

# Big Data and Distributed System

gta 3

gta 5

When the complexity or size of our data increases, we will be needing more resources.

## Resources

1. Storage  $\rightarrow$  hdd, ssd
2. Memory  $\rightarrow$  RAM
3. Performance  $\rightarrow$  cores

### Single system

#### Monolithic System



increase the resource

mono  $\rightarrow$  1

$X \rightarrow 2X \rightarrow 10X$

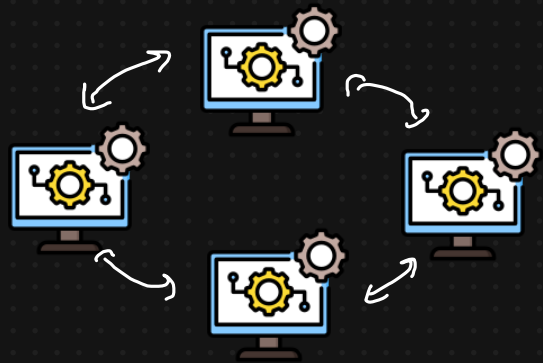
$P \rightarrow 2P \rightarrow 6P$

hardware limitation

they are not scalable

### Multiple systems

#### Distributed systems



Group of machines  $\rightarrow$  Cluster

We can add more m/c here as & when required.

### True Scaling

$X \rightarrow 2X \rightarrow 10X$

$P \rightarrow 2P \rightarrow 10P$

All good big data systems are based on distributed systems.