

Rate Limiting

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.

Policy Type

Rate Limiting Policy

Rate Limiting Policy Configuration

I want to limit request rates to

of requests

requests per

Granularity

per

Period

Configure the rate limiting related response headers below. These headers will convey useful information to clients such as imposed limits and when to reset the rate period. Override the default header names by supplying your own in the fields below.

Limit Response Header

X-RateLimit-Limit

Remaining Response Header

X-RateLimit-Remaining

Reset Response Header

X-RateLimit-Reset

Add Policy

Cancel

On this Page:

- [Configuration Options](#)

Related Pages:

- [Policies](#)
 - [Assigning Policies](#)
 - [Policy Configuration](#)

Configuration Options

Option	Type	Description	Possible Values	Default
# of requests	Integer	Number of requests that must be received before the policy will trigger.	The maximum value you can specify is 9007199254740991 (2 ⁵³ - 1).	-
Granularity	Enum	The element for which the requests are counted.	Client User API IP Address	None
Period	Enum	The time period over which the policy is applied.	Second Minute Hour Day Month Year	None
Limit Response Header	String	Optional. HTTP response header the API Management will use to store the limit being applied.	-	X-RateLimit-Limit
Remaining Response Header	String	Optional. HTTP response header the API Management will use to store how many requests remain before the limit is reached.	-	X-RateLimit-Remaining
Reset Response Header	String	Optional. HTTP response header the API Management will use to store the number of seconds until the limit is reset.	-	X-RateLimit-Reset