



MICROSOFT FABRIC CAPACITY REVIEW

The Five-Question Fabric Capacity Checklist

A working guide for your next Microsoft Fabric capacity review

Five questions, one test for each: can your team answer it with data, or only with a best guess?

CAPACITY REVIEW SCORECARD

- | | | |
|---|------------------|------------|
| 1 | Who consumes it | With data |
| 2 | Demand vs. spike | With data |
| 3 | What shares | Partly |
| 4 | Who owns it | Best guess |
| 5 | Value | Partly |

HOW TO USE THIS CHECKLIST

Built for one meeting: the capacity review

Who should be in the room



The Fabric capacity administrator or data platform lead



The BI lead for the reports leadership relies on



Business owners for your largest workspaces and workloads



Someone who speaks for the budget: finance, FinOps, or technology finance

How the review runs

BEFORE

Assign each question to one person. They bring the evidence; most of it comes from the Capacity Metrics App.

DURING

Work through the questions in order. A gap early in the list usually explains gaps later.

AFTER

Record ratings on the scorecard, give every follow-up an owner and a date, and book the next review.

SCORING GUIDE

Rate each question one of three ways



Yes, with data

You can show the evidence, not just describe it. Someone outside the platform team could check it.



Partly

You have some of the evidence, or you have it for some workloads but not others.



No, or a best guess

The answer depends on someone's memory or instinct.

Don't treat the total as a grade

The first question you rate below "Yes, with data" is where to start. The questions after it depend on it.

Do track ratings over time

Moving from "Partly" to "Yes, with data" between reviews is real progress, and the easiest way to show leadership it's working.

THE FIVE QUESTIONS

Each question builds on the ones before it

- | | | |
|---|--|------------|
| 1 | Do we know which workspaces and workloads consume our Fabric capacity? | Baseline |
| 2 | Can we distinguish expected demand from problematic spikes? | Demand |
| 3 | Have we decided which workloads should share capacity? | Boundaries |
| 4 | Does every major workload have a business and technical owner? | Ownership |
| 5 | Can we explain how Fabric consumption maps to business value? | Value |

QUESTION 1 OF 5 · BASELINE

Do we know which workspaces and workloads consume our Fabric capacity?

Why it matters. This is your baseline. Without it, you can't tell what normal looks like, and every spike feels like an emergency.

IF YOU CAN'T ANSWER IT YET

Establish a capacity baseline. Then export snapshots after each review: the Metrics App keeps only a short window of history.



A confident answer sounds like

- "These are our top five consumers by workspace and item, and roughly what share each uses."
- "Three are recurring and predictable. One was a one-time spike last month, and we know why."



Warning signs

- Top consumers are described from memory, not shown
- Nobody can say whether this week's biggest consumer is the usual one



Bring to the review

- Top consumers by workspace and item, from the Capacity Metrics App
- A note on whether each is recurring or a one-off
- A saved export or screenshot, so the next review can compare

QUESTION 2 OF 5 · DEMAND

Can we distinguish expected demand from problematic spikes?

Why it matters. Some peaks are planned, like the Monday refresh wave or month-end close. Mix them up with real problems and you'll overreact or miss one.

IF YOU CAN'T ANSWER IT YET

Use the shape diagnostic. High around the clock points to sizing. Spikes in specific windows point to workload management.



A confident answer sounds like

- "Our peaks are Monday mornings and month-end. We see them coming, and we're sized for them."
- "Tuesday's slowdown wasn't one of those. It traced to a background job that ran long Monday night."



Warning signs

- Problems diagnosed from one current utilization number
- No clear split of interactive and background workloads
- Warehouse queries assumed interactive (they're background, smoothed over 24 hours)



Bring to the review

- Recurring peak windows, lined up against the business calendar
- Incidents since the last review: date, workload, type, and how far up the throttling ladder it went

KEEP IN MIND FOR QUESTION 2

**The same 250% overage.
Two very different recoveries.**

~90 min

to clear if the overage is interactive work

~36 hrs

to clear if it's background work

"Expected" demand includes the backlog your own overnight jobs leave behind. Background work is smoothed across the following 24 hours, so part of tomorrow morning's load was scheduled last night.

QUESTION 3 OF 5 · BOUNDARIES

Have we decided which workloads should share capacity?

Why it matters. If you don't decide, Fabric effectively decides for you: everything shares. That's how critical dashboards end up competing with an experimental pilot.

IF YOU CAN'T ANSWER IT YET

Pick from four common splits: dev/test vs. production, business domain, workload type, AI vs. reporting. Protect high-value interactive work first.



A confident answer sounds like

- "We have a few intentional boundaries, and we can say what each one protects and why."
- "Copilot and Data Agent workloads are isolated until we have enough usage history."



Warning signs

- "Everything's on one capacity. That's where it landed."
- Leadership reports share hours with heavy background jobs
- Or the opposite: too many small capacities, compute sitting unused



Bring to the review

- A map of which workspaces sit on which capacity
- Leadership's daily reports, and the background jobs sharing their capacity and hours
- A one-line reason for each boundary. No reason? Flag it.

QUESTION 4 OF 5 · OWNERSHIP

Does every major workload have a business and technical owner?

Why it matters. Fixes nobody owns come undone, one new workload at a time. When the next backlog hits, “whose job caused this?” needs an immediate answer.

IF YOU CAN'T ANSWER IT YET

Start with showback: make consumption visible to the team that creates it. You don't need chargeback on day one.



A confident answer sounds like

- “Every major workload has two names next to it. The business owner can explain its value, and the technical owner can explain its consumption.”



Warning signs

- The platform team is listed as the owner of everything
- Owners exist on paper but have never seen the consumption their workload creates



Bring to the review: an owner register

Workload	Business owner	Technical owner	Sees its usage?

QUESTION 5 OF 5 · VALUE

Can we explain how Fabric consumption maps to business value?

Why it matters. Your CDO or CFO will ask, especially as Copilot and Data Agent workloads arrive. The first four questions make this one answerable.

IF YOU CAN'T ANSWER IT YET

Get Questions 1–4 in place first. The goal is matching spend to value, not driving spend as low as it will go.



A confident answer sounds like

- “For each top consumer, we can say what it costs in capacity and what the business gets.”
- “Our AI workloads have had showback since launch, so we can judge whether to scale them.”



Warning signs

- Fabric cost is an IT line item no one else can explain
- A new AI workload is scaling before anyone checks its consumption
- The only lever discussed is making the bill smaller



Bring to the review

- For each top consumer: the business owner’s one-line value statement
- For AI workloads: consumption, trend since launch, and the expected outcome

CHOOSING THE RIGHT FIX

Work through the fixes in order

The first two usually cost nothing but time. The last one shows up on your bill every month afterward.

1

Optimize

the workload

One workload consumes far more than it should, even when it runs alone.

Likely cause: inefficient workload design

2

Reschedule

the background work

Utilization is fine most of the day but spikes when background jobs overlap busy hours.

Likely cause: scheduling collision

3

Isolate

the workloads that need protecting

Critical interactive workloads keep absorbing the impact of other teams' jobs.

Likely cause: lack of isolation

4

Resize

the capacity

Utilization stays high around the clock after the first three are ruled out.

Likely cause: genuinely insufficient capacity

NEWER FABRIC LEVERS

Precision tools for specific problems

Each is useful in the right situation. None of them is a strategy on its own.



Autoscale Billing for Spark

Moves Spark jobs to separate pay-as-you-go compute, so they stop competing with BI.

KEEP IN MIND

Spark is then billed differently, and you still need base capacity for everything else.



Surge protection

Caps background consumption so a runaway job can't drag down interactive work.

KEEP IN MIND

It limits background use, but doesn't guarantee interactive users never see a delay.



Capacity overage

Pay a metered rate instead of being throttled, within a spending limit you set.

KEEP IN MIND

It's a premium rate, meant for a genuine crunch rather than a standing fix.

SUGGESTED AGENDA

A 60-minute capacity review

Time	Topic	Led by
0:00–0:05	Purpose, and actions carried over from the last review	Review chair
0:05–0:20	Q1 and Q2: top consumers, peak windows, incidents	Capacity administrator
0:20–0:30	Q3: are boundaries still right? New AI workloads?	Data platform lead
0:30–0:45	Q4 and Q5: owner gaps and value statements	Business owners, budget lead
0:45–0:55	Actions: optimize, reschedule, isolate, resize, in order	Whole group
0:55–1:00	Record the scorecard, assign owners, book the next review	Review chair

How often: Recommended cadence, e.g. a monthly consumption review with workload owners and a quarterly look at sizing and boundaries. Whatever you choose, keep a regular rhythm, not only after a throttling incident or a surprise bill.

REVIEW SCORECARD

Record it, then compare next time

Question	Yes, with data	Partly	No / best guess	Follow-up action	Owner	Due
1. Who consumes our capacity?						
2. Expected demand or problem spike?						
3. What shares capacity?						
4. Who owns each workload?						
5. How does consumption map to value?						

First question below “Yes, with data”

Our starting point for the next review

Fabric Capacity Health Check

Not sure your team could answer all five? That's exactly what the Lantern Fabric Capacity Health Check is for.

The first move probably isn't buying more capacity. It's a clear baseline. We work with your team to find what's driving your consumption, where performance or cost risk is emerging, and which actions to prioritize. It's built on the same framework as this checklist: sizing, organization, controls, and accountability.



What's actually driving your consumption



Where the risk is building



What to fix first



Book your Health Check: lanternstudios.com/fabric-health-check



KEEP GOING

Go deeper on every question



Watch the webinar on demand

Control Microsoft Fabric Costs: A FinOps Playbook for Capacity and Throttling



Read the Fabric FinOps series

Sizing decisions, smoothing and throttling, monitoring, the four-step playbook, and real-world scenarios.

