

Rate Limiting

Feature Description

UPDATED: 28 July 2023

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

These materials and all Progress® software products are copyrighted and all rights are reserved by Progress Software Corporation. The information in these materials is subject to change without notice, and Progress Software Corporation assumes no responsibility for any errors that may appear therein. The references in these materials to specific platforms supported are subject to change.

#1 Load Balancer in Price/Performance, 360 Central, 360 Vision, Chef, Chef (and design), Chef Habitat, Chef Infra, Code Can (and design), Compliance at Velocity, Corticon, Corticon.js, DataDirect (and design), DataDirect Cloud, DataDirect Connect, DataDirect Connect64, DataDirect XML Converters, DataDirect XQuery, DataRPM, Defrag This, Deliver More Than Expected, DevReach (and design), Driving Network Visibility, Flowmon, Inspec, Ipswitch, iMacros, K (stylized), Kemp, Kemp (and design), Kendo UI, Kinvey, LoadMaster, MessageWay, MOVEit, NativeChat, OpenEdge, Powered by Chef, Powered by Progress, Progress, Progress Software Developers Network, SequeLink, Sitefinity (and Design), Sitefinity, Sitefinity (and design), Sitefinity Insight, SpeedScript, Stylized Design (Arrow/3D Box logo), Stylized Design (C Chef logo), Stylized Design of Samurai, TeamPulse, Telerik, Telerik (and design), Test Studio, WebSpeed, WhatsConfigured, WhatsConnected, WhatsUp, and WS_FTP are registered trademarks of Progress Software Corporation or one of its affiliates or subsidiaries in the U.S. and/or other countries.

Analytics360, AppServer, BusinessEdge, Chef Automate, Chef Compliance, Chef Desktop, Chef Workstation, Corticon Rules, Data Access, DataDirect Autonomous REST Connector, DataDirect Spy, DevCraft, Fiddler, Fiddler Classic, Fiddler Everywhere, Fiddler Jam, FiddlerCap, FiddlerCore, FiddlerScript, Hybrid Data Pipeline, iMail, InstaRelinker, JustAssembly, JustDecompile, JustMock, KendoReact, OpenAccess, PASOE, Pro2, ProDataSet, Progress Results, Progress Software, ProVision, PSE Pro, Push Jobs, SafeSpaceVR, Sitefinity Cloud, Sitefinity CMS, Sitefinity Digital Experience Cloud, Sitefinity Feather, Sitefinity Thunder, SmartBrowser, SmartComponent, SmartDataBrowser, SmartDataObjects, SmartDataView, SmartDialog, SmartFolder, SmartFrame, SmartObjects, SmartPanel, SmartQuery, SmartViewer, SmartWindow, Supermarket, SupportLink, Unite UX, and WebClient are trademarks or service marks of Progress Software Corporation and/or its subsidiaries or affiliates in the U.S. and other countries. Java is a registered trademark of Oracle and/or its affiliates. Any other marks contained herein may be trademarks of their respective owners.

Please refer to the NOTICE.txt or Release Notes – Third-Party Acknowledgements file applicable to a particular Progress product/hosted service offering release for any related required third-party acknowledgements.

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

Table of Contents

1 Introduction **4**

2 Limiting UI Options **5**

2.1 Global Limits 5

2.2 Limiter Options 5

2.3 Client Limiting 6

2.3.1 Maximum Client Concurrent Connection Limit 6

2.3.2 Client Connections/sec Limit 7

2.3.3 Client Requests/sec Limit 7

2.3.4 URL Based Limiting 8

2.4 Client Limit Statistics 11

3 Rate Limiting Log Information **12**

4 RESTful API Details **13**

4.1 Maximum Client Concurrent Connection Limit 14

4.2 Client CPS Limit 14

4.3 Legacy Client CPS Limit Commands 15

4.4 Client RPS Limit 15

4.5 URL-Based Limiting Rules 15

4.6 Statistics 16

Last Updated Date **17**

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

1 Introduction

In LoadMaster firmware version 7.2.52, a new **QoS/Limiting** feature was introduced. The terms Quality of Service (QoS) and limiting are used interchangeably. Throughout the remainder of this document, this feature is referred to as limiting. This is a system-level limit/rate controller. It tracks ingress activity. The purpose of the limiting feature is to protect the machine as a whole. Rate limiting can guard against certain types of attacks, for example Distributed Denial of Service (DDoS) or brute-force password-guessing attacks. You can also use rate limiting to protect servers from being overwhelmed by too many requests at once.

An example scenario may be that a machine becomes resource-saturated, for example, 100% CPU utilization at 1,000 Connections Per Second (CPS) and 10,000 Requests Per Second (RPS). You may never want a machine to saturate. With the limiting feature in the LoadMaster, you can apply a system-level controller to cap or curtail levels of ingress traffic to the LoadMaster (for example, 800 CPS and 8,000 RPS).

You can configure:

- Max connections (the maximum number of established connections)
- Connections Per Second (CPS) rate
- Requests Per Second (RPS) rate

A log is generated every five seconds (this is configurable and is off by default) to include the following information:

- Current active connections
- Current CPS
- Current RPS
- Current CPS being rate-controlled (that is, the number being rejected)
- Current RPS being rate controlled (that is, the number being rejected)

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

2 Limiting UI Options

This section describes the limiting options available in the LoadMaster User Interface (UI). To access the limiting screen in the UI, go to **System Configuration > QoS/Limiting**.

2.1 Global Limits

In the **Global Limits** section, you can configure the following options:

- **Maximum Concurrent Connections:** Limit the maximum number of simultaneous connections (combined total of TCP and UDP connections) allowed to the LoadMaster. Setting the limit to 0 disables this option. Valid values are 0 - 100000000.

  > The maximum values are based on the hardware or Virtual LoadMaster that is in use and may vary per model.

- **Global Connections/s Limit:** Limit the maximum number of connection attempts (per second). Setting the limit to 0 disables this option. Valid values are 0 - 1000000.
- **Global HTTP Requests/s Limit:** Limit the maximum number of HTTP request attempts (per second). This has no effect on non-HTTP traffic. Setting the limit to 0 disables this option. Valid values are 0 - 1000000.

  > The **Global Limits** take precedence over the other limits configured. For example, if you set the **Client Concurrent Connection Limit** to **5000** but the global **Maximum Concurrent Connections** limit is set to **50**, then 50 is the limit that is enforced.

  > If the total number of connections from all clients exceed the global limit, they will be dropped.

2.2 Limiter Options

In the **Limiter Options** section, you can configure the following options:

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

- **Error Responses:** By default, the LoadMaster simply drops any connections when the RPS limit is reached. The system can send a 429 or 503 HTTP error response instead (followed by a close) if you select the appropriate option in this drop-down list.
- **Fail on RS/Sub-VS Rate Limiting:** If rate limiting is activated for a Real Server (RS) or a SubVS, the LoadMaster normally tries to select a different RS/SubVS to use for the connection. Enabling this check box forces the request to fail if the RS that was selected (for example, by persistence) was rate limited. An error response is sent back if one is selected in the **Error Responses** drop-down list.
- **Generate Limiter Statistics:** Enabling this option generates a global summary syslog message every five seconds containing the current state of the limiting subsystem.

  > This option is disabled by default. Depending upon your client limiting configuration, this can generate a lot of log messages which could be resource intensive.

- **Client Message Repeat Delay:** Set the minimum time after a client is no longer limited before a new message is generated. If a client generates a message and continues to be blocked for continuously hitting the limit, no new message is generated. Only if the client goes quiet for the delay period will a new message be generated. Valid values range from 10 - 86400 seconds. The default value is 60 seconds.

2.3 Client Limiting

Refer to the sections below for details on the client limiting options.

2.3.1 Maximum Client Concurrent Connection Limit

In the **Maximum Client Concurrent Connection Limit** section, you must configure the global **Client Concurrent Connection Limit** before you get options to configure concurrent connection limits for particular addresses or networks. The **Client Concurrent Connection Limit** limits the default maximum number of concurrent connection attempts (per second) from a specific host. Setting the limit to 0 disables this option. Valid values range from 0 - 1000000.

When you set a **Client Concurrent Connection Limit**, each client has this limit unless you have a specific entry for that client. If there is a specific limit entry for a client, the client-specific limit is applied. The options allow you to specify addresses or networks with particular limits for the concurrent connection attempts (per second) from that specific host/network. If you specify a subnet, all clients in the subnet get the same limit.

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

> The global **Maximum Concurrent Connection** value takes precedence over the client concurrent connection limits. If you try to set a client concurrent connection limit to a value greater than what is currently configured as the Maximum Concurrent Connections limit, you will get an error message.

> If you attempt to set a new specific concurrent connection limit for a particular address or network that has a limit that is greater than the **Client Concurrent Connection Limit**, you will get a warning message and will be asked if you want to continue and confirm the change.

2.3.2 Client Connections/sec Limit

In the **Client Connections/sec Limit** section, you must configure the global **Client Connection Limit** before you get options to configure the CPS limits for particular addresses or networks. The **Client Connection Limit** limits the default maximum number of connection attempts (per second) from a specific host. Setting the limit to 0 disables this option. Valid values range from 0 - 1000000.

When you set a **Client Connection Limit**, each client has this limit unless you have a specific entry for that client. If there is a specific limit entry for a client, the client-specific limit is applied. The options allow you to specify addresses or networks with particular limits for connection attempts (per second) from that specific host/network. If you specify a subnet, all clients in the subnet get the same limit. When there are multiple subnets, the lower limit applies.

> The **Global Connections/s Limit** value takes precedence over the client CPS limits. If you try to set a client CPS limit to a value greater than what is currently configured as the **Global Connections/s Limit**, you will get an error message.

> If you attempt to set a new specific CPS limit for a particular address or network that has a limit that is greater than the **Client Connection Limit**, you will get a warning message and will be asked if you want to continue and confirm the change.

2.3.3 Client Requests/sec Limit

In the **Client RPS Limit** section, you must configure the global **Client HTTP Request Limit** before you get options to configure the RPS limits for particular addresses or networks. The **Client HTTP**

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

Request Limit limits the default maximum number of HTTP request attempts (per second) from a specific host. This has no effect on non-HTTP traffic. Setting the limit to 0 disables this option. Valid values range from 0 - 1000000.

When you set a **Client HTTP Request Limit**, each client has this limit unless you have a specific entry for that client. If there is a specific limit entry for a client, the client-specific limit is applied. The options allow you to specify addresses or networks with particular limits for HTTP request attempts (per second) from that specific host/network. If you specify a subnet, all clients in the subnet get the same limit. When there are multiple subnets, the lower limit applies.

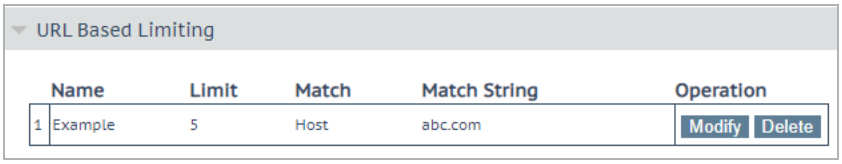
> The **Global HTTP Requests/s Limit** value takes precedence over the client RPS limits. If you try to set a client RPS limit to a value greater than what is currently configured as the **Global HTTP Requests/s Limit**, you will get an error message.

> If you attempt to set a new specific RPS limit for a particular address or network that has a limit that is greater than the **Client HTTP Request Limit**, you will get a warning message and will be asked if you want to continue and confirm the change.

2.3.4 URL Based Limiting

The **URL Based Limiting** is based on options in a HTTP request. A request consists of a URL, Method, and request headers. **Host** and **User-Agent** are request headers. The LoadMaster URL-based limiting rules inspect based on what is selected in the **Match** drop-down list (**Request URL**, **Host**, **User Agent**, **Method**, **!Request URL**, **!Host**, **!User Agent**, or **!Method**. If the limit is hit the LoadMaster sends a response code (as set in the **Error Responses** drop-down list in the **Limiter Options** section).

> The values with an exclamation mark (**!**) before them matches the inverse, for example, not a specific request or not a specific user agent.



© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

The above screenshot shows a simple example. If a request comes into the LoadMaster with a host header of **abc.com** this rule gets triggered and if the requests per second is greater than the limit set on the rule, the LoadMaster limits the request and sends out the response depending on what is selected in the **Error Responses** drop-down list.

Here is a further breakdown of this example:

- A rule exists for a host **abc.com** with a **Limit** of **5** RPS
- The **Error Responses** drop-down list is set to **Send 429 Too Many Requests**
- Requests or traffic with the specified host header is hitting the RPS limit (sending 10 RPS)
- For the requests breaching the limit, a **429 Rate Limited Rate Limit exceeded** response is sent

The request should have the **host** header as **abc.com**, for example:

```
GET /a.html HTTP/1.1
Accept: */*
Accept-Encoding: gzip, deflate
Connection: keep-alive
host: abc.com
```

Here is the example response when the limit is not hit:

```
HTTP/1.1 200 OK
Date: Tue, 08 Sep 2020 15:15:33 GMT
Server: Apache/2.4.7 (Ubuntu)
Last-Modified: Fri, 15 Feb 2019 07:40:17 GMT
ETag: "1d-581e9e2e8d033"
Accept-Ranges: bytes
Content-Length: 29
Accept: */*
User-Agent: qa-agen
Accept-Encoding: gzip, deflate
Connection: keep-alive, Keep-Alive
host: abc.com
Via: 1.1 172.16.178.55:80
X-Forwarded-For: 172.16.128.217
```


© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

```
X-Forwarded-For-Port: 54385
MyHeader1: D=138 t=1599578133851755
Keep-Alive: timeout=150, max=100
Content-Type: text/html
```

Here is the example response when the RPS limit is hit:

```
HTTP/1.1 429 Rate Limited
Date: Tue, 08 Sep 2020 15:15:33 GMT
X-Frame-Options: SAMEORIGIN
X-XSS-Protection: 1; mode=block
X-Content-Type-Options: nosniff
Connection: close
Content-Length: 84
Content-Type: text/html
<html><head><title>429 Rate Limited</title></head><body>Rate limit exceeded</body>
```

The **User-Agent** header works similarly to the **Host** example provided above.

In the **URL Based Limiting** section, you can configure the following options for a specific URL-based limiting rule:

- **Name:** The name of the new request limit. This must be unique, alpha-numeric (underscores are also allowed) and it must not start with a number.
- **Limit:** Limit the number of attempts (per second) to a specific request/URL. Valid values range from 0 - 1000000. Setting the value to 0 disables the rule (but does not delete it). This can be useful when testing, for example, if a rule has a limit of 0 it does not incur an performance impact on the system.
- **Match:** The request field/URL to match. This drop-down list contains the following values:
    - **Request URL**
    - **Host**
    - **User Agent**
    - **!Request URL**

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

- **!Host**
- **!User Agent**

- **Match String:** The pattern (regular expression) to use to match the request field/URL.

When processing HTTP traffic (non-HTTP traffic is not affected), the URL is matched against the set of rules that contain regular expressions. Each rule has a limit associated with it. If the number of requests per second exceeds the specified limit, the request is blocked and the connection is closed (an error response can be sent if an appropriate selection is made in the **Error Responses** drop-down list).

If a specific request could match more than one rule, the limit is applied to the first rule that matches in the list. You can change the order of the rules using the **Move** option.

You can also modify or delete any existing rules.

2.4 Client Limit Statistics

You can view statistics relating to the client limits by going to **Statistics > Real Time Statistics > Client Limits**. Statistics are updated every 10 seconds. The **Client Limits** button is only displayed if there is at least one client limit enabled in the **System Configuration > QoS/Limiting** screen. Statistics are only generated if the **Generate Limiter Statistics** check box is enabled in **System Configuration > QoS/Limiting**. There are buttons on the right of the **Client Limits** statistics screen where you can select different pages for **Connections/s**, **HTTP Requests/s**, **Total Connections**, and **Bandwidth Usage**.

> These buttons are only displayed if the corresponding client limits are set in **System Configuration > QoS/Limiting**.

The top 10 clients are displayed for the **Last 30 seconds**, **Last 5 minutes**, and **Last 30 minutes**. There are separate columns to show the number of **Ok** and **Blocked** connections. Based on these insights, you can configure specific rate controls for specific client IP addresses.

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

3 Rate Limiting Log Information

Logging is off by default. To enable logging, select the **Generate Limiter Statistics** check box in **System Configuration > QoS/Limiting > Limiter Options**. When this is enabled, a log is generated every five seconds. The following information is recorded in the logs:

- **curconns:** The total number of active connections on the system (at this specific moment in time).
- **totconns:** The total number of attempted connections (of all time). If you reset the statistics this value is cleared.
- **totreqs:** The total number of successful requests (of all time).
- **totrulereq:** The total number of requests that have matched a rule (of all time).
- **totcblocked:** The total number of connections that have been blocked (of all time). This is client connections (no HTTP processing).
- **totrblocked:** The total number of HTTP requests that have been blocked (of all time), which were not blocked by the client connection blocking.
- **totruleblock:** The total number of URL rules that have been blocked (of all time), which were not blocked by the client connection blocking and not blocked by HTTP request blocking.
- **totmaxcblocked:** Total cumulative number of client requests blocked due to the maximum number of connections being exceeded.

> All counters are 64-bit

When the bandwidth limit is exceeded, logs similar to the ones below are generated:

```
2020-11-03T10:08:05+00:00 lb100 kernel: L7: Bandwidth limit exceeded in the last 10 seconds
2020-11-03T10:08:35+00:00 lb100 kernel: L7: Bandwidth limit exceeded in the last 60 seconds
```


© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

4 RESTful API Details

This section contains details about the limiting API commands and parameters. You can retrieve or configure each of these parameters using the **get** or **set** RESTful API commands.

Here is an example of a **get** command to retrieve the **TotalConnectionLimit** parameter value:

```
/access/get?param=TotalConnectionLimit
```

Here is an example of a set command to set the **TotalConnectionLimit** to **80000**:

```
/access/set?param=TotalConnectionLimit&value=80000
```

The following limiting parameters can be retrieved or configured using the **get** or **set** commands:

- **MaxConnsLimit:** The maximum number of simultaneous connections (TCP and UDP).
- **MaxCPSLimit:** The global connection limit (per second).
- **MaxRPSLimit:** The global request limit (per second).
- **MaxBandwidthLimit:** The global bandwidth limit (kilobits per second)
- **SendRateLimitError:** This parameter accepts the following values:
    - 0 - no error response (the connection is simply dropped)
    - 1 - **Send 429 Too Many Requests** error response
    - 2 - **Send 503 Service Unavailable** error response
- **RateLimitFail:** Fail on rate limit. This parameter accepts the following values:
    - 0 - disabled (the LoadMaster attempts to select a different RS or SubVS to use for the connection)
    - 1 - enabled (forces an error)
- **LimitLogging:** Generate a summary log entry every 5 seconds. This parameter accepts the following values:
    - 0 - disabled
    - 1 - enabled

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

- **ClientRepeatDelay:** Set the minimum time after a client is no longer limited before a new message is generated. If a client generates a message and continues to be blocked for continuously hitting the limit, no new message is generated. Only if the client goes quiet for the delay period will a new message be generated. Valid values range from 10 - 86400 seconds.
- **ClientMaxConnsLimit:** This limits the default maximum number of concurrent connection attempts (per second) from a specific host. Setting the limit to 0 disables this option. Valid values range from 0 - 1000000.
- **ClientCPSLimit:** The global client connection limit.
- **ClientRPSLimit:** The global client request limit.
- **ClientMaxBandwidthLimit:** The global client maximum bandwidth limit.

For further details on the RESTful API in general, refer to the Long Term Support (LTS) RESTful API Interface Description.

4.1 Maximum Client Concurrent Connection Limit

To list the existing client concurrent connection limits, run the **clientmaxclimitlist** command, for example:

```
/access/clientmaxclimitlist
```

To add a new client concurrent connection limit, run the **clientmaxclimitadd** command, for example:

```
/access/clientmaxclimitadd?l7addr=<Address>&l7limit=<Limit>
```

To delete an existing client concurrent connection limit, run the **clientmaxclimitdel** command, for example:

```
/access/clientmaxclimitdel?l7addr=<Address>
```

4.2 Client CPS Limit

To list the existing CPS limits, run the **clientcpslimitlist** command, for example:

```
/access/clientcpslimitlist
```

To add a new CPS limit, run the **clientcpslimitadd** command, for example:

```
/access/clientcpslimitadd?l7addr=<Address>&l7limit=<Limit>
```

To delete an existing CPS limit, run the **clientcpslimitdel** command, for example:

```
/access/clientcpslimitdel?l7addr=<Address>
```


© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

4.3 Legacy Client CPS Limit Commands

Before the rate limiting functionality was improved in LoadMaster version 7.2.52, you could run the following commands relating to client CPS limits.

> Although these commands still work - Kemp recommends using the new commands outlined in the **Client CPS Limit** section above.

Client limiting can be used to limit the default maximum number of connection attempts (per second) from a specified host. The limit can be set using the **set** parameter, for example:

```
https://<LoadMasterIPAddress>/access/set?param=l7limitinput&value=25
```

Setting the limit to zero disables the option.

A number of addresses or networks can be specified to be limited. To add an address, run the command below:

```
https://<LoadMasterIPAddress>/access/afeclientlimitadd?l7addr=<L7Address>&l7limit=<L7Limit>
```

To delete an address, run the command below:

```
https://<LoadMasterIPAddress>/access/afeclientlimitdel?l7addr=<L7Address>
```

To list the addresses and their limits, run the command below:

```
https://<LoadMasterIPAddress>/access/afeclientlimitlist?
```

4.4 Client RPS Limit

To list the existing RPS limits, run the **clientrpslimitlist** command, for example:

```
/access/clientrpslimitlist
```

To add a new RPS limit, run the **clientrpslimitadd** command, for example:

```
/access/clientrpslimitadd?l7addr=<Address>&l7limit=<Limit>
```

To delete an existing RPS limit, run the **clientrpslimitdel** command, for example:

```
/access/clientrpslimitdel?l7addr=<Address>
```

4.5 URL-Based Limiting Rules

You can list the existing URL-based limiting rules by running the **listlimitrules** command, for example:

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

/access/listlimitrules

You can add a new URL-based limiting rule by running the **addlimitrule** command, for example:

/access/addlimitrule?name=ExampleRule&pattern=/test/a.html&limit=5&match=0

You can modify an existing URL-based limiting rule by running the **modlimitrule** command, for example:

/access/modlimitrule?name=ExampleRule&pattern=/test/b.html&limit=5&match=0

Valid values for the **match** parameter are as follows:

- 0 - Request
- 1 - Host
- 2 - User Agent
- 64 - !Request
- 65 - !Host
- 66 - !UserAgent

> The values with an exclamation mark (**!**) before them matches the inverse, for example, not a specific request or not a specific user agent.

You can delete an existing URL-based limiting rule by running the **dellimitrule** command, for example:

/access/dellimitrule?name=ExampleRule

You can move the position of an existing URL-based limiting rule by running the **movelimitrule** command, for example:

/access/movelimitrule?name=ExampleRule3&position=1

Setting the **position** parameter to a value larger than the size of the list will move the rule to the end of the list.

4.6 Statistics

Global and Virtual Service statistic information (including some related to limiting) is available in the **stats** RESTful API command:

/access/stats

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.

Last Updated Date

This document was last updated on 28 July 2023.

© 2022 Progress Software Corporation and/or one of its subsidiaries or affiliates. All rights reserved.