Scaling

Allow more users to use the service

A single computer has limited resources, we can't serve every single user with the same computer. So we need to scales the compute in some way.

Two ways:

① Verticle Scaling - Bigger computer
② Horizontal Scaling - More computer.

- What does it mean for a system to be scalable?

① Able to handle increased load.
② If isn't too complex (idk)
③ There shouldn't be a hit in performance.

Example

Vertical	Horizontal
+ Upgrade AWS instance	+ Get more AWS instances
+ Good for small scale (smaller firms)	+ Required for large # requests. (Adds some complexity)