scieneers
DRIVEN BY DATA

Build your data platform on Azure

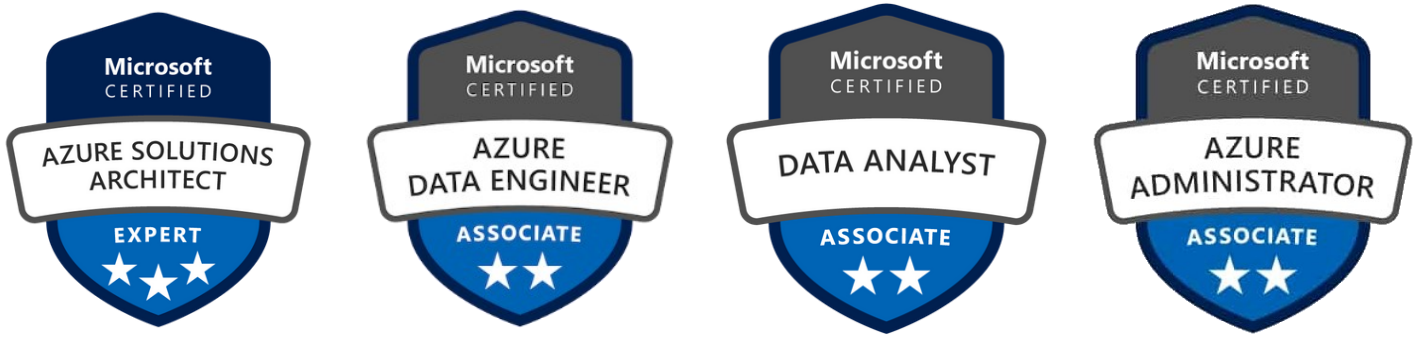
...but secure please!



Stefan Kirner



- > 20 years experience using Microsoft Data Platform
- PASS Chapter Lead Karlsruhe & Beirat
- Director Business Intelligence scieneers GmbH
- Twitter: @KirnerKa





Agenda

- General security rules
- Data Platform architectures and security topics
- Safe connectivity on-prem / Azure
- Keeping secrets secret
- Networks, Firewalls and Endpoints
- Data view permissions
- Summary



General Security Rules

General Security Rules



KISS: Keep It simple, stupid

Avoid over-engineering



SoC: Separation of Concerns

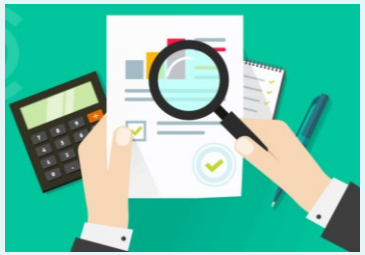
Create intelligent structure for Azure resources



DRY: Don‘t repeat yourself

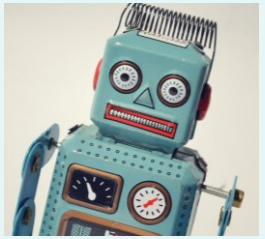
permissions: Prefer function based groups

General Security Rules



Monitoring

Automate testing and security audits



Automate processes

Code to replace frequent manual processes

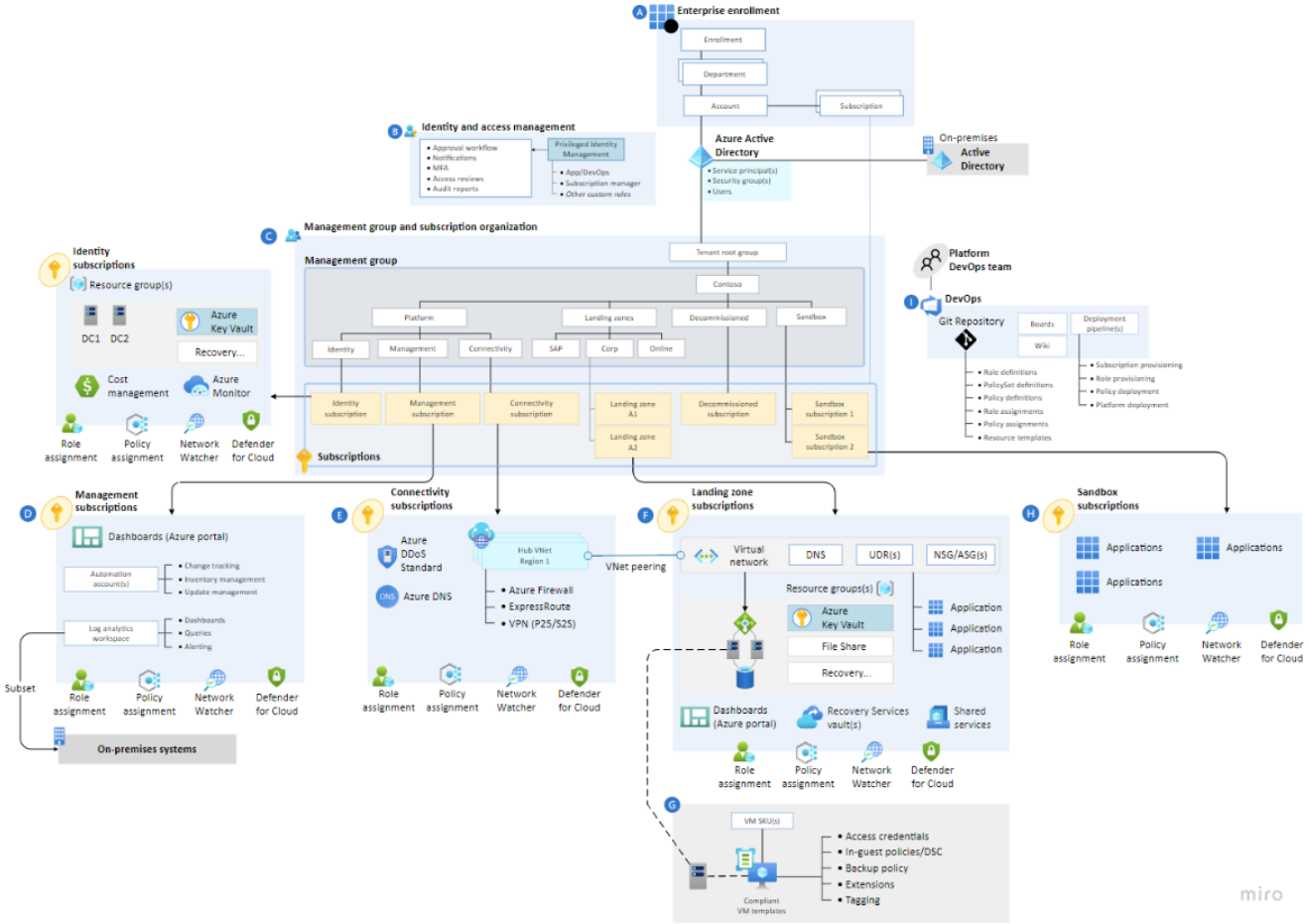


Documention for transparancy

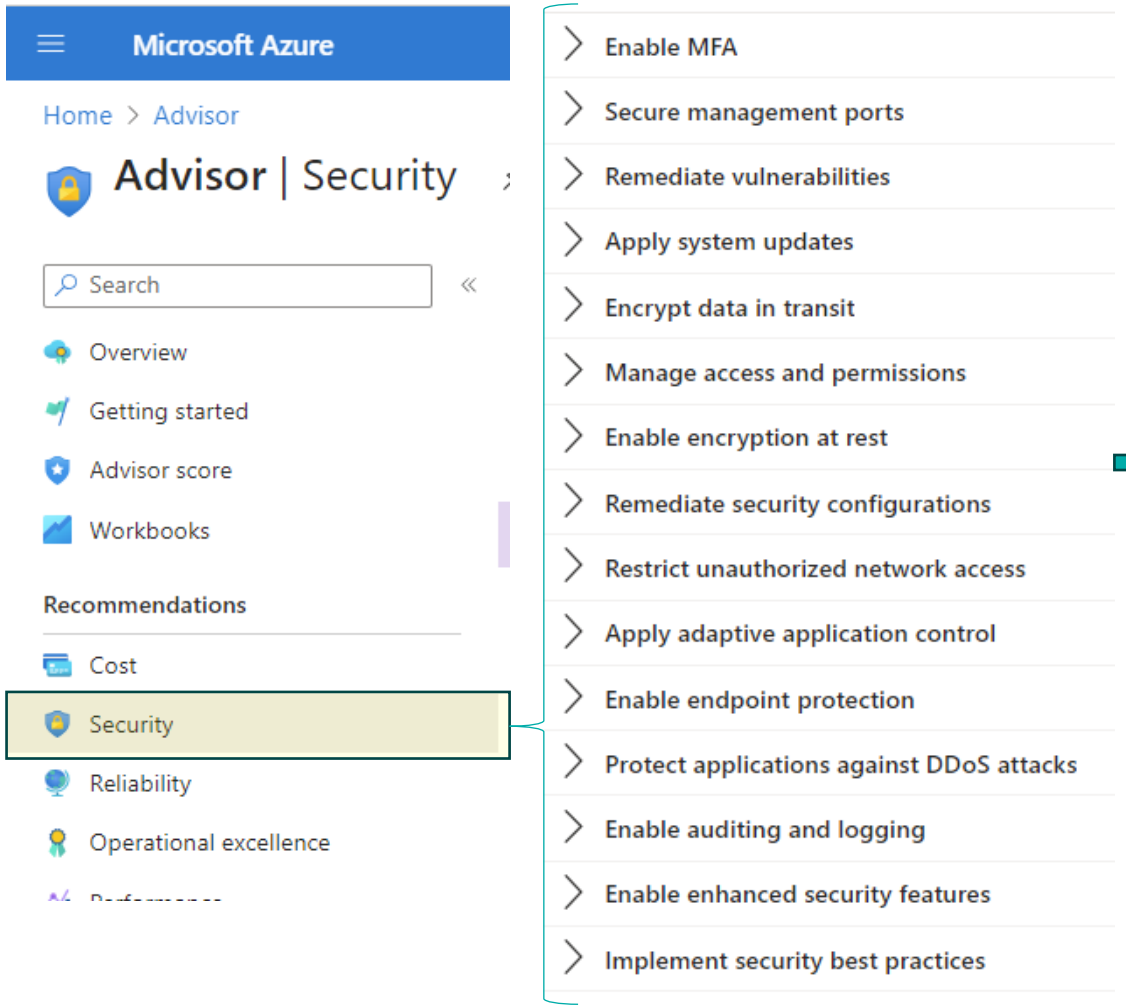
Focus on concepts explanation of decisions

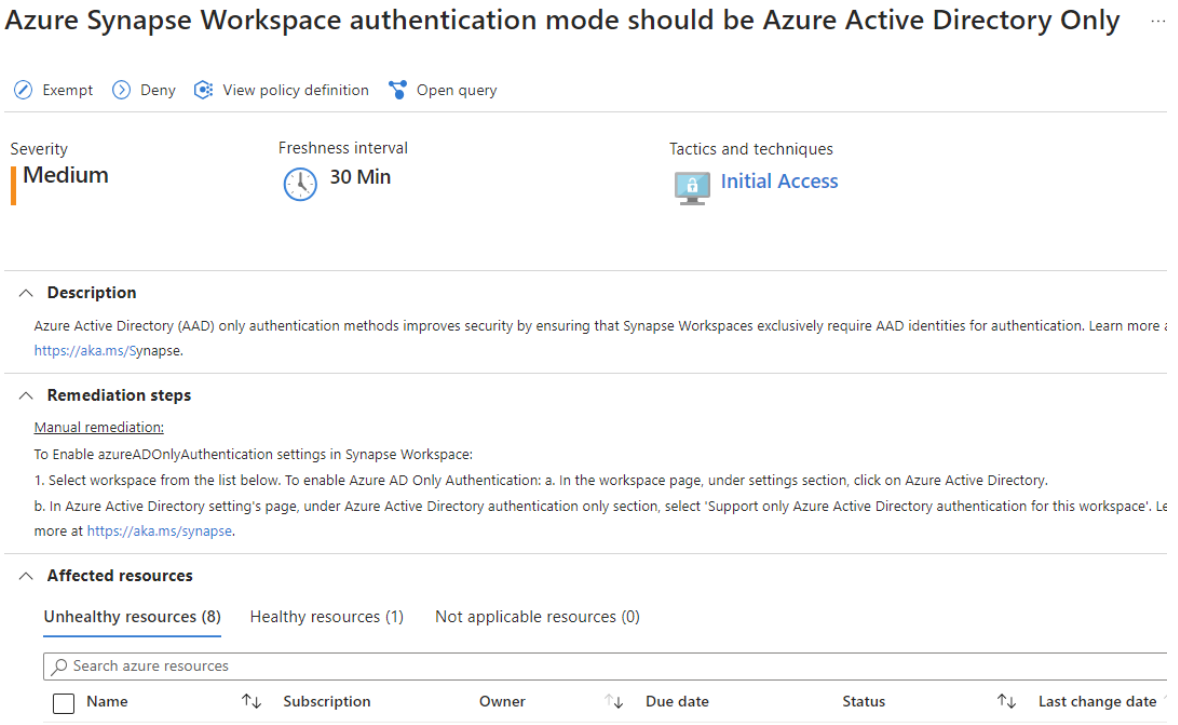
Good starting point - using Azure Landing zones

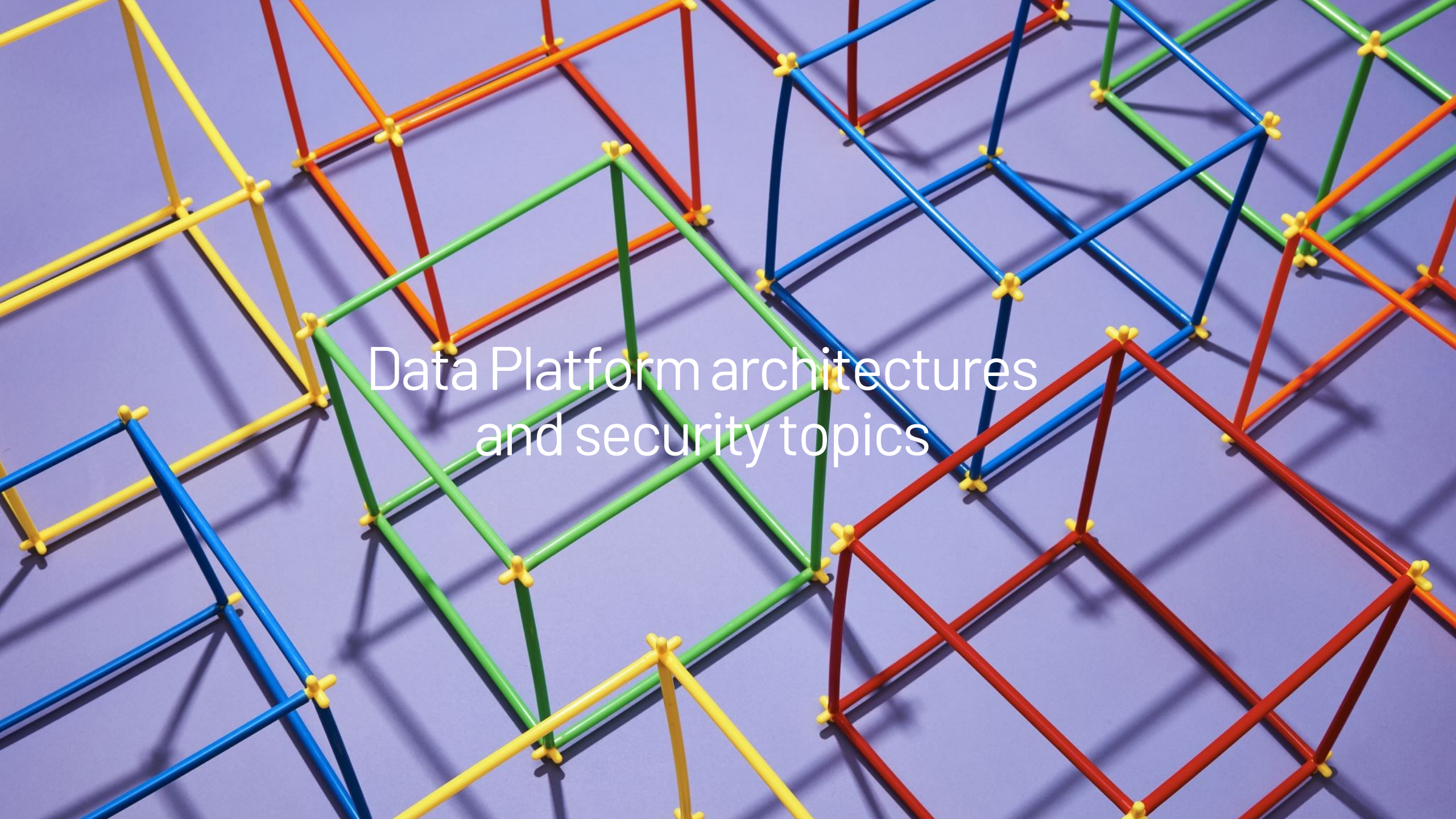
- Best Practice recommended by Microsoft
- Tree structure of logical containers allows policies and roles to be applied at different levels, reducing redundancy
- Contains many useful policies
- SoC: Azure platform vs. applications
- Sandbox as environment in which users can learn (separated and secure)
- Similar structures avoid mistakes and simplify onboarding of new project members



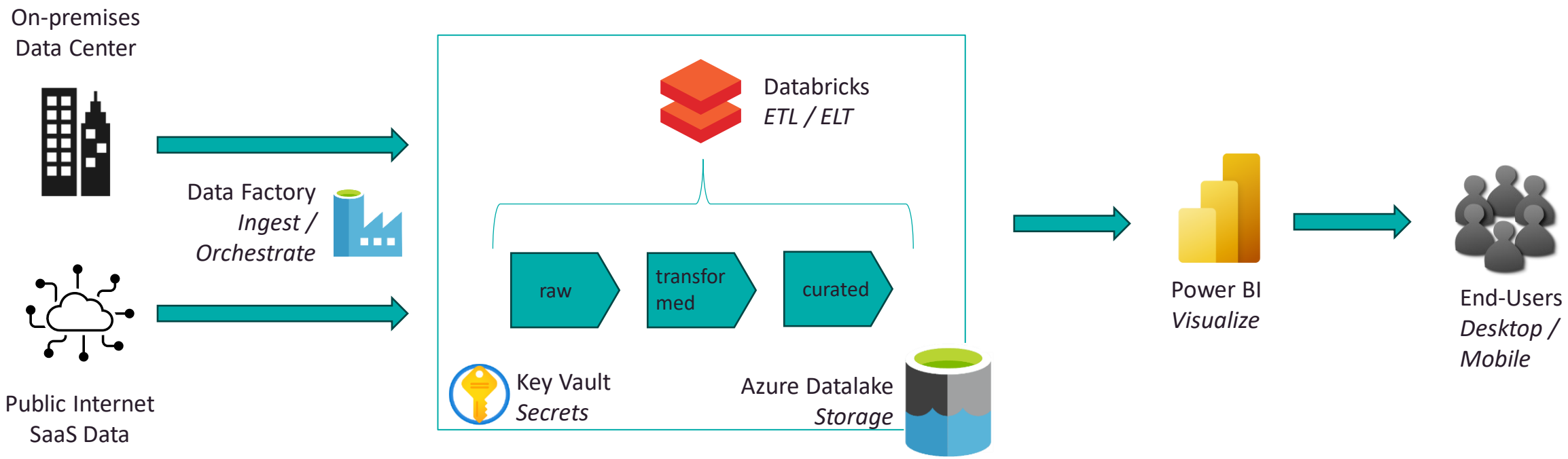
Use Azure Security Advisor to check for existing issues (and fix them)



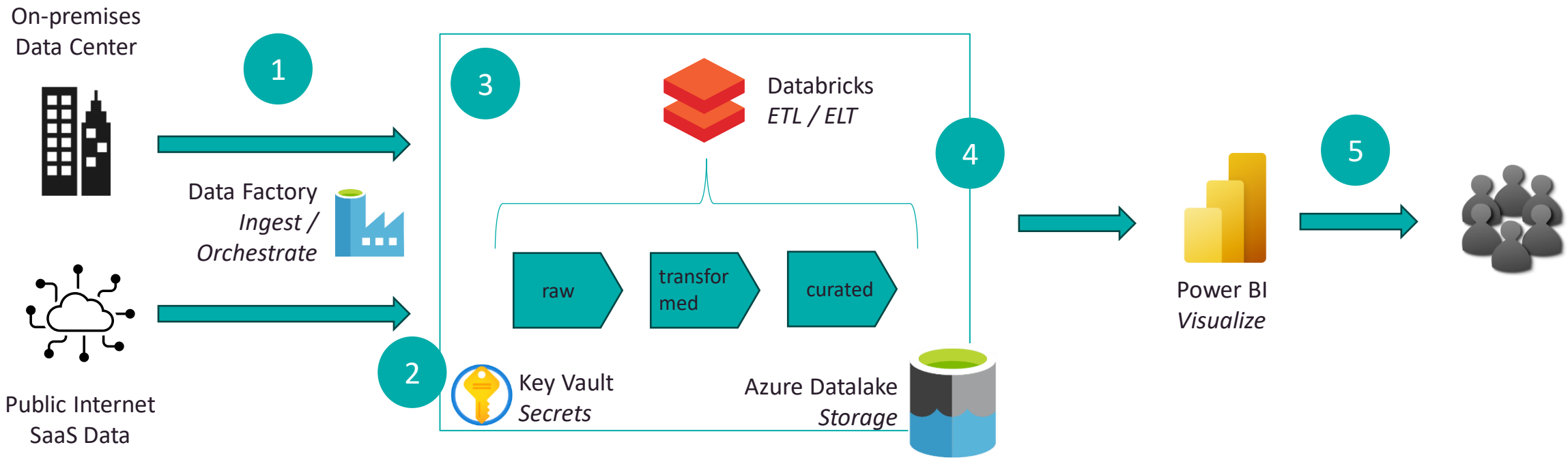



Data Platform architectures
and security topics

Azure Data Services – a typical setup



Azure Data Services – security topics



1 Safe connectivity on-prem / Azure

2 Keeping Secrets Secret

3 Authorization

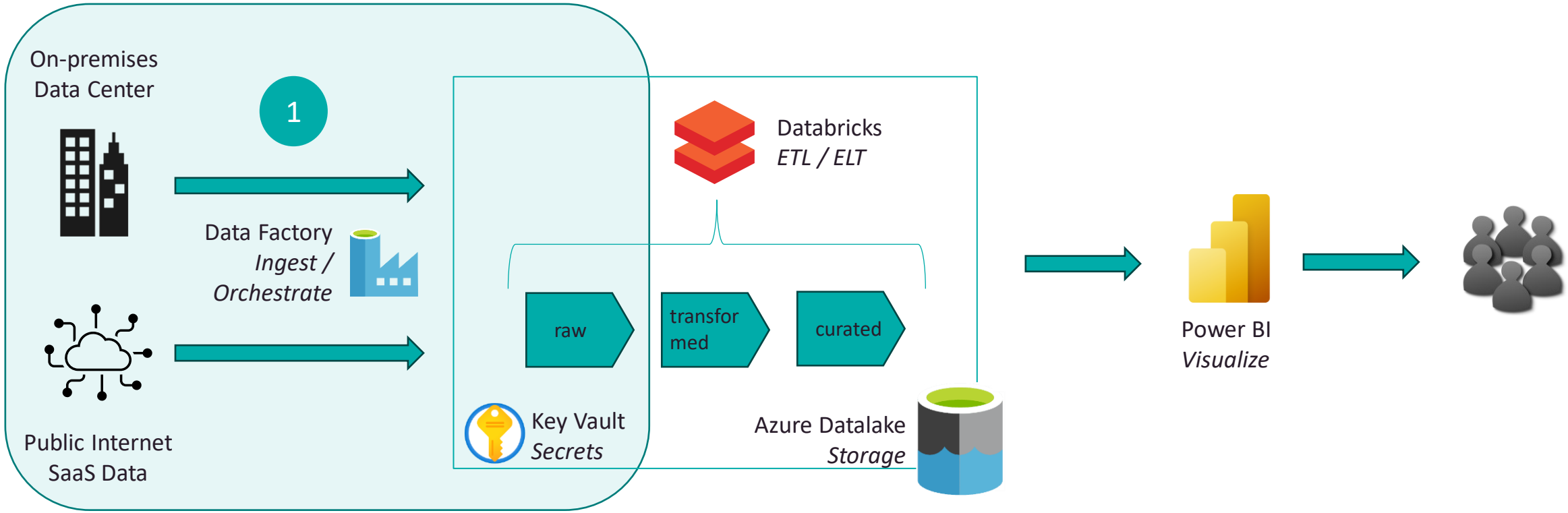
4 Networks, Firewalls and Endpoints

5 Data view permissions



Safe connectivity on-prem / Azure

Azure Data Services – safe connectivity



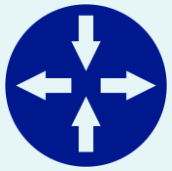
1 Safe connectivity on-prem / Azure

Connectivity options



SHIR for Azure Data Factory / Synapse

Connect Azure Data Factory to on-prem using the „Self-hosted Local Integration Runtime“



On-premises data gateway

Connect Power BI, Azure Analysis Services and other data services to on-prem

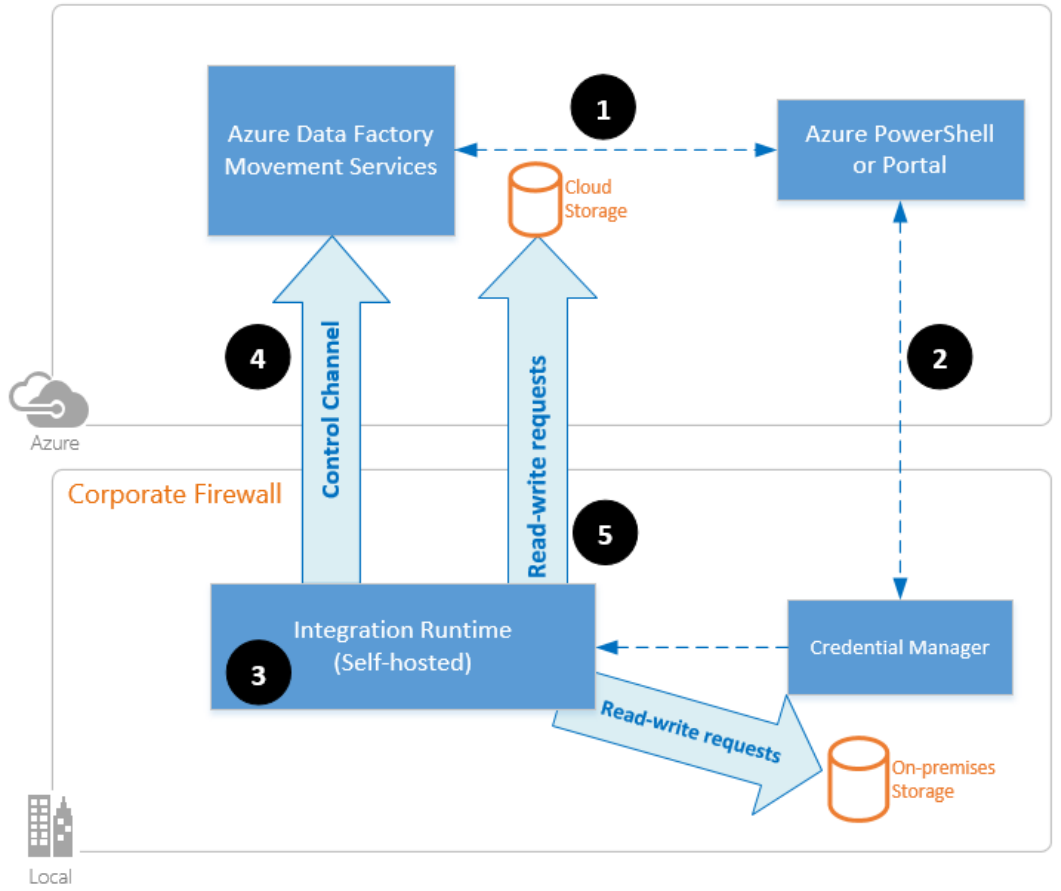


Virtual Private Network

Connect Azure to on-prem permanently by site2site VPN tunnels

Self-hosted Local Integration Runtime

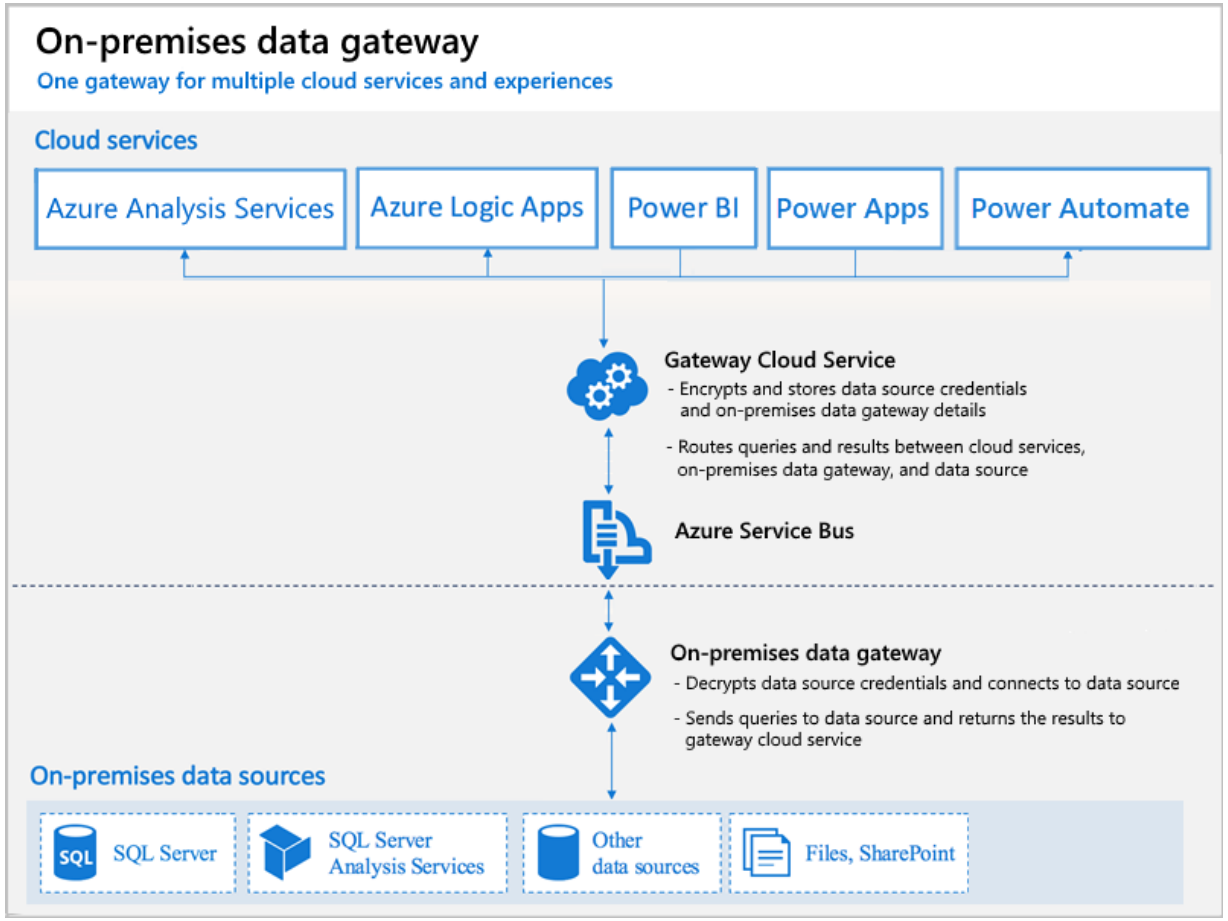
- Connectivity for Azure Data Factory (Synapse) to on-prem
- Gateway software from Microsoft installed on Virtual Machine in your corporate network
- Permissions to local resources provided to service account at VM
- Secured, encrypted https connection between on-prem and Azure by default ports



Vgl. https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime#command-flow-and-data-flow

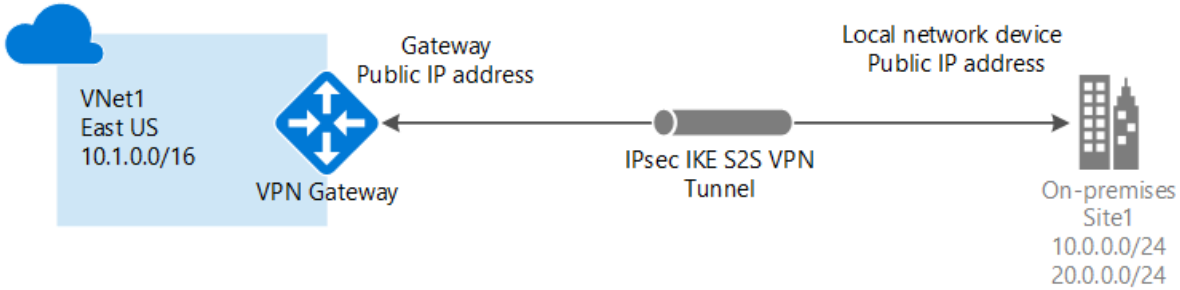
On-premises data gateway

- Connectivity for following services:
  - Power BI, Azure Analyses Services, Azure Automation, Power Apps, Power Automate, Azure Logic Apps
- Gateway software from Microsoft installed on Virtual Machine in your corporate network (not same as SHIR)
- Personal / non-personal mode
- Direct Query / Live Connect for SSAS in Power BI



Virtual private network – VPN gateway

- Site2Site VPN
- encrypted traffic over public internet or Microsoft network
- VPN gateway vs. Express Route gateway
- Needs compatible VPN device*
- Open necessary ports for data services
- SHIR could be necessary on Top



https://docs.microsoft.com/en-us/azure/vpn-gateway/vpn-gateway-howto-site-to-site-resource-manager-portal

* https://docs.microsoft.com/en-us/azure/vpn-gateway/vpn-gateway-about-vpn-devices

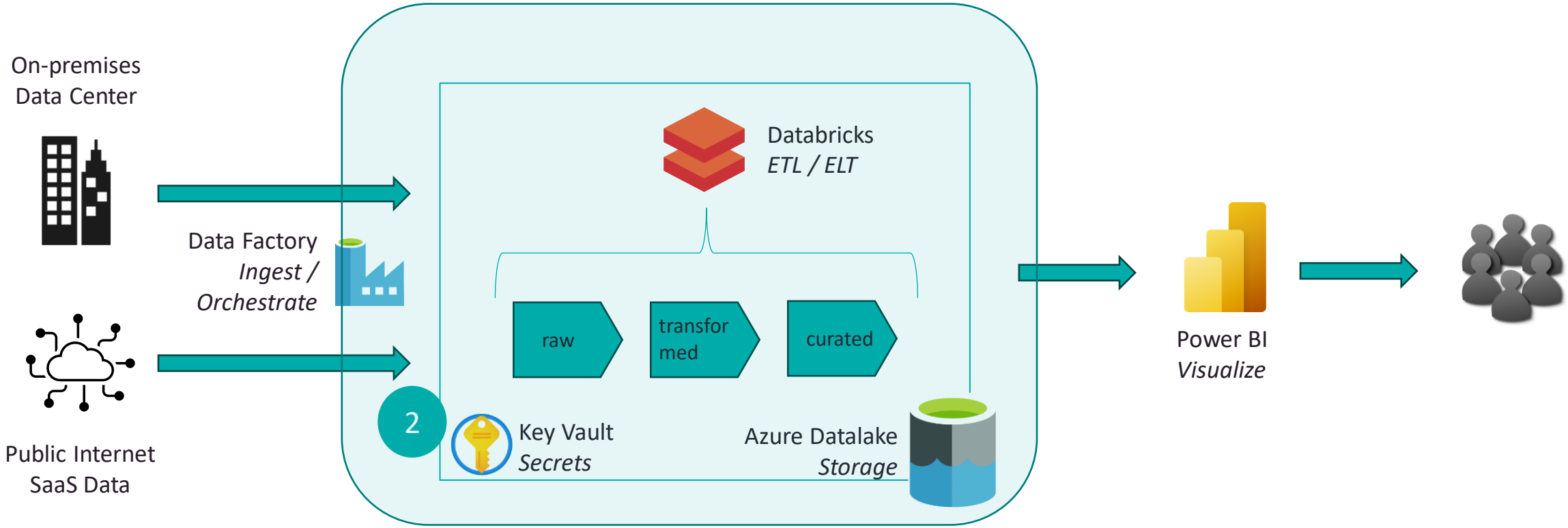
Conclusion & Takeaways

- Use SHIR for Azure Data Factory / Synapse or On-premises gateway for other services as quick but safe option for data projects
- Use VPN for more integration into your IT Infrastructure, but more overall effort to establish



Keeping secrets secret

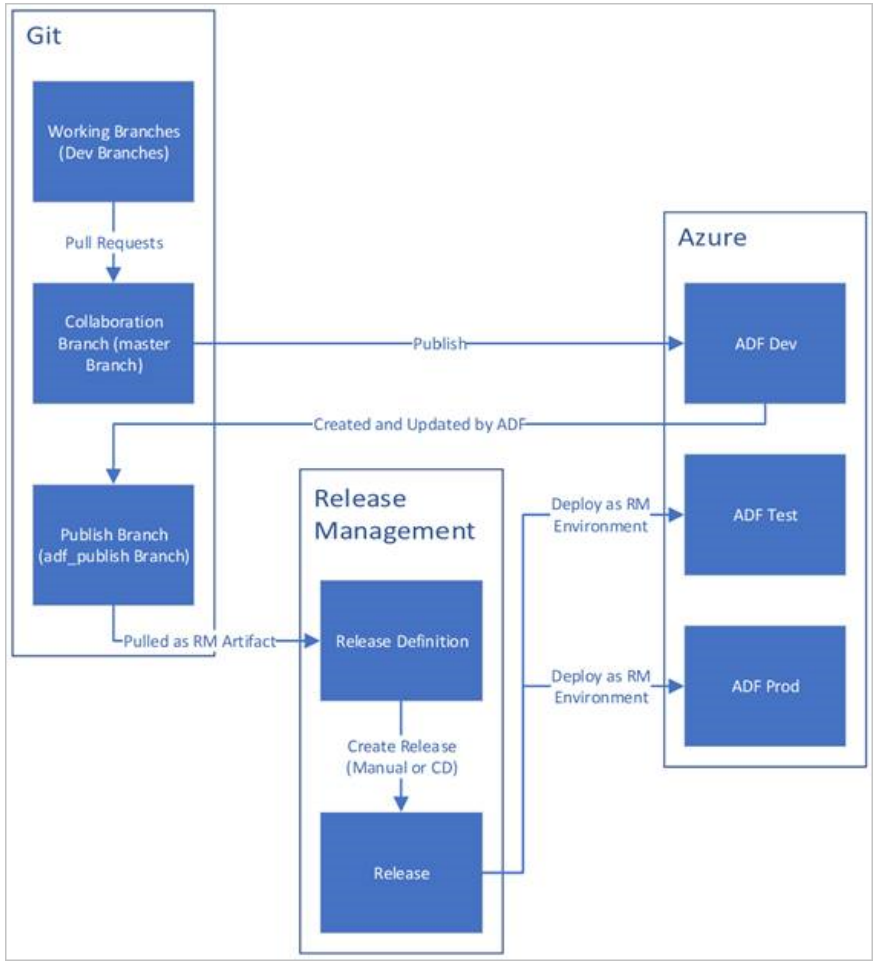
Azure Data Services – Keeping secrets secret



2 Keeping Secrets Secret

Secrets

- Inside Azure use Managed Identities whenever possible for Azure Service 2 Service communication
- But Passwords and other secrets for accessing data sources and services have to be used by ETL and Deployment Pipelines
- # of entries above multiplied by different environments (Dev, Test, Prod..)
- Raising # of people who know passwords and the ways of sharing them means risk



Example Deployment Process

Technical users – The glue of your data application

- Advantages:
  - No MFA (Problem with service)
  - No password sharing (between colleagues)
  - No expiring passwords
- Types
  - All types are "Service Principals" in the background
  - Managed Identities
  - Azure AD Application with certificate or password in Key Vault

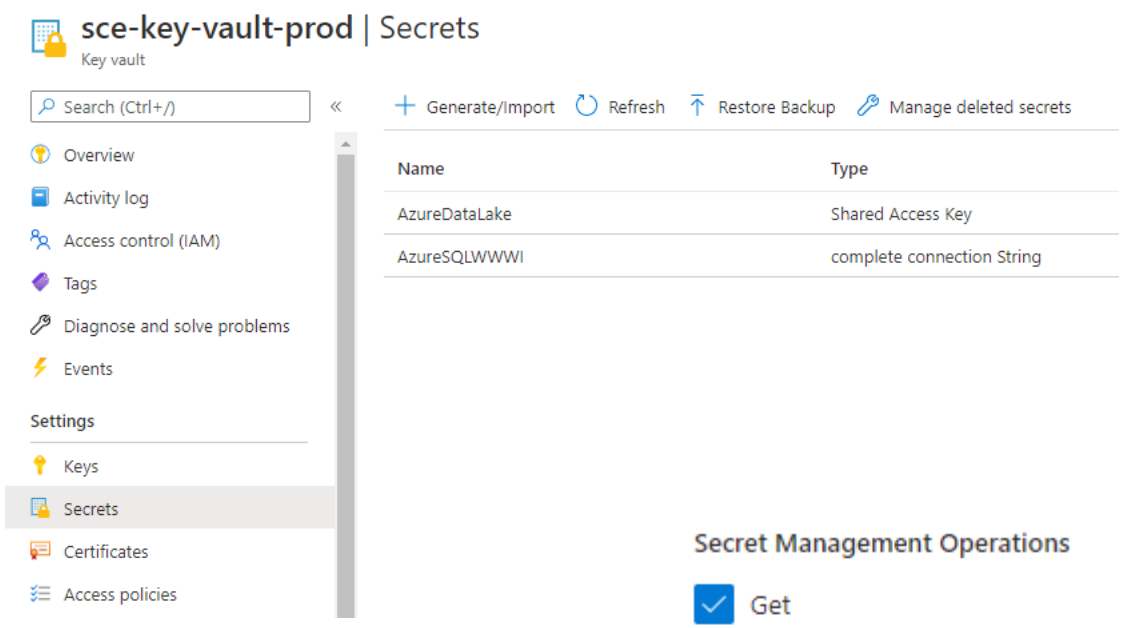


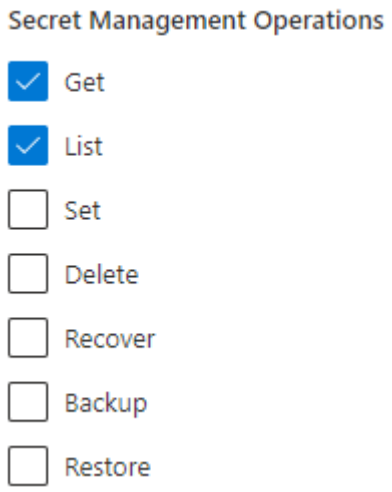
Azure Key Vault

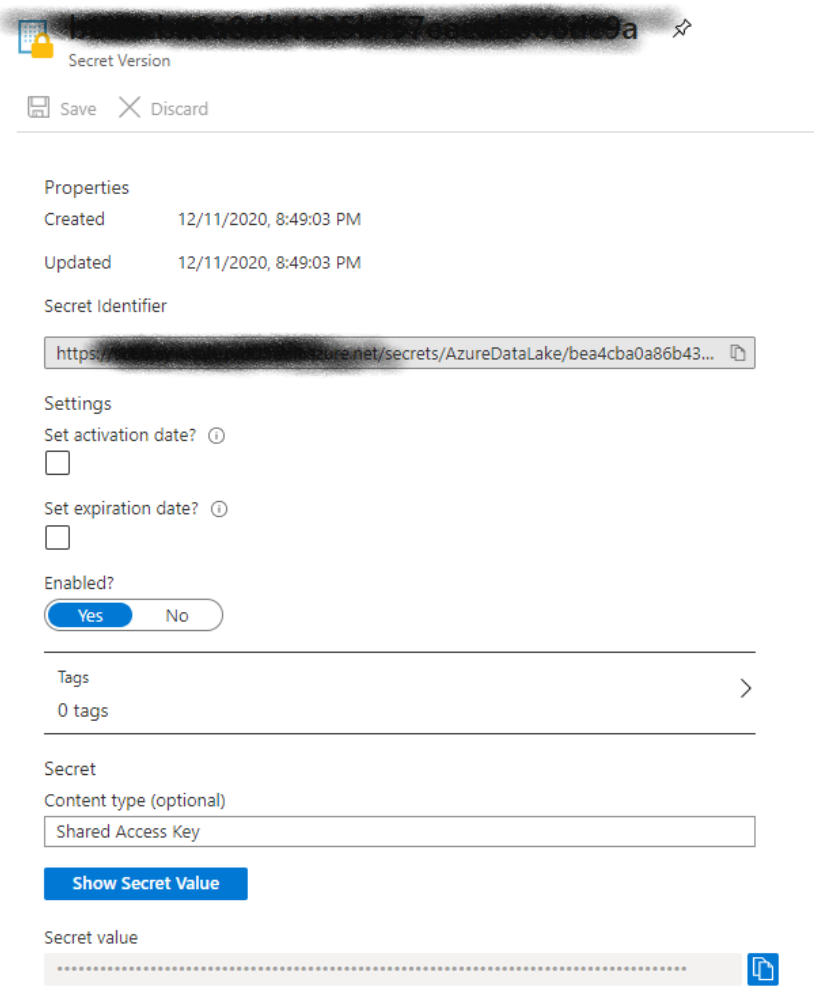


- **No passwords in code!**
- Use Key vault for any needed secrets in processes
  - Stores credentials, certificates in one save place
  - Password content hidden by default
  - Validity period and history available to support change workflows
  - Use it by day one in development
  - Use it for deployment automation
  - Use it in connections if possible

Azure Key Vault







Additional security features to use with care

- Azure Resource Locks – functionality breaking Databricks & Synapse Analytics Workspace
- Additional Infrastructure encryption – Performance Impact
- Customer Managed Keys – Management overhead
- Datalake: Soft Delete - Cost Impact
- Databricks: Encryption between worker nodes - Performance Impact

Conclusion & Takeaways

- Prefer technical users without passwords for processes
- Use Azure Key Vault as secure store for all secrets
- Permit Key Vault access to managed identities of services (ADF..)
- Minimize given permissions on Key Vault entries
- Use additional Security features with care


Authorization

Use Azure Active Directory as prefered Identity Manager in Azure

scieneers
DRIVEN BY DATA

- Identities
  - Users (employees, guests)
  - Service Principal (technical)
  - Groups
- are members of roles
  - Define allowed activities

are assigned to scopes

- Management Groups
  - Subscriptions
    - Resource Groups
      - Resources

- Single-SignOn by AAD should be first option for any access
  - Set services to use AAD only
  - Use Sync features of service e.g. Databricks Credential pass through or SCIM

About Azure AD Premium

P1

- Sync with on-prem AD
- dynamic groups
- self-service group management
- Microsoft Identity Manager
- cloud write-back capabilities
- Included in Office 365 Enterprise subscriptions

P2

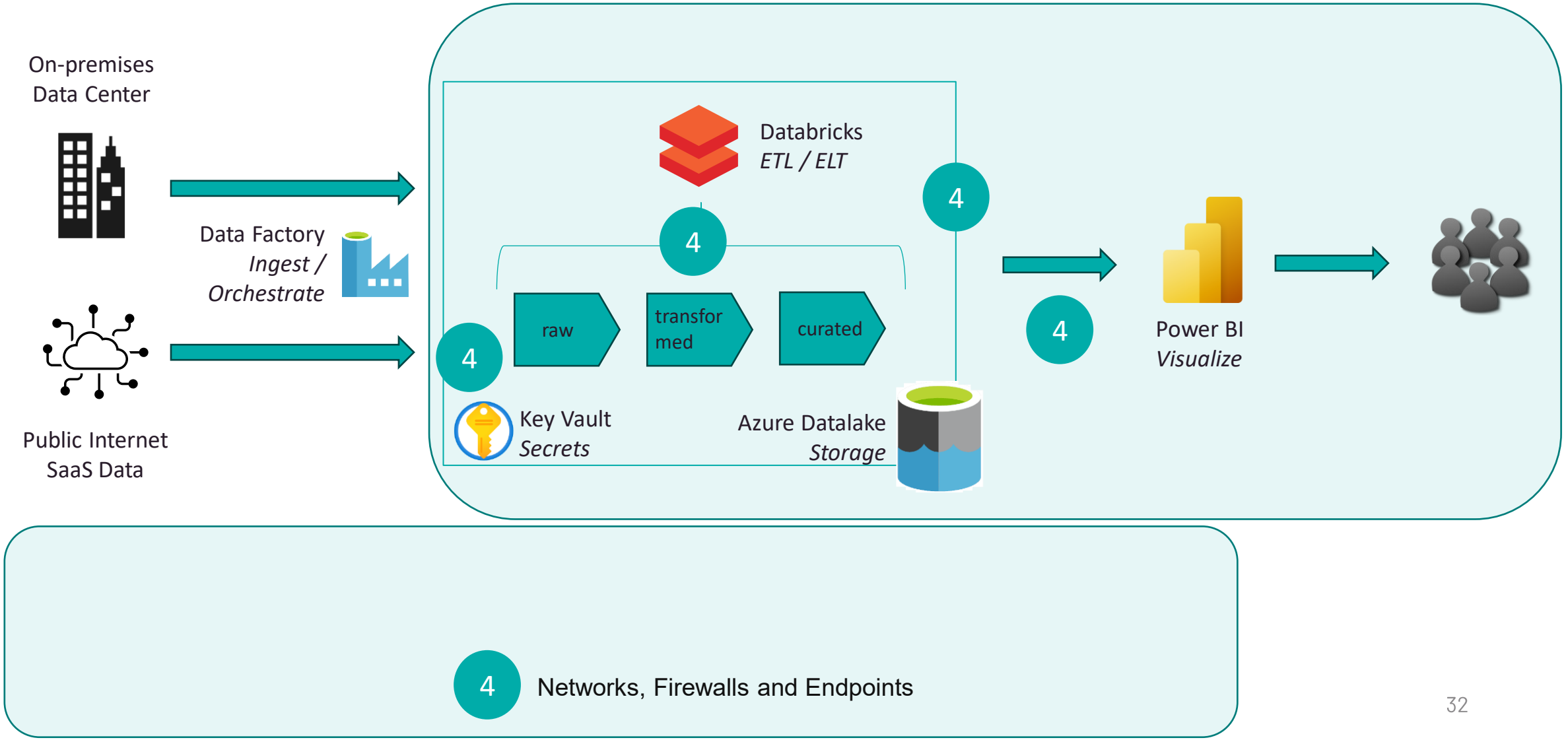
- All from P1 and
- AAD Identity Protection (risk based conditional access)
- Privileged Identity Management (restrict admins)

Conclusion & Takeaways

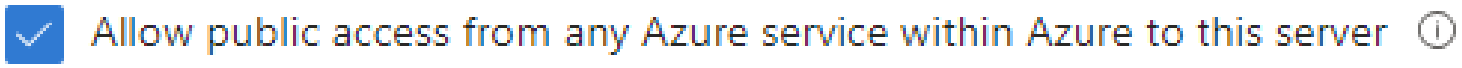
- Azure AD SSO preferred authentication mode for services
- RBAC preferred over other authorization mechanisms
- Groups preferred over Single User assignments
- Management Groups enable Multi-Subscription assignments
- Prefer Azure B2B for collaboration with other companies to reduce identity management
- Prefer technical users for processes in data applications
- Use personal identities (users) for data viewing and exploration
- Most companies get along using Azure AD P1


Networks, Firewalls and Endpoints

Azure Data Services – security topics

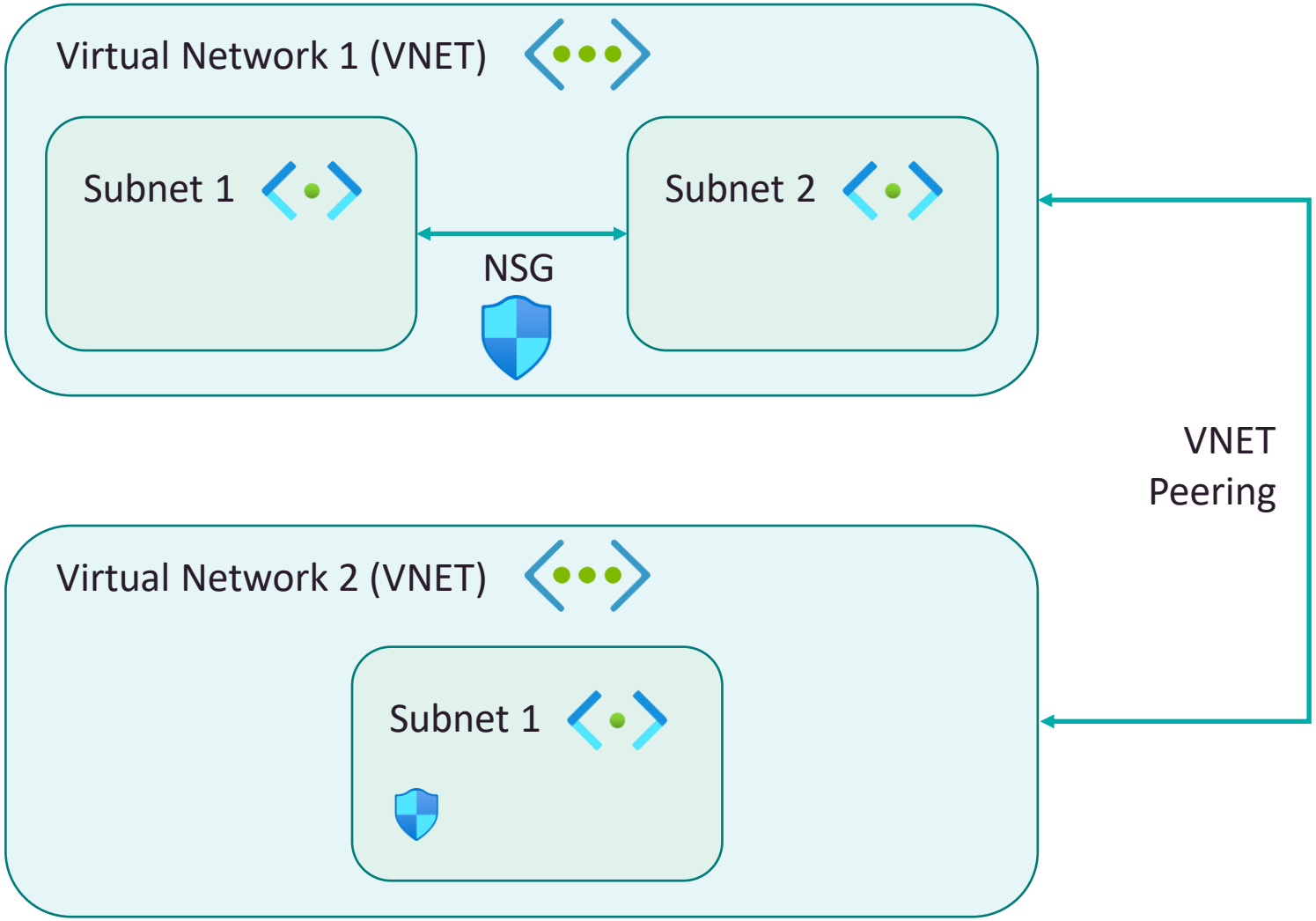


How to access data behind firewalls of PaaS services?



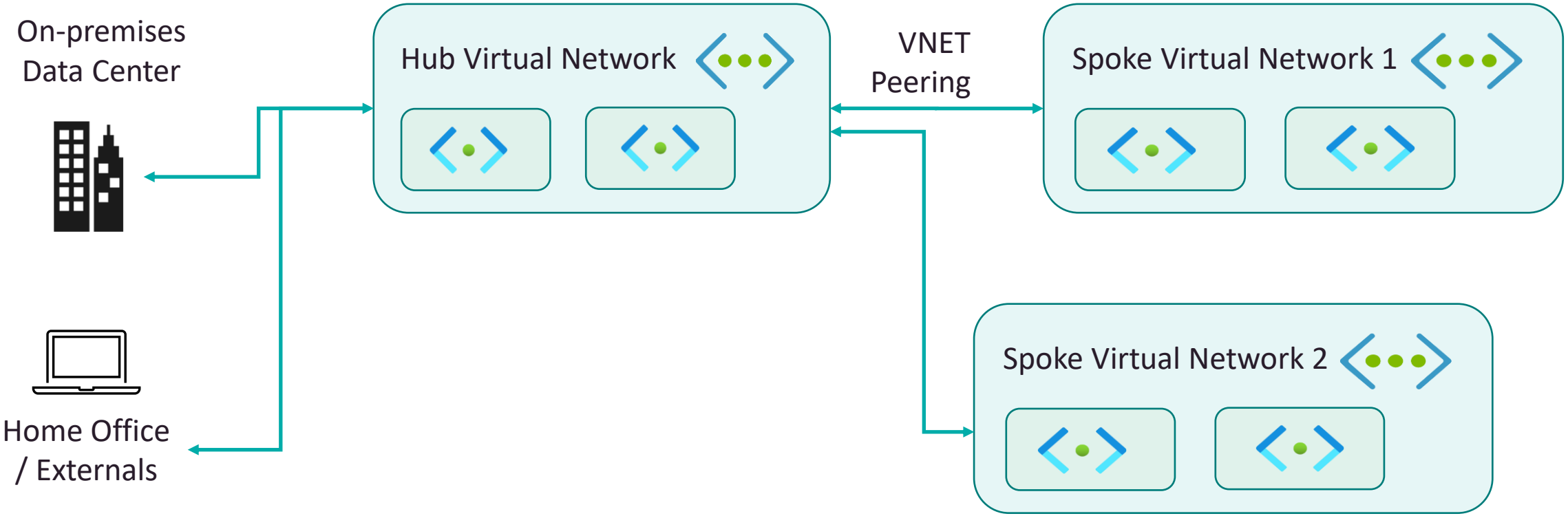
- Best solution?
  - Any service means any on Azure, not only yours!
  - Anyone could spin up a VM and try to access your data
  - Try does not mean anyone can, still needs authentication…
- Another topic is where the network traffic is routed
  - Microsoft backbone
  - Public Internet

Some Basic Network Structures on Azure



Grouped by countries, divisions, applications, services…

The common Spoke & Hub Architecture



But how is this network stuff build for PaaS?

- 1 (hidden) network per service with endpoint(s)
- Public endpoints are default for PaaS services – this could be a risk, some companys have rules which prohibit this
- Some Services use Managed VNETs
- Azure Virtual Network integration optional over private Link/Endpoint
  - service endpoint (enable technology to be connected e.g. sql db)
  - private endpoint – network interface for service
- Network routing options
  - Microsoft Backbone or Internet routing

How are PaaS Services which use public IPs protected by Azure?

scieneers
DRIVEN BY DATA

- Firewall per Service instance filter traffic by different criteria
  - Exceptions possible by IP address ranges, resource instances, general list of trusted resources (options variying by service)
- Endpoints secured by RBAC security (best case)
  - AAD identities are hardened by MFA / Certificates
  - Managed identities provide secured identity
- Endpoints using custom security are easier to attack
  - e.g. MySQL using custom users with weak password policy
- Stored data is encrypted on Azure by default
  - But different by services
- Usage of additional security features like Policies, Auditing, Defender and Sentinel to check for unusual events and activities make sense

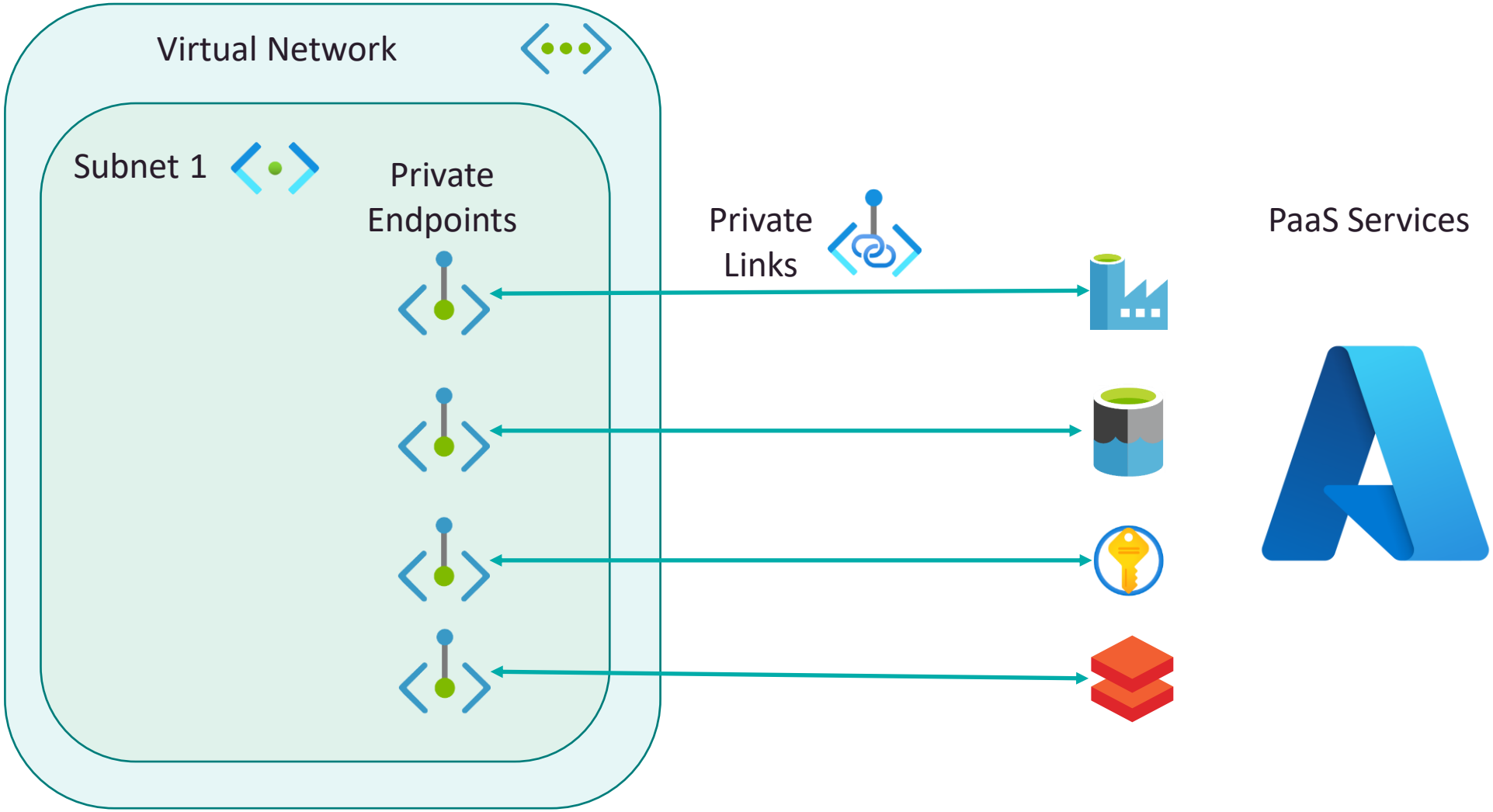
Why using private networking and endpoints also for PaaS Services?

- To secure traffic on data platform within organization avoiding public internet
- Avoid possible attacks on public endpoints
- For integration with other secured services on-prem and Azure (Hub & Spoke)

Why is using private networking / endpoints not a no-brainer?

- it adds some complicated overhead, due to
  - Different settings and options for any PaaS Service
  - Networking and firewalling know-how not „home-turf“ of data engineers or data scientists
  - Coordination with IT about IP-address spaces and firewall rules could be time-consuming
  - Additional costs and performance impact are to be considered
  - Connectivity restrictions when disabling public endpoint
    - Jumphost in VNET / Bastion Host for access (IaaS)
    - Deployment from Azure DevOps via Self-Hosted Agent (IaaS)

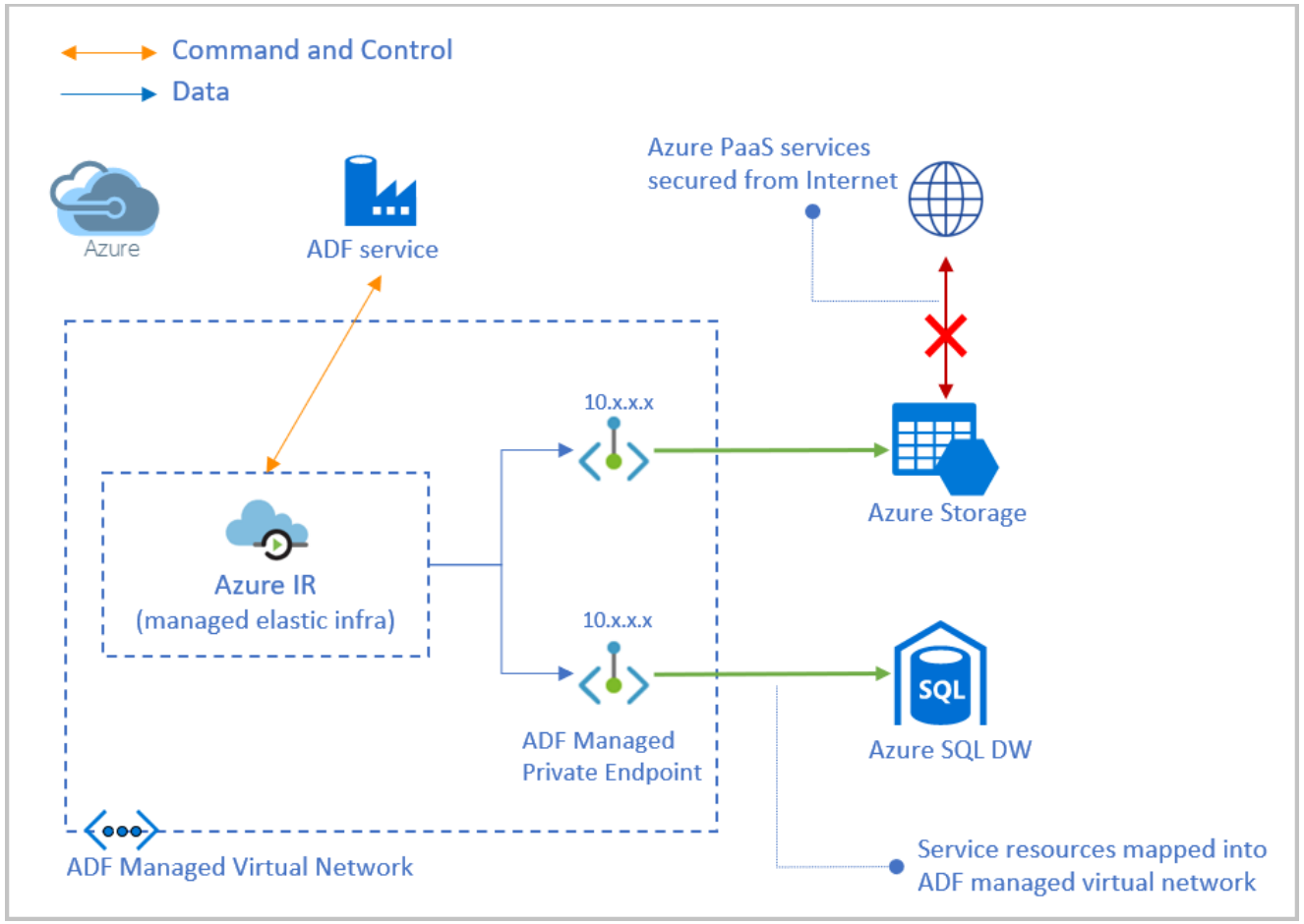
PaaS Integration using private links / endpoints



PaaS Integration using Private Endpoint in Managed VNET in Azure Data Factory

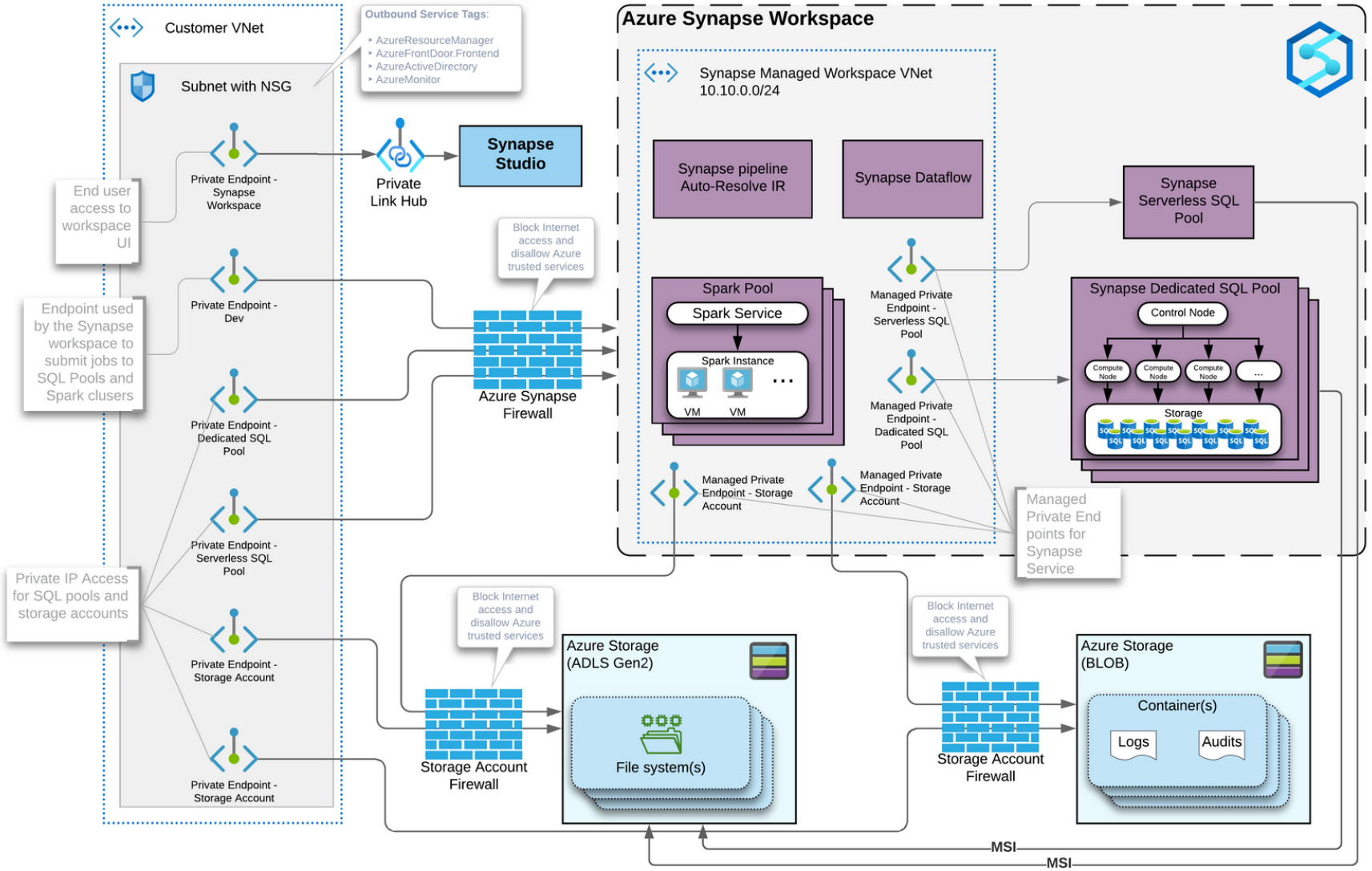
scieneers
DRIVEN BY DATA

- Primarly to offload the burden for management of virtual networks to ADF
- enables Private Endpoint Management in ADF to create outbound connections
- Network stuff less configurable due to „managed“



https://learn.microsoft.com/en-us/azure/data-factory/managed-virtual-network-private-endpoint

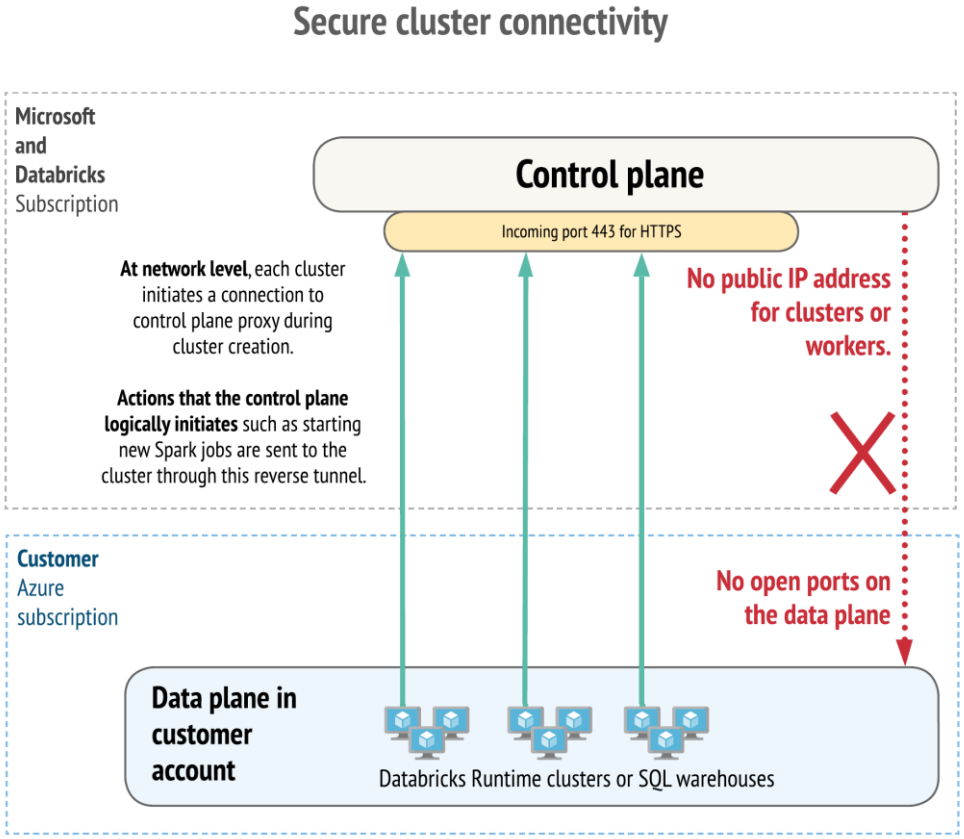
Synapse Managed networking overview – a lot of services add complexity and tons of private endpoints



https://techcommunity.microsoft.com/t5/azure-architecture-blog/understanding-azure-synapse-private-endpoints/ba-p/2281463?lightbox-message-images-2281463=274870iAF555B44999C3EC7

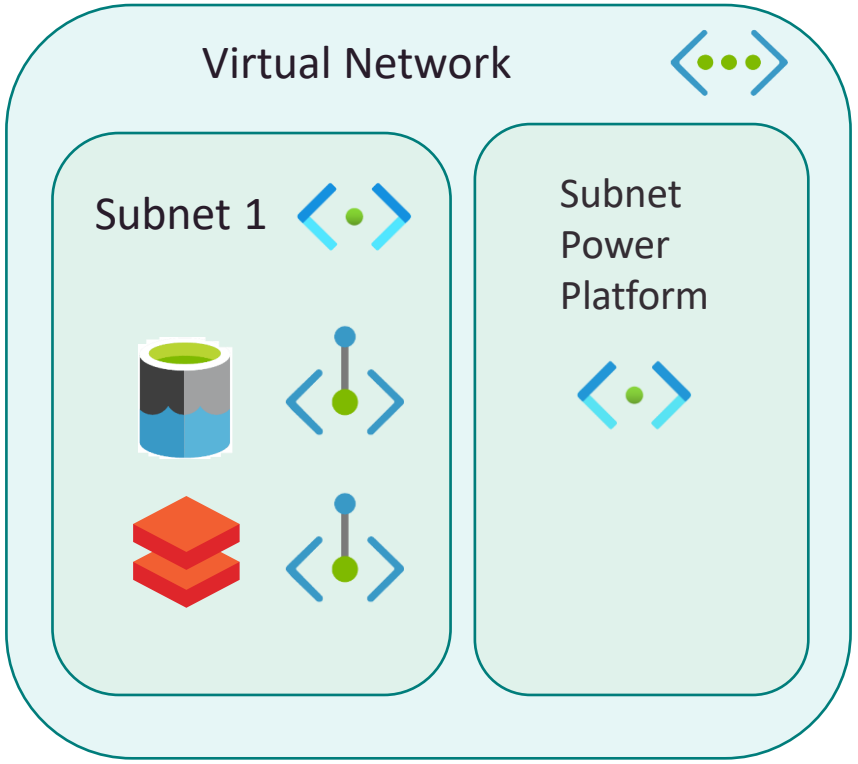
Azure Databricks networking security options

- Any databricks [internal] traffic goes over Microsoft network backbone – not the public internet (but for not serverless option)
- Secure Cluster Connectivity for additional **encryption between data and control** plane
- 2 Options for Egress connectivity
  - Default Managed VNET
  - Vnet injection (custom VNET)
- Also additional **encryption between cluster worker nodes** possible



https://learn.microsoft.com/en-us/azure/databricks/security/network/secure-cluster-connectivity

How to access Azure Data secure from Power BI?

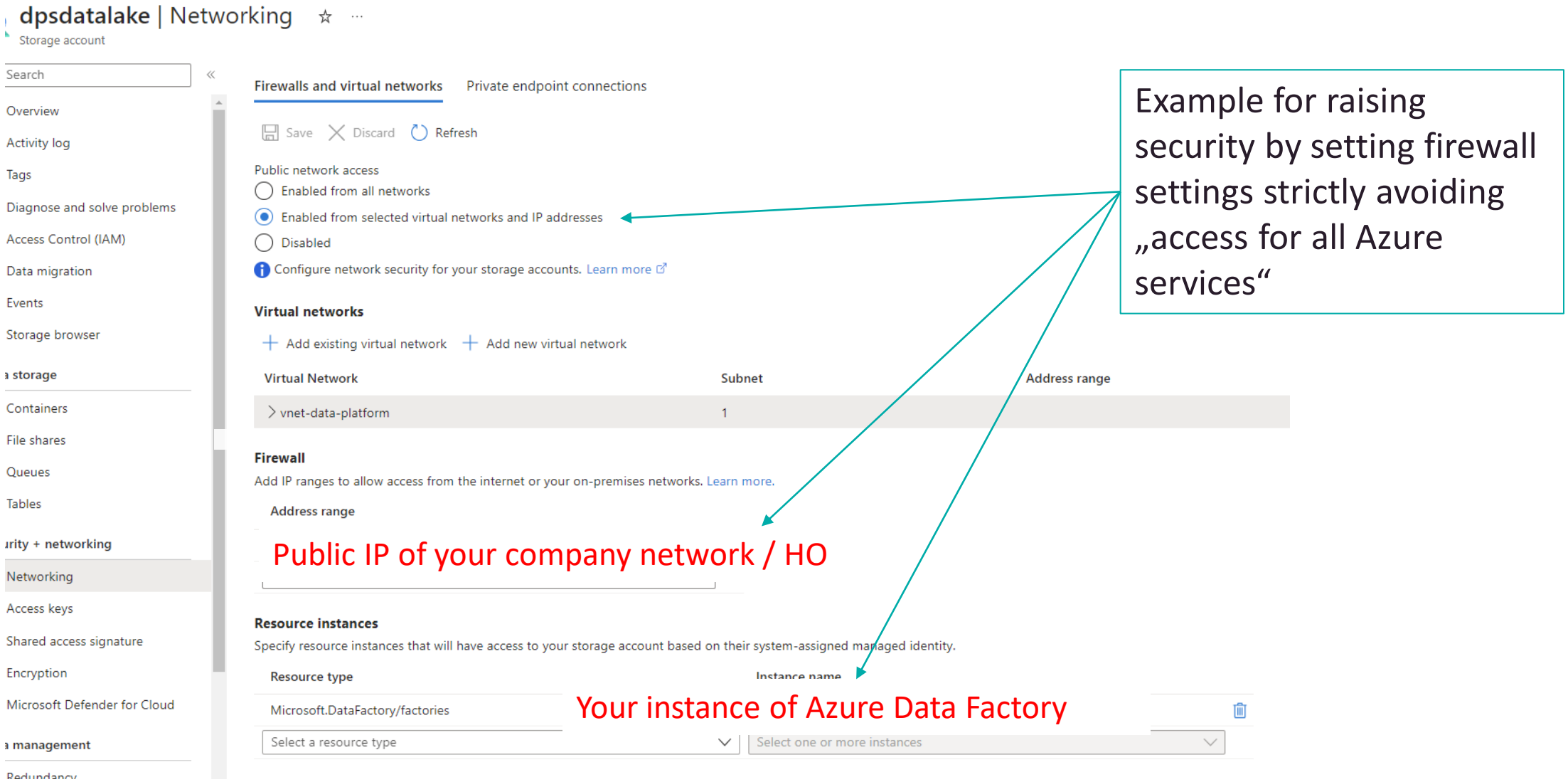




You have only Power BI Pro?
Use Power BI Gateway on VM in VNET

Good step-by-step guide here:
https://www.datahai.co.uk/power-bi/connecting-power-bi-to-azure-sql-database-using-private-endpoints/

Also any KISS Solutions possible?



Example for raising security by setting firewall settings strictly avoiding „access for all Azure services“

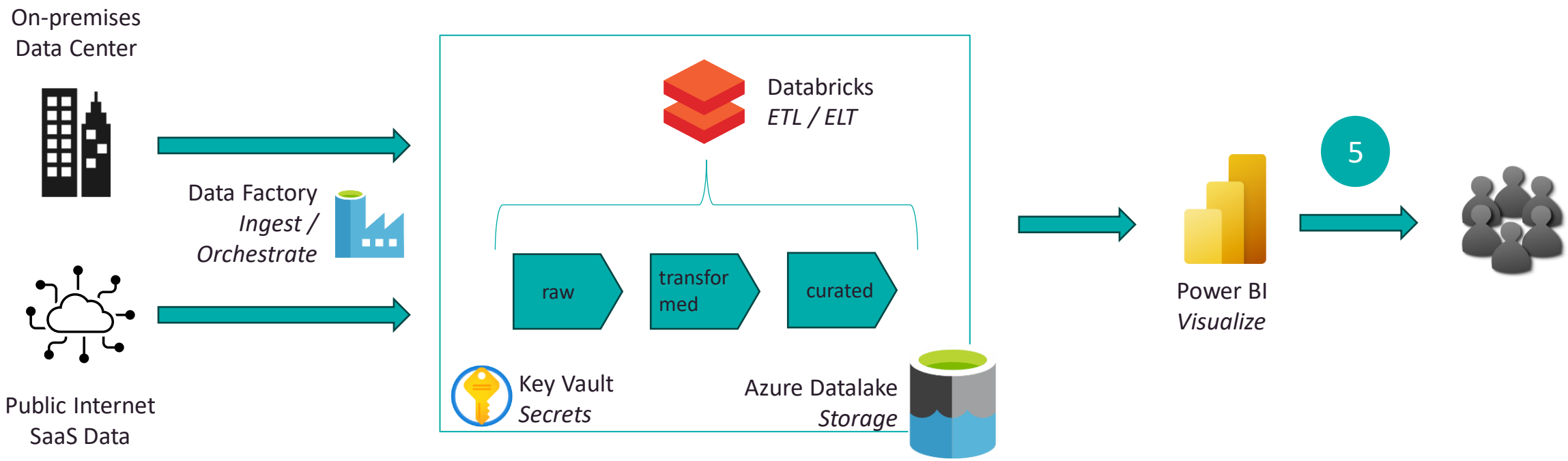
Conclusion & Takeaways

- Decision about routing of network traffic (public internet vs. MS backbone / Express Route)
- Check each relevant PaaS service‘s options for networking
- Using RBAC as default for permission management reduces risks when using public endpoints
- Limit allowed access in firewall settings for public endpoints (or disable them)
- Optional limit connectivity to specific networks instead of public internet by firewall using private endpoints, but be aware of the overhead / complexity



Data view permissions

Azure Data Services – Data view permissions



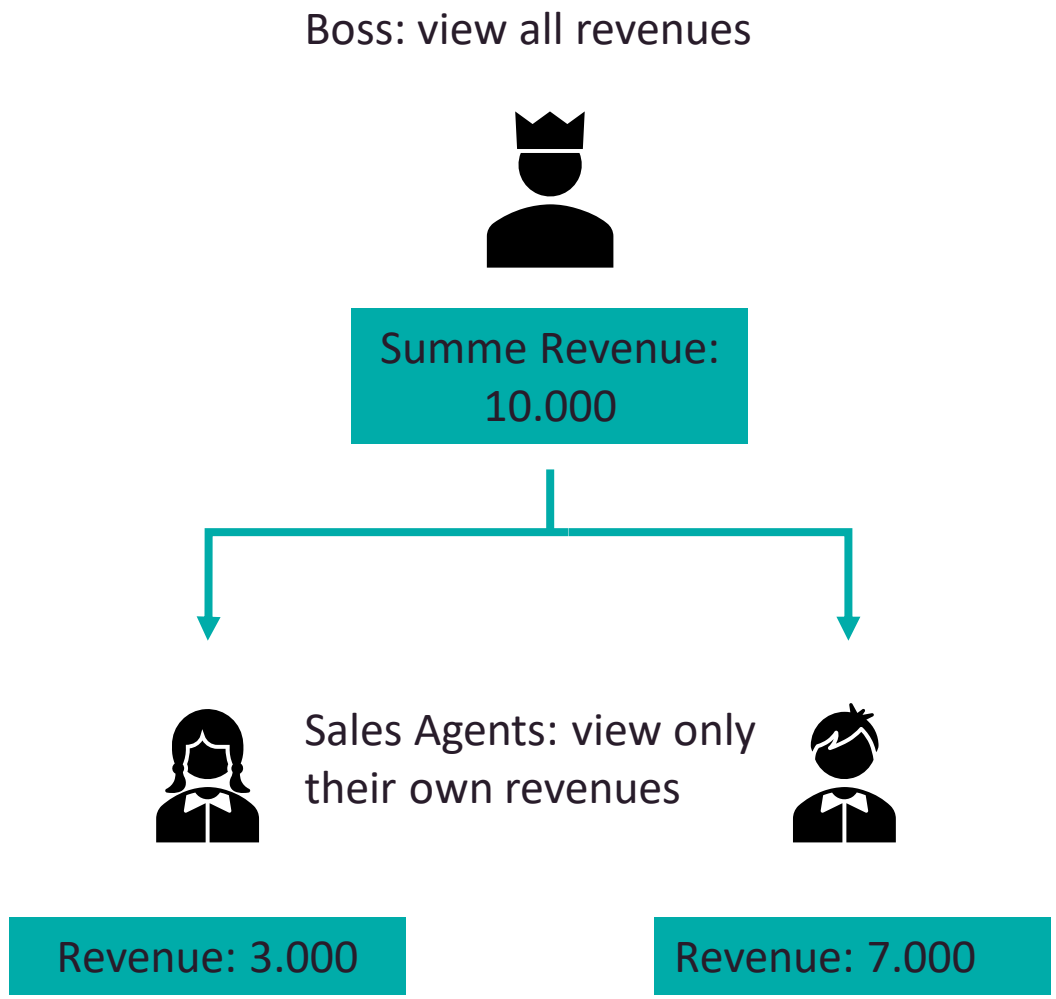
5 Data view permissions

First differentation – Object vs. Data level based Security

- Samples for permissions on object level
  - Reader permission for Power BI Report, App or Workspace
  - Read, Write Permission in whole Database or Schema
  - Role based Access Storage Blob Data Contributor for Data Lake in Azure
- Samples for permissions on data level
  - Sales person could only see sales data for regions he is responsible for
  - Sales area manager could see data for all regions of his employees
  - The CEO sees all data for all regions

Data view permissions

- Delivering value means delivering data in our context
- Filtering values predicate based could be implemented in different services using diverse concepts
- Static or dynamic definition of filters



Permission technologies in data services

Only the most relevant



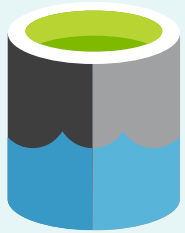
Power BI / Tabular

Row-Level Security on models in PBI service



Azure SQL / Synapse

Row-Level Security in relational database



Azure Data Lake

Hierarchical permissions on folders

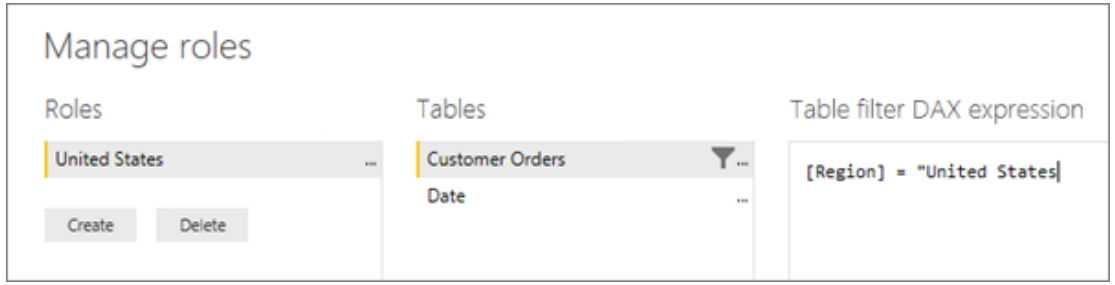


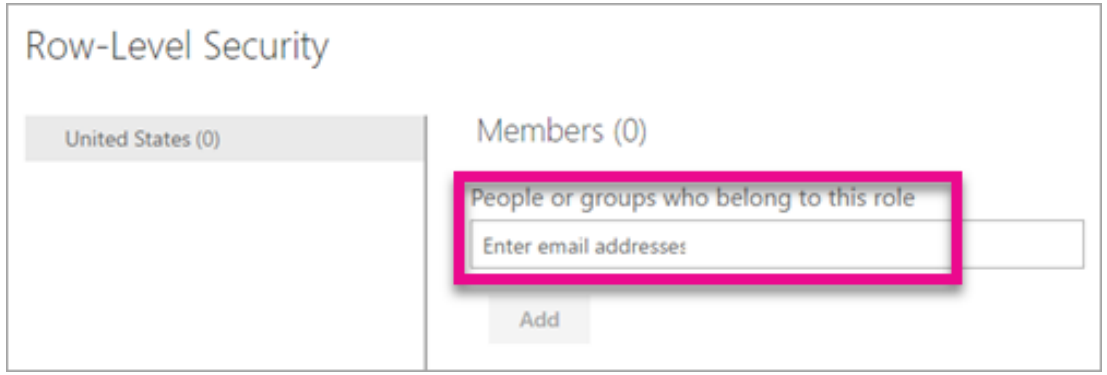
Azure Databricks

Row-Level Security in Delta Lake

Row-Level Security in Power BI

- Different roles possible
- Rules defined in DAX in Power BI Desktop, model deployed to service
- Current User identified by userprincipalname() function
- View as role in PBI Desktop
- Azure AD & guest users (B2B) possible
- Same technology used in Azure Analysis Services





https://docs.microsoft.com/en-us/power-bi/admin/service-admin-rls

Conclusion & Takeaways data view permissions

- If object security is sufficient this will keep efforts low and reduces complexity (KISS)
- Best place to implement row-level security depends on requirements and tooling.
  - Less effort in Power BI but avoid duplication of code (DRY)
  - Backend implemenation enables usage auf RLS by other frontends
  - But Direct vs. Import Mode in Power BI to be considered



Summary

Summary Azure Data Platform Security

- Connectivity on-prem / Azure by SHIR (fast track) or VPN for deeper integration
- Prefer technical users and use Azure Key Vault for secrets
- Azure AD and RBAC as first choice for authentication and authorization
- Clarify networking requirements early and collaborate with IT
- KISS principle also valid for data view permissions

How about your data platform on Azure?

Feel free to check out a security assessment with scieneers



Data is valueable.
Data needs security!

A data application built on Azure PaaS components like Data Lake, Data Factory or SQL Azure can be set up in a short time. Thanks to various documentation available, even beginners can build great functionality with it. A secure baseline and rule set is mandatory for internet exposed services and live data, but it's often missing or incomplete in real world projects or planned to be done later.

Since much data is still stored on premises the connectivity to load this data secured to Azure is one of the first and most important questions to be answered for any project. Other themes are the proper handling of secrets in a public cloud, how to secure network connectivity between the components and implementing data viewing thus everyone sees only his data.

We want that your data is secure on Azure!

Read more

scieneers
DRIVEN BY DATA



Thanks for your attention, I appreciate your feedback!