Peer To Peer Networks

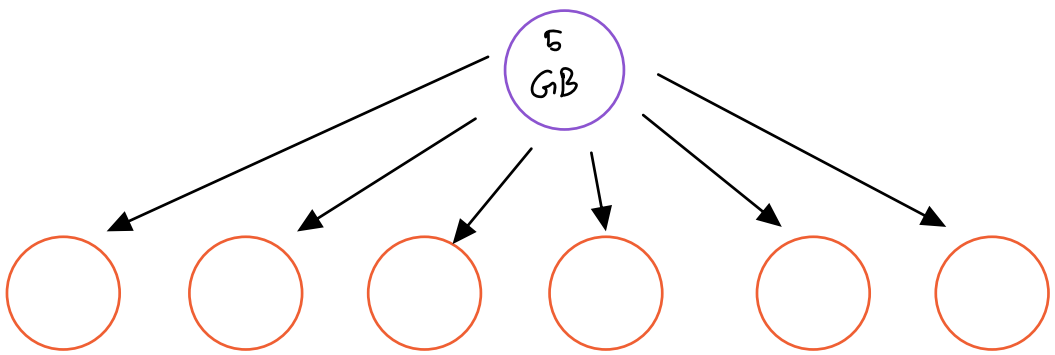
Peer to peer network:

A collection of machines referred to as peers that divide a workload amongst themself to complete it faster. Peer to peer networks are generally used in file-distribution systems

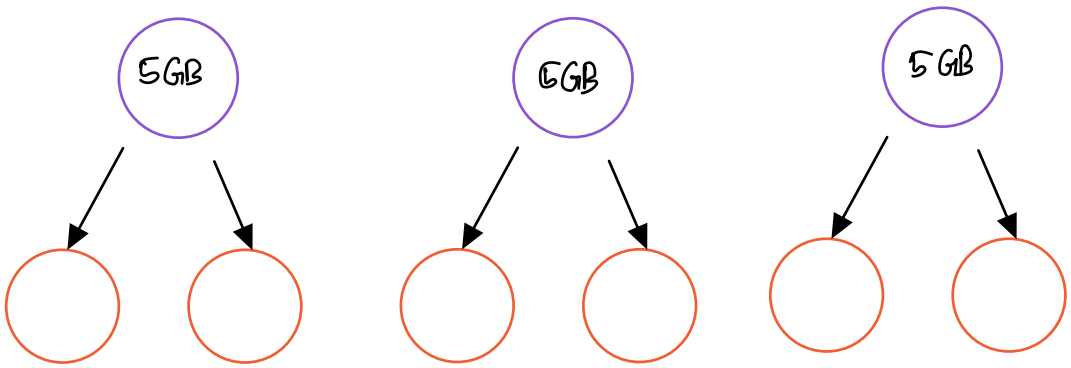
Assume for every 15 mins we have to upload/transfer a large file from a single machine (say 5GB video file) to thousand machines at once

Assume the network speed is 5 GBps. So it will take ~17 mins to transfer the 5GB video file to all the thousand machines.

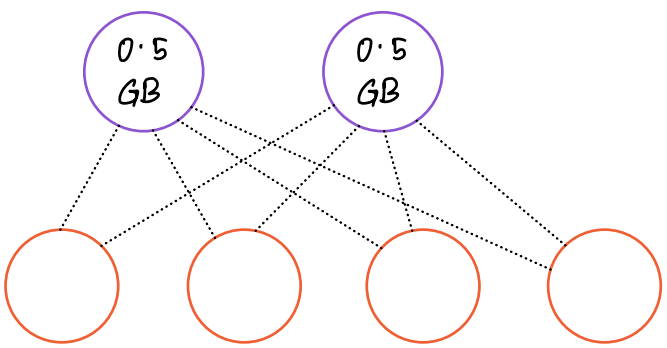
Single Node System



If replication is used (ie instead of using a single machine to transfer the 5GB video file we use 10 machines) it would still take 1.7 mins Moreover this method is more expensive.

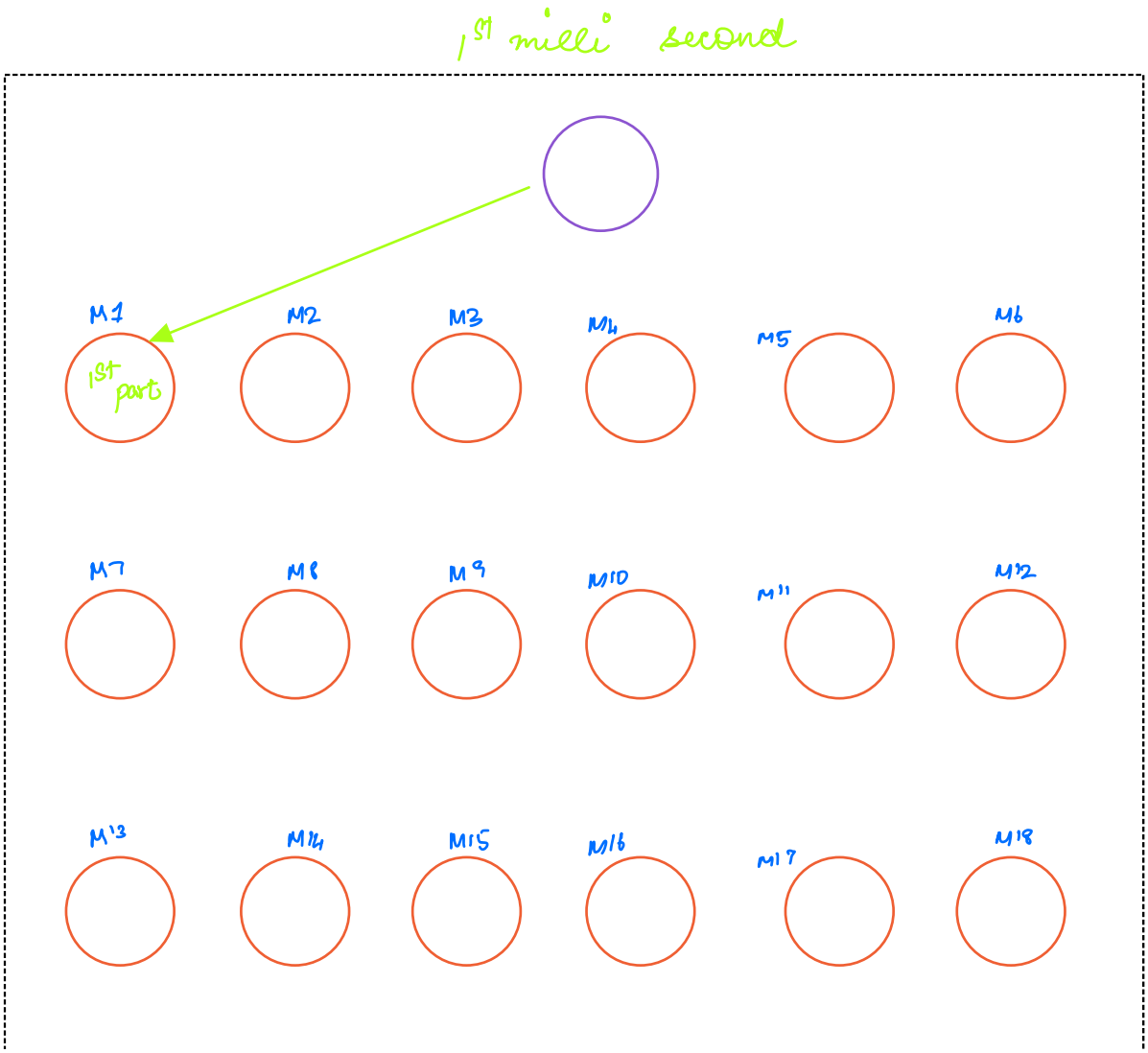


If sharding is used it would still take time greater than that for replication and lesser than the time taken in a single node system

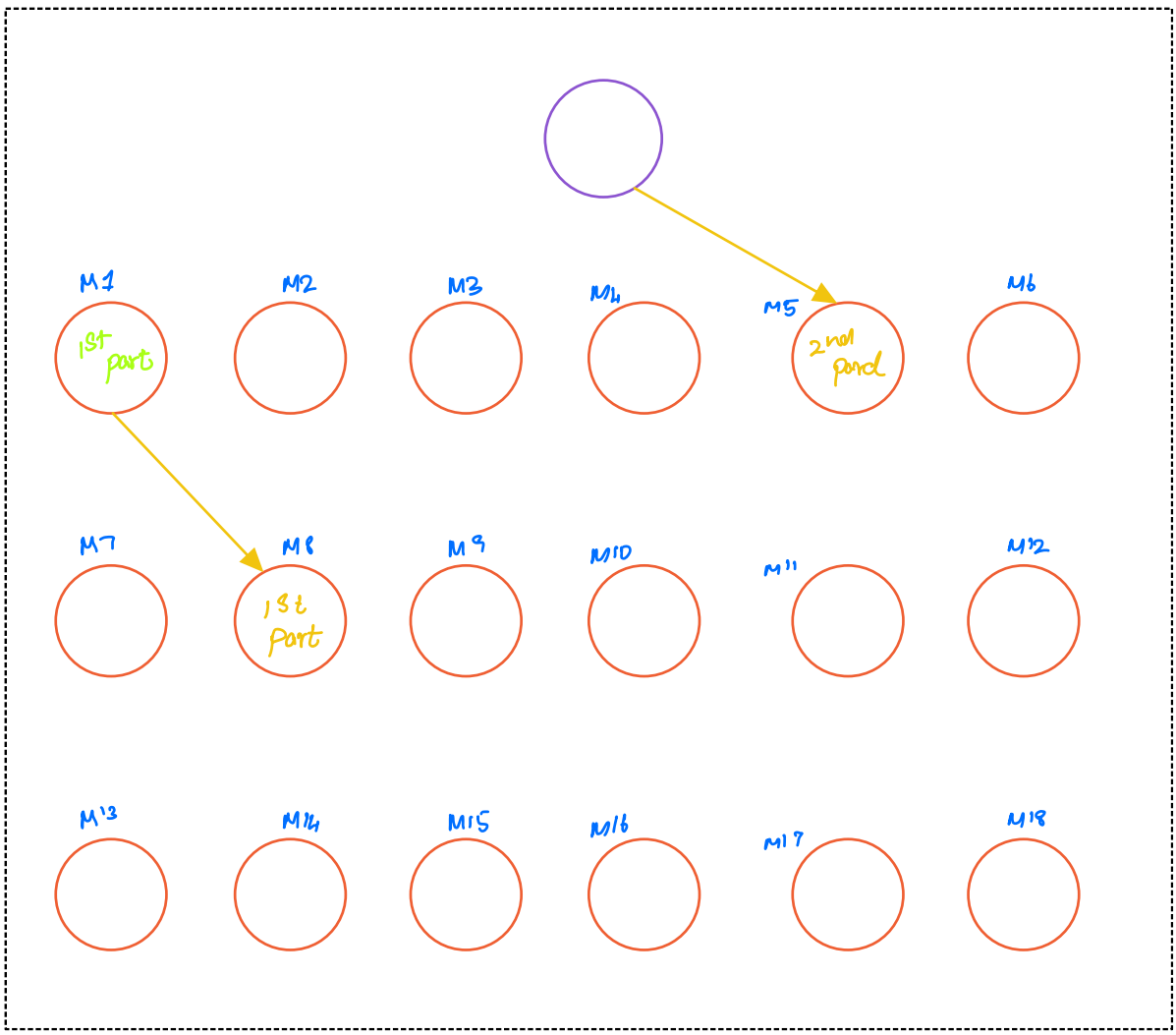


In peer to peer networks the 5 GB data gets split into 1000 parts (The size of each part is 5mb). Each part would be sent to each machine (Eg: The 1st part would be sent to 1st machine, 2nd part to 2nd machine and so on It would take 1 sec for this process to complete.

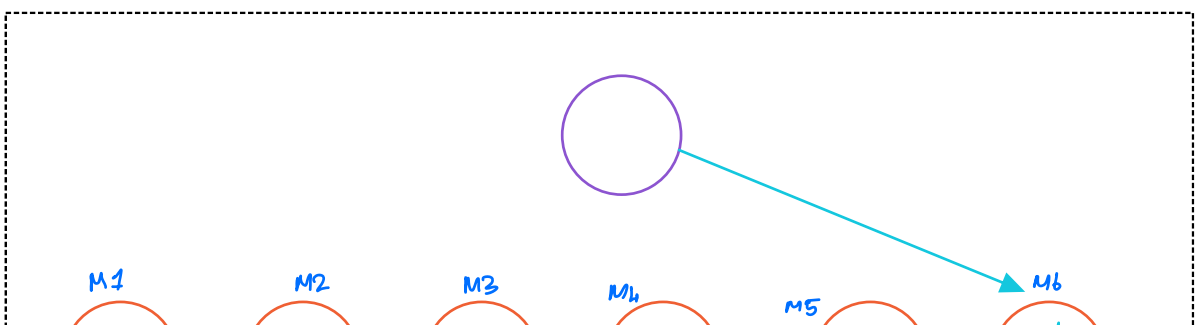
M1 → machine 1

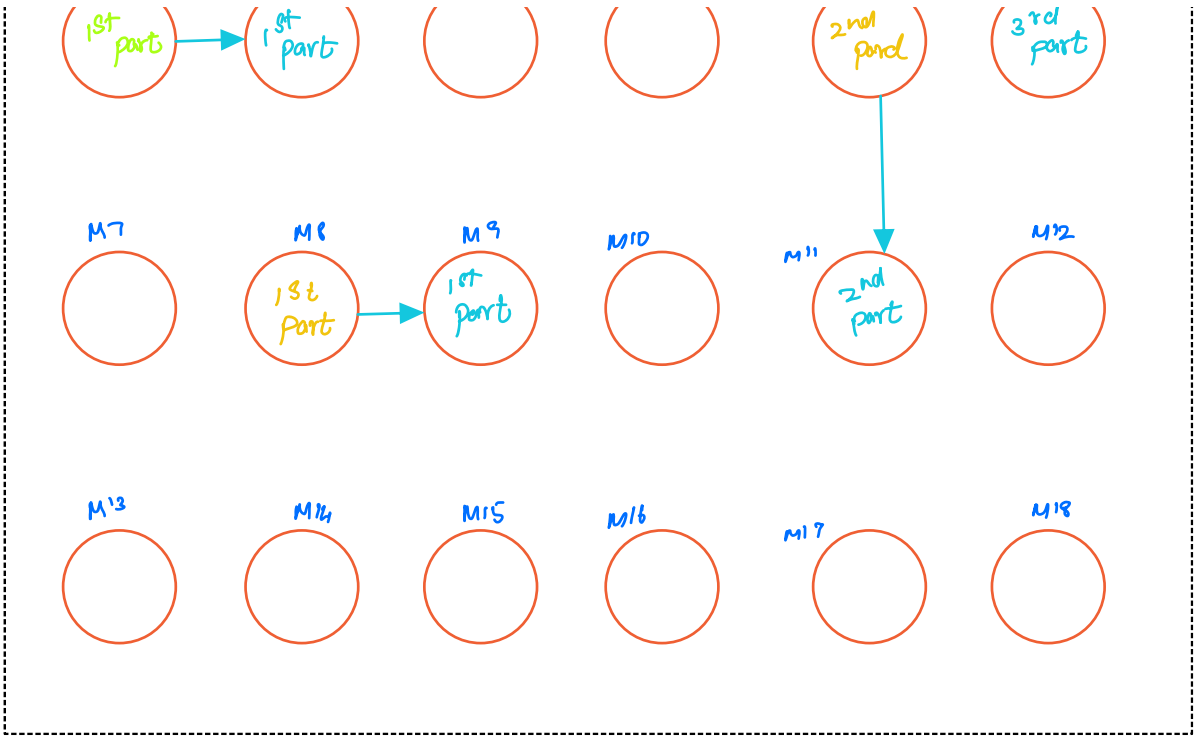


2nd milli second

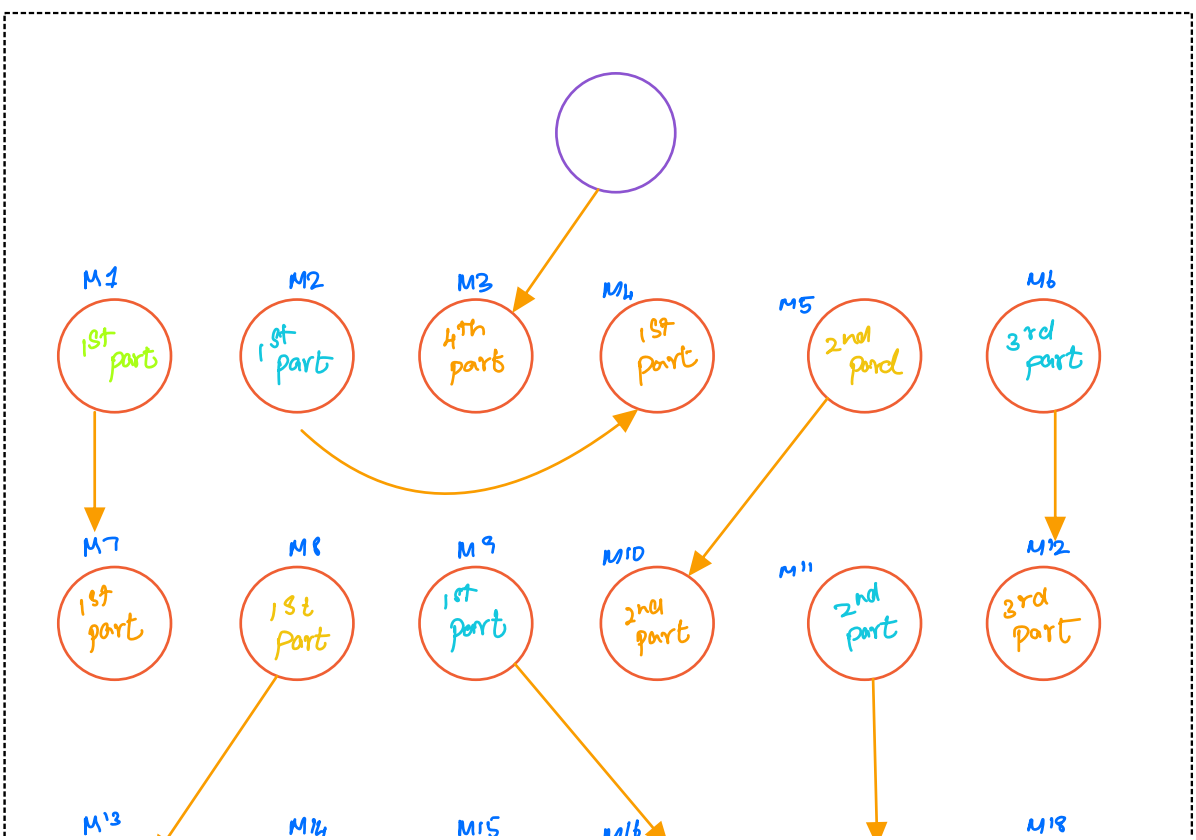


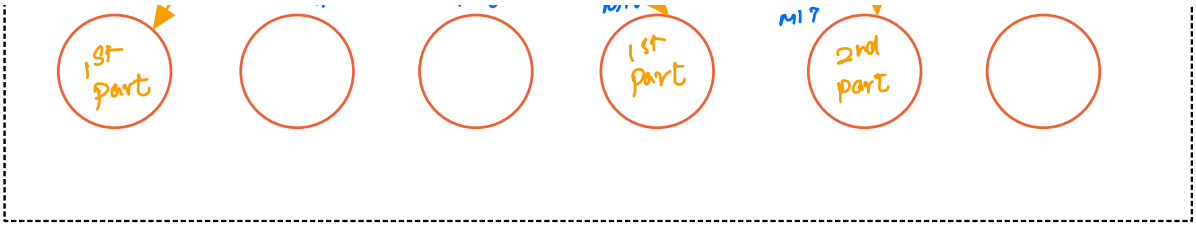
3rd milli second

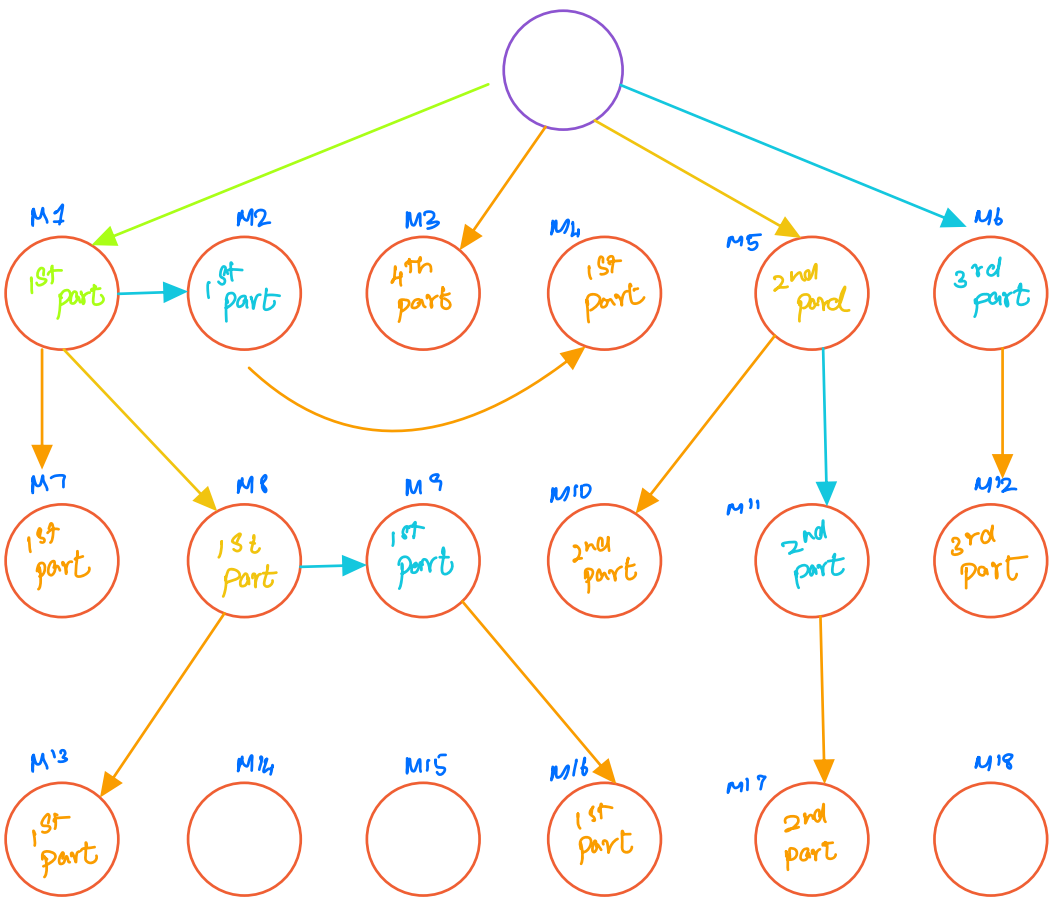




2,4th multi second







=> 1st milli second (1)

=> 2nd milli second (2)

↳ no of nodes affected

Total no of nodes affected = 15

For 0.004 seconds 75 mb of data is transfered. So it should take 4.44 mins for the file to get transfered to all the 1000 machines. But it wont take 4.44 mins. instead the entire process will take just a few seconds because as the number of seconds increases data transfered doesn't increase linearly instead it increases exponentially

Peer Discovery / Peer Selection

In the above examples the peers communicate randomly. But inorder for the peer to peer network to work properly a peer at a point should know to what peer it has to communicate next. This is done by peer

discovery / peer selection. The peer discovery is done in two ways.

① A central database known as Tracker is used. The tracker orchestrates the entire peer to peer network (ie it would tell a peer to what peer it has to communicate next)

② Gossip Protocol / Epidemic Protocol is used. Gossip protocol is used in a decentralized system that does not have any central node to keep a track of all the nodes. Gossip protocol allows state sharing in a distributed system All the peers will have a hash table that holds the data of the chunk it has received and also has data like "I have communicated with peer 4 & 5. peer 4 has chunk no. 99 and peer 5 has chunk no. 101."

This hash table is known as a DHT ⇒ Distributed Hash Table

Peer to peer networks is used in Torrents

Eg : Kraken

https://www.uber.com/en-IN/blog/introducing-kraken/ => See the graph to understand the power of P2P