Govern your Fabric environment

Hylke Peek

Hylke Peek

Data Consultant

I'm working on a lot of different stuff. Focus on **Azure**, **Fabric**, and **Power BI**.

- Data Platform
- Data warehousing
- Analytics
- Azure data solutions

linkedin.com/hylkepeek

github.com/hylkepeek

PeekData.nl

Agenda

Components

Administration

12:00 – 12:45 - Lunch break

Security

Data Governance

Development

Recap

Components

Understand Fabric architecture

Single SaaS lake

Provisioned automatically with the tenant

All workloads automatically store data in OneLake workspace folders

All data is organized in a hierarchical namespace

Data in OneLake is automatically indexed for discovery, MIP labels, lineage, PII scans, sharing, governance and compliance

Microsoft Fabric

Data Integration	Data Engineering	Data Warehouse	Data Science	Real Time Analytics	Business Intelligence	Observability
Data Factory	Synapse	Synapse	Synapse	Synapse	Power BI	Data Activator

- **Data Factory**: Data integration combining Power Query with the scale of Azure Data Factory to move and transform data.
- **Synapse Data Engineering**: Data engineering with a Spark platform for data transformation at scale.
- **Synapse Data Warehouse**: Data warehousing with industry-leading SQL performance and scale to support data use.
- **Synapse Data Science**: Data science with Azure Machine Learning and Spark for model training and execution tracking in a scalable environment.
- **Synapse Real-Time Analytics**: Real-time analytics to query and analyze large volumes of data in real-time.
- **Power BI**: Business intelligence for translating data to decisions.
- **Data Activator**: Real-time detection and monitoring of data that can trigger notifications and actions when it finds specified patterns in data.

Store data
Lakehouse
Warehouse
KQL Database
Notebook
Experiment
Model
Train and use ML Models
Prep & query data
Dataflow gen2
KQL Queryset
Data pipeline
Spark Job Definition
All-in-one
Fabric Items
Manage real-time data
Eventstream
Streaming Dataflow
Streaming Dataset
Scorecard
Report
Dashboard
Paginated Report
Present data
Realtime Dashboard
Aka.ms/FabricNotes
@cmaneu

OneLake

"The OneDrive for your data"

Visually Explained

All analytical engines can access OneLake Data

SQL | Spark | Power BI | KQL DB

DataLake + warehouse = Lakehouse

Each Fabric Organization has <u>one</u> Data Lake, entirely managed for you
It's unified across all regions & you pay per GB stored (no scaling needed).

With Fabric, your **warehouse** Data is also stored in OneLake, and accessible in an open format

This is a **Lakehouse**

It allows you to distribute ownership to organizations, and easily collaborate on data

It contains

A files area

A tables area

Shortcuts

Allow you to query data from one lakehouse to another <u>without data movement</u>

Shortcuts link to other clouds too!

ADLS Gen2 SDK & APIs can natively access OneLake

The Files area can contain unstructured & semi-structured data.

It's not uncommon to organize them in three "areas": Bronze, Silver & Gold

OneLake

Fabric Lakehouse

Data *Lake*

- Scalable, distributed file storage
- Flexible schema-on-read semantics
- Big data technology compatibility

Data Ware*house*

- Relational schema modelling
- SQL-based querying
- Proven basis for reporting and analytics

Fabric Warehouse

Warehouse

- Centred on single data lake
- Powered by Synapse Analytics
- Fully supports T-SQL
- Parquet file format
- Analyze and process structured data

Warehouse vs. Lakehouse

Warehouse

- Structured data
- Multi-table transactions
- High performance on complex queries and large data volumes
- Expansive security (object-level, DDL/DML, dynamic data masking)
- T-SQL, Spark

Lakehouse

- Semi-structured or unstructured data
- Scalable and cost-effective
- Supports Delta Lake features
- T-SQL security (row/table level)
- T-SQL, Spark (Scala, PySpark, Spark SQL, R)
- Large-scale data processing and analytics, including machine learning and predictive modelling
- Centralized location for data engineers, data scientist, and data analysts

SQL Analytics Endpoint

‘Turn your lakehouse into a SQL datawarehouse’

- Put a warehouse on top of your Lakehouse
- Query the tables that reference data in your Delta Lake folders in the lake
- Create views, inline TVFs, and procedures to encapsulate your semantics and business logic in T-SQL
- Manage permissions on the objects
- No DML

Two SQL Endpoints

In your workspace

SQL endpoint for the lakehouse

	ContosoThreadsLH	SQL endpoint
	ContosoThreadsLH	Lakehouse

Automatically created with your lakehouse

Datawarehouse

Default API for Warehouse

	ContosoThreadsDWH	Warehouse

They both store data in the Open Delta Parquet format

Is read-only. Updates through Spark

Read-write

Full transactional DDL and DML support

Spark, pipelines, dataflows, shortcuts

Ingestion

SQL, Pipelines, dataflows

Medallion architecture (bronze, silver, gold)

Recommended structure

Analytical schema with Tables, views,...

You can query one from another with the "three dot" structure (e.g. Select * From dwh.dbo.customers)

Real-Time Intelligence

Set of capabilities that offers streaming data solutions

- Capture existing events (e.g. event hub)
- Process event-data (business logic)
- Store events
- Create real-time alerts based on the events

microsoft.com

Real-Time Intelligence

Set of capabilities that offers streaming data solutions

Capture existing events

- Azure Event hub
- Custom app for all other data
- Many in preview (e.g. Azure Blob Storage events, CDC)

Process event-data

- Transformations like filter, group by time, join etc.

Store events in a destination

- KQL database
- Lakehouse
- Reflex
- Custom app

Real-Time Intelligence

Set of capabilities that offers streaming data solutions

Workspace for KQL databases

- Manage multiple databases to optimize performance and reduce costs
- A KQL-database is designed for time-based streaming events
- Expose event data to other sources using shortcuts and views

Real-Time Intelligence

Set of capabilities that offers streaming data solutions

Stored KQL query

- Analyze data stored in a KQL database
- Create, run and change queries for other items like Real-Time dashboard

Real-Time Intelligence

Set of capabilities that offers streaming data solutions

Real-time visualizations of event data

- Visualizations based on KQL queries
- Collections of tiles
- Organized into multiple pages

Dataflow Gen2

Low-code graphical environment for defining ETL solutions

- Extract data from multiple sources, transform it, and load it into a destination
- Run dataflows independently or as an activity in a Pipeline

Benefits

- Allow **self-service** users access to a subset of data warehouse separately.
- Optimize performance with dataflows, which enable extracting data once for reuse, reducing data refresh time for slower sources.
- Simplify data source complexity by **only exposing** dataflows to larger **analyst groups**.
- Ensure consistency and quality of data by enabling users to clean and transform data before loading it to a destination.

Limitations

- Not a replacement for a data warehouse.
- Row-level security isn't supported.
- Fabric capacity workspace is required.

Pipelines in Microsoft Fabric

Pipeline concepts

- Activities
  - Data transformation
  - Control flow
- Parameters
- Pipeline runs

Fabric notebooks

Fabric notebook concepts

- Code (PySpark, Scala, R, Spark SQL)
- Markdown (comments)
- Run individual or multiple cells
- Ingest and transform
- Support automation

Spark job definition

Spark job definition concepts

- Automate script execution
- Non-interactive script run in Spark at a scheduled time
- To edit and run Spark code in Fabric, use *notebooks* or create a *Spark Job Definition*

Power BI

Data analytics and visualization environment for business users

Semantic model

- Ingest data from Fabric and other sources
- Model relationships, hierarchies, measures etc.
- Create scalable dataset for performance optimization

Report

- Transform data to actionable insights
- Create standardized reports

And many more…

- Sharing inside and outside the organization
- Streaming datasets
- Power Query
- …

Experiment

Experiment concepts

- Controlling ML-runs - Code version, output files, parameters
- Compare runs to identify the best subset of parameters
- Use notebooks to execute runs

Model

Model concepts

- Algorithm to recognize certain types of patterns
- Apply this model from notebooks or the UI

Copilot

General available

- Power BI

Preview

- Data Engineering, Data Science, Data Factory, Data Warehouse, and Real-Time Intelligence

Considerations

- Data maybe processed outside your capacity's geographic region
- Capacity admins can also manage this setting
- F64 +
- Not available in all regions yet

Data Science and data engineering

- Code completion, automate routine tasks, contextual code suggestion, support on errors

Data factory

- Code generation and code explanations

Data warehouse

- Natural Language to SQL, code completion and quick actions

Real-time intelligence

- Natural Language to KQL

Power BI

- Generates reports, summary's, Q&A, write DAX

Assignment (2 or 3 persons)

- Take an environment that you can create or have created with Fabric.

Components

Ask yourself the following questions

- Which components are relevant to your environment?
- How are they connected?

Administration

Licensing concepts

Tenant

- One tenant with one OneLake

Capacity

- Your own distinct pool of resources
- Workspace belongs to one capacity
- You can have multiple capacities

SKU's

- Corresponds to a amount of compute power, measured in Capacity Units (CU's)

SKU*	Capacity Units (CU)	Power BI SKU	Power BI v-cores
F2	2	-	0.25
F4	4	-	0.5
F8	8	EM/A1	1
F16	16	EM2/A2	2
F32	32	EM3/A3	4
F64	64	P1/A4	8
Trial	64	-	8
F128	128	P2/A5	16
F256	256	P3/A6	32
F512	512	P4/A7	64
F1024	1024	P5/A8	128
F2048	2048	-	256

Feature differences

Feature	F SKU	P SKUs
Power BI Autoscale	✗	✓
Copilot	F64 or higher	✓
Managed Private Endpoints	All F SKUs or trial capacity	✗
Trusted workspace access	All F SKUs	✗
ARM APIs and Terraform	✓	✗
Pause and resume your capacity	✓	✗
On-demand resizing	✓	✗
Bring your own key for Power BI	✓	✓
Private links	✓	✓
View Power BI items without a per user license	F64 or higher	✓

Per-user license

- You need one F-license
- Each user needs one of these per-user license

Free

- Create, share and use non-Power BI-items only
- End-users: Use Power BI reports in a F64 capacity or higher
- Platform developers: Non Power BI-items

Pro

- Create, share and use all Fabric items
- End-users: Use Power BI reports in a F32 capacity or lower
- Power BI creators: Semantic models and reports
- Platform developers: All Fabric items

PPU

- Use all premium features in combination with a small F-license
- Use Power BI Premium-features for small businesses

Capabilities	Free	Pro	PPU
Access Microsoft Fabric web application	✓	✓	✓
Create Fabric capacity workspaces	✓	✓	✓
Create Power BI Premium workspaces	✗	✓	✓
Create Pro workspaces	✗	✓	✓
Create, update, delete or manage Power BI items in workspaces other than their "My Workspace"	✗	✓	✓
Create PPU workspaces	✗	✗	✓
Create non-Power BI Fabric items in Fabric / Trial capacity workspaces	✓	✓	✓
Create non-Power BI Fabric items in Power BI Premium capacity workspaces	✓	✓	✓
Share non-Power BI Fabric items	✓	✓	✓
View Power BI items in Pro workspaces or Fabric Capacity workspaces (where the Fabric Capacity SKU is less than a F64)	✗	✓	✓
Users signing in with their own account can view Power BI items in Power BI Premium Per Capacity or Fabric Capacity workspaces (where the Fabric capacity SKU is greater than or equal to a F64)	✓	✓	✓
Anyone can view Power BI items in Power BI Premium Per Capacity or Fabric Capacity workspaces when they're embedded using the embed for your customers solution	✓	✓	✓

Trial license

What you get

- A free Fabric capacity (F64) for 60 days
- 1 TB of OneLake storage
- The most features

What you don’t get

- Copilot
- Trusted workspaces
- Managed private endpoints

Considerations

- You have a limit of free licenses in one tenant
- Capacity admins can also manage this setting
- F64 +
- Not available in all regions yet

Pricing

Where do you pay for?

- Fabric Capacity
- OneLake Storage
- Per-user licenses
- Networking (future)

Fabric Capacity

Region: West Europe | Currency: Euro Zone – Euro (€) EUR | Display pricing by: Month

1 USD = 0.8955 EUR

SKU	Capacity unit (CU)	Pay-as-you-go	Reservation
F2	2	€287.647/month	€171.048/month ~41% savings
F4	4	€575.293/month	€342.095/month ~41% savings
F8	8	€1,150.585/month	€684.190/month ~41% savings
F16	16	€2,301.169/month	€1,368.379/month ~41% savings
F32	32	€4,602.338/month	€2,736.758/month ~41% savings
F64	64	€9,204.675/month	€5,473.515/month ~41% savings
F128	128	€18,409.350/month	€10,947.030/month ~41% savings
F256	256	€36,818.699/month	€21,894.059/month ~41% savings
F512	512	€73,637.398/month	€43,788.117/month ~41% savings
F1024	1024	€147,274.796/month	€87,576.233/month ~41% savings
F2048	2048	€294,549.591/month	€175,152.465/month ~41% savings

OneLake storage

Storage	Price
OneLake storage/month**	€0.0215 per GB
OneLake BCDR storage/month	€0.0412 per GB
OneLake cache/month*	€0.2329 per GB

Per-user license

Pro – 9,40 user/month

PPU – 18,70 user/month

Fabric admin monitoring workspace

Monitor your environment including

- Capacities
- Workspaces
- Items
- Activities
- Users

Measures

- Usage
- Performance

Fabric capacity metrics app

Monitor your capacities

- Charts display capacity usage over a 14-day period
- Helps you understand which workloads use more capacity

Essential for

- Informed decisions on how to use your capacity resources

Monitoring Hub

Monitor your activities

- List of all succeeded, failed and in progress activities
- Cancel (long-running) activities

Bursting, Smoothing and Throttling

Bursting and Smoothing

- Use more resource (CU's) than the capacity limit – Using future CU's
- Handling spikes
- Job classification
  - Interactive: Short-term spikes
  - Background: Long-running high demand

Bursting and Smoothing

- Regardless of SKU, Fabric *bursting* will automatically allocate resources as needed to execute at maximum performance
- As such, one query could consume all the quota of a single time window and much more!
- To avoid an overload, *smoothing* kicks in

Bursting and Smoothing

- No one (or few) queries can trigger an overload
- Instead of allowing runaway queries to create a local overload, Fabric smooths the queries reported usage to future time windows
- Kind of "Buy now, pay in the future" installment plan

Bursting and Smoothing

- Large batch processes traditionally were a threat to interactive queries as they would overload the compute resource
- DBAs traditionally had to carefully schedule these jobs to off-hours to avoid interference with interactive user experiences

Bursting and Smoothing

- A similar "installment plan" logic is applied for batch jobs
- But for batch jobs the smoothing is applied uniformly for the next 24 hours
- This completely liberates the DBA from any consideration of job scheduling. The load will be uniform regardless of the schedule.
- Most importantly, regardless of when batch jobs are scheduled, there will not be any degradation on interactive query performance

Bursting and Smoothing

- Even after smoothing, there are times when many jobs may build up and cause a capacity limit to be exceeded.
- When this happens, a capacity does not move into an overload state immediately. It "carries forward" usage above limits into future periods where there is available capacity.

Bursting and Smoothing

- Carrying usage above limits into future periods <u>in addition to smoothing</u> further lessens the impacts of compute spikes.
- When thinking of capacity load there are two dimensions
  - The percent (%) consumed
  - The amount of future periods filled

Bursting, Smoothing and Throttling

Bursting and Smoothing

- Use more resource (CU's) than the capacity limit – Using future CU's
- Handling spikes
- Job classification
  - Interactive: Short-term spikes
  - Background: Long-running high demand

Throttling

- Compensating smoothing
- Degraded user performance

Usage	Policy Limits	Platform Policy Experience Impact
Usage <= 10 minutes	Overage protection	Jobs can consume 10 minutes of future capacity use without throttling.
10 minutes < Usage <= 60 minutes	Interactive Delay	User-requested interactive jobs are delayed 20 seconds at submission.
60 minutes < Usage <= 24 hours	Interactive Rejection	User-requested interactive jobs are rejected.
Usage > 24 hours	Background Rejection	All requests are rejected.

Domains

Logical groups of Fabric items

Organize and manage data according to your business

- Group Fabric item into logical groups
- Assign workspaces to groups, bulk assignment and one at the time
- Control by using new roles: Domain admin and contributor
- Managed within the admin portal

Workspace best-practice

There's no one way, it depends on…

- Permission
  - Who works on which items?
  - Workspace contributor (and higher) can edit all items within the workspace. You don't want to give a report builder edit permissions to a data pipeline.
- Data security
  - Who is allowed to see which data?
  - Workspace contributor (and higher) has access to the underlying data. Data security like RLS doesn't apply.
- Development process
  - Which environments do you need (dev, test, acc, and prod)?
  - You might have different users for each environment.
- Deployment process
  - Which items should be deployed together?
  - You might have different deployment policies and/or frequency.

Yes, you can put everything in 1 workspace, but you don't want to!

OneLake Shortcuts

Extend your lake with virtual data

What is it?

- Artifacts that point to other storage locations (folder), so you can add data from other sources
- It's virtual, so it doesn't copy data

Shortcut types

- Internal
  - Link to other Fabric items like Lakehouse, KQL database and Warehouse
  - You can cross item-types and workspaces
- External
  - Link to ADLS Gen2, Amazon S3, Google Cloud Storage, and Dataverse
  - Caching occurs for Amazon S3 and Google Cloud Storage

OneLake Shortcuts

Shortcuts allow you to create a virtualized data lake, eliminating copies of data between organization domains, analytical engines, or clouds

Internal

From a lakehouse folder to a lakehouse folder

From a lakehouse Table to a lakehouse Table

From a Warehouse Table to a lakehouse Table

From a KQL Database Table to a lakehouse Folder

Two Types Of Shortcuts

External

From an ADLS Gen2/S3 folder to a lakehouse Folder

From an ADLS Gen2/S3 folder to a Table

S3 Shortcuts are read-only

Fabric admin monitoring workspace

Fabric capacity app

Domains

Shortcuts

Administration

Ask yourself the following questions

- Which licence(s) are relevant to you?
- Do you have external resources you want to access via shortcuts?
- How would you set up domains and workspaces?
- What kind of user groups can you define?

Security

Manage Fabric Security

Data handling

Data **at Rest** Encryption

Data at Rest

- Encrypted by default using Microsoft-managed keys
- Customer managed keys (CMK) is on the roadmap
- Workaround: Use ADLS Gen2 and shortcuts
- For Power BI you still have BYOK

Data **in Transit** Encryption

Data in Transit

- Across the public internet between Microsoft services: Encrypted by minimum Transport Layer Security (TLS) 1.2.
- Inbound: Encrypted by minimum Transport Layer Security (TLS) 1.2.
- Outbound: Enforces TLS 1.2, but can be configured to use TLS 1.0 if the external infrastructure doesn’t support newer protocols

Secure inbound traffic

1. Private links

- Uses a private ip-address
- Requests follows the configured network path (VNet)
- When you want to block all access from the public internet

Private link considerations

- You need a VNet data gateway for all data behind the firewall.
- Warehouse: No visual query and no copy data from and into a warehouse.
- Power BI: No live connection to a semantic model and dataflow. No direct lake mode, no subscriptions and export to PDF.
- Eventhouse: No ingesting data from OneLake, no connection from data pipelines, no shortcut.
- F64 and higher

Secure inbound traffic

2. Entra conditional access

- Access through the public internet
- Microsoft Entra handles the authentication
- Configure conditional access based on user, groups, device, IP location, application

Applied policy examples

- Requiring multifactor authentication for users with administrative roles
- Blocking or granting access from specific locations
- Requiring organization-managed devices for specific applications

Secure outbound traffic

Trusted workspace access

1. The ADLS Gen 2 account has network access enabled from selected VNets and IP addresses. This means it is closed off from the public internet.
2. A resource instance rule is set up in the Azure portal to allow access from specific Fabric workspaces.
3. Only the workspace configured in Fabric to have access to the ADLS Gen 2 account will have access.
4. The workspace can now create a shortcut to the ADLS account or you can create a data pipeline to access the data directly in Fabric.

Secure outbound traffic

Managed VNets

- Network isolation for spark workloads
- The managed VNet is provisioned automatically when the job starts
- No use of starter pools (pre-warmed clusters in a shared hosted VNet)

Enable managed VNet by adding managed private endpoint to the workspace

Also when you enable private link for inbound traffic

Data gateway

Why you need one?

- Your data source is located on-premises
- Your data source is located in a private network (like an Azure VNet)
- You require a host for connector software (custom data connectors)

Which type of data gateway do you need?

- Personal gateway
- On-premises data gateway (standard mode)
- Virtual network (VNet) gateway

microsoft.com

Data gateway

On-premises data gateway (standard mode)

1. The gateway transfers data from on-premises data source to cloud services.
2. The gateway transfers data from cloud data source to cloud services in specific scenarios like custom data connectors.
3. The data gateway is installed on an always-on VM.

4/5/6. Users can connect to the data sources behind the gateway

microsoft.com

Data gateway

Virtual network (VNet) gateway

1. The gateway transfers data from data source in a private network to cloud services.
2. The VNet data gateway is a Microsoft managed service. You centrally manage the VNet data gateway from the Azure portal and the Power Platform admin portal.

3/4/5. Users can connect to the data sources behind the gateway

Typically used in in conjunction with private endpoints for Azure data sources to ensure that no traffic is ever exposed to a public endpoint.

microsoft.com

Data gateway

Why you need one?

- Your data source is located on-premises
- Your data source is located in a private network (like an Azure VNet)
- You require a host for connector software (custom data connectors)

Which type of data gateway do you need?

- Personal gateway
- On-premises data gateway (standard mode)
- Virtual network (VNet) gateway

Other options to consider for on-premises data gateway:

- Gateway cluster for better availability and performance
- DTAP environments to split workloads
- Consider the location of your gateway. Ensure residency requirements

microsoft.com

Fabric data security

Multiple layers of security and access control

Platform level access. Authentication through Microsoft Entra ID. Optional private link or conditional access.

Workspace level access. Managing artifacts and data access through these artifacts.

Compute security. User specific security applied during compute. e.g. RLS, OLS on a SQL analytics endpoint, warehouse), and semantic model

Item level access. e.g. ReadAll access on a Lakehouse or read access to only a subset of data within a Lakehouse (schemas, tables, rows, columns).

But what about OneLake level security?

Workspace level access

Viewer

- Can view all content in the workspace

Contributor

- Can also modify content in the workspace

Member

- Can also share content in the workspace

Admin

- Can also manage permission and delete the workspace

Access through apps

Preferred way to share reports within your organization

- One place to develop reports, with the option to share with multiple user groups
- No need to give permissions on the workspace
- Users get read access on data

Compute security

SQL analytics endpoint and Warehouse

Column-level security

- Restrict access to certain columns of a table, e.g. PII data
- This only applies to queries on a Warehouse or SQL analytics endpoint
- Use T-SQL to grant or deny permissions to roles

Row-level security

- Restrict access to rows, e.g. only their department
- This only applies to queries on a Warehouse or SQL analytics endpoint
- Use T-SQL to grant or deny permissions to roles. This is similar to RLS in SQL Server

Dynamic data masking

- Limit sensitive data exposure
- Change sensitive data to untraceable values
- Use T-SQL to apply masking.

Compute security

Semantic model

Object-level security

- Restrict access to certain columns or tables, e.g. PII data
- This only applies to queries on a semantic model
- Use tabular editor to implement OLS

Row-level security

- Restrict access to rows, e.g. only their department
- This only applies to queries on a semantic model
- Use Power BI desktop and DAX to implement RLS

Item level access

Lakehouse

No additional permissions (Read)

- ⊘ SQL Analytics Endpoint
- ⊘ Lakehouse
- ⊘ Default semantic model

User can connect, but can't read any data

No additional permissions (Read) and GRANT-permissions

- ✓ SQL Analytics Endpoint
- ⊘ Lakehouse
- ⊘ Default semantic model

User can read tables for which he/she received GRANT-permissions.

ReadAll SQL Analytics Endpoint

- ✓ SQL Analytics Endpoint
- ⊘ Lakehouse
- ⊘ Default semantic model

User can read all tables with the SQL Analytics Endpoint.

ReadAll Apache Spark

- ⊘ SQL Analytics Endpoint
- ✓ Lakehouse
- ⊘ Default semantic model

User can read all files with Spark.

Build reports on the semantic model

- ⊘ SQL Analytics Endpoint
- ⊘ Lakehouse
- ✓ Default semantic model

User can create Power BI reports on the default semantic model.

Item level access

Warehouse

No additional permissions (Read)

- ✗ SQL Analytics Endpoint
- ✗ Lakehouse
- ✗ Default semantic model

User can connect, but can't read any data

No additional permissions (Read) and GRANT-permissions

- ✓ SQL Analytics Endpoint
- ✗ Lakehouse
- ✗ Default semantic model

User can read tables for which he/she received GRANT-permissions.

Read all data using SQL (ReadData)

- ✓ SQL Analytics Endpoint
- ✗ Lakehouse
- ✗ Default semantic model

User can read all tables with SQL.

Read all OneLake (ReadAll)

- ✗ SQL Analytics Endpoint
- ✓ Lakehouse
- ✗ Default semantic model

User can read all files with Spark.

Build reports on the semantic model (Build)

- ✗ SQL Analytics Endpoint
- ✗ Lakehouse
- ✓ Default semantic model

User can create Power BI reports on the default semantic model.

OneLake data access role

- Grant access to specific folders
- Lakehouse items only
- Restricts access to user with read permission only

Data access role security only applies to users accessing OneLake directly!

OneLake data access role

1. Create a new role

2. Assign people to this new role

- Directly: Based on Entra-group or Entra-ID
- Virtual: Based on Lakehouse permissions

OneLake data access role

How is OneLake access determined?

Shortcut security

- You need access on the shortcut location and on the target location
- The most restrictive permission of the 2 locations is applied
- Apply permissions to the shortcut location like any folder in OneLake, like Lakehouse sharing, OneLake data access roles etc.
- For external authorization (including ALDS Gen2), one credential is used to secure the target location

Build permissions

Usage

- Create reports
- Analyze in Excel
- Export underlying data
- Usage XMLA-endpoint

How to get build permission

- Workspace role
- App audience
- (Semantic model) Item permission
- Share report
- Request permission

Collaborate with external users

Setup

- Same as before with Power BI
- Create guest user in your Entra
- Guest user can get the same permissions as internal users

Keep in mind

- Guest users need to use Power BI service, so they can't
  - Publish to Power BI
  - Connect to semantic model in Power BI service with Power BI desktop or Excel

External data sharing

- Share folders with external users
- Data is not copied
- Read-only access to the data
- Other security policies should be defined in consumer tenant
- No reshare within consumer tenant
- Provider can’t control access in other tenant
- Lakehouse and KQL database only
- External user accepts the share
- Data might cross geographic boundaries

Item level access + Column-level security for SQL analytics endpoint and Warehouse

Create a OneLake RBAC role and assign people

Row-level security for semantic model

Security

Ask yourself the following questions

- Why is the basic Fabric sufficient or do you need more inbound and outbound security?
- Are you going to set up security at item level? And on which one?
- Do you need OLS/RLS?
- Would you also set up OneLake RBAC? For whom and why?
- What workspace roles would you give to which groups?
- Why do/don’t you need a data gateway?

Data Governance

Data Governance

Sensitivity labels

Classify and protect your data

Different ways of labelling

- Manual, default, and programmatic

Inheritance

- Downstream and based on other item with label during creation

Export

- Excel, Excel with live connection to Power BI dataset, PowerPoint and Power BI Desktop

Limitations

- Never overwrites manually applied labels
- Inheritance of data sources only works for semantic models with import mode

Name	Type	Sensitivity
Datamodel Sales	Report	General
Datamodel Sales	Semantic model	General
Fanshop Sales Data	Dataflow	—
Reporting Sales	Report	Confidential

Endorsement and discovery

Identify your high quality and trustworthy data

Promotion

- Promote your fabric item

Certification

- This fabric item meets company standards

Name	Type	Endorsement
Datamodel Sales	Report	—
Datamodel Sales	Semantic model	Certified
Fanshop Sales Data	Dataflow	—
Reporting Sales	Report	Promoted

Lineage

Understand and document the flow of your data

Find related Fabric items

- All items within the workspace

Limitations

- Only one level up outside the workspace

Impact Analysis

Identify potential impact when making changes

Find related Fabric items

- All downstream items

Notify relevant people

- Workspace contact list

Limitations

- Only direct children for data sources
- You don’t see names
- Not on dataflows

Tags

Enhance discoverability

Categorize your items

- Tags are created by the Fabric admin
- Each item can have multiple tags

Limitations

- Up to 10 tags for each item
- Filter on tags only available in workspace

Metadata scanning

Catalog all metadata of your analysis and report items

Extract info about your Fabric items

- Workspace, report, dataset, lakehouse, pipeline, governance, measures and more…

Make it more powerful

- Combine this with the Power BI activity log

Limitations

- No UI, just a set of api's


```
"Notebook": [
  {
    "id": "8f1f8d5f-6df9-40a1-8f04-e241321cac0e",
    "name": "Load_dm_customer",
    "description": "Business logic for customer",
    "state": "Active",
    "lastUpdatedDate": "2023-12-22T17:20:24.8291782",
    "createdDate": "2023-12-20T19:59:40.5861766",
    "modifiedBy": "hpeek@peekdata.nl",
    "createdBy": "hpeek@peekdata.nl",
    "modifiedById": "f9e59873-feed-451d-abf4-5337d4c9ca4a",
    "createdById": "f9e59873-feed-451d-abf4-5337d4c9ca4a",
    "sensitivityLabel": { "labelId": "3491b186-cd90-41bd-863e-6dc99599faac" },
    "relations": [
      {
        "dependentOnArtifactId": "5e51f72e-4fe3-48b6-a98e-d4b08b64431d",
        "workspaceId": "a9374840-2398-4da6-95ec-6881daf2bedd",
        "relationType": "Datasource",
        "settingsList": "None",
        "usage": "Datasource"
      }
    ],
    "extendedProperties": {}
  },
```

Purview Hub

Create an overview of all items in your data estate

Catalog all your Fabric items

- Workspace, report, dataset, lakehouse, pipeline and more…

Metadata of your items

- Creator, sensitivity and endorsement

Admin view and non-admin view

- All items vs only owned semantic datamodels

Limitations

- No activity insights
- No data access information
- To create a really good overview, you need more

Sensitivity labels

Purview Hub

Data governance

Ask yourself the following questions

- What is currently set up regarding data governance? Think about both strategy, process, people trained and technology?
- What in Fabric can help you with this and how?

Development

GIT-integration

Why CI/CD?

- Central source location
- Version control
- Track and merge changes
- Automated deployments
- Scalability and disaster recovery

Not all items can use Git.

Deployment Pipelines

- Automated deployments
- Friendly UI
- Selective deployment
- Limited compared to Azure DevOps

Deployment Pipelines

External development tools

PBI Tools

Git-integration

Deployment pipelines

Development

Ask yourself the following questions

- Do the capabilities in Git with Fabric suit the current way of working?
- If not, could you find a way in this to do work with Git in Fabric? Or what else do you need?

Govern your Fabric environment

github.com/hylkepeek

Hylke Peek