

# Limites de uso técnico da plataforma

Versão: v1.0

Última modificação em: 4 de abril de 2022.

## Índice

<b>Índice</b>	<b>1</b>
<b>Histórico de revisões</b>	<b>2</b>
<b>Visão Geral</b>	<b>2</b>
Categorias dos limites de uso técnico	2
Escopos de limites de uso técnico	3
<b>Limites globais de uso técnico da Plataforma Digibee</b>	<b>3</b>
Definições de limites de uso técnico	3
Capacidade de processamento	3
Tráfego de Saída (Egress Traffic)	5
Taxas de Mensagem e Enfileiramento	5
Mensagem	5
MPS (Mensagens por Segundo)	5
Enfileiramento	5
Retenção de Log de Execução	6
Digibee Storage	6
Object Store	6
Gestão de relacionamento	7
VPNs	7
Limites de uso da capacidade de RTU de testes	8
Limites de uso da capacidade de RTU de produção	9

## Histórico de revisões

Versão	Descrição da mudança (seção, página ou texto)	Data
1.0	Criação e Publicação do Documento	4 de abril de 2022

## Visão Geral

Limite de uso técnico refere-se aos limites dos recursos da Plataforma Digibee alocados no seu *realm*. Essas limitações são aplicadas para garantir que a Plataforma Digibee seja usada de forma apropriada por todos os *realms* SaaS para proteger a integridade do ecossistema SaaS da Digibee. Esses limites foram calculados com base na utilização média da plataforma de acordo com o número de RTUs e subscrições ativas.

Dependendo do número de Pipeline Subscriptions e RTUs de teste e produção, seu *realm* terá disponível um limite de uso específico. Esses limites incluem, mas não se limitam, a quantos usuários, implantações de pipeline, recursos de infraestrutura e infraestrutura dedicada você possui na Plataforma Digibee.

Este documento detalha quais são os limites de uso técnico para a Plataforma Digibee.

## Categorias dos limites de uso técnico

Existem duas categorias de limites dentro da Plataforma Digibee:

- **Proporcional:** os limites proporcionais são aqueles que se aplicam em proporção direta ao número de RTUs em cada ambiente de teste ou produção. Eles crescem em função linear do número de RTUs.
- **Intervalo:** limites de intervalo são aqueles que se aplicam a intervalos específicos de RTUs em cada ambiente de teste ou produção. Eles não crescem linearmente mas em blocos.

## Escopos de limites de uso técnico

Cada limite de uso técnico pode ser aplicado em um escopo específico dentro do *realm* do cliente:

- **Um único RTU:** o escopo da aplicação do limite acontece para cada RTU individualmente. Por exemplo, a memória ou CPU para uma implantação de pipeline (incluída em um RTU) não pode ser compartilhada com outra implantação.
- **Todos os RTUs no *realm*:** o escopo da aplicação do limite é a soma de todos os limites de RTU individuais; quando aplicados, cada um conta na soma e não individualmente.
- **RTUs implantados para um único pipeline:** o escopo da aplicação do limite é a soma de todos os limites de RTU para essa implantação de pipeline específica; quando aplicado, cada uso conta na soma e não individualmente.
- **Compartilhado entre *realms*:** esse é um caso muito específico quando o limite é aplicado através de todos os *realms* e a capacidade é implementada pelo time de infraestrutura da Digibee.

## Limites globais de uso técnico da Plataforma Digibee

Os limites globais de uso técnico se aplicam independentemente do número de RTUs na assinatura atual. São eles:

- Até 50 usuários têm acesso ao portal de integração e operação.
- 1 integração *SAML* personalizada com o diretório de identidade.- 1 endpoint de API dedicado por ambiente; ex: <customer name>.test.digibee.io or <customer name>.api.digibee.io

## Definições de limites de uso técnico

### Capacidade de processamento

Cada RTU contém 64MB de memória e 20% do núcleo de um CPU moderno para execução do pipeline. Ele também fornece 10 threads de processamento para execução de fluxo simultâneo.

Para dimensionar a capacidade de processamento, a plataforma permite que os usuários o façam de duas formas diferentes e complementares:

- **Escala vertical:** significa adicionar mais capacidade de processamento a uma única instância de execução de fluxo.
- **Escala horizontal:** significa adicionar mais cópias de uma instância de execução de fluxo.

#### Escala vertical (tamanho de implantação do pipeline)

A plataforma permite que os pipelines sejam implantados em diferentes tamanhos que consumirão mais RTUs. Atualmente, três tamanhos de pipeline estão disponíveis na Plataforma Digibee: pequeno, médio e grande.

A tabela a seguir mostra como a capacidade de processamento cresce em cada um dos tamanhos de implantação:

Tamanho de implantação	CPU*	Memória	Execuções simultâneas	RTUs
Pequeno	20%	64MB	10	1
Médio	40%	128MB	20	2
Grande	80%	256MB	40	4

\* normalmente um núcleo de CPU moderno fornecido por processadores como Intel Xeon

As alocações de memória e CPU definidas devem acomodar a estrutura de fluxo do pipeline (todos os componentes, seus relacionamentos, cápsulas etc.), serviços subjacentes que ajudam o pipeline a executar suas tarefas e as cargas úteis que devem ser processadas pelo pipeline. Definir o tamanho exato da implantação do pipeline para uma determinada integração é uma tarefa a ser executada durante o projeto do pipeline e os testes iniciais. Cada fluxo de pipeline tem suas próprias particularidades que podem exigir mais capacidade.

#### Escala horizontal (réplicas)

Cada pipeline pode ser implantado em várias réplicas, o que basicamente adiciona mais capacidade horizontal ao pipeline. Cada réplica é uma cópia completa de uma implantação de pipeline. Dito isso, se um pipeline for implantado como tamanho pequeno com 1 réplica, isso representaria 1 RTU. Da mesma forma, se o pipeline for implantado como tamanho pequeno com 2 réplicas, seriam 2 RTUs. A fórmula geral para o cálculo do RTU é:

# RTUs = # RTUs para implantação do pipeline de tamanho X # réplicas

## Tráfego de Saída (Egress Traffic)

O tráfego de saída é responsável por todo o tráfego que sai da Plataforma Digibee SaaS. Isso pode incluir, mas não se limita a: chamadas para APIs externas, sistemas, protocolos que não estão hospedados na Plataforma Digibee, conexões VPN, uso de qualquer componente ou cápsula que faça chamadas para provedores externos.

## Taxas de Mensagem e Enfileiramento

### Mensagem

Uma mensagem é a carga processada sempre que um pipeline recebe uma solicitação para processá-la, realizando o processamento e produzindo uma resposta.

### MPS (Mensagens por Segundo)

MPS é a métrica que calcula quantas mensagens são processadas por um determinado pipeline em uma única janela de 1 segundo. O número máximo de MPS é calculado para todo o *realm*. Por exemplo, se um *realm* contiver 20 RTUs para um ambiente de produção e a quantidade de MPS por RTU for 1, a quantidade total de MPS suportada pelo Realm será de 20 MPS. Esse número é compartilhado em todos os pipelines implantados.

Observe que isso é diferente do máximo de execuções simultâneas permitidas em um tamanho de implantação de um pipeline específico. Por exemplo, um pipeline com no máximo 10 execuções simultâneas processará 1 mensagem por segundo quando cada mensagem for processada com um tempo médio de processamento de 10 segundos. Por outro lado, o mesmo pipeline processaria 10 mensagens por segundo quando cada mensagem fosse processada com um tempo médio de processamento de 1 segundo. Assim, o MPS está diretamente correlacionado com o tempo médio de processamento e não necessariamente com o número máximo de execuções simultâneas.

### Enfileiramento

A plataforma permite que as filas de pipeline sejam preenchidas com mensagens pendentes a serem processadas.

O limite imposto ao enfileiramento é calculado por todas as RTUs implantadas em um único pipeline. Por exemplo, se um pipeline for implantado com 4 RTUs e cada RTU fornecer 1.000.000 de mensagens na fila, essa fila de pipeline oferecerá suporte a até 4.000.000 mensagens na fila a qualquer momento.

## Retenção de Log de Execução

A plataforma rastreia a execução de cada mensagem processada pelo sistema. As mensagens são retidas de acordo com o tempo, número de mensagens e quantidade de bytes.

Para registros de execução, que são mostrados na tela Monitor da Plataforma em Execuções Concluídas, o limite é dado pelo tempo de retenção ou quantidade de mensagens, o que ocorrer primeiro. A quantidade de mensagens é contabilizada por número de RTUs implantadas para um determinado pipeline.

Para logs de execução, que são mostrados na tela Monitor da Plataforma em Pipeline Logs ou nos logs de cada execução concluída, o limite é dado pelo tempo de retenção ou pela quantidade de bytes armazenados, o que ocorrer primeiro. A quantidade de bytes armazenados é contabilizada por número de RTUs implantadas para um determinado pipeline.

## Digibee Storage

A Plataforma Digibee fornece um sistema de armazenamento em nuvem para que os pipelines leiam e gravem arquivos. Esses arquivos são acessados através do uso do Componente Digibee Storage. A plataforma aplica limites no número de bytes armazenados neste sistema Cloud Storage.

O limite é um limite global no Realm calculado por RTU.

## Object Store

A Plataforma Digibee fornece acesso a um sistema de armazenamento temporário de objetos que pode armazenar qualquer tipo de objeto JSON. Esses objetos podem ser consultados com base em regras específicas.

Object Stores para o ambiente de produção são fornecidos em camadas que crescem à medida que o número de RTUs aumenta. Eles são instâncias dedicadas para cada *realm*.

Camadas típicas do Object Store e sua correspondência com o número de RTUs:

- Até 60 RTUs: 2 vCPU e 4GB RAM
- Até 120 RTUs: 2 vCPU e 8GB RAM
- Até 180 RTUs: 4 vCPU e 12GB RAM
- E em um padrão semelhante para cada novo bloco de 60 RTUs

Embora a equipe do Digibee Capacity sempre revise essas camadas para que os pipelines possam fazer o melhor uso dos Object Stores, os clientes devem projetar seus pipelines de acordo com as práticas recomendadas dos Object Stores. O uso excessivo do Object Store pode incorrer em penalidades de desempenho para os pipelines.

Object Stores para o ambiente de testes são compartilhados entre *realms* e crescem de acordo com as definições do time Digibee Capacity. Os Object Stores para testes não devem ser muito usados. Eles devem ser usados para validar os aspectos funcionais dos pipelines.

## Gestão de relacionamento

A Plataforma Digibee fornece acesso a um sistema de gerenciamento de relacionamento de ids que pode armazenar mapeamentos entre dados em diferentes sistemas. O limite implícito no sistema de gerenciamento de relacionamento está relacionado à quantidade de mapeamentos exclusivos armazenados.

## VPNs

A Plataforma Digibee fornece acesso aos sistemas locais de nossos clientes por meio de conexões VPN. Os gateways VPN são dimensionados em intervalos de acordo com o número de RTUs na assinatura existente. As VPNs de teste são fornecidas como uma única instância de *gateway*, enquanto as VPNs de produção são fornecidas em pares para redundância.

## Limites de uso da capacidade de RTU de testes

Os limites a seguir se aplicam a RTUs de teste sob a assinatura do cliente atual.

Limite	Categoria	Valor limite	Crescimento por	Escopo
Capacidade de processamento	Proporcional	64 MB / 20% de CPU / 10 execuções simultâneas	Cada RTU	Um único RTU
Tráfego de saída	Proporcional	10GB por RTU (mensal)	Cada RTU	Todos os RTUs no <i>realm</i>
Taxa de mensagens	Proporcional	1 MPS por RTU	Cada RTU	Todos os RTUs no <i>realm</i>
Enfileiramento	Proporcional	Máximo de 100.000 na fila	Cada RTU	RTUs implantados para um único pipeline
Retenção de registro de execução	Proporcional	7 dias ou 604.800 (registro de execução) e 3 dias ou 250 MB (logs de execução)	Cada RTU	RTUs implantados para um único pipeline
Digibee Storage	Proporcional	10MB	Cada RTU	Todos os RTUs no <i>realm</i>
Object Store	Intervalo	Veja a definição de Object Store para mais detalhes	N/A	Compartilhado entre <i>realms</i>
Gestão de Relacionamento	Proporcional	100.000 objetos (base) e 10.000 objetos (por RTU)	Cada RTU	Todos os RTUs no <i>realm</i>
VPNs	Intervalo	1 gateway de VPN (A capacidade do gateway)	Grupo de 10 RTUs	Todos os RTUs no <i>realm</i>

		VPN cresce)		
--	--	-------------	--	--

## Limites de uso da capacidade de RTU de produção

Limite	Categoria	Valor limite	Crescimento por	Escopo
Capacidade de processamento	Proporcional	64MB / 20% CPU / 10 concurrent executions	Cada RTU	Um único RTU
Tráfego de saída	Proporcional	1 TB por RTU (mensal)	Cada RTU	Todos os RTUs no <i>realm</i>
Taxa de mensagens	Proporcional	1 MPS por RTU	Cada RTU	Todos os RTUs no <i>realm</i>
Enfileiramento	Proporcional	Máximo de 1.000.000 de mensagens na fila	Cada RTU	RTUs implantados para um único pipeline
Retenção de registro de execução	Proporcional	30 dias ou 2.592.000 (registro de execução) e 10 dias ou 1 GB (registros de execução)	Cada RTU	RTUs implantados para um único pipeline
Digibee Storage	Proporcional	100MB	Cada RTU	Todos os RTUs no <i>realm</i>
Object Store	Intervalo	Veja a definição de Object Store para mais detalhes	Grupo de 60 RTUs	Todos os RTUs no <i>realm</i>
Gestão de Relacionamento	Proporcional	10.000.000 objetos (base) 1.000.000 objetos (por RTU)	Cada RTU	Todos os RTUs no <i>realm</i>
VPNs	Intervalo	2 gateways VPN (redundantes) (A capacidade do gateway VPN cresce)	Grupo de 10 RTUs	Todos os RTUs no <i>realm</i>



--	--	--	--	--