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.

The Azure Admin Engineers

Marisol Steinau

Tool Chest for Data



Who am I?

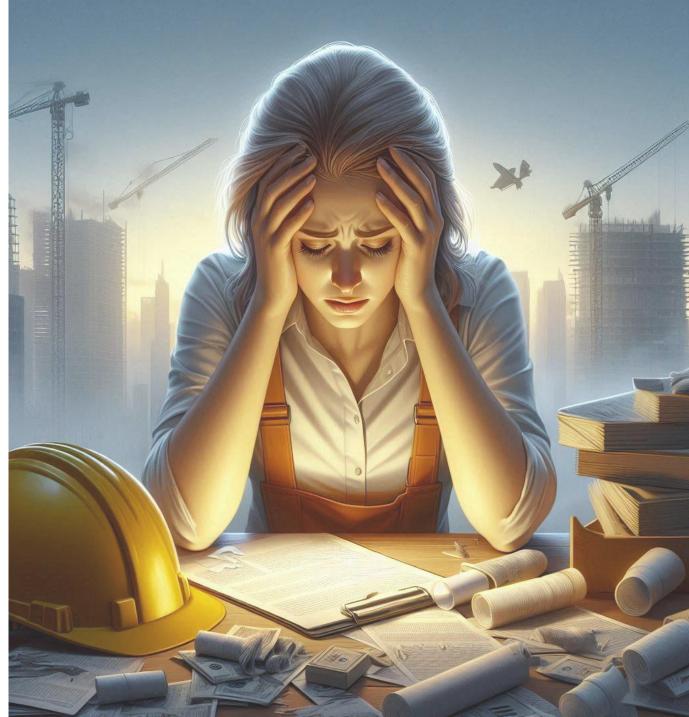
Marisol Steinau

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



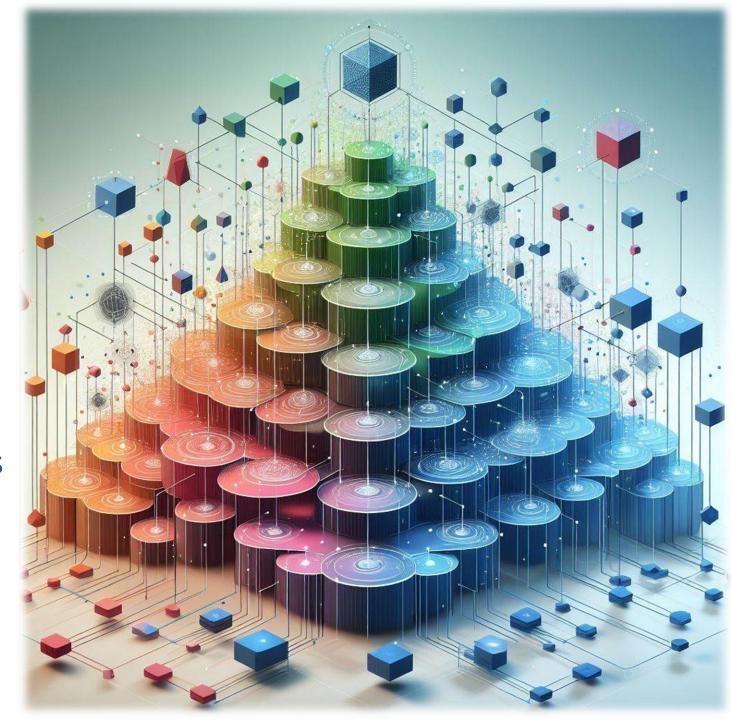
Agenda

- 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

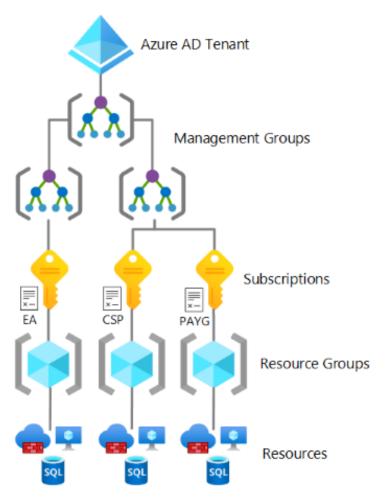


Recap: Azure Hierarchy

Organizing structure for resources



Azure Resource Hierarchy



Tenant: organization in Azure Entra ID

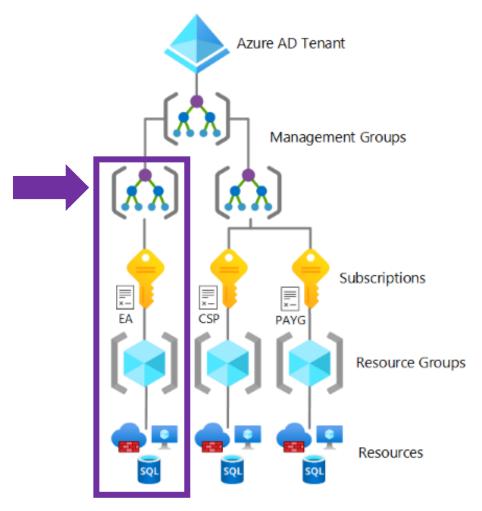
Management Groups: manage access, policy, and compliance for multiple subscriptions

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

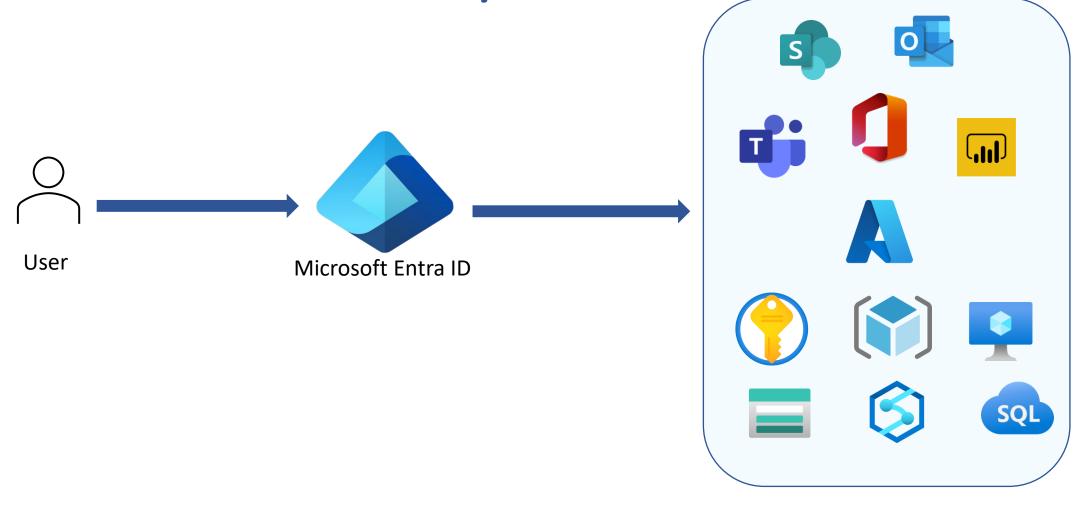
Subscriptions: manage costs and the resources that are created by users, teams, or projects

Resource Groups: logical containers for deployment and management of resources

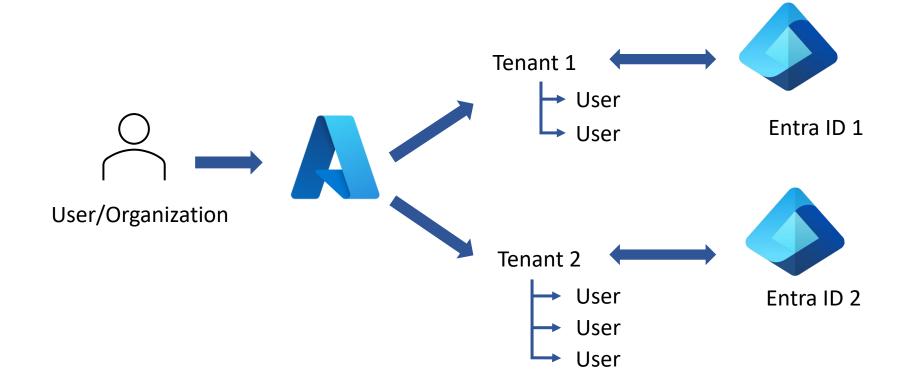
Resources: instances of services



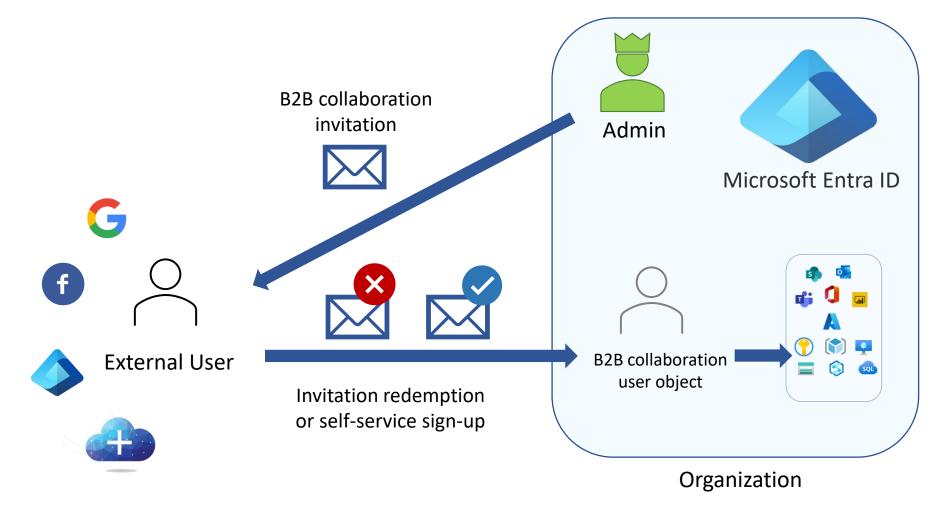
Azure Active Directory Microsoft Entra ID



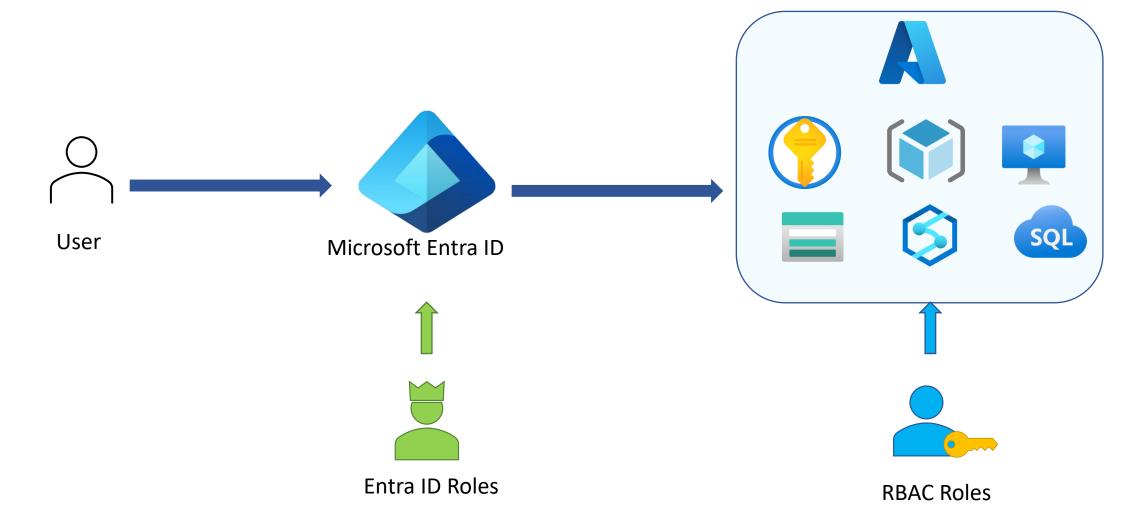
Microsoft Entra ID



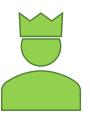
Azure B2B – External Users



Entra ID Roles vs Azure RBAC Role-Based Access Control



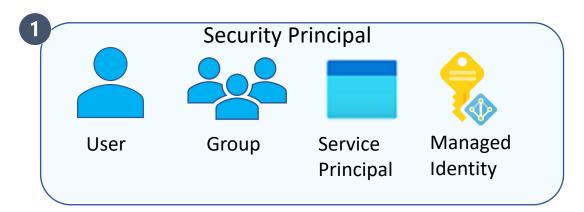
Microsoft Entra ID Roles



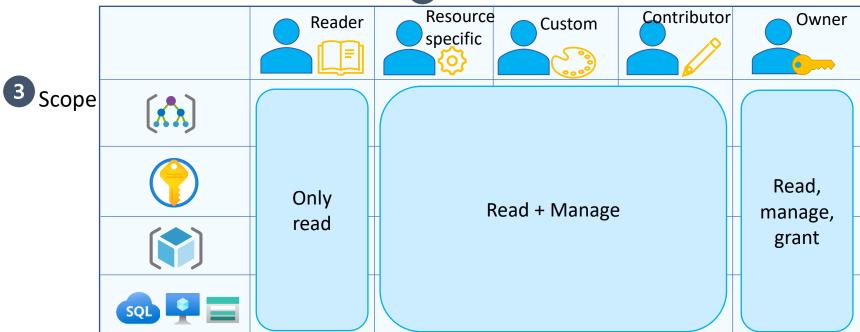
Most important Entra ID Roles from the data perspective:

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

Role Based Access Control

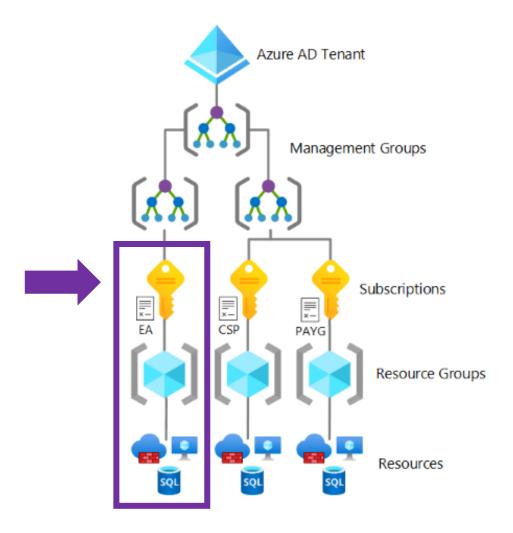


2 Role



- 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

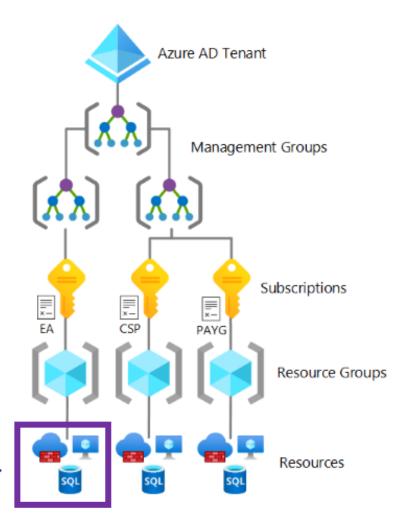
Management Groups: manage access, policy, and compliance for multiple subscriptions

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

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

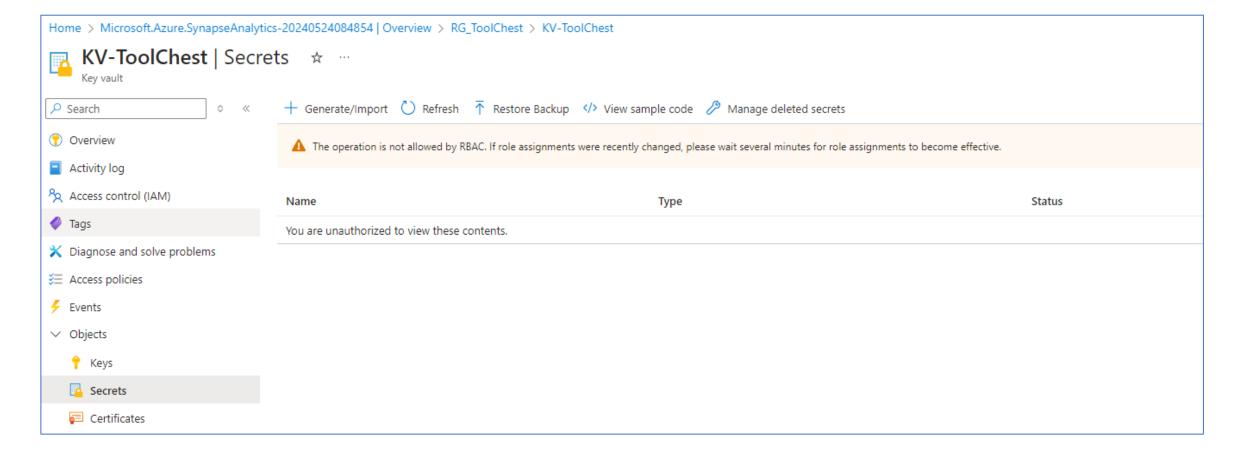
Resources: instances of services

Resource Specific RBAC 🍎 🥸



An example for resource specific RBAC is Azure Key Vault





Resource Specific RBAC 😂



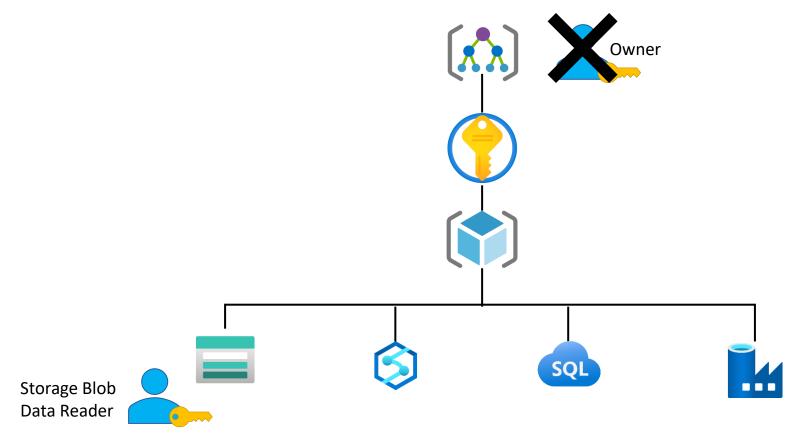
An example for resource specific RBAC is Azure Key Vault



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 k
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.

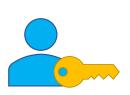
Principle of least privilege

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



Tips when you are not





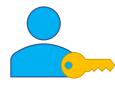
Privileged Identity Management





Stay on good terms with the people who can grant roles

User Access Admin (RBAC)



Principle of least privilege

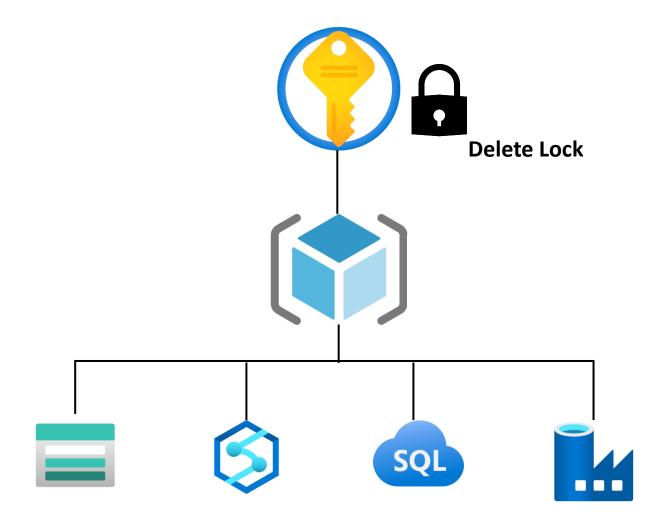


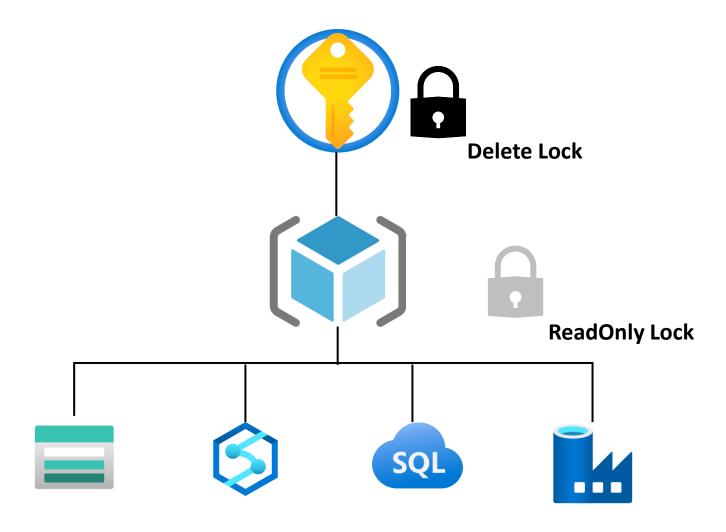


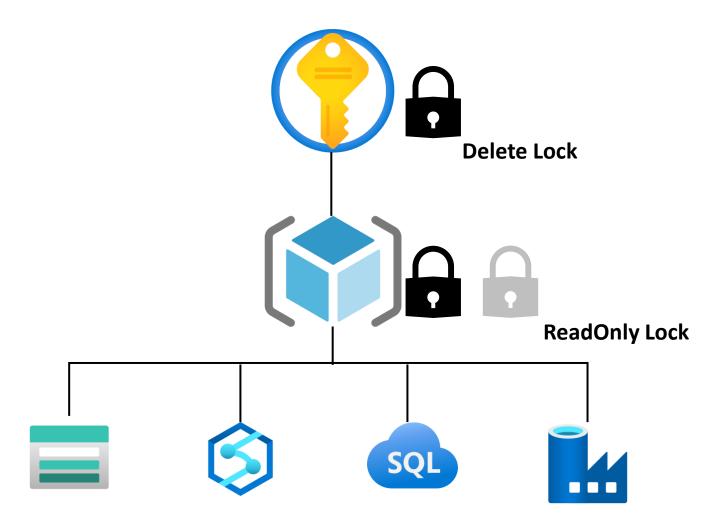
Locking resources can help ensure users don't accidently delete or modify resources.

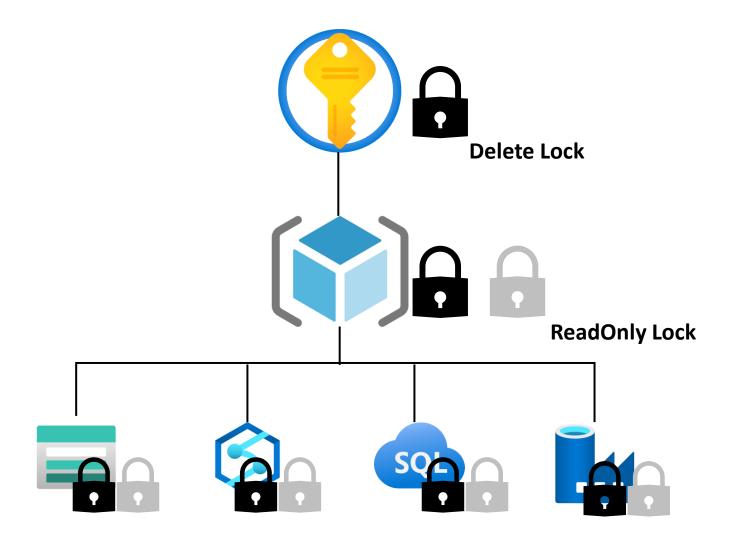


Owners or **user access administrators** may create or remove locks. They must remove the lock before they can delete a resource.

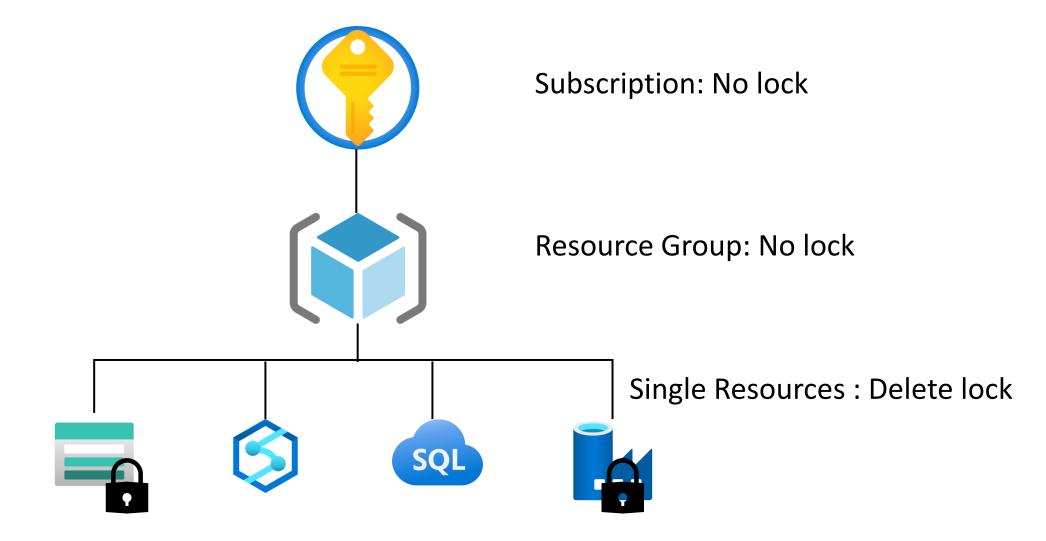






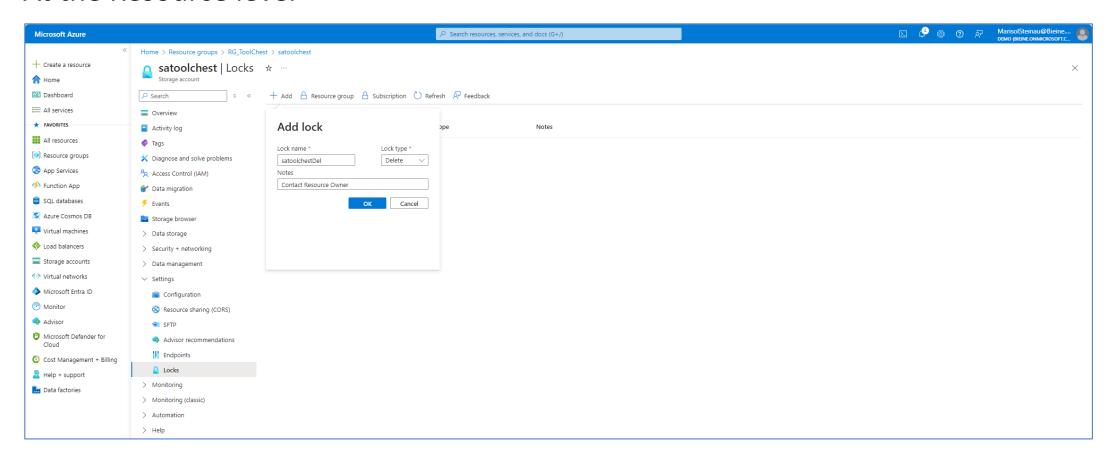


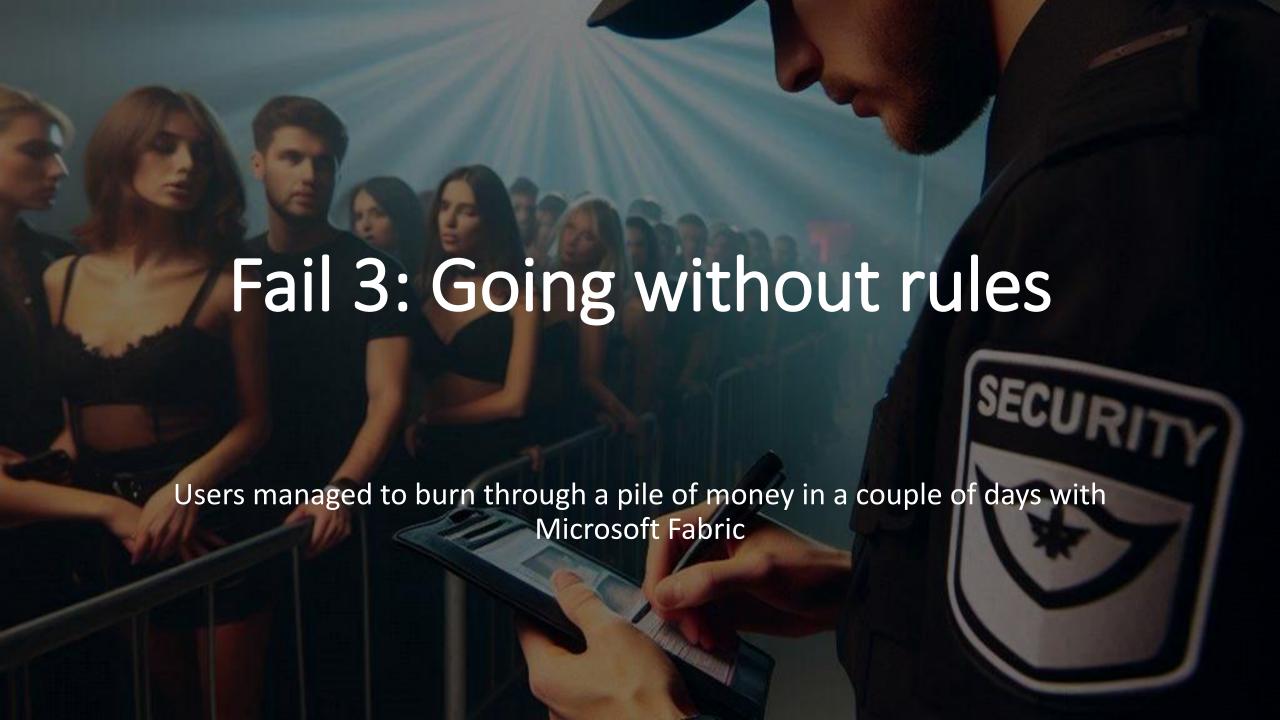
Recommendation



Applying locks

At the Resource level





Azure Policies

Corporate Standards

Do's

Fabric SKU < F64

HTTPs

Azure Tags

• •

Dont's

Fabric SKU >= F64

Spark Pool SKU > Medium

Location <> Europe

• • •

Azure Policies

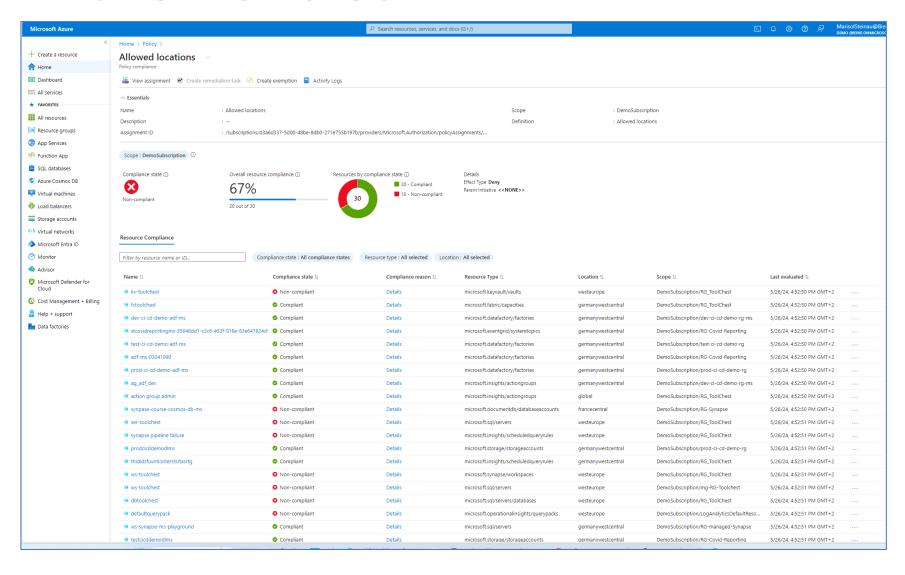
Scope **Azure Policy** Azure Policies help in controlling, restricting and auditing resources

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)

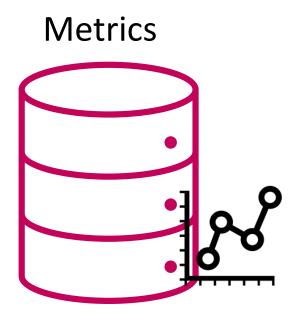
Azure Policies





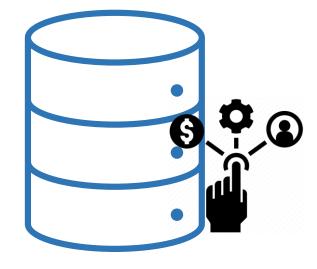


Azure Monitor



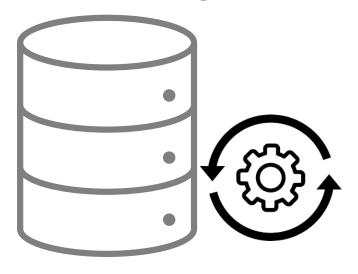
Numerical values

Activity Logs



Operations on resources

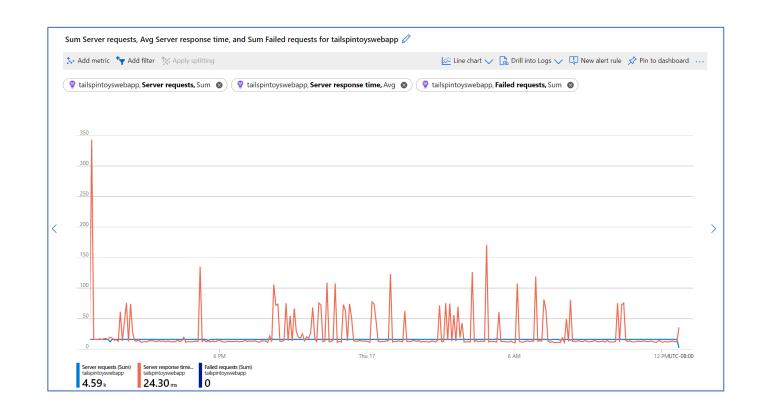
Resource Logs



Operations by resources

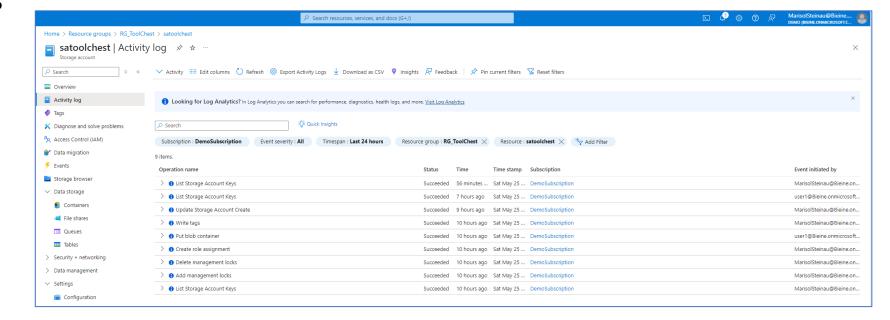
Metrics

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



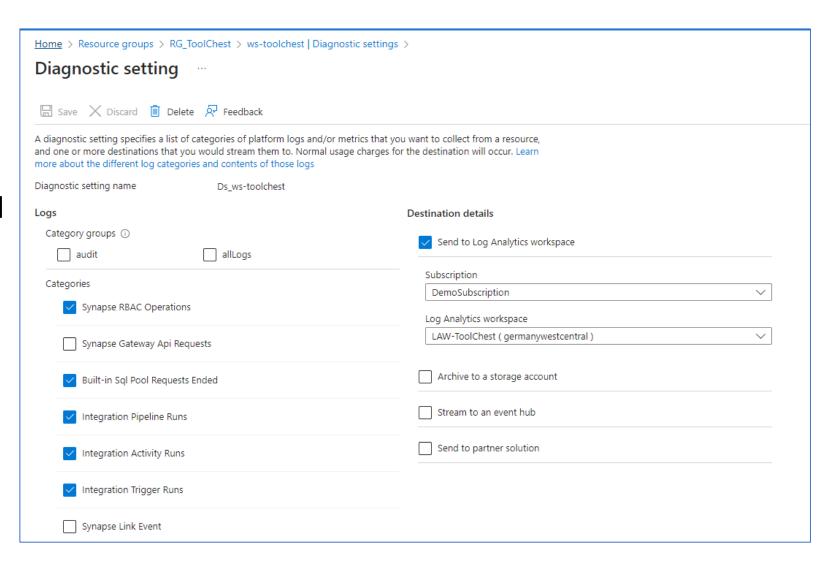
Activity Logs

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

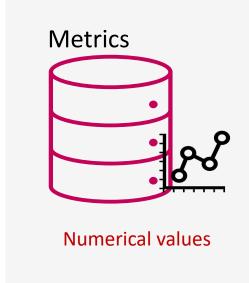


Resource Logs

- 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

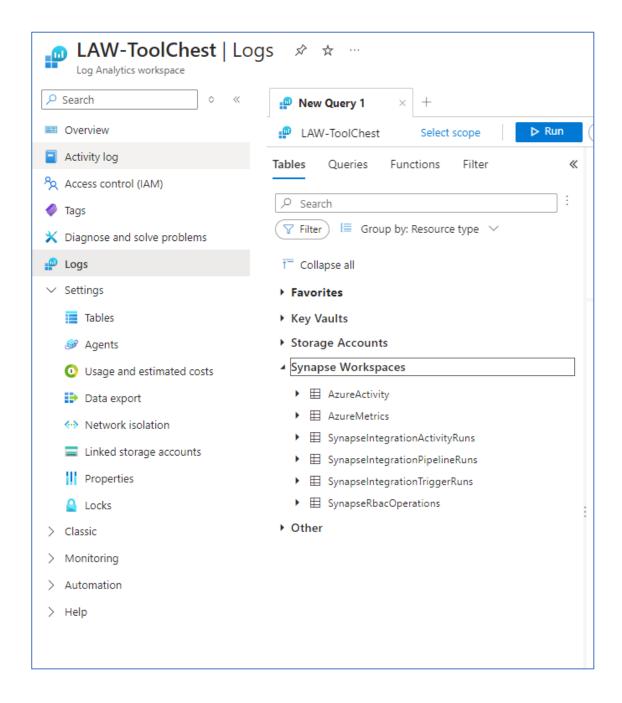


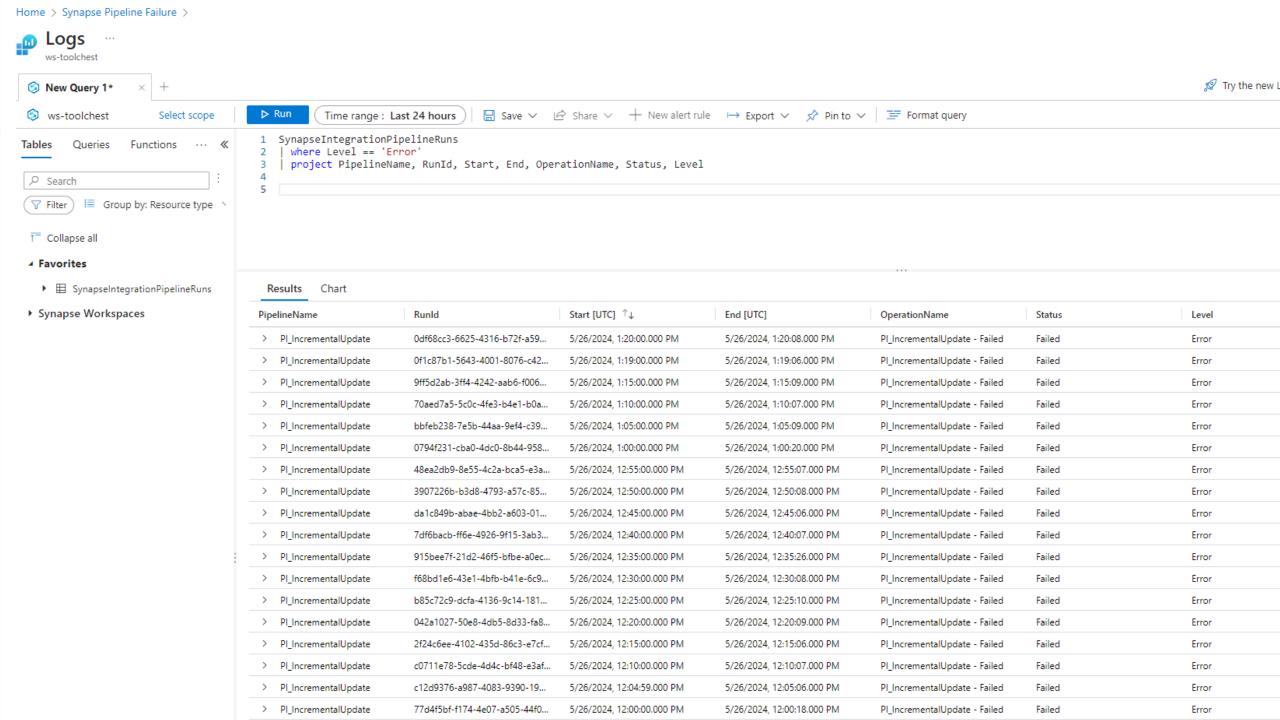
Azure Storage



Partner Solution

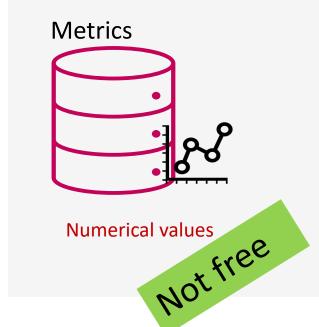






Alerts

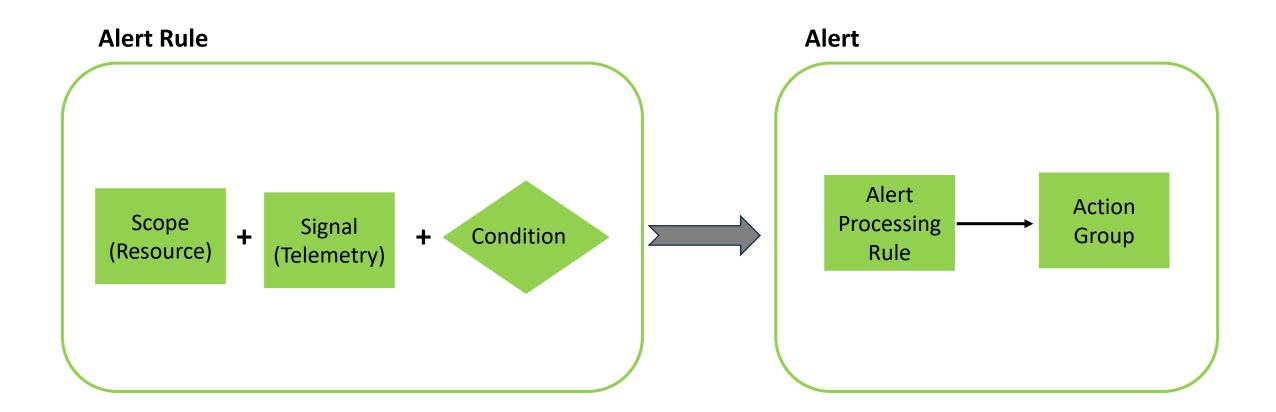








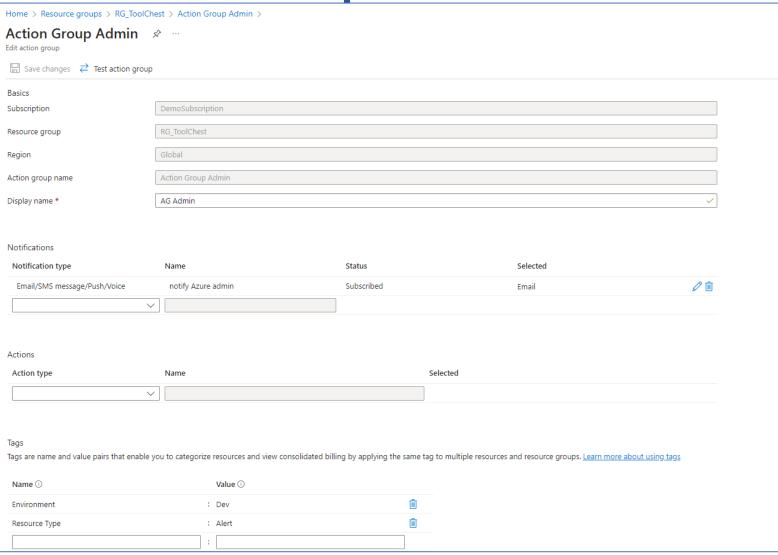
Alerts



Alerts

Home > Resource groups > RG_ToolChest > Synapse Pipeline Failure >				
Edit alert rule				
Scope Condition Actions	Details Tags Rev	riew + save		
Configure when the alert rule should	trigger by selecting a sigr	al and defining its logic.		
Signal name * ①	P Custom log s	earch	<u> </u>	
	See all signals			
Define the logic for triggering an aler	t. Use the chart to view tro	ends in the data. Learn more		
	gs. The results returned b	y this query are used to populate the aler	t definition below.	
Search query *			•	
SynapseIntegrationPipeline where Level == 'Error'	eRuns			
	unId, Start, End, Op	perationName, Status, Level		
View result and edit query in Logs 🛮	D			
Measurement				
Select how to summarize the results	We try to detect summari	zed data from the query results automatic	cally	
Selection to summarize the results	vic try to detect summan	sed data from the query results datematic	oury.	
Measure ①	Table rows	~		
Aggregation type ①	Count			
Aggregation granularity ①	1 minute	<u> </u>		
Split by dimensions				
		ntext to the fired alert. Dimensions can be e combination will trigger its own alert ar	e either number or string columns. If you select more and will be charged separately. ①	
Dimension name	Operator	Dimension values	Include all future values	
Select dimension	✓ =	∨ 0 selected	✓ □	
		Add custom value		

Action Group

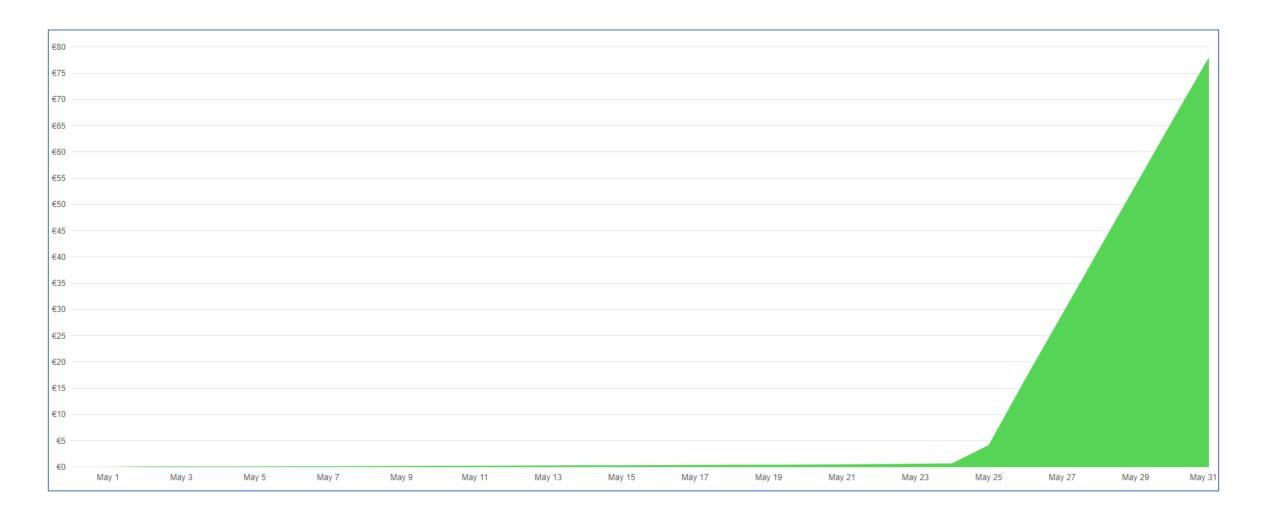


Notification

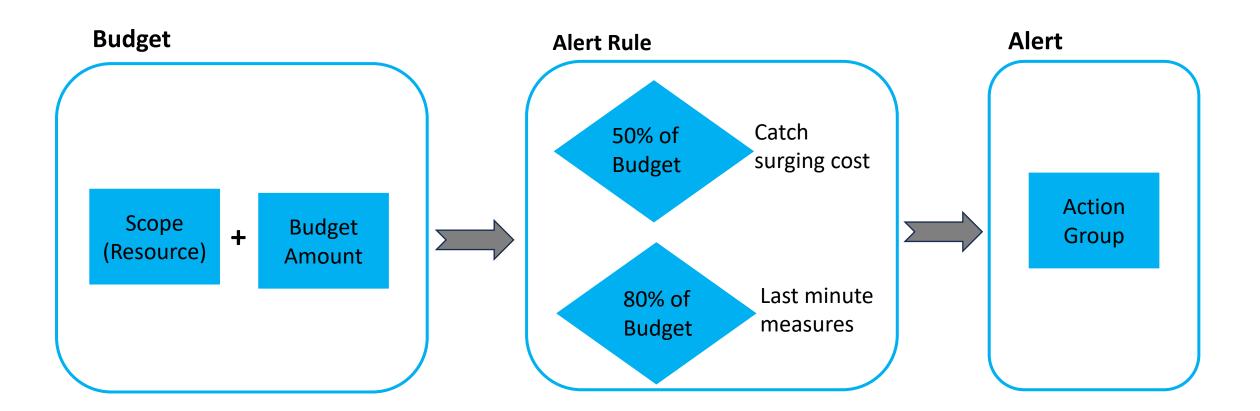
Fired:Sev1 Azure Monitor Alert Synapse Pipeline Failure on ws-toolchest (microsoft.synapse/workspaces) at 5/26/2024 2:21:01 PM View the alert in Azure Monitor > Investigate > Summary Alert name Synapse Pipeline Failure Severity Monitor condition Fired Affected resource ws-toolchest microsoft.synapse/workspaces Resource type Resource group rg_toolchest Log Alerts V2 Monitoring service Signal type Log May 26, 2024 14:21 UTC Fired time 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 ['Microsoft.Synapse/workspaces'] Target resource types Time aggregation Count Operator GreaterThan Threshold Metric value Number of violations Number of examined periods



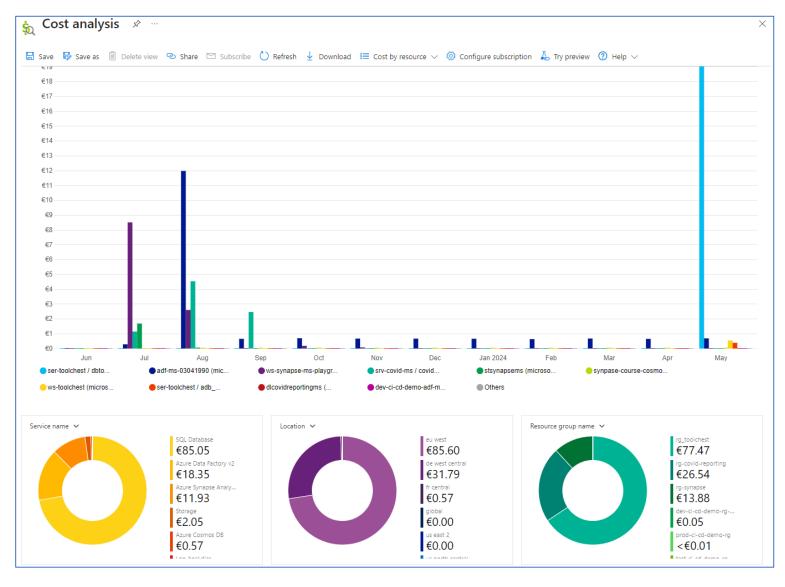
Cost Management



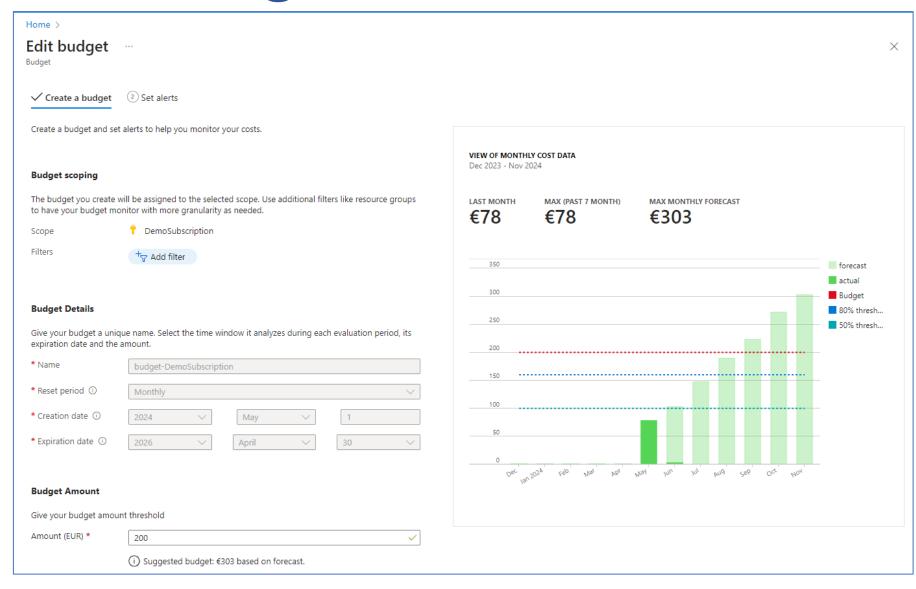
Cost Management



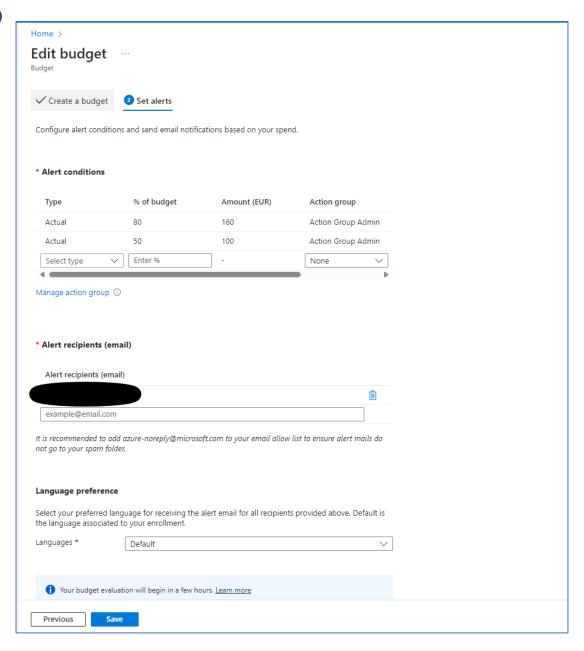
Cost Management



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 >

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

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

Thank you for your attention. Any questions?

Feedback is kindly appreciated

