

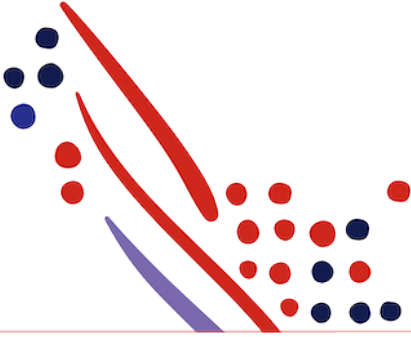


Guide

# Web API Gateway Rate Limit Policy

Published on  
Jul 20, 2021 4:20PM

Last modified  
Aug 04, 2021 3:49PM



## ADP Copyright Information

ADP, the ADP logo, and Always Designing for People are trademarks of ADP, Inc.

Windows is a registered trademark of the Microsoft Corporation.

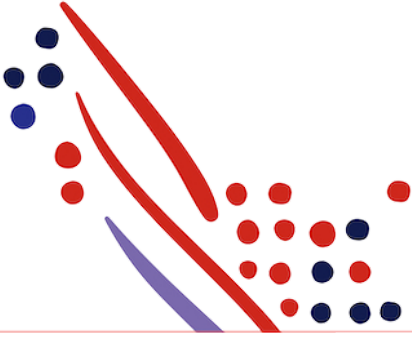
All other trademarks are the property of their respective owners.

Copyright © 2021 ADP, Inc. ADP Proprietary and Confidential - All Rights Reserved. These materials may not be reproduced in any format without the express written permission of ADP, Inc.

These materials may not be reproduced in any format without the express written permission of ADP, Inc. ADP provides this publication "as is" without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability or fitness for a particular purpose. ADP is not responsible for any technical inaccuracies or typographical errors which may be contained in this publication. Changes are periodically made to the information herein, and such changes will be incorporated in new editions of this publication. ADP may make improvements and/or changes in the product and/or the programs described in this publication.

Published on  
Jul 20, 2021 4:20PM

Last modified  
Aug 04, 2021 3:49PM



# Table of Contents

## Chapter 1

### Web API Gateway Rate Limit Policy

#### **Rate limiting**

#### **Concurrency limiting**

#### **Current Production Limits**

#### **Rate Limits Across Nodes**

# Web API Gateway Rate Limit Policy

The Marketplace Web API Gateway limits all calls to API endpoints using two types of Rate limiting: Rate Limiting and Concurrency. The information in this article is subject to change but is accurate as of 6/13/2018.

## In This Article

- Rate Limiting
- Concurrency

## Rate limiting

The API gateway limits the total number of calls to API endpoints per second or per minute and provides a means of controlling the volume of inbound API traffic that is processed by the API gateway. The rate limits are applied to each integration profile (internally known as a consumer application) and enforced per subscribed client to the associated integration profile. The specific calls per minute value is defined in the integration profile. This rate is then propagated to the CA API Gateway and stored as a per-second value (per-second = per-minute/60). When this rate is exceeded, a 429 error will be returned.

## Concurrency limiting

Any given API call will receive a response from the system, regardless of whether the initial call is successful. The API call is considered to be open during the period of time between when the API call is made and when the system responds. Concurrent calls are the number of unique open API calls or transactions through the gateway at any given time.

Concurrency throttling limits the number of such open transactions enforced per subscribed client per associated integration profile. If a client has 10 open calls at given instance, then there will be 10 concurrent calls through the API Gateway. This number may increase until the concurrency limit is reached at which point a 429 error will be returned by the API Gateway for that client.

## Current Production Limits

Tiers	Rate (calls/min)	Concurrency (# of calls)
Default	300	50
Tier 2	480	50
Tier 3	780	50

**Important:** The above rate limits are applied across four nodes in production.

## Rate Limits Across Nodes

Rate/Concurrency	Single Node	All Nodes
300/50	75/12	300/50

480/50	120/12	480/50
780/50	195/12	780/50

The single node values are important because if you reach either of the limits on a single node, you may receive a 429 response. The most common cause for the 429 response is violating the single node concurrency limit.

For example, if you have 12 open connections on the same node, the next open connection would result in a 429 response if none of the 12 open connections closed (completed) before the 13th call was opened.

Last updated: Aug 4, 2021