

# Enterprise Scale Analytics - What's in it for me?

Tillmann Eitelberg , oh22

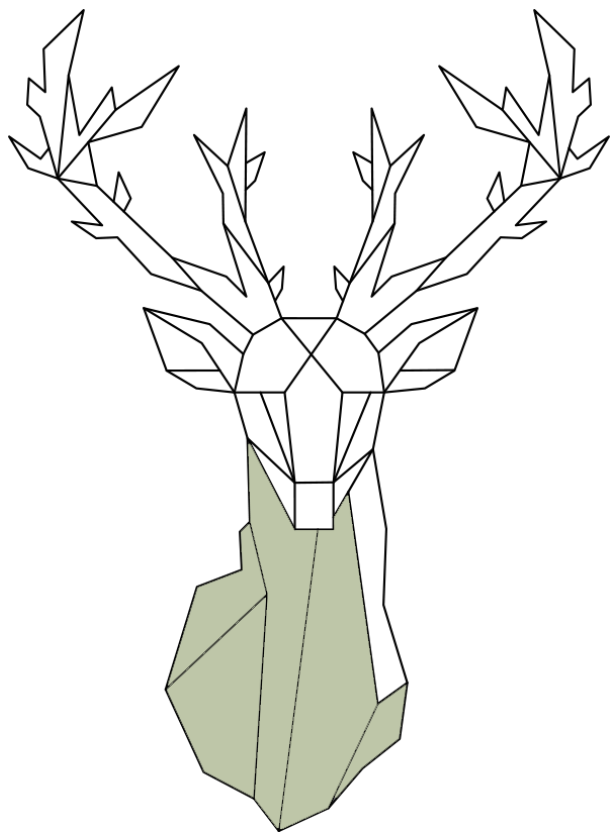
Oliver Engels, oh22

for



Lingen 2022





**WE CREATE  
SMART AND CASUAL  
DATADESIGN.  
EVERY DAY.**

# About Us



## Tillmann Eitelberg

CEO  
oh22information services GmbH

[t.eitelberg@oh22.net](mailto:t.eitelberg@oh22.net)

Kellerstr. 3  
53772 Königswinter  
Deutschland

➤ [www.oh22.is](http://www.oh22.is)



Microsoft  
Partner



Gold Cloud Platform  
Gold Data Analytics  
Gold Data Platform  
Gold Application Development  
Gold Datacenter  
Gold Application Integration  
Silver Cloud Productivity  
Silver Enterprise Resource Planning



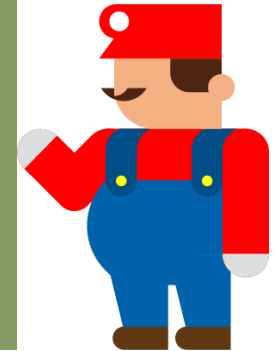
## Oliver Engels

CEO  
oh22data AG

[o.engels@oh22.net](mailto:o.engels@oh22.net)

Otto-Hahn-Str. 22  
65520 Bad Camberg  
Deutschland

➤ [www.oh22.net](http://www.oh22.net)

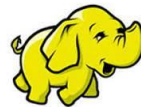


# Timetravel

DWH



BigData



Cloud



Lake  
House



1980's

1990's

2000's

2010's

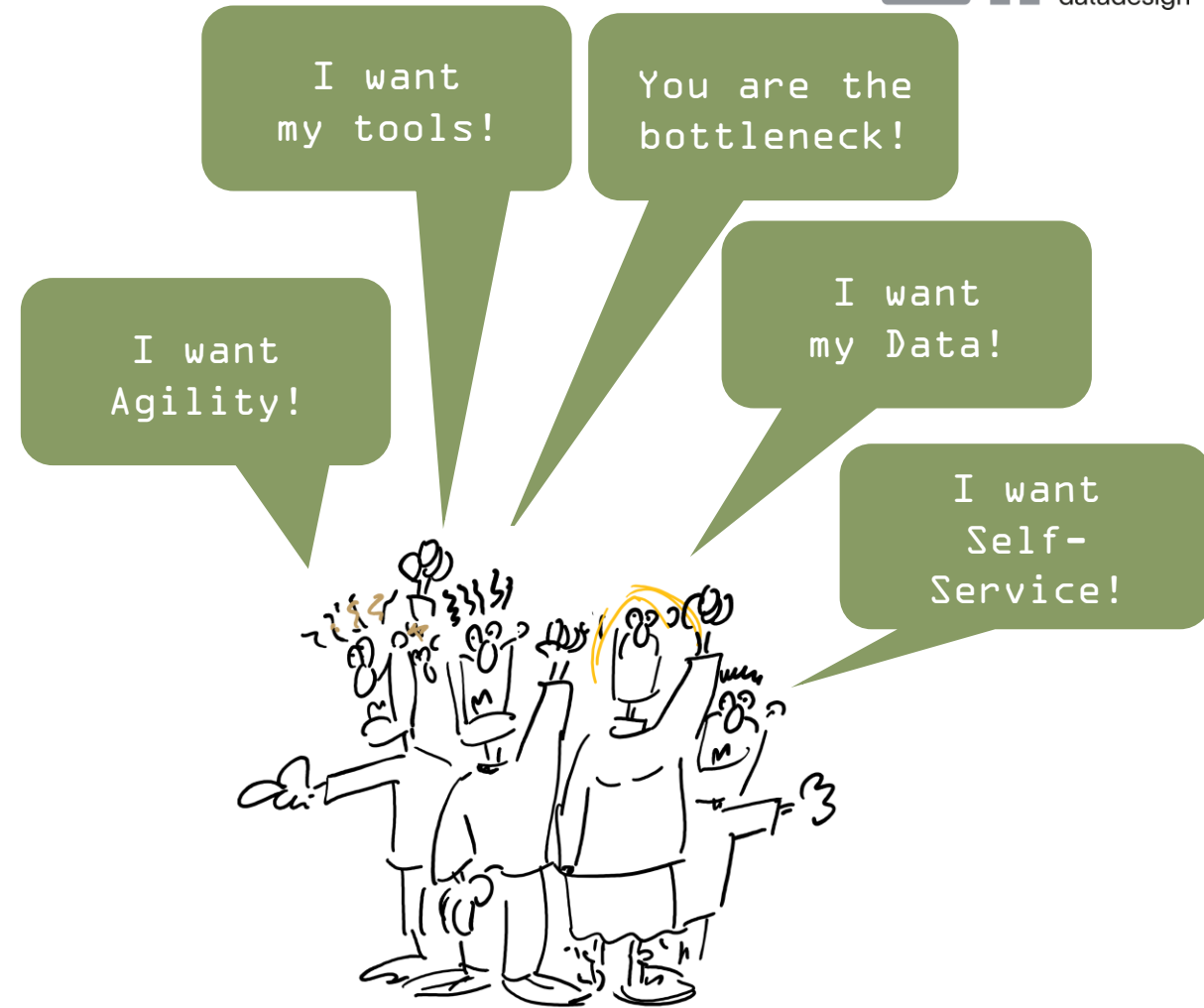
2020's

## Still the same challenges

- + Centralized
  - + Infrastructure centric
- + Ownership
  - + Who owns the data?
- + Quality
  - + Who is responsible?
- + Scaling
  - + Tech: yes! Org: no!



Central  
IT



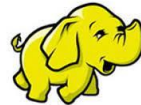
Projects  
& Users

# Timetravel

DWH



BigData



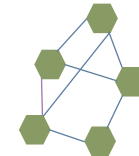
Cloud



Lake  
House



Data  
Mesh



1980's

1990's

2000's

2010's

2020's

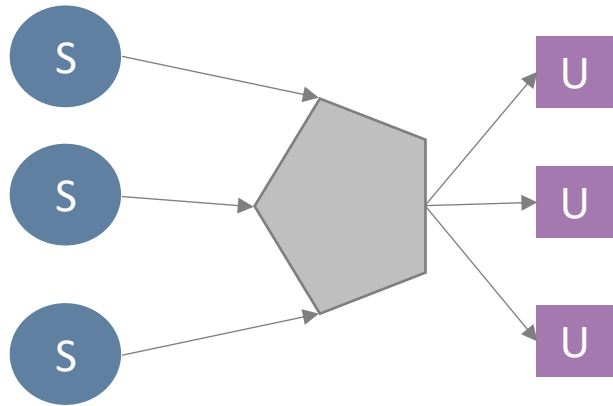
...

## Data Mesh

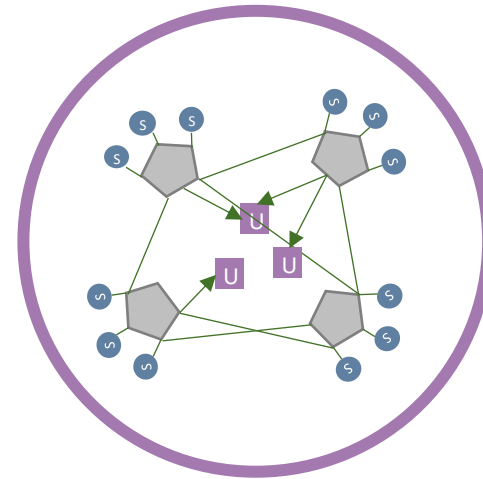
- + Intentionally designed **distributed** data architecture
- + **Centralized governance** and standardization
- + Shared and harmonized infrastructure
- + No fragmented silos, no inaccessible data
- + **Lakes and DWH** are simply **nodes** of the mesh
- + **Data Product** more important as tooling and pipeline



# Monolithic to Data Mesh



Centralized

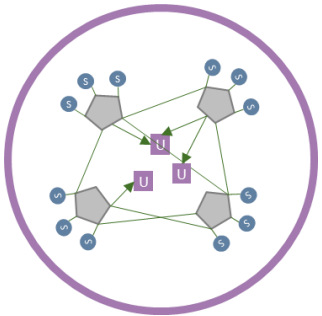


Distributed



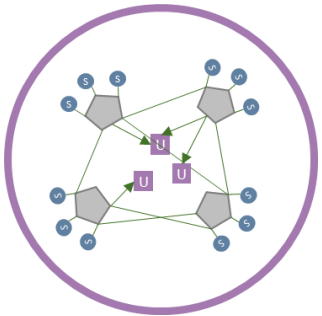


## Data Mesh



- + Multiple groups working with their own analytics platform  
→ nodes (**data domains**) in a mesh
- + All nodes are controlled by common policies and standards and based on templates / blueprints  
→ **data infrastructure as platform**
- + A central hub is providing **global governance**, logging, cost management and interoperability
- + All on an **@scale** cloud platform providing automation, security and availability

# Data Mesh: What do we need technically?



- + Some requirements
  - + Scalable polyglot big data storage
  - + Encryption for data at rest and in motion
  - + Data product versioning
  - + Data product schema
  - + Data product de-identification
  - + Unified data access control and logging
  - + Data pipeline implementation and orchestration
  - + Data product discovery, catalog registration and publishing
  - + Data governance and standardization
  - + Data product lineage
  - + Data product monitoring/alerting/log
  - + Data product quality metrics (collection and sharing)
  - + In memory data caching
  - + Federated identity management
  - + Compute and data locality

## Data Mesh with Microsoft?



- + For sure!
  - + Azure as Hypervisor
  - + Infrastructure as code
  - + Data Platform
    - + Storage
    - + Pipelines & Transformation
    - + Analytics
    - + Security
  - + Governance tools (?)
  - + Virtualization tools (?)



## Data Mesh with Microsoft



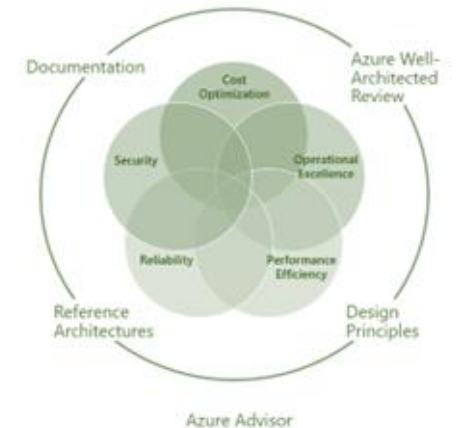
### + How can Microsoft help us to create a Data Mesh?

#### + Cloud Adoption Framework

The Cloud Adoption Framework is a collection of documentation, implementation guidance, best practices, and tools that are proven guidance from Microsoft designed to accelerate your cloud adoption journey

#### + Azure Well-Architected Framework

The Azure Well-Architected Framework architecture contains principles to design and optimize workloads running in your data estate



# Data Mesh with Microsoft

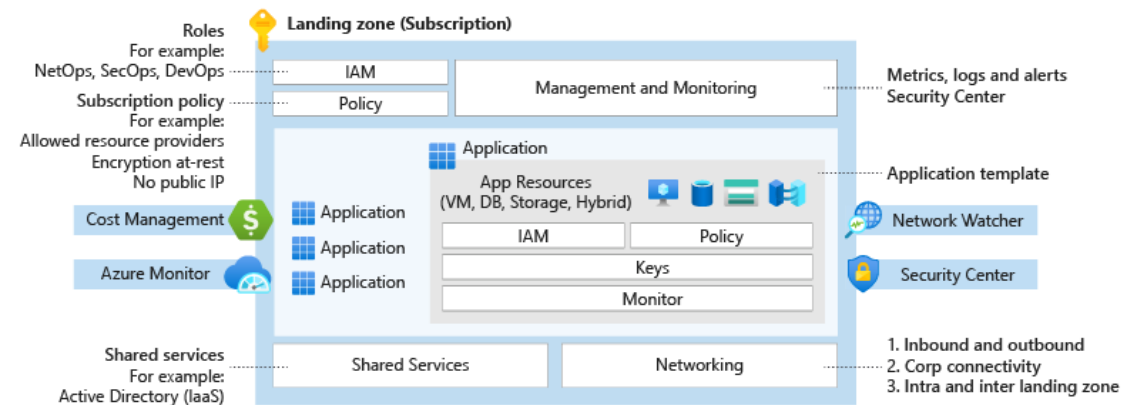


## + How can Microsoft help us to create a Data Mesh?

### + Azure Landing Zone

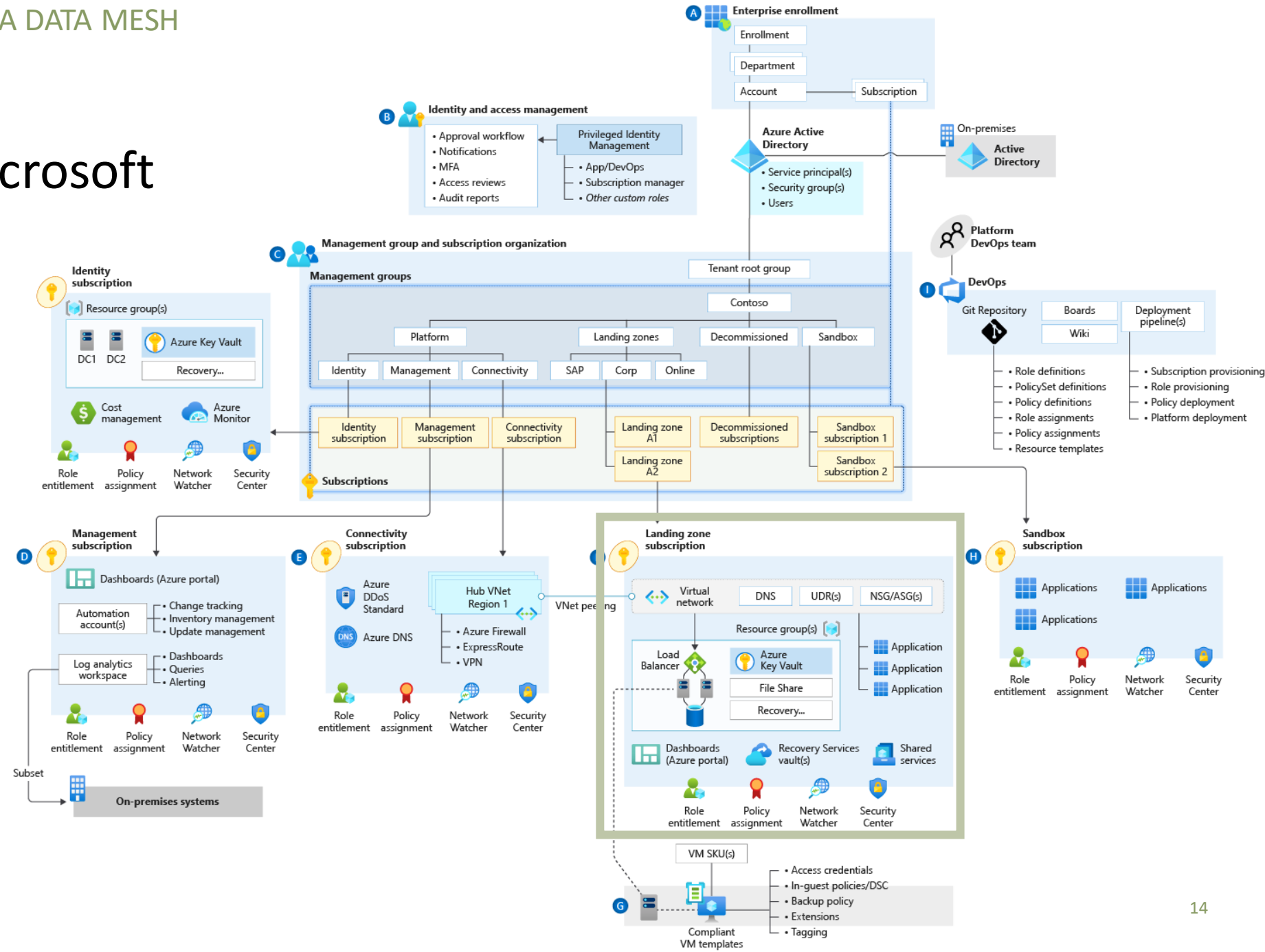
Azure landing zones are the output of a multisubscription Azure environment that accounts for scale, security governance, networking, and identity

Azure landing zones enable application migration, modernization, and innovation at enterprise-scale in Azure.



# Data Mesh with Microsoft

## Azure Landing Zone Conceptional Architecture



## Data Mesh with Microsoft



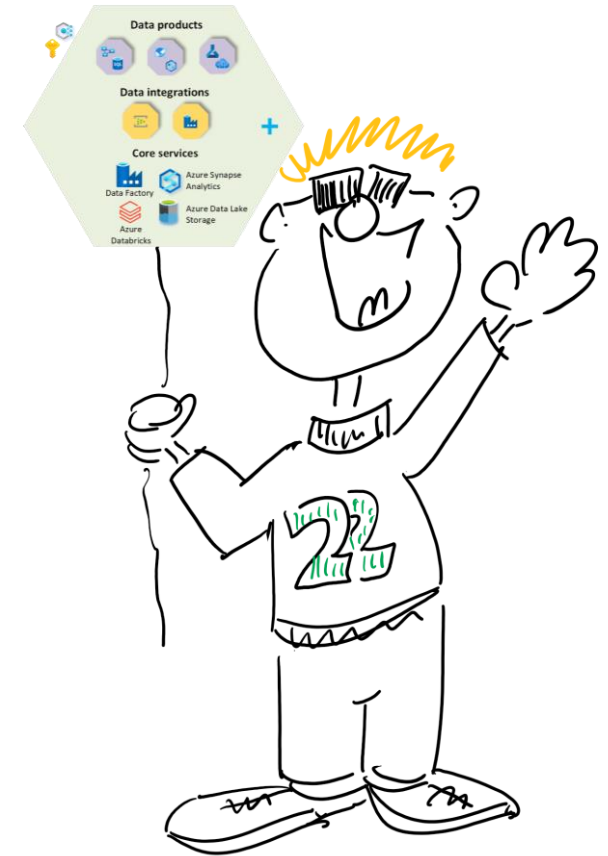
- + How can Microsoft help us to create a Data Mesh?
- + What's in it so far:
  - + CAF: Strategy, Financial consideration, Organisational alignment, Skill readiness etc.
  - + WAF: Cost optimization, Performance efficiency, Operational excellence
- + Get "Ready" with "Enterprise Scale Analytics" Template
- + OK: New name: "Cloud Scale Analytics"

I want  
something  
hands-on!



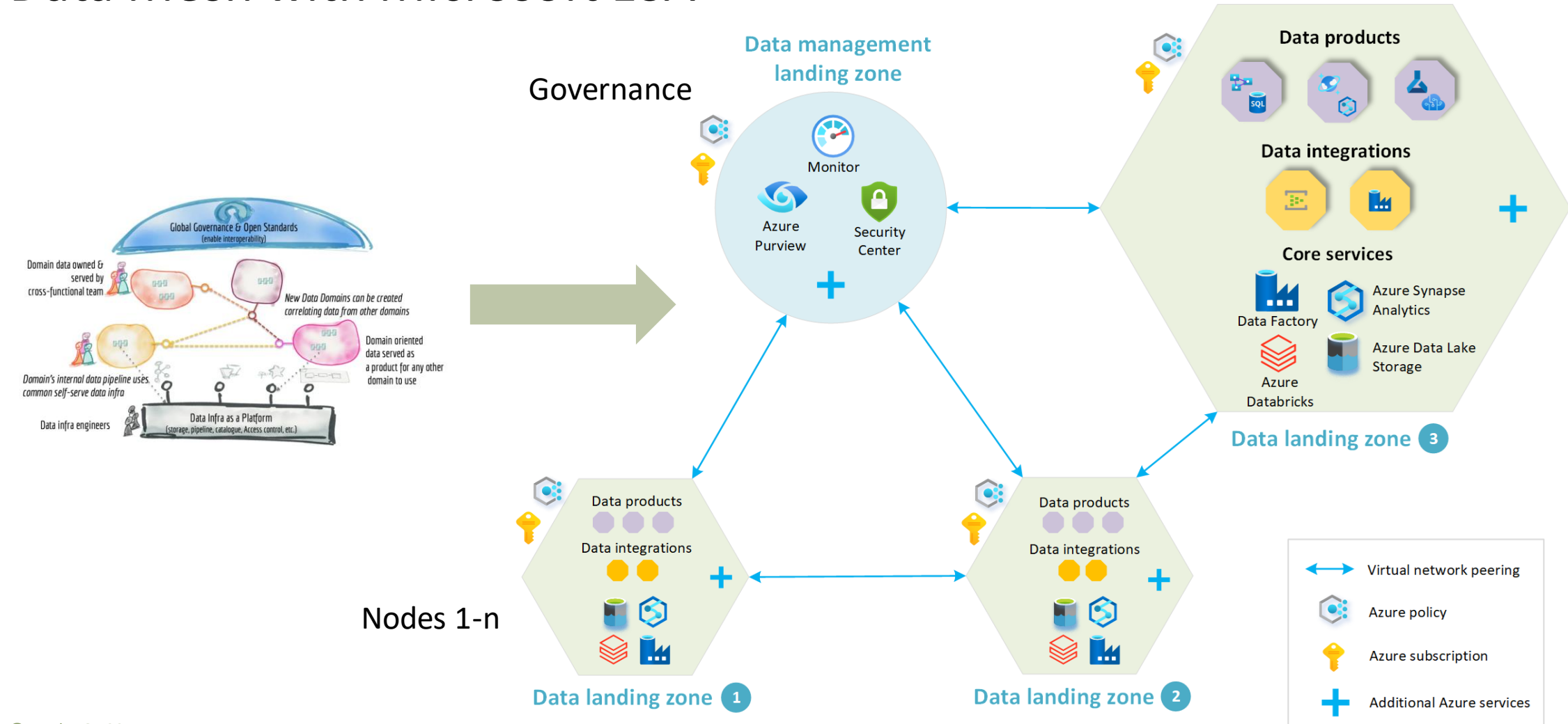
## Data Mesh with Microsoft ESA

- + What's in it for me?
- + Based on the Enterprise Scale Template, the analytics version incorporates principals of the data mesh
- + Data Landing Zone → Nodes
- + Data Management Zone → Global Governance
- + Data Products with workloads templates for
  - + Batch
  - + Streaming
  - + Analytics

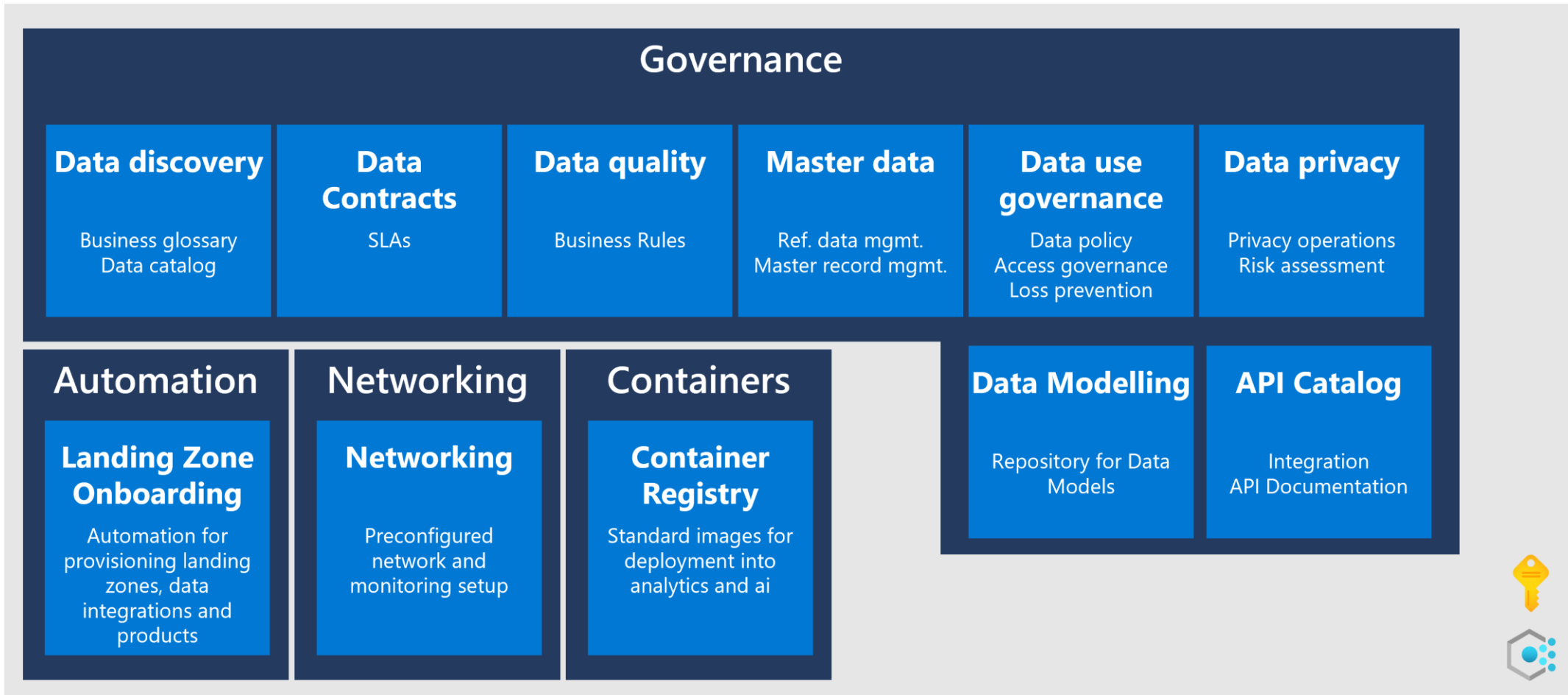




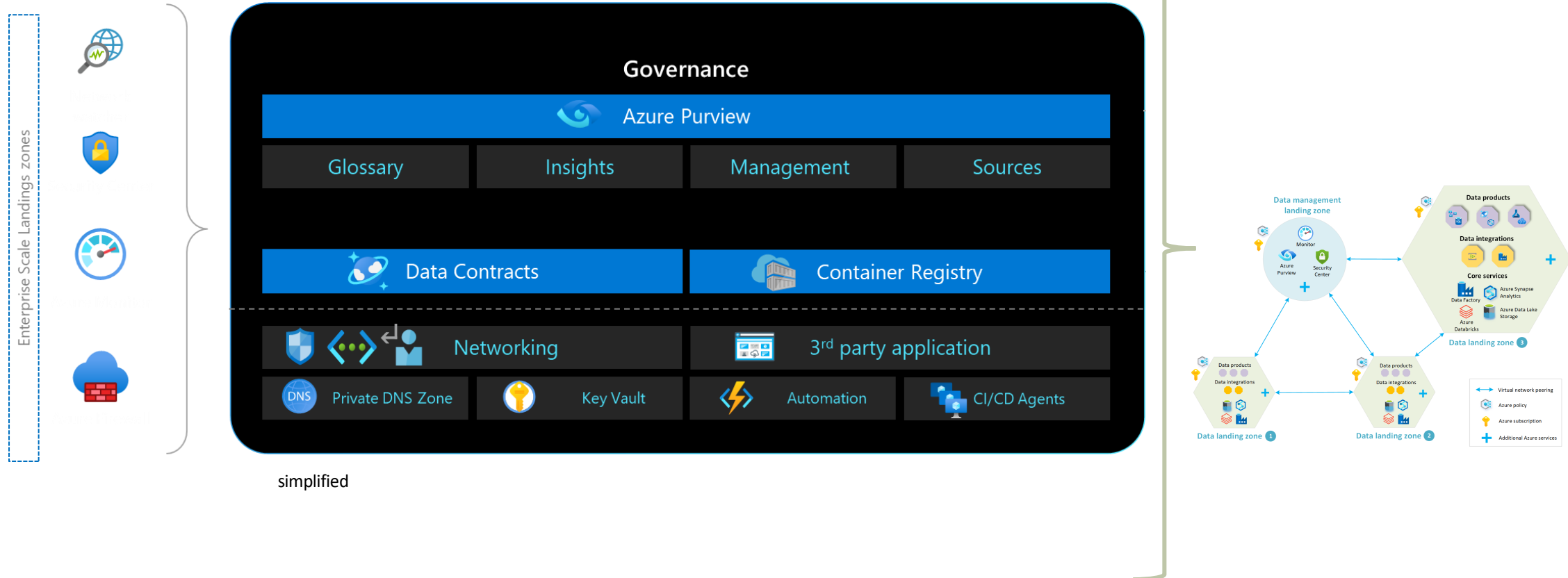
# Data Mesh with Microsoft ESA



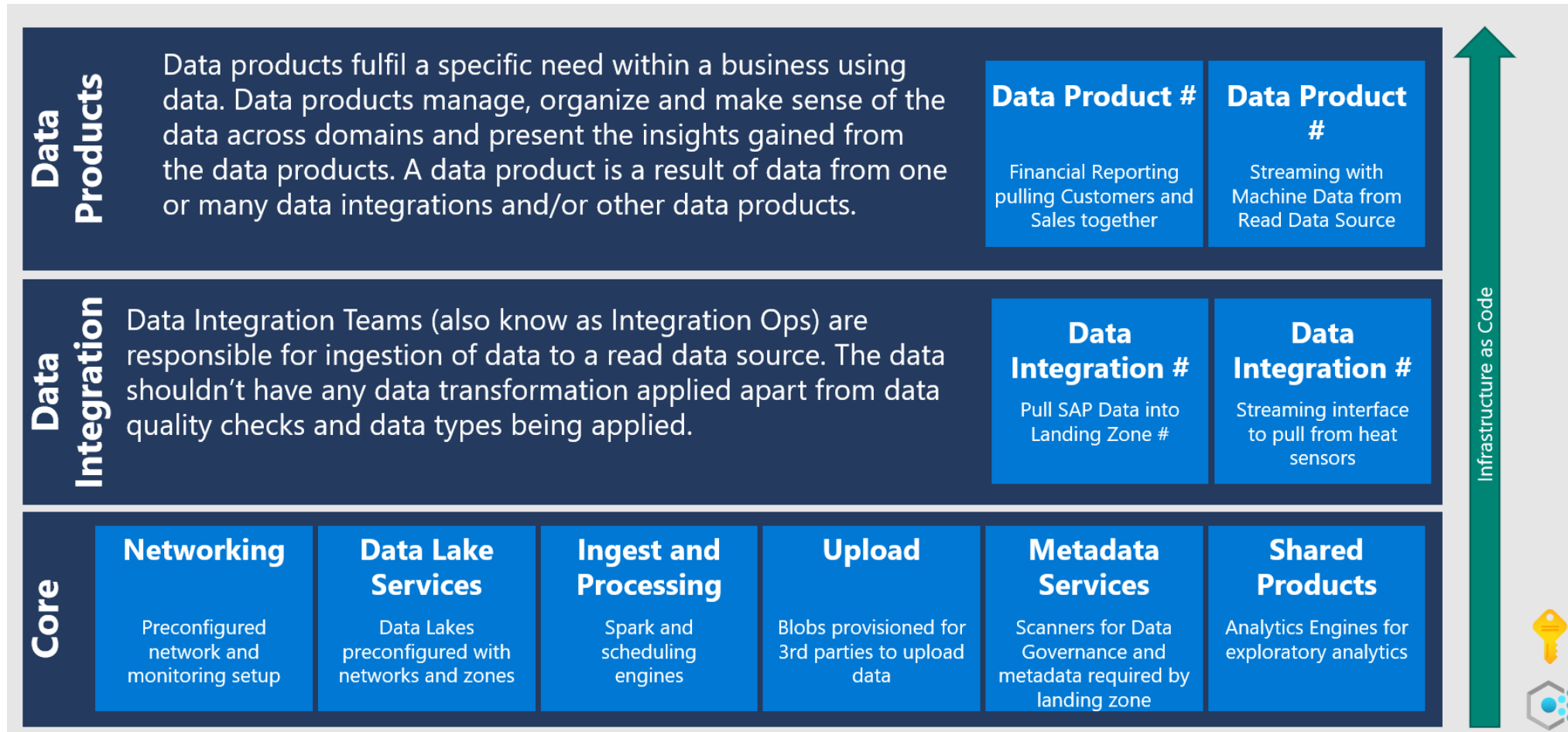
# Microsoft ESA: Data Management Zone



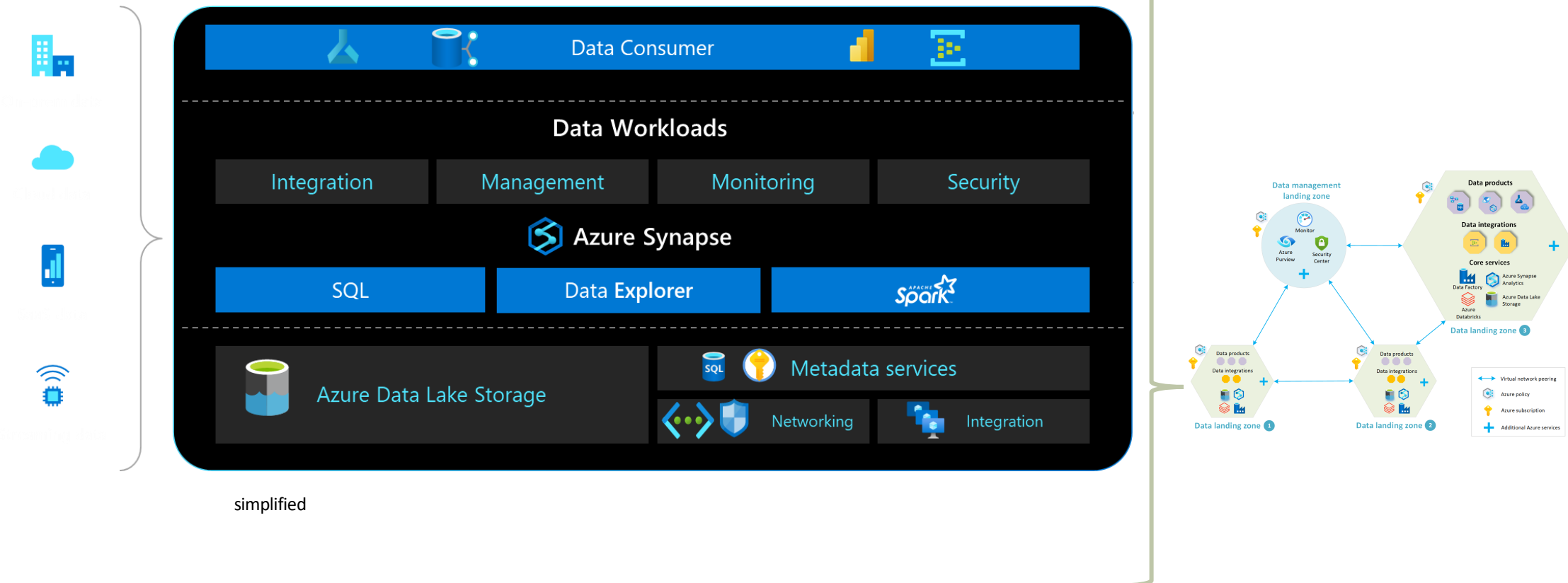
# Microsoft ESA: Data Management Zone



## Microsoft ESA: Data Landing Zone



# Microsoft ESA: Data Landing Zone



## Data Mesh with Microsoft ESA – Ready to go

- + To implement these Data Mesh architecture baselines to your landscape templates are available via GitHub Repos:
- + <https://github.com/Azure/data-management-zone>
- + <https://github.com/Azure/data-landing-zone>
- + What you need?
  - + Azure Subscriptions (min. 2)
  - + Owner or User Access Admin to create SP and RBACs
  - + GitHub or Azure DevOps

Now we are  
talking!





# ESA REPO & DEPLOYMENT

## Overview

# Data Mesh with Microsoft ESA – Git Repo

Azure / data-management-zone Public template

<> Code Issues 17 Pull requests 5 Discussions Actions Projects Wiki Security Insights

main 9 branches 3 tags Go to file Code

marvinbuss Updated Diagnostic Settings Policies for Synapse (#271) 818cd4c on Jan 24 639 commits

|                    |  |               |
|--------------------|--|---------------|
| .ado/workflows     | Added What-If Controlled Rollout (#185)                  | 5 months ago  |
| .devcontainer      | Bicep (#81)  | 9 months ago  |
| .github            | Add Data Landing Zone Subs as Input (#267)               | 2 months ago  |
| code               | Automated Purview Root Collection Role Assignment (#189) | 4 months ago  |
| docs               | Fix Portal UI Bug (#272)                                 | 2 months ago  |
| infra              | Updated Diagnostic Settings Policies for Synapse (#271)  | 2 months ago  |
| .gitattributes     | added gitignore  | 2 years ago   |
| .gitignore         | Review feedback (#72)                                    | 10 months ago |
| CODE_OF_CONDUCT.md | Initial CODE_OF_CONDUCT.md commit                        | 2 years ago   |
| CONTRIBUTING.md    | Update CONTRIBUTING.md (#273)                            | 2 months ago  |
| LICENSE            | Initial LICENSE commit                                   | 2 years ago   |
| README.md          | Rename Data Management Zone (#269)                       | 2 months ago  |
| SECURITY.md        | Validation of readme (#43)                               | 13 months ago |

Azure DevOps YML Pipeline

Services to deploy

Azure / data-landing-zone Public template

<> Code Issues 6 Pull requests 4 Discussions Actions Projects Wiki Security Insights

main 6 branches 3 tags Go to file Code

marvinbuss Fix Portal UI Bug (#226) f7fcb35 on Jan 14 444 commits

|                    |   |               |
|--------------------|---|---------------|
| .ado/workflows     | Added SHIR Deployment via Custom Data (#214)          | 3 months ago  |
| .devcontainer      | Bicep (#123)  | 9 months ago  |
| .github            | Add Data Landing Zone Subs to Storage Firewall (#223) | 2 months ago  |
| code               | Updated All Docs (#142)                               | 8 months ago  |
| docs               | Fix Portal UI Bug (#226)                              | 2 months ago  |
| infra              | Rename data management zone (#224)                    | 2 months ago  |
| .gitattributes     | updated gitattributes                                 | 13 months ago |
| .gitignore         | Feedback review changes (#113)                        | 10 months ago |
| CODE_OF_CONDUCT.md | Initial CODE_OF_CONDUCT.md commit                     | 2 years ago   |
| CONTRIBUTING.md    | Espen/contribution_guide (#227)                       | 2 months ago  |
| LICENSE            | Initial LICENSE commit                                | 2 years ago   |
| README.md          | Rename data management zone (#224)                    | 2 months ago  |
| SECURITY.md        | Tagging, Linting, RG Deployment (#121)                | 9 months ago  |

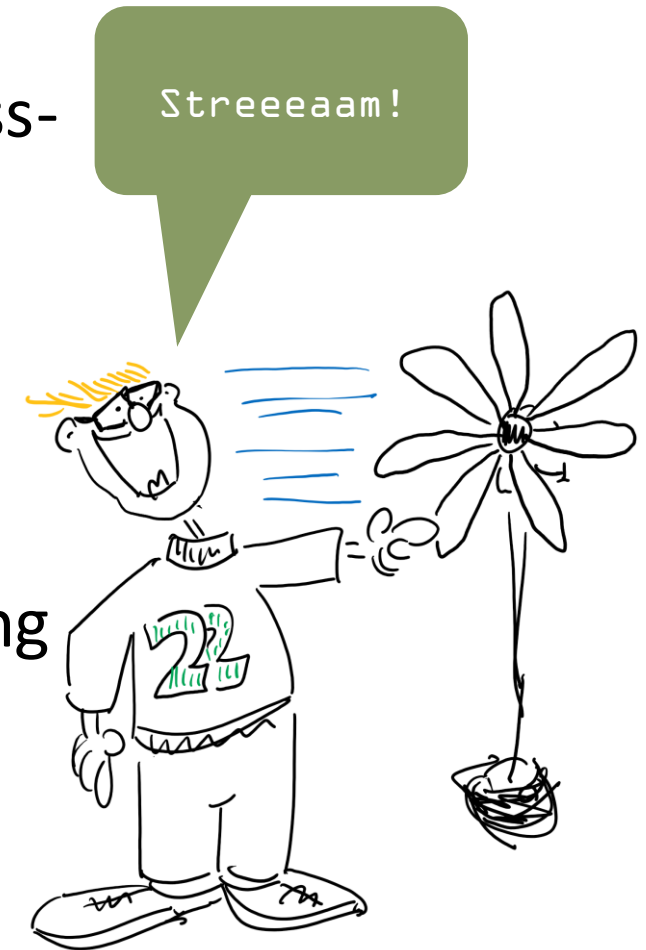
GitHub Action YML

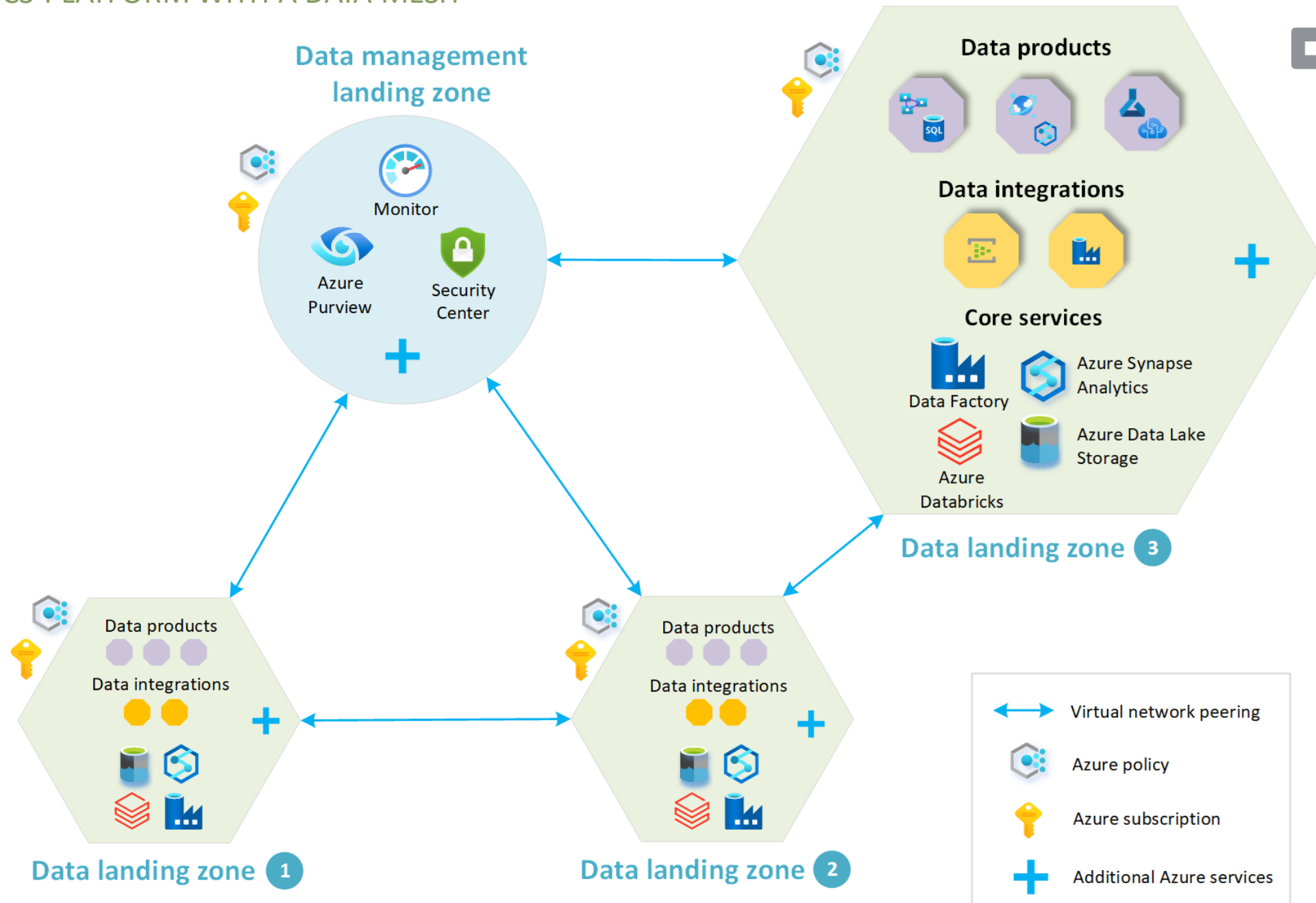
Shell Scripts PS/SH



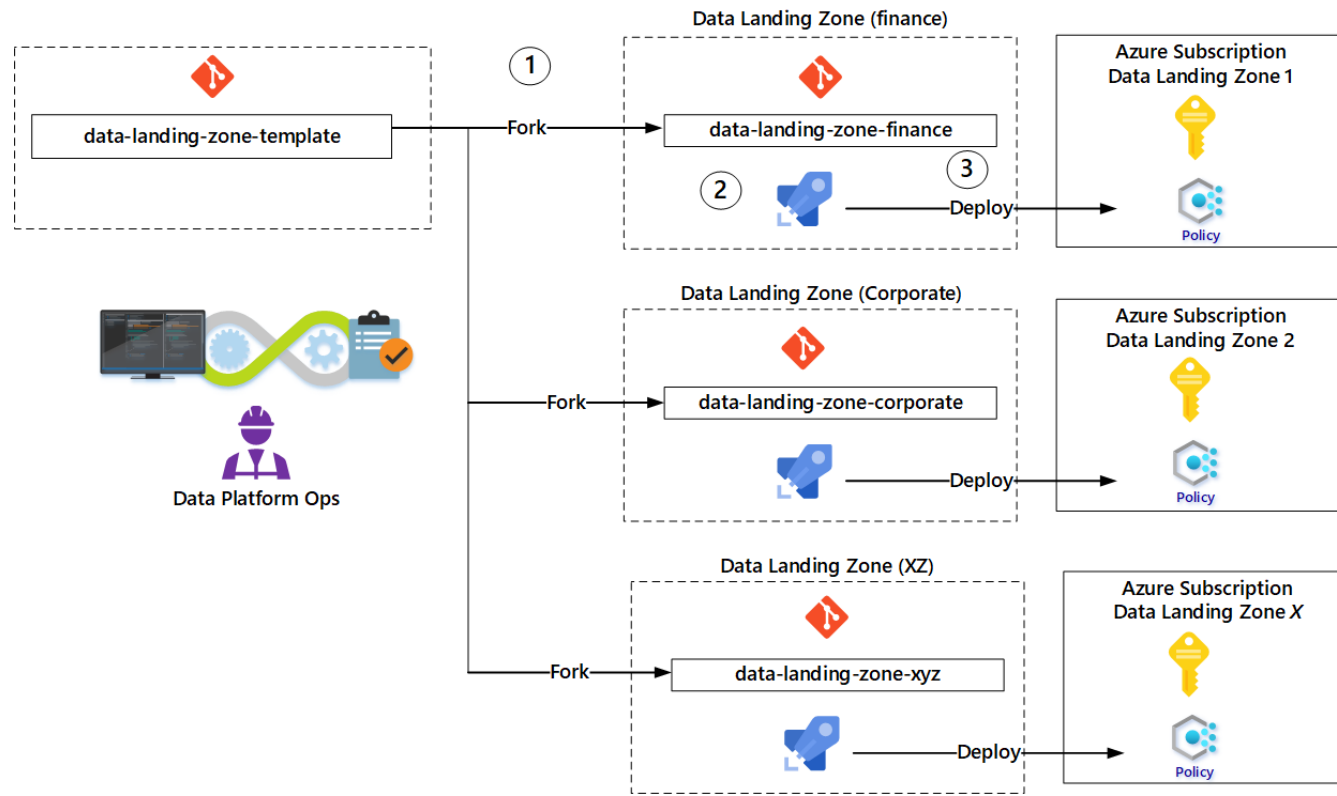
## Data Mesh with Microsoft ESA – Git Repo

- + Additional templates for the data mesh to enable cross-functional teams
- + <https://github.com/Azure/data-product-streaming>
- + <https://github.com/Azure/data-product-batch>
- + <https://github.com/Azure/data-product-analytics>
- + These templates can be deployed into the Data Landing zone





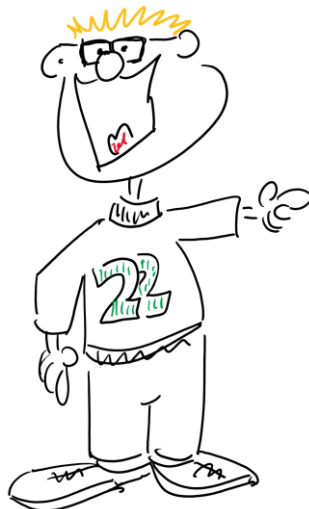
# Data Mesh with Microsoft ESA – Deployment



## Data Mesh with Microsoft ESA – Costs

- Pricing Calculator templates:
  - Management Zone: <https://azure.com/e/658478643d4b46fdbf8b1972c4b0704b>
  - Data Landing Zone: <https://azure.com/e/55cb6feafcc24cec8cfefb10486d54ab5>

Data Management Zone w/ Azure Firewall and Private DNS Zones



| Storage Accounts         | i | Block Blob Storage, General Purpose V2, LRS Redun...      | + - | Upfront: \$0.00 | Monthly: \$0.00     |
|--------------------------|---|---|-----|-----------------|---------------------|
| Azure Container Registry | i | Premium Tier, 1 units x 30 days, 0 GB Bandwidth, 0 ...    | + - | Upfront: \$0.00 | Monthly: \$50.00    |
| Key Vault                | i | Vault: 0 operations, 0 advanced operations, 0 renew...    | + - | Upfront: \$0.00 | Monthly: \$0.00     |
| Azure Private Link       | i | 8 Endpoints X 730 Hours, 100 GB Outbound data pr...       | + - | Upfront: \$0.00 | Monthly: \$60.40    |
| Virtual Network          | i | East US (Virtual Network 1): 0 GB Outbound Data Tr...     | + - | Upfront: \$0.00 | Monthly: \$0.00     |
| Azure DNS                | i | Zone 1, Private; 46 hosted DNS zones, 0 DNS queries       | + - | Upfront: \$0.00 | Monthly: \$14.60    |
| Azure Firewall           | i | Premium tier, 1 Logical firewall units x 730 Hours, 0 ... | + - | Upfront: \$0.00 | Monthly: \$1,277.50 |
| IP Addresses             | i | 0 Dynamic IP Addresses, 1 Static IP Addresses, 0 Re...    | + - | Upfront: \$0.00 | Monthly: \$2.63     |

... we need to talk

We need to  
discuss this  
later!



Yeah!  
all  
distributed!



## Is a data mesh always useful?

- + As usual, all depends:
- + More suitable for Enterprises
  - + Number of data domains
  - + Number of different data teams
  - + Data as a product mindset
  - + Quantity of data sources
  - + Importance of data governance
- + Scoring sample by Barr Moses (link below)

We need to  
discuss this  
later!



Yeah!  
all  
distributed!



## Is the ESA approach useful?

### + Pro

- + Great Source of Knowledge and Best Practises
- + High flexibility
- + Leverages the actual state of the art data platform
- + Can be adjusted to your needs
- + Open Source
- + BICEPs Template approach
- + Templates can be adopted to smaller footprints
- + Git and AzDO pipelines

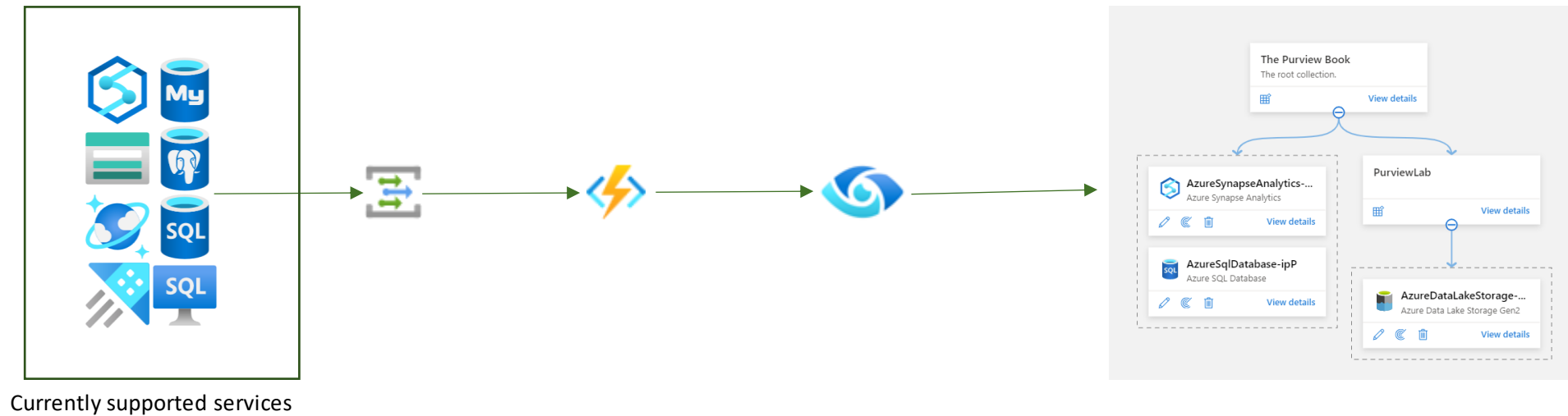


### - Con

- Complexity for small and medium enterprises
- Governance functionality not completely covered by MS
- Data Virtualisation needs are complex to achieve
- Centralize Management or Standards for Data Access (not yet) available
- Can be difficult to apply to existing Azure implementations



# Purview Automation

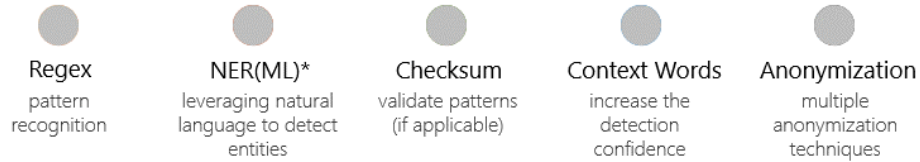


[marvinbuss/PurviewAutomation: Event-Driven Onboarding, Scanning and Classification of Data Sources and Setup of Lineage to Azure Purview. \(github.com\)](https://github.com/marvinbuss/PurviewAutomation)



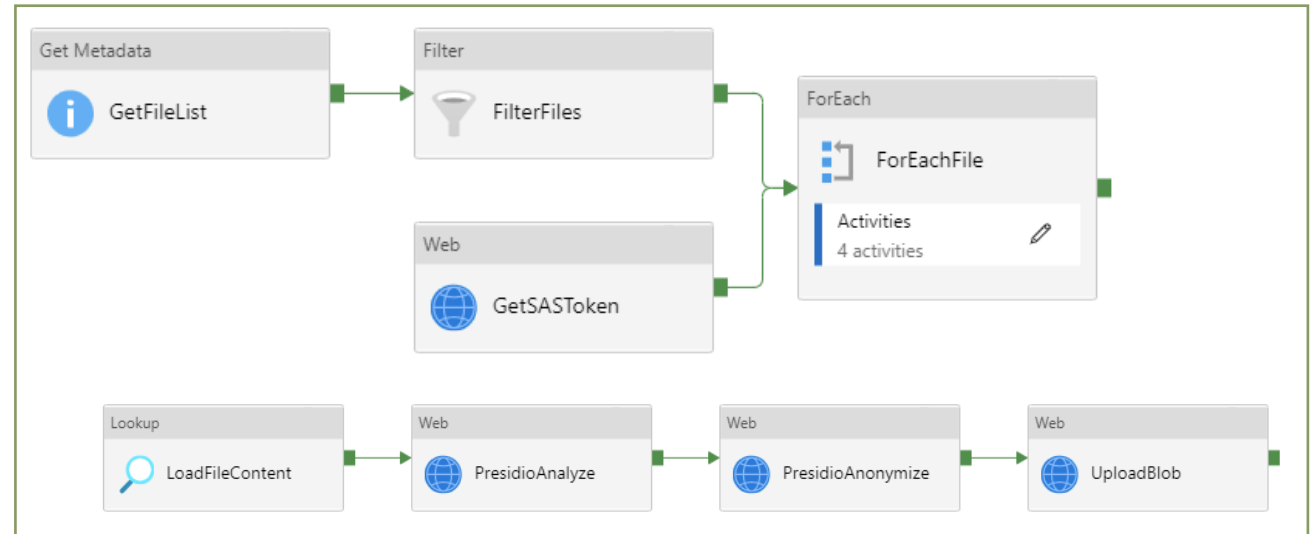
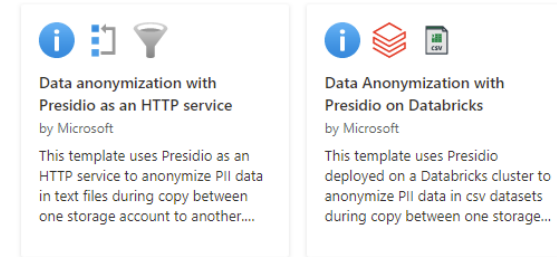
# Microsoft Presidio

## Presidio Detection Flow



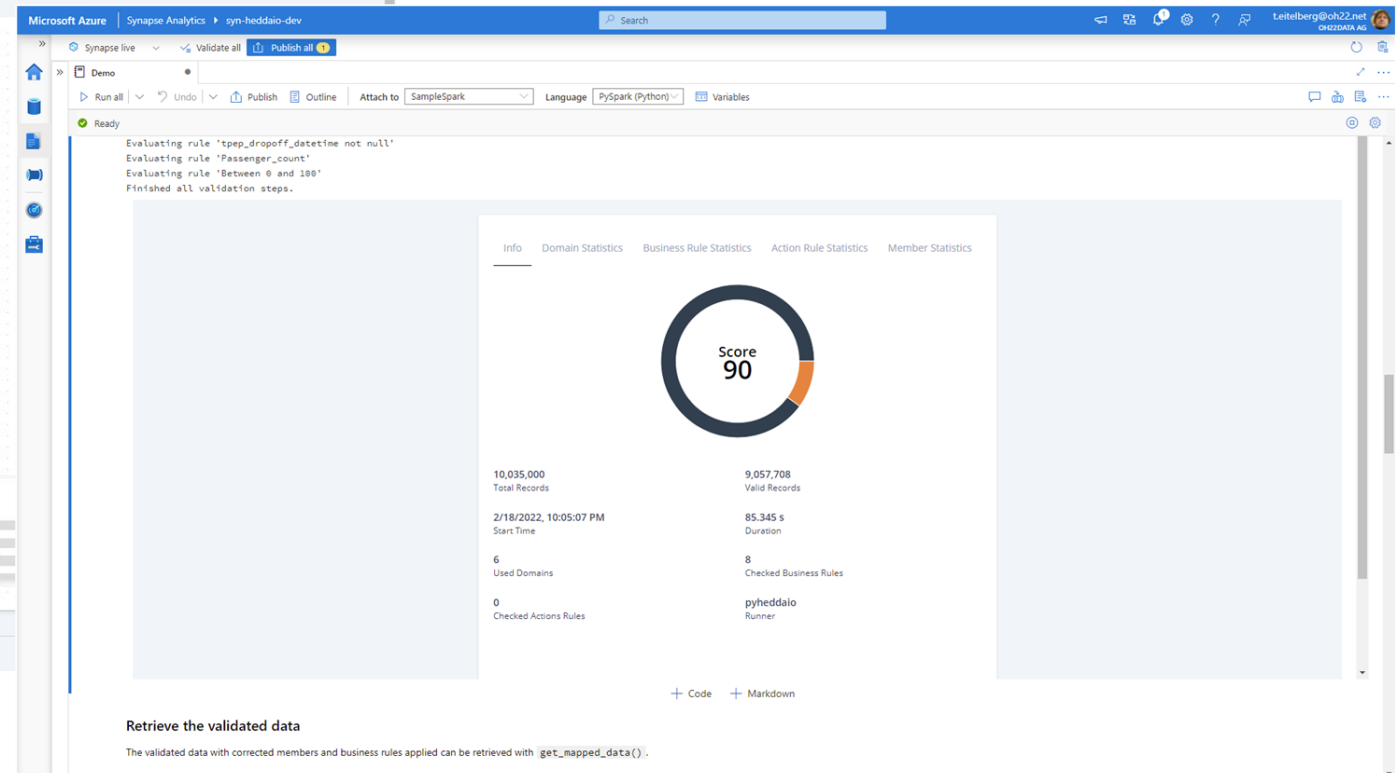
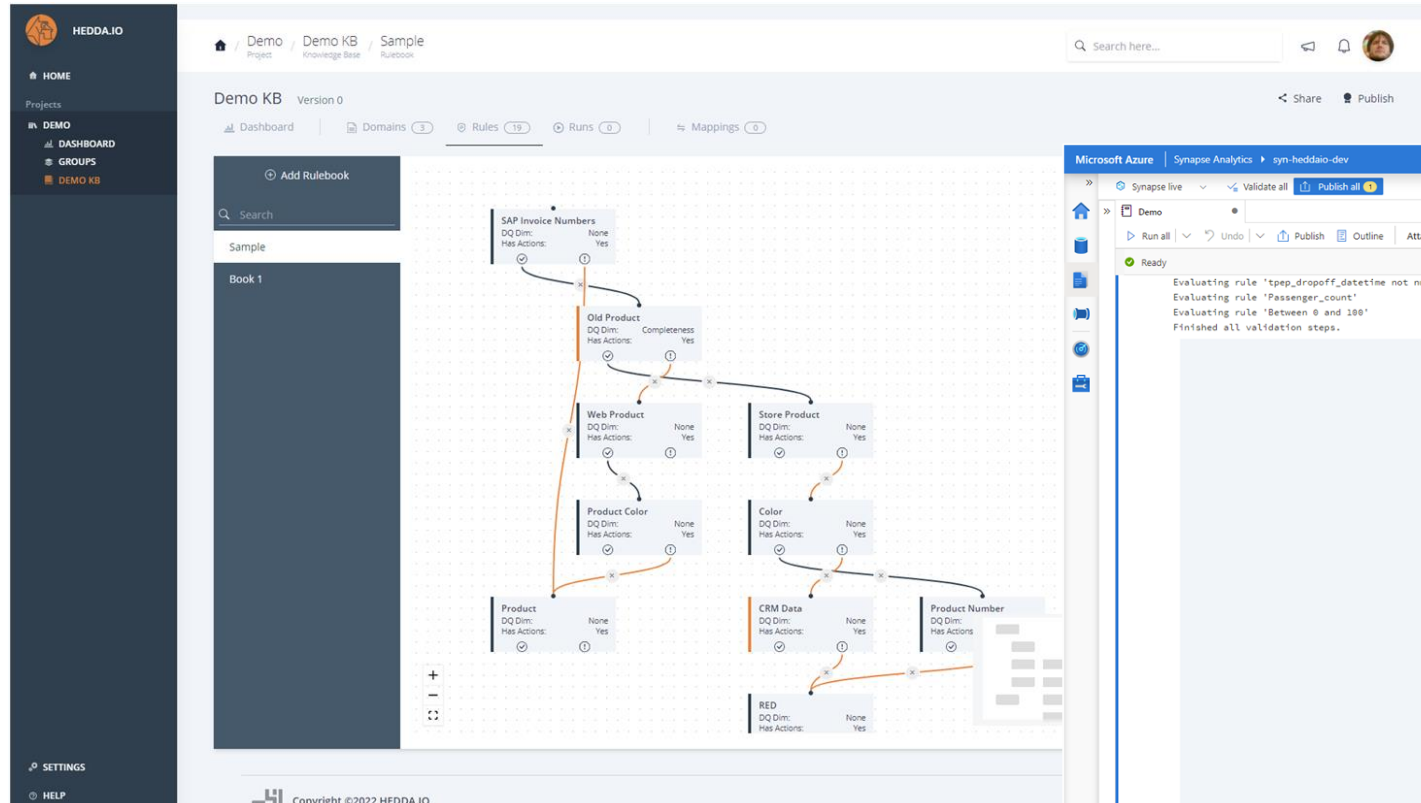
\*NER – Named Entity Recognition

<https://microsoft.github.io/presidio/>



# HEDDA.IO

Central cloud native data quality solution with  
standalone workers for Databricks and Synapse.



## Some useful links

- + [Data Mesh Principles and Logical Architecture \(martinfowler.com\)](https://martinfowler.com/data/data_mesh/)
- + [Data Mesh defined | James Serra's Blog](#)
- + [Data Mesh: Centralized vs decentralized data architecture | James Serra's Blog](#)
- + [Implementing Data Mesh on Azure. Practical design considerations when... | by Piethein Strenghtolt | Jan, 2022 | Towards Data Science](#)
- + [Data Domains — Where do I start?. Practical guidance from the field | by Piethein Strenghtolt | Towards Data Science](#)
- + [Building a Data Mesh Architecture in Azure – Part 1 – Welcome to the Community Blog of Paul Andrew \(mrpaulandrew.com\)](#)

<https://evals.datagrillen.com/evals.aspx>

