One Lake

Deep Dive

(or at least a little deeper!)

Martin Catherall

Data Analytics Consultant
ADAICO.ai

@MartyCatherall

www.MartinCatherall.com

- Data Architect

- SQL Server
  - Since 2002
- Microsoft Azure
  - Since 2015
- Microsoft Fabric
  - Since 2022

MVP Microsoft Most Valuable Professional

Microsoft Certified: Azure Data Fundamentals

Microsoft Certified: Power BI Data Analyst Associate

Microsoft Certified: Azure AI Engineer Associate

Microsoft Certified: Azure Data Engineer Associate

Microsoft Certified: Azure Solutions Architect Expert

Agenda and content

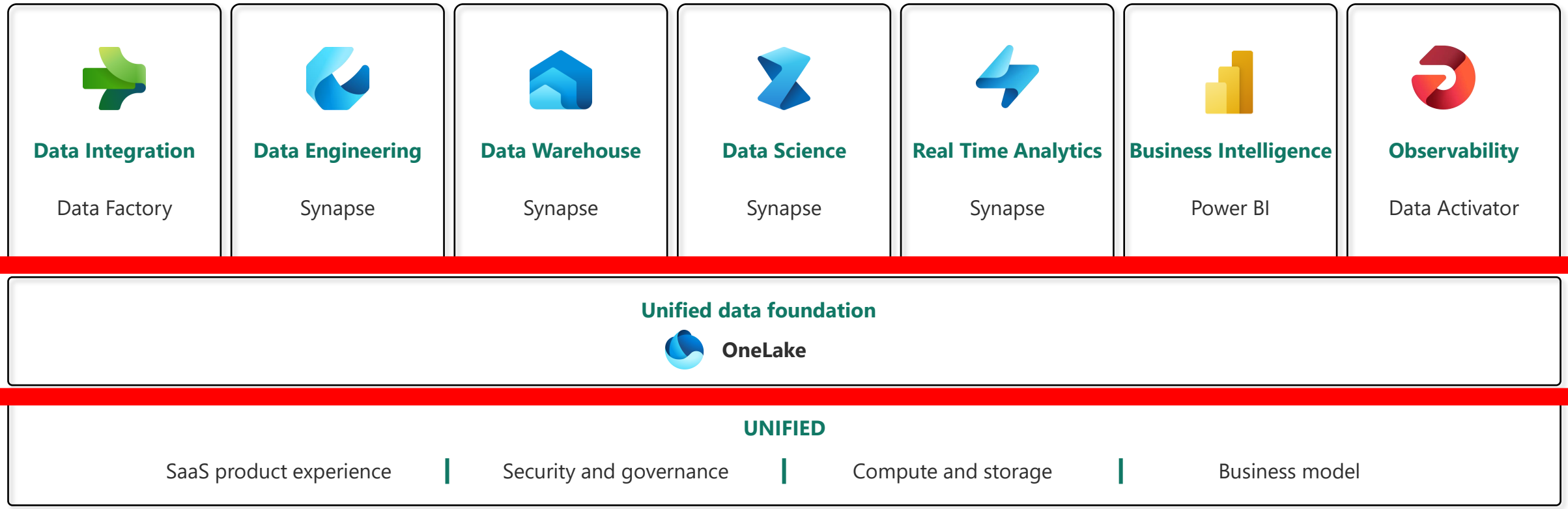
- OneLake – short recap
  - Real Time Intelligence
- Diving into the Lake (hopefully not the swamp)
- Going between lakehouses
- DELTA PARQUET
- OPTIMIZE
- VACUUM
- ?????
- Direct Lake and Power BI (DAX)

Resources

- Data Engineering documentation in Microsoft Fabric

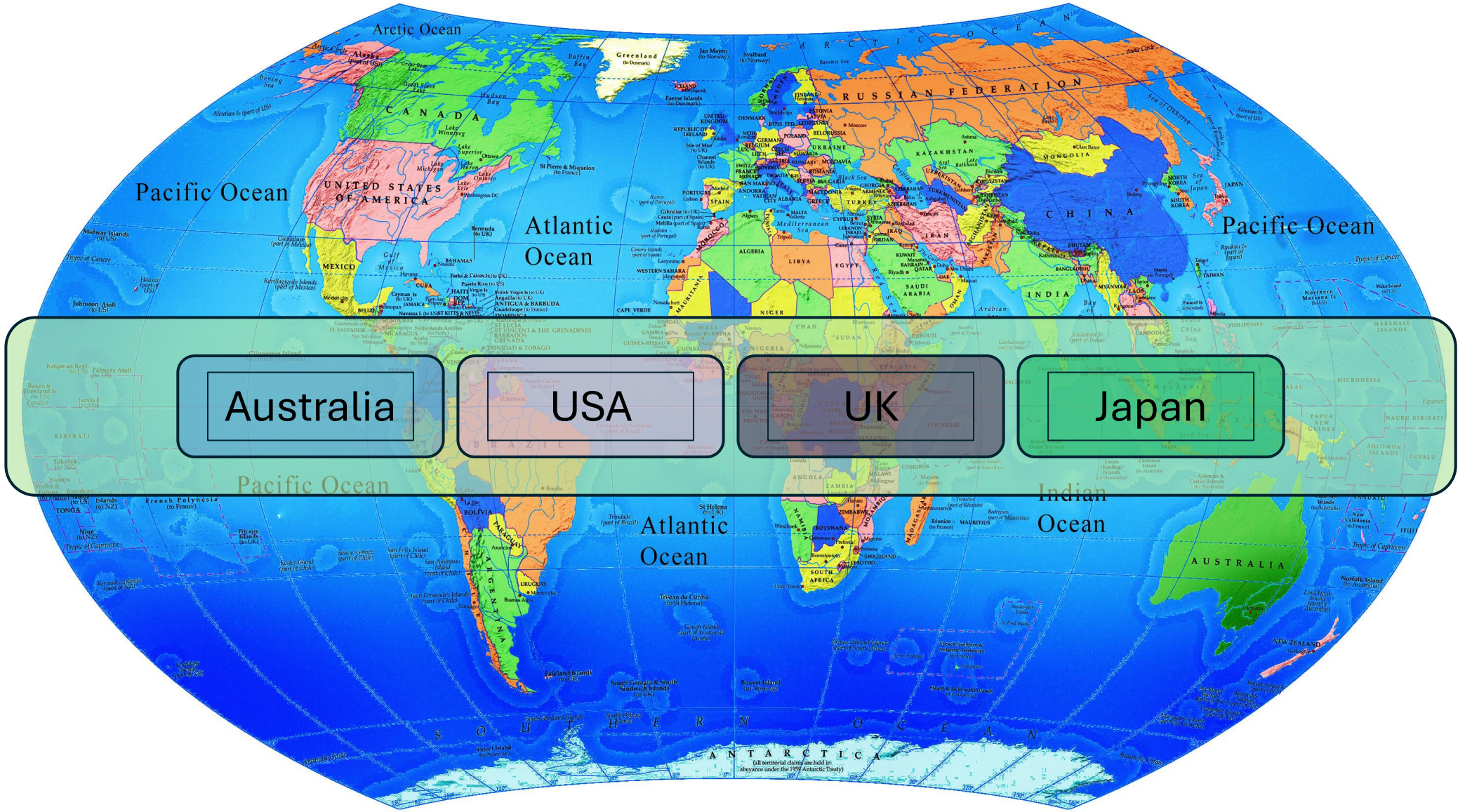
Microsoft Fabric does it all—in a unified solution

An end-to-end analytics platform that brings together all the data and analytics tools that organizations need to go from the data lake to the business user



OneLake – How many do I have?

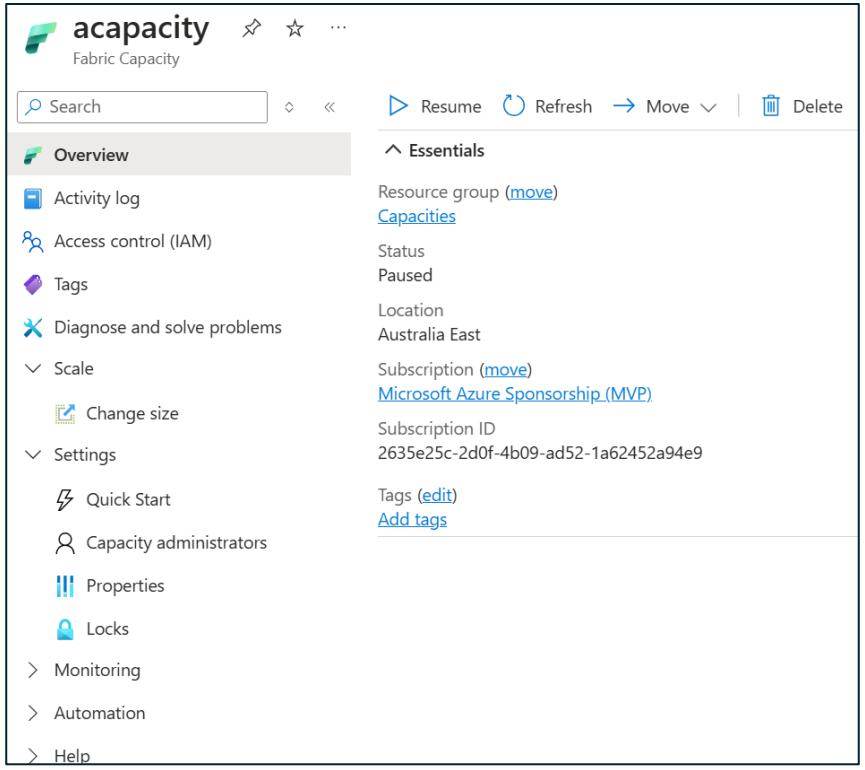
OneLake – How many do I have?



OneLake – How many do I have?

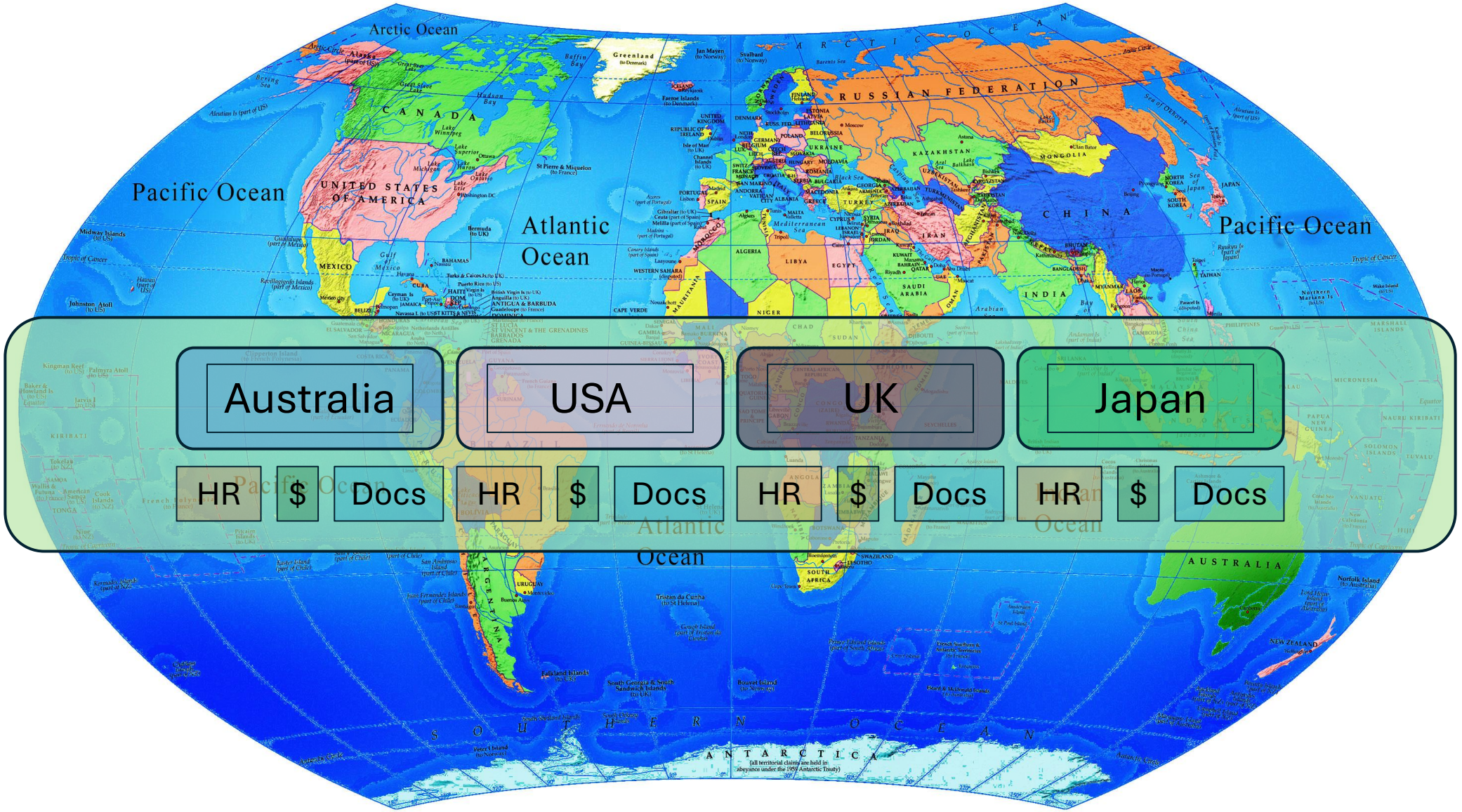
Use a Fabric Capacity

Set up in the Azure Portal



SKU	Capacity Units
F2	2
F4	4
F8	8
F16	16
F32	32
F64	64
F128	128
F256	256
F512	512
F1024	1024
F2048	2048

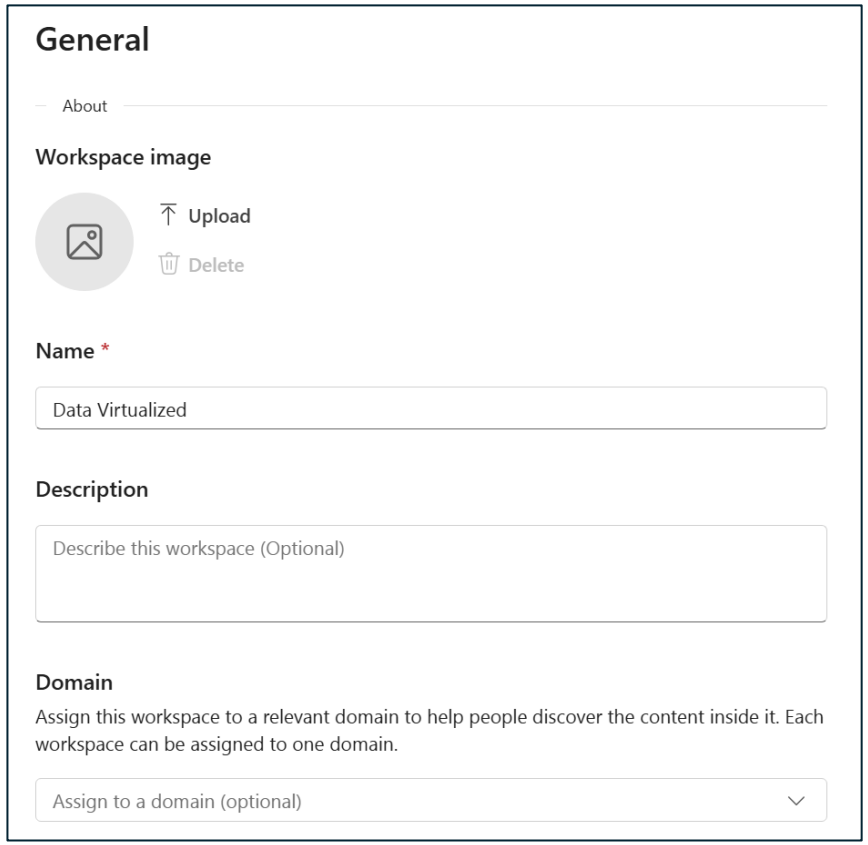
OneLake – How to divide this up further?

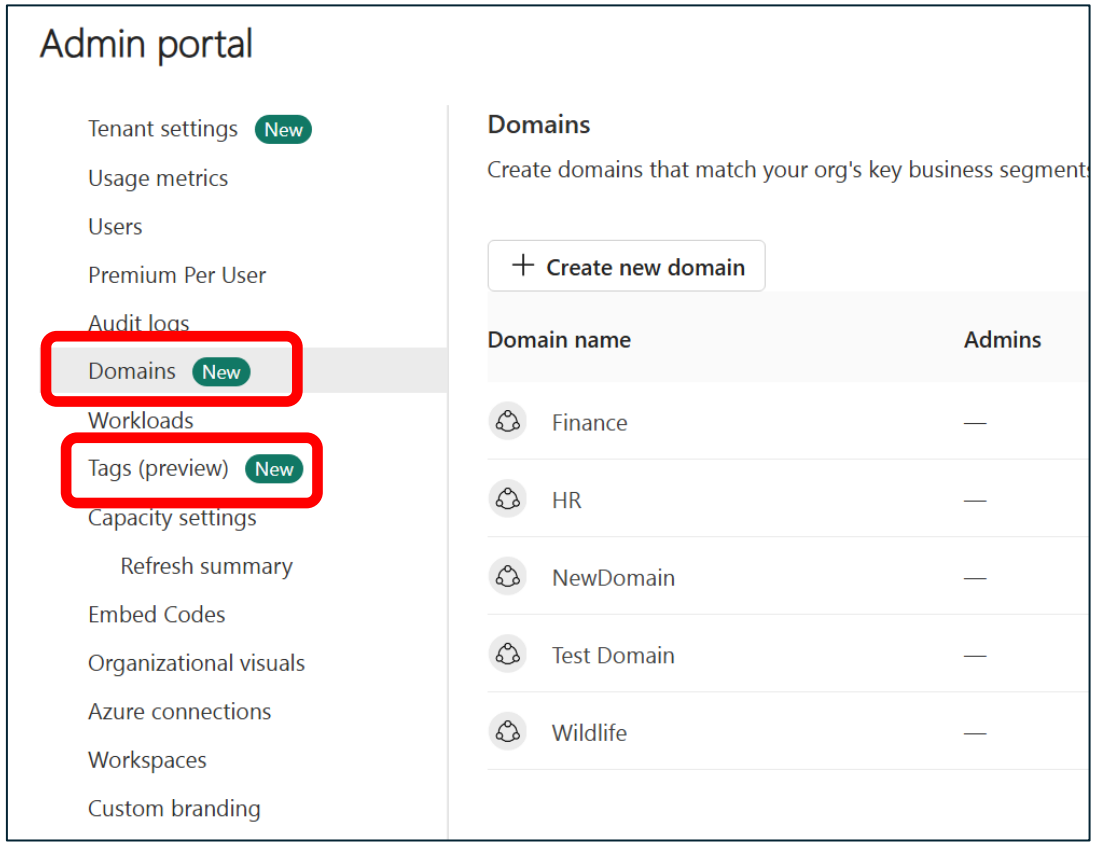


OneLake – How many do I have?

Use further Fabric Capacities

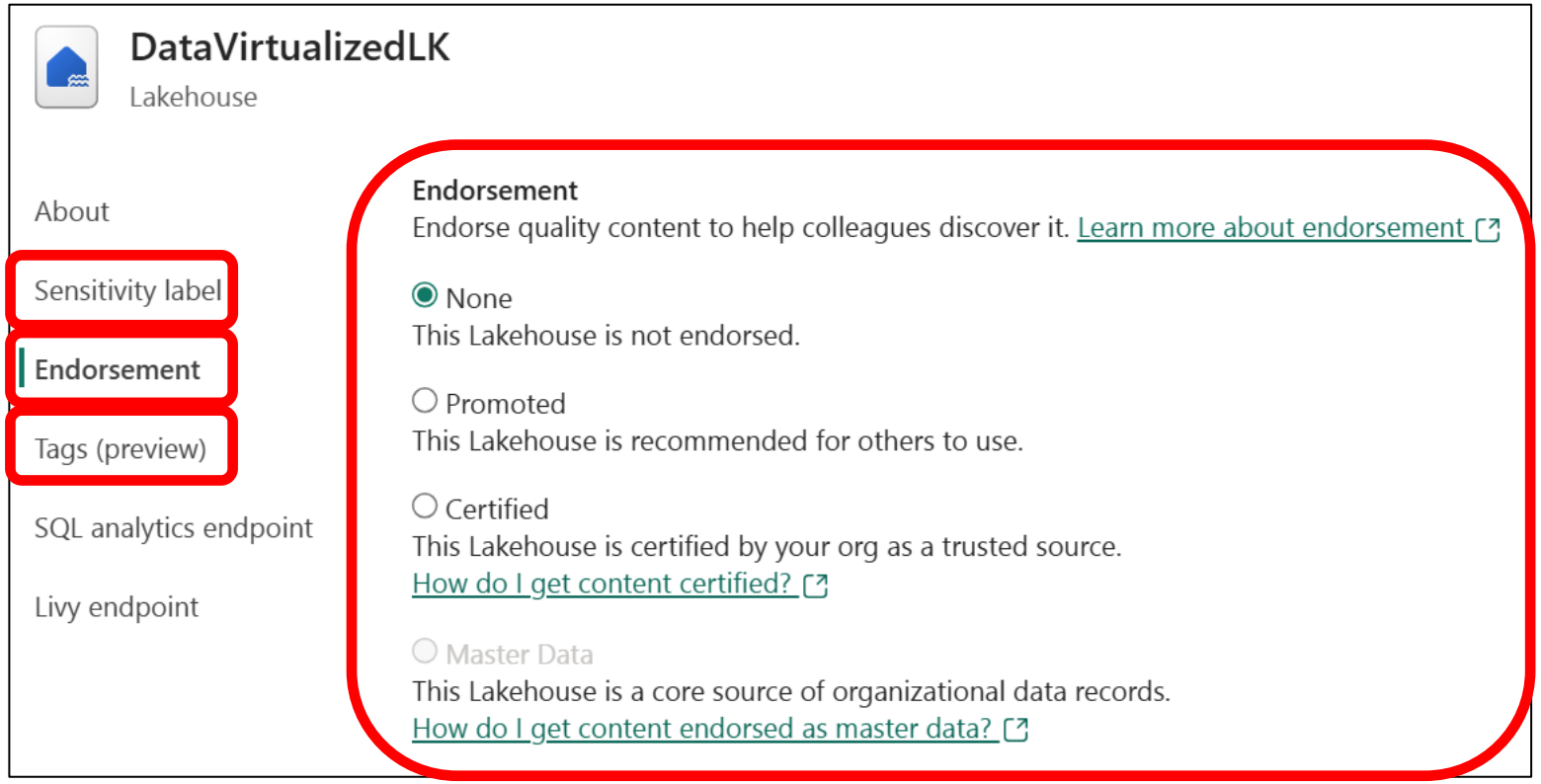
Use Domains



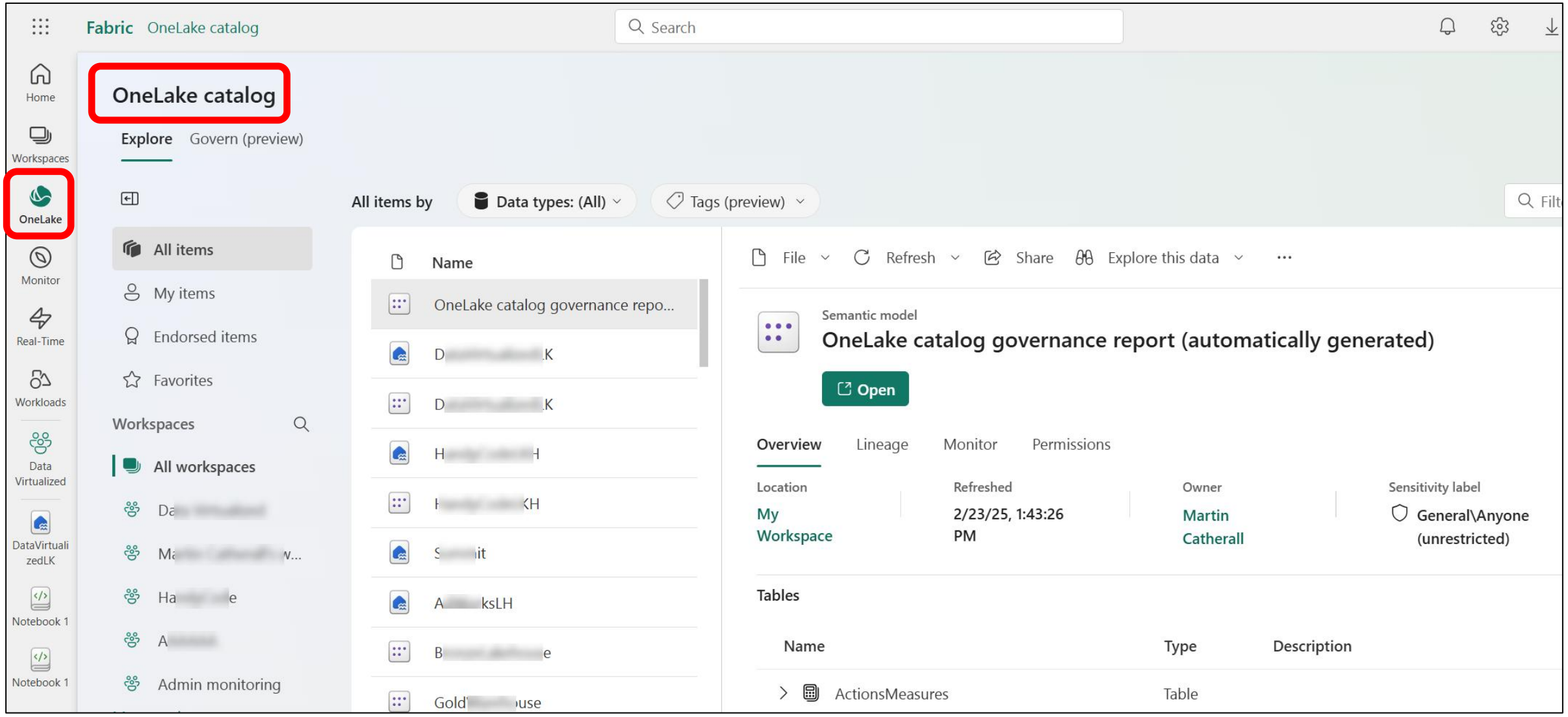


Other feature to aid your Data Governance

- Endorsements
  - None
  - Promoted
  - Certified
  - Master Data
- Sensitivity Labels
- Tagging
- Task Flows
- OneLake catalog governance report (automatically generated)



OneLake Catalog



OneLake Data

- The Lakehouse
  - Files (any type)
  - Structured Data
    - Tables – Parquet files (delta)
- The Warehouse
  - Structured Data
    - Tables – Parquet files (delta)

Demo

Accessing data and using shortcuts

Why Parquet?

- Data is stored in columns (not rows)
- Can be accessed for most languages
- Has great compression
- Is an Open Standard (no vendor lock in)
- Good data type support
    - And the ***delta*** adds more

Demo

Let’s see those parquet files

OneLake file explorer

OneLake Data – Accessing data

Optimize

OneLake Data – Maintenance of Delta

Vacuum

Sukrija
Dziekuje
Danke
Obrigado
THANK YOU
Hvala
Thanks
Grazie
Gracias
Dank U
Tak
Dankeschön
Merci
Mahalo
Gratias Tibi
Terima Kasih
Dakujem
Kitos
Köszönöm