**What are some basic concepts here?**

1. Vertical scaling: add more resources to the same server
2. Horizontal scaling: add more servers in general

What’s the difference between an **API gateway and a load balancer?**

An API gateway is focused on 𝗿𝗼𝘂𝘁𝗶𝗻𝗴 requests to the appropriate microservice, while a load balancer is focused on 𝗱𝗶𝘀𝘁𝗿𝗶𝗯𝘂𝘁𝗶𝗻𝗴 requests evenly across a group of backend servers.