WEB INFRASTRUCTURE DESIGN

Q1. Distributed web infrastructure

You must be able to explain some specifics about this infrastructure:

1. For every additional element, explain why you are adding it.
   Added 2 more servers. To increase redundancy effectively avoiding SPOF, which would arise from using just one server.

   Added a HAproxy load balancer to reduce the workload on the initial one server. Load balancer works to eliminate SPOF, that is in case one server crashes, your website remains up, since the other remaining running servers continue to serve it.

2. What distribution algorithm your load balancer is configured with and how it works
   The Round Robin Distribution Algorithm - Sequentially each request gets assigned to each server one by one. After sending the request to the last server, it starts from the first server again.

3. Is your load-balancer enabling an Active-Active or Active-Passive setup, explain the difference between both
   Load Balancer is enabling an Active-Active Setup. This is because workloads are evenly distributed across all nodes/servers ensuring load balancing.
   The difference between this setup and an active-passive setup is that in the latter not all nodes are active, with the other node(s) on standby to take over if the active node fails, ensuring service continuity without load distribution.

4. How a database Primary-Replica (Master-Slave) cluster works
   The master/primary database serves as the primary source of data.
   It is responsible for handling write operations, meaning any data modifications (such as inserts, updates, or deletes). It holds the genuine data.

   There can be multiple slave databases connected to the master.Slaves are primarily used for queries or read operations.
   When a user requests data, the slave databases handle the read operations.
   These slave databases act as backups to the master. If the master goes down, the data remains accessible through the slaves.

5. What is the difference between the Primary node and the Replica node in regard to the application
   The primary node is the master.It handles all write operations.It is responsible for copying data to the replicas. The replica node is the secondary node. It receives copied data from the primary node. They provide read only access to the data. They act as backup, a safety net in case the primary node fails.

You must be able to explain what the issues are with this infrastructure:

1. **Where are SPOF:**
   Just one load balancer

2. **Security issues (no firewall, no HTTPS)**

   Communication between the end user and the application is over HTTP Protocol which is textual and not secure.Anyone including an attacker in the middle can view sensitive personal information such as passwords and credit card information.

   No firewall - This acts as a protective shield between the application and the external environment. Having no firewall can lead your application susceptible to threats such as malicious actors having direct access to servers and services, Denial of Service attacks(where excessive traffic which is mitigated by a firewall overwhelms your application leading to service disruptions), exposure to malware, viruses, worms, amongst other threats.

   No monitoring - The servers, the application and other components of this system are not being monitored. Monitoring helps to identify performance issues among other problems, security threats (where an alert is issued incase of unusual activity) or downtime and fix these issues as they arise before they escalate, hence preventing decreased productivity and costs that may be brought about by not capturing problems early enough.