ALX PROJECT

Web Infrastructure Design

Task 1

Explanation of Terms

1. For every additional element, why you are adding it: we added a new server in order to be able to add a load balancer which will help to handle too much incoming traffic, and also make it possible for us to eliminate a single point of failure which occurs by having only one server.

2. What distribution algorithm your local balancer is configured with and how it works: the load balancer we used is configured with Round Robin algorithm which ensures that each server receives an equal share of incoming requests over time unless the server is down. Requests are served by the server sequentially one after another. After sending the request to the last server, it starts from the first server again. This algorithm is used when servers are of equal specifications and there are not many persistent connections.

3. Is your load-balancer enabling an Active-Active or Active-Passive setup, explain the difference between both: our load balancer enables Active-Active setup where multiple servers or nodes are actively serving clients requests simultaneously. All nodes are actively processing and responding to requests, sharing the overall load. Each node can independently handle requests and perform computations unlike Active-Passive setup where there is a primary active node that handles all client requests, while the passive node remains idle and ready to take over in case of a failure or when the primary node becomes unavailable. The Active-Active setup helps in scaling the system and handling increased traffic.

4. How a database Primary-Replica (Master-Slave) cluster works: master-slave replication enables data from one database server (the master) to be replicated to one or more other database servers (the slaves). The master logs the updates, which then ripples through the slaves. If the changes are made to the master and slaves at the same time, It is synchronous. If changes are queued up and written later, it is asynchronous. It is usually used to spread read access on multiple servers for scalability, although it can also be used for other purposes such as for failover, or analyzing data on the slave in order not to overload the master.

5. What is the difference between the primary node and the Replica node in regards to the application: a replica node is a copy of the primary node, they provide redundant copies of the application codebase to protect against hardware failure and increase capacity to serve read requests like searching or retrieving a document.