

Request Limits and Allocations considerations

Dynamics 365 FastTrack Architecture Insights

Jason du Plessis Dynamics 365 FastTrack



Agenda

- Why are we introducing Requests limits and allocations?
- What are the requests being counted?
- What are the allocations?
- What will happen when the limits are exceeded?
- How to get started

Why are Request limits and allocations being introduced?

Why are Request and allocations being introduced?



To help ensure service levels, availability and quality, there are entitlement limits to the number of requests users can make each day across model-driven apps in Dynamics 365 (such as Dynamics 365 Sales and Dynamics 365 Customer Service), Power Apps, and Power Automate.



By providing a single view of resource consumption per user to customers we are empowering them to understand what their usage looks like, and providing a clear conversation with regards to their architectural design and best usage of tools and features.

What are the requests being counted?

What are the requests?

Requests are counted across these services, measured within any 24-hour period

Power Automate Flow

All API requests to connectors, process advisor analysis, HTTP actions, and built-in actions from initializing variables to a simple compose action. Both succeeded and failed actions count towards these limits. Additionally, retries and other requests from pagination count as action executions as well.



Microsoft Dataverse

All CRUD Operations, Share, Assign, Associate, Disassociate



Power Virtual Agent

API requests (or calls) to Power Automate flows from within a chatbot conversation.



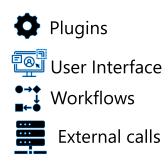
Power Apps

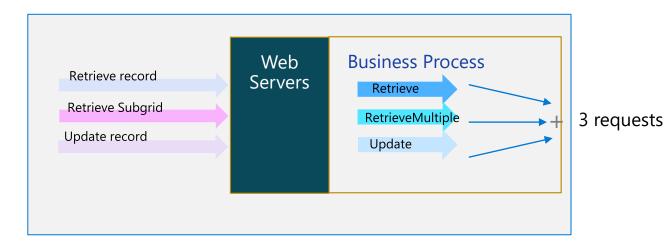
All API requests to connectors and Microsoft Dataverse

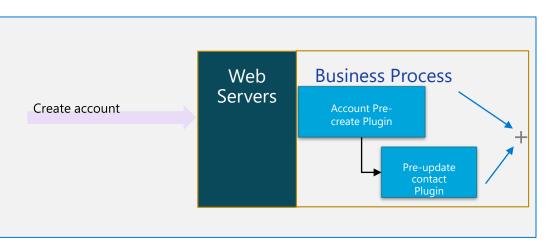


Microsoft Dataverse

- Requests consideration is at the business operation level, not at the external API interface
- **ALL** operations whether directly from external API, from plugin, workflow etc are counted, for all users, and for all entities except for a list of excluded entities.
- External individual requests, batch requests, User Interface requests, and any plugins or workflows they trigger are counted



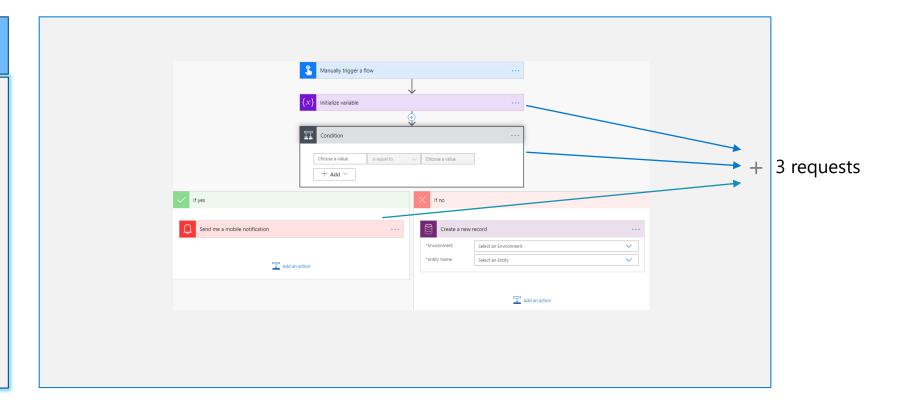




2 requests

Power Automate Flow

- All API requests to connectors, process advisor analysis, HTTP actions, and built-in actions from initializing variables to a simple compose action.
- Both succeeded and failed actions count towards these limits.
- Additionally, retries and other requests from pagination count as action executions as well.



- These are counted towards either the shared request allocation pool if Flow is executing per User Plan, of separate per flow request pool if the Flow is running Per App Plan.
- Note that request from Power Automate flow to Microsoft Dataverse are not double-counted
- Note that Power Automate Per Flow Plan does not consume from the tenant-wide limit



Power Apps

All API requests to connectors and Microsoft Dataverse

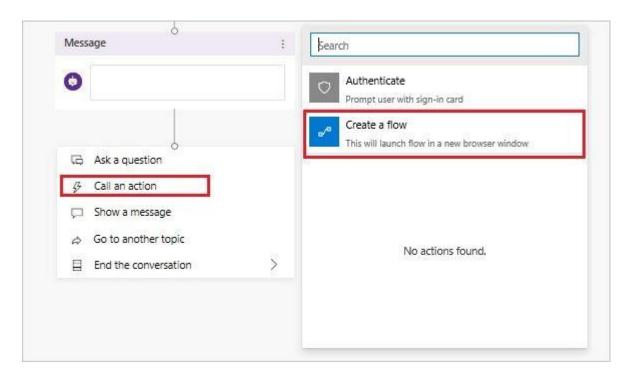


Any usage of a Connector will be counted towards the per user request allocation pool.



Power Virtual Agent

API requests (or calls) to Power Automate flows from within a chatbot conversation.





What are the allocations?

Request allocations

Microsoft Power Platform requests allocations based on user licenses

All the users of Microsoft Power Platform can use a certain number of requests based on the license they are assigned.

The following table defines the number of requests a user can make in a 24-hour period.

Products	Requests per paid license per 24 hours
Paid licensed users for Power Platform (excludes Power Apps per App, Power Automate per flow, and Power Virtual Agents) and Dynamics 365 excluding Dynamics 365 Team Member	40,000
Power Apps pay-as-you-go plan, and paid licensed users for Power Apps per app, Microsoft 365 apps with Power Platform access, and Dynamics 365 Team Member	6,000
Power Automate per flow plan ³ , Power Virtual Agents base offer, and Power Virtual Agents add-on pack ⁴	250,000
Paid Power Apps Portals login	200

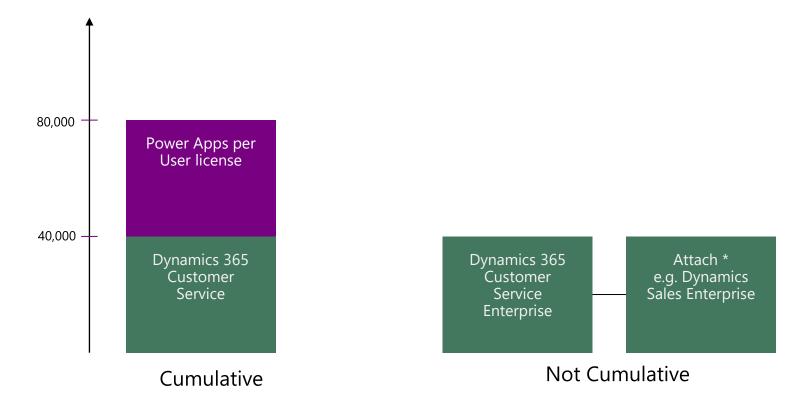
Entitlements subject to change, please refer to the licensing guides for the latest information: Dynamics 365 Licensing Guide and Power Apps, Power Automate, and Power Virtual Agents Licensing Guide

Request Allocations

User licenses: Cumulative if different Product lines

If a user has multiple paid licenses assigned to them, the total number of requests allowed would be the sum of requests allowed for each license. For example, if a user has both a Dynamics 365 Customer Service Enterprise base license and a Power Apps per user license then that user will have a total of 40,000 + 40,000 = 80,000 requests available per 24 hours.

Power Platform Request limits are only included with paid base licenses in the Dynamics 365 'base + attach' licensing model. Attach licenses do not include seperate limits



Request Allocations

Non-licensed users/application users/Users with special free licenses

Microsoft Dataverse also provides the ability to have identities that do not require any user license to interact with the service.

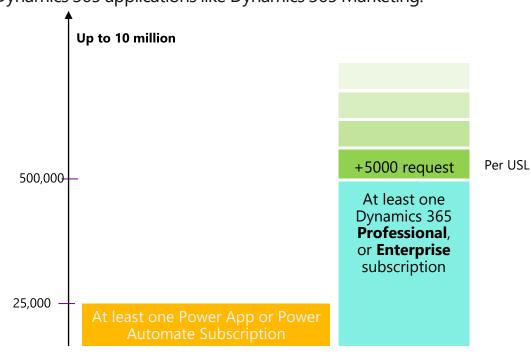
These include

- Application users
- Non-interactive users
- Administrative users.
- SYSTEM user

Additionally, there are special free (\$0) licenses, which are used to interact with Dynamics 365 applications like Dynamics 365 Marketing.

For these users, every **tenant** will get base request capacity per tenant that can only be used by these users and not by users with standard licenses, plus accrued limits determined by the quantity of Dynamics 365 Enterprise and Professional Licenses

Products	Pooled non-licensed tenant-level requests per 24 hours
Dynamics 365 Enterprise & Professional applications	500,000 base requests + 5,000 requests accrued per USL ¹ up to 10,000,000 max ²
Power Apps (all licenses)	25,000 base requests with no per-license accrual for the tenant
Power Automate (all licenses)	25,000 base requests with no per-license accrual for the tenant



Per tenant

Let's look at an example

A customer has 1000 Dynamics 365 Customer Service Enterprise+Sales Enterprise Attach licenses, and 500 Power App Platform licenses

- One user has both the Dynamics 365 Customer Service Enterprise+Sales Enterprise Attach licenses, and Power App Platform license assigned, providing 80 000 requests in 24 hour period
- There are 50 service account is used for integration. The total request allocation for non-interactive requests is $500\ 000+5000*1000 = 5\ 500\ 000\ requests$

In one 24 period for the user

- The user launches a Power App that connects to 3 services
- The user launches a Power Automate Flow 30 times that executes 100 steps.
- The user creates 100 new case using Customer Service, each triggers a plugin under the user context that retrieves a custom Validation table to check access, total requests per create including Form load requests is 20 (for example)
- Total consumption for the user for 24 hour period is

3+30*100+100*20 = 5003 requests of entitlement of 80 000

In one 24 period for the non-licensed users

- For each 100 case create, there is has another plugin that runs under the SYSTEM account that performs 5 request operations
- There is an integration the executes 50 000 retrieves in 24 hours
- Total consumption for the 24-hour period is:

100*5+50000 = 50 500 requests of entitlement of 5 500 000

What will happen if the allocation is exceeded?

Impact of reaching request allocation limits

- Power Platform Request limits have been updated and substantially increased in late 2021 to be at levels that
 are significantly higher than typical usage for most customers. With the updated limits, expectations are that
 very few users would exceed the documented limits. If you anticipate exceeding the non-licensed user limits,
 reach out to your reseller or Microsoft Account team to discuss a custom solution.
- Any possible high usage enforcement will not happen until six months after Power Platform Request usage reporting has been generally available in the Power Platform Admin Center. Admin reports are expected to be available in the first quarter of calendar 2022.
- Microsoft reserves the right to enforce limits for overages. If a customer encounters high usage enforcement, they can expect some form of throttling. Customers can purchase additional capacity to avoid high usage enforcement or move their environment to pay-as-you-go and pay for actual usage above daily limits.

How to get started

EXPLORE

Talk about architecture and consumption

Where do I start?

- Determine your request consumption across the tenant and identify licensed and nonlicensed users that are highest consumers
- Investigate further by environment, application, and table.
- Identify the usage patterns and review alternative design options that can reduce overall usage

Create an action plan

- Based on the usage pattern highlight the key areas that need to be invested in.
- Review if the approach can be optimized to reduce requests, and if it makes use of modern cloud capabilities.
- Measure your changes and review impact and iterate.

Explore consumption via reporting

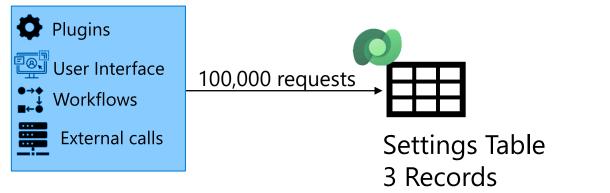
- Reporting for Power Platform Request usage will be in the Power Platform admin center when generally available, expected in the first quarter of calendar 2022. This reporting will initially be a downloadable Excel format with daily reporting of request usage versus limits.
- Until those reports are available, the Power Platform admin center contains reports on Dataverse API
 requests. This reporting accounts for interactive and non-interactive traffic. This helps you to quickly view
 adoption and user metrics for your organization. If your apps or flows primarily use Dataverse, then these
 reports can serve as good approximations of the total usage of your solutions.
- Additionally, for Power Automate usage specifically, you can see the action usage for a given flow by selecting the Analytics action from the flow properties page, and this works across all types of actions.

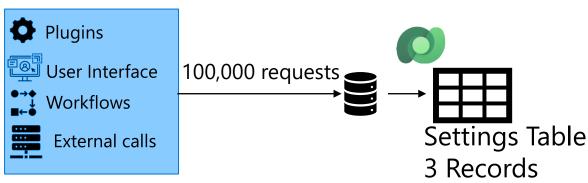
Review architectural approaches

- Review current architectural approach for any optimizations that can reduce overall resource usage
- Considering making use of cloud-services to store data outside of the Dataverse (general rule is store only data in the Dataverse that is needed by the Dataverse Applications)

As an example consider a commonly seen pattern is where a table with only a few records is being queried many 100 000's of times a day

- Typically this would be to store dynamic configuration, either for integration, user experience or plugin purposes. This is an easy approach, but very resource intensive.
- Typically this type of data is slow changing, so rather consider other more scalable solution that makes use of a caching mechanism or similar





Thank you