



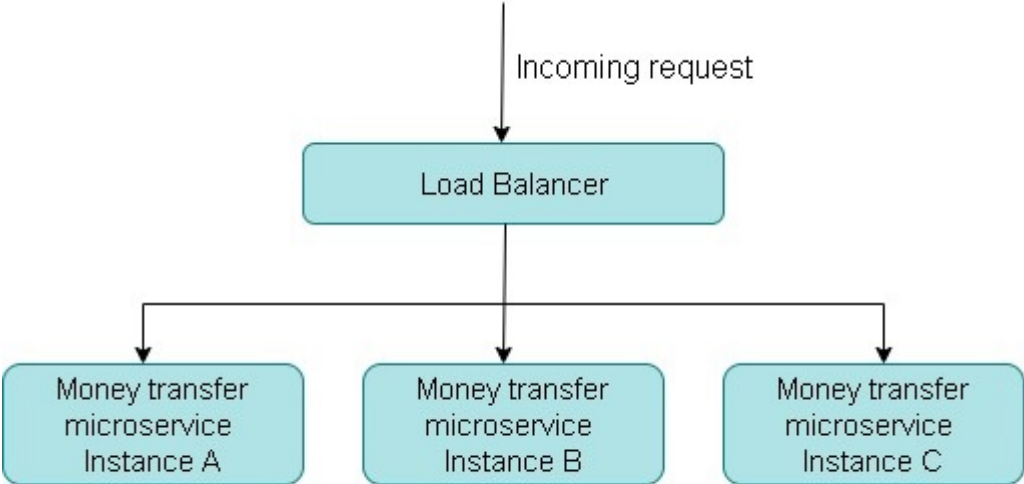
What is loadbalancing and algorithms used for loadbalancing

Description

Introduction

Load balancers in the context of microservices are used to distribute the incoming traffic to the instances that are running on various hosts, and these hosts sit behind the load balancer.

Looking at the below image you see Money transfer microservice instance is present in 3 hosts and whenever customers try to invoke the web url to transfer their money, instead of the request directly coming to a microservice it has to pass load balancer.



Some of the salient features of load balancers are as below.

- 1. Load balancers distribute calls sent to them to one or more instances based on some algorithm.
- 2. Reliability and high availability is maintained by redirecting the requests only to the server which

are available.

3. We have the ease of adding or removing the servers in the network based on the demand. E.g. you must have seen some giant eCommerce websites launching their biggest sale where most of the items are sold at big discounts. And with this it's obvious that the website traffic will also increase and this gives the flexibility to increase the number of servers running the application. In this case load balancers will take the job of distributing the traffic evenly across the servers.

We will discuss 3 common types of load balancing algorithms.

Round Robin

In this algorithm the incoming requests are distributed in round robin fashion. Let's say there are 3 instances of application running. The first request will go to the 1st instance, the second will go to the 2nd instance and the 3rd will go to the 3rd instance. If 5 other requests come in, the request will be redirected to the 1st then 2nd and likewise.

Least connection

In this algorithm the incoming request will be redirected to the server that has the least no of requests. Load balancers should know which server has the least number of requests. Load balancer has the additional overhead of getting the data about the available servers and the request count they are processing based on which it has to make the decision.

IP hashing

Load balancer will redirect the request based on the IP address of the client from where the request originated. All requests from Client A will always go to that specific server as long as the server is alive.

Category

1. Design
2. Microservices

Date Created

March 29, 2021

Author

kk-ravi144gmail-com