

RATELIMITING

The **Rate Limiting Policy** governs the number of times requests made to an API within a specified time period. The requests can be filtered by user, application or API and can set the level of granularity for the time period to second, minute, hour, day, month or year. The intended use of this policy is for fine grained processing, for example limiting the requests to 10 per second.

Configuration Options

Rate Limiting

Basic Configuration

I want to limit request rates to

Number

 requests per

Granularity

 per

Period

Advanced Configuration

Use the advanced options to configure the rate-limiting response headers. These headers convey useful information to clients, such as the limits imposed and when to reset the rate period. Override the default header names by entering your own.

Limit Response Header

Remaining Response Header

Reset Response Header

Cancel

Back

Next

On this Page:

- [Configuration Options](#)
 - [Basic Configuration](#)
 - [Advanced Configuration](#)

Related Pages:

- [Policies](#)
 - [Attaching Policies](#)

Basic Configuration

Option	Description	Possible Values	Default
Number	Number of requests that must be received before the policy will trigger.	The maximum value you can specify is 9007199254740991 (2 ⁵³ - 1).	-
Granularity	The element for which the requests are counted.	<ul style="list-style-type: none">• Client• API• User• IP Address	None
Period	The time period over which the policy is applied.	<ul style="list-style-type: none">• Second• Minute• Hour• Day• Month• Year	None

Advanced Configuration

Use the advanced options to configure the rate-limiting response headers. These headers convey useful information to clients, such as the limits imposed and when to reset the rate period. Override the default header names by entering your own.

Option	Description	Possible Values	Default
Limit Response Header	Optional. HTTP response header the API Management will use to store the limit being applied.	a string	X-RateLimit-Limit
Remaining Response Header	Optional. HTTP response header the API Management will use to store how many requests remain before the limit is reached.	a string	X-RateLimit-Remaining
Reset Response Header	Optional. HTTP response header the API Management will use to store the number of seconds until the limit is reset.	a string	X-RateLimit-Reset