Microsoft Dynamics 365

Finance & Operations
Priority-based throttling

Davy Vliegen
Senior FastTrack Solution Architect

Agenda

- Priority-based throttling explained
- Current status
- Guidance
- What’s next?
- References

Priority-based throttling explained

Improve availability, reliability and reduce operational costs by controlling the frequency of operations and consumption of underlying resources.

Priority-based service protection framework that allows:

- Monitoring and controlling service usage per user or AAD application and return a 429 HTTP response when thresholds are reached
- Customers to set relative priority for specific business-critical integration workloads and users (vs. less-critical)

For Customers: a graceful approach to deal with workload and failing requests

For Microsoft: safeguarding resources and avoiding outages, reduction in incidents and focus on 'real' service degradation issues

Current status

Since v10.0.19 (PU43) priority-based throttling enabled by default

- For OData and custom service-based integrations
- Not applicable for interactive users
- Only resource-based throttling (AOS CPU/Memory)
- 429 responses are automatically triggered

Internal Microsoft services excluded from throttling

Temporary exempt for throttling on a case-by-case basis

Guidance

Implement retry operations

- When a request is throttled and a 429 error occurs, the response header will include a Retry-After interval
- Can be used to retry the request after a specific number of seconds

Adjust and distribute workload

- Use dedicated integration accounts
  - Because these service protection settings are set up for user- or AAD application-specific values
- Configure priorities
  - Setting appropriate priorities ensures that low-priority integrations will be throttled before high-priority integrations

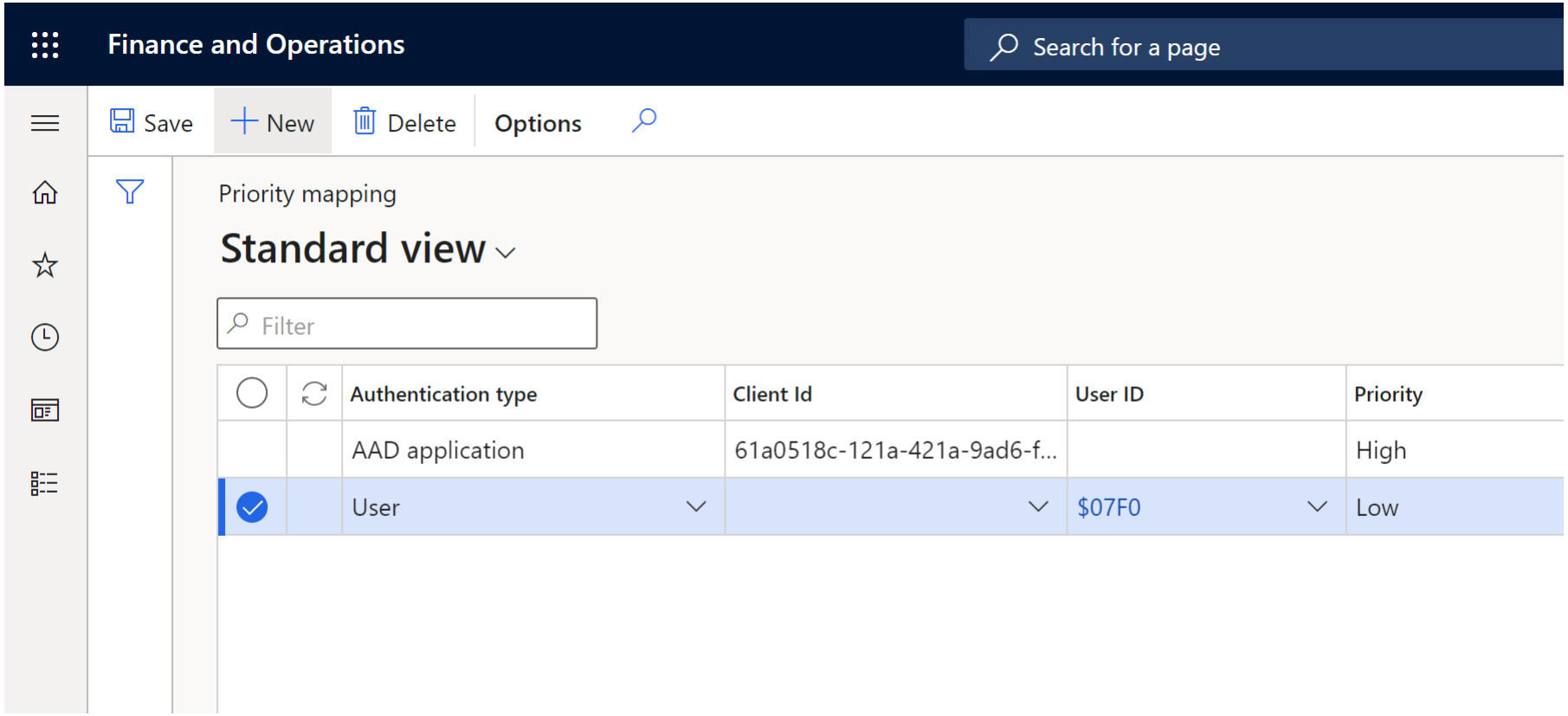
Test and Monitor

- Requests throttled view in LCS Environment monitoring

Open support ticket for additional help

Configure priorities

System administration > Setup > Throttling priority mapping



You must be assigned the System administrator or Integration priority manager role to complete the set up

What's next?

Usage-based throttling limits

Default service protection API limits:

Measure	Description	Service protection limit
Number of requests	The cumulative number of requests made by the user	6000 within the 5-minute sliding window
Execution time	The combined execution time of all requests made by the user	20 minutes (1200 seconds) within the 5-minute sliding window
Number of concurrent requests	The number of concurrent requests made by the user	52

Important: These limits are subject to change and may vary between different environments. These numbers represent default values and are provided to give you some idea of what values you can expect

Timeline

No fixed timeline yet

Rollout incrementally to environments

- Starting with environments with lowest API usage
- Increased throttling limits on a case-by-case basis

Proactive communications

- Updated Docs content
- Webinars (e.g. TechTalk)
- Direct conversations

References

Priority-based throttling - Finance & Operations | Dynamics 365 | Microsoft Docs

Priority-based throttling FAQ - Finance & Operations | Dynamics 365 | Microsoft Docs

Thank you