



Microsoft Fabric

This presentation is the property of Microsoft and is intended for informational and educational purposes only. You may use, copy, and distribute this presentation for your personal, non-commercial purposes. You may not modify, alter, or create derivative works from this presentation without the prior written consent of Microsoft. You may not use this presentation to misrepresent, defame, or disparage Microsoft or its products, services, or affiliates. You may not use this presentation to endorse or promote any other products, services, or organizations without the prior written consent of Microsoft.

By using this presentation, you agree to abide by these terms. If you do not agree, you must not use this presentation. Microsoft reserves the right to change these terms and conditions at any time without notice. Microsoft disclaims any and all warranties, express or implied, relating to this presentation, including but not limited to the accuracy, completeness, timeliness, or suitability of the information contained herein. Microsoft is not liable for any damages, losses, or liabilities arising from your use of or reliance on this presentation.

Please review the terms of use posted in the content library.

#FABCONSQLCON2026

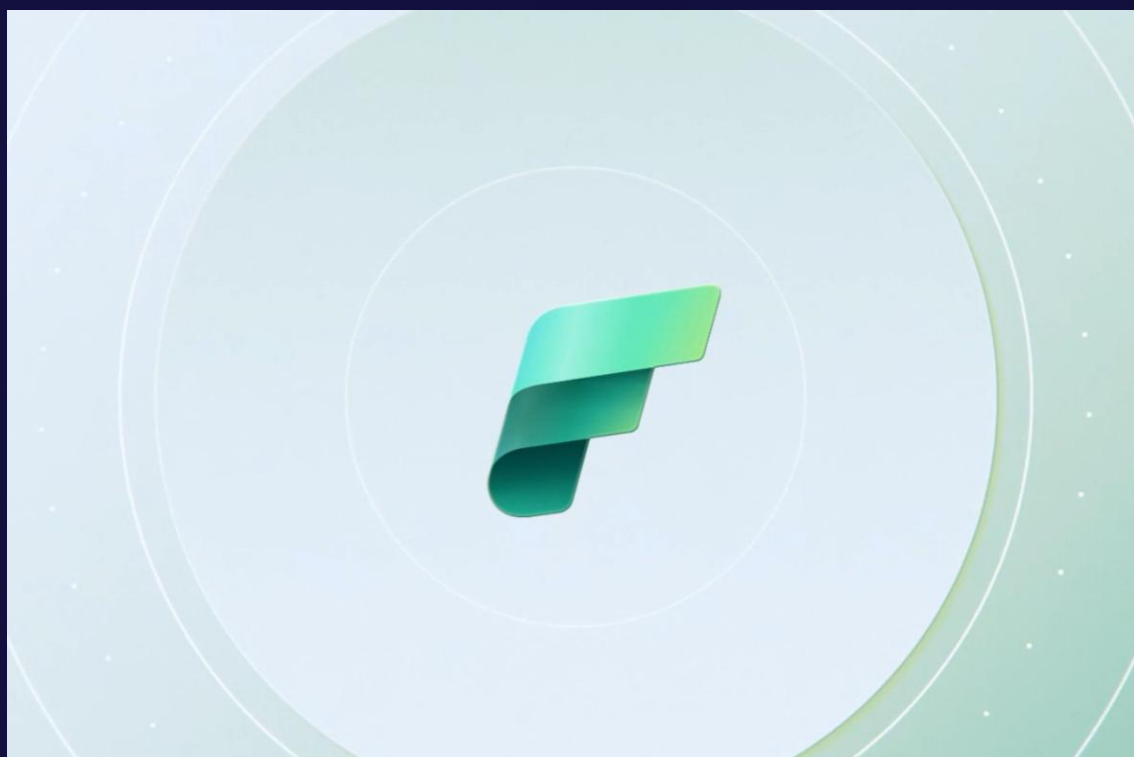
FABCON

Microsoft Fabric
COMMUNITY CONFERENCE

SQLCON

Microsoft SQL
COMMUNITY CONFERENCE

ATLANTA MARCH 16 - 20, 2026



Fabric Capacity Management & Insights

Fabric Platform Team

Pankaj Arora • Arsalan Yarveisi

Let's talk about Capacities...

- 1 What are they and why do they matter?
- 2 How are we investing in making them even more powerful?
- 3 What are some tactics and best practices for managing capacities?

We'll also show some demos and save time for Q&A!

What are Capacities?

Capacities Enable Develop, Deploy, and Manage Compute

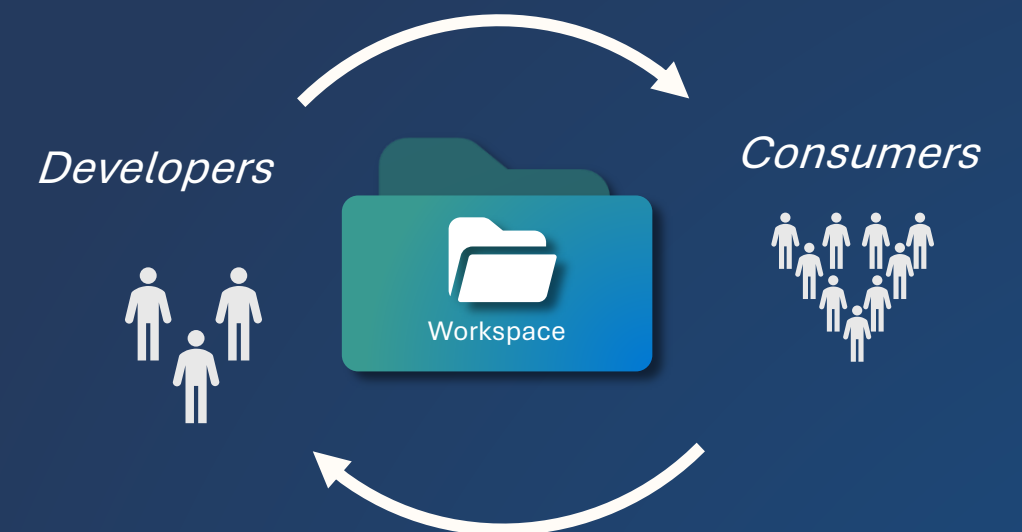
Fabric Capacities provide the compute you need

Like a “CPU” in the cloud
Single Capacity Unit (CUs)
Self-managing
Universal
Flexible

Complete Solutions
Any workload, same capacity

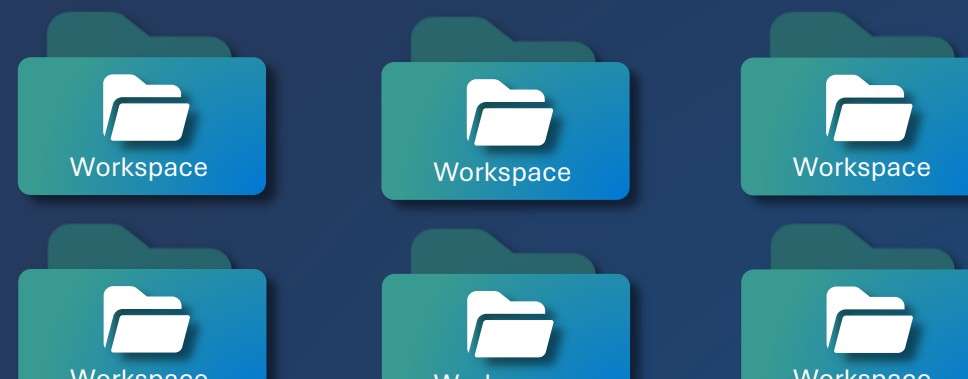


Empower all users
Engineers, Developers, Citizen Dev,
and Consumers can collaborate



Many Projects

Numerous workspaces owned by different business
units

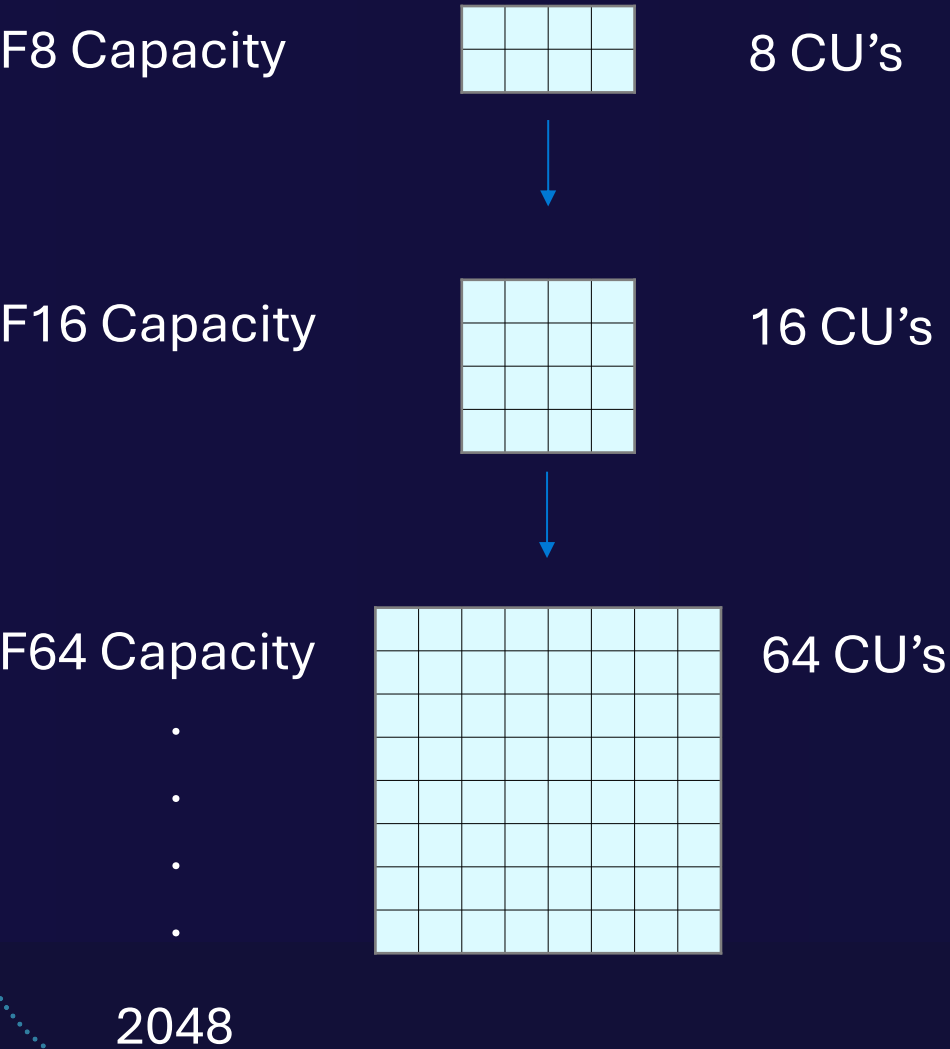


Capacities are flexible building blocks for growth

Capacities can be configured in endless ways to meet scale, usage and governance requirements while tuning to minimize total cost of ownership and achieve performance goals

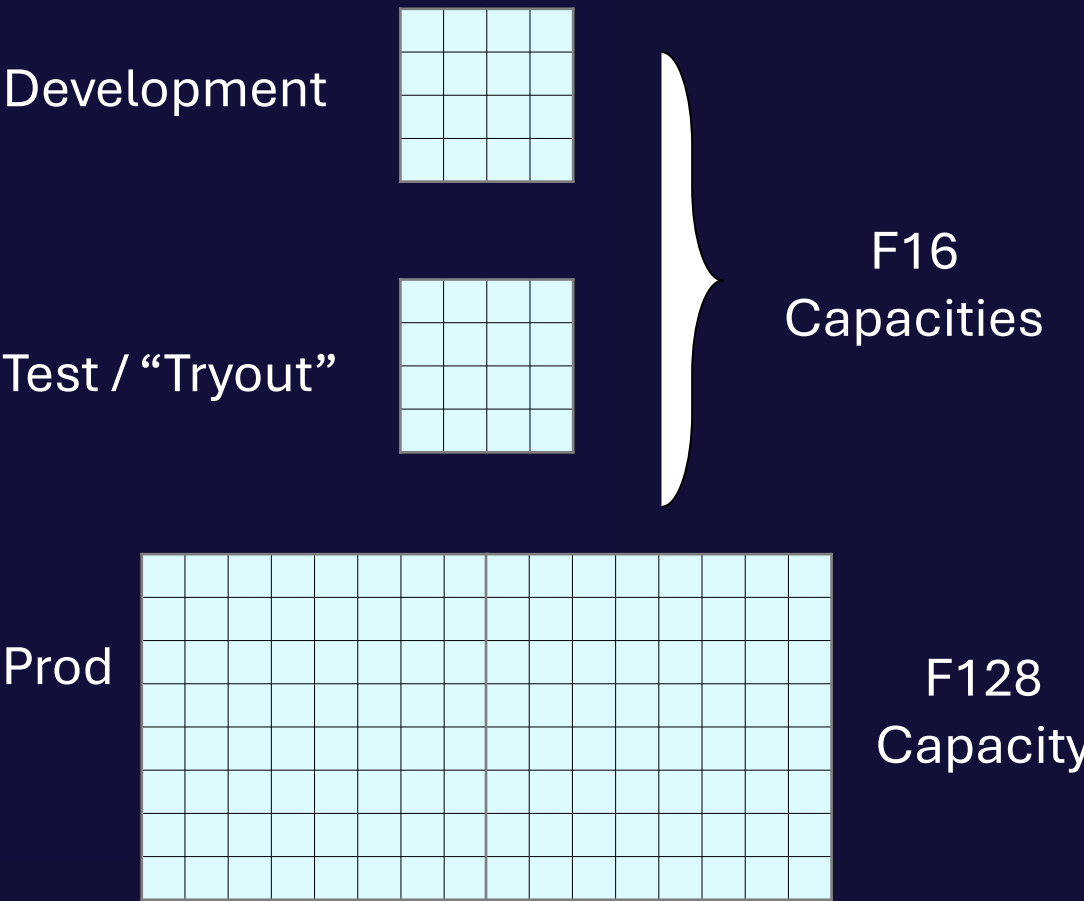
Scale Vertically

Increased capacity size provides more throughput



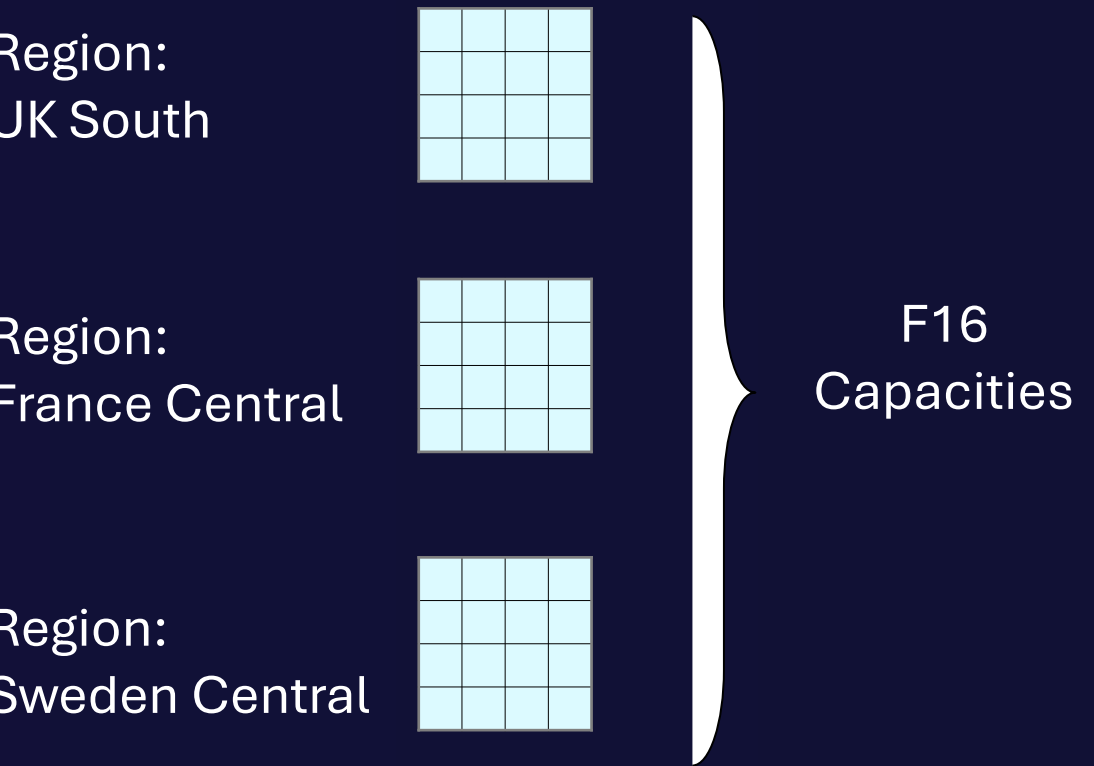
Scale Horizontally

Scale horizontally using the benefits of modular design for hardened isolation and governance



Regional Availability

Use different capacities for different regions to support GDPR / Data residency requirements



Throttling

Overuse can result in three states of throttling...

Usage	Policy Limits	Platform Policy Experience Impact
Usage <= 10 minutes	Overage protection	Jobs can consume 10 minutes of future capacity use without throttling.
10 minutes < Usage <= 60 minutes	Interactive Delay	User-requested interactive jobs are delayed 20 seconds at submission.
60 minutes < Usage <= 24 hours	Interactive Rejection	User-requested interactive jobs are rejected.
Usage > 24 hours	Background Rejection	All requests are rejected.

... however, life is better than it sounds

Fabric is different, in a good way...

Bursting

Smoothing

Surge Protection



Great performance and simplified management with bursting and smoothing



Bursting

Bursting allows you to consume extra compute resources beyond what you've purchased to speed the execution of a workload. For example, instead of running a job on 64 CU in 60 seconds, bursting could use 256 CUs to complete in 15 seconds.



Smoothing

Fabric runs smoothing on all capacities when small usage bursts occur, flattening compute usage over time to help you avoid being penalized with additional cost.

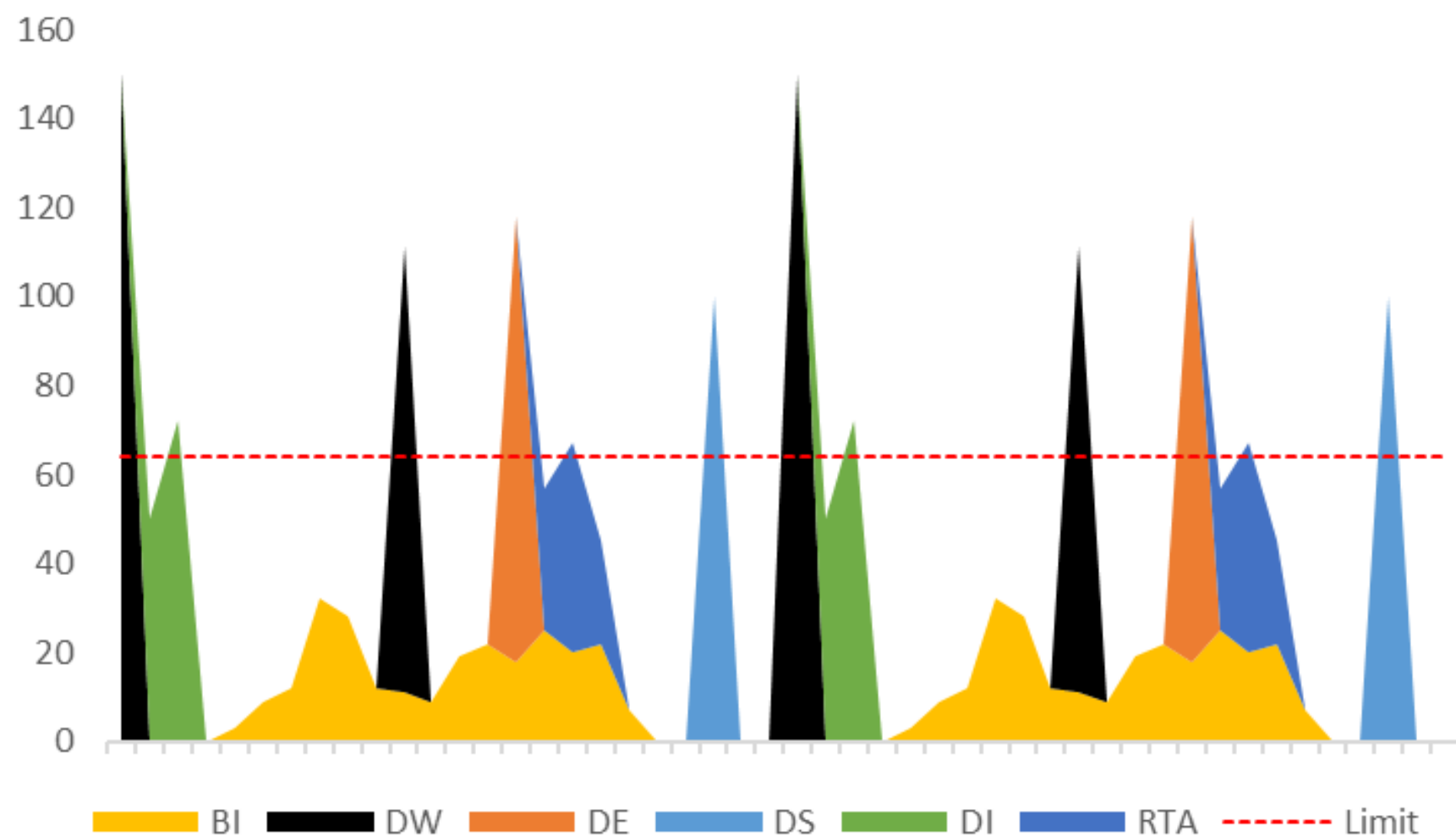
- **Interactive jobs:** Demand is smoothed over a minimum of 5 minutes (5–64 mins).
- **Scheduled/background jobs:** Demand is spread over 24 hours.

Bursting and smoothing | before and after

- Click to edit Master text styles

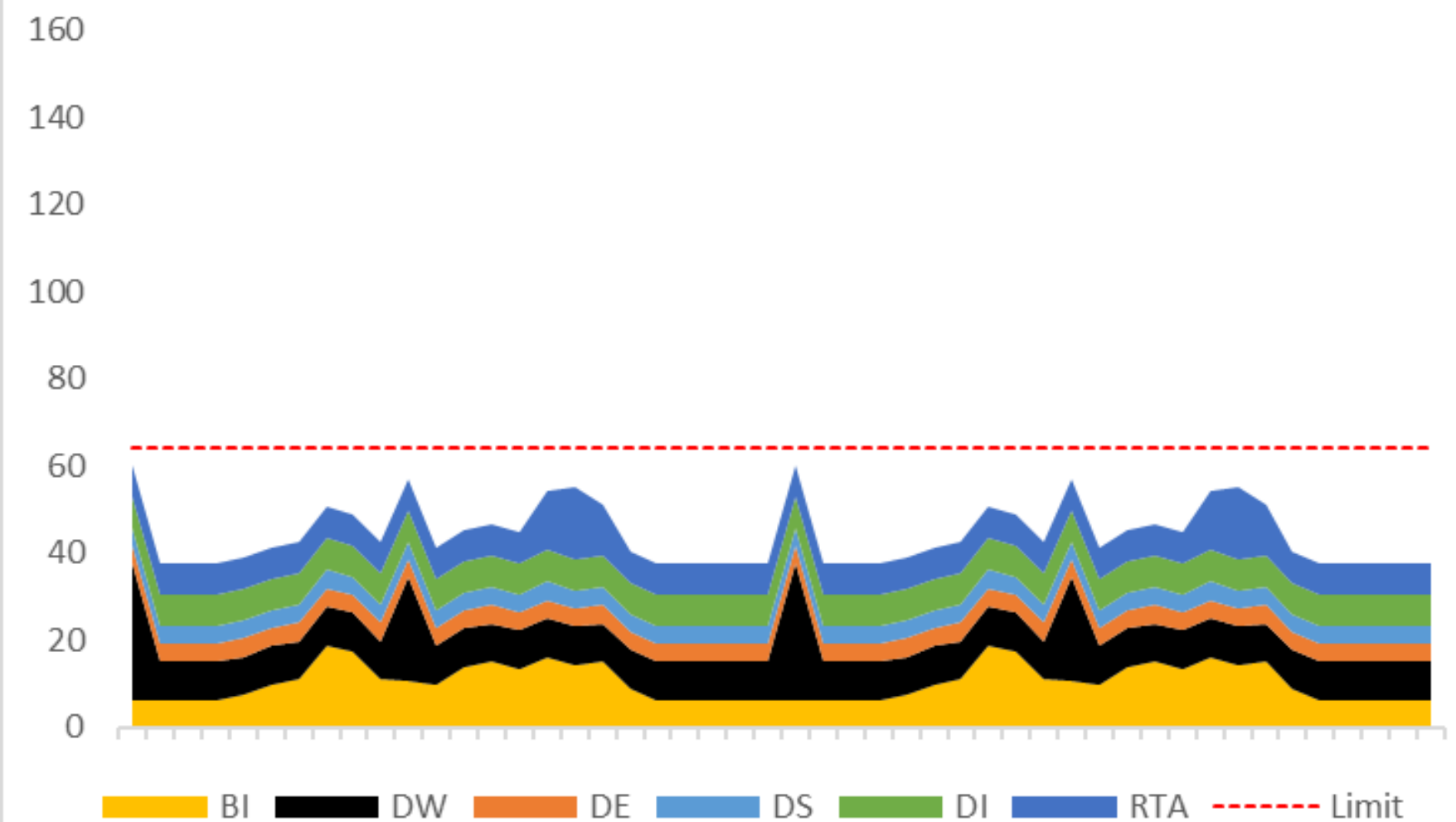
Before Smoothing

- Actual load as executed on the capacity before smoothing
- *Bursting* accelerates jobs execution by resource boosting
- The capacity could be overloaded 25% of the time
- Some of the overloads are more than 2x the limit
- There are periods of no/low usage



After Smoothing

- Shows the reported load (not runtime execution) against the capacity limits
- There is NO overload, and consumption is more stable
- The smoothing of usage fills in gaps



Capacity Management

Capacity-level Surge Protection



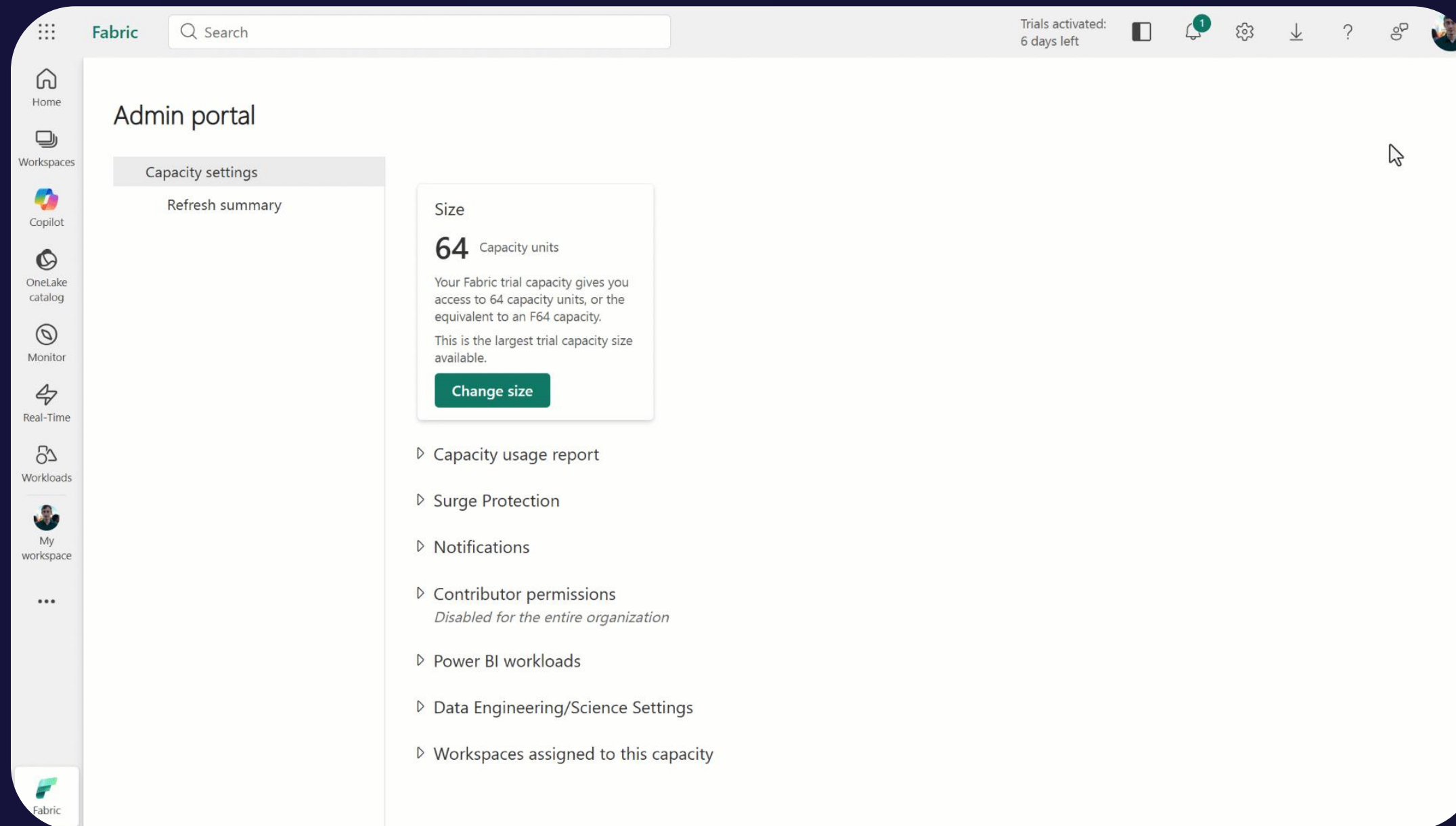
Protects capacities from
unexpected surges

Admins can set limits on
**background activity
consumption**, preventing
background jobs from
starting when limit is
reached

New in Public Preview:
**Workspace-level surge
protection...**

Surge Protection (Workspace-level)

Surge Protection V2 (Workspaces): Empowering Capacity Admins to avoid throttling through capacity limits



Apply workspace-level surge limits to accommodate capacity demand

Block noisy workspaces when needed during usage spikes

Keep essential workloads running with Mission Critical mode

Manage custom behavior directly from the Admin Portal

Introducing... Capacity Overage Protection

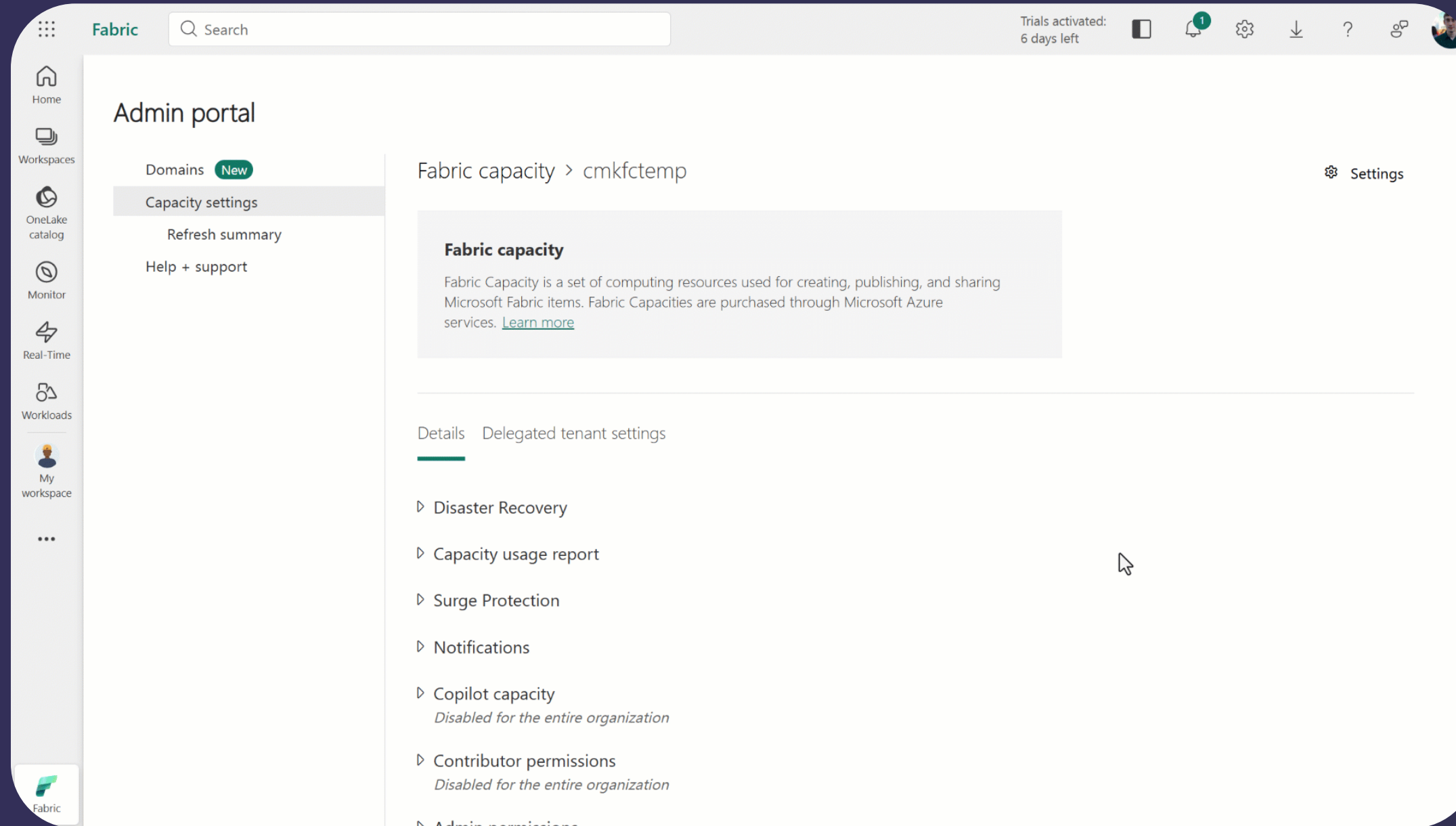
Empowering Capacity Admins to avoid throttling through paying for overages

Automatically prevent throttling during periods of peak demand

Pay only for extra usage when capacity limits are exceeded

Set 24-hour spending limits directly in the Admin Portal

Right-size your SKU—no need to over-provision for spikes



+ New report

▼

New items saved to:

My workspace



Uncover insights in your data with the help of AI

Ask a question about your data

+

Find reports about [a topic]

Prep a summary for my team about report

What can Copilot help me with?

See more ▼

- Recent
- Favorites
- My apps
- From external orgs

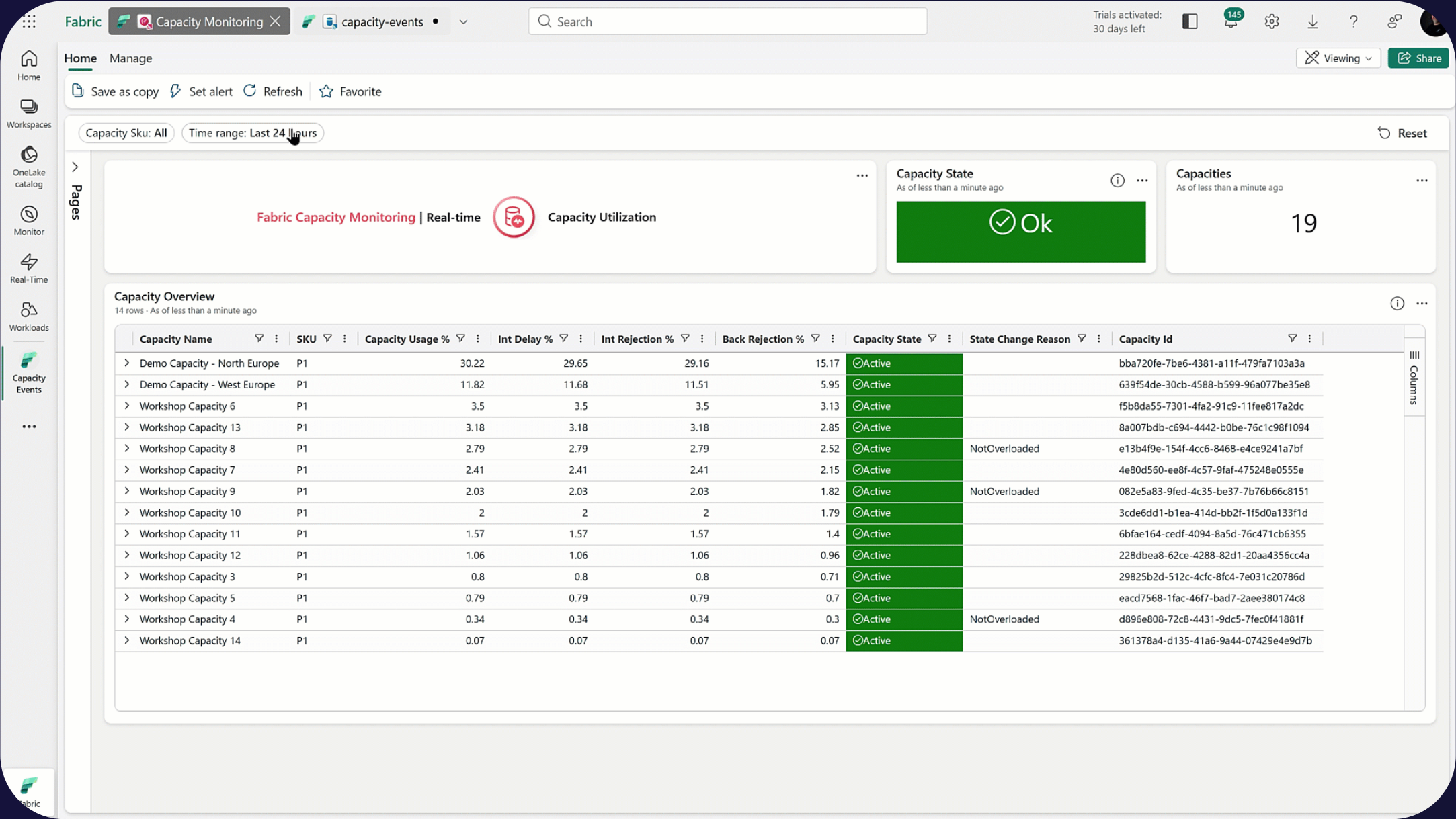
Filter by keyword

Filter ▼

	Name	Type	Opened	Location	Endorsement	Sensitivity
	ad-Test	Workspace	3 hours ago	Workspaces	—	—
	OneLake Catalog Governance for Admins	Report	2 days ago	Admin monitoring		

Capacity Events in Real-Time Hub

Empowering Capacity Admins to manage their capacities with data



Harness near real-time telemetry to monitor capacity health

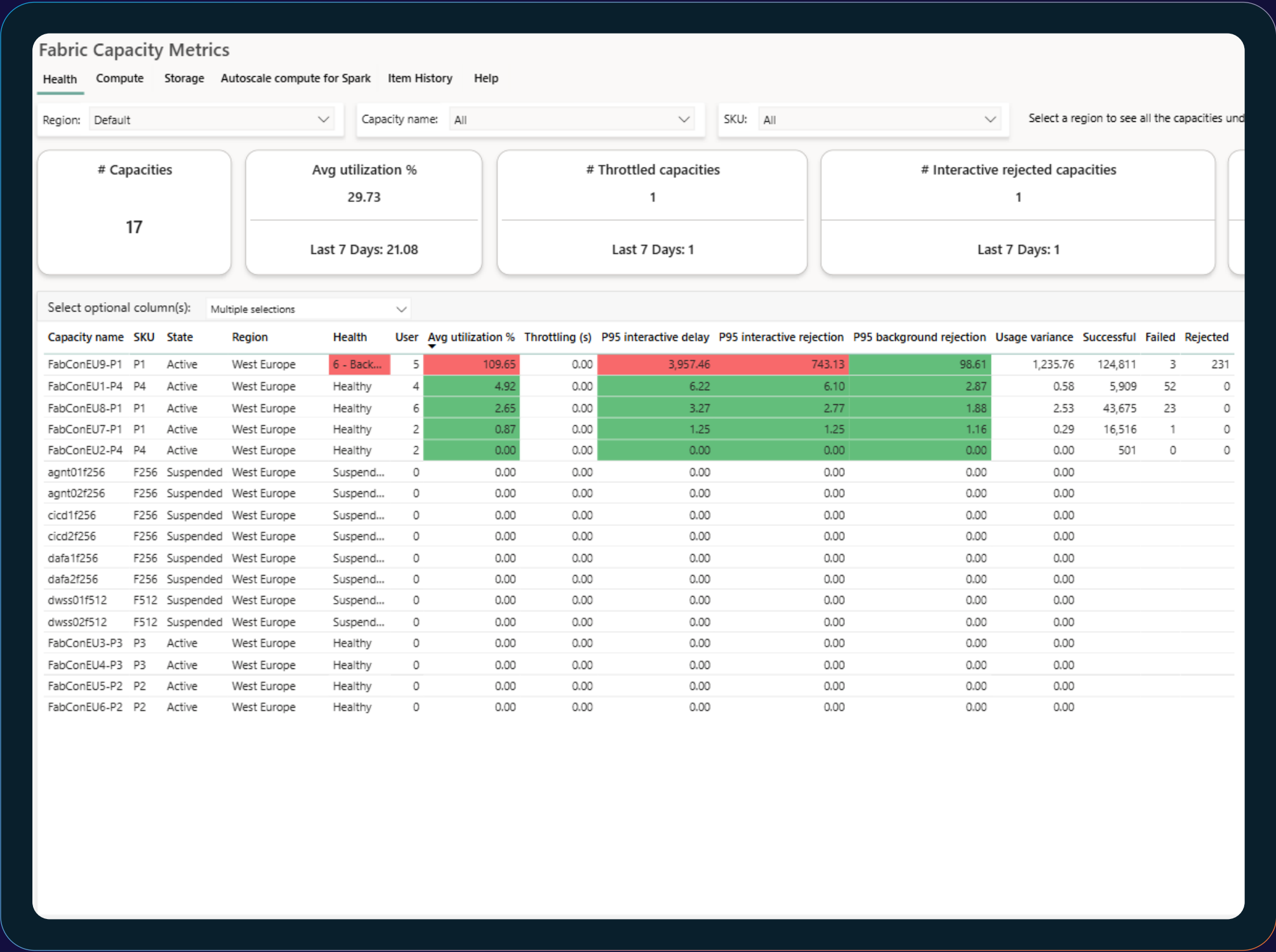
Trigger alerts and actions using RTI, Eventhouse, and Activator

Integrate capacity signals with custom systems of your choice

Coming soon

Track granular capacity usage with Operation Events and more

Capacity Metrics App Overview



Using Fabric’s capacity metrics app, admins can monitor usage and make data-driven scale-up decisions to unlock the full potential of their data



Capacity metrics in Fabric is a tool for admins to track capacity consumption, monitor the performance of their workloads and identify performance issues

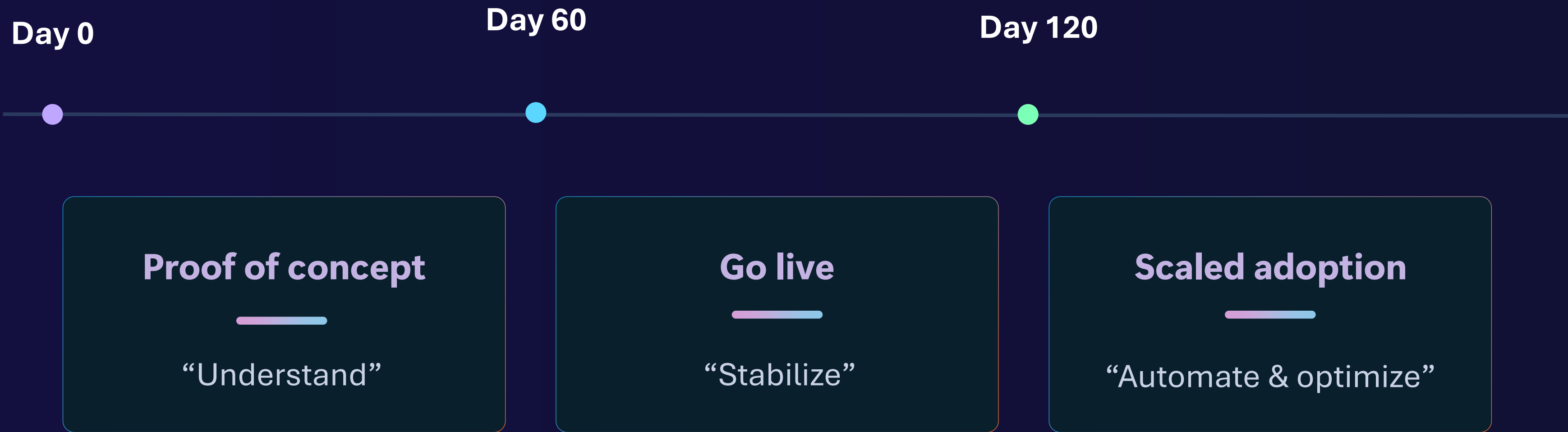


Recent Investments: Health page, Timepoint Summary, Timepoint item detail, and History



Best Practices

Zava’s journey with Fabric





Phase 1: Proof of concept

Objectives

Understand

Usage pattern insights

Early guardrail signals

Starter SKU hypothesis

Key questions

What does consumption look like for each workload?

Which workloads create spikes?

What are the peak CU windows?

Where do we see throttling?

What is a realistic SKU based on total projected CU usage?



Best practices for “proof of concept” phase

- | | |
|------------------------------|---|
| Start small | Narrow-scoped project for a limited set of users, in an isolated capacity/workspace |
| Fabric trial capacity | Costs nothing to get started and provides compute throughput equivalent to F64 |
| Capacity Metrics app | Your source of truth for detailed usage data |
| SKU Estimator tool | Starting point for capacity planning based on anticipated requirements |



Phase 2: Go live

Objectives

Stabilize

Predictable performance

Validated SKU size

Key questions

Are users constantly experiencing interactive delays?

Which workloads should run isolated?

Are we throttled due to consistent overload or occasional spikes?

Do we need multiple capacities?

Are we underutilized or constantly throttled?

What is the best “value SKU”?



Best practices for “go live” phase

Monitoring

Identify throttling patterns, top consumption drivers, and peak times in Capacity Metrics

Isolation

Separate production and development capacities

Leverage a separate, smaller capacity to gauge CU consumption of experimentations

Isolate critical workloads from the rest to ensure a reliable experience

Spark Auto-scale

Enable it if you observe occasional spikes in Spark

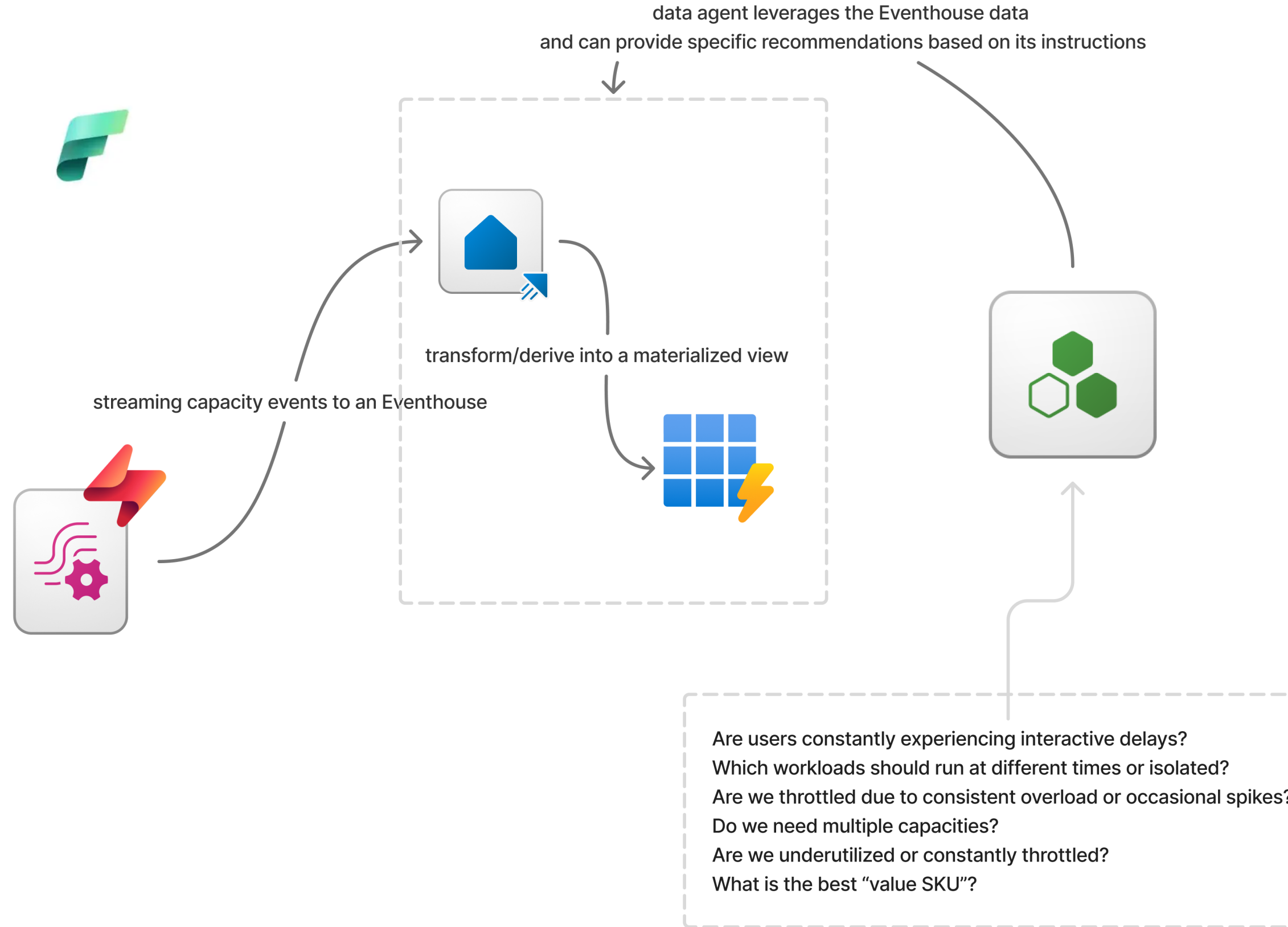
Iterative sizing

Observe → hypothesize → resize (scale up / down / out) → validate

Forecast future needs based on pilot usage extrapolation

Build a **Data Agent** on top of RTH to help with capacity data analysis and recommendation

Example: Capacity Analyst Agent





Phase 3: Scaled adoption

Objectives

Automate & Optimize

Optimized cost/performance

Automated governance

Key questions

Can we detect overload early enough to avoid user impact?

What actions can we fully automate?

Which workloads are critical path and must be protected?

How do we optimize CU utilization and idle capacity cost?

Which workloads should run on different SKUs?

Can we scale dynamically without surprise costs?



Best practices for “scaled adoption” phase

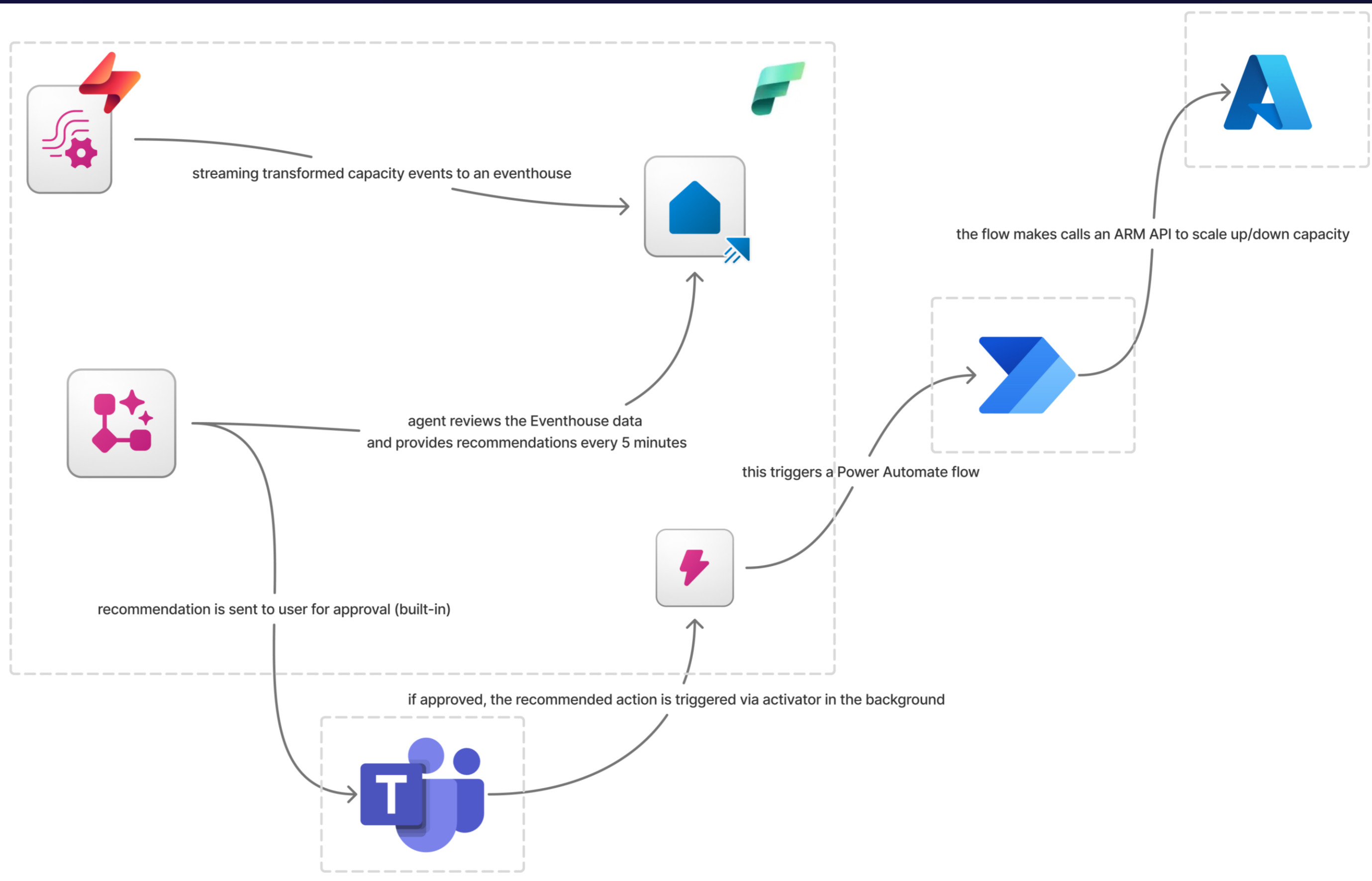
Real-time monitoring Leverage **Capacity Events in Real-Time Hub** capability for proactive management

Cost optimization Leverage **pause/resume** to minimize idle time costs
Consolidate capacities when possible (considering region, SLA, cost centers, etc.)
Lock in a 1-3 year commitment to benefit from **RI discounts** (~40%)

Performance protection Enable **Surge Protection** on capacity or workspace level to manage over-utilization
Configure **Capacity Overage** to pay for excess CU usage and not get throttled

Automations Set up **alerts** on certain capacity events to stay on top
Leverage **time-based** or **event-driven automations** (scheduled pause/resume, off-peak scale down, etc.)
Build a **capacity governor agent** to manage the capacity based on your goals!

Example: Capacity Governor Agent





Data preview Data insights Refresh

Hide settings Search Show details

Data type: Json

Last refreshed: 02/25/26 09:06:24 AM

source	s...	subject	d...	id
72f988bf-86f1-41af-91ab-2d7cd011db47	1.0	/capacities/f74fe639-0f7d-48e0-8bf8-194f9500c685	1.1	8046b425-a382-4b83-83cf-81
72f988bf-86f1-41af-91ab-2d7cd011db47	1.0	/capacities/f74fe639-0f7d-48e0-8bf8-194f9500c685	1.1	b30dc39d-f931-4629-b426-a4
72f988bf-86f1-41af-91ab-2d7cd011db47	1.0	/capacities/f74fe639-0f7d-48e0-8bf8-194f9500c685	1.1	6a7f5838-575e-4c16-ace4-c77
72f988bf-86f1-41af-91ab-2d7cd011db47	1.0	/capacities/f74fe639-0f7d-48e0-8bf8-194f9500c685	1.1	855e8307-837a-4964-9d2c-10
72f988bf-86f1-41af-91ab-2d7cd011db47	1.0	/capacities/f74fe639-0f7d-48e0-8bf8-194f9500c685	1.1	0ec44ea5-9ae6-4855-a072-3b
72f988bf-86f1-41af-91ab-2d7cd011db47	1.0	/capacities/f74fe639-0f7d-48e0-8bf8-194f9500c685	1.1	a0a5c9f5-1175-47ae-b9b0-b6
72f988bf-86f1-41af-91ab-2d7cd011db47	1.0	/capacities/f74fe639-0f7d-48e0-8bf8-194f9500c685	1.1	344d59db-6596-4a13-aede-64
72f988bf-86f1-41af-91ab-2d7cd011db47	1.0	/capacities/f74fe639-0f7d-48e0-8bf8-194f9500c685	1.1	12bc35e0-2c06-4c19-a5eb-56

Sound off.
The mic is all yours.
Influence the product roadmap.

Join the Fabric User Panel



Share your feedback directly with our
Fabric product group and researchers.

<https://aka.ms/JoinFabricUserPanel>

Join the SQL User Panel



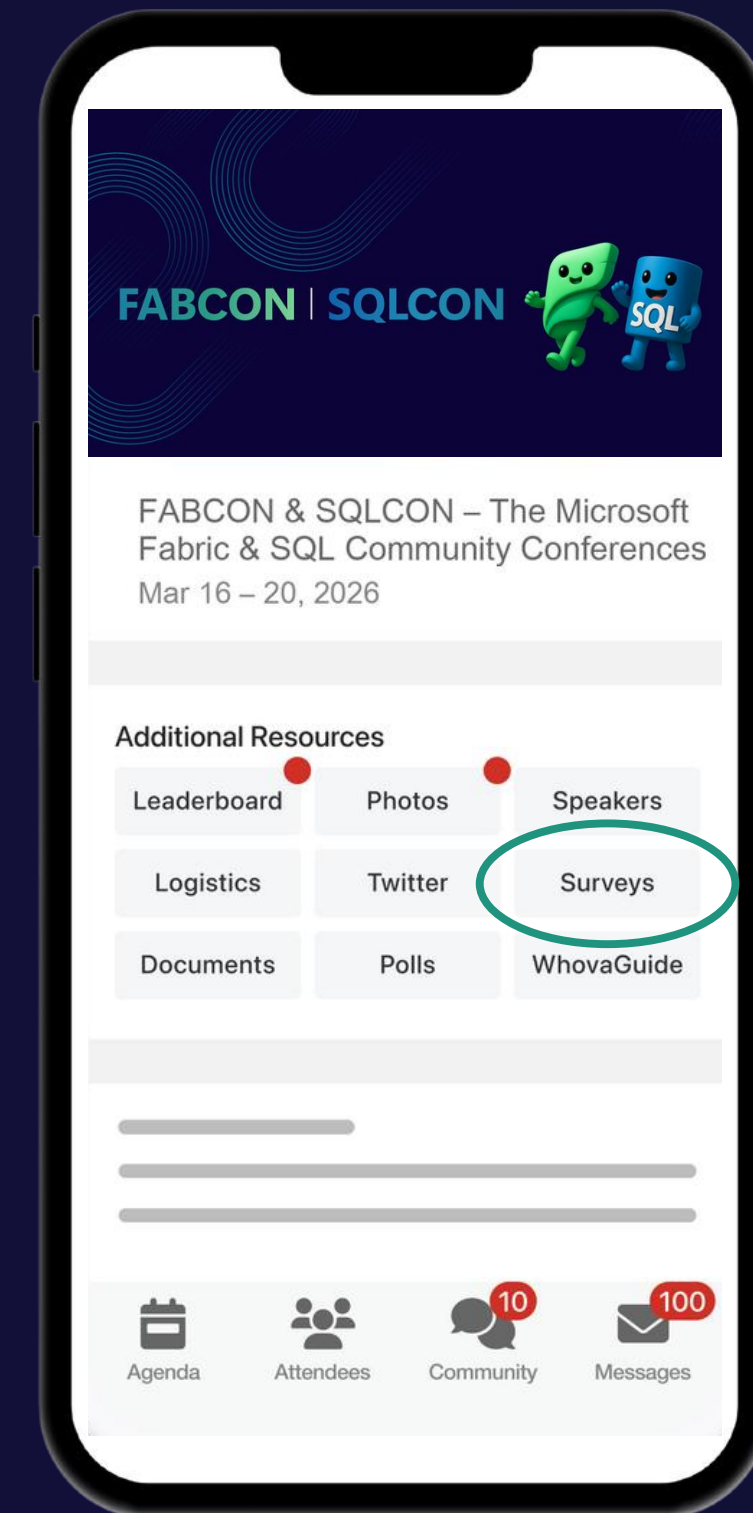
Influence our SQL roadmap and ensure
it meets your real-life needs

<https://aka.ms/JoinSQLUserPanel>

How was the session?



Complete Session Surveys in
Uhova for your chance to WIN
PRIZES!

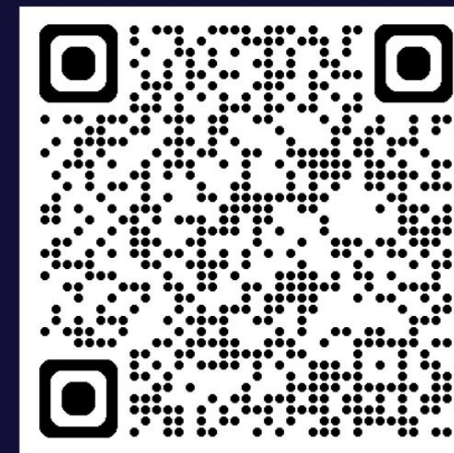


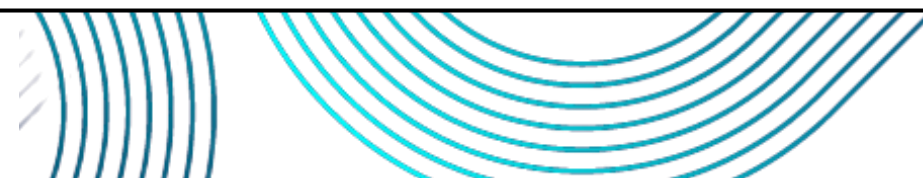
Get Two Fabric Certifications for FREE

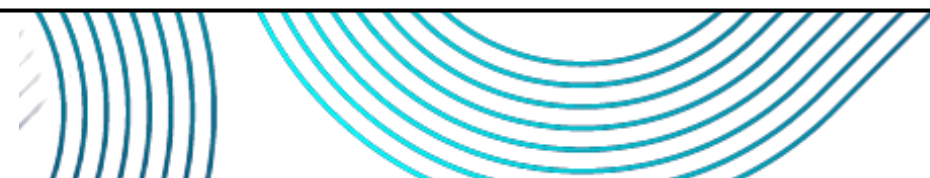
Attendees of FABCON can take the Fabric Analytics Engineer or Fabric Data Engineer exam for free. Be part of the 2 fastest growing role-based certifications in Microsoft history.

Request your voucher by March 23, 2026.

<https://aka.ms/fabcon/cert100>







Sound off.
The mic is all yours.
Influence the product roadmap.

Join the Fabric User Panel



Share your feedback directly with our
Fabric product group and researchers.

<https://aka.ms/JoinFabricUserPanel>

Join the SQL User Panel



Influence our SQL roadmap and ensure
it meets your real-life needs

<https://aka.ms/JoinSQLUserPanel>

How was the session?



Complete Session Surveys in
Whova for your chance to WIN
PRIZES!

