Tying Together Loose Threads:

Capacity Monitoring in Microsoft Fabric

Stephen Babcock

Introduction

Change is the only constant in business intelligence, yet the last year has brought a significant shift that is reshaping the very foundation of the ecosystem's most popular platform. The introduction of Microsoft Fabric is transforming how organizations manage Power BI, with implications for how enterprises evaluate, budget, and ultimately optimize their data programs to balance reliability and results. In particular, the migration of Microsoft's data and analytics environment to Fabric has brought about a number of changes in the area of capacity monitoring. One is a branding shift that is visible at surface level to BI admins, with the Power BI Premium Capacity Metrics App now appearing as the Fabric Capacity Metrics App. But the changes go deeper when examining areas including pricing structure, metrics, and overall management concepts. Understanding the difference between the Fabric Capacity Metrics App and the legacy Power BI Premium Capacity Metrics App is crucial for admins seeking to rein in BI sprawl, and be proactive about instituting change. Equally, the Fabric Capacity Metrics App falls short in a number of capability areas that are critical for BI admins as they seek to understand spend and make budget decisions, including the ability to compare metrics over time and pinpoint specifics about individualized jobs that are causing issues. In this article, we'll drill down on what BI admins need to know about how the transition to Microsoft Fabric affects capacity

What's Microsoft Fabric Means for Power BI

Let's be clear from the start: The introduction of Microsoft Fabric doesn't mean that Power BI is going away.

After all, Power BI is beloved among data users, creators, and leaders. Its combination of cost-effectiveness and user-friendliness, when coupled with the ubiquity of Microsoft products in enterprise settings, has made it the cornerstone of many data and analytics programs, while helping to democratize data in ways we never thought possible.

Instead of replacing Power BI, Fabric is now the umbrella under which Power BI sits. Thanks in no small part to the success of Power BI, the data ecosystem has grown exponentially since Power BI was launched in 2015. In the second half of 2023, Microsoft launched Fabric to unify all of its data and analytics tools in one platform, and make room for all of the AI-driven innovation coming down the pipe.

In this new paradigm, Power BI is the business intelligence application of Fabric. BI is a core function of Fabric, but sits alongside other functions and productions, such as data lakes (OneLake), data integration (Azure Data Factory), and data engineering (Azure Synapse Analytics). Power BI also continues to operate as a standalone product, though that relationship is shifting (more on that later).

In a sense, the relationship between Power BI and Fabric is similar to that of Microsoft Excel and Microsoft Office. Excel stands on its own and is widely associated with spreadsheets, yet it is also part of Microsoft Office (now 365), where it can be made more powerful via interoperability with the other apps, all of which are under a single umbrella for access and governance.

Capacity Monitoring with Microsoft Fabric

With Fabric, Microsoft is building on the success of Power BI in many important ways, and capacity is among the most prominent. Under Power BI, Microsoft made two pricing models available. One was a licensing model for individual workspaces. The second, geared toward enterprises, provided a pricing model called the Power BI Premium Capacity. This granted a certain allocation of compute to an organization for all reports, datasets, and dashboards. With the shift toward the cloud, this model based pricing off of the resources consumed, as opposed to a fixed price based on the number of workstations.

With Fabric, Microsoft is applying the Premium Capacity pricing model to Power BI, and extending it to the other services under the Fabric umbrella. Now, enterprises purchase one allocation pool of compute for all of the Fabric products they use, and this is able to be scaled up and down as needed.

To be sure, there is plenty that is consistent with this change. Power BI users maintained access to the same analysis and analytics tools they always enjoyed, and Microsoft is building off a model that should be familiar to Power BI admins.

But for Power BI admins, the rollout of a connected experience has a couple of important implications. First, it centralized tenant and capacity level administration under the Fabric brand. This means there are a host of new names for Power BI Premium users to learn. The Power BI Portal was renamed to the Fabric Admin portal. The Power BI Administrator role was renamed to Fabric Administrator.

When it comes to capacity for Power BI Premium users, Power BI was purchased in Power BI per capacity SKUs, known as P-SKUs, and measured in v-cores. With Fabric, it is purchased in F-SKUs and measured in CUs, or capacity units per hour. Microsoft removed P-SKUs as a purchase option on July 1, 2024, and set a sunset date of January 1, 2025.

There are also important changes to how capacity is monitored. With the shift to CUs, there is now one capacity measure for all of the different applications of Fabric. So Power BI compute is measured in the same unit as Synapse. If there is a performance issue in Azure Data Factory, it could affect Power BI. Previously, admins could focus only on Power BI Premium Capacities.

Capacity across all Microsoft Fabric products is now also viewed in the same place. Microsoft's recently-released Fabric Capacity Metrics App is designed to take the capabilities of the Power BI Premium Capacity Metrics App, and extend them to all of Microsoft Fabric's tools. Through this app, a Power BI report, a Dataflow or Synapse Notebook are all shown in one view.

This means that capacity monitoring for all Microsoft Fabric components are measured by a single metric, and that metric the helps admins understand compute is monitored in a single view.

Microsoft made these changes with the goal of bringing together apps under all data and analytics functions under the Fabric umbrella and further coalescing the business model to be more in line with a typical SaaS delivery.

However, adding new things often means that others will be taken away, and that is the case with Fabric's new monitoring schema, particularly the Fabric Capacity Metrics App.

Fabric Capacity Metrics, Defined

At the end of the day, the Fabric Capacity Metrics App is a Power BI dashboard. And, like any Power BI dashboard, it is only as useful as the decision making it empowers. As users dig in, they are confronting new challenges when it comes to fully understanding Power BI performance.

The reason gets at how the Fabric Capacity Metrics App is built, particularly how it structures data, and how it operationalizes the pay-as-you-go model.

On-demand flexibility is key to a SaaS model, and Fabric is no different. Like many cloud services, subscribers effectively rent storage and compute from Microsoft, who in turn provide the ability to scale up or scale down, as needed. However, BI differentiates itself from the needs of a typical cloud customer due to the intensive compute resources associated with sheer volume of reports, dashboards, and datasets being refreshed. This high volume of jobs creates significant strain on the environment, so BI admins strategize jobs and corresponding allocation of compute based on available resources. They may want to space out refreshes during weekends to avoid the risk of overloading the system during normal business hours. In order to monitor all of this, BI admins can benefit from the ability to see trends over longer periods of time, pinpoint times that capacity was maxed, and correlate it with the specific resources being used at that time. That way, they can plan compute allocation over time in order to avoid potential overages.

With Fabric, Microsoft has introduced its own system to measure capacity, and even implemented its own fixes leveraging unallocated compute and the pay-as-you-go model.

Capacity Metrics are as follows:

Interactive vs. Background: Not all operations are created equal, nor do they use capacity equally. The Capacity Metrics App breaks down operations into two categories.

Interactive activities are the front-end requests triggered by users, usually when they interact directly with the UI. These are on-demand, often shorter requests, such as opening a report or DAX Queries.

Background activities are longer and often not directly triggered by users, such as data refreshes. Typically, background activities can be scheduled at non-peak times in order to prevent capacity from spiking.

Throttling. Interactive and Background classifications are useful for Microsoft as it looks to categorize compute, but they are also especially important for BI admins to understand. That's because this distinction governs the progression of performance degradation that is triggered when the system is over-capacity. This is known as throttling.

When capacities are overloaded with jobs, and the system experiences compute or performance strains, the first thing typically impacted is the interactive experience. This is when users start seeing the loading circle while waiting for dashboard graphs and visuals to populate.

The second level of saturation affects background activities, like when refreshes fail. When this happens, it indicates that the environment's saturation is even worse than anticipated.

Smoothing and Bursting: Capacities go through peaks and valleys. There are times when more is used, and times when less is used. When there is a spike, it can potentially cause capacity to be expended beyond the amount that is purchased. This could trigger throttling, which degrades performance in a variety of ways, from delays to outright rejection. However, instead of throttling first, Fabric accounts for this with features called bursting and smoothing. With bursting, CU usage for specific tasks is automatically increased in order to speed up operations, ensuring they can run at the highest level of performance. This could eat up capacity, so, for the purposes of billing, smoothing spreads out the usage over the period of the spike, by taking the average of capacity used over the duration of an operation. Smoothing is applied over 5 minutes for an interactive operation, and over 24 hours for background operation. So, effectively, users are not penalized for an overage, even in the event that an operation would consume all of the available compute at a given time. In turn, this creates a mechanism to prevent throttling, even when capacity is over the SKU limit.

Carryforward: Smoothing is a fix, but at times it does not completely remedy the overage. In the event that smoothing does not distribute enough compute to prevent the spike over multiple jobs, Fabric applies a further fix called Carryforward. In the event of an overage, this automatically allocates usage that is above the limit to a future time period when a company has available usage.

Limitations in the Visualization

The Fabric Capacity Metrics App represents the above techniques within its reports, allowing users to view compute usage and where overages occurred. However, at times these views do not allow users to zoom out and take a holistic look at the data, while at other times it does not provide BI admins with comparison capabilities between times and capacities. Both are important. Ultimately, the data provided in the Fabric Capacity Metrics App is not an end unto itself. BI admins use data to zero in on the root cause of any overages, and take what they learn to plan jobs and make budget decisions.

The Fabric Capacity Metrics App has a number of drilldown capabilities that allow data leaders to examine compute, including views that show which operations are having the largest impact on compute. However, these often fall short of providing the comprehensive picture that's needed for true cost monitoring.

Overall, the App's limitations can be traced to how it structures data. Let's break down why that presents challenges:

The Fabric Capacity Metrics App **only shows data corresponding with one Fabric Capacity Unit at a time**, making it difficult to understand how the accumulated impact of multiple capacities running at once impacts compute. Many large enterprises have dozens of premium capacities, and understanding aggregate compute data is necessary to observe the causes of throttling. They want to see how the compute used in one capacity is performing relative to another, and

how they are influencing each other. Remember, it is the accumulation of many BI activities that lead to compute overages, not a single job. Admins would benefit from capabilities to understand the environment as a whole, rather than what was happening in specific capacities at specific times. The Capacity Metrics App displays **timepoint data**, challenging the ability to pinpoint problems and recognize trends. In drilldown views, the App shows CU usage that was applied over a 30-second interval, or timepoint. While this offers granularity, it is often insufficient to reflect how the accumulated volume of BI operations impacted compute and potentially caused throttling over time. In data provided by Microsoft, BI admins see lots of different data. They see data on interactive jobs, on background jobs, on compute data from other operations. Given this large amount of data, a view that showed a longer time period would help BI admins spend time on delivering quality insights that revealed trends, rather than digging through timestamps.

Similarly, **the App's standard 14-day view of data** falls short of providing the historical data necessary to gauge trends over a long period of time. Typically, BI admins seek to cover a month or longer as they seek patterns that help to budget resources. They also seek to make month-over-month comparisons in order to observe trends that can inform long-term planning and budgets. Microsoft proved that improvement in the area of data retention was possible with Fabric Warehouse, where it extended the retention period from 7 to 30 days in July. The cited reason was customer feedback. No such change has come to the Capacity Metrics App, yet.

The struggles associated with these limitations are **compounded by smoothing**. In the Fabric Capacity Metrics App, users effectively see a visualization of how smoothing happens. Remember, smoothing spreads capacity usage over time, rather than displaying the actual usage in the moment it occurred. Due to smoothing, the App displays a visual where jobs are shown running, even at times when they weren't actually running. The visual is displaying the accurate time and usage for how Microsoft Fabric is reading Many admins want to explore the data behind a spike to discover where it occurred. With event-based data, they would be able to see what operations were running during the period when a spike occurred, hone in on what was causing the issue, and make a fix accordingly. But with timepoint data, they won't see the exact event when a spike occurs. Rather, they will only see the smoothed CU usage. A number of operations could be shown as running during that time. But because of how the Capacity Metrics App presents data, those operations may not have actually been running at that time, and the source of the spike remains obscured.

Catching problems early keeps them from becoming worse. But with the Fabric Capacity Metrics App's visualization, admins may not even see spikes until after the fact, and they have consumed all of the unused capacity that is available for smoothing. By that point, they may be faced with throttling, while still being left in the dark about the source of the spike.

Capacity Monitoring, Re-Engineered

In the absence of hard data and expansive views that reveal trends, admins are extracting the data from Fabric and turning to their own calculations and fixes, or even creating their own reports. But these workarounds in turn present efficiency challenges. They are time-consuming, and displace precious resources that could be spent on analyzing data that helps other parts of their business, not just the management of data resources.

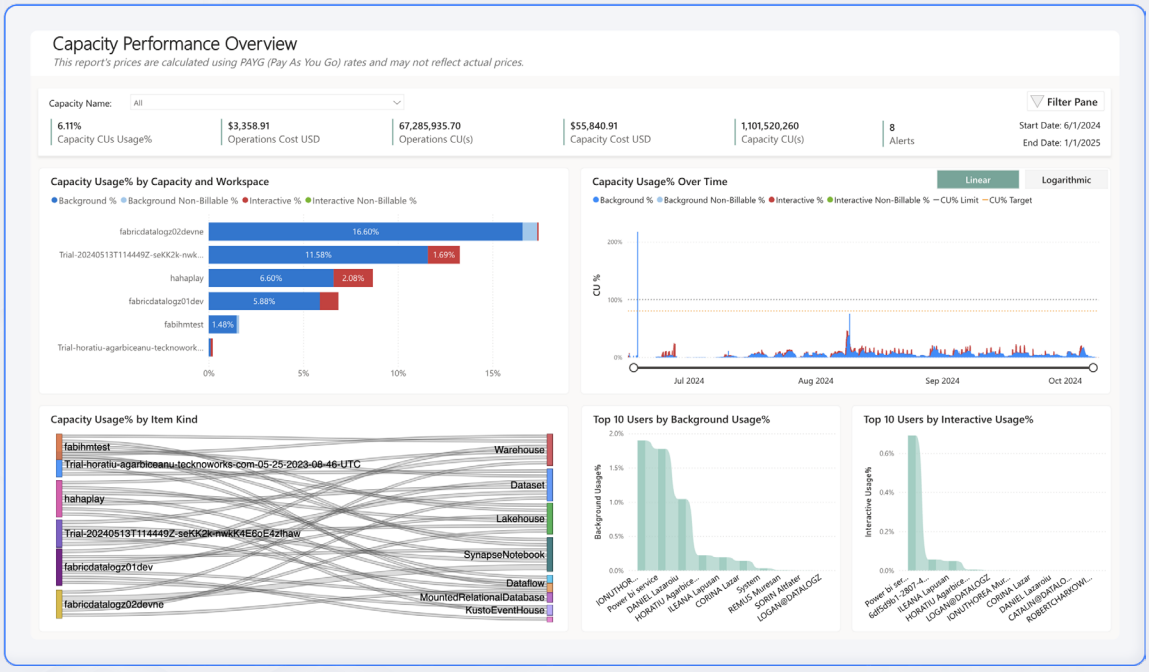
At Datalogz, **we've heard from BI admins** about the issues, and the hacks. That's why we've taken action to help create the comprehensive Capacity Monitoring capability that data leaders can depend on. We've re-engineered the functionality of the Fabric Capacity Metrics App to surface the data you need, and provide useful insights to reduce costs.

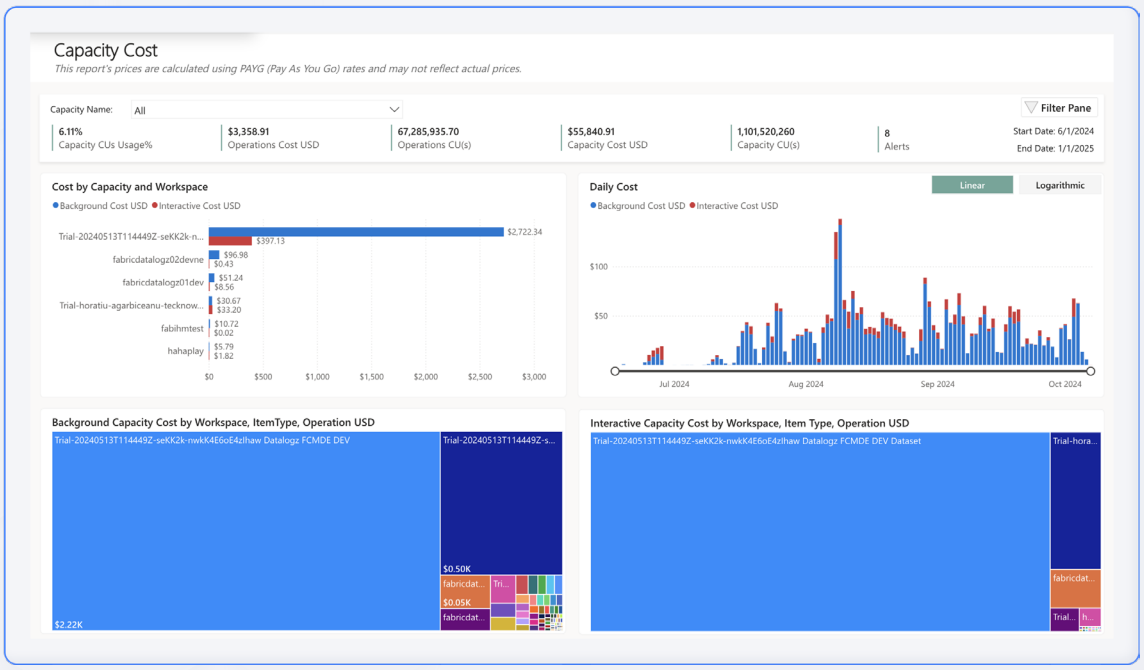
Here's a look at key features:

Comprehensive Capacity Insights help you gain a full understanding of your Power BI capacity usage with historical data retention and advanced analytics that allow for deep dives into specific issues.

This includes a dashboard that contains a **macro-level view of an entire environment**, with a view that shows all of the premium capacities being used, including alerts, CUs, costs, and more. This allows for trend analysis, allowing admins to compare how capacities are performing side-by-side, and visualize how one capacity impacts another.

This wider view also reveals **noisy neighbors**, which are reports that are devoting compute resources to datasets that are not actually needed for the report. These noisy neighbors not only use large amounts of compute, but they also cause other refreshes to fail as a result.





From this wide view, users can then **drill deep into the specific times, capacities and items** that are **causing spikes and overages**. Toggle to specific capacities, or view capacity usage by item types to see what is driving overage events. Break down usage between background and interactive costs.

Click into an Event View that shows what is **causing saturation at a specific timestamp.** This shows macro-level states, as well as the background and interactive operations, usage stats, and items being used at that time.

Rather than searching for a timepoint in a specific capacity to try to construct the bigger picture, admins can see the bigger picture, then drill down to get specifics from there.

Elsewhere, users can access Advanced Cost Analysis tools to help track and manage your data and analytics expenses, ensuring that you stay within budget while optimizing performance. This breaks down daily costs, as well as costs by capacity and workspace. Custom-built metrics provide the ability to tailor your monitoring experience with metrics designed specifically for your organizational needs, giving you the insights that matter most.

Get notified immediately with a **Smart Alerting System** that provides proactive notifications when capacity issues arise, not just a look at what has happened in the past, along with actionable recommendations to address them before they impact performance.

View **Throttling Stages** that provide details about affected users and operations, as well as carryforward and capacity usage over time.

To make all of these capabilities even more powerful, our Capacity Monitoring solution will soon be **integrated within the Datalogz Control Tower**, providing a unified platform for managing your entire Power BI environment and other BI tools, such as Tableau and Qlik, that may also be in use.
We're currently in R&D with early customers and are excited to roll out these features to all our users soon. Interested in learning more? Check out our video demo here and let's discuss how Datalogz can help your organization optimize Power BI capacity management.

End BI sprawl with Datalogz, book a demo at datalogz.io/book

Resources:

Understand the metrics app compute page, Microsoft; Important Update coming to Power BI licensing, Microsoft; Capacity metrics in; Microsoft Fabric, Microsoft; Microsoft Fabric pricing, Microsoft; Fabric: The Secrets of Capacity Consumption, redgate Simple Talk; A closer look at Microsoft Fabric pricing, billing, and autoscaling dataroots; Microsoft Fabric: Understanding Capacity, Covenant Technology Partners; How to extract data from the Fabric Metrics App – Part 1, PBI Guy; How to Use the Power BI Premium Capacity Metrics App, PhData; What to Know About Microsoft Fabric if you Already Have Power BI, Cloud Servus.