Microsoft Dynamics 365

What's new in resource governance for Dynamics 365 Business Central online

Agenda

- Resource Governance Mission
- Resource Governance Story
- Shared/Multi-Tenant Architecture
- Resource Consumption Tracking
- Key Improvement Areas
- New Operational Limits
- Preview Practicalities

Resource Governance Mission

Resource Governance – Mission

*To offer **optimal** and **fair** resource consumptions to all Dynamics 365 Business Central online customers*

- ***Optimal** means we offer maximal availability & performance for all consumptions w/ minimal redundancy*
- ***Fair** means we treat all customers and their users equally with respect to capacities and proportionately according to their licenses, ensuring that they can consume…*
  - *HOW (WHAT)? In their entitled manner (license-specific **features**)*
  - *HOW MUCH? As much as possible… as long as it's covered by their licenses (entitlement **quotas**)*
  - *HOW FAST? As fast as possible (top speed for all users)… as long as their consumptions don't adversely impact others' (operational **limits**)*
  - *QUOTAS vs. LIMITS?*
    - *In governance terminology, quotas are "entitlements/rights" for customers and their users to claim/exercise, limits are "laws/rules" to follow*
    - *At the user level, quotas are **tiered**, **cumulative** w/ each license purchased, and **transferable** among users*
    - *At the user level, limits are **fixed**, **equal** for all users, and **non-transferable** among users*

Resource Governance Story

Resource Governance – Story

- **2018**: We lifted & shifted Dynamics 365 Business Central into Azure to run on a **shared/multi-tenant** architecture as Software-as-a-Service (SaaS)

- We started to build an elastic service using the following techniques: **mounting** environments on multiple server instances/hosts, **balancing** workloads across hosts, **grouping** hosts to serve different environments, and adding new hosts/**scaling** out

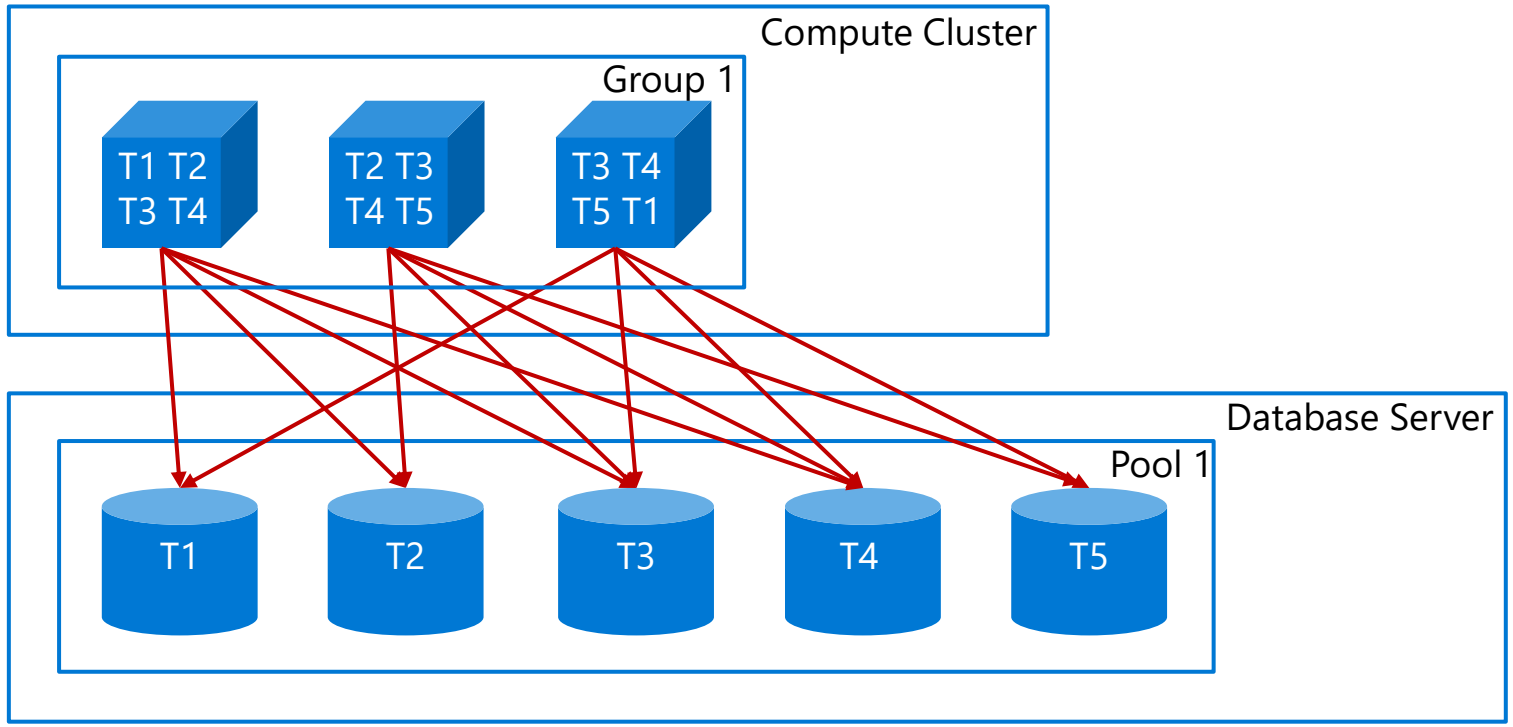
- While we're building our service, we must already govern resource consumptions, so we started w/ **conservative** limits and **indefinite** quotas

- **2022**: We started to **track** resource consumptions
  - Track all **workload/capacity units** (e.g. tasks, requests, sessions, execution time, memory, query time, storage, etc.) consumed by any **user**, **environment** (NAV tenant), and **customer** (Microsoft Entra tenant) at any given moment/period

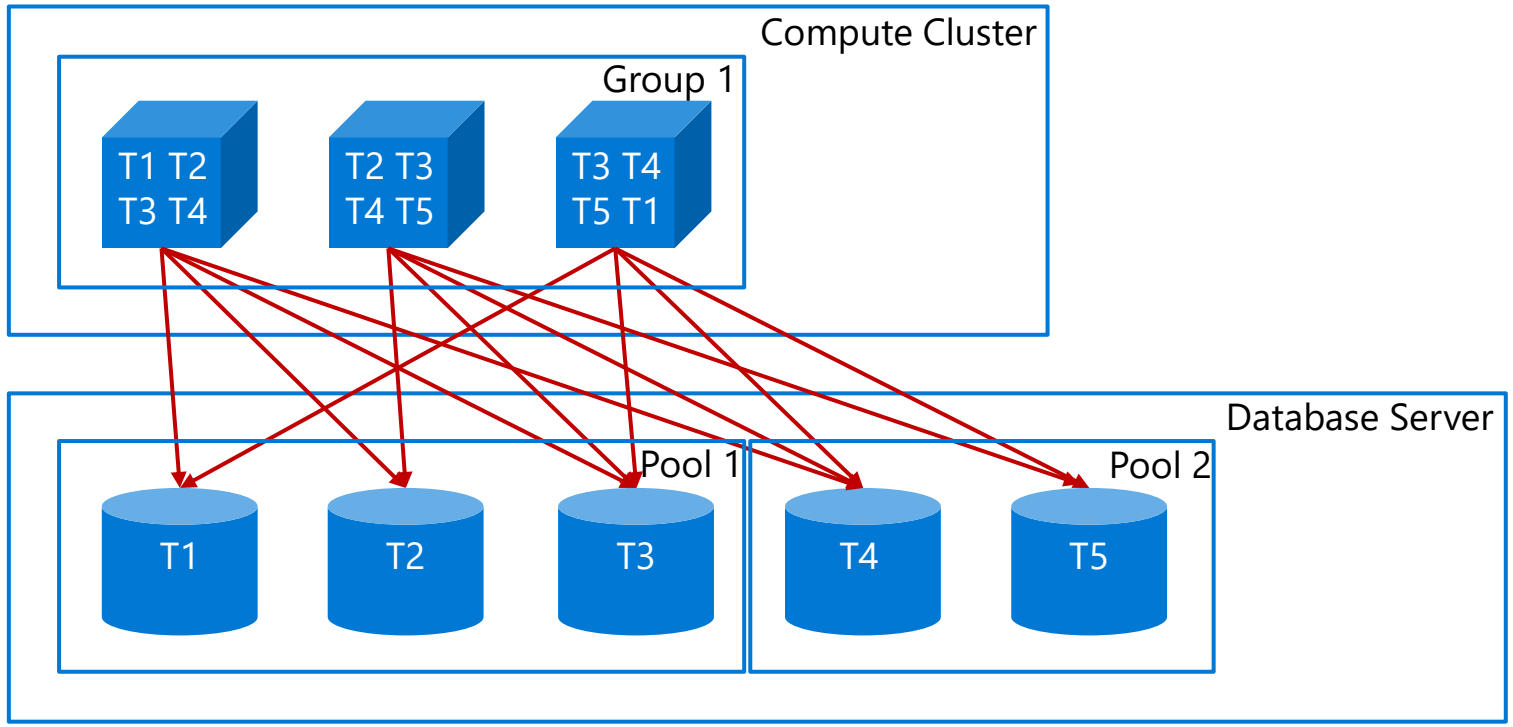
- As our service matures, resource consumption tracking also allows us to **improve** our elastic techniques, as well as offer more **expansive** limits and **quantifiable** quotas

Shared/Multi-Tenant Architecture

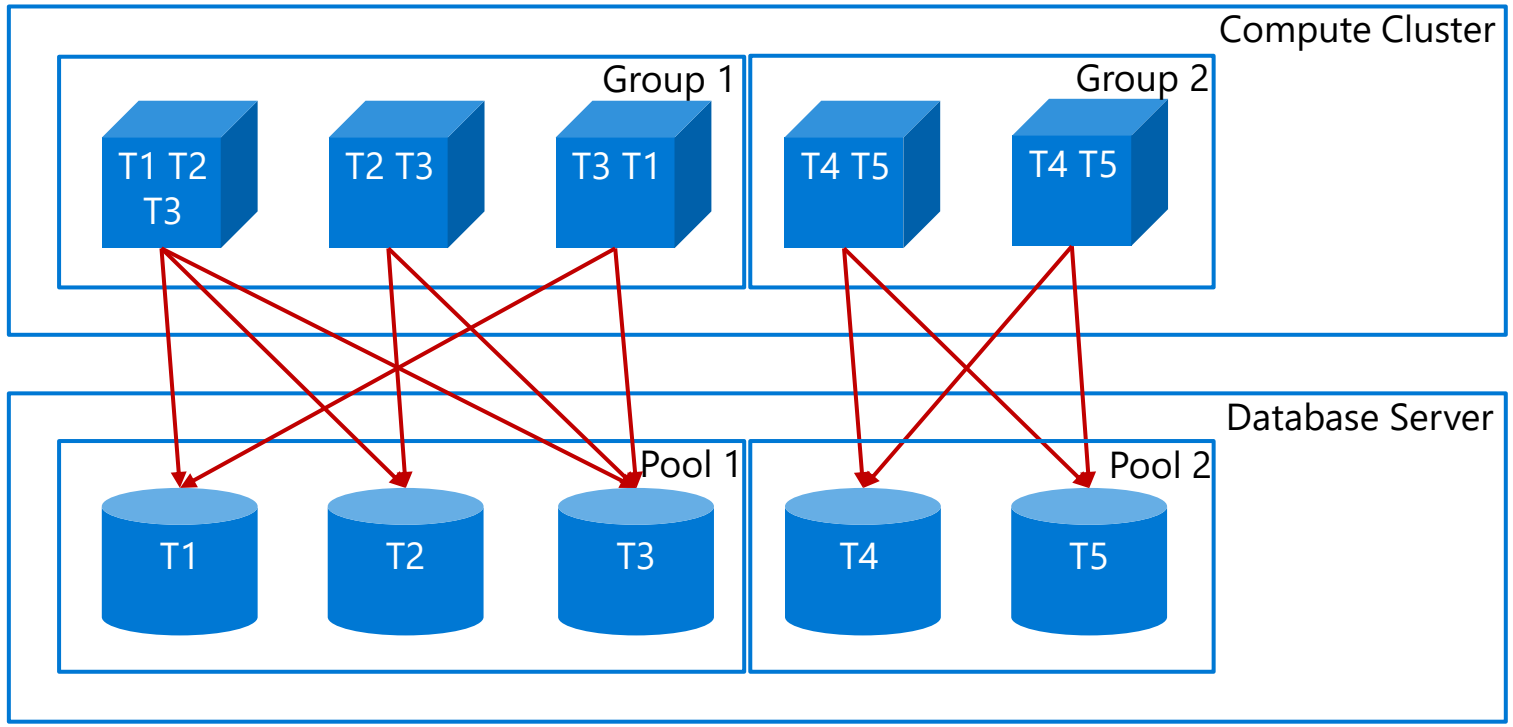
Resource Governance – Shared/Multi-Tenant Architecture



Resource Governance – Shared/Multi-Tenant Architecture



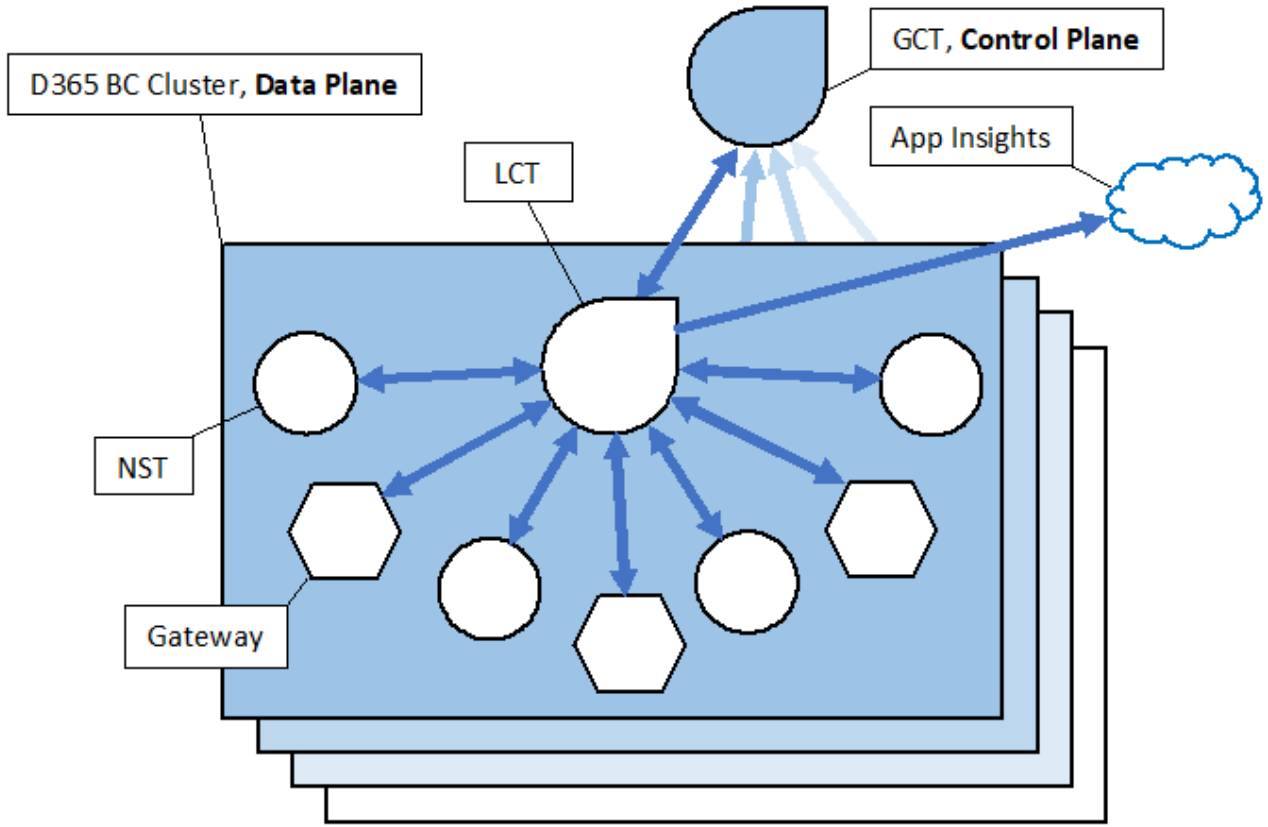
Resource Governance – Shared/Multi-Tenant Architecture



Resource Consumption Tracking

Resource Governance – Consumption Tracking

- Track all **workload/capacity units** (e.g. tasks, requests, sessions, execution time, memory, query time, storage, etc.) consumed by any **user**, **environment** (NAV tenant), and **customer** (Microsoft Entra tenant) at any given moment/period
  - Create **Local Consumption Tracker (LCT)** service to locally track consumptions by each user/environment in each cluster
  - Create **Global Consumption Tracker (GCT)** service to globally aggregate consumptions for each customer across all users/environments



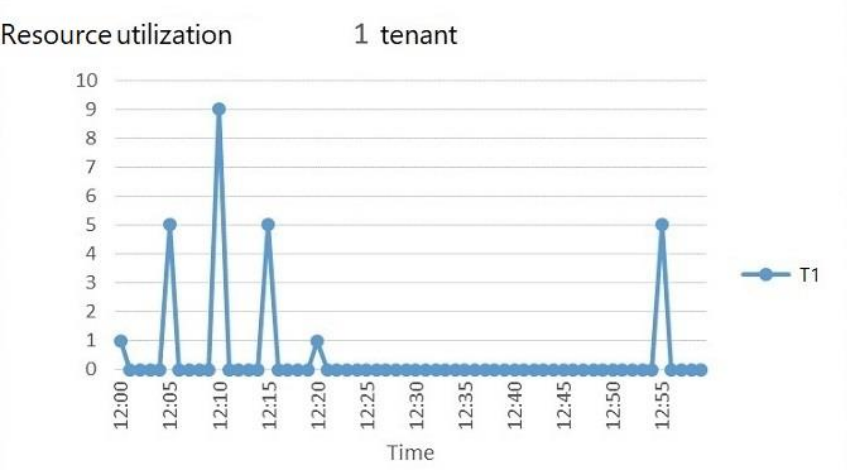
Resource Governance – Consumption Tracking

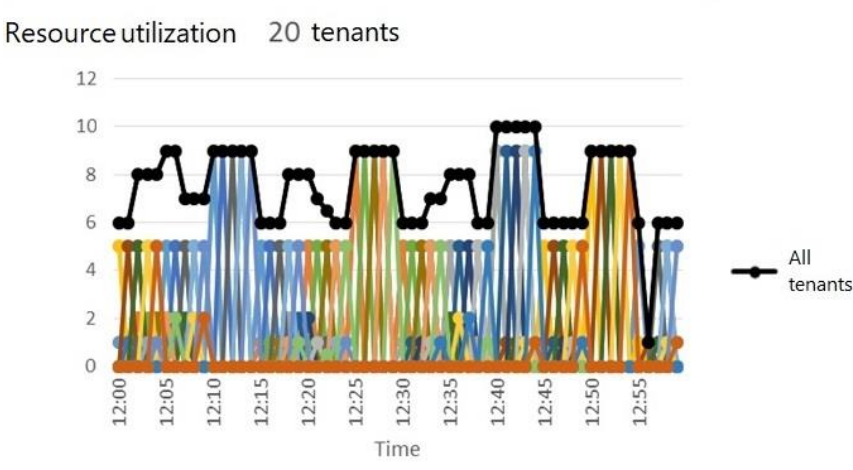
- Track all **workload/capacity units** (e.g. tasks, requests, sessions, execution time, memory, query time, storage, etc.) consumed by any **user**, **environment** (NAV tenant), and **customer** (Microsoft Entra tenant) at any given moment/period

no/short-term memory

consumption tracking

long-term memory





Key Improvement Areas

Resource Governance – Key Improvement Areas

1. Operational limits that are fair and expansive
   - Publish clear **user-based limits** that protect our service, ensure fairness among users, and inspire elasticity/scalability (ETA: **Q4CY23**)
   - Apply dynamic **resource-based limits** that protect our service and ensure fairness among customers (environments) (ETA: **Continuous/ongoing** improvements)
2. Intelligent balancing, mounting, grouping, and scaling
   - Use consumption patterns and limit exhaustions to distribute/rebalance workloads across current hosts, mount environments on more hosts, group hosts to serve different environments, and scale out, before **throttling as a last resort** (ETA: **Continuous/ongoing** improvements)
3. Entitlement quotas and flexible excess consumptions
   - Publish clear **user-based/tenant-wide quotas** for licensed/unlicensed users, respectively, and allow buying **excess consumptions** (ETA: **CY24/25**)

For these improvements, we'll start w/ the following types of resources:

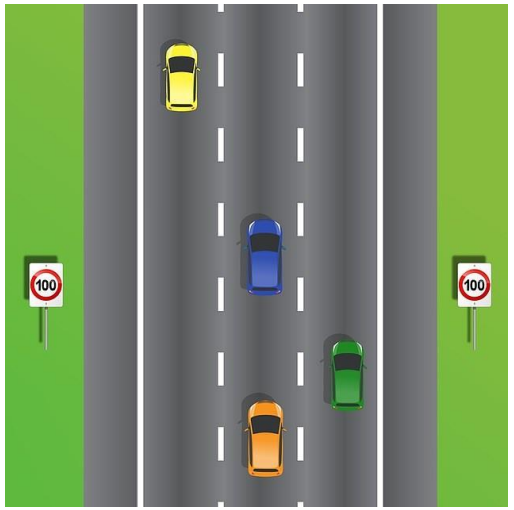
- **Web service request handling** resources
- **Compute** resources
- **Database** resources

New Operational Limits

Resource Governance – New Operational Limits

- For complete governance, we need to set up new **speed, concurrency, and execution time** limits

A “loose” highway analogy:

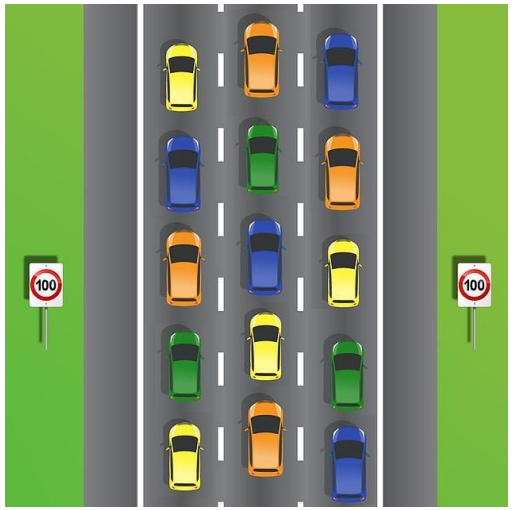


- Imagine a *stretch of highway* that represents our *limited host capacity*
- *Traversing the highway* represents *consuming a portion of host capacity*
- *Cars* represent *workloads submitted by our customers and their users*, which are color-coded as Blue/Green/Orange/Yellow
- All users must obey the *same rule/speed limit t*o enter/merge into the highway
- *Higher toll charges* allow travelling further and they represent *premium licenses* that allow consuming *higher entitlement quotas*

Resource Governance – New Operational Limits

- For complete governance, we need to set up new **speed, concurrency, and execution time** limits

A "loose" highway analogy:

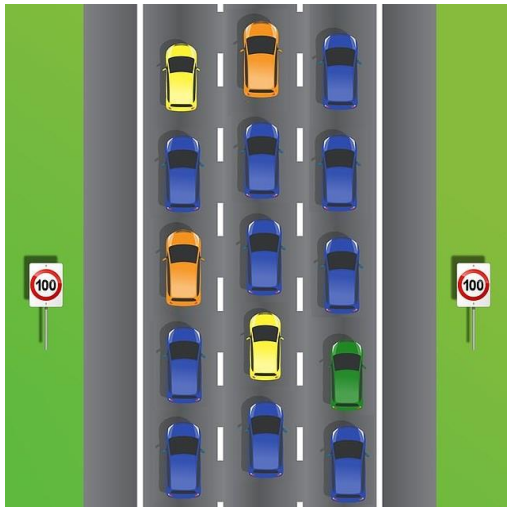


Speed limit:

- To control how **fast** cars enter the highway.

In Business Central online/SaaS:

- To control how **fast** workloads enter our system.

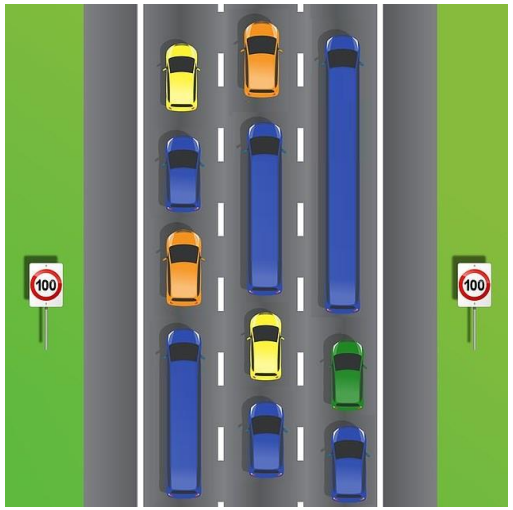


Concurrency limit:

- To control how **many** cars are present in the highway.

In Business Central online/SaaS:

- To control how **many** workloads are concurrently processed in our system.



Execution time limit:

- To control how **long** cars are present in the highway.

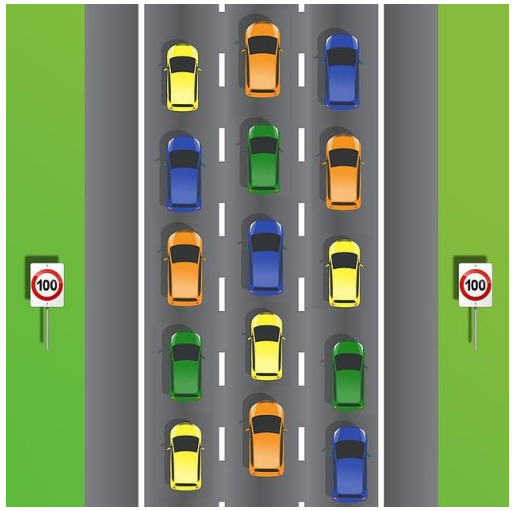
In Business Central online/SaaS:

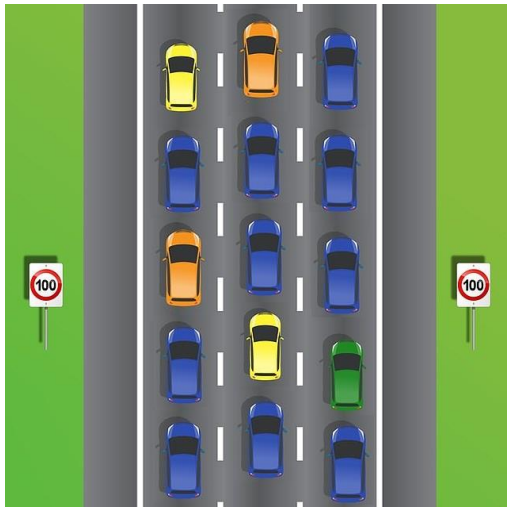
- To control how **long** workloads are processed in our system.

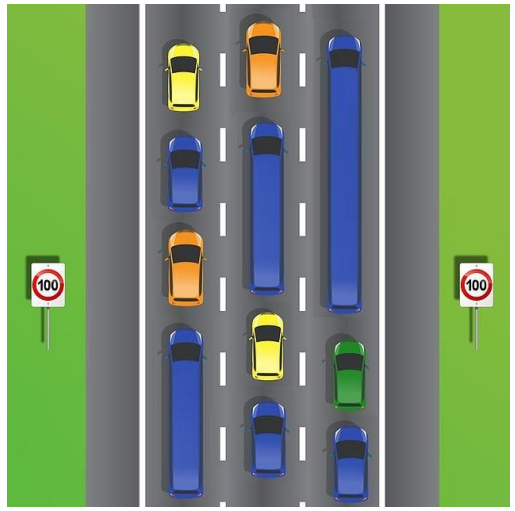
Resource Governance – New Operational Limits – PREVIEW

- For complete governance, we need to set up new **speed, concurrency, and execution time** limits

A "loose" highway analogy:







Speed limit:

For web service requests:
- OLD: **600** web service requests per **environment** per **minute**
- NEW: **6000** web service requests per **user** in the previous **5-minute sliding window**

(**ETA: Mid Q4CY23**)

Concurrency limit:

For scheduled tasks:
- OLD: **3** concurrently running tasks per **environment**
- NEW: **5** concurrently running tasks per **user**

(**ETA: Early Q4CY23**)

For web service requests:
- OLD: **100** concurrently handled (5 processed & 95 queued) web service requests per **environment**
- NEW: **100** concurrently handled (5 processed & 95 queued) web service requests per **user**

(**ETA: Late Q4CY23**)

Execution time limit:

- NEW: **X** minutes per **user** in the previous **5-minute sliding window**
- NEW: Each **environment** can consume up to **Y%** of the total execution time in one **host**
- NEW: Each **environment** can consume up to **Z%** of the total execution time in one **cluster**

(**ETA: TBD**)

NOTES: Limits are applied the same way to **ALL** users, including app (S2S)/device license users

Preview Practicalities

Resource Governance – PREVIEW Practicalities

- Throughout **Q4CY23**, our new operational limits are **subject to change** as we continue **monitoring** and **fine-tuning** them to provide **higher throughput** than the old ones

- Throughput will be gradually increased throughout out the PREVIEW period:

- We'll communicate any changes to these limits in our **release plans** , e.g. minor 23.x releases, and **documentations**, e.g. Operational Limits in Dynamics 365 Business Central - Business Central | Microsoft Learn

- As always, we appreciate your **feedbacks** and **suggestions**, e.g. via https://aka.ms/BCYammer

Resources

General Business Central resources, learn more!

Have a question?
aka.ms/BCYammer

Join the conversation
twitter.com/
MSDyn365BC

Looking for resources?
aka.ms/BCAll

Submit your ideas
aka.ms/BCIdeas

Join the office hours
aka.ms/BCOfficeHours

Thank you