

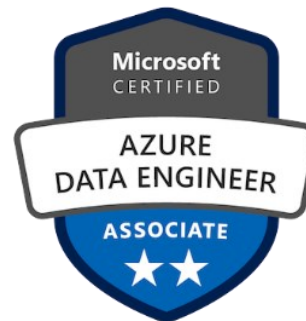
Build your data platform on Azure

...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
- Authorization
- 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

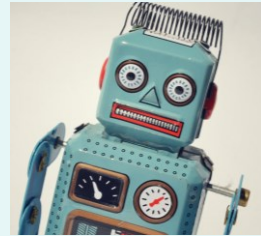
permissions:
Prefer function
based groups

General Security Rules



Monitoring

Automate testing
and
security audits



Automate processes

Code to replace
frequent manual
processes

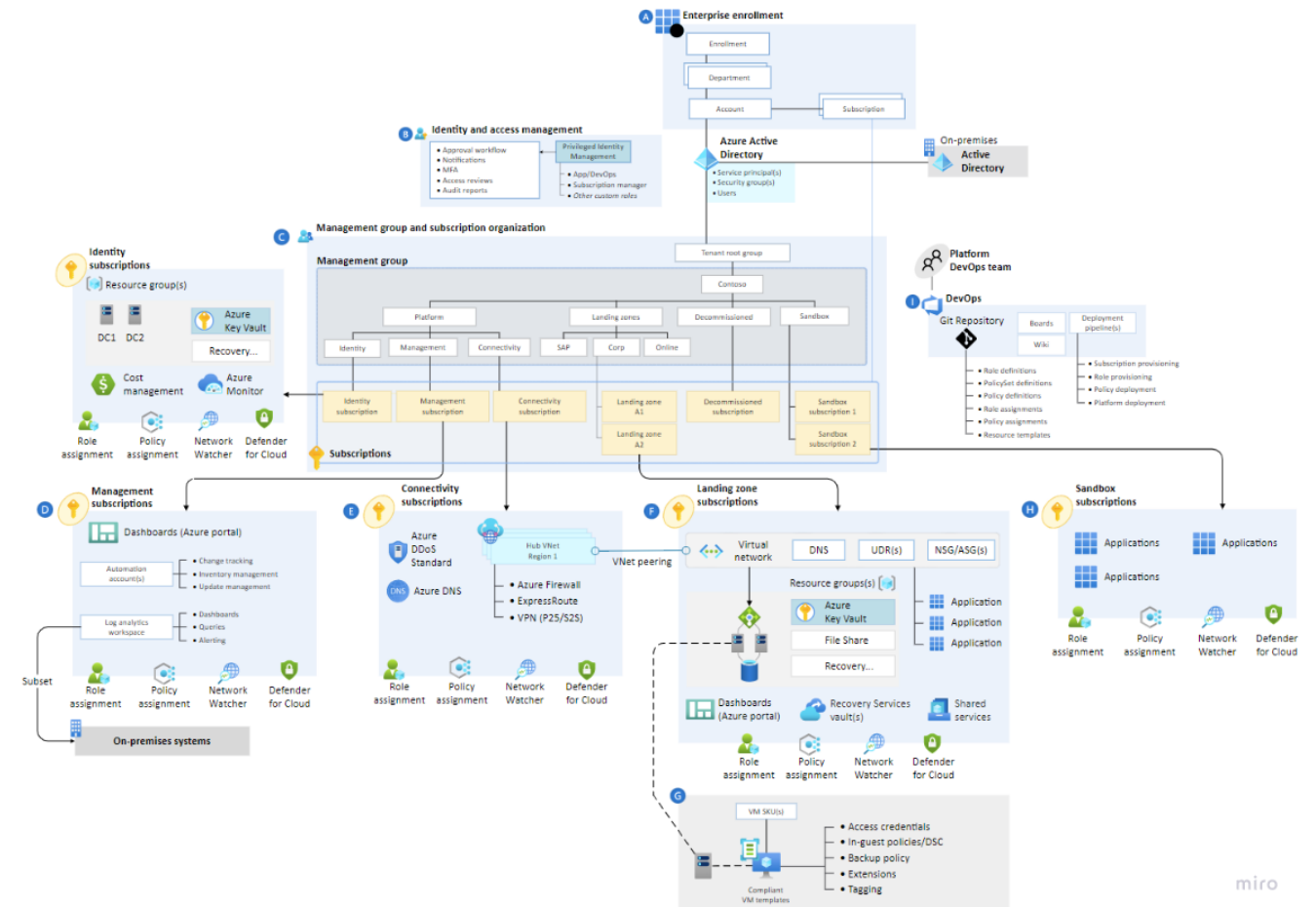


Documentation for transparency

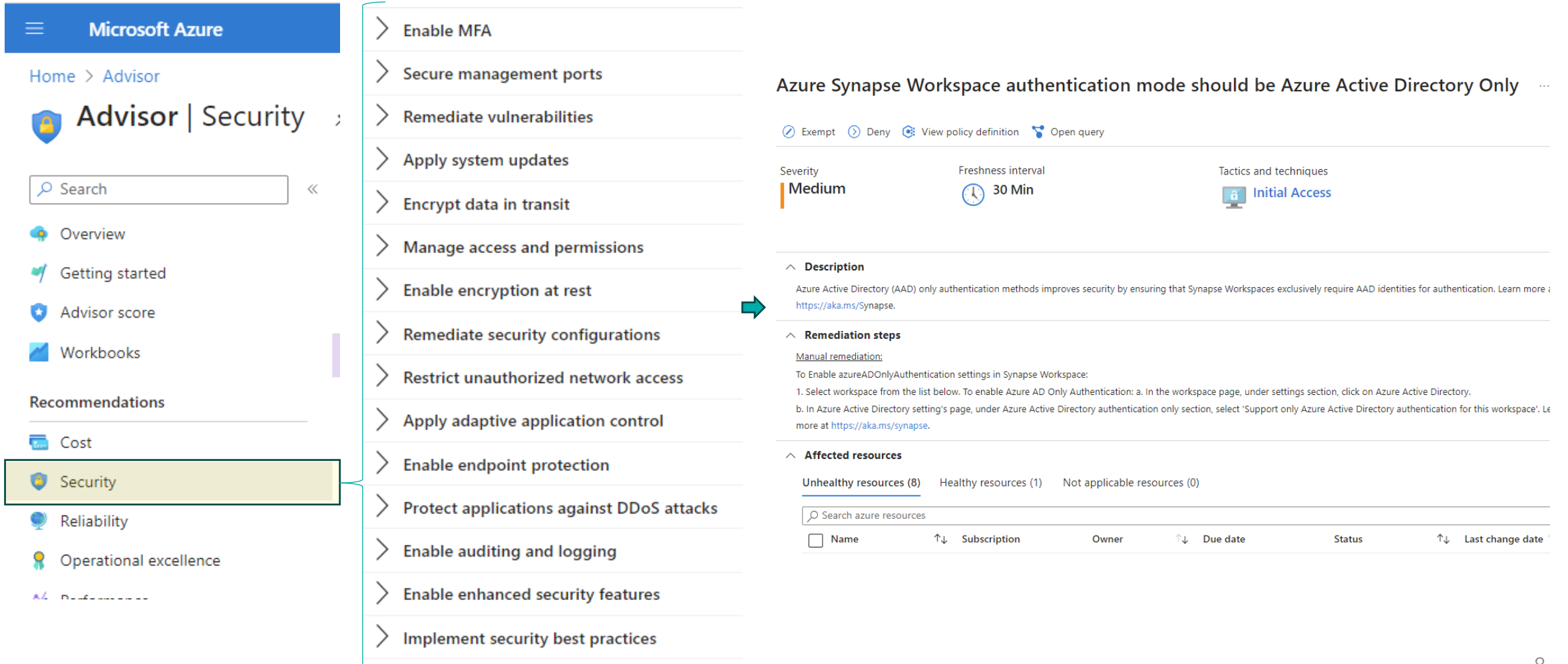
Focus on
concepts
explanation of
decisions

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)



The screenshot displays the Microsoft Azure Security Advisor interface. On the left, the navigation pane shows the 'Security' category selected under 'Recommendations'. The main content area is divided into two sections: a list of recommendations on the left and a detailed view of a specific issue on the right.

Recommendations List:

- Enable MFA
- Secure management ports
- Remediate vulnerabilities
- Apply system updates
- Encrypt data in transit
- Manage access and permissions
- Enable encryption at rest
- Remediate security configurations
- Restrict unauthorized network access
- Apply adaptive application control
- Enable endpoint protection
- Protect applications against DDoS attacks
- Enable auditing and logging
- Enable enhanced security features
- Implement security best practices

Issue Detail View:

Azure Synapse Workspace authentication mode should be Azure Active Directory Only

Exempt Deny View policy definition Open query

Severity: **Medium** Freshness interval: 30 Min Tactics and techniques: Initial Access

Description: Azure Active Directory (AAD) only authentication methods improves security by ensuring that Synapse Workspaces exclusively require AAD identities for authentication. Learn more at <https://aka.ms/Synapse>.

Remediation steps:

Manual remediation:

To Enable azureADOnlyAuthentication settings in Synapse Workspace:

- Select workspace from the list below. To enable Azure AD Only Authentication: a. In the workspace page, under settings section, click on Azure Active Directory.
- In Azure Active Directory setting's page, under Azure Active Directory authentication only section, select 'Support only Azure Active Directory authentication for this workspace'. Learn more at <https://aka.ms/synapse>.

Affected resources: Unhealthy resources (8) Healthy resources (1) Not applicable resources (0)

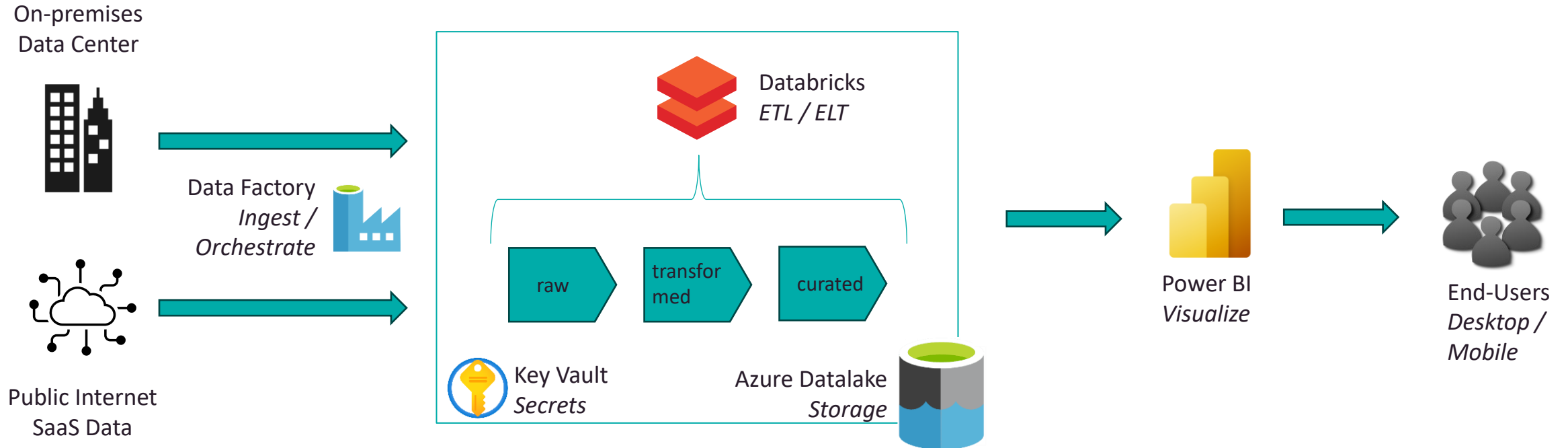
Search azure resources

Name	Subscription	Owner	Due date	Status	Last change date
------	--------------	-------	----------	--------	------------------

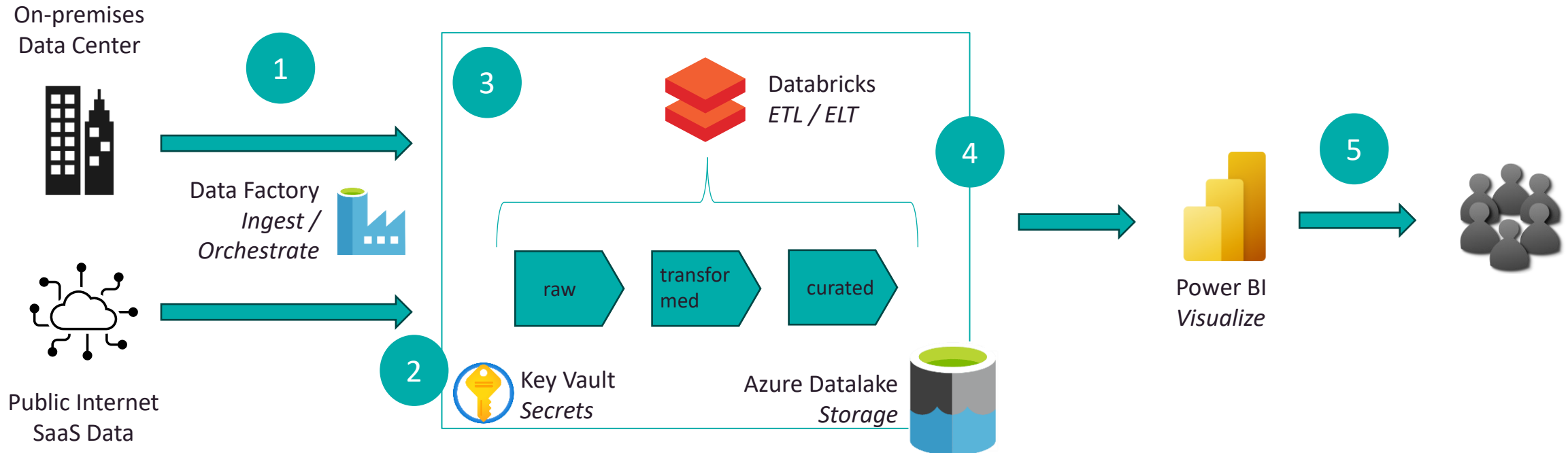
A 3D geometric structure composed of colorful sticks (yellow, green, blue, red) and yellow connectors, arranged in a complex, interconnected pattern on a blue background. The structure consists of various polygons, including triangles and quadrilaterals, connected at their vertices by small yellow plastic connectors. The sticks are of different colors and are arranged in a way that creates a sense of depth and complexity. The background is a solid, light blue color.

Data Platform architectures and security topics


Azure Data Services – a typical setup



Azure Data Services – security topics

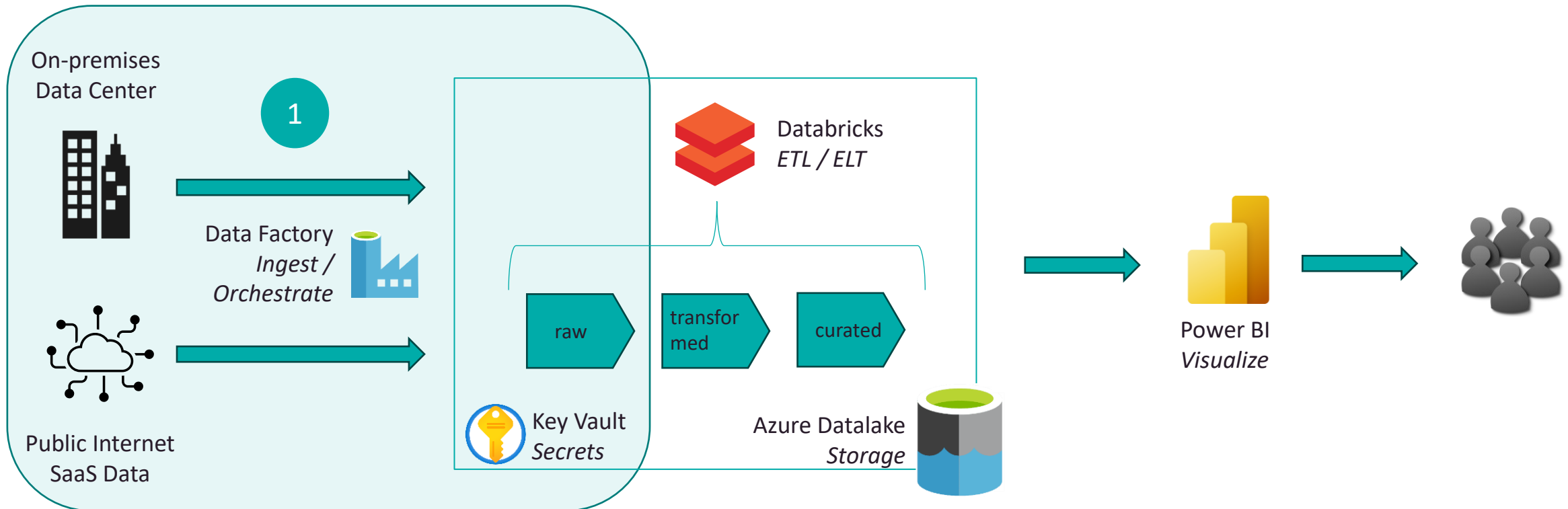


- | | | |
|--|--|--------------------------------|
| 1 Safe connectivity on-prem / Azure | 3 Authorization | 5 Data view permissions |
| 2 Keeping Secrets Secret | 4 Networks, Firewalls and Endpoints | |



Safe connectivity
on-prem / Azure

Azure Data Services – safe connectivity



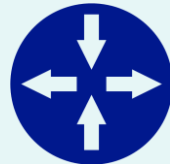
1 Safe connectivity on-prem / Azure

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

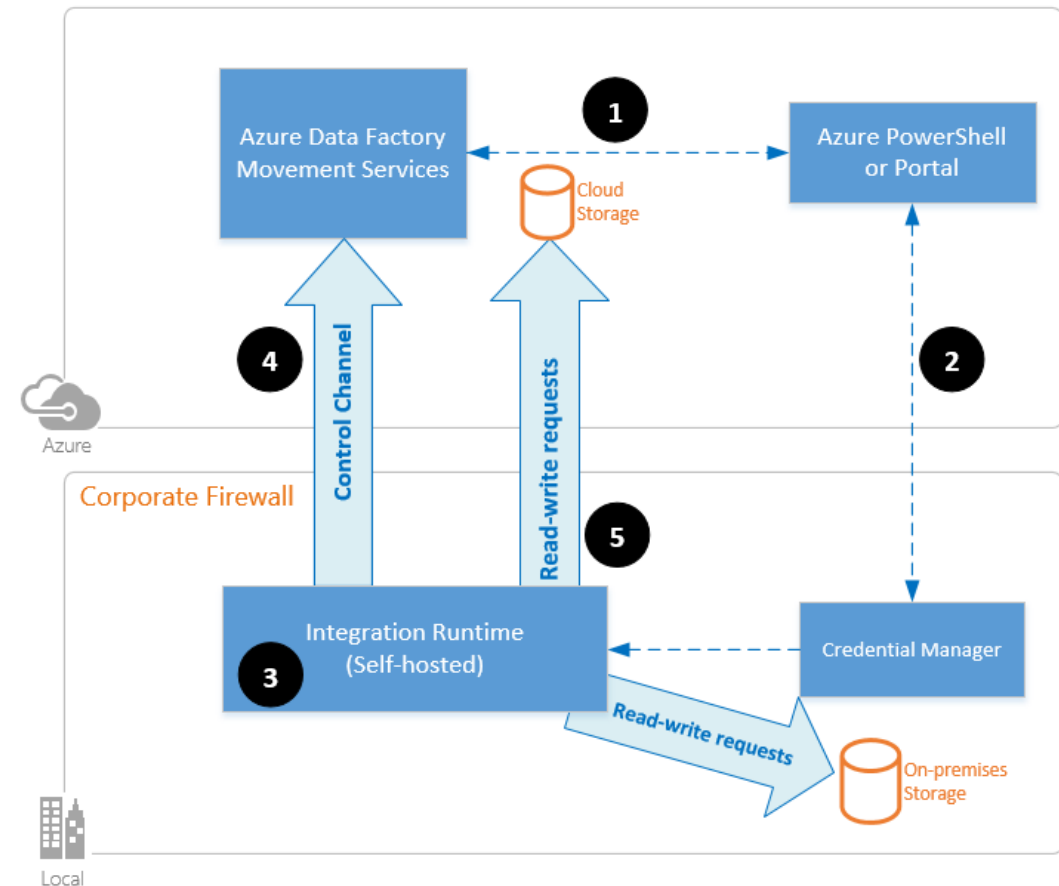


**Virtual Private
Network**

Connect Azure to
on-prem
permanently by
site2site VPN
tunnels

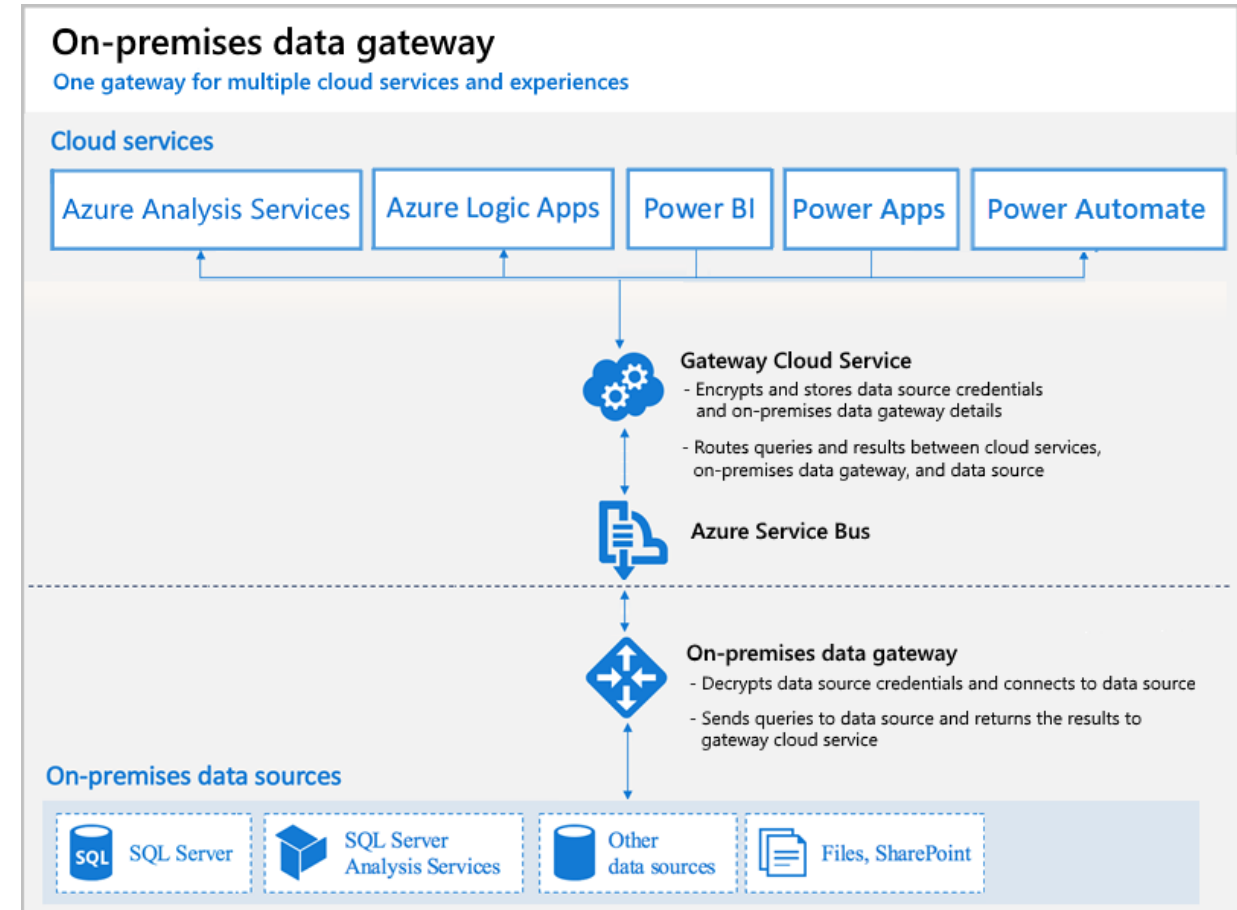
Self-hosted Local Integration Runtime

- 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



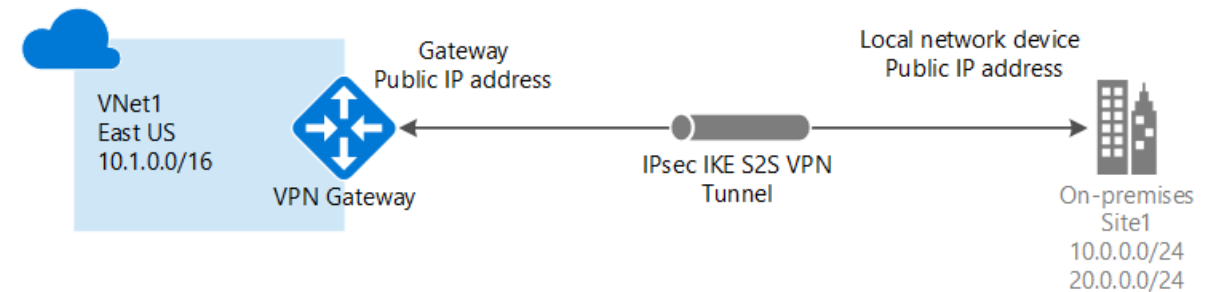
On-premises data gateway

- 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>

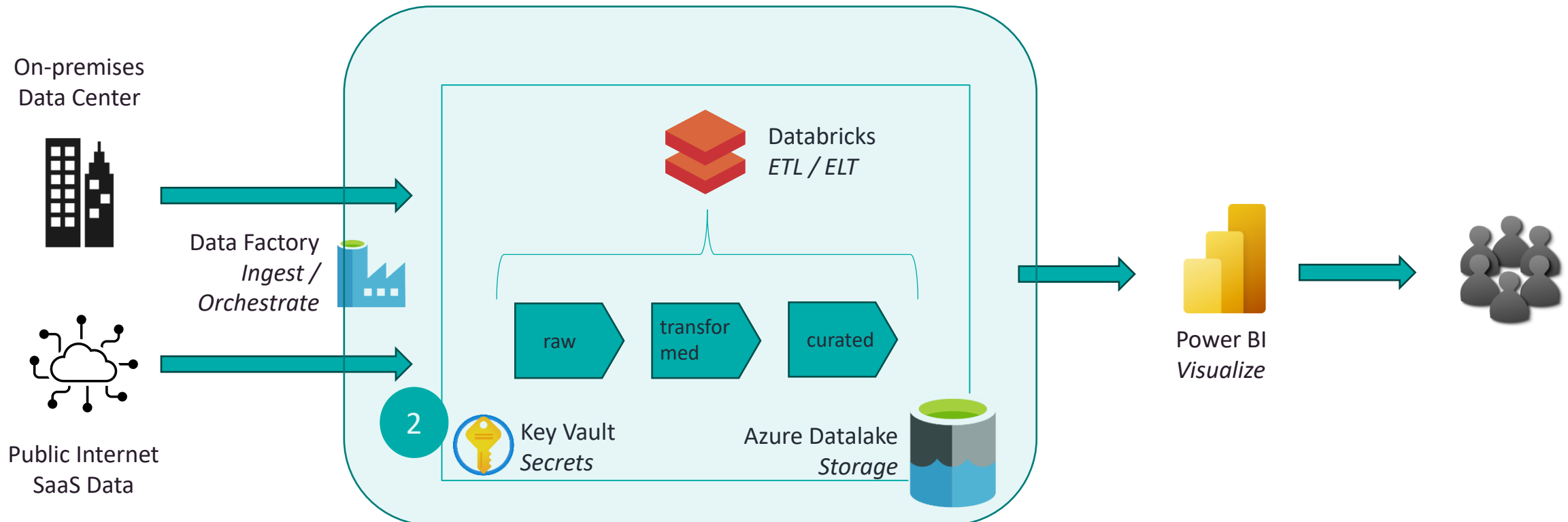
Conclusion & Takeaways

- 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

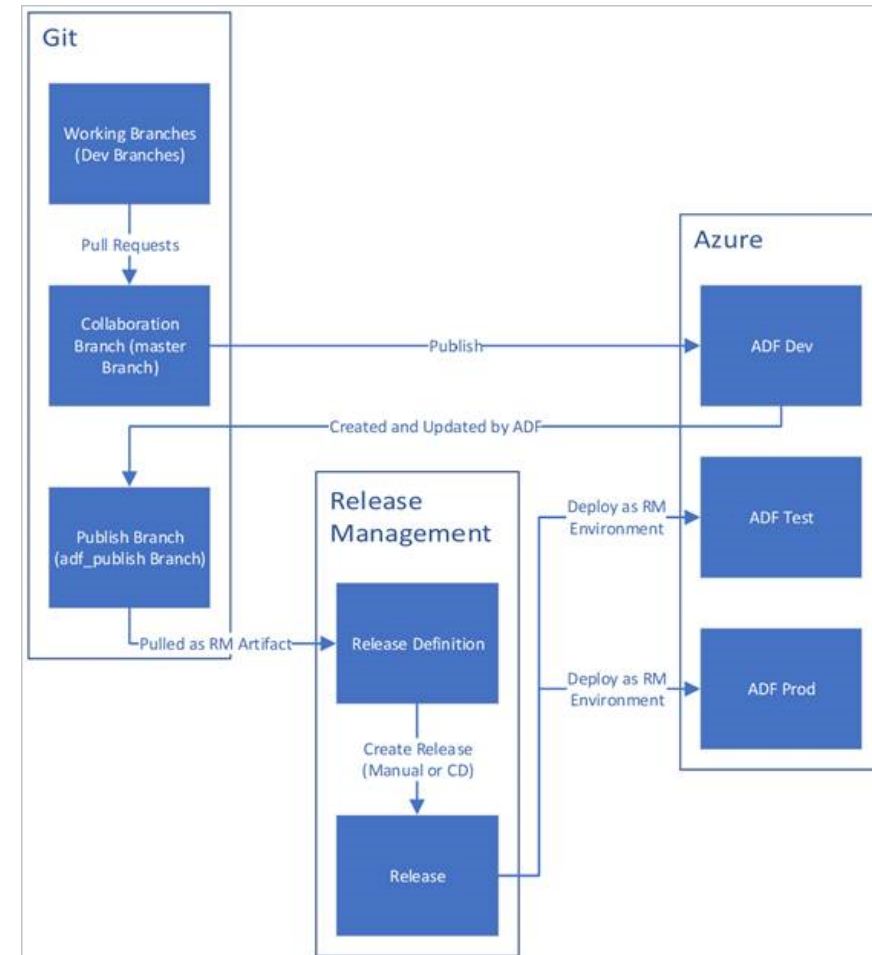
Azure Data Services – Keeping secrets secret



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

sce-key-vault-prod | Secrets

Key vault

Search (Ctrl+)

Generate/Import Refresh Restore Backup Manage deleted secrets

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Events
- Settings
- Keys
- Secrets**
- Certificates
- Access policies

Name	Type
AzureDataLake	Shared Access Key
AzureSQLWWWI	complete connection String

Secret Management Operations

- Get
- List
- Set
- Delete
- Recover
- Backup
- Restore

Secret Version

Save Discard

Properties

Created 12/11/2020, 8:49:03 PM
Updated 12/11/2020, 8:49:03 PM

Secret Identifier

https://.../secrets/AzureDataLake/bea4cba0a86b43...

Settings

- Set activation date?
- Set expiration date?

Enabled?

Yes No

Tags
0 tags

Secret

Content type (optional)

Shared Access Key

Show Secret Value

Secret value

.....

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

A red velvet rope stanchion with a gold ball top, set against a dark background with bokeh lights. The stanchion is made of polished metal and has a red velvet rope draped over it. The background is dark with several out-of-focus light sources, creating a bokeh effect. The word "Authorization" is written in white text in the lower-left corner.

Authorization

Use Azure Active Directory as preferred Identity Manager in Azure

- 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)

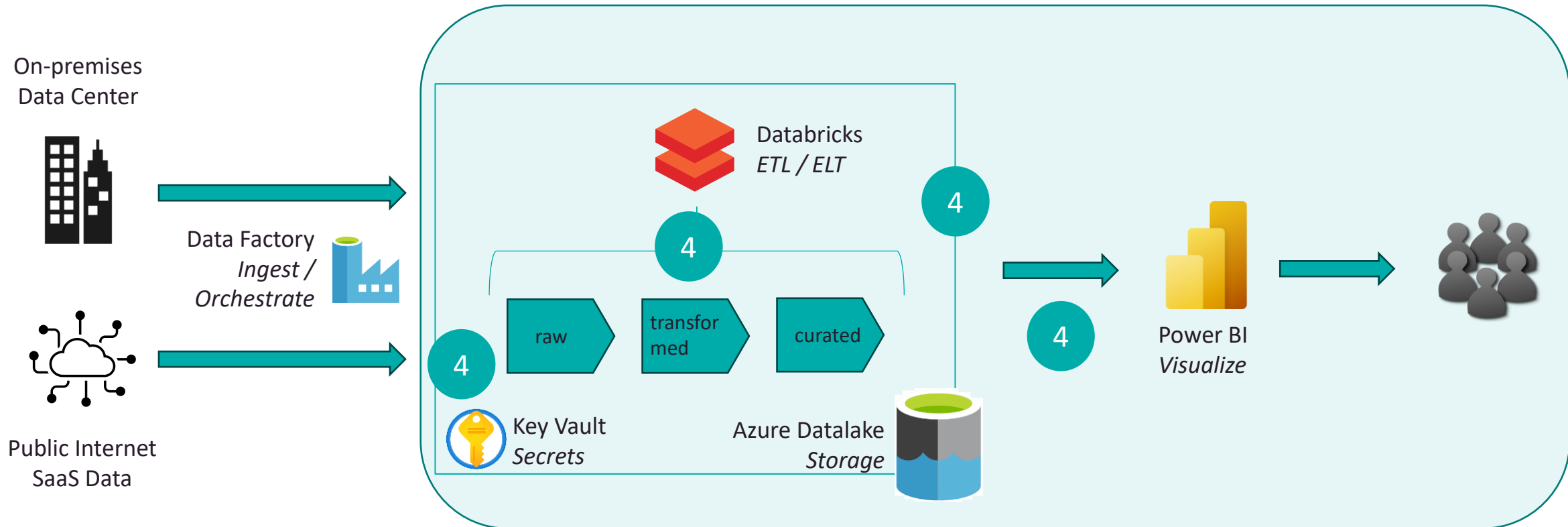
Conclusion & Takeaways

- 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
- Use Azure B2B for collaboration
- Prefer technical users for processes in data applications
- Most companies get along using Azure AD P1

Networks, Firewalls and Endpoints



Azure Data Services – security topics



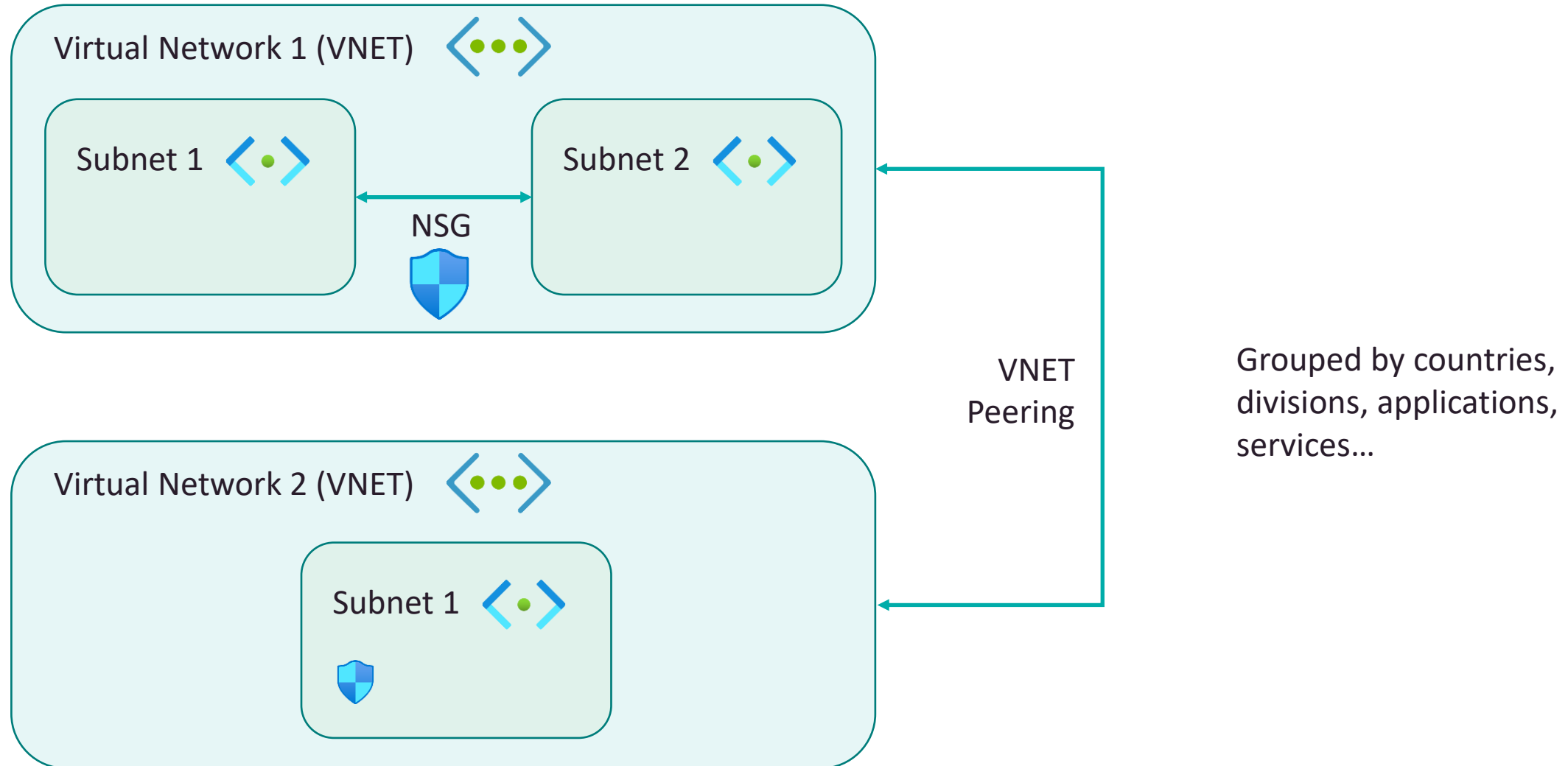
4 Networks, Firewalls and Endpoints

How to access data behind firewalls of PaaS services?

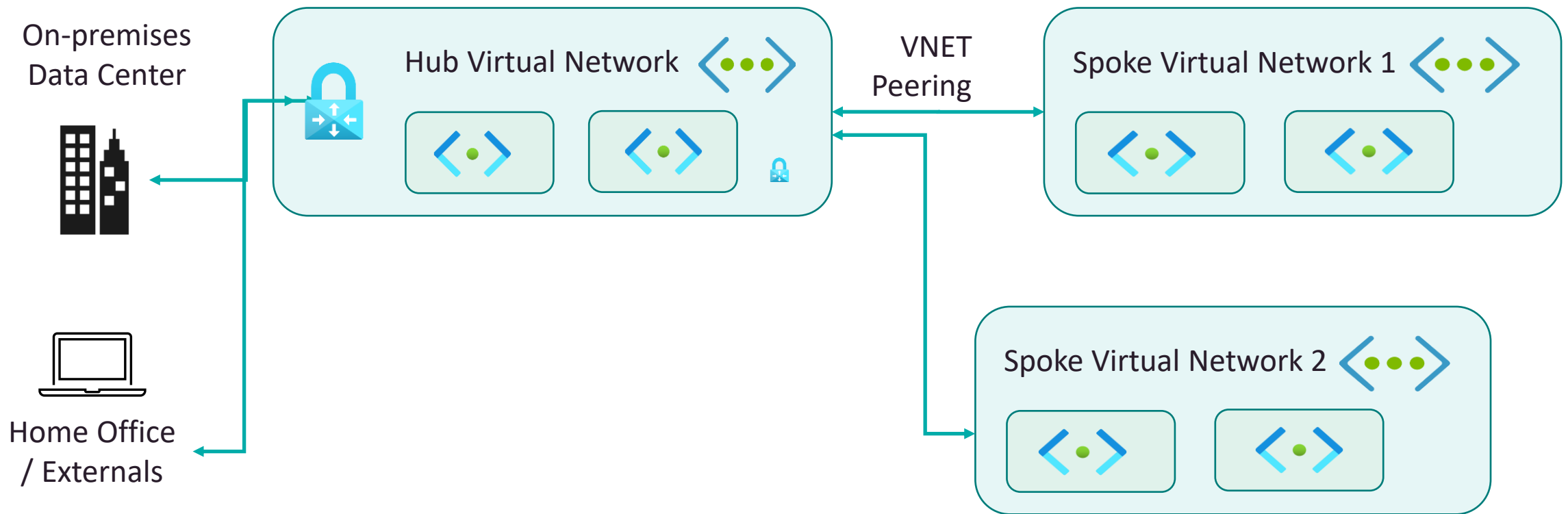
Allow public access from any Azure service within Azure to this server ⓘ

- 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



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?

- Firewall per Service instance filter traffic by different criteria
 - Exceptions possible by IP address ranges, resource instances, general list of trusted resources (options varying 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

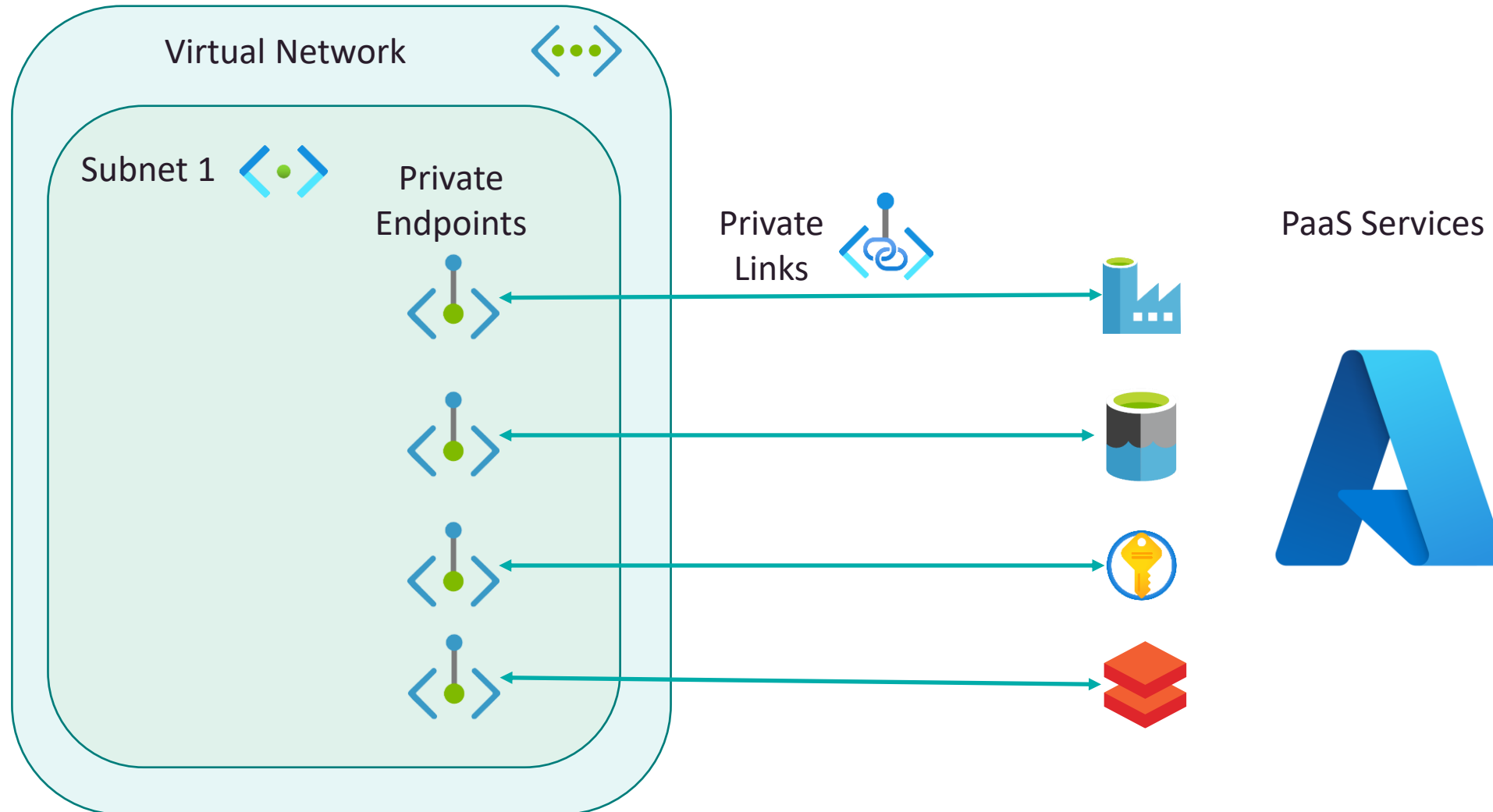
Why using private networking and endpoints also for PaaS Services?

- 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?

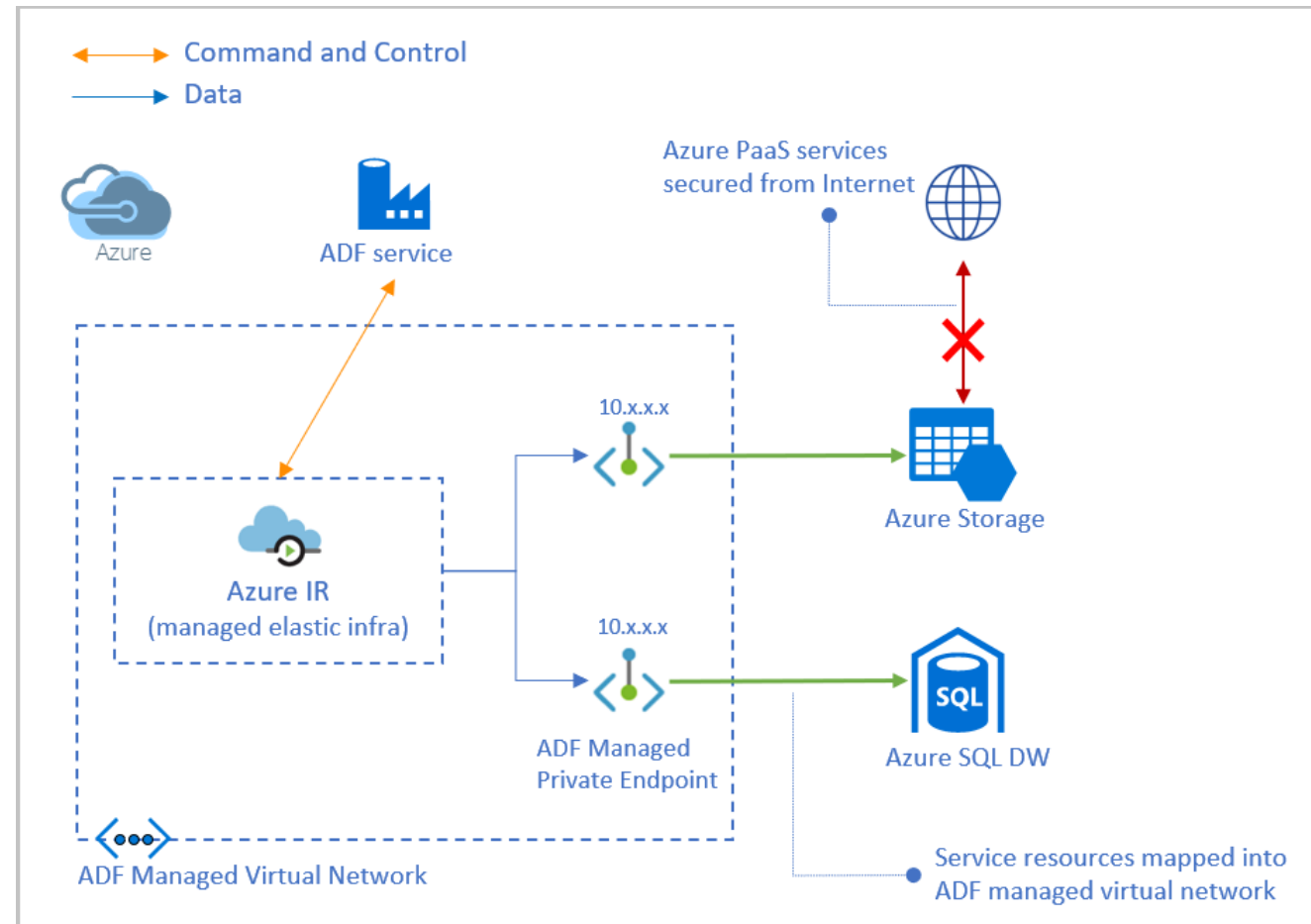
- it adds some complicated overhead, due to
 - Different settings and options for any PaaS Service
 - Networking and firewalling know-how not „home-turf“ of data engineers or data scientists
 - Coordination with IT about IP-address spaces and firewall rules could be time-consuming
 - Additional costs 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

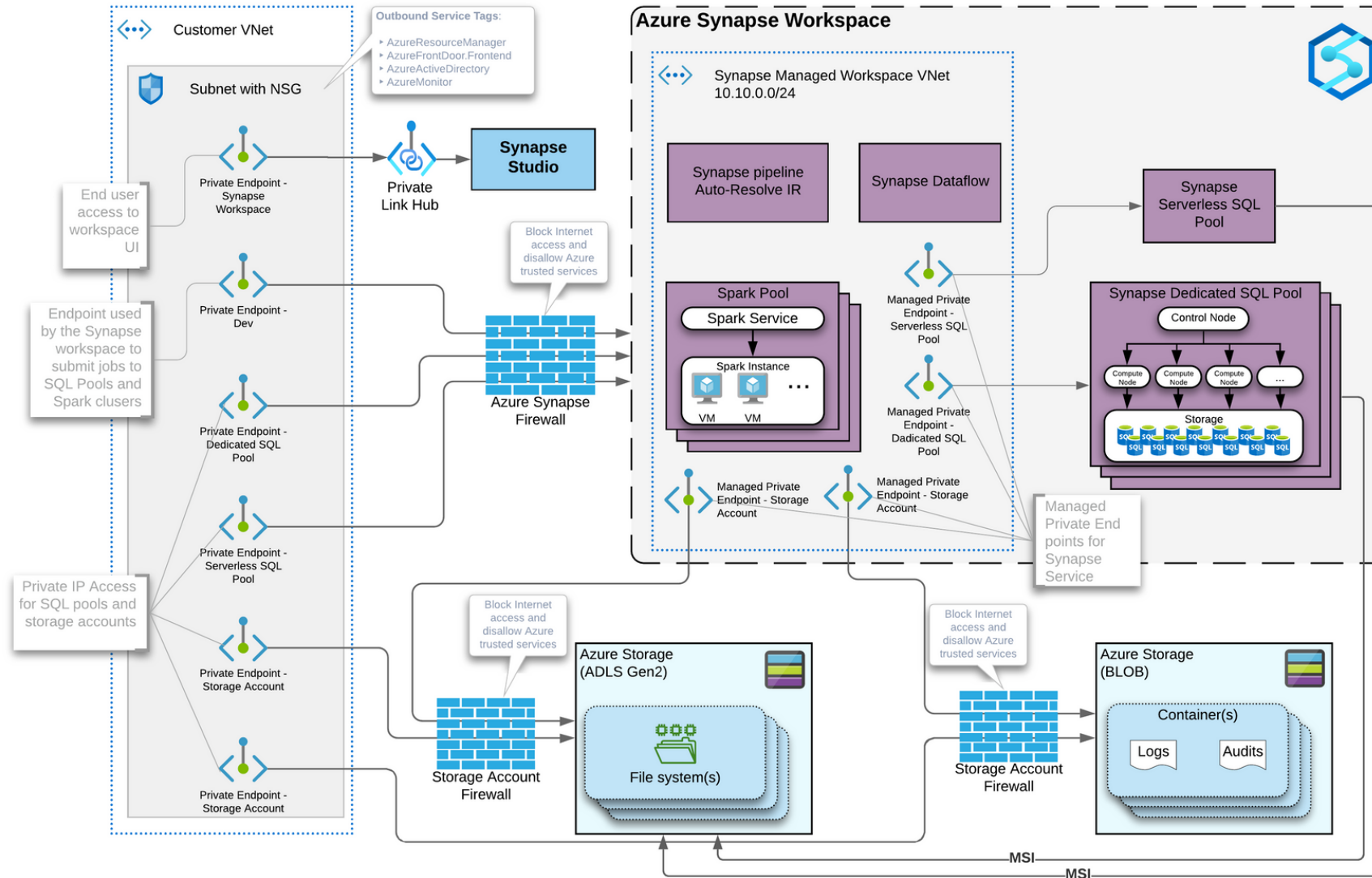


PaaS Integration using Private Endpoint in Managed VNET in Azure Data Factory

- Primarily 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“



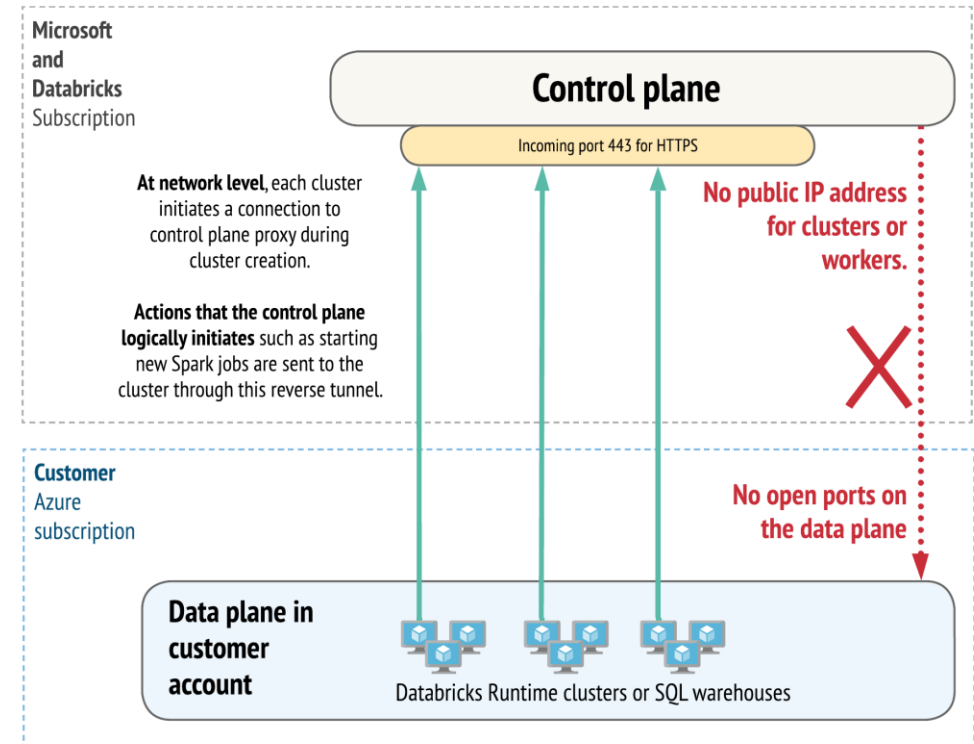
Synapse Managed networking overview – a lot of services add complexity and tons of private endpoints



Azure Databricks networking security options

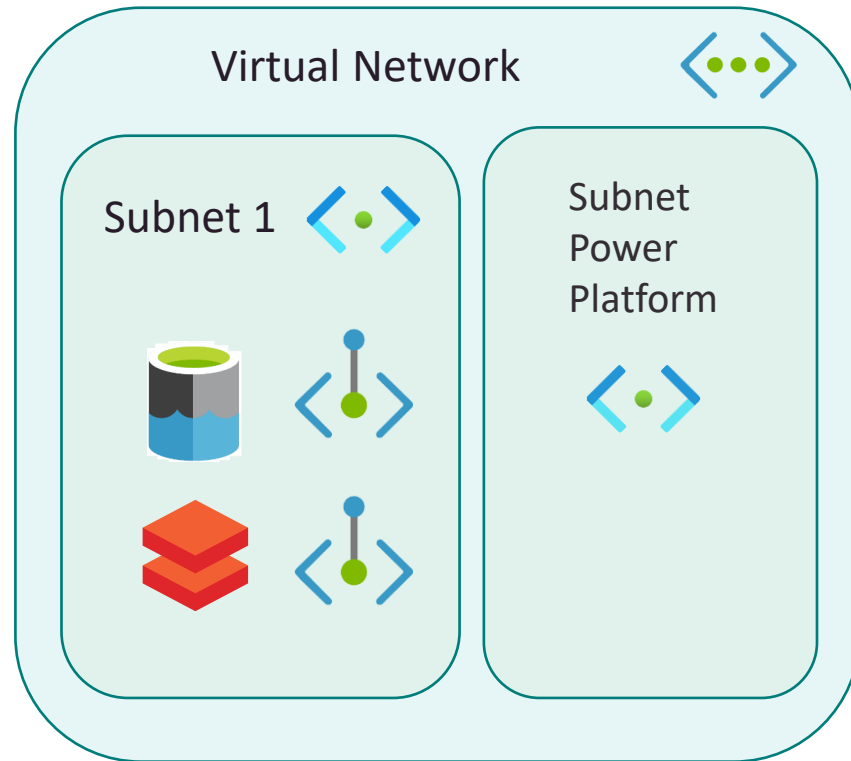
Secure cluster connectivity

- 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



On Azure Databricks, network traffic between the data plane and the control plane traverses the Microsoft network backbone not the public Internet, independent of whether secure cluster connectivity is enabled.

How to access Azure Data secure from Power BI?



You have only Power BI Pro or your data store is not supported?
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?

dpsdatalake | Networking ☆ ...
Storage account

Search

Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser

storage

Containers
File shares
Queues
Tables

Priority + networking

Networking

Access keys
Shared access signature
Encryption
Microsoft Defender for Cloud

management

Redundancy

Firewalls and virtual networks Private endpoint connections

Save Discard Refresh

Public network access

Enabled from all networks

Enabled from selected virtual networks and IP addresses

Disabled

Configure network security for your storage accounts. [Learn more](#)

Virtual networks

+ Add existing virtual network + Add new virtual network

Virtual Network	Subnet	Address range
> vnet-data-platform	1	

Firewall

Add IP ranges to allow access from the internet or your on-premises networks. [Learn more](#).

Address range

Public IP of your company network / HO

Resource instances

Specify resource instances that will have access to your storage account based on their system-assigned managed identity.

Resource type	Instance name
Microsoft.DataFactory/factories	Your instance of Azure Data Factory

Select a resource type Select one or more instances

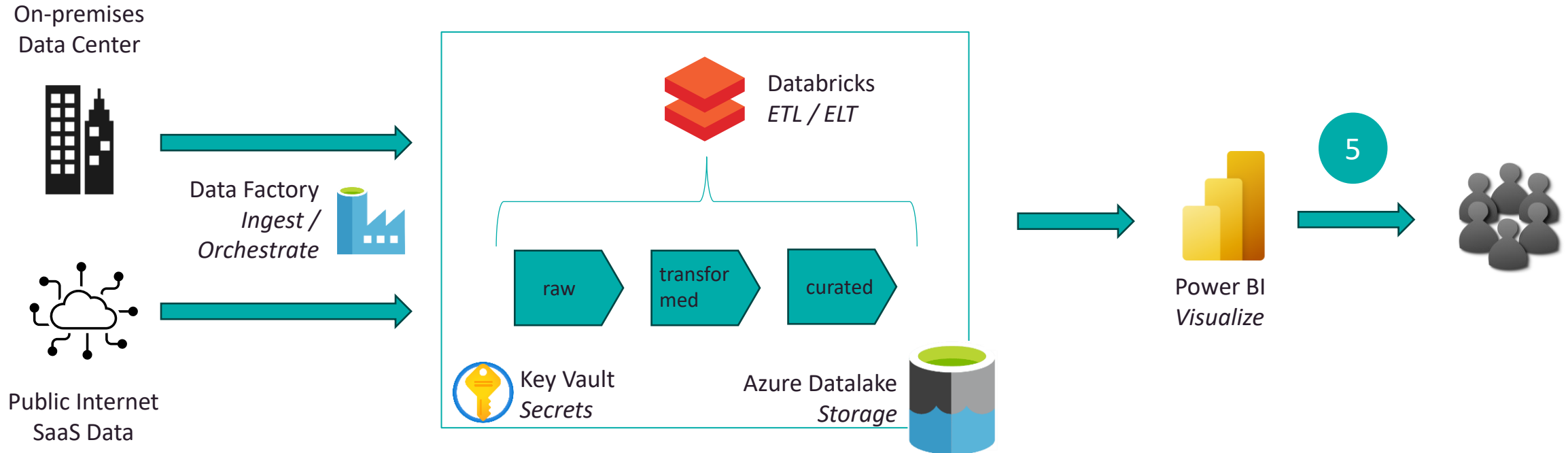
Example for raising security by setting firewall settings strictly avoiding „access for all Azure services“

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



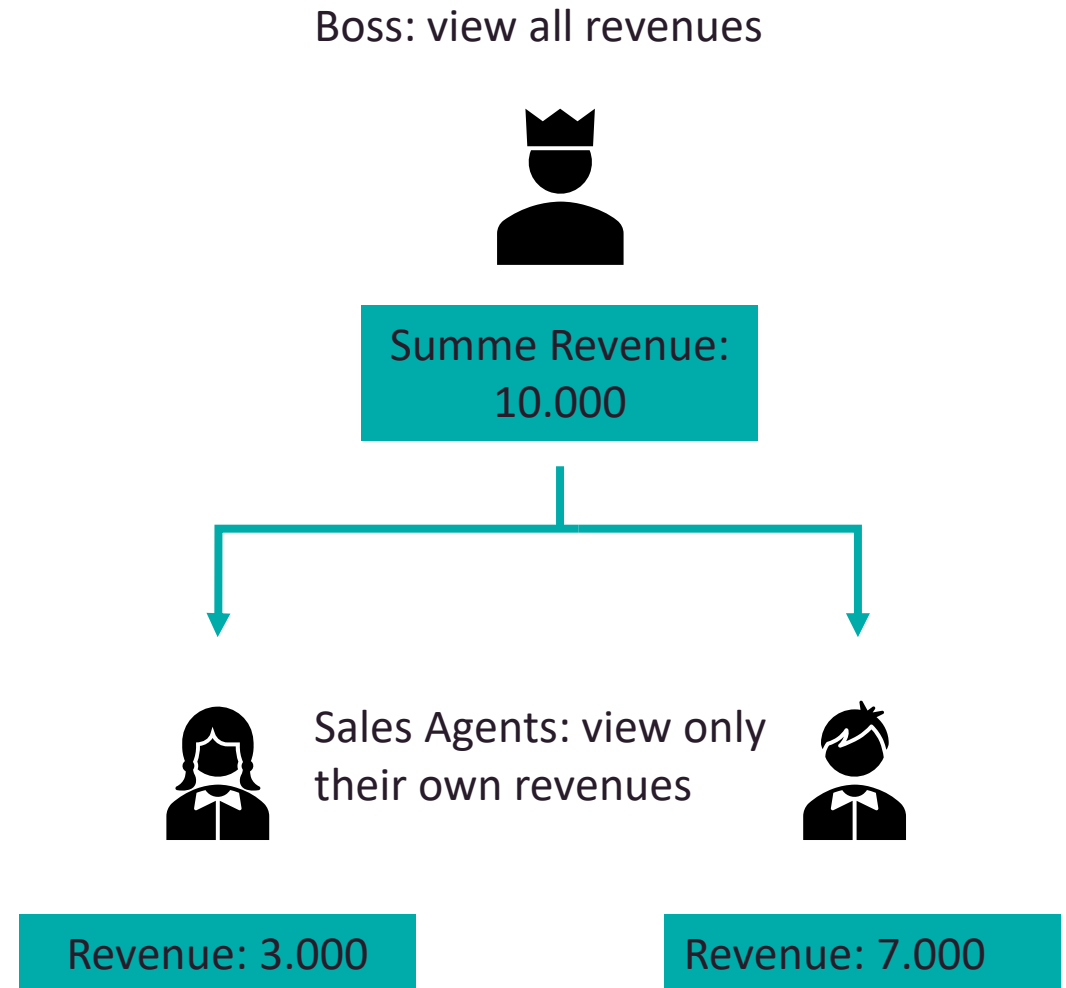
5 Data view permissions

First differentiation – 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

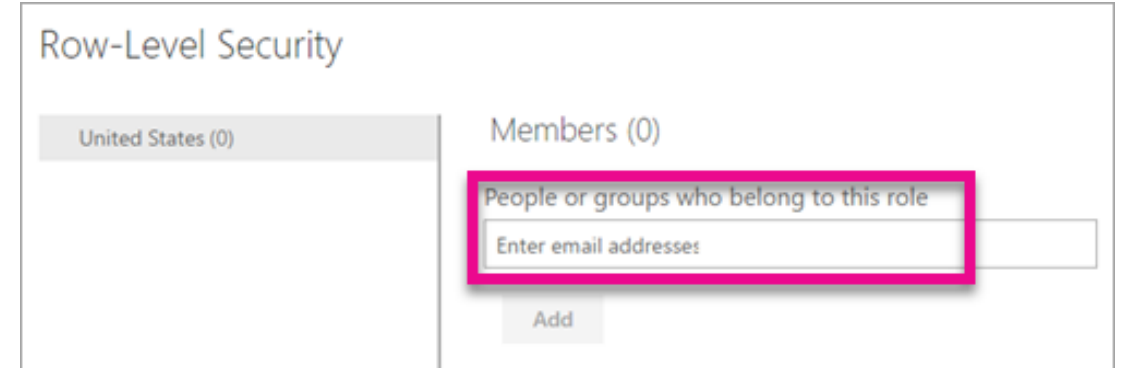
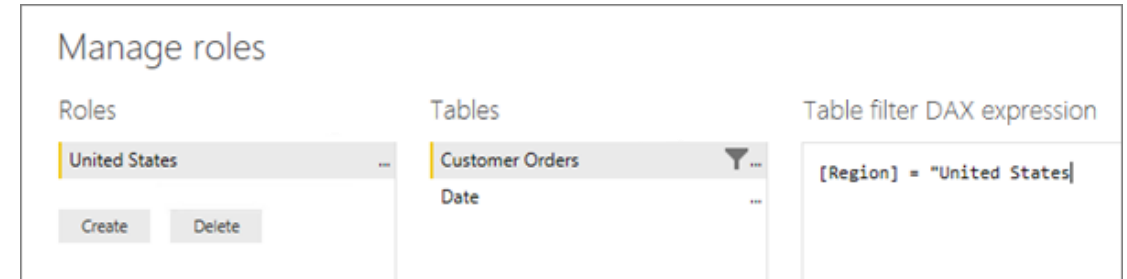


Azure
Databricks

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



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 implementation 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 valuable. 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](#)



Thanks for your attention, I appreciate your feedback!