



Microsoft Power BI Premium

Whitepaper

By Amir Netz, Technical Fellow, Microsoft Corp.

©2017 Microsoft Corp. All rights reserved. This document is provided "as-is." Information and views expressed in this document, including URL and other internet website references, may change without notice. You bear the risk of using it.

Some examples are for illustration only and are fictitious. No real association is intended or inferred.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

Introduction

Since Power BI was introduced in July 2015, it has enjoyed a fantastic reception in the market. The service's "seconds to signup, minutes to wow" approach has helped Power BI deploy in hundreds of thousands of organizations.

As Power BI's adoption evolves from individuals and teams and into large enterprises, the need for new capabilities continues to evolve with it. Enterprises that are adopting Power BI are demanding larger scale, increased performance, complete governance and control, and the ability to deploy reports and dashboards to a very large number of employees, in a cost-effective way.

Power BI Premium is designed from the ground up to address the challenges of large enterprise deployments and demanding workloads. A new service architecture was created to optimize for these enterprise scenarios, including the ability for organizations to serve a large number of "readers" — who simply need to view and explore dashboards and reports — in a cost-effective manner and without needing to license each user individually. This white paper will review the basics of Power BI Premium, its vision, its capabilities, its road map, and its various usage scenarios.

This white paper describes Power BI Premium, both as it exists when launched and also how it will evolve. We believe that by articulating the full vision you can better understand how to leverage Premium most effectively in your organization. Although we at Microsoft are working hard to release those features as soon as possible, know that some unexpected difficulties may arise that delay or change our plans.

Dedicated capacity with Power BI Premium

The Power BI Premium offering introduces the concept of dedicated capacity, and it is the heart of the offering.

What is dedicated capacity?

Power BI Premium enables dedicated capacity, a set of resources (read: hardware) reserved for the exclusive use by you—the customer. To understand why having this dedicated capacity is so important, we first have to discuss how Power BI works internally.

Microsoft operates large clusters of machines that handle sign-ins, data load and refresh, query processing, rendering of dashboards and reports, natural language understanding, machine learning, sharing content and securing it, connecting to gateways, and powering all the other features of Power BI. In each of these clusters hundreds of machines are working together in concert silently to provide you with the familiar Power BI experience you know and love.

Since all the users of Power BI are sharing the same set of machines (also referred to as the “shared capacity”), controls are in place to make sure that a few users cannot monopolize all the shared capacity in the cluster and disrupt the quality of service for others. Many of the published restrictions of Power BI, like the 1GB model size or the limitation on the number of refreshes per day, are driven by the need to secure excellent quality of service for everyone. Enter Power BI Premium—with dedicated capacity we are offering the opportunity to move away from shared capacity and instead secure your own set of hardware resources to be used exclusively by you and your users.

Since the dedicated resources for Power BI Premium will be used only by you, all of the throttles and limitations that we at Microsoft put in to protect you and the other tenants can now be removed. The removal of the restrictions opens Power BI to a whole range of new features that allows for much larger scale, better performance, and fresher data.

And just as exciting as the new features, having your own dedicated capacity also allows you to publish BI content to users throughout the organization and beyond, without having to purchase licenses for them.

With Power BI Premium you can distribute all your BI apps to users throughout your organization without requiring per-user licensing.

With Power BI Premium we took great care to make sure that although some resources in our cluster become dedicated to your exclusive use, they are still completely integrated into the overall Power BI experience. The end users in your organization will continue to log in to Power BI as usual. They will find all their dashboards and reports as usual. They don't have to ever be aware that some (or all) of the BI content is hosted in your dedicated capacity. For your end users, everything works exactly as before.

Power BI Premium is transparent to end users and does not change their experience.

Lastly, we know that some organizations are not fully ready to transition all their data and BI architecture to the cloud. We understand the complexities of the real world and the need of some organizations, especially in regulated industries, to keep sensitive data behind the firewall. To accommodate these situations, we also included an on-premises server in Power BI Premium—the Power BI Report Server. Power BI Report Server allows you to keep sensitive data and reports fully behind your firewall, creating a hybrid deployment mixing cloud and on-premises.

Dedicated capacity nodes: the v-Cores

Of course, different BI workloads will require different compute power. Supporting hundreds of users of few reports and dashboards is very different than enabling 100,000 users to consume many BI applications.

So when you decide to purchase Power BI Premium, you will need to provision enough capacity to support your workload. Capacity is provisioned through purchasing capacity nodes. The capacity nodes you purchase have a certain number of v-Cores (which stand for “virtual cores”), memory, and bandwidth that will need to be sufficient to power your BI system. Each node is composed of a number of v-Cores: frontend cores and backend cores.

The frontend cores are responsible for the web service, dashboard and report document management, access rights management, scheduling, APIs, uploads and downloads, and generally for everything that relates to the user experience.

The backend cores are responsible for the heavy lifting: query processing, cache management, running R servers, data refresh, natural language processing, real-time feeds, and server-side rendering of reports and images. With the backend cores, a certain amount of memory is reserved as well. Having sufficient memory becomes especially important when dealing with large data models or with a large number of active datasets.

The backend cores that do the heavy lifting are fully dedicated to you, and they will not be shared by others. The frontend cores are shared with others, and they are computed as your portion in the pool of machines that handles the shared frontend services in Power BI.

At launch, Microsoft will offer three sizes for Premium capacity—P1, P2, and P3. Each comes with a different number of v-Cores and memory size:

CAPACITY NODE	CORES	BACKEND CORES	FRONTEND CORES
P1	8 v-Cores	4 cores, 25 GB RAM	4 cores
P2	16 v-Cores	8 cores, 50 GB RAM	8 cores
P3	32 v-Cores	32 cores, 100 GB RAM	16 cores

How much capacity?

Calculating how much capacity you’ll need is an inexact science. You should plan on provisioning enough capacity to be able to comfortably handle your peak usage. Make sure to plan for some offline scheduled activities such as data refresh, smart alerting, or broad email distributions. If you have not reserved sufficient capacity, your cores could become overloaded and the responsiveness of the system will suffer.

Microsoft has provided a Capacity Calculator on its website to help you plan your Power BI Premium deployment. The calculator uses a simple usage model based on Microsoft experience across many different organizations. However, the variability of the usage patterns is very significant and will depend on many factors including the specific data models used, the number of queries and their complexity, the hourly distribution of the usage of the service, the data refresh rates, the use of advanced features like Q&A, R and email distributions, and other usage patterns. As such you must use the calculator with caution and with tempered expectations—estimating capacity is not an exact science and performance cannot be guaranteed.

Most organizations will experience gradual adoption of the BI system over time. As adoption grows, the capacity utilization will increase. Another factor that will increase the capacity utilization is the planned future functionality of Power BI services on the capacity. Microsoft will gradually be moving more Power BI features and services after the launch to the capacity, and these will consume more of the horsepower. Therefore, Microsoft is providing multiple node size options, e.g., P1, P2, and P3, for customers as their usage grows over time.

All this is to explain that estimating the capacity is not a trivial exercise. But as a cloud service, Power BI provides the flexibility to increase capacity as needed.

Performance and responsiveness are critical to user satisfaction. We strongly recommend that you plan for provisioning enough capacity to secure high performance—and with it the success of the deployment and the happiness of your customers.

Make sure to provision enough capacity to secure peak performance of your BI system, and with it its success.

Scaling up

At launch, Microsoft is offering the three initial capacity node sizes shown above. However, additional node sizes available in the future could support data model sizes in the hundreds of gigabytes and tens of processor virtual cores.

Scaling out

When deploying to a very large number of users, even the largest node size may not be enough. To address large deployments, Microsoft is preparing a scale-out capacity deployment option. With the scale-out option, you will be able to purchase multiple capacity nodes and deploy them together as a single capacity cluster. Power BI will autonomously manage your capacity cluster and will automatically balance the load across the nodes using the same technology it is using in the shared capacity. No extra management overhead will be involved when deploying scale-out capacity.

Power BI application distribution

For many, the most important capability of Power BI Premium is the ability to build and distribute BI solutions broadly to users, without the need for additional user licensing. This option is similar to the “per-core” licensing for on-premises servers.

Although you have Power BI Pro licenses to collaborate, create, and publish BI content, using Power BI Premium is perfect when you need to serve a large number of “readers” (aka “viewers”), who simply need to view and explore dashboards and reports. For this class of users, a Power BI Pro license is not necessary, and there is no need to name or count these users when you use the dedicated capacity.

To use this option, your BI content will be packaged in a Power BI App and distributed to users. Building Apps is very simple—create an App Workspace and collect all the dashboards, reports, and datasets for distribution. When completed, simply publish the app to your selected audience with a couple of clicks. Note that to create apps and publish them, you'll need a Power BI Pro license.

Once apps are created and distributed, any user, regardless of whether they have a Pro user license or not, can view the content contained in the app, which uses the resources of the Power BI Premium capacity. These capabilities include navigating the dashboards, interacting with the reports, and exploring using Q&A and Quick Insights. There is no restriction on the type of content the applications can include. Any supported data source can be used, both on-premises (through the gateway) and in the cloud. Power BI Apps are also available on Power BI Mobile and through Cortana.

But for all this to work, you will need to make sure that the Power BI apps are hosted in Power BI Premium capacity. To do that you will need to configure the capacity to include the app workspaces for those apps you'd like to distribute. Once the app workspaces are migrated to the dedicated capacity, the apps can be used freely by end users.

Migrating app workspaces to Power BI Premium is very simple and takes only a couple of minutes. When the migration is completed, your dedicated capacity will provide the compute power to serve the application(s) to the users.

You can also create a Capacity Assignment List and specify the Power BI Pro users, either individually or via a security group, whose applications will automatically be hosted in your dedicated capacity. From that point on, any application workspace created by one of the users in the Assignment List will be automatically migrated to the dedicated capacity without any manual migration. In addition, new users joining your organization who are in the Assignment List will automatically have their personal workspaces created in your capacity. Assigning all the authors in the organization is common when creating a baseload capacity (more on that later).

Power BI for developers

Power BI apps are ideal for users who use the Power BI service directly. However, sometimes you will need to build other SaaS applications where Power BI content is embedded in them. With these SaaS applications your users don't log in to Power BI at all. Instead your application will log in to Power BI to serve them their reports or dashboards. Going forward, Power BI for developers will be purchased via the Power BI Premium node offers. Developers can choose to use their capacity in "embedded mode" instead of user authenticated "SaaS mode."

Compared with the previous version of Power BI's offering for developers, called Power BI Embedded, developers pay for Power BI Premium on a capacity basis yielding predictable pricing compared with the variable session-based model. In addition, developers can now use all the features of Power BI, including dashboard embedding, cached models, and on-premises data gateways.

The principles, pricing model, and APIs for building such applications are consistent for both internal-facing enterprise portals or applications and external SaaS applications built by ISVs for their customers.

The Power BI Premium capacity features and road map

Beyond the ability to distribute BI applications broadly throughout the organization, Power BI Premium offers a range of other very desirable capabilities. The initial release of Power BI Premium will not yet have some of these capabilities, but we intend to iterate following the offering's general availability. Additional planned Premium Capacity capabilities include the following:

Noisy neighbors isolation

The query performance of your reports and dashboards is now protected from the activities of the other users and tenants in the Power BI service. The hardware is exclusively for your use, so only your own workload will affect the performance. (This capability is available at launch.)

Huge dataset storage and no user quotas

With Power BI Premium you are freed from the standard user quotas of Power BI. Previously, we allowed up to 10 GB per Power BI Pro license in your tenant. With Power BI Premium, you get 100 TB of storage for each capacity you provision. (This capability is available at launch.)

Higher dataset refresh rates

We increased the data refresh frequency limits in Power BI Premium, from eight per day to 48 per day. So the standard limits on refresh (hourly) is gone, and now you can have your datasets refreshed on a minute-based schedule! This is a great way to ensure that your users are always seeing the latest data. (This capability is available at launch.)

Large datasets

The standard limitation on the dataset size cached in Power BI is 1 GB. But with Power BI Premium, we will be removing that limitation. You will be able to build models as large as the Power BI Premium dedicated capacity memory can hold. This means that the Power BI Premium P3 SKU will be able to support a dataset with 50 GB of data or more, and upcoming SKUs will offer capacity sizes that allow datasets with hundreds of gigabytes of data!

The increase of the dataset size limits will be done gradually after launch and will be going through thorough testing to ensure that the performance of the system remains high. In addition, to effectively support large datasets, a set of additional features, listed below, will need to be introduced. (Not available at launch.)

Incremental refresh

When the datasets get large, constantly refreshing all the data becomes a lengthy and highly inefficient way of managing the system. An upcoming feature for Power BI Premium is the ability to incrementally refresh the data so that only the newest data from the last day (or hour) is loaded into Power BI.

With Incremental Update all you have to do is set the refresh policy for the various tables in the model and Power BI will take care of everything else. (Not available at launch.)

Pinning Datasets to Memory

The Power BI system dynamically manages the memory use of the backend cores. When datasets stand idle without any queries hitting them, Power BI will silently evict them from memory to make room for more popular datasets. This dynamic memory management mechanism allows Power BI to host millions of datasets very effectively. However, there are situations when you may want to override this behavior.

When datasets are large, bringing them into the backend node's memory may take a while (how long depends on the dataset's size). This could mean that users opening a report find themselves waiting for the datasets to be loaded to memory before the first query can be executed. If so, consider pinning the dataset to memory, ensuring that Power BI will not evict it. This could be important for datasets where you want to secure the absolute highest quality of service—perhaps the ones powering the CEO's dashboard.

The Pinning Datasets to Memory capability allows you to gain new control over the performance of the BI system. (Not available at launch.)

Dedicated Data Refresh Nodes

Data refresh is a significant CPU-consuming activity that can put a heavy load on the node. Many administrators will schedule the data refresh processing during off-peak hours to avoid affecting the visible performance of the system.

If this is not an option, or you want to secure the optimal performance at all times, you will be able to set some of the nodes (when deploying scale-out capacity) as Data Refresh Nodes. This will separate the interactive query activities from the data refresh activities, ensuring that the interactive performance is maintained at all times. (Not available at launch.)

Read-Only Replicas

When you need to deploy BI applications to a very large user population and you are worried that the number of queries run on the node will overload it, you should consider deploying Read-Only Replicas. With this capability you will be able to mark datasets requiring replicas, and with that copies of the datasets will be automatically created and deployed into the other nodes in your Power BI Premium dedicated capacity. From this point on, Power BI will balance the query load across the nodes.

With Read-Only Replicas you'll be able to handle the largest query load. (Not available at launch.)

Geographic distribution

Capacities can be provisioned in any of the datacenters worldwide where Power BI is available. This supports BI projects where content must reside in certain geographic regions for regulatory or compliance reasons.

In addition, capacities can be set up as geo-replicas to handle the scenarios of global deployments of BI applications. If the dataset is stored in the U.S. and the end user is residing in Australia or China or South Africa, the distance between the user and the data can create performance issues due to bandwidth limitations and the speed of light (seriously).

Geo-Replicas are very similar to Read-Only Replicas, but with this capability the replica is copied to a node in another region. By setting Geo-Replicas you can ensure that users from around the world will connect to a copy of the data close to them and will enjoy the optimal performance and interactivity without concern for the physical distance between the user and the data. (Not available at launch.)

Road map cautions

The discussion above covers some features that are have not yet been released. Although we at Microsoft are working hard to release those features as soon as possible, one must also take into account that some unexpected difficulties can come up that may delay or change the road map outlined.

“Predictions are hard, especially of the future.”
– Yogi Berra

Additional Power BI Premium features not discussed above are being worked on, so keep your ears to the ground because more goodies are coming.

In the cloud and on-premises: Power BI Report Server

The IT industry is moving to cloud architecture. Although the trend is clear, we do understand that different organizations are moving at different paces, and some have not fully reached the comfort zone or the regulatory approval required to start moving sensitive data to the cloud.

Which BI product architecture should you select—cloud architecture or on-premises? Power BI Premium supports both, so you don't have to choose.

When you purchase Power BI Premium you not only get dedicated capacity nodes in the cloud, but also the new Power BI Report Server to be used on-premises.

Power BI Report Server is an incredibly rich and mature product that traces its roots to the popular SQL Server Reporting Services. On top of the rich paginated reporting functionality, Microsoft added the ability to serve Power BI Reports authored using Power BI Desktop, and deliver them in the Power BI mobile application—or through modern browsers or embedded in applications.

The Power BI Report Server is part of the Power BI Premium offering to allow you to build distributed hybrid BI systems that include cloud and on-premises mixed deployment. With the ability to pin visuals from reports on Power BI Report Server into the Power BI cloud service, you can now create dashboards that span both cloud and on-premises reports.

With Power BI Report Server, you have full flexibility to decide what portions of the data and reports you want to keep on-premises and what portions should reside in the cloud. The same number of v-Cores you have provisioned in the cloud can also be deployed on-premises through the Power BI Report Server. And you don't have to split the capacity you purchased: When you deploy eight v-Cores in the cloud, you also get to use eight cores on-premises for a total of 16 cores!

We at Microsoft went to great lengths to ensure compatibility between Power BI Report Server and the Power BI cloud service. This means that whenever you are ready to transition more of the content to the cloud, all you have to do is simply upload the reports from Power BI Report Server to the Power BI cloud service and you are ready to rock.

Power BI Premium — common scenarios

Power BI Premium is ideal for several distinct analytics use cases. These use cases are not mutually exclusive and can be used in concert in one organization.

Baseload capacity

Perhaps the most common use case for Power BI Premium is the centralized provisioning of a baseload capacity by IT or a central BI function for the general use by an entire organization ("baseload" comes from the world of power utilities and means the constant power capacity available at any given time). In this scenario we expect many authors to create diverse BI content and applications to be shared with a large number of people.

The baseload capacity is intended to support a large number of projects, some now and some in the future, and there is no way (or desire) to specify in advance the specific content or applications that will be created or the intended user base for those. Premium Capacity administrators typically do not get involved in specific projects; instead, their job is to set general policies and most important to make sure that there is enough baseload capacity provisioned for the org to support all the ongoing activities.

Baseload capacity can also be used to support enterprise self-service use cases, where self-service users need higher data size limits or geographic dispersion. For this use case, Premium will complement Power BI Pro licenses for self-service users who are creating or collaborating on content.

To enable the baseload scenario, the Power BI Premium administrator will add all the employees into the capacity Assignment List, and from that point on all the content authored in the organization will be hosted and served through the dedicated capacity.

With the capacity in place, the organization immediately enjoys the ability to have unlimited Power BI apps creation and distribution by any Pro user to the whole organization, and the consumption of the apps by any user throughout the organization. In addition, all the user quotas and the default restrictions on dataset size and refresh rates are removed and are now in the control of the Capacity Administrator who can set new policies and have firm controls and governance of the performance and SLA of the BI system.

Application capacity

When a BI project needs to distribute a set of dashboards and reports to a large number of users, and in the absence of baseload capacity, Power BI Premium dedicated capacity can be deployed for the narrow purpose of supporting one specific application.

For example, the HR team may want to distribute a set of dashboards and reports to all the employees in the organization. The project team in HR will purchase and provision dedicated capacity and assign to it only their project. This narrow application targeting will allow the team to ensure that only the exact capacity needed for their Power BI app is deployed and they don't have to take any "freeloaders" onto their capacity. The HR application can then be distributed to all the employees with no additional per- user licensing costs.

In this scenario, the usage is limited to one (or few) known app(s) and the capacity provisioned can be fully tailored to the needs of that app. The administrator is typically part of the project team and will size and provision the capacity that will support the scale and load of that specific app.

With the capacity in place, the application team enjoys the unlimited application distribution of the Power BI app to the whole organization without the need to license the individual users. The capacity in this scenario is optimized for the scale, performance, and budget of the application project.

Solution capacity allows each project or team in the organization to own and manage the optimized Premium Capacity specifically needed to support their application.

Customized capacity

Even when a baseload capacity is available in the organization, certain projects or applications may have unique needs that go beyond the design point of the baseload capacity. In this case the project team will provision their own customized dedicated capacity and will host their own application.

Note: Different applications can be hosted on different capacity units inside the org without any effect on the user experience. The experience remains the same regardless of the content hosting topology.

Examples when customized capacity may be needed include the following:

- **Extreme Scale** – the BI project requires very large Capacity Nodes that the baseload capacity does not have deployed.
- **Extreme Distribution** – the project is designed to reach a very large population, much larger than the baseload capacity was designed to support.
- **Extreme Peak** – the project is designed for a specific moment in time—perhaps the Olympic Games or the presidential election. For that time an extreme load is projected on the system, which the baseload capacity cannot support.
- **Extreme SLA** – BI solution owners want to ensure the most optimized performance for their project and they don't want any interference from other applications in the organization. An example would be the BI solution for the CEO dashboard.

In these scenarios, the project team will provision and manage their own customized capacity that will be used by their application only, while the rest of the organization will continue to use the baseload capacity. Power BI consolidates the content and applications hosted on all the Capacity Nodes and provides end users with a single pane of glass view of all the BI content in the organization.

New economic disruption enables Data Culture everywhere

We hope you can now appreciate the breakthrough capabilities enabled by the Power BI Premium offering. It sets a new bar for scale, performance, and deployment size for cloud BI.

But the Power BI Premium disruption is much more than just feeds and speeds—it is a market changer. When Power BI was introduced to the market, with a freemium model and a price of only \$9.99 per user, per month for a Pro license, it completely disrupted a market accustomed to very high price points.

Power BI Premium is unleashing the next economic disruption in the BI market. Powered by VertiPaq technology, the Premium capacity nodes are designed to be incredibly flexible—which means that you now have the ability to support very large deployments in a very affordable way.

We encourage you to spend some time with our Premium Capacity calculator on our website and see the estimated capacity needed for your large deployment. With Power BI, we have a solution that fits your budget need.

With Power BI, every organization can now afford to reach the state of Data Culture—the ability of every employee to read, use, reason, and act on data as the core decision process of the business.