



How to plan capacity for Power BI Embedded

Whitepaper

By Aviv Ezrachi, Senior PM, Microsoft

©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

Power BI Embedded offers app developers/ISVs a predictable and affordable embedded analytics' comprehensive set of BI capabilities.

This paper provides guidance for the app developer/ISV to determine the most appropriate capacity for its businesses. This guidance is specific for Power BI Embedded and intended for app developers/ISVs who leverage a multi-tenant system.

Additional documentation and guidance for Power BI Embedded is available [online](#).

Licensing requirements for Power BI Embedded

Two elements are required to take embedded analytics to production. First, Power BI Pro licenses are required for administration, content publishing, and development. Then, Power BI Embedded is required for testing and deployment.

Development

Development requires a Power BI Pro license for each developer on the team. Anyone using the Power BI APIs is required to have a Power BI Pro license. Power BI Pro licenses contain all APIs and can be used to develop a full proof of concept.

Pre-production and production

For testing and deployment, Power BI Embedded must be acquired. Power BI Embedded enables full testing of the solution with embed tokens that allow multi-user access to the embedded Power BI reports and dashboards. When deploying to a production environment, options include using the same Power BI Embedded capacity used for pre-production or a dedicated capacity for pre-production and another for production.

How much capacity do I need?

Calculating how much capacity is needed is based on multiple parameters like the specific data models used, the number and complexity of queries, the hourly distribution of the usage of your SaaS application, the data refresh rates, and other usage patterns that are hard to predict.

This paper serves as guidance when planning capacity. Capacity estimation is not an exact science and performance cannot be guaranteed.

APP DEVELOPER/ISVS SHOULD USE THE GUIDANCE IN THIS PAPER. PLAN ON PROVISIONING ENOUGH CAPACITY TO COMFORTABLY HANDLE PEAK USAGE. IF SUFFICIENT CAPACITY IS NOT RESERVED, CORES COULD BECOME OVER LOADED AND PERFORMANCE WILL SLOW.

Dedicated capacity nodes: The V-Cores

Purchase enough capacity to support the workload.

Power BI Embedded Capacity Options

Name	Virtual Cores	Memory	Frontend / Backend Cores
Power BI Embedded A1	1	3 GB RAM	.5/.5 ¹
Power BI Embedded A2	2	5 GB RAM	1/1 ¹
Power BI Embedded A3	4	10 GB RAM	2/2
Power BI Embedded A4	8	25 GB RAM	4/4
Power BI Embedded A5	16	50 GB RAM	8/8
Power BI Embedded A6	32	100 GB RAM	16/16

1. Shared resources

Power BI Embedded planning for app developers/ISVs

Plan on provisioning enough capacity to comfortably handle peak usage. Because Power BI Embedded is a solution for report rendering and data querying, app developer/ISV capacity calculations are based on a **page renders' range** at peak hour.

We do not include data refresh as a parameter for the peak hour load. To keep data fresh without affecting performance, run data refresh during off-peak hours.

What is a page render?

A page render is counted any time Power BI visuals are loaded on a page. A page refresh counts as a page render, as does any other page interactivity, like slice and dice, filtering, etc. Typically, multiple queries running on the backend cores¹ translate to a page render.

The load generated on the cores depends on the number of visual elements contained on a page. A typical page contains up to 10 visual elements.

¹ You can learn more about cores in the [Power BI Premium white paper](#).

	Virtual Cores	Memory	Price/hour	Peak Renders per hour
Power BI Embedded A1	1	3 GB RAM	\$1.008	300
Power BI Embedded A2	2	5 GB RAM	\$2.00081	600
Power BI Embedded A3	4	10 GB RAM	\$4.0242	1,200
Power BI Embedded A4	8	25 GB RAM	\$8.0565	2,400
Power BI Embedded A5	16	50 GB RAM	\$16.1210	4,800
Power BI Embedded A6	32	100 GB RAM	\$32.2506	9,600

Managing costs

Power BI Embedded Azure resource consists of several features that enable different usage patterns for our vast range of ISV customers. Mainly this is the case for scaling and utilizing the service based on usage to control costs.

Pause/Resume of the Power BI Embedded service

On creation of the Power BI Embedded resource, it is automatically being deployed and started. At any time, the resource owner can pause the service so that there will be no charges for it when not at use. Behind the scenes we will automatically evacuate the models loaded into memory and release the resource. When choosing to go back and restart the service, models would be loaded into memory again according to the loaded reports used by logged in customers.

Scaling of the Power BI Embedded service

One of the advantages of running Power BI Embedded in Azure is being able to dynamically manage it to adapt to changing workload demands. You can scale-up and scale-down the SKUs (CPU/memory resources) as you need.

Examples

Example 1

The ISV knows that the SaaS App with Power BI Embedded handles 100 users in the peak hour. It is assumed that these users will trigger a total of 250 page renders for that hour because each user will load a report and interact with it 2.5 times during the peak hour. Per the table, the ISV should choose Power BI Embedded A1.

Example 2

The ISV reading its telemetry data knows the App usage patterns of the application with Power BI Embedded reports. The ISV assumes there is potential for 100,000 monthly page renders. A month consists of 220 hours of work which is spread across 22 days with 10 hours of work in a day.

By dividing the expected page renders in a month by the hours in a month, we show that an average hour handles 455 page renders. From telemetry data, the ISV knows the ratio between peak and average use is 4:1. Using this ratio they calculate page renders at the busiest hour in the month as 1820, for which the ISV should choose Power BI Embedded A4.