# **Abstract**

### The Azure Admin Tool Chest for Data Engineers

I am a data engineer working with Databases, Data Lakes, Synapse and Databricks in Microsoft Azure. Some of the problems I encountered could not be solved with SQL or Python, but instead required the skills of an Azure Admin. And since all the real ones were busy, I had to administrate the things myself.

I will show with real world examples I encountered what parts of the Azure Admin Tool Chest can be relevant for the humble data engineer. I will explain:

- what headache could have been prevented with delete locks
- why I should have adhered to the principle of least privileged when assigning RBAC roles to access my Synapse workspace.
- what is RBAC anyway
- how to set up cost management to prevent Databricks or Azure Synapse stampeding over your budget
- where monitoring your resources and setting up basic alerts makes sense to prevent common catastrophes

In this session, we'll look beyond the traditional tasks of the data engineer and see what's there in the realm of Azure Administration. This knowledge can empower you to solve problems when others are too busy. The Azure Admin Tool Chest makes your skillset more well-rounded.

# DATA SATURDAY CROATIA 2024

The Azure Admin
Tool Chest for Data
Engineers



Marisol Steinau



### **EVENT SPONSORS, THANKS!!!**

### **GOLD**









**BRONZE** 









# Have you seen a dragon?



Help him!





# Who am 1?

# **Marisol Steinau**

- Autodidact Data Enthusiast
- > 8 years experience using Microsoft Data Platform
- Senior Data Engineer @Be-terna
- Passionate about Architecture
- Linkedin: linkedin.com/in/marisol-steinau-bb1253253



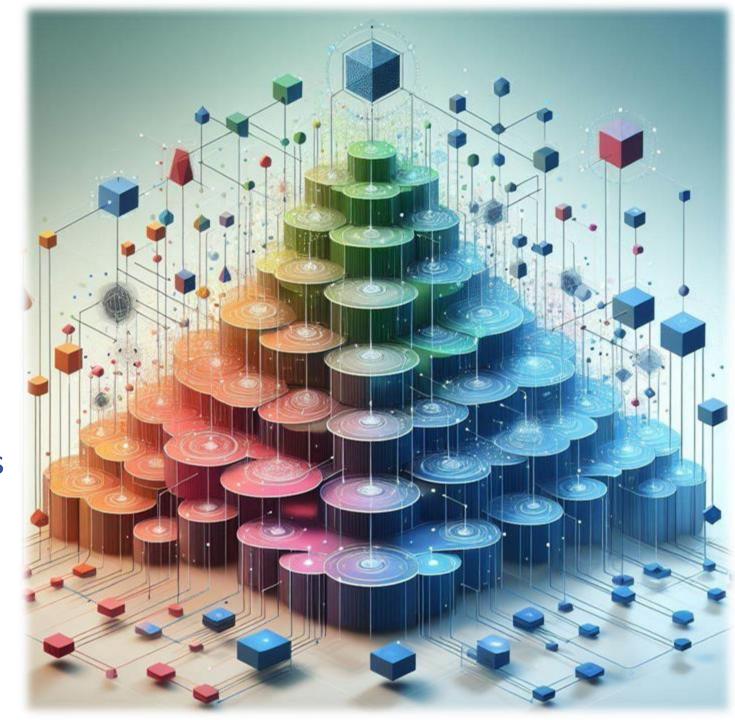
# Agenda

- Recap: Azure Hierarchy
- Fail 1: Misconfiguring access control
- Fail 2: Granting resource deletion rights
- Fail 3: Going without rules
- Fail 4: Forgetting to setup monitoring
- Fail 5: Being careless about cost

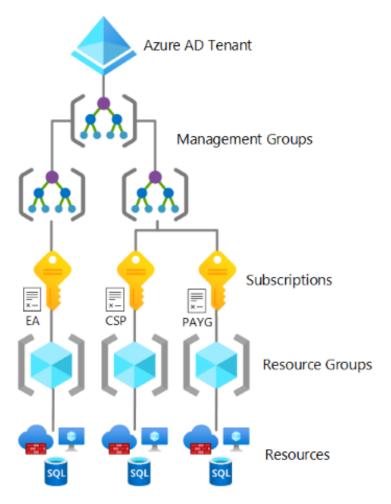


# Recap: Azure Hierarchy

Organizing structure for resources



# Azure Resource Hierarchy



**Tenant:** organization in Azure Entra ID

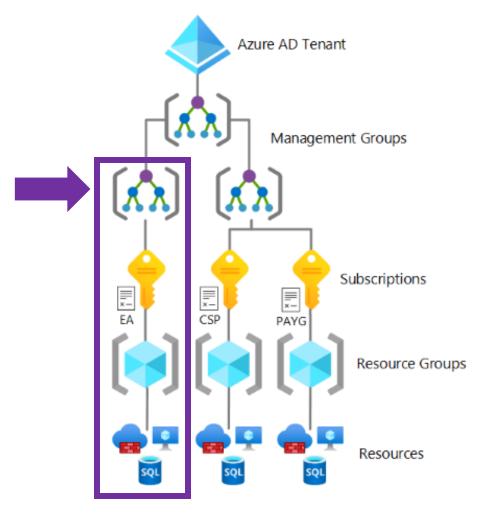
**Management Groups:** manage access, policy, and compliance for multiple subscriptions

**Subscriptions:** manage costs and the resources that are created by users, teams, or projects

**Resource Groups:** logical containers for deployment and management of resources

**Resources:** instances of services

# Azure Resource Hierarchy



**Tenant:** organization in Azure Entra ID

**Management Groups:** manage access, policy, and compliance for multiple subscriptions

**Subscriptions:** manage costs and the resources that are created by users, teams, or projects

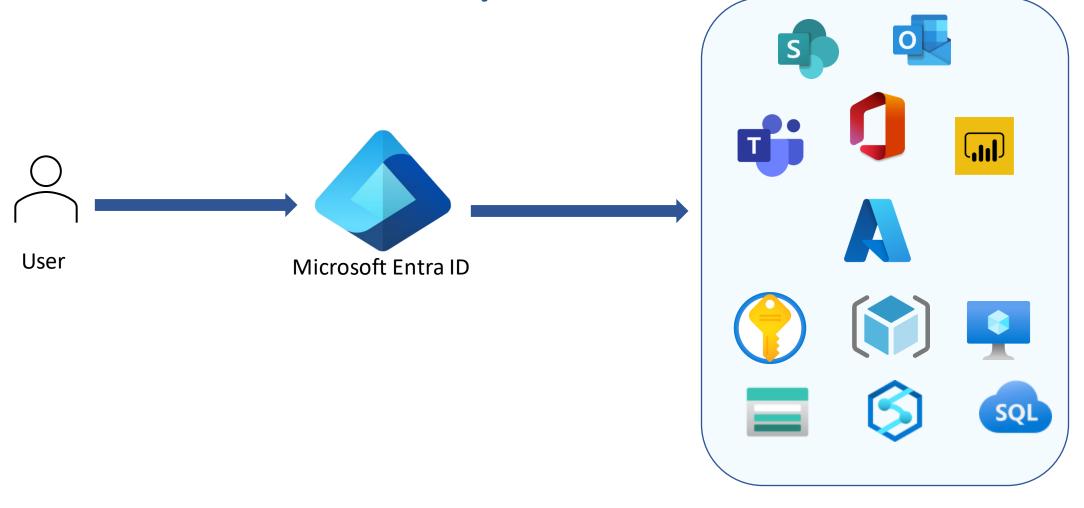
**Resource Groups:** logical containers for deployment and management of resources

**Resources:** instances of services

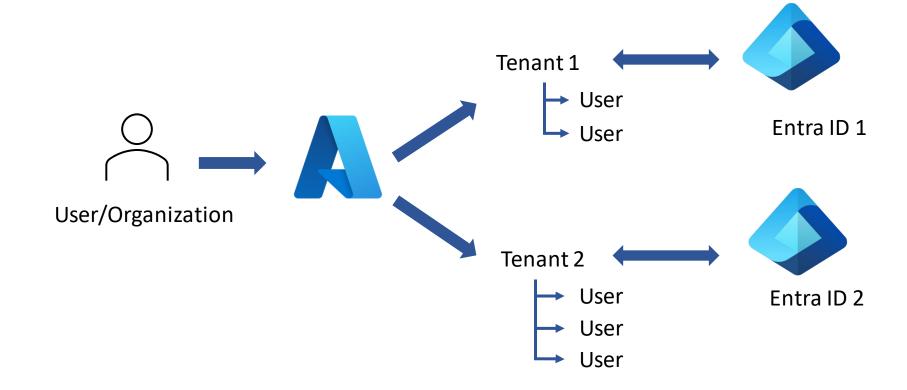




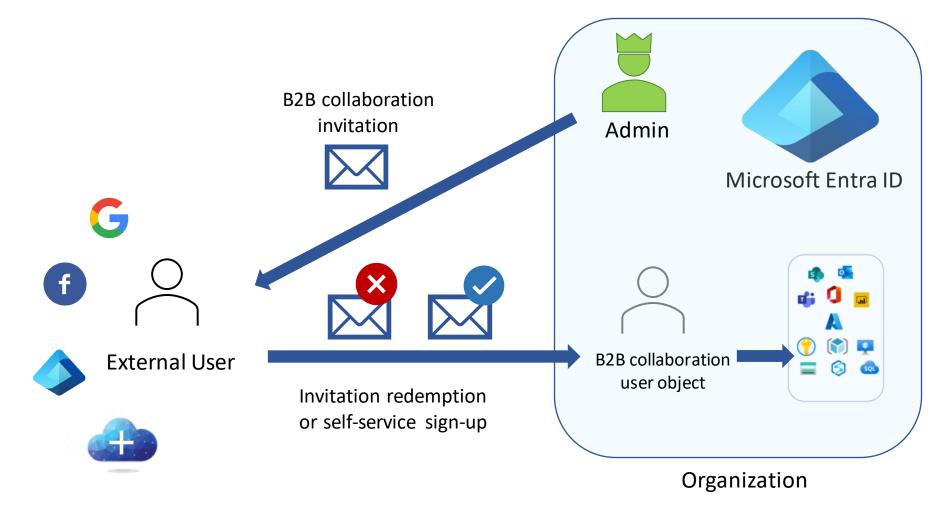
# Azure Active Directory Microsoft Entra ID



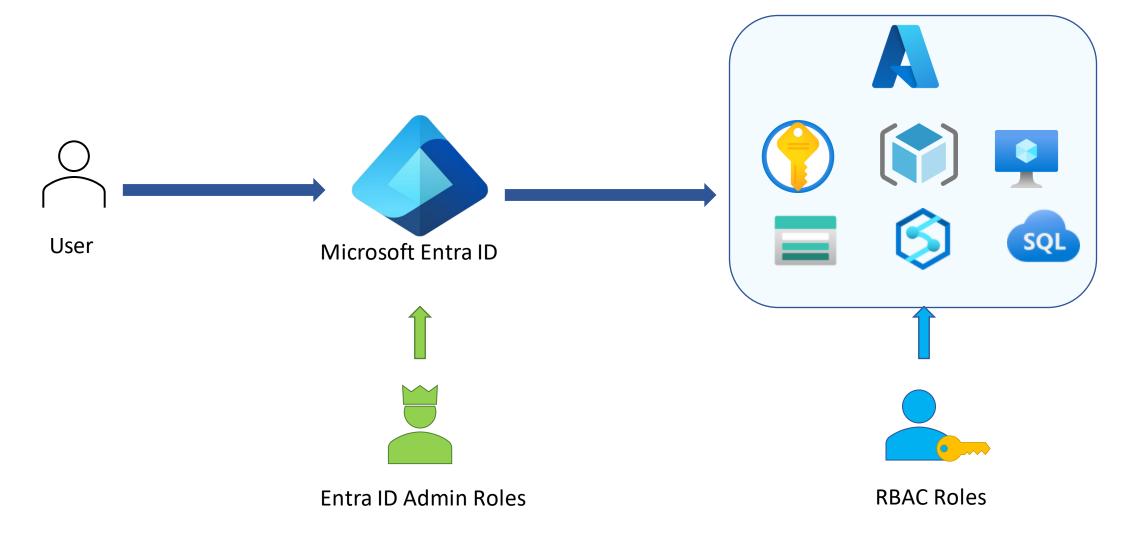
# Microsoft Entra ID



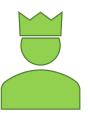
# Azure B2B – External Users



# Entra ID Roles vs Azure RBAC



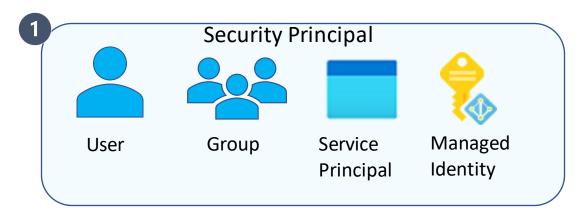
# Microsoft Entra ID Roles



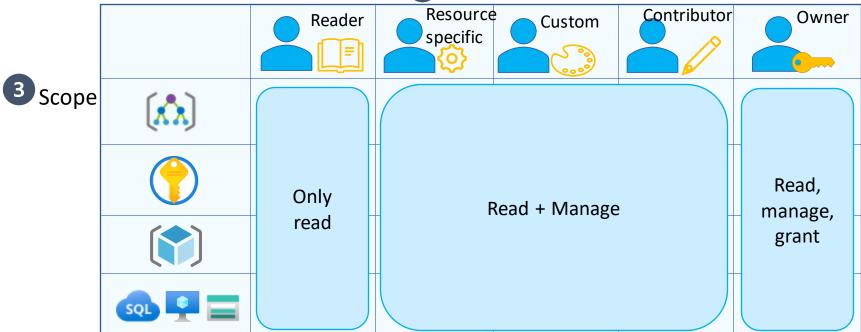
### Most important Entra ID Roles from the data perspective:

- Global Administrator
- User Administrator
- Application Administrator
- Fabric Administrator
- Power Platform Administrator
- Azure DevOps Administrator

# Role Based Access Control

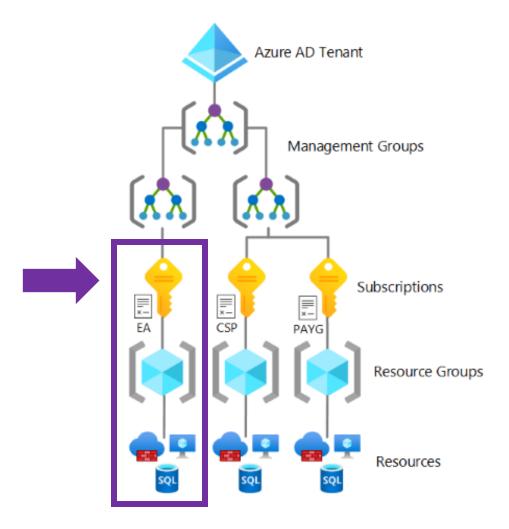


2 Role



- Who has access to resources?
- What can they do with resources?
- What is the scope of access?

# Azure Resource Hierarchy



**Tenant:** organization in Azure Entra ID

**Management Groups:** manage access, policy, and compliance for multiple subscriptions

**Subscriptions:** manage costs and the resources that are created by users, teams, or projects

**Resource Groups:** logical containers for deployment and management of resources

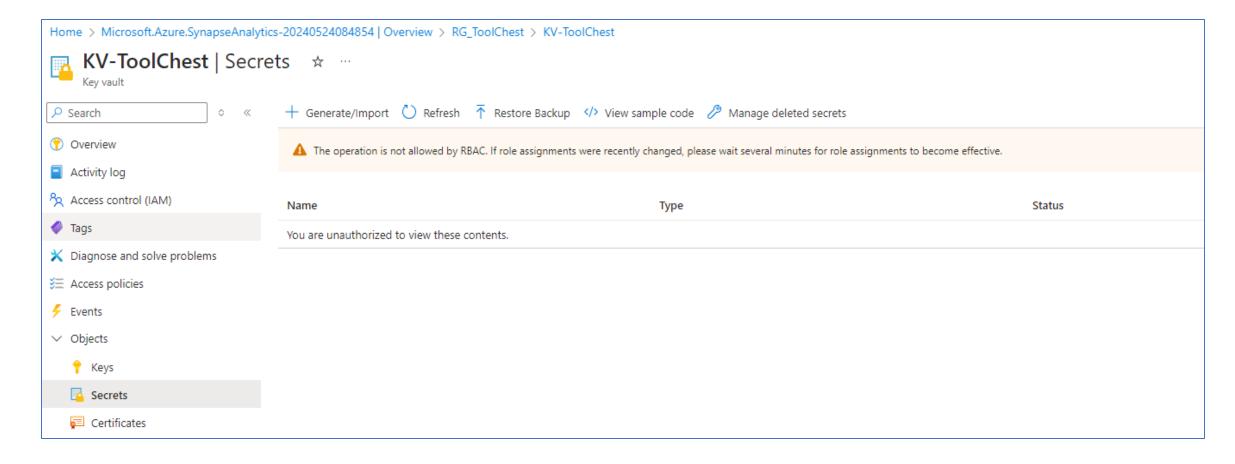
**Resources:** instances of services

# Resource Specific RBAC 🍎 🐯



### An example for resource specific RBAC is Azure Key Vault





# Resource Specific RBAC 👵



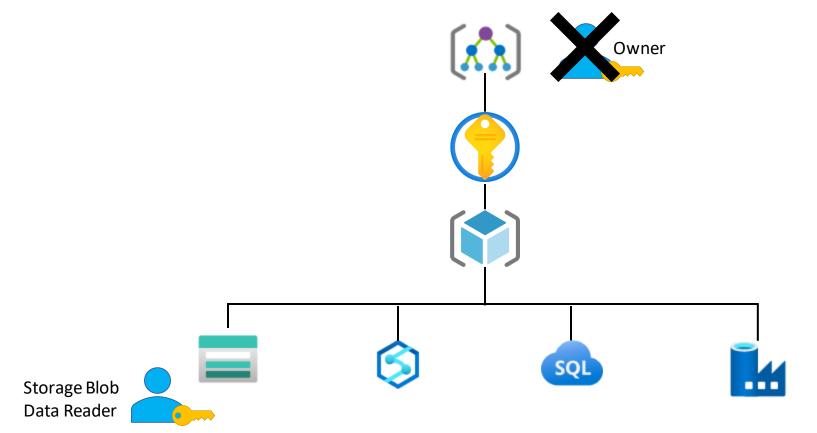
# An example for resource specific RBAC is Azure Key Vault



Name ↑↓	Description ↑↓
Key Vault Administrator	Perform all data plane operations on a key vault and all objects in it, including certificates, keys, and secrets. Cannot manage
Key Vault Certificate User	Read certificate contents. Only works for key vaults that use the 'Azure role-based access control' permission model.
Key Vault Certificates Officer	Perform any action on the certificates of a key vault, except manage permissions. Only works for key vaults that use the 'Az
Key Vault Contributor	Lets you manage key vaults, but not access to them.
Key Vault Crypto Officer	Perform any action on the keys of a key vault, except manage permissions. Only works for key vaults that use the 'Azure ro
Key Vault Crypto Service Encryption User	Read metadata of keys and perform wrap/unwrap operations. Only works for key vaults that use the 'Azure role-based acc
Key Vault Crypto Service Release User	Release keys. Only works for key vaults that use the 'Azure role-based access control' permission model.
Key Vault Crypto User	Perform cryptographic operations using keys. Only works for key vaults that use the 'Azure role-based access control' pern
Key Vault Data Access Administrator	Manage access to Azure Key Vault by adding or removing role assignments for the Key Vault Administrator, Key Vault Certi
Key Vault Reader	Read metadata of key vaults and its certificates, keys, and secrets. Cannot read sensitive values such as secret contents or keys, and secrets.
Key Vault Secrets Officer	Perform any action on the secrets of a key vault, except manage permissions. Only works for key vaults that use the 'Azure
Key Vault Secrets User	Read secret contents. Only works for key vaults that use the 'Azure role-based access control' permission model.
<u> </u>	

# Principle of least privilege

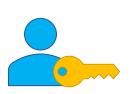
Each Security Principal should only have as many permissions as it needs to perform its tasks .





# Tips when you are not





Privileged Identity Management





Stay on good terms with the people who can grant roles

User Access Admin (RBAC)



Principle of least privilege

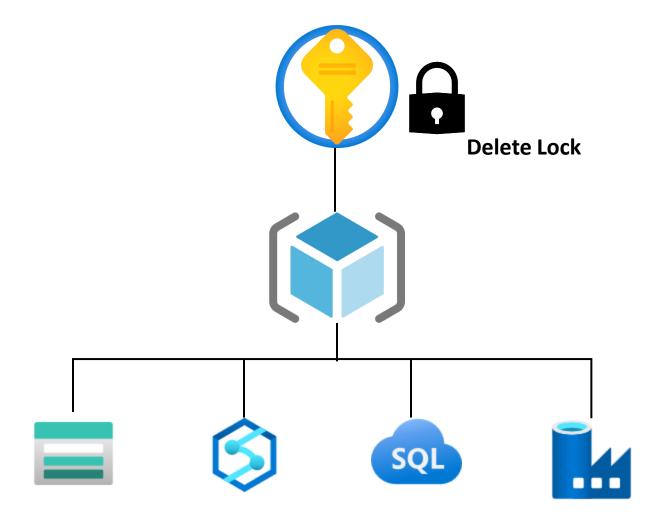


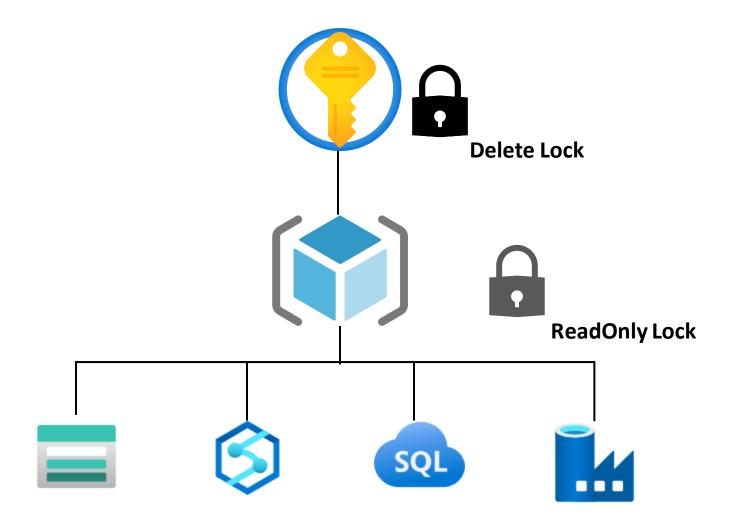
# Fail 2: Granting resource deletion rights A colleague deleted the storage account of our prod environment

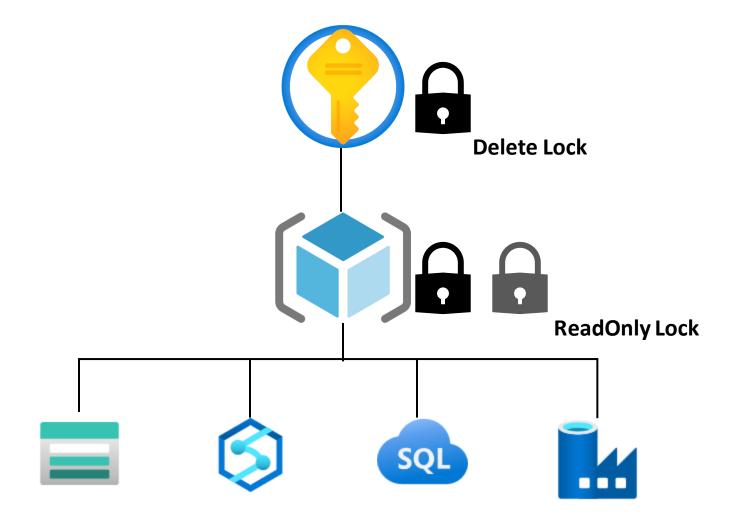
Locking resources can help ensure users don't accidently delete or modify resources.

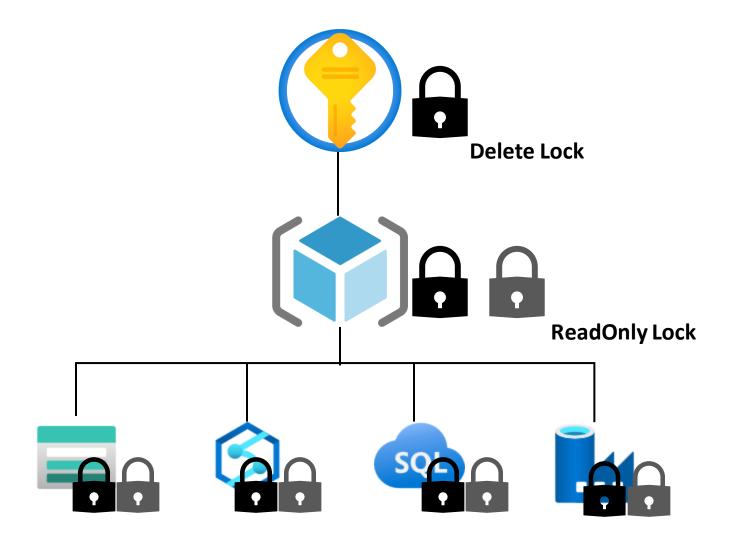


**Owners** or **user access administrators** may create or remove locks. They must remove the lock before they can delete a resource.



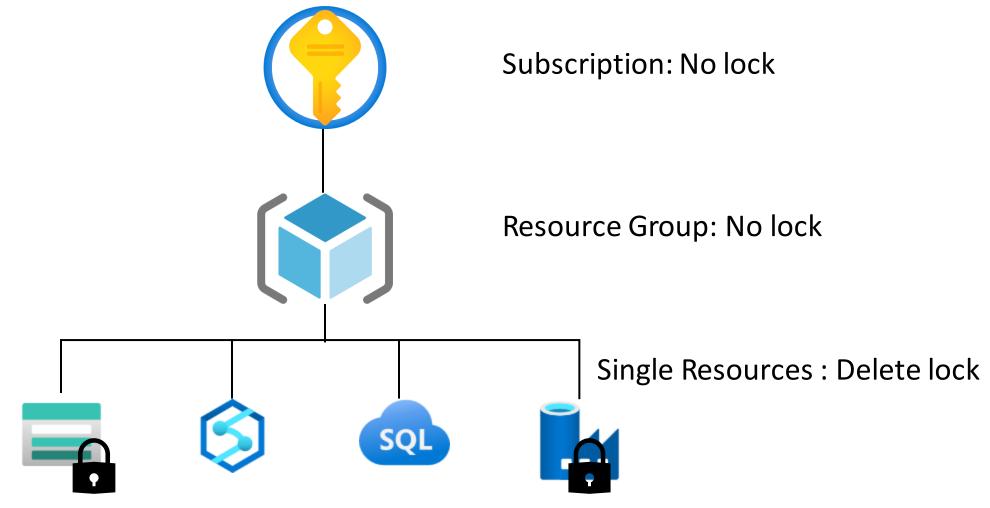






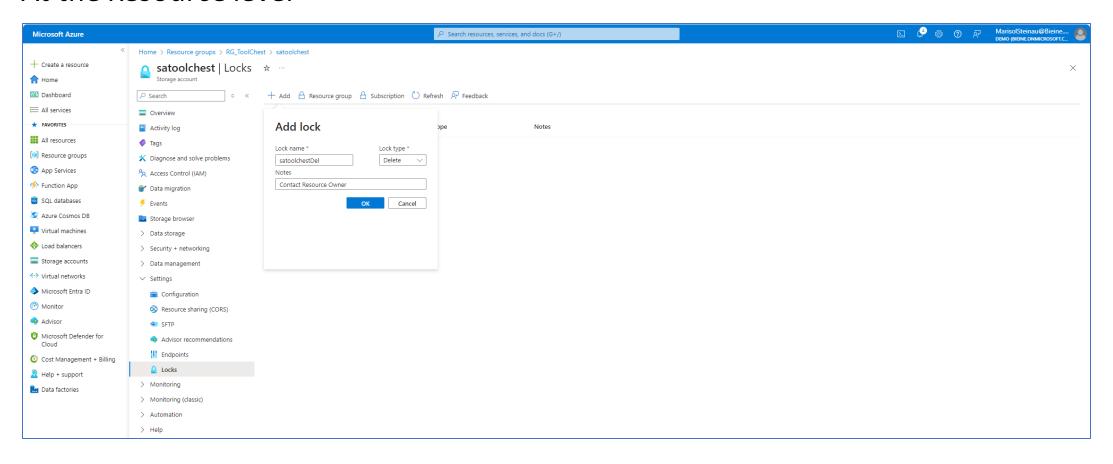
# Recommendation

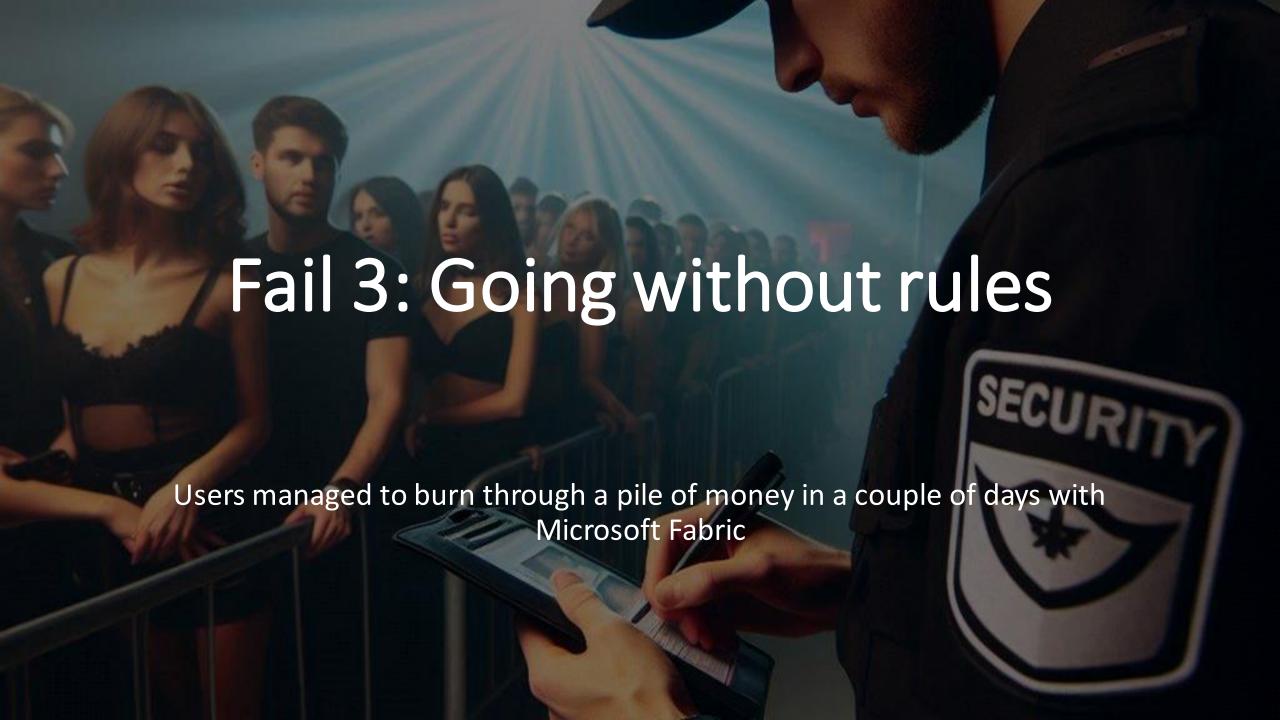




# Applying locks

### At the Resource level





# **Azure Policies**

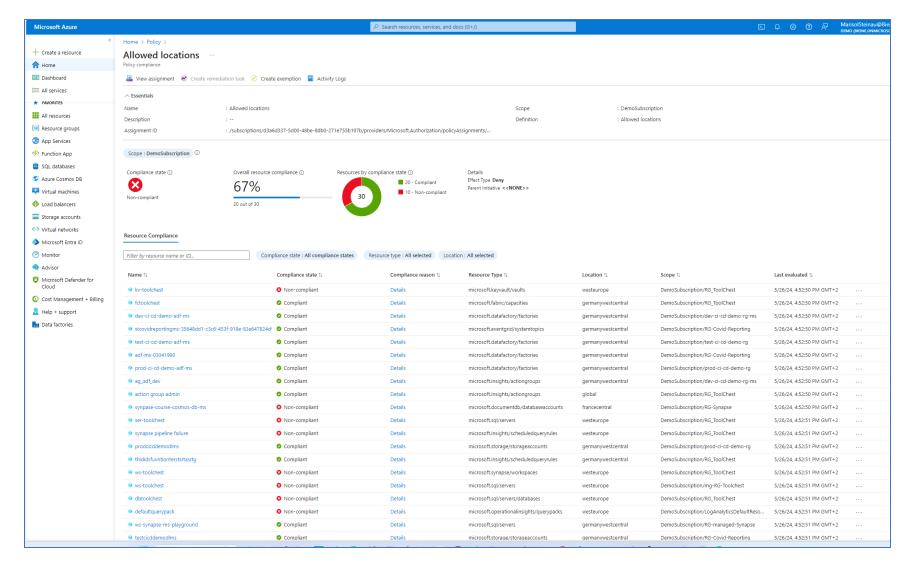
Scope **Azure Policy** Azure Policies help in controlling, restricting and auditing resources

# **Azure Policies** Inheritance Assignment with Exemption

# **Azure Policies**

### Instance details satoolchestpolicydemo Storage account name \* (i) (Africa) South Africa North Region \* (i) Deploy to an Azure Extended Zone Resources can only be created in the locations Germany North, Germany West Central and Germany (Policy details) Performance \* (i) Standard: Recommended for most scenarios (general-purpose v2 account) Premium: Recommended for scenarios that require low latency. Locally-redundant storage (LRS) Redundancy \* (i)

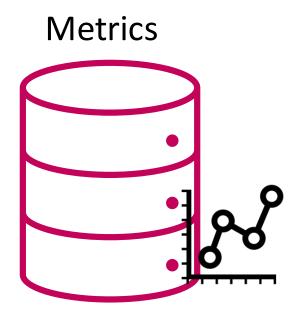
# **Azure Policies**





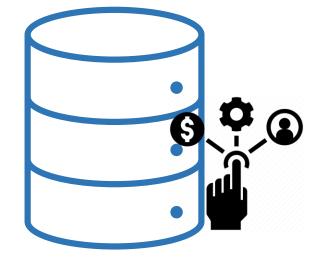


# **Azure Monitor**



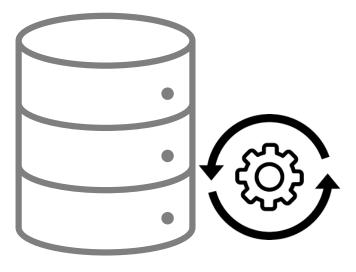
Numerical values





Operations on resources

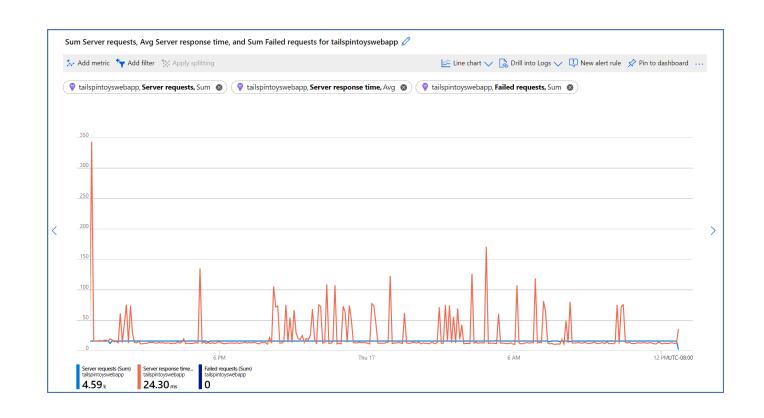
Resource Logs



Operations by resources

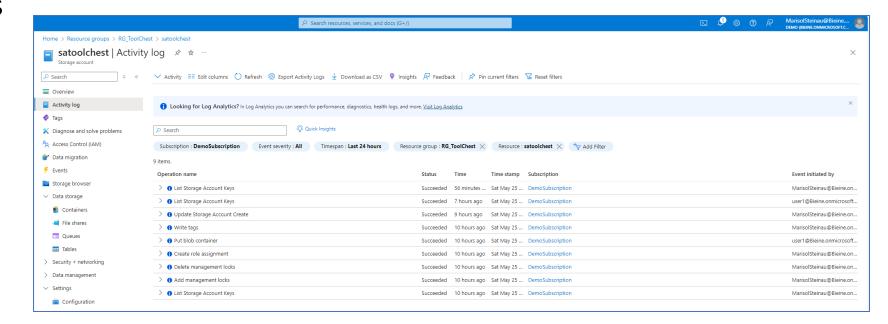
#### Metrics

- Collection of numeric data
- Support of near-real time
- Saved for 30 days
- Useful for alerts
- Based on resources



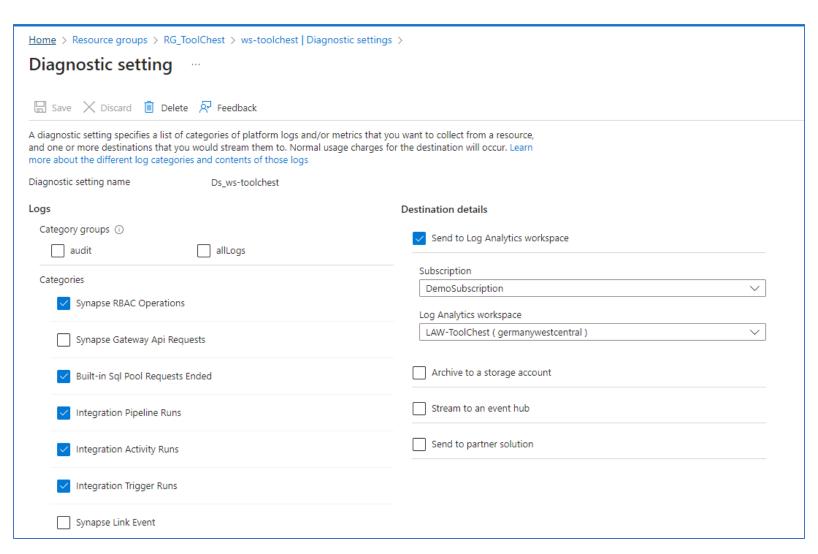
# **Activity Logs**

- Operations on resources
- What? When? Who?
- Saved for 90 days
- Useful for alerts
- Based on resources

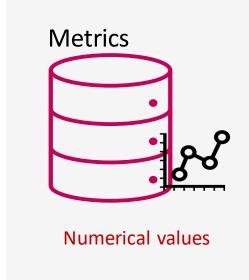


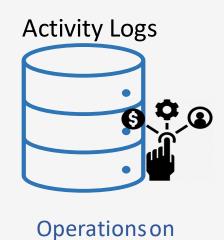
### Resource Logs

- Operations by resources
- Logs are generated automatically but not stored
- Storage has to be set-up
- Based on resources



# **Collect Monitoring Data**









resources



**Diagnostic Settings** 

Log Analytics Workspace



**Event Hubs** 

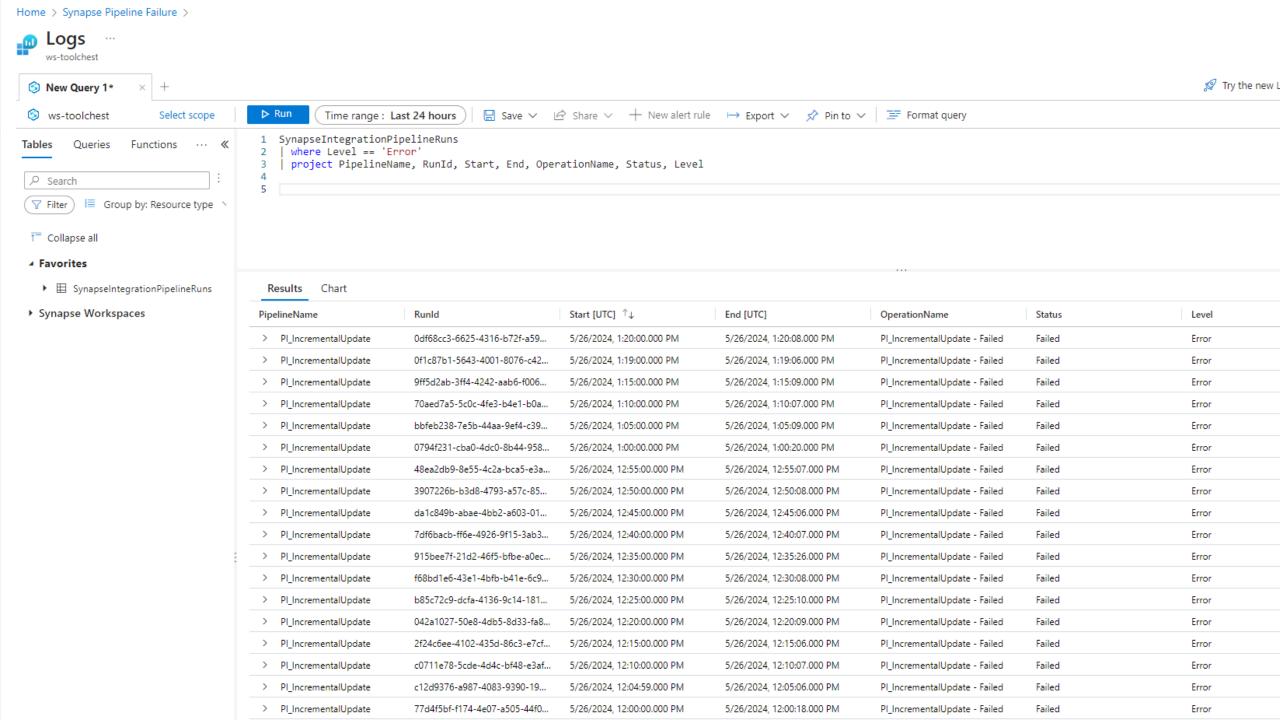


Azure Storage



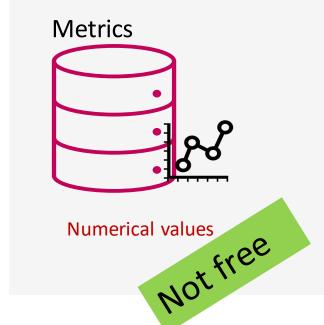
**Partner Solution** 





### Alerts



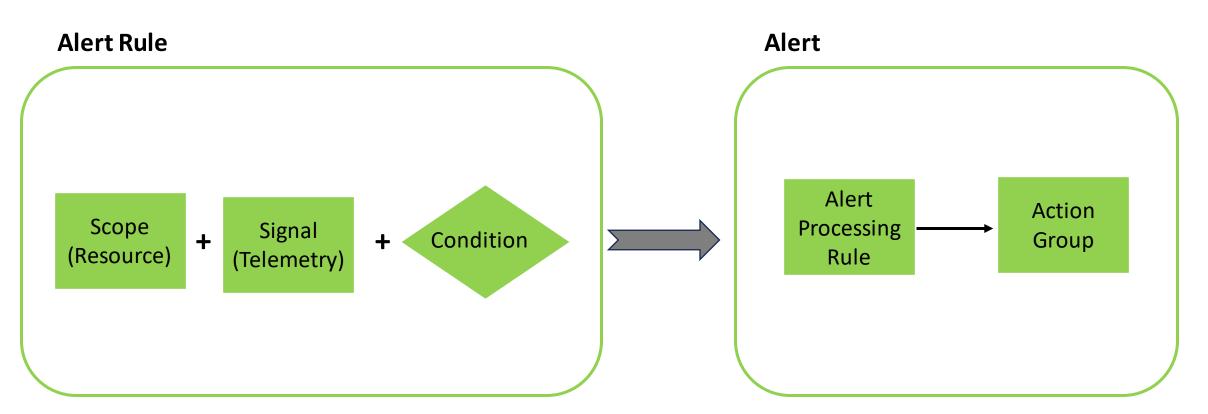






### Alerts

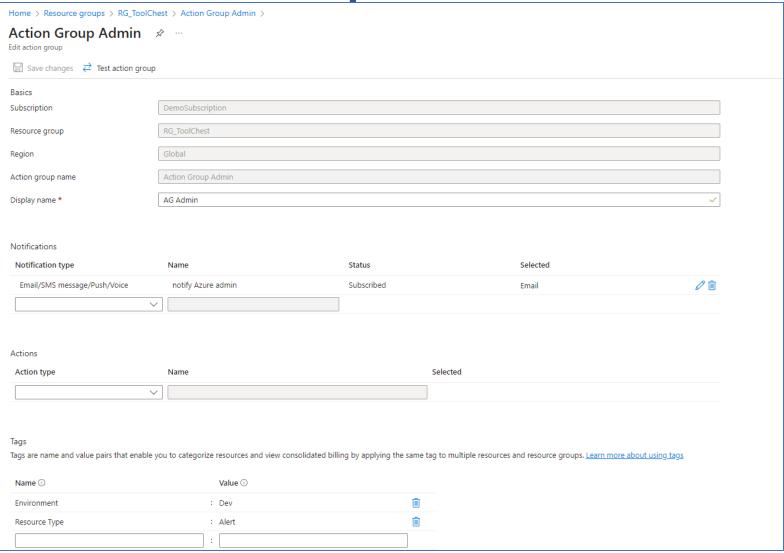




## Alerts

lome > Resource groups > RG_To	olChest > Synapse Pipe	line Failure >	
dit alert rule			
Scope Condition Actions	Details Tags Revi	ew + save	
Configure when the alert rule should	trigger by selecting a signa	al and defining its logic.	
Signal name * ①		arch	~
	See all signals		
Define the logic for triggering an aler	t Use the chart to view tre	nds in the data Learn more	
Jenne the logic for triggering an aler	t. Ose the chart to view tie	nus in the data. <u>Learn more</u>	
	gs. The results returned by	this query are used to populate the aler	rt definition below.
Search query *			
SynapseIntegrationPipeline   where Level == 'Error'			
project PipelineName, Ru	ınId, Start, End, Op	erationName, Status, Level	
	h		
/iew result and edit query in Logs 🗜			
Measurement			
Select how to summarize the results.	We try to detect summariz	ed data from the query results automati	cally.
Measure (i)	Table rows	~	
Aggregation type ①	Count		
aggregation type ()	Count		
Aggregation granularity ①	1 minute	~	
Split by dimensions			
Use dimensions to monitor specific ti	me series and provide cont	text to the fired alert. Dimensions can be	e either number or string columns. If you select more
		combination will trigger its own alert ar	
Dimension name	Operator	Dimension values	Include all future values
Select dimension	<u> </u>	O selected  Add custom value	
		Add custom value	

### **Action Group**



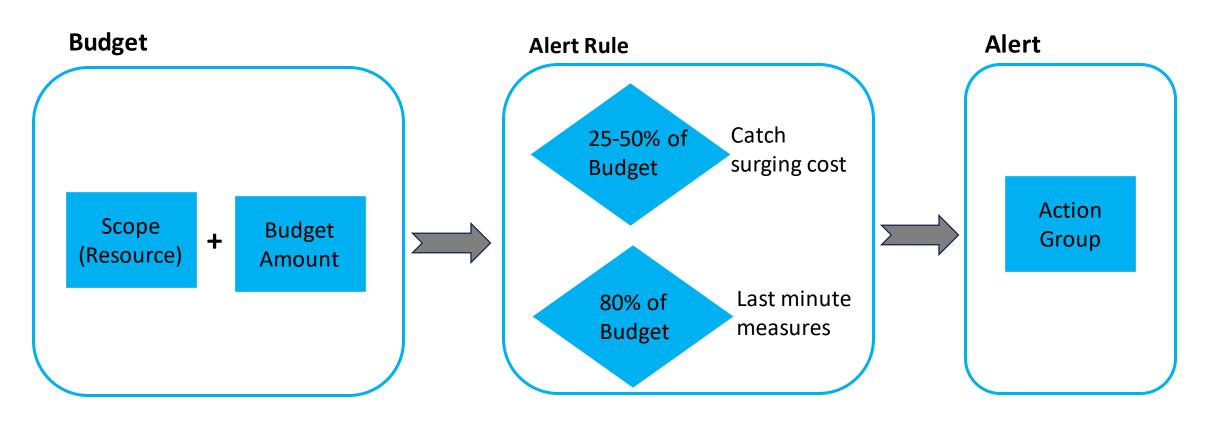
#### Notification

Fired:Sev1 Azure Monitor Alert Synapse Pipeline Failure on ws-toolchest ( microsoft.synapse/workspaces ) at 5/26/2024 2:21:01 PM View the alert in Azure Monitor > Investigate > Summary Alert name Synapse Pipeline Failure Severity Monitor condition Affected resource ws-toolchest Resource type microsoft.synapse/workspaces Resource group rg\_toolchest Monitoring service Log Alerts V2 Signal type Log Fired time May 26, 2024 14:21 UTC Alert ID 181a6301-35e4-1e94-e21b-184889f5000e Alert rule ID https://portal.azure.com/#resource/subscriptions/d3a6 d337-5d00-48be-8db0-271e755b197b/resourceGroup s/RG\_ToolChest/providers/microsoft.insights/scheduled queryrules/Synapse Pipeline Failure Filtered search results View query results Search results View query results Search query SynapseIntegrationPipelineRuns | where Level == 'Error' project PipelineName, Runld, Start, End, OperationNa me, Status, Level Target resource types ['Microsoft.Synapse/workspaces'] Time aggregation Count GreaterThan Operator Threshold Metric value Number of violations Number of examined 1 periods

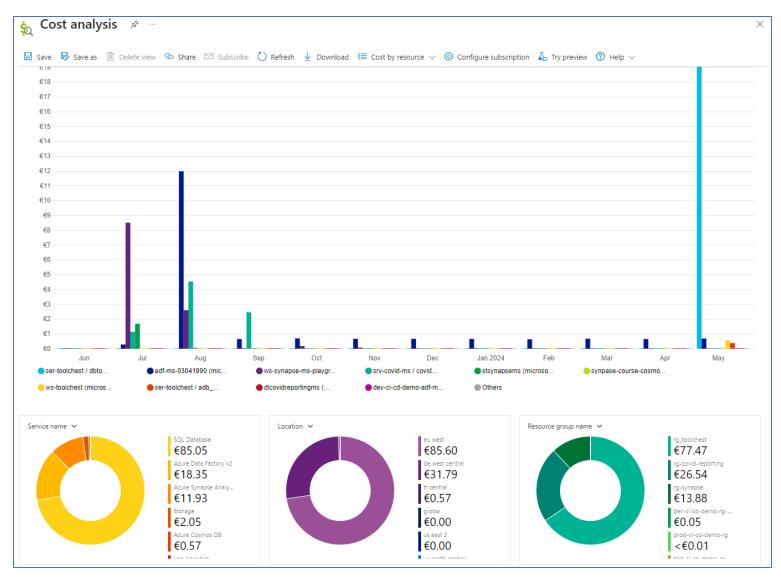


#### Cost Management & Billing

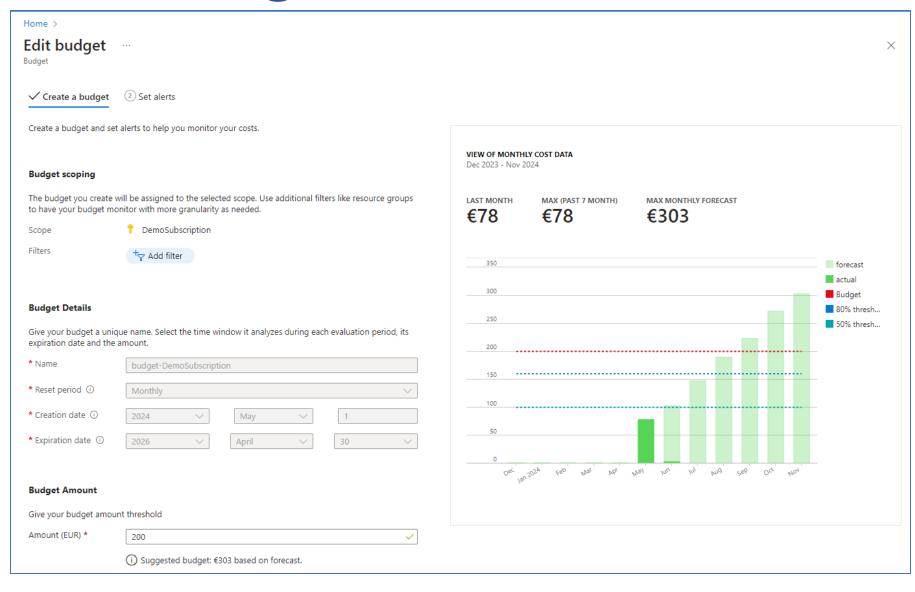




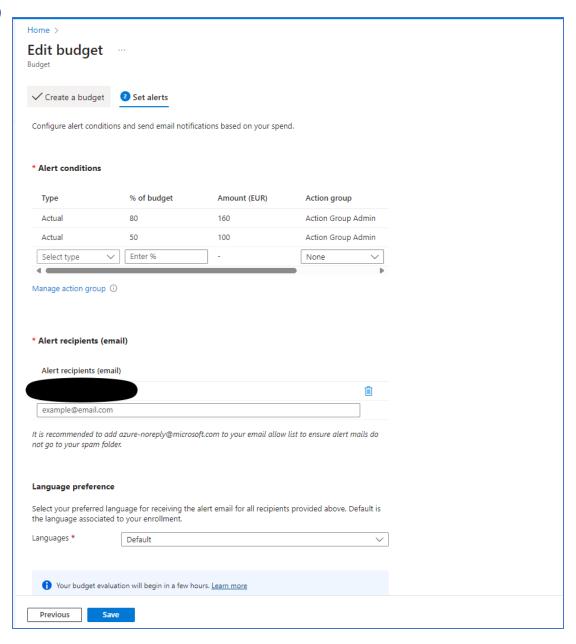
# Cost Management & Billing



## Cost Budget



#### **Cost Alerts**



#### Notification

#### Your cost is forecasted to exceed its threshold value for your Azure budget, budget-DemoSubscription

The total spend for your budget, budget-DemoSubscription, is forecasted to reach 367,58 € before the end of the period. This amount exceeds your specified budget threshold of 200,00 €. Please review your budget in the Azure portal and analyze your cost and usage.

#### Review budget >

Budget name:	budget-DemoSubscription	
Budget start date:	June 1, 2024	
Budget type:	Cost	
Budget amount:	200,00 €	
Notification threshold for forecast alert:	200,00 €	
Total forecasted for period:	367,58 €	
Alert generation time:	June 1, 2024 1:28 UTC	

To see more information about this alert, view the alert details in the Azure portal. If you need additional help, visit budgets tutorial. If you need help, please contact support.

# Tip for Exploring Costs



#### **Use Azure Tags for organizing resources:**

- Name-value pairs that can be attached to resources
- No inheritance of tags at the Resource Group level
- Fundamental for proper governance of any environment
- Azure policy can be used to enforce tagging rules

# YAGNI

# You Ain't Gonna Need It



# Session evaluation, thanks!



https://eval.datasaturdays.com/event/16636280



