

PGP in Cloud Computing

Containers

- Time for provisioning of capacity=time to initialize(infra + app)
- The challenge is to minimize the time for provisioning

Class -----> Object

Images -----> Container

- Disk utilization is significantly reduced if using containers instead of VMs.
- Users interact with **Docker client** and ask for a container
Docker client sends request to the docker daemon
Docker daemon reads the Image and creates a process with a process ID. This process ID is called **container**.
[https://en.wikipedia.org/wiki/Docker_\(software\)](https://en.wikipedia.org/wiki/Docker_(software))
- We can have multiple containers bound to different ports.
- Spinning up processes (**containers**) will be faster than spinning up VMs. Time for provisioning can be significantly reduced.
- **Kubernetes** is an orchestration and container management tool.
<https://en.wikipedia.org/wiki/Kubernetes>
- Kubernetes + Docker daemon working on top of a farm of VMs which can be autoscaled is a great responsive system.
- Managed Kubernetes services are **ECS** and **EKS**.
<https://docs.aws.amazon.com/AmazonECS/latest/developerguide/Welcome.html>
- Docker daemon creating a container :
 downloads docker image >> checks the local disk first
 If it doesnt find the image there >> goes to **DOCKERHUB** or creates a private repository called the **Elastic container Registry (ECR)**
 Spins up the containers (specifying the cpu,memory)
<https://aws.amazon.com/ecr/>
- A **task** is nothing but a group of related containers or processes that exist together (started and stopped together).
- In Kubernetes, it is called a **POD**.
<https://kubernetes.io/docs/concepts/workloads/pods/pod-overview/>
- Best practice is to have 1 task per instance and scale it horizontally.