Why distribute a database across multiple machines?
1. Scalability: if the data volumne, read load, or write load grows bigger than a single machine can handle
2. Fault tolerance/high availability: if one machine goes down, to still have the service online
3. Latency: users worlwide -- can be serviced by a data centre closest to them

Scaling to higher load

Scaling up/vertical scaling: get a machine with 2x memory 2x compute etc

- trouble is that cost grows superlinearly

Shared nothing architecture/horizontal scaling:

- each machine or VM running the db software is a node
- each node uses own CPU RAM disk etc
- any coordination across nodes is done at the software level using a conventional network
- no special hardware needed

Replication vs Partitioning
Two common ways data is distributed across nodes
1. Replication: keep a copy of the same data on different nodes -- provides redundancy
2. Partitioning: split a big db into smaller subsets called partitions so that diff partitions are on diff nodes -- called *sharding*