

What's the average response time of a Cloud API request?

Our Cloud services process requests from all products including API. Thus, we are not able to distinguish API requests from the other requests. The average response time for all types of requests is 300 ms. The response time depends on the list of factors: size of the text, language, a number of mistakes, spelling or/and grammar type.

The entry point for all requests is the AWS load balancer that distributes the amount of the request to the running instances. Besides, there is an auto-scaling mechanism that tracks the CPU load on each instance and if the CPU consumption increases on all instances above a certain level, a new instance will be launched.