Software Consulting

DevOps Projects

Corporate Training

"UpGrade Your Business with BEST BRAINS"

Docker

Kubernetes

Rajesh G

Master Trainer & CTO,
Brain Upgrade Academy,
A division of Unisuraksha Tracking Systems Pvt Ltd

Kubernetes

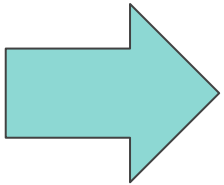
Rajesh G

CTO, Managing Partner
https://uniqps.in

Training Objectives

At the end of training,

participants should be able to



- ❏ Know Kubernetes and Be a Helmsman
- ❏ Create and run PODs
- ❏ Bundle applications & Deploy
- ❏ Service apps using Load Balancers
- ❏ Troubleshoot

Table of Contents

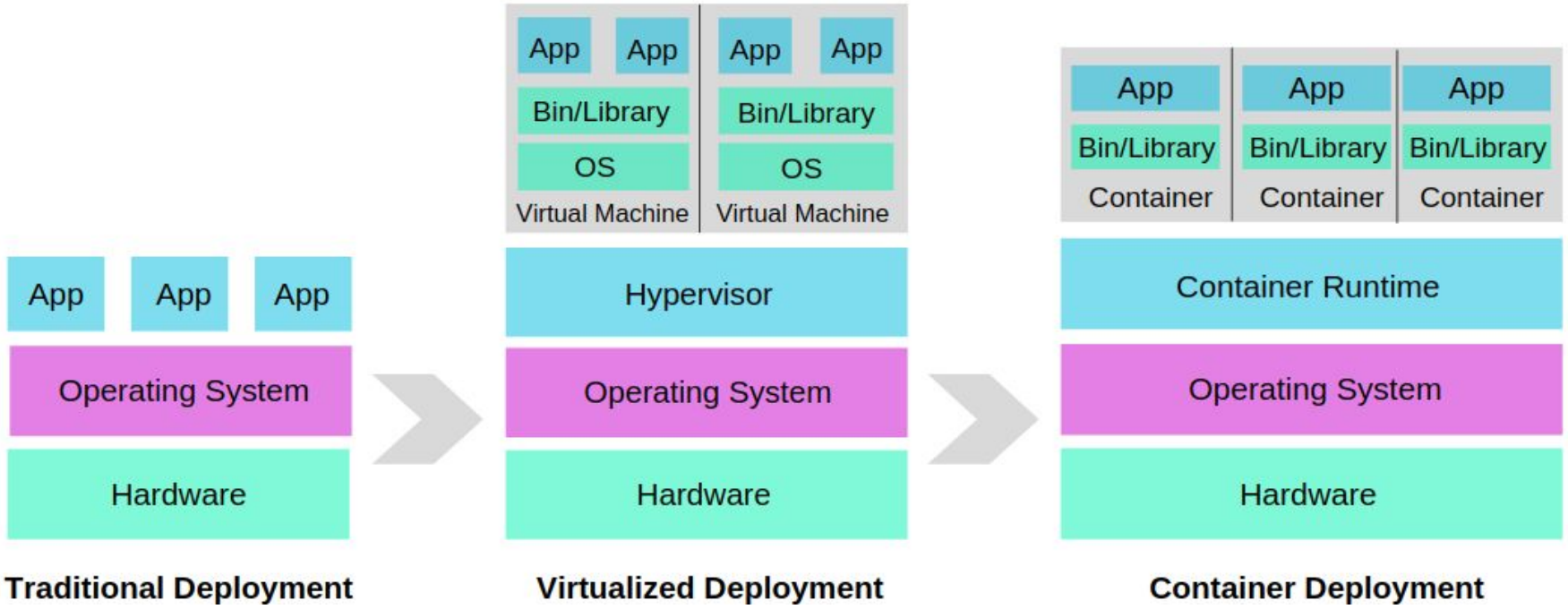
BrainUpgrade Academy

Part ONE		
Kubernetes Core	**Pod**	**Pod - Advanced**
Architecture	Overview	Patterns - Sidecar, Adapter etc
Components	Lifecycle	Jobs, CronJob
Objects	Probes, Init Containers	Labels, Selectors, Annotations
Kubectl	Topology	Demo & Practicals
Demos & Practicals	Demo & Practicals	
Part TWO		
Scalability	**Persistence**	**Configuration**
Deployments	PersistenceVolume	ConfigMaps
Rolling Updates	PersistenceVolumeClaim	Secrets
Rollbacks	Statefulset	SecurityContexts
Auto Scaling	Daemonset	Accounts
Demos & Practicals	Demos & Practicals	Demo & Practicals
Part THREE		
Scheduling	**Services & Networking**	**Monitoring & Misc**
Node Name	ClusterIP	Dashboard
Node Selector	NodePort	Logging & Debugging
Affinity	LoadBalancer	Port Forwarding
Taints & Tolerations	Ingress	Tips - CKAD & CKA (CNCF)
Demo & Practicals	Demos & Practicals	Monitoring & Debugging

Kubernetes Core

- Container Journey
- Why Kubernetes
- Architecture
- Core Components
- API Primitives
- Kubectl
- Demo
- Practicals

Container - Journey



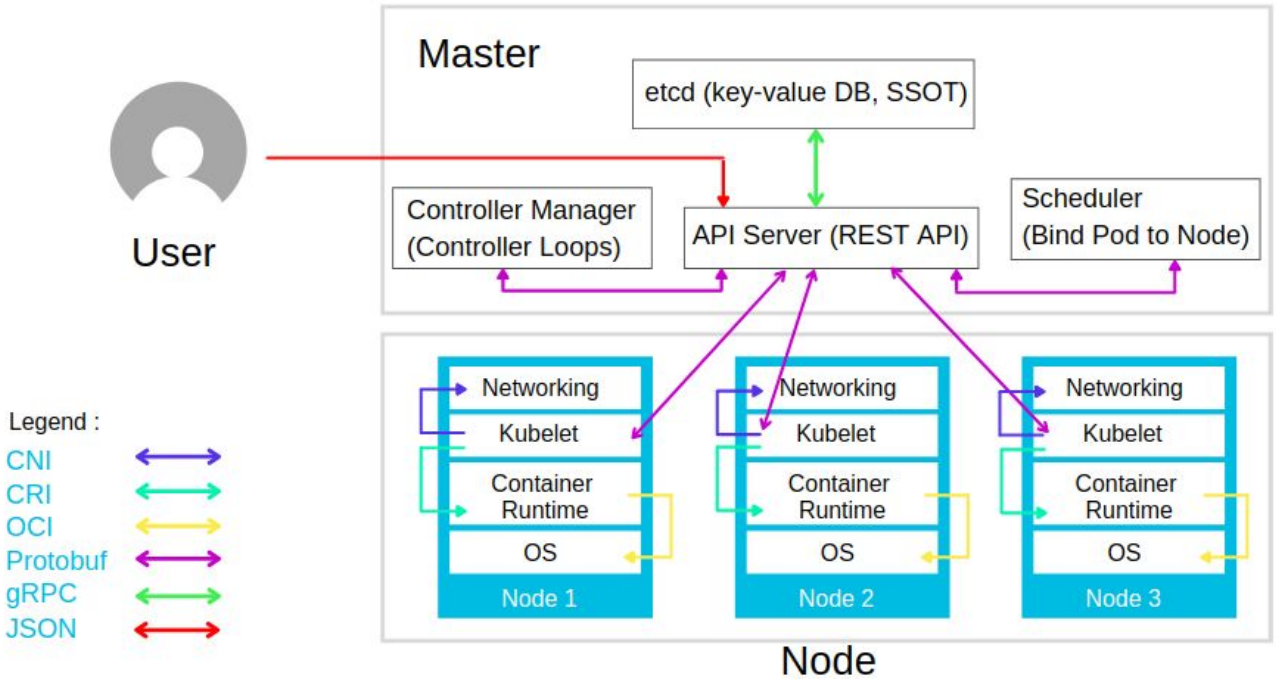
What is / why Kubernetes

Kubernetes is a portable, extensible, open-source platform for managing containerized workloads and services, that facilitates both declarative configuration and automation.

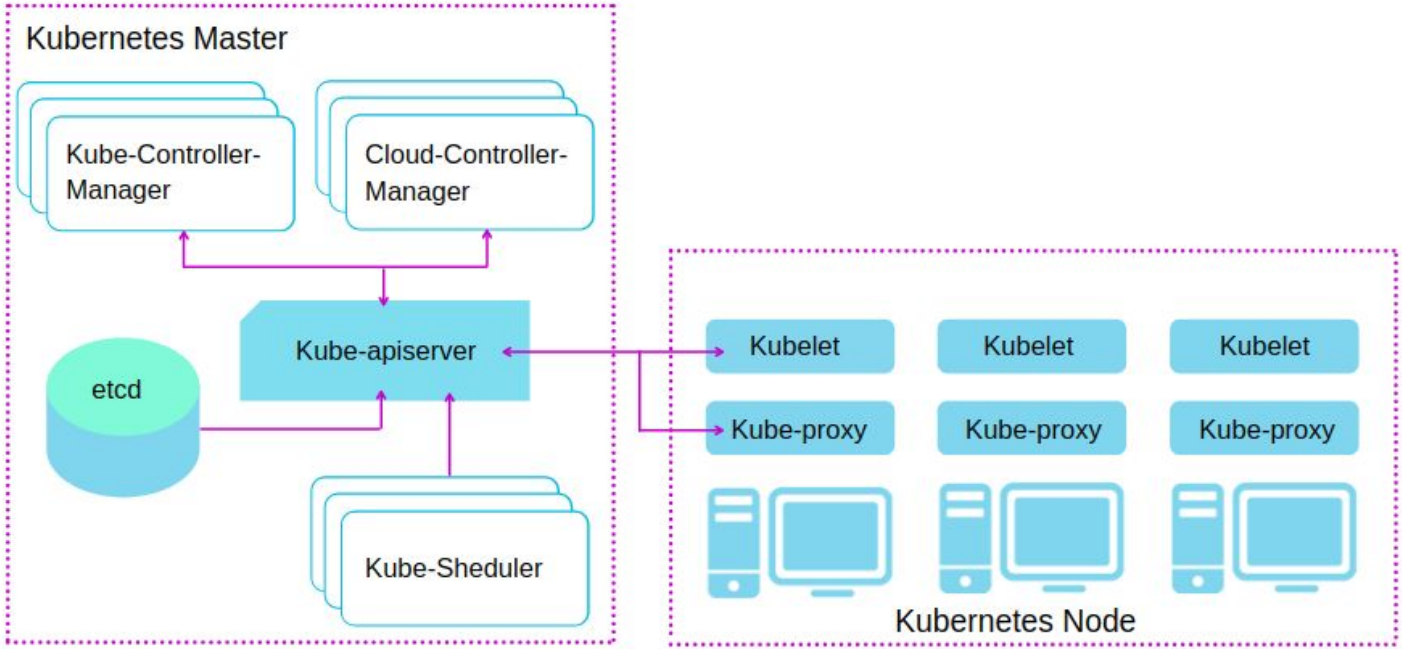
Why?

- Service Discovery & Load Balancing
- Storage Orchestration
- Automated rollouts & rollbacks
- Automatic bin packing
- Self-healing
- Secret and configuration Management

Architecture Overview



Architecture Overview



Master Components - ETCD

- Distributed reliable key-value store that is simple, secure & fast
- Uses RAFT based consensus algorithm to work in distributed environment
- Key value store distributed database
- Runs on port 2379

Master Components - API Server

- The central management entity
- Only component that connects to ETCD
- Designed for horizontal scaling

Connectivity:

- External: kubectl
- Internal: kubelet
- Persistent Storage: ETCD

Master Components - Scheduler

Schedules pods on appropriate Node(s)

Watches for newly created PODs that have no nodes assigned

Decision Parameters:

- Resource requirements (memory, cpu, disk type say SSD)
- Hardware, Software, Policy requirements
- Affinity, Anti-affinity
- Data locality
- Inter workload interference
- Deadlines

Master Components - Kube Controller

- Node Controller
  - Responsible for noticing and responding when nodes go down
- Replication Controller
  - Responsible for maintaining the correct number of pods for every replication controller object in the system
- Endpoints Controller
  - Populates the Endpoints object (that is, joins Services & Pods)
- Service Account & token Controller
  - Create default accounts and API access tokens for new namespaces

Master Components - Cloud Controller

- Node Controller
  - For checking the cloud provider to determine if a node has been deleted in the cloud after it stops responding
- Route Controller
  - For setting up routes in the underlying cloud infrastructure
- Service Controller
  - For creating, updating and deleting cloud provider load balancers
- Volume Controller
  - For creating, attaching, and mounting volumes, and interacting with the cloud provider to orchestrate volumes

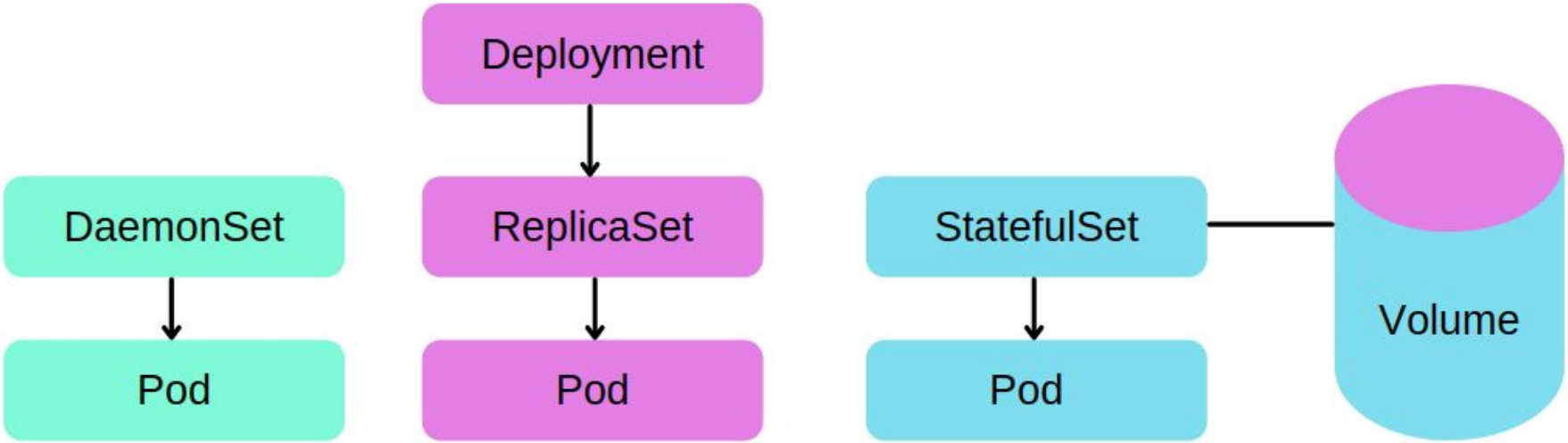
Node Components

- KUBE-PROXY
  - Network proxy that runs on every node in cluster
  - Maintains network rules on nodes
  - Uses OS packet filtering layer else forwards traffic itself
- KUBELET
  - Runs on every node
  - Ensures containers are running & healthy in PODs
  - Doesn’t manage container not created by K8S

Addon Components

- Cluster DNS
  - Cluster DNS is a DNS server, in addition to the other DNS server(s) in your environment, which serves DNS records for Kubernetes services
- Web UI
  - General purpose, web-based UI for Kubernetes clusters to view and manager cluster
- Container Resource Monitoring
  - Generic time-series metrics about containers in a central database, and provides a UI for browsing that data
- Cluster level Logging
  - Mechanism responsible for saving container logs to a central log store with search/browsing interface

Objects


Deployment
DaemonSet
ReplicaSet
StatefulSet
Volume
Pod
Pod
Pod

Kubectl

Command line tool to control kubernetes cluster

- Imperative commands to manage objects (basic & intermediate)
- Deploy commands
- Cluster Management commands
- Troubleshooting and Debugging
- Advanced, Settings and Other

kubectl - commands

- kubectl get pods
- kubectl describe pod hello-world
- kubectl describe pod/nginx
- kubectl delete pod nginx
- kubectl cluster-info
- kubectl get pods -o yaml
- kubectl get services -o json
- kubectl get pods --sort-by=.metadata.name
- kubectl get rs,deployments,service
- kubectl describe pods
- kubectl get pod/<pod-name> svc/<svc-name>
- kubectl get pod -l name=<label-name>
- kubectl delete pods --all
- kubectl get nodes -o json | jq '.items[] | {name:.metadata.name, cap:.status.capacity}'
- kubectl get nodes -o yaml | egrep '\sname:|cpu:|memory:'
- kubectl get all
- kubectl run hello-world --image=tutum/hello-world --port=80
- kubectl run -it busybox --image=busybox --restart=Never
- kubectl run nginx --image=nginx

Exercises

All lab work is available on Google classroom

Join link:

https://classroom.google.com/c/MzE3MjgxODUzMDUx?cjc=kun7p7g

POD

- Overview
- Lifecycle
- Init Containers
- Preset
- Topology Spread

POD - Overview

- ❖ Smallest deployable unit
- ❖ Supports multiple cooperating processes (containers) that form cohesive unit of service
- ❖ Ephemeral Entity

Encapsulates

- application container(s)
- Storage resources
- Unique network IP

Shared Resources:

- Networking
- Storage

Example

```yaml
apiVersion: v1
kind: Pod
metadata:
  creationTimestamp: null
  labels:
    run: nginx
  name: nginx
spec:
  containers:
  - image: nginx
    name: nginx
    resources: {}
  dnsPolicy: ClusterFirst
  restartPolicy: Always
status: {}
```

Example

```yaml
apiVersion: v1
kind: Pod
metadata:
  labels:
    run: pod-busybox
  name: pod-busybox
spec:
  containers:
  - command:
    - sh
    - -c
    - echo App is running! && sleep 30
    image: busybox
    name: pod-busybox
    resources: {}
  restartPolicy: Never
```

POD - Lifecycle

- Phase
  - Pending (waiting to be scheduled, image downloading)
  - Running (all containers started and ready to serve)
  - Succeeded (all containers exited with success)
  - Failed (all containers exited but at least one with failure)
  - Unknown (unable to fetch status as node is unreachable)
- Container States
  - Waiting, Running, Terminated
- Restart Policy (**Always**, Never, OnFailure)
- Conditions
  - Type: PodScheduled, ContainersReady, Initialized, Ready
    (lastProbeTime, lastTransitionTime, Message, reason, status)
- Probes
  - Startup, Readiness, Liveness
- Lifecycle hooks

Phase - Pending

```yaml
apiVersion: v1
kind: Pod
metadata:
  creationTimestamp: null
  labels:
    run: nginx
  name: nginx
spec:
  containers:
  - image: nginx
    name: nginx
    resources:
      requests:
        cpu: "1000m"
        memory: "1Gi"
  dnsPolicy: ClusterFirst
  restartPolicy: Never
status: {}
```

Phase - Running

```yaml
apiVersion: v1
kind: Pod
metadata:
  creationTimestamp: null
  labels:
    run: busybox
  name: busybox
spec:
  containers:
  - command:
    - ping
    - google.com
    image: busybox
    name: busybox
    resources: {}
  dnsPolicy: ClusterFirst
  restartPolicy: Always
status: {}
```

Phase - Succeeded

```yaml
apiVersion: v1
kind: Pod
metadata:
  creationTimestamp: null
  labels:
    run: busybox
  name: busybox
spec:
  containers:
  - image: busybox
    name: busybox
    resources: {}
  dnsPolicy: ClusterFirst
  restartPolicy: Never
status: {}
```

Probes

- Types
  - Startup
  - Readiness
  - Liveness
- Methods
  - Http
  - Tcp
  - Command

- Settings
  - initialDelaySeconds
  - periodSeconds
  - timeoutSeconds
  - successThreshold
  - failureThreshold

- Http
  - Host
  - Scheme
  - Path
  - Port
  - Headers

Probe - Liveness - Exec

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: probe-liveness-exec
spec:
  containers:
  - name: probe-liveness-exec
    image: k8s.gcr.io/busybox
    args:
    - /bin/sh
    - -c
    - touch /tmp/healthy; sleep 30; rm -rf /tmp/healthy; sleep 600
    livenessProbe:
      exec:
        command:
        - cat
        - /tmp/healthy
      initialDelaySeconds: 5
      periodSeconds: 5
```

Probe - Liveness - http

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: probe-liveness-http
spec:
  containers:
  - name: probe-liveness-http
    image: k8s.gcr.io/liveness
    args:
    - /server
    livenessProbe:
      httpGet:
        path: /healthz
        port: 8080
        httpHeaders:
        - name: Custom-Header
          value: Awesome
      initialDelaySeconds: 3
      periodSeconds: 3
```

Probe - Liveness - readiness - tcp

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: probe-liveness-readiness-tcp
spec:
  containers:
  - name: probe-liveness-readiness-tcp
    image: k8s.gcr.io/goproxy:0.1
    ports:
    - containerPort: 8080
    readinessProbe:
      tcpSocket:
        port: 8080
      initialDelaySeconds: 5
      periodSeconds: 10
    livenessProbe:
      tcpSocket:
        port: 8080
      initialDelaySeconds: 15
      periodSeconds: 20
```

Probe - Liveness - startup - http

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: probe-liveness-startup-http
spec:
  containers:
  - name: probe-liveness-startup-http
    image: k8s.gcr.io/liveness
    args:
    - /server
    livenessProbe:
      httpGet:
        path: /healthz
        port: 8080
      failureThreshold: 1
      periodSeconds: 10
    startupProbe:
      httpGet:
        path: /healthz
        port: 8080
      failureThreshold: 30
      periodSeconds: 10
```

POD Init Containers

- Always run to completion
- Must complete successfully before next one
- Readiness probes not supported
- Run(s) before application containers

Examples:

- Custom code / utilities to run before app containers
- Block / delay app container startup
- App container image building can be separate

POD Init - Statuses

- Init:N/M
- Init:Error
- Init:CrashLoopBackOff
- Pending
- PodIniliazing
- Running

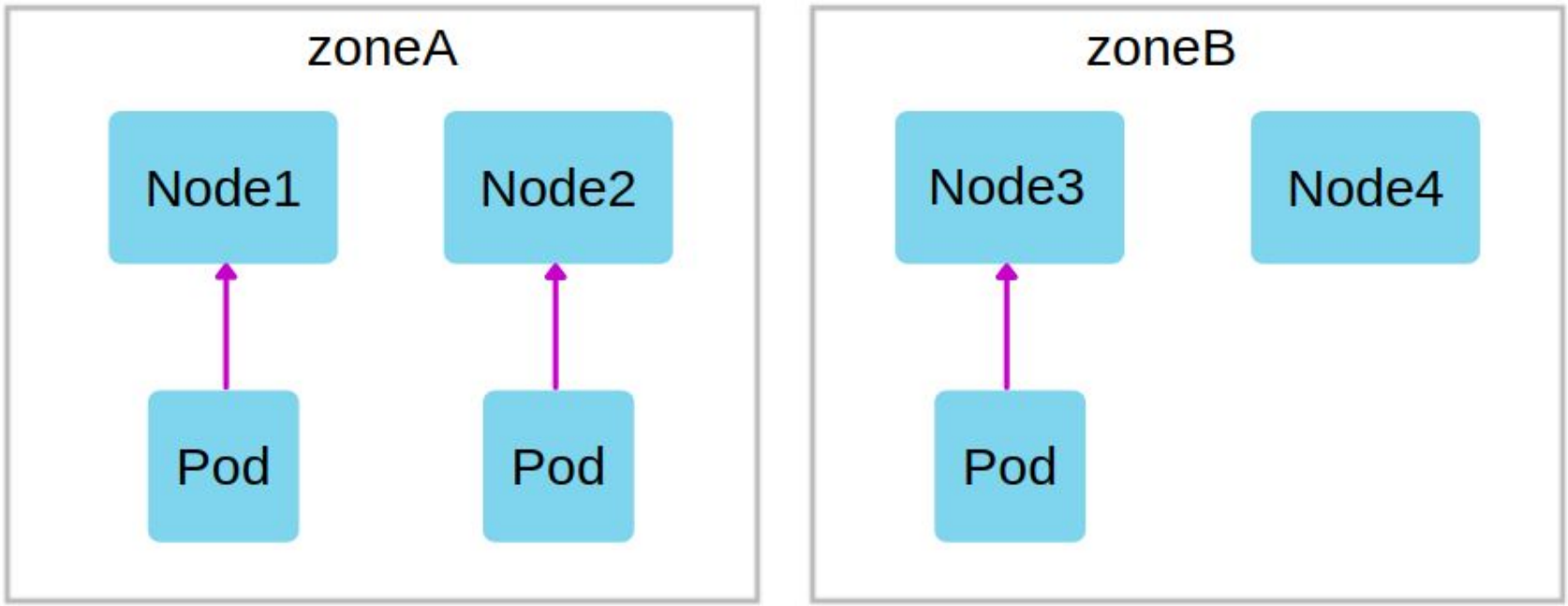
Example

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: init-containers
spec:
  containers:
  - name: main-container
    image: busybox:1.28
    command: ['sh', '-c', 'echo The app is running! && sleep 3600']
  initContainers:
  - name: init-service
    image: busybox:1.28
    command: ['sh', '-c', "until nslookup myservice.$(cat /var/run/secrets/kubernetes.io/serviceaccount/namespace).svc.cluster.local; do echo waiting for myservice; sleep 2; done"]
  - name: init-mydb
    image: busybox:1.28
    command: ['sh', '-c', "until nslookup mydb.$(cat /var/run/secrets/kubernetes.io/serviceaccount/namespace).svc.cluster.local; do echo waiting for mydb; sleep 2; done"]
```

POD - Topology Spread

- Objectives
  - To control how Pods are spreads across regions, zones, nodes and other user defined topology domains
  - To achieve high availability
  - To achieve efficient resource utilization
- Spread Constraints
  - maxSkew
  - topologyKey
  - whenUnsatisfiable (DoNotSchedule / ScheduleAnyway)
  - labelSelector

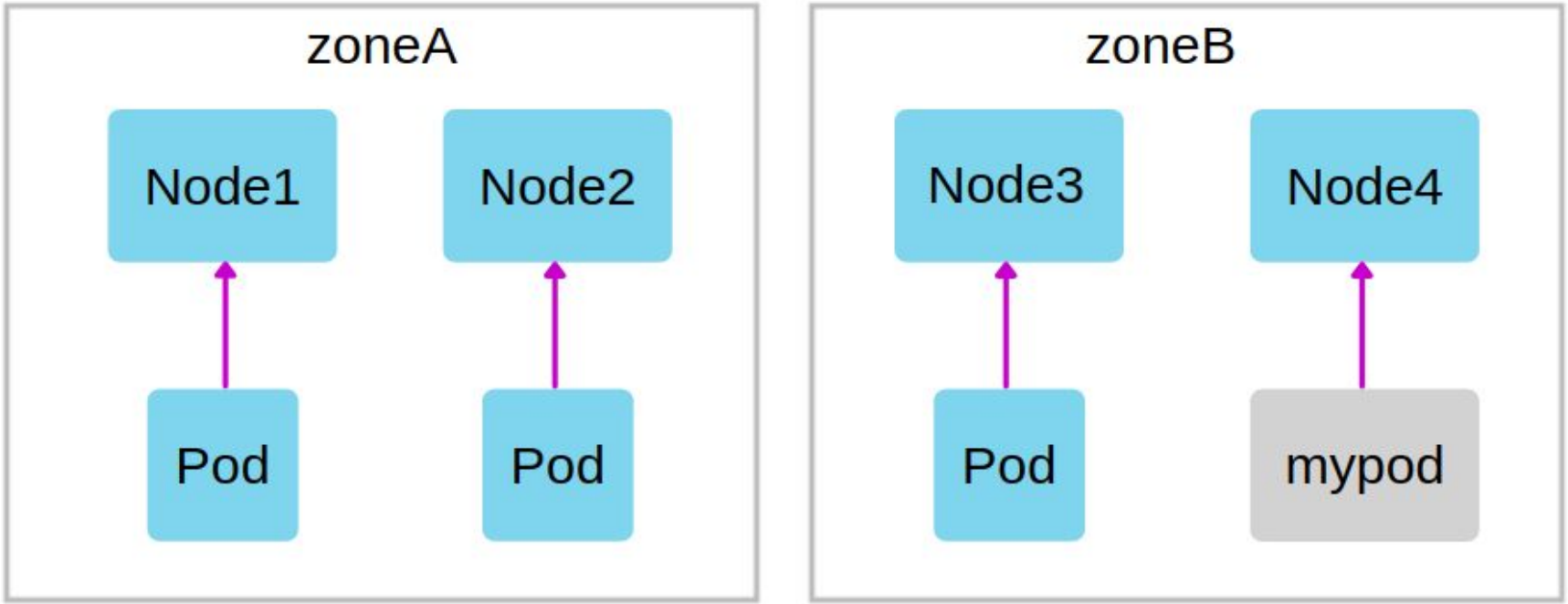
POD - Topology Spread



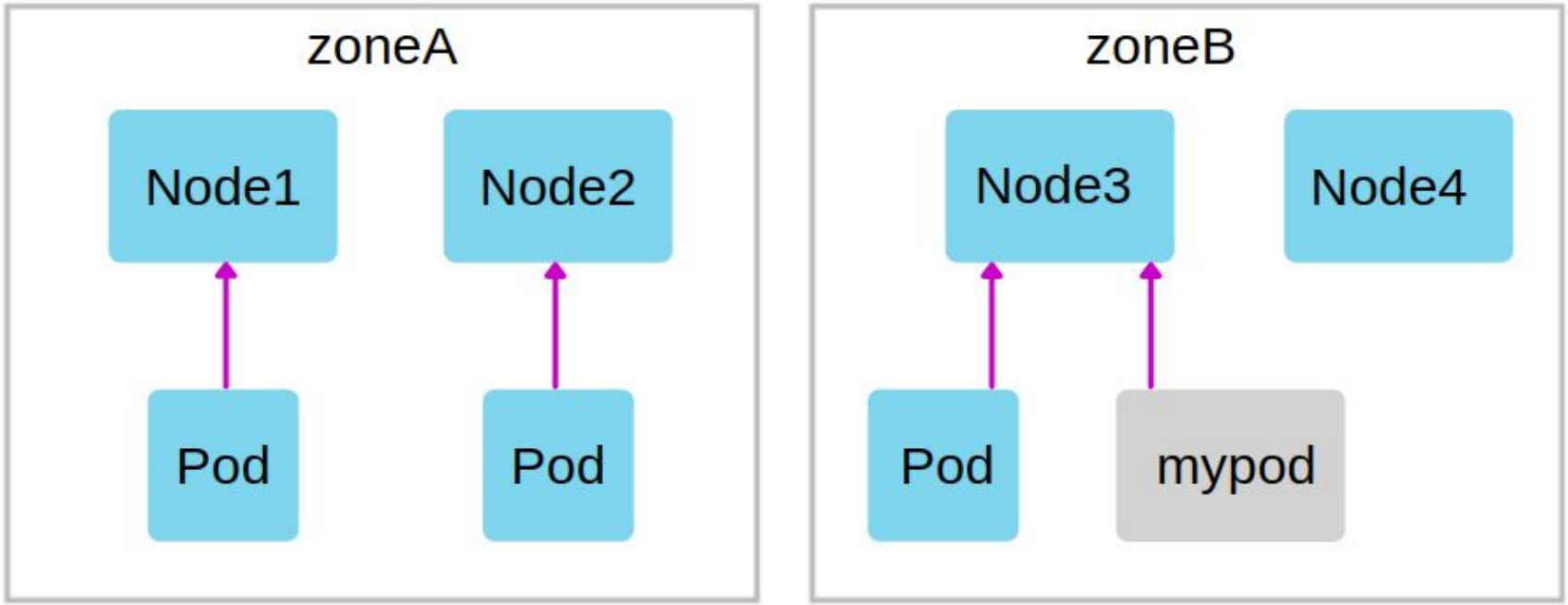
POD - Topology Spread

```yaml
kind: Pod
apiVersion: v1
metadata:
  name: topology-zone-constraint-1
  labels:
    foo: bar
spec:
  topologySpreadConstraints:
  - maxSkew: 1
    topologyKey: failure-domain.beta.kubernetes.io/zone
    whenUnsatisfiable: DoNotSchedule
    labelSelector:
      matchLabels:
        foo: bar
  containers:
  - name: pause
    image: k8s.gcr.io/pause:3.1
```

POD - Topology Spread



POD - Topology Spread



Topology - Multiple Constraints

```yaml
kind: Pod
apiVersion: v1
metadata:
  name: topology-contraints-two
  labels:
    foo: bar
spec:
  topologySpreadConstraints:
  - maxSkew: 1
    topologyKey: topology.kubernetes.io/zone
    whenUnsatisfiable: DoNotSchedule
    labelSelector:
      matchLabels:
        foo: bar
  - maxSkew: 1
    topologyKey: kubernetes.io/hostname
    whenUnsatisfiable: DoNotSchedule
    labelSelector:
      matchLabels:
        foo: bar
  containers:
  - name: pause
    image: k8s.gcr.io/pause:3.1
```

Topology - Constraint & Node Affinity

```yaml
kind: Pod
apiVersion: v1
metadata:
  name: topology-contraint-nodeaffinity
  labels:
    foo: bar
spec:
  topologySpreadConstraints:
  - maxSkew: 1
    topologyKey: zone
    whenUnsatisfiable: DoNotSchedule
    labelSelector:
      matchLabels:
        foo: bar
  affinity:
    nodeAffinity:
      requiredDuringSchedulingIgnoredDuringExecution:
        nodeSelectorTerms:
        - matchExpressions:
          - key: zone
            operator: NotIn
            values:
            - zoneC
  containers:
  - name: pause
    image: k8s.gcr.io/pause:3.1
```

Exercises

All lab work is available on Google classroom

Join link:

https://classroom.google.com/c/MzE3MjgxODUzMDUx?cjc=kun7p7g

Pod - Advanced

- Patterns
  - Ambassador
  - Sidecar
  - Adapter
- Jobs
- Cron Jobs
- Labels, Selectors, Annotations
- Demo
- Practicals

Patterns - POD

- To extend the functionality of the existing container
- To have helper process enhancing work of the existing container
- To send logs to external server

- Types
  - Sidecar - To export logs
  - Ambassador - To proxy connection
  - Adapter - To standardise and normalize output

Pattern - Sidecar

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: sidecar
spec:
  volumes:
  - name: shared-logs
    emptyDir: {}
  containers:
  - name: main-container
    image: alpine
    command: ["/bin/sh"]
    args: ["-c", "while true; do date >> /var/log/index.html; sleep 10;done"]
    volumeMounts:
    - name: shared-logs
      mountPath: /var/log
  - name: sidecar-container
    image: nginx
    ports:
      - containerPort: 80
    volumeMounts:
    - name: shared-logs
      mountPath: /usr/share/nginx/html
```

Pattern - Adapter

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: adapter
spec:
  volumes:
  - name: shared-logs
    emptyDir: {}
  containers:
  - name: main-container
    image: alpine
    command: ["/bin/sh"]
    args: ["-c", "while true; do date > /var/log/top.txt && top -n 1 -b >> /var/log/top.txt; sleep 10;done"]
    volumeMounts:
    - name: shared-logs
      mountPath: /var/log
  - name: adapter-container
    image: alpine
    command: ["/bin/sh"]
    args: ["-c", "while true; do (cat /var/log/top.txt | head -1 > /var/log/status.txt) && (cat /var/log/top.txt | head -2 | tail -1 | grep
 -o -E '\\d+\\w' | head -1 >> /var/log/status.txt) && (cat /var/log/top.txt | head -3 | tail -1 | grep
-o -E '\\d+%' | head -1 >> /var/log/status.txt); sleep 5; done"]
    volumeMounts:
    - name: shared-logs
      mountPath: /var/log
```

Pattern - Ambassador

```yaml
apiVersion: v1
kind: ConfigMap
metadata:
  name: ambassador-nginx-config
data:
  nginx.conf: |
    worker_processes 1;
    worker_rlimit_nofile 4096;
    events {
      worker_connections 512;
    }
    http {
      proxy_set_header HOST $host;
      proxy_set_header X-Forwarded-Proto $scheme;
      proxy_set_header X-Real-IP $remote_addr;
      proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
      upstream backend {
        server msn.com:80;
      }
      server {
        listen 80;
        location / {
          proxy_pass http://backend;
        }
      }
    }
---
apiVersion: v1
kind: Pod
metadata:
  name: multi-pod-ambassador
spec:
  containers:
  - name: main-app
    image: busybox
    imagePullPolicy: IfNotPresent
    command: ["/bin/sh"]
    args: ["-c","while true;do wget -O /tmp/app.txt localhost ;sleep 30;done"]
  - name: ambassador
    image: nginx
    imagePullPolicy: IfNotPresent
    ports:
    - containerPort: 80
    volumeMounts:
      - name: nginx-config
        mountPath: /etc/nginx/nginx.conf
        subPath: nginx.conf
  volumes:
    - name: nginx-config
      configMap:
        name: ambassador-nginx-config
```

Job

- To provide reliable parallel execution of tasks
- Examples:
  - Send emails, transcode files, Scan database for a set of rows,
- Patterns
  - Non parallel job
  - Fixed completion count job
  - Work queue job

Job

```yaml
apiVersion: batch/v1
kind: Job
metadata:
  name: job
spec:
  template:
    spec:
      containers:
      - name: perl
        image: perl
        command: ["perl",  "-Mbignum=bpi", "-wle", "print bpi(2000)"]
      restartPolicy: Never
  backoffLimit: 4
```

Job - Timeout

```yaml
apiVersion: batch/v1
kind: Job
metadata:
  name: job-timeout
spec:
  backoffLimit: 5
  activeDeadlineSeconds: 100
  template:
    spec:
      containers:
      - name: perl
        image: perl
        command: ["perl",  "-Mbignum=bpi", "-wle", "print bpi(2000)"]
      restartPolicy: Never
```

CronJob

- Creates jobs on a repeating schedule
- Schedule times are based on kube-controller-manager
- Useful for tasks like migrating data to reporting server, sending emails, creating backups etc
- Schedule tasks at specific time (like when cluster is idle)

Key Configurations:

- startingDeadlineSeconds - Missed occurrences in last X seconds will be counted
- concurrencyPolicy
    - If Allow, then job will run at least once
    - If Forbid, will be missed if previous instance is still running

CronJob - Expression

```
# ┌──────────── minute (0 - 59)
# │ ┌──────────── hour (0 - 23)
# │ │ ┌──────────── day of the month (1 - 31)
# │ │ │ ┌──────────── month (1 - 12)
# │ │ │ │ ┌──────────── day of the week (0 - 6) (Sunday to Saturday;
# │ │ │ │ │                                     7 is also Sunday on some systems)
# │ │ │ │ │
# │ │ │ │ │
  # * * * * * <command to execute>
```

Examples:

- */15 0,8,16 * * * echo running backup (every 15 minutes of 0,8 & 16th hour)
- 30 0 * * 6 /home/oracle/scripts/export_dump.sh (last day of week at 00:30)
- 1 0 * * * printf "" > /var/log/apache/error_log (everyday at 00:01)

CronJob - Expression

```yaml
apiVersion: batch/v1beta1
kind: CronJob
metadata:
  name: cron-job
spec:
  schedule: "*/1 * * * *"
  jobTemplate:
    spec:
      template:
        spec:
          containers:
          - name: cron-job
            image: busybox
            args:
            - /bin/sh
            - -c
            - date; echo Migrating data to reporting server...
          restartPolicy: OnFailure

```

Labels

- Labels
  - Key value pairs attached to objects
  - To specify identifying attributes of objects
  - To organize and select subset of objects
  - To query objects efficiently (cli as well gui monitoring tools)
  - Attached at creation time and can be added / modified at any time
  - Label key must be unique per object
- Example labels:
  - `"release" : "stable", "release" : "canary"`
  - `"environment" : "dev", "environment" : "qa", "environment" : "production"`
  - `"tier" : "frontend", "tier" : "backend", "tier" : "cache"`
  - `"partition" : "customerA", "partition" : "customerB"`
  - `"track" : "daily", "track" : "weekly"`

Labels

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: pod-labels
  labels:
    environment: production
    app: nginx
spec:
  containers:
  - name: nginx
    image: nginx
    ports:
    - containerPort: 80
```

Selectors

- Equality Based
  - `environment = production`
  - `tier != frontend`
- Set Based
  - `environment in (production, qa)`
  - `tier notin (frontend, backend)`
  - `partition`
  - `!partition`

Selectors - Examples

- `kubectl get pods -l environment=production,tier=frontend`
- `kubectl get pods -l 'environment in (production),tier in (frontend)'`
- `kubectl get pods -l 'environment in (production, qa)'`
- `kubectl get pods -l 'environment,environment notin (frontend)'`

Jobs, Deployments, ReplicaSet, Daemonset

```
selector:
  matchLabels:
    component: redis
  matchExpressions:
    - {key: tier, operator: In, values: [cache]}
    - {key: environment, operator: NotIn, values: [dev]}
```

Selectors - Examples

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: selector-pod-node
spec:
  containers:
    - name: cuda-test
      image: "k8s.gcr.io/cuda-vector-add:v0.1"
      resources:
        limits:
          nvidia.com/gpu: 1
  nodeSelector:
    accelerator: nvidia-tesla-p100
```

Annotations

- To attach non-identifying arbitrary metadata to objects
- Usage
  - Pointers for debugging purposes
  - Build, release, image hashes etc
  - Author info, contact details
  - Metadata to help tools for deployment, management, introspection

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: annotations-pod
  annotations:
    imageregistry: "https://hub.docker.com/"
spec:
  containers:
  - name: nginx
    image: nginx
    ports:
    - containerPort: 80
```

Exercises - Reference

All lab work is available on Google classroom

Join link:

https://classroom.google.com/c/MzE3MjgxODUzMDUx?cjc=kun7p7g

Scalability

- Deployments
  - Rolling Updates & Rollbacks
  - Auto scaling pods
- Demo
- Practicals

Deployments

Use Cases

- To rollout a set of PODs
- To declare a new set of PODs
- To rollback to an earlier version of deployment
- To scale up deployment to facilitate more load
- To pause the deployment / rollout
- To autoscale deployment when cpu usage threshold reached

Deployment - Example

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: test-app
spec:
  replicas: 1
  selector:
    matchLabels:
      component: test-app
  template:
    metadata:
      labels:
        component: test-app
    spec:
        containers:
          - name: test-app
            image: brainupgrade/test-app:all-tiers-in-one
            imagePullPolicy: IfNotPresent
            ports:
              - containerPort: 8080
            resources:
              requests:
                cpu: "100m"
                memory: "250Mi"

```

Deployment - Commands

- kubectl create deployment nginx --image=nginx:1.15 --replicas=5
- kubectl get deployment/nginx
- kubectl describe deployment/nginx
- kubectl rollout history deployment/nginx
- kubectl set image deployment/nginx nginx=nginx:1.16
- kubectl rollout history deployment/nginx
- kubectl rollout undo deployment/nginx
- kubectl rollout undo deployment/nginx --to-revision=2
- kubectl scale --replicas=50 deployment/nginx
- kubectl rollout pause deployment/nginx
- kubectl rollout status deployment/nginx
- kubectl rollout resume deployment/nginx
- kubectl autoscale deployment/nginx --min=2 --max=10

Exercises

- Create a deployment to run 3 replicas of nginx container
- Scale down the replicas to 1
- Scale up replicas to 10
- View the roll out history
- Switch to rollout version 2
- Scale down replicas to 5
- Update image to nginx:1.18 and immediately try another rollout with nginx:1.17
- Observe if rollout with nginx:1.17 was completed or not
- Autoscale pod to max 5 min 1

Persistence

- Persistence Volume
- Persistence Volume Claim
- Statefulset
- Daemonset

Overview - PV / PVC

- Ephemeral
  - Tightly coupled with POD lifetime
  - Deleted when POD is removed
  - Example: emptydir
- Persistent
  - Survives POD reboots
  - Meant for long term and independent of POD / Node lifecycle
  - Examples: hostpath, NFS, Cloud storage (EBS etc)
- The access modes are:
  - ReadWriteOnce -- the volume can be mounted as read-write by a single node
  - ReadOnlyMany -- the volume can be mounted read-only by many nodes
  - ReadWriteMany -- the volume can be mounted as read-write by many nodes

Examples - emptyDir

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: volume-emptydir
spec:
  containers:
  - image: nginx
    name: test-container
    volumeMounts:
    - mountPath: /cache
      name: cache-volume
  volumes:
  - name: cache-volume
    emptyDir: {}
```

Examples - hostpath (file/dir)

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: hostpath-volume
spec:
  containers:
  - image: nginx
    name: test-container
    volumeMounts:
    - mountPath: /data-mounted-as
      name: hostpath-volume
  volumes:
  - name: hostpath-volume
    hostPath:
      # directory location on host
      path: /data
      # this field is optional
      type: DirectoryOrCreate
```

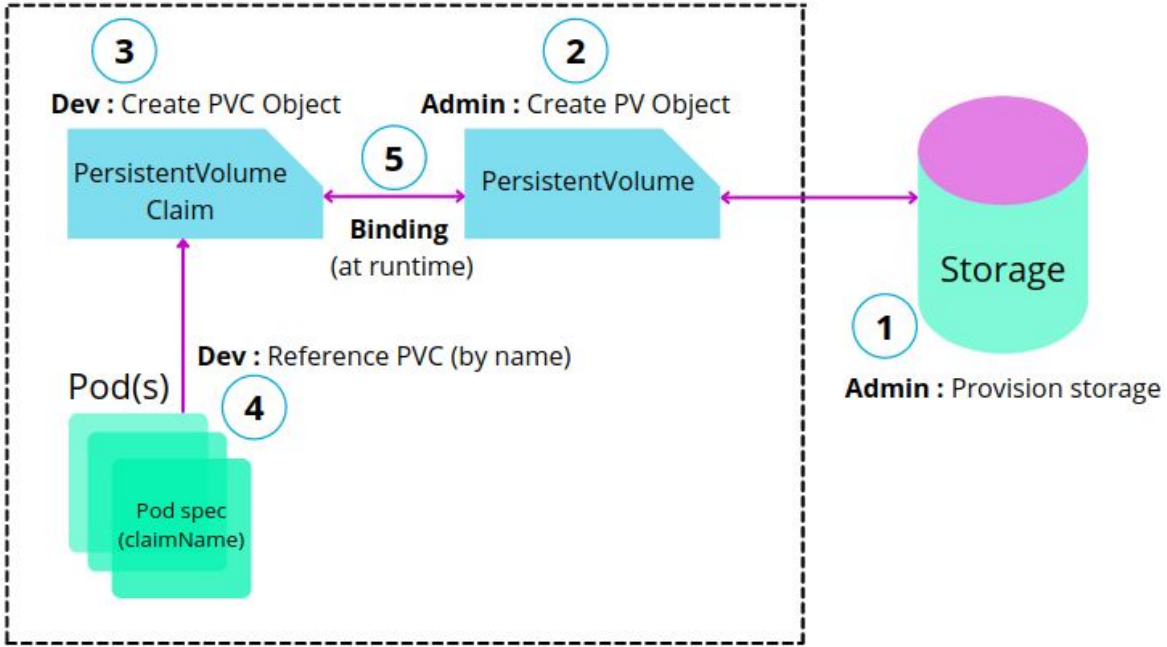
Persistent Volume - local & node

```yaml
apiVersion: v1
kind: PersistentVolume
metadata:
  name: pv-local-node-affinity
spec:
  capacity:
    storage: 10Gi
  # volumeMode field requires BlockVolume Alpha feature gate to be enabled.
  volumeMode: Filesystem
  accessModes:
  - ReadWriteMany
  persistentVolumeReclaimPolicy: Delete
  storageClassName: local-storage
  local:
    path: /mnt/disks/ssd1
  nodeAffinity:
    required:
      nodeSelectorTerms:
      - matchExpressions:
        - key: kubernetes.io/hostname
          operator: In
          values:
          - ip-172-31-87-231.ec2.internal
```

Persistent Volume - EBS

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: volume-ebs
spec:
  containers:
  - image: k8s.gcr.io/test-webserver
    name: test-container
    volumeMounts:
    - mountPath: /test-ebs
      name: ebs-volume
  volumes:
  - name: ebs-volume
    # This AWS EBS volume must already exist.
    awsElasticBlockStore:
      volumeID: <volume-id>
      fsType: ext4
```

Persistent Volume - static



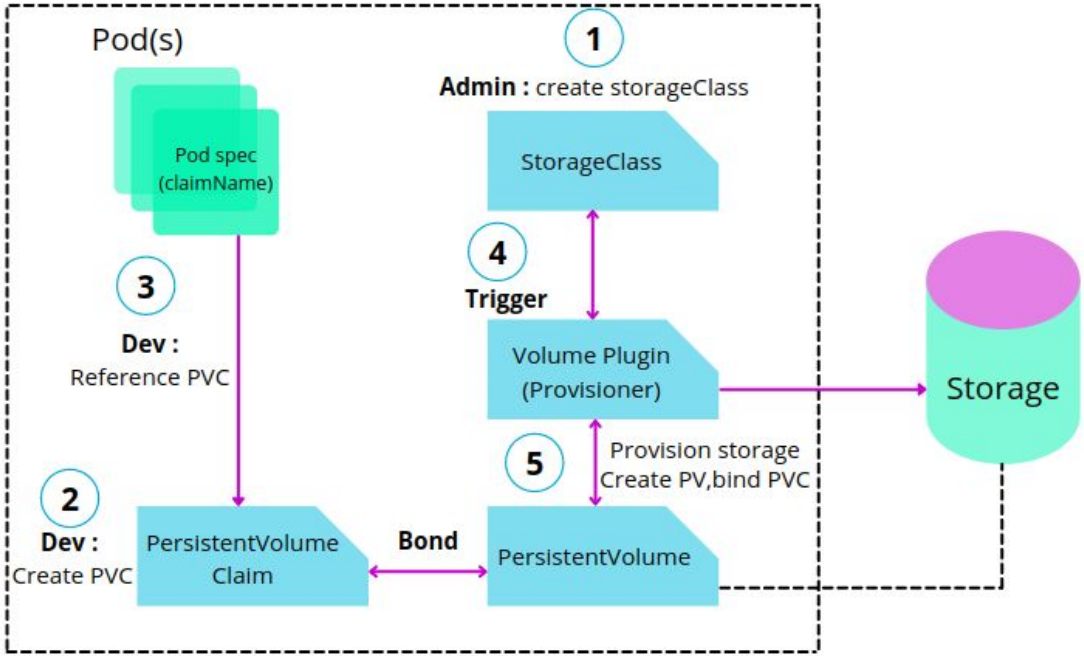
Example

```yaml
apiVersion: v1
kind: PersistentVolume
metadata:
  name: pv-hostpath
  annotations:
    pv.beta.kubernetes.io/gid: "1234"
  labels:
    type: local
spec:
  storageClassName: manual
  capacity:
    storage: 10Gi
  accessModes:
    - ReadWriteOnce
  hostPath:
    path: "/mnt/data"
```

```yaml
apiVersion: v1
kind: PersistentVolumeClaim
metadata:
  name: pvc-test
spec:
  storageClassName: manual
  accessModes:
    - ReadWriteOnce
  resources:
    requests:
      storage: 3Gi
```

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: pod-pv-pvc
spec:
  volumes:
    - name: pv-storage
      persistentVolumeClaim:
        claimName: pvc-test
  containers:
    - name: task-pv-container
      image: nginx
      ports:
        - containerPort: 80
          name: "http-server"
      volumeMounts:
        - mountPath: "/usr/share/nginx/html"
          name: pv-storage
```

Persistent Volume - Dynamic



Example

```yaml
kind: StorageClass
apiVersion: storage.k8s.io/v1
metadata:
  name: storageclass-generic
provisioner: kubernetes.io/aws-ebs
parameters:
  type: gp2
  zones: us-east-1a,us-east-1b,us-east-1c
  iopsPerGB: "10"
  fsType: ext4
```

```yaml
apiVersion: v1
kind: PersistentVolumeClaim
metadata:
  name: pvc-dynamic
  labels:
    app: nginx
spec:
  storageClassName: storageclass-generic
  accessModes:
    - ReadWriteOnce
  resources:
    requests:
      storage: 1Gi
```

```yaml
kind: Pod
apiVersion: v1
metadata:
  name: pod-volume-dynamic
  labels:
    app: nginx
spec:
  containers:
    - name: nginx
      image: nginx
      volumeMounts:
      - mountPath: "/var/www/html"
        name: external
  volumes:
    - name: external
      persistentVolumeClaim:
        claimName: pvc-dynamic
```

Persistent Volumes

- GCEPersistentDisk
- AWSElasticBlockStore
- AzureFile
- AzureDisk
- CSI
- FC (Fibre Channel)
- FlexVolume
- Flocker
- NFS
- iSCSI
- RBD (Ceph Block Device)
- CephFS
- Cinder (OpenStack block storage)
- Glusterfs
- VsphereVolume
- Quobyte Volumes
- HostPath (Single node testing only – local storage is not supported in any way and WILL NOT WORK in a multi-node cluster)
- Portworx Volumes
- ScaleIO Volumes
- StorageOS

StatefulSet

Use Cases

- Stable, unique network identifiers
- Stable, persistent storage
- Ordered, graceful deployment and scaling
- Ordered, automated rolling updates

Limitations

- No automatic deletion of referenced volumes
- No PODs deletion guarantee when StatefulSet is deleted
- Rolling Updates not consistent always

StatefulSet - Example

```yaml
apiVersion: apps/v1
kind: StatefulSet
metadata:
  name: sts-web
spec:
  serviceName: "nginx"
  replicas: 2
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
      - name: nginx
        image: k8s.gcr.io/nginx-slim:0.8
        ports:
        - containerPort: 80
          name: web
        volumeMounts:
        - name: www
          mountPath: /usr/share/nginx/html
  volumeClaimTemplates:
  - metadata:
      name: www
    spec:
      accessModes: [ "ReadWriteOnce" ]
      resources:
        requests:
          storage: 1Gi
```

- Scale Up
- Scale Up
- Update (image)

StatefulSet - Example

- Scale Up
- Scale down
- kubectl set image sts/sts-web nginx=nginx:1.18
- Staged Update
    - kubectl patch statefulset sts-web -p '{"spec":{"updateStrategy":{"type":"RollingUpdate","rollingUpdate":{"partition":3}}}}'
    - kubectl patch statefulset sts-web --type='json' -p='[{"op": "replace", "path": "/spec/template/spec/containers/0/image", "value":"nginx:1.17"}]'
    - kubectl get pod sts-web-1 --template '{{range $i, $c := .spec.containers}}{{$c.image}}{{end}}'

DaemonSet

Purpose

- To run a copy of a POD on all / some node(s)

Use Cases

- Storage cluster daemon (gluster, ceph)
- Log Collectors (fluentd, logstash)
- Node Monitoring daemons (Prometheus, Dynatrace, collectd)

DaemonSet - Example

```yaml
apiVersion: apps/v1
kind: DaemonSet
metadata:
  name: fluentd-elasticsearch
  namespace: kube-system
  labels:
    k8s-app: fluentd-logging
spec:
  selector:
    matchLabels:
      name: fluentd-elasticsearch
  template:
    metadata:
      labels:
        name: fluentd-elasticsearch
    spec:
      tolerations:
      - key: node-role.kubernetes.io/master
        effect: NoSchedule
      containers:
      - name: fluentd-elasticsearch
        image: quay.io/fluentd_elasticsearch/fluentd:v2.5.2
        resources:
          limits:
            memory: 200Mi
          requests:
            cpu: 100m
            memory: 200Mi
        volumeMounts:
        - name: varlog
          mountPath: /var/log
        - name: varlibdockercontainers
          mountPath: /var/lib/docker/containers
          readOnly: true
      terminationGracePeriodSeconds: 30
      volumes:
      - name: varlog
        hostPath:
          path: /var/log
      - name: varlibdockercontainers
        hostPath:
          path: /var/lib/docker/containers
```

Exercises

- Create POD (nginx / redis) to use volume emptyDir
- Create nginx POD that uses pvc for serving web files
- Define pvc that uses pv
- Define pv that refers to host path /mnt/data

Exercises

- Create StatefulSet having image=nginx:1.16 and replicas 1
- Scale replicas to 5
- Change image to nginx:1.17
- Set the rolling update partition to 3
- Change the image to nginx:1.17
- Watch the pods scaling up / down using
    - kubectl get -w pods

Configuration

- Config Maps
- Secrets
- Security Contexts
- Accounts
- Demo
- Practicals

ConfigMap

- To store non-confidential key-value pairs
- Can be consumed as env variables, command line args or config files in volume
- To decouple env specific config from images for portability
- Max data 1MB

```yaml
apiVersion: v1
kind: ConfigMap
metadata:
  name: cm-game-demo
data:
  # property-like keys; each key maps to a simple value
  player_initial_lives: "3"
  ui_properties_file_name: "user-interface.properties"
  # file-like keys
  game.properties: |
    enemy.types=aliens,monsters
    player.maximum-lives=5
  user-interface.properties: |
    color.good=purple
    color.bad=yellow
    allow.textmode=true
```

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: pod-configmap
spec:
  containers:
    - name: demo
      image: alpine
      command: ["sleep", "3600"]
      env:
        # Define the environment variable
        - name: PLAYER_INITIAL_LIVES # Notice that the case is different here
                                     # from the key name in the ConfigMap.
          valueFrom:
            configMapKeyRef:
              name: cm-game-demo           # The ConfigMap this value comes from.
              key: player_initial_lives # The key to fetch.
        - name: UI_PROPERTIES_FILE_NAME
          valueFrom:
            configMapKeyRef:
              name: cm-game-demo
              key: ui_properties_file_name
      volumeMounts:
      - name: config
        mountPath: "/config"
        readOnly: true
  volumes:
    # You set volumes at the Pod level, then mount them into containers inside that Pod
    - name: config
      configMap:
        # Provide the name of the ConfigMap you want to mount.
        name: cm-game-demo
        # An array of keys from the ConfigMap to create as files
        items:
        - key: "game.properties"
          path: "game.properties"
        - key: "user-interface.properties"
          path: "user-interface.properties"
```

Secret

- To manage sensitive info like password, oauthkeys, docker login, ssh keys, tls etc

- Examples

```
kubectl create secret docker-registry secret-tiger-docker \
  --docker-username=tiger \
  --docker-password=pass113 \
  --docker-email=tiger@acme.com

kubectl create secret tls my-tls-secret \
  --cert=path/to/cert/file \
  --key=path/to/key/file
```

Secret

```yaml
apiVersion: v1
data:
  username: YWRtaW4=
  password: MWYyZDFlMmU2N2Rm
kind: Secret
metadata:
  name: pod-secret
  namespace: default
  resourceVersion: "164619"
  uid: cfee02d6-c137-11e5-8d73-42010af00002
type: Opaque
```

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: pod-secret
spec:
  containers:
  - name: mypod
    image: redis
    env:
      - name: SECRET_USERNAME
        valueFrom:
          secretKeyRef:
            name: mysecret
            key: username
      - name: SECRET_PASSWORD
        valueFrom:
          secretKeyRef:
            name: mysecret
            key: password
    volumeMounts:
    - name: foo
      mountPath: "/etc/foo"
      readOnly: true
  volumes:
  - name: foo
    secret:
      secretName: pod-secret
      items:
      - key: username
        path: my-group/my-username
```

Security Context

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: security-context-demo
spec:
  securityContext:
    runAsUser: 1000
    runAsGroup: 3000
    fsGroup: 2000
  volumes:
  - name: sec-ctx-vol
    emptyDir: {}
  containers:
  - name: sec-ctx-demo
    image: busybox
    command: [ "sh", "-c", "sleep 1h" ]
    volumeMounts:
    - name: sec-ctx-vol
      mountPath: /data/demo
    securityContext:
      allowPrivilegeEscalation: false
      capabilities:
        add: ["NET_ADMIN", "SYS_TIME"]

```

User Accounts & Service Accounts

- User Accounts
  - User accounts are for humans.
  - User accounts are intended to be global. Names must be unique across all namespaces of a cluster.
- Service Accounts
  - Service accounts are for processes, which run in pods.
  - Service accounts are namespaced.
  - Service account creation is intended to be more lightweight

Scheduling

- Node Name
- Node Selector
- Affinity (Pod, Node)
- Taints & Tolerations
- Demo
- Practicals

Node Name

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: schedule-node
spec:
  replicas: 1
  selector:
    matchLabels:
      component: schedule-node
  template:
    metadata:
      labels:
        component: schedule-node
    spec:
        nodeName: ip-172-31-102-18.ec2.internal
        containers:
          - name: test-app
            image: brainupgrade/test-app:all-tiers-in-one
            imagePullPolicy: IfNotPresent
            ports:
              - containerPort: 8080
            resources:
              requests:
                cpu: "100m"
                memory: "250Mi"
```

Node Selector

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: schedule-nodeselector
spec:
  replicas: 1
  selector:
    matchLabels:
      component: schedule-nodeselector
  template:
    metadata:
      labels:
        component: schedule-nodeselector
    spec:
        nodeSelector:
          node-role.kubernetes.io/spot-worker: "true"
        containers:
          - name: test-app
            image: nginx
            imagePullPolicy: IfNotPresent
            ports:
              - containerPort: 80
```

POD - Affinity & Anti Affinity (example 1)

```yaml
spec:
    affinity:
      podAffinity:
        requiredDuringSchedulingIgnoredDuringExecution:
        - labelSelector:
            matchExpressions:
            - key: tier
              operator: In  # NotIn, Exists, NotExists
              values:
              - cache
          topologyKey: topology.kubernetes.io/zone
      podAntiAffinity:
        preferredDuringSchedulingIgnoredDuringExecution:
        - weight: 100
          podAffinityTerm:
            labelSelector:
              matchExpressions:
              - key: tier
                operator: In
                values:
                - messaging
            topologyKey: topology.kubernetes.io/hostname
    containers:
      - name: pod-affinity-middle-tier
        image: nginx
        imagePullPolicy: IfNotPresent
        ports:
          - containerPort: 80
```

POD - Affinity & Anti Affinity (example 2)

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: app-cache-store
spec:
  selector:
    matchLabels:
      app: cache-store
  replicas: 2
  template:
    metadata:
      labels:
        app: cache-store
    spec:
      affinity:
        podAntiAffinity:
          requiredDuringSchedulingIgnoredDuringExecution:
          - labelSelector:
              matchExpressions:
              - key: app
                operator: In
                values:
                - cache-store
            topologyKey: "kubernetes.io/hostname"
      containers:
      - name: redis-server
        image: redis:3.2-alpine
```

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: app-web-server
spec:
  selector:
    matchLabels:
      app: web-server
  replicas: 2
  template:
    metadata:
      labels:
        app: web-server
    spec:
      affinity:
        podAntiAffinity:
          requiredDuringSchedulingIgnoredDuringExecution:
          - labelSelector:
              matchExpressions:
              - key: app
                operator: In
                values:
                - web-server
            topologyKey: "kubernetes.io/hostname"
        podAffinity:
          requiredDuringSchedulingIgnoredDuringExecution:
          - labelSelector:
              matchExpressions:
              - key: app
                operator: In
                values:
                - cache-store
            topologyKey: "kubernetes.io/hostname"
      containers:
      - name: web-app
        image: nginx:1.16-alpine
```

Node - Affinity

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: node-affinity
spec:
  replicas: 1
  selector:
    matchLabels:
      tier: middle-tier
  template:
    metadata:
      labels:
        tier: middle-tier
    spec:
      affinity:
        nodeAffinity:
          requiredDuringSchedulingIgnoredDuringExecution:
            nodeSelectorTerms:
            - matchExpressions:
              - key: topology.kubernetes.io/zone
                operator: In  #In, NotIn, Exists, DoesNotExist, Gt, Lt
                values:
                - us-east-1b
                - us-east-1c
          preferredDuringSchedulingIgnoredDuringExecution:
          - weight: 1
            preference:
              matchExpressions:
              - key: kops.k8s.io/instancegroup
                operator: In  #NotIn,  DoesNotExist for AntiAffinity
                values:
                - spotnodes
      containers:
        - name: nginx
          image: nginx
          imagePullPolicy: IfNotPresent
          ports:
            - containerPort: 80
```

Taints & Tolerations

- Objective

  To steer pods away from nodes or evict pods that should not be running on particular nodes

- Use Cases
  - Dedicated Nodes
  - Nodes with special hardware
- Notes
  - Taints - Node (to reject pods that don’t tolerate taints)
  - Tolerations - Pod (to allow pods to schedule onto nodes matching taints)

Taints

```
lab@rajesh-Gazelle:~$ kubectl get nodes -o json | jq ".items[]|{name:.metadata.name, taints:.spec.taints}"
{
  "name": "ip-172-31-102-18.ec2.internal",
  "taints": null
}
{
  "name": "ip-172-31-34-27.ec2.internal",
  "taints": [
    {
      "effect": "NoSchedule",
      "key": "node-role.kubernetes.io/master"
    }
  ]
}
{
  "name": "ip-172-31-87-231.ec2.internal",
  "taints": null
}
```

Taints

- Examples (Taints)
  - To apply

    kubectl taint nodes qanode release=qa:NoSchedule

    kubectl taint nodes devnode release=dev:PreferNoSchedule

    kubectl taint nodes prodnode release=prod:NoExecute
  - To remove

    kubectl taint nodes node-name key=value:Effect-

Tolerations

```yaml
apiVersion: v1
kind: Pod
metadata:
  labels:
    release: qa
  name: tolerations-release-qa
spec:
  containers:
  - image: nginx
    name: tolerations-unreachable
  dnsPolicy: ClusterFirst
  restartPolicy: Always
  tolerations:
    - key: "release"
      operator: "Exists"
      value: "qa"
      effect: "NoSchedule"
```

Services

- Cluster IP
- Node Port
- Load Balancer
- Demo
- Practicals

Service Types

- Cluster IP
  - Service exposed on cluster internal IP
  - Reachable only within cluster
- Node Port
  - Exposed on each Node IP at static port
- Load Balancer
  - Exposed through external cloud load balancer
- External Name
  - Exposed through the contents of external field via CNAME record

Service

- An abstract way to expose an application running on pod as network service.
- Frontends and backends of application can connect without worrying about POD IPs
- Uses session affinity while connecting to backend PODs

```
1  apiVersion: v1
2  kind: Service
3  metadata:
4    labels:
5      app: nginx
6    name: nginx
7  spec:
8    ports:
9    - port: 80
10     protocol: TCP
11     targetPort: 80
12   selector:
13     app: nginx
```

Service - External IP

```yaml
apiVersion: v1
kind: Service
metadata:
  name: my-service
spec:
  selector:
    app: MyApp
  ports:
    - name: http
      protocol: TCP
      port: 80
      targetPort: 9376
  externalIPs:
    - 80.11.12.10
```

Use Cases:

- External DB Cluster in production
- To point a service in another namespace / cluster

Ingress

- Provides load balancing, SSL Termination and Name based virtual hosting
- Provides externally reachable URLs to Services
- Used for HTTP / HTTPS protocols

```yaml
apiVersion: networking.k8s.io/v1beta1
kind: Ingress
metadata:
  name: test-ingress
  annotations:
    nginx.ingress.kubernetes.io/rewrite-target: /
spec:
  rules:
  - http:
      paths:
      - path: /testpath
        backend:
          serviceName: test
          servicePort: 80
```

Examples

- API Services rollout
- Spring Boot with DB

Exercises

- Create deployment based on nginx image with 3 replicas
- Expose deployment via Service
- Create Ingress to access service over the internet

Exercise - Scenario

Assume that based on your recently acquired K8S expertise, you are tasked by your firm to develop real time video based facility monitoring service with below high level Objectives:

- New video service should be independent of any other earlier services (/API) developed
- Deployment should be as easy as possible
- New service should be provided to end customers via <existing-url>/video
- You are expected to use current k8s setup and extend on it

Outcome expected:

- Yaml based definitions of deployment, service and domain based routing and load balancing.

Monitoring & Misc

- Kubernetes Dashboard
- Cluster config
- Port Forwarding
- Introspection & Debugging
- Container Logging
- Best Practices
- Demo
- Practicals

Monitoring Dashboard

```
kubectl apply -f
https://raw.githubusercontent.com/kubernetes/dashboard/v2.0.0-beta8/aio/deploy/recommended.yaml

kubectl proxy
```

http://localhost:8001/api/v1/namespaces/kubernetes-dashboard/services/https:kubernetes-dashboard:/proxy/

Cluster Access

To view cluster configuration

```
kubectl config view
```

Reverse proxy to API server

```
kubectl proxy --port=8080
```

Port Forwarding

```
kubectl port-forward <pod> 7000:6379
```

```
kubectl port-forward <deployment> 7000:6379
```

```
kubectl port-forward <svc> 7000:6379
```

To access LoadBalancer service on localhost

```
minikube tunnel
```

Introspection & debugging

kubectl get pods
kubectl get pod <pod-name> -o yaml
kubectl describe <pod-name>
kubectl describe <pod-name> -o yaml
kubectl get events
kubectl get events --namespace=my-namespace (--all-namespaces)
kubectl get nodes
kubectl get node <node-name>
kubectl get node <node-name> -o yaml
kubectl describe node <node-name>
kubectl describe node <node-name> -o yaml

Shell to running container

rajesh@rajesh-Gazelle:~/git/kubernetes/debugging/shell$ kubectl apply -f shell-demo.yaml

kubectl get pod shell-demo
kubectl exec -it shell-demo -- /bin/bash

root@shell-demo:/# ls /

root@shell-demo:/# echo Hello shell demo > /usr/share/nginx/html/index.html
root@shell-demo:/# apt-get update
root@shell-demo:/# apt-get install curl
root@shell-demo:/# curl localhost

kubectl exec shell-demo env

kubectl exec -it my-pod --container main-app -- /bin/bash

Best Practices

- Configuration - specify latest stable API version
- Keep config files in version control before pushing to cluster
- Prefer YAML over JSON
- Group related objects into one file whenever it makes more sense
- Don’t specify default values unnecessarily
- Put Object descriptions as part of annotations
- Don’t use naked PODs
- Create service before deployments
- Avoid using hostPort for POD
- Use labels effectively
- Use image tag instead of using latest as the default
- Use kubectl run and expose to launch single container deployments & services

Tips for CKAD & CKA (CNCF)

Recap

- Review
- Q & A

Misc: Spring boot app monitoring

```
- name: management.endpoints.web.exposure.include
  value: "*"
- name: spring.application.name
  value: fleetLytics
- name: management.server.port
  value: "8888"
- name: management.metrics.web.server.request.autotime.enabled
  value: "true"
- name: management.metrics.tags.application
  value: fleetLytics
```

```
annotations:
  prometheus.io/scrape: "true"
  prometheus.io/port: "8888"
  prometheus.io/path: /actuator/prometheus
```

```
<dependency>
        <groupId>org.springframework.boot</groupId>
        <artifactId>spring-boot-starter-actuator</artifactId>
</dependency>
<dependency>
        <groupId>io.micrometer</groupId>
        <artifactId>micrometer-registry-prometheus</artifactId>
</dependency>
```

Thank You for your active participation!

info@brainupgrade.in

https://www.facebook.com/brainupgrade.in

https://www.linkedin.com/company/brains-upgrade/

https://www.linkedin.com/in/rajesh-gheware/