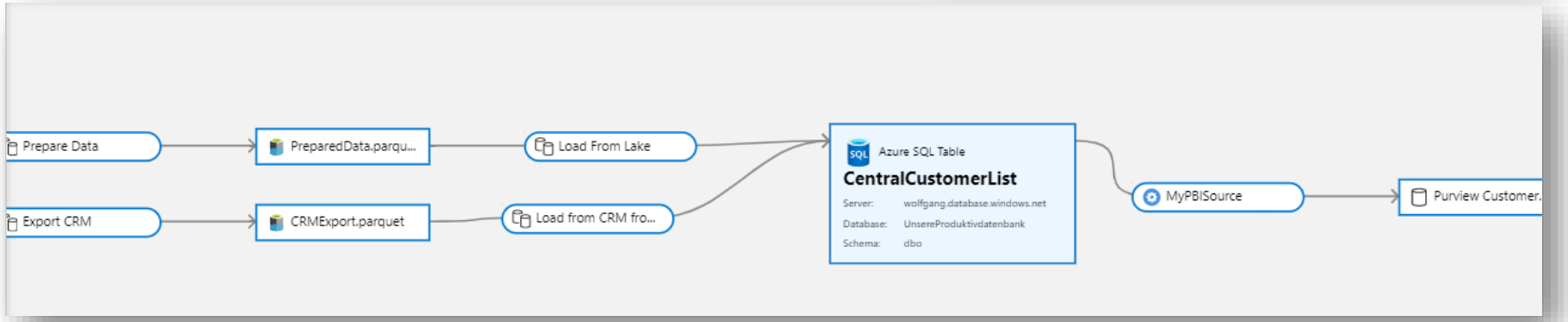
Scans

Lineage

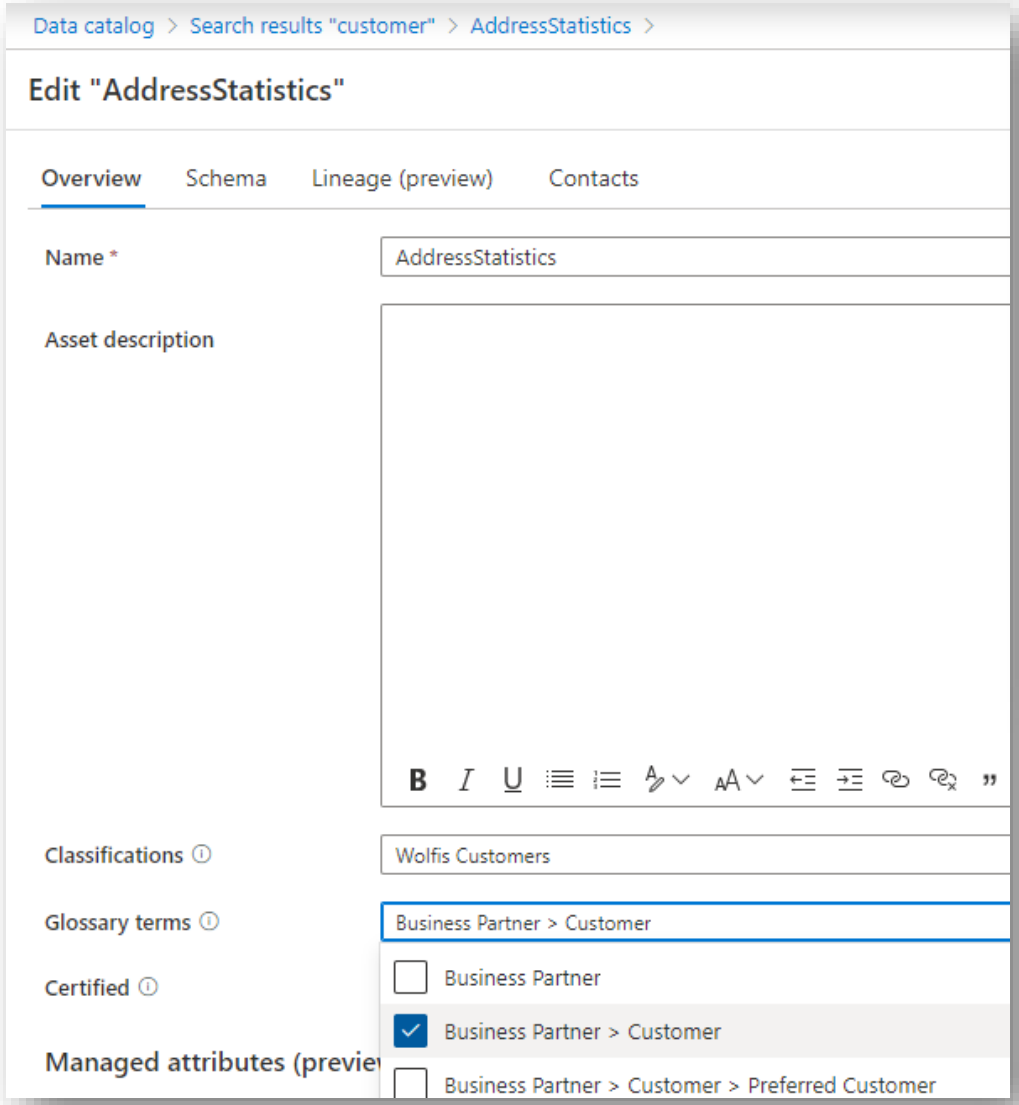


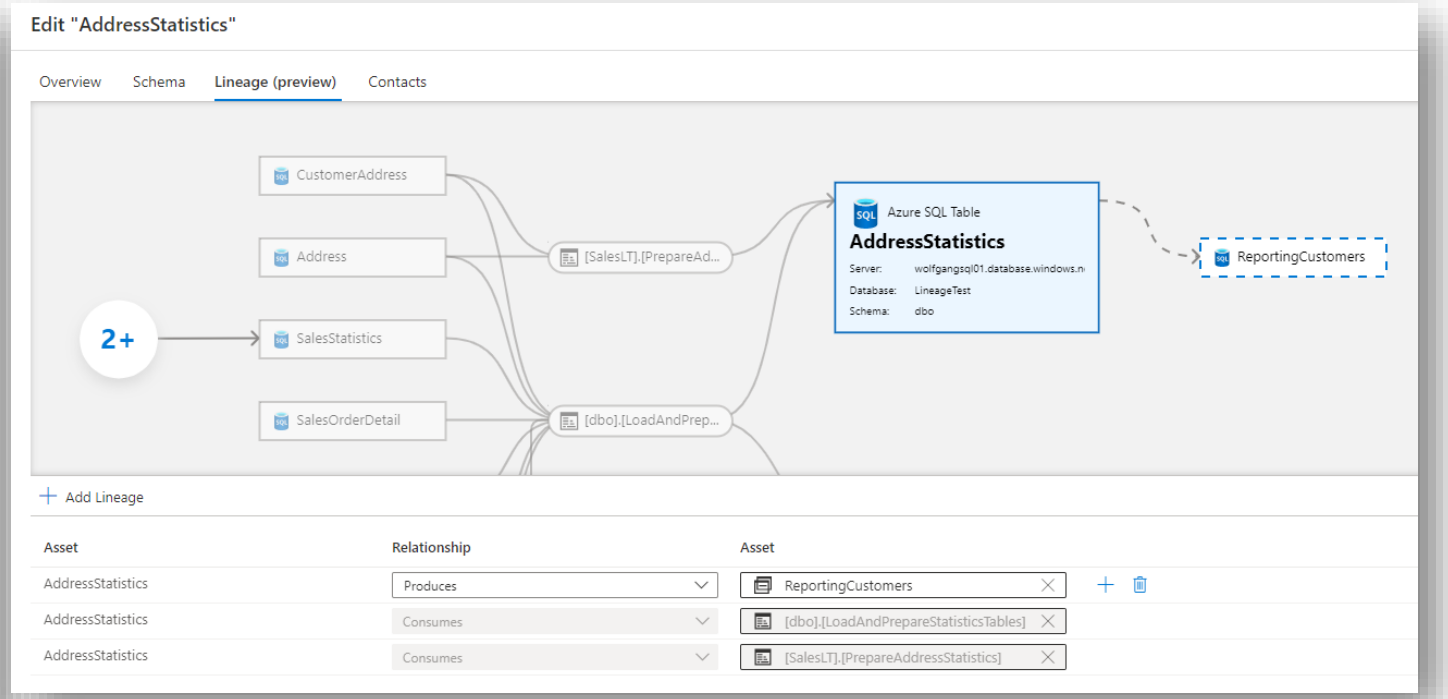






Manual Curation






?
Automation
Bulk Actions
Reporting

APIs

Application Programming Interface (API)

An application programming interface (API) is a way for **two or more computer** programs to **communicate with each other**.

It is a type of **software interface**, offering a service to other pieces of software.

Reddy, Marathi (2011). *API Design for C++*. Elsevier Science. p. 1. ISBN 9780123850041.

REST APIs

REST APIs
REpresentational State Transfer

Architectureal approach to design web services

Resources – each one has an identifier

```
https://adventure-works.com/orders/1
```

Clients interact by exchanging
representations of resources

```
{
  "orderId":1,
  "orderValue":99.90,
  "productId":1,
  "quantity":1
}
```

API operations

GET retrieve representation of the resource

POST creates a new resource

PUT create or replace resource

PATCH partial update

DELETE removes the resource

Resource	POST	GET	PUT	DELETE
/customers	Create a new customer	Retrieve all customers	Bulk update of customers	Remove all customers
/customers/1	Error	Retrieve the details for customer 1	Update the details of customer 1 if it exists	Remove customer 1
/customers/1/orders	Create a new order for customer 1	Retrieve all orders for customer 1	Bulk update of orders for customer 1	Remove all orders for customer 1

https://learn.microsoft.com/en-us/azure/architecture/best-practices/api-design

Apache ATLAS API

Apache ATLAS
Data Governance and Metadata framework for Hadoop

Build a catalog of their data assets

Classify and govern these assets

Provide collaboration

https://atlas.apache.org/#/

Key Features

Metadata types & instances

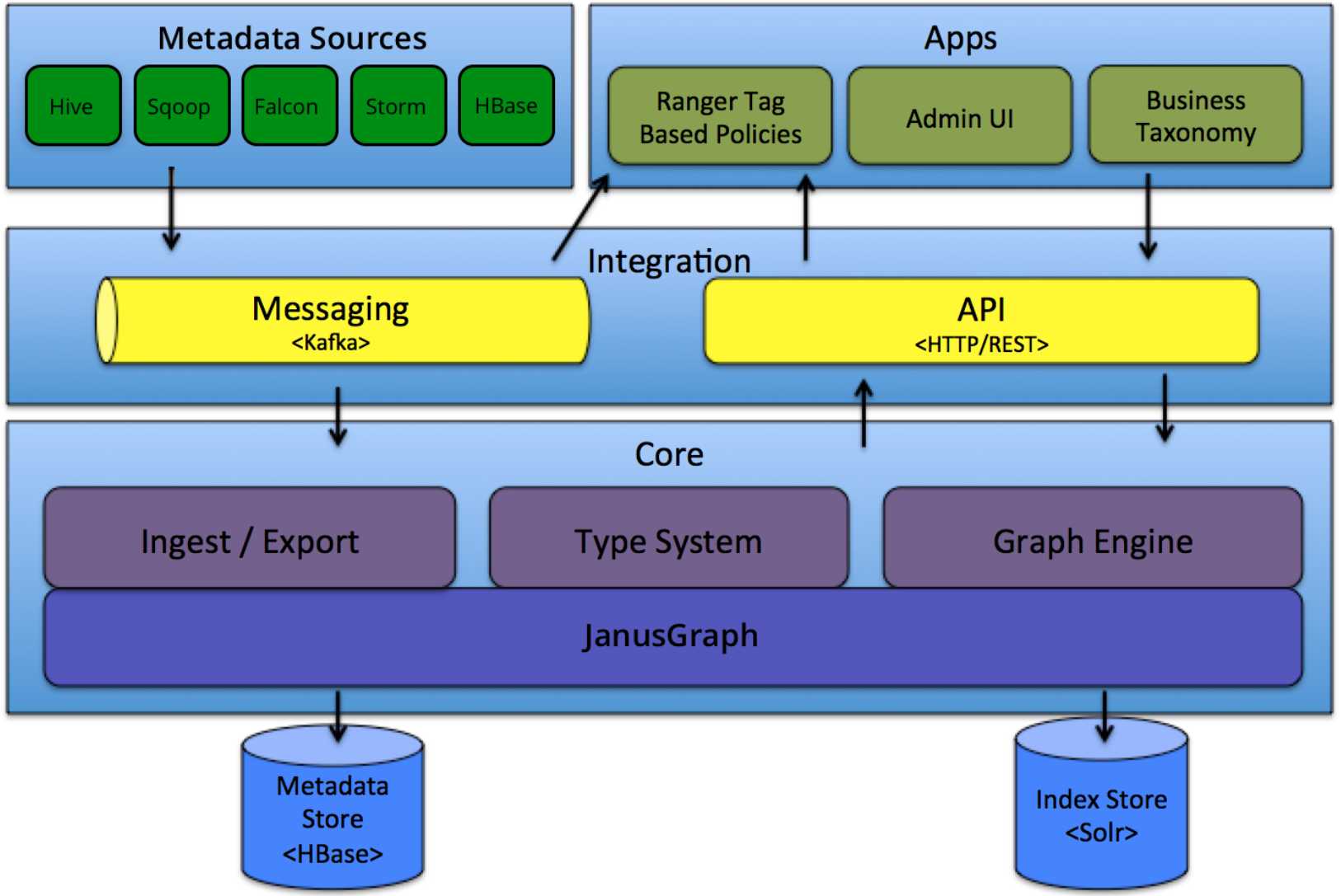
Classification

Lineage

Search/Discovery

Security / Data Masking

Apache Atlas High Level Architecture



https://atlas.apache.org/#/Architecture

Atlas Type System

Types	~ „Class“
Attributes	~ Members
Entities	~ „Objects“ (instance)
Relationships	~ References/Pointers

Types


```
{
  "category": "ENTITY",
  "guid": "68dfc42c-6e6d-b117-2554-991d0df22766",
  "createdBy": "admin",
  "updatedBy": "admin",
  "createTime": 1666089091994,
  "updateTime": 1666089091994,
  "version": 1,
  "name": "azure_sql_db",
  "description": "azure_sql_db",
  "typeVersion": "1.0",
  "serviceType": "Azure SQL Database",
  "options": {
    "purviewEntityExtDef": "{\"qualifiedNamePattern\":\"mssql://(?<Server>[^/{}]+)[.]database[.]windows[.]net/(?<Database>[^/
  },
  "lastModifiedTS": "1",
  "attributeDefs": [
    {
      "name": "collation",
      "typeName": "string",
      "isOptional": true,
      "cardinality": "SINGLE",
      "valuesMinCount": 0,
      "valuesMaxCount": 1,
      "isUnique": false,
      "isIndexable": false,
      "includeInNotification": false
    },
    ...
  ],
```

Types extends Types (super- / subtypes)


```
{
  "category": "ENTITY",
  "guid": "68dfc42c-6e6d-b117-2554-991d0df22766",
  "createdBy": "admin",
  "updatedBy": "admin",
  "createTime": 1666089091994,
  "updateTime": 1666089091994,
  "version": 1,
  "name": "azure_sql_db",
  "description": "azure_sql_db",
  "typeVersion": "1.0",
  "serviceType": "Azure SQL Database",
  "options": {},
  "lastModifiedTS": "1",
  "attributeDefs": [],
  "superTypes": [
    "azure_resource",
    "Database"
  ],
  "subTypes": [],
```



```
{
  "category": "ENTITY",
  "guid": "b613f685-7134-5d37-0477-1fe937b087ca",
  "createdBy": "admin",
  "updatedBy": "admin",
  "createTime": 1666089091985,
  "updateTime": 1666089091985,
  "version": 1,
  "name": "azure_resource",
  "description": "azure_resource",
  "typeVersion": "1.0",
  "lastModifiedTS": "1",
  "attributeDefs": [],
  "superTypes": [
    "Asset"
  ],
  "subTypes": [
    "azure_sql_server",
    "ads_share",
    "azure_synapse_dedicated_sql_db",
    "ads_account",
    "azure_mysql_server",
    "azure_datalake_gen1_account",
    "azure_synapse_workspace",
    "azure_machine_learning_account",
    "azure_data_explorer_database",
    "azure_sql_dw",
    "azure_storage_service",
    "azure_cognitive_search_service",
    "arcsql_instance",
    "azure_data_factory",
    "azure_postgresql_server",
    "azure_cosmosdb_account",
    "ads_share_subscription",
    "azure_sql_mi",
    "azure_data_explorer_cluster",
    "azure_storage_account",
    "azure_sql_db",
    "azure_mariadb_server"
  ],
```

Relationships


```
{
  "category": "ENTITY",
  "guid": "68dfc42c-6e6d-b117-2554-991d0df22766",
  "createdBy": "admin",
  "updatedBy": "admin",
  "createTime": 1666089091994,
  "updateTime": 1666089091994,
  "version": 1,
  "name": "azure_sql_db",
  "description": "azure_sql_db",
  "typeVersion": "1.0",
  "serviceType": "Azure SQL Database",
  "superTypes": [
    "azure_resource",
    "Database"
  ],
  "subTypes": [],
  "relationshipAttributeDefs": [
    ...,
    {
      "name": "server",
      "typeName": "azure_sql_server",
      "isOptional": false,
      "cardinality": "SINGLE",
      "valuesMinCount": -1,
      "valuesMaxCount": -1,
      "isUnique": false,
      "isIndexable": false,
      "includeInNotification": false,
      "relationshipTypeName": "azure_sql_server_databases",
      "isLegacyAttribute": false
    },
    ...,
    ...
  ],
  "businessAttributeDefs": {
    "General Data System Attributes": ...,
    "test group": ...
  }
},
```

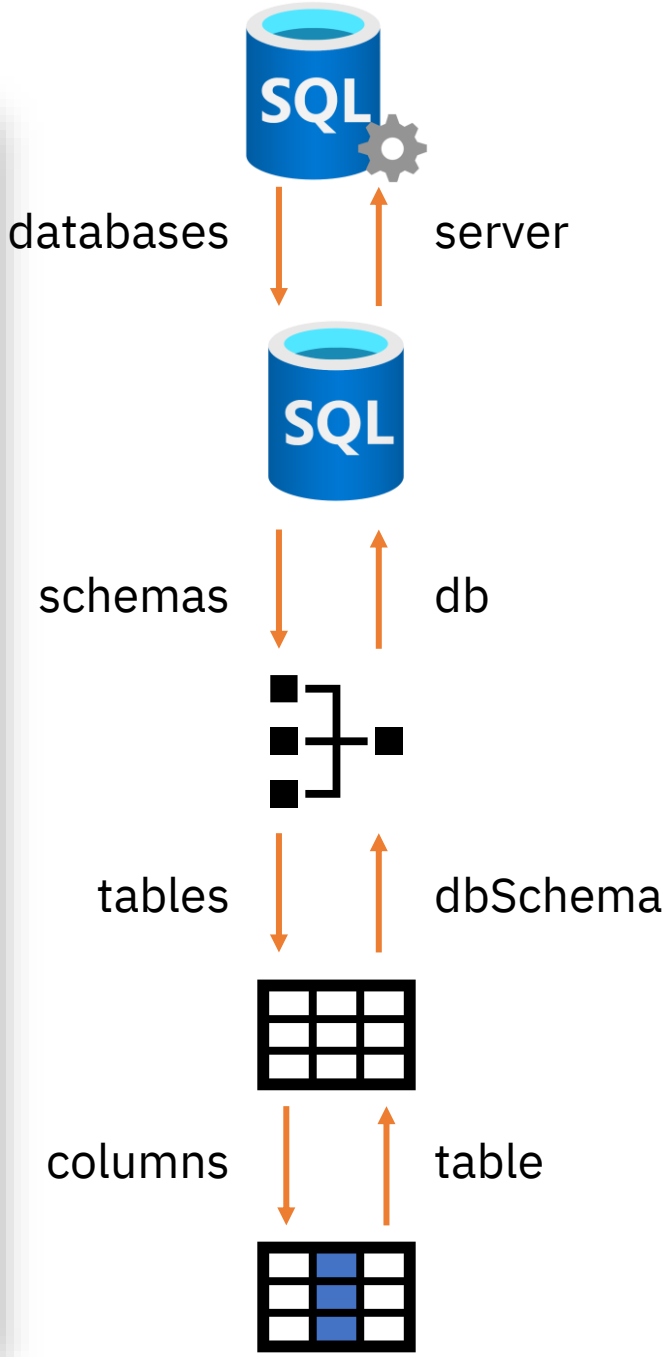

Hierarchy

- demos
  Azure SQL Server
- UnsereProduktivdatenbank
  Azure SQL Database
- dbo
  Azure SQL Schema
- CentralCustomerList
  Azure SQL Table

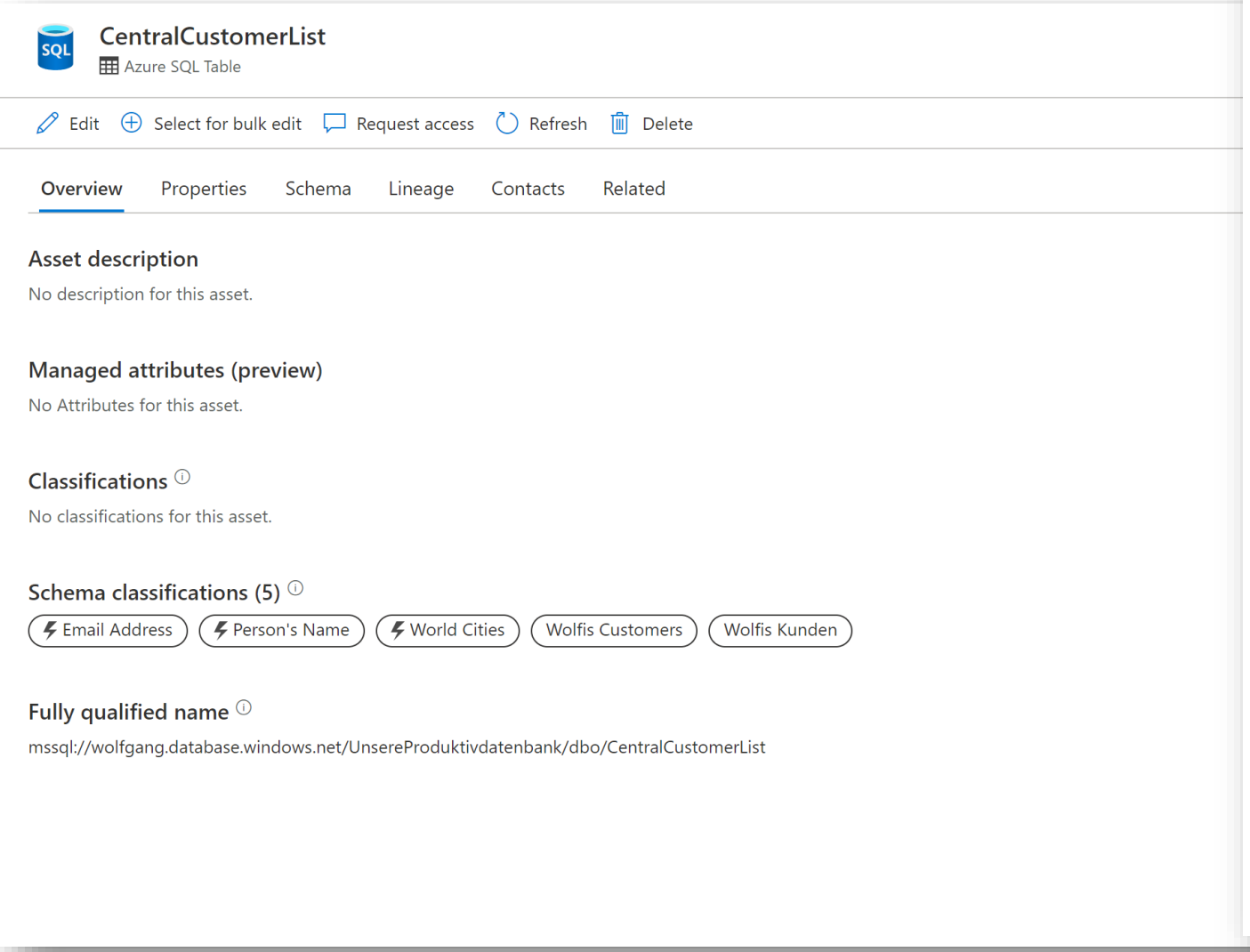
Relationships

```
"relationshipAttributeDefs": [
    {
        "name": "server",
        "typeName": "azure_sql_server",
        "isOptional": false,
        "cardinality": "SINGLE",
        "valuesMinCount": -1,
        "valuesMaxCount": -1,
        "isUnique": false,
        "isIndexable": false,
        "includeInNotification": false,
        "relationshipTypeName": "azure_sql_server_database
        "isLegacyAttribute": false
    },
    {
        "name": "schemas",
        "typeName": "array<azure_sql_schema>",
        "isOptional": true,
        "cardinality": "SET",
        "valuesMinCount": -1,
        "valuesMaxCount": -1,
        "isUnique": false,
        "isIndexable": false,
        "includeInNotification": false,
        "constraints": [
            {
                "type": "ownedRef"
            }
        ],
        "relationshipTypeName": "azure_sql_db_schemas",
        "isLegacyAttribute": false
    },
```

```
{
    "category": "RELATIONSHIP",
    "guid": "cd47211b-0de9-05ec-4e1e-ea513f062747",
    "createdBy": "admin",
    "updatedBy": "admin",
    "createTime": 1667400144438,
    "updateTime": 1667400144438,
    "version": 1,
    "name": "azure_sql_db_schemas",
    "description": "azure_sql_db_schemas",
    "typeVersion": "1.0",
    "serviceType": "Azure SQL Database",
    "lastModifiedTS": "1",
    "attributeDefs": [],
    "relationshipCategory": "COMPOSITION",
    "propagateTags": "NONE",
    "endDef1": {
        "type": "azure_sql_db",
        "name": "schemas",
        "isContainer": true,
        "cardinality": "SET",
        "isLegacyAttribute": false
    },
    "endDef2": {
        "type": "azure_sql_schema",
        "name": "db",
        "isContainer": false,
        "cardinality": "SINGLE",
        "isLegacyAttribute": false
    }
},
```

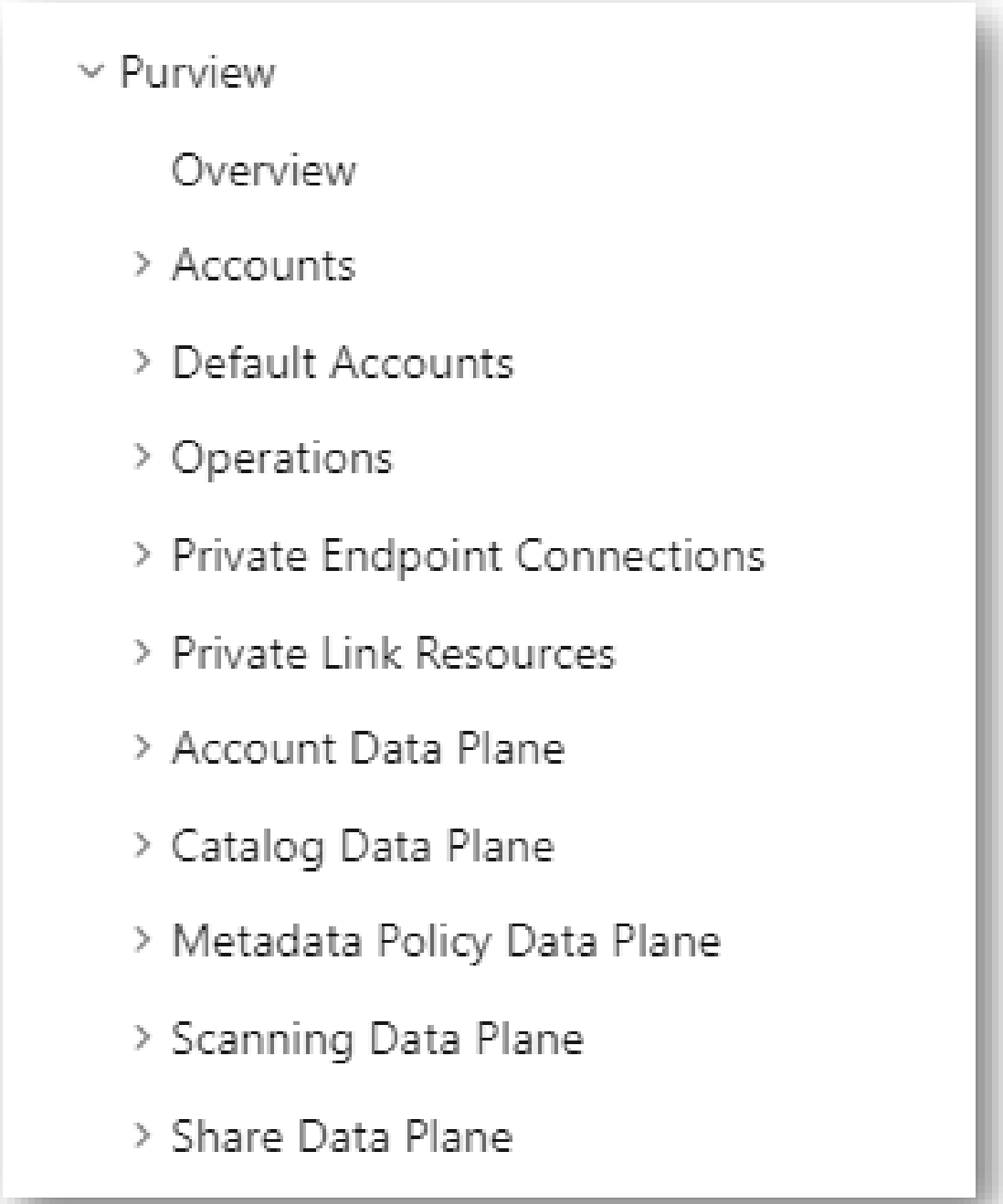


Entities = Data Assets

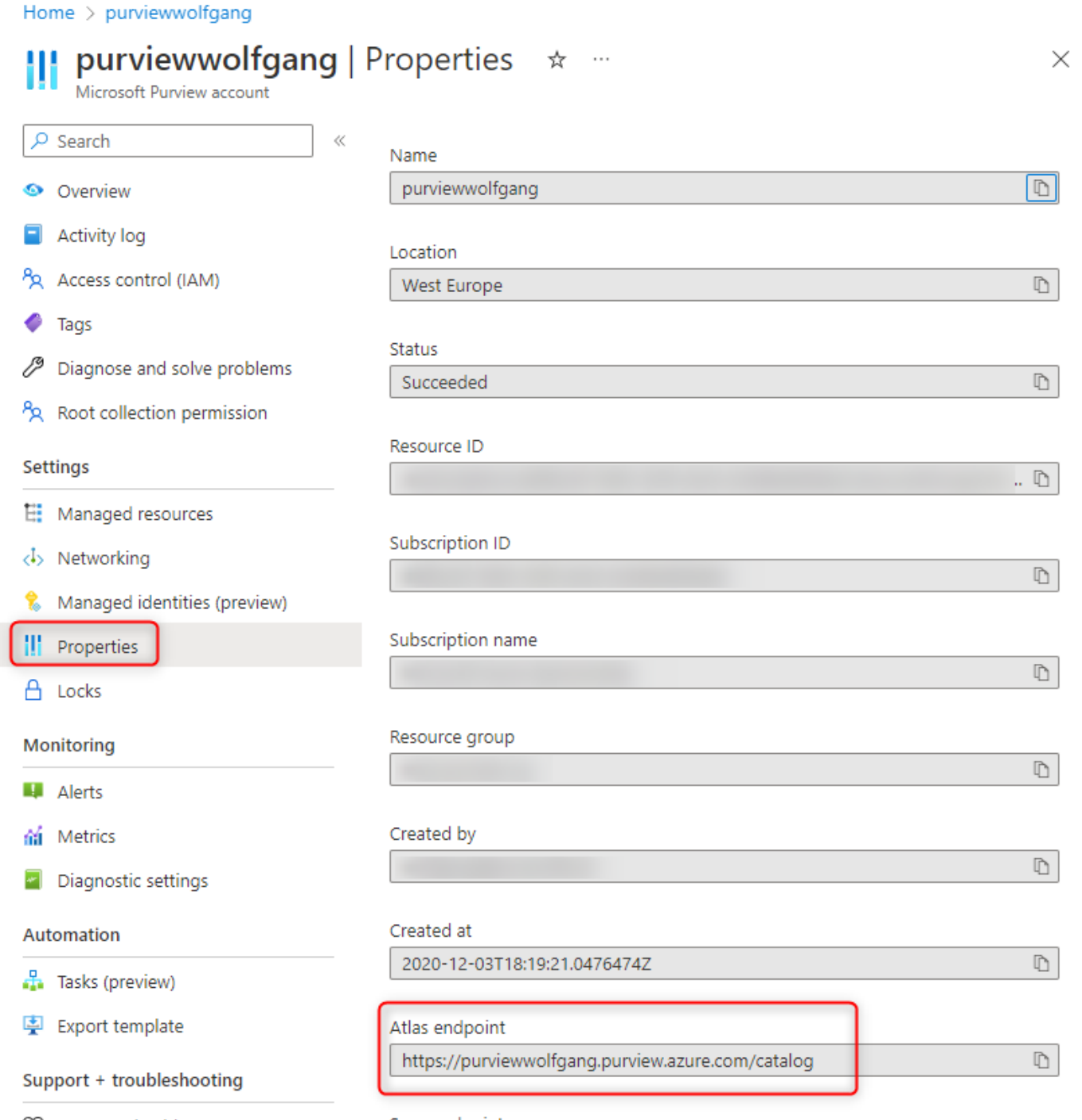



```
"entity": {
    "typeName": "azure_sql_table",
    "attributes": {
        "owner": null,
        "modifiedTime": 1631730943000,
        "replicatedTo": null,
        "userDescription": null,
        "replicatedFrom": null,
        "createTime": 1631730943000,
        "qualifiedName": "mssql://wolfgang.database.windows.net/UnsereProdu
        "displayName": "City",
        "name": "CentralCustomerList",
        "description": null,
        "principalId": 0,
        "objectType": "U "
    },
    "lastModifiedTS": "30",
    "guid": "e0455127-1635-44d1-9ffb-b6f6f6f60000",
    "isIncomplete": false,
    "status": "ACTIVE",
    "createdBy": "ServiceAdmin",
    "updatedBy": "ServiceAdmin",
    "createTime": 1631731187164,
    "updateTime": 1665568041660,
    "version": 0,
    "source": "DataScan",
    "sourceDetails": {
        "ScanResourceId": "datasources/AzureSqlDatabase-wolfgang/scans/Scan
        "ScanLastModifiedAt": "2021-12-15T16:07:44.6322435Z",
        "ScanRuleSetResourceId": "scanrulesets/AzureSQL_CustomScanRuleSet"
        "ScanRuleSetLastModifiedAt": "2021-10-11T15:11:51.1581093Z",
        "__schemaModifiedBy": "DataScan",
        "__schemaModifiedTime": 1667238984912
    },
```

Microsoft Purview and Apache ATLAS



https://learn.microsoft.com/en-us/rest/api/purview/

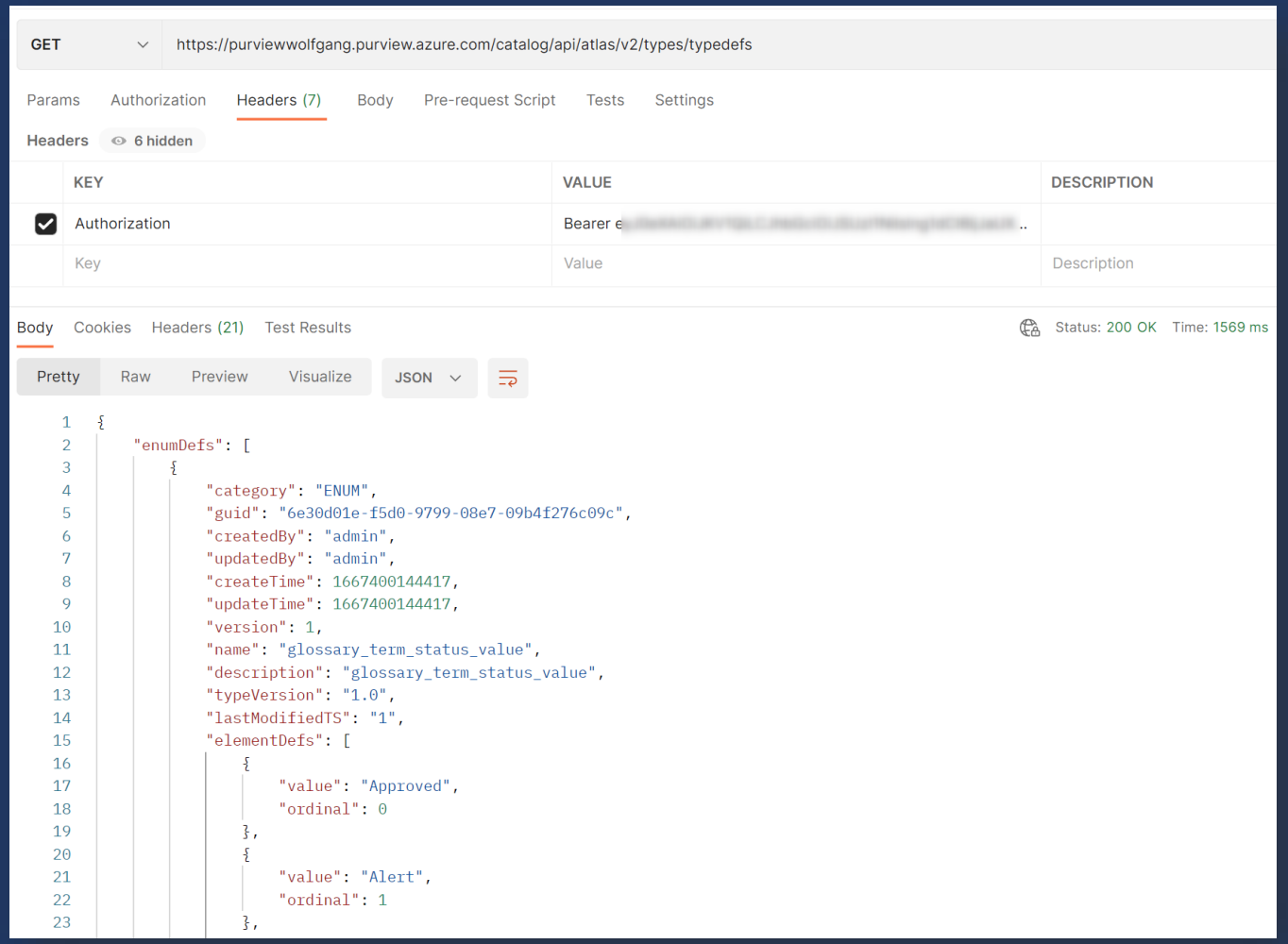


https://learn.microsoft.com/en-us/azure/purview/tutorial-atlas-2-2-apis

DEMOs

Purview – ATLAS API

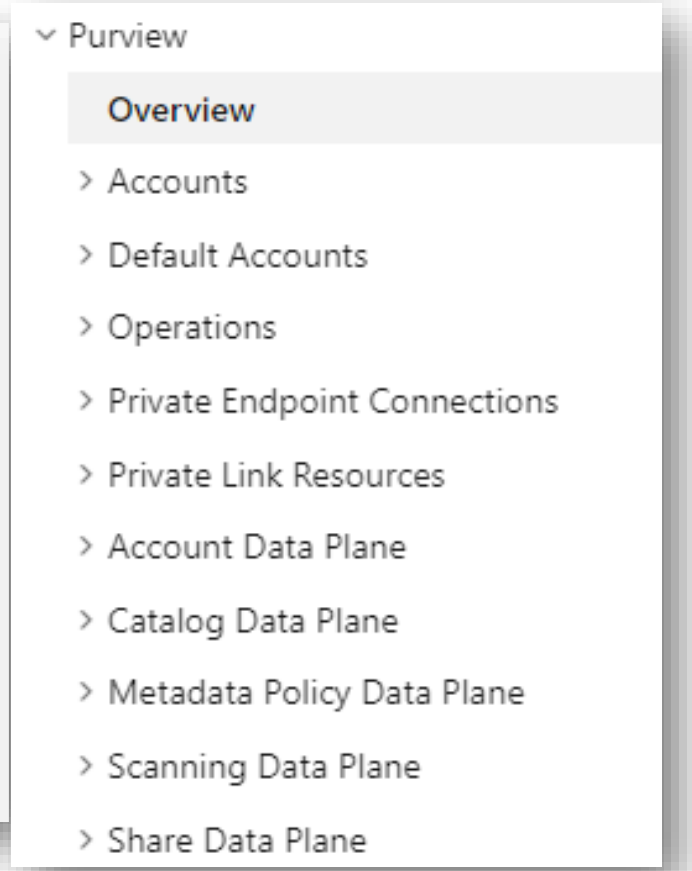
Postman



Other Ways to “talk” to your Purview instance

REST Operation Groups

Operation Group	Description
Accounts	Provides operations to manage purview accounts including account name availability, creating or updating account resources and listing accounts by resource group or subscription.
Default Accounts	Provides operations to manage default account for a scope (tenant).
Operations	Provides operations to manage Microsoft Purview resource.
Private Endpoint connections	Provides operations to manage Private endpoint connections for Microsoft Purview accounts including approving or rejecting private endpoint connection request.
Private Link Resources	Provides operations to manage private link resources for Microsoft Purview.

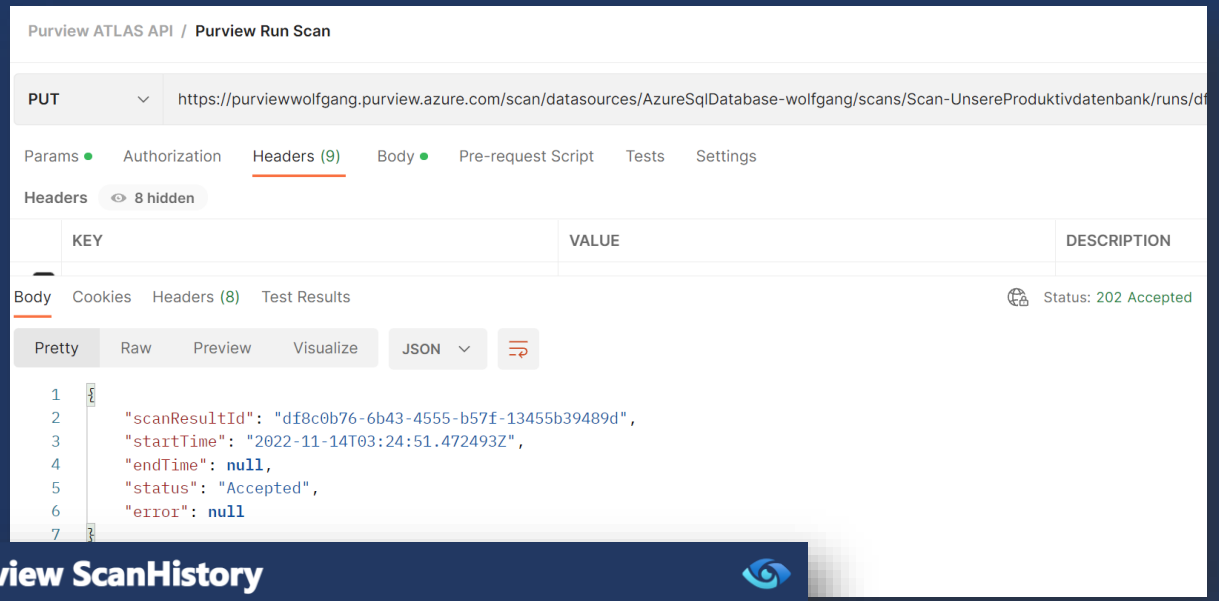


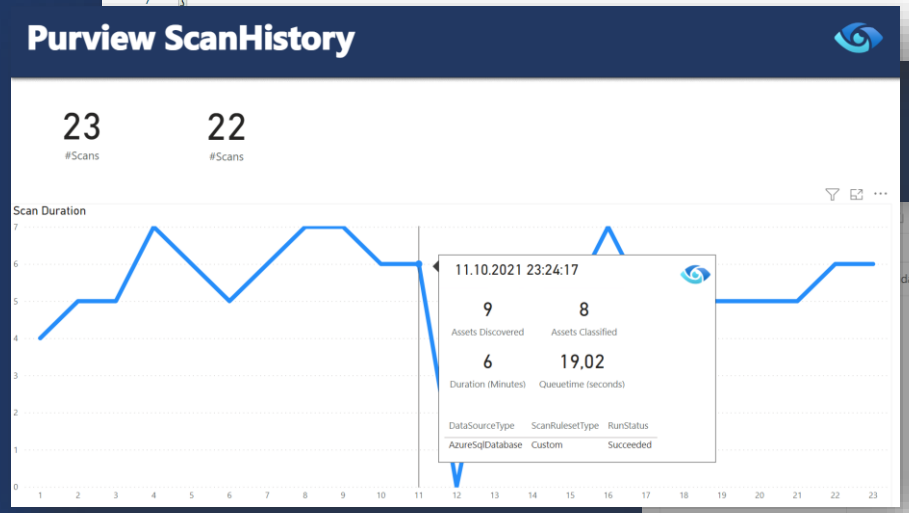
https://learn.microsoft.com/en-us/rest/api/purview/

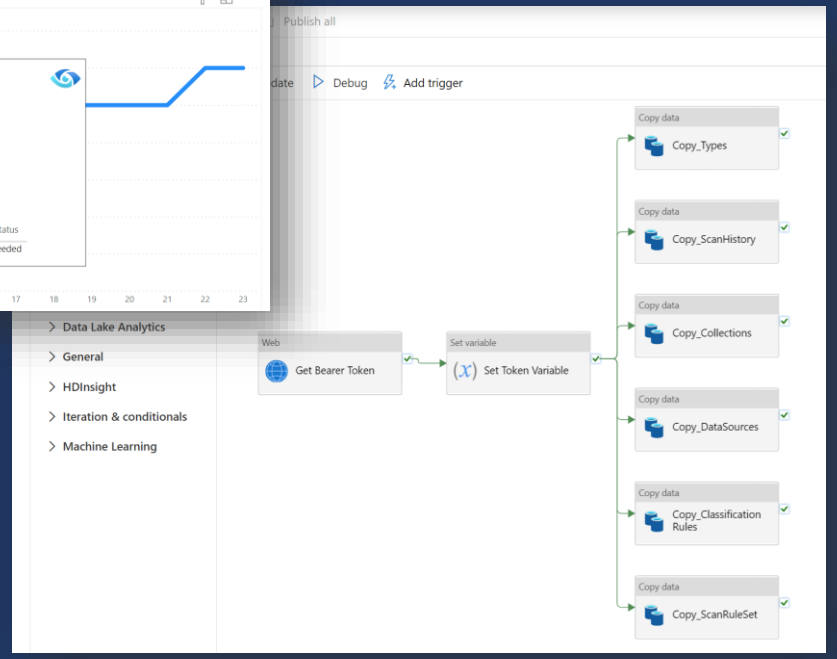
DEMOs

Purview – REST API

Trigger Scans, Synapse, Power BI

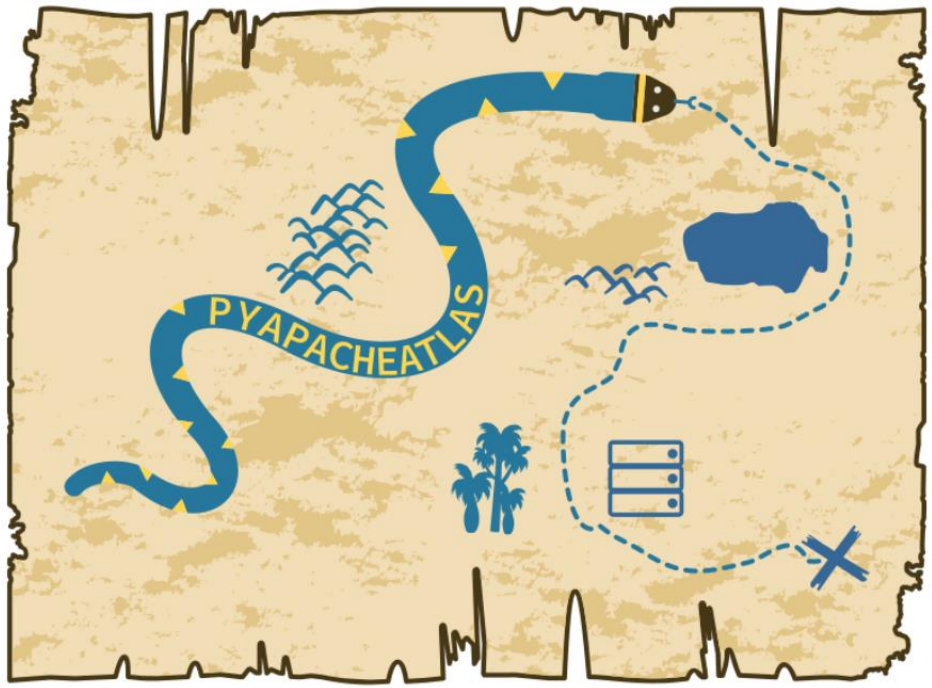






pyapacheatlas

PyApacheAtlas: A Python SDK for Azure Purview and Apache Atlas



PyApacheAtlas lets you work with the Azure Purview and Apache Atlas APIs in a Pythonic way. Supporting bulk loading, custom lineage, custom type definition and more from an SDK and Excel templates / integration.

https://github.com/wjohnson/pyapacheatlas

DEMOs

(Purview CLI https://github.com/tayganr/purviewcli)

PyApacheAtlas: https://github.com/wjohnson/pyapacheatlas

Summary

Purview UI is not the end..

Apache ATLAS

REST APIs

C#, CLI, python, …

Dankeschön
Thank you

Gibt’s noch Fragen?
Any Questions?



Wolfgang Strasser
Data Consultant @ ACP CUBIDO Digital Solutions
workingondata.wordpress.com
https://www.youtube.com/@WolfgangStrasser
Microsoft® MVP Most Valuable Professional