Kubernetes Foundations

Part One

Version 1.1.0

1-23-2020PublicTraining

About Strigo

- Strigo is a web-based platform that provides the classroom environment for our courses.
- Let's walk through the features of the platform.
- We'll also get your lab environment initialized

7-23-2020PublicTraining

Introductions

- About your instructor(s)

- Tell us about yourself
  - Would you classify yourself as a
    - developer
    - systems administrator
    - architect
    - It depends on the day/hour
  - What are your goals for learning and adopting Kubernetes?

7-23-2020Public Training

Agenda

Part One

1. Introduction to Containers
2. Kubernetes Fundamentals
3. Kubernetes Architecture & Troubleshooting
4. Deployment Management
5. Pod & Container Configurations

7-23-2020 Public Training

Agenda

Part Two

6. Kubernetes Networking
7. Resource Organization
8. Storage & Stateful Applications
9. Dynamic Application Configuration
10. Additional Workloads
11. Security

7-23-2020PublicTraining

Course Format

- This is a lab-intensive, hands-on course
- Each section will begin with the introduction of a new concept
- Each section contains a lab exercise where you will explore each new concept

7-23-2020PublicTraining

Introduction to Containers

Chapter 01

7-23-2020PublicTraining

Agenda

Part One

1. Introduction to Containers
2. Kubernetes Fundamentals
3. Kubernetes Architecture & Troubleshooting
4. Deployment Management
5. Pod & Container Configurations

7-23-2020 Public Training

Cloud Native Principles

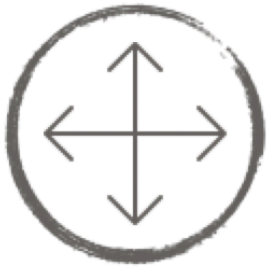
- Container Packaged
  - isolated unit of work that does not require OS dependencies
- Dynamically Managed
  - actively scheduled and managed by an orchestration process
- Microservice Oriented
  - Loosely coupled with dependencies explicitly described

7-28-2020PublicTraining

Why Containers?

Velocity

enables business to develop and roll out new offerings faster



Portability

predictable execution in any linux based environment. Move from your laptop to your datacenter to the cloud

Reliability

hermetically sealed image makes production deployments very simple and less error-prone

Efficiency

maximize resource utilization

Self-Service

developer productivity

Isolation

avoid dependency conflict

7-23-2020 Public Training

Containers ≠ VMs

Virtual Machines

- Full OS on virtual hardware
- Deployed as a unit
- Require a hypervisor
- VM performance critical to cloud performance
- Overhead cannot be removed
- Hotplugging requires support from guest OS

Containers

- Decouple application and OS
- Can be composed together
- Modifies existing OS to provide isolation
- A concept built on kernel namespace feature
- Contain as little as one process
- An application container consumes less RAM
- Operate at a higher abstraction level, offer more insight into behaviors without deploying additional agents

7-23-2020Public Training

Isolation - All Levels

- Cost to create, store, and run a container is very low

7-23-2020 Public Training

Programmatic Construction

- VM Creation is not guaranteed to be scripted/reproducible
- Container build process enforces reproducibility

```
FROM ubuntu:16.04
RUN apt-get update && apt-get install -y apache2
RUN apt-get clean && rm -rf /var/lib/apt/lists/*
ENV APACHE_RUN_USER www
ENV APACHE_RUN_GROUP www
ENV APACHE_LOG_DIR /var/log/apache2
EXPOSE 80
CMD ["/usr/sbin/apache2", "-D", "FOREGROUND"]
```

7-23-2020PublicTraining

Image vs Container

- An image is the result of a build
- A container is a running instance from an image

```
$ docker images
REPOSITORY      TAG         IMAGE ID          CREATED             SIZE
busybox         latest      8c811b4aec35      2 weeks ago         1.15 MB
mysql           latest      a8a59477268d      4 weeks ago         445 MB
foo.com/mysql   latest      a8a59477268d      4 weeks ago         445 MB

$ docker ps
CONTAINER ID  IMAGE  COMMAND    CREATED              STATUS              PORTS      NAMES
Acc5c8f58392  mysql  “mysqld”   About a minute ago   Up About a minute   3306/tcp   myWebsiteDB
```

7-23-2020PublicTraining

Image/Container – VM comparison

- Model similar to linux process invocation
  - **image** – one immutable set of bits on disk
  - **container** – running instances launched from the image
  - conceptual like running multiple processes of a command line process at once
- Container filesystem is ephemeral
  - promotes cloud native and immutable deployments

7-23-2020PublicTraining

Starting and Stopping Containers

`docker run`

- start a new container from an image

`docker stop`

- stop a running container

`docker rm`

- delete a container (must be stopped)
- add `-f` to both stop and remove a container

7-23-2020PublicTraining

Managing Images

`docker images`

- display a list of images on the machine

`docker rmi`

- delete an image

`docker build`

- build an image from a Dockerfile

`docker tag`

- add tags to an image

`docker pull/push`

- pull and push images to/from a registry

7-23-2020PublicTraining

Docker Troubleshooting Commands

`docker ps`

- retrieve a list of running containers
- add `-a` to include non-running containers

`docker logs`

- view a container's log output

`docker exec`

- run a command within a container
- can also start a shell within a container (if available)

1-23-2020Public Training

Container Registry

- A container registry provides a central location for container image storage
- Types of registries
  - Hosted: Docker Hub, Google Container Registry
  - Self-Hosted: Artifactory, Harbor, Quay
- Images are built on a build host and pushed to a registry

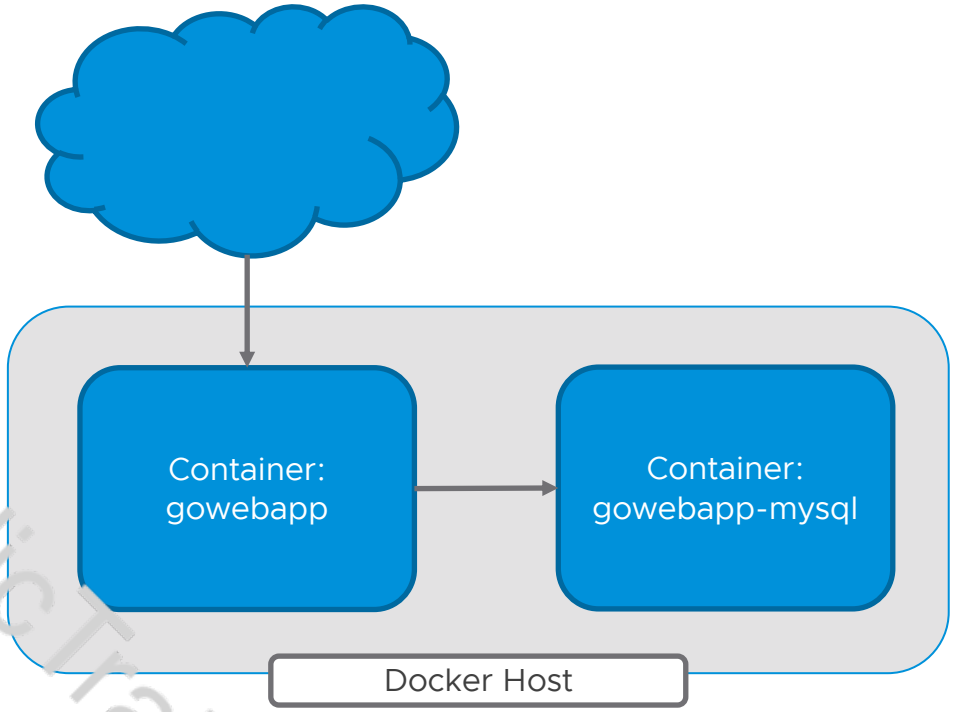
7232020PublicTraining

What We're Building

Deploying Go Web App using Docker

Go Web App:

- A web–based notepad application
- Frontend written in Golang
- MySQL database backend



7-23-2020PublicTraining

Lab 01

Containerize Applications

Prepare your lab

Write Dockerfile for frontend web app

Write Dockerfile for backend data store

Build, test, and publish Docker images

7-23-2020PublicTraining

Kubernetes Fundamentals

Chapter 02

7-23-2020PublicTraining

Agenda

Part One

1. Introduction to Containers
2. Kubernetes Fundamentals
3. Kubernetes Architecture & Troubleshooting
4. Deployment Management
5. Pod & Container Configurations

7-23-2020 Public Training

Cloud Native Principles

- Container Packaged
  - isolated unit of work that does not require OS dependencies
- Dynamically Managed
  - actively scheduled and managed by an orchestration process
- Microservice Oriented
  - Loosely coupled with dependencies explicitly described

7-29-2020PublicTraining

About Kubernetes

- Open-source cluster management tool
  - Automates: Deploying, managing, scaling applications
- Managed by Cloud Native Computing Foundation (CNCF); originally created at Google
- Under active development by a well-supported community, including former Borg engineers
- Lessons learned from Borg in production for more than a decade
- Can run on both bare metal and on various cloud providers

4-23-2020Public Taining

Imperative vs. Declarative

Imperative	Declarative
defines actions	defines desired state

7-23-2020 Public Training

Kubernetes Concepts

- Pods
- Services
- Deployments

7-23-2020PublicTraining

Why Pods?

- Legacy applications
  - new apps have luxury of implementing many features within the app
- Separation of Concerns
  - isolate logic and changes to different apps
- Reusability
  - if designed appropriately containers can be re-used by different apps

7-28-2020PublicTraining

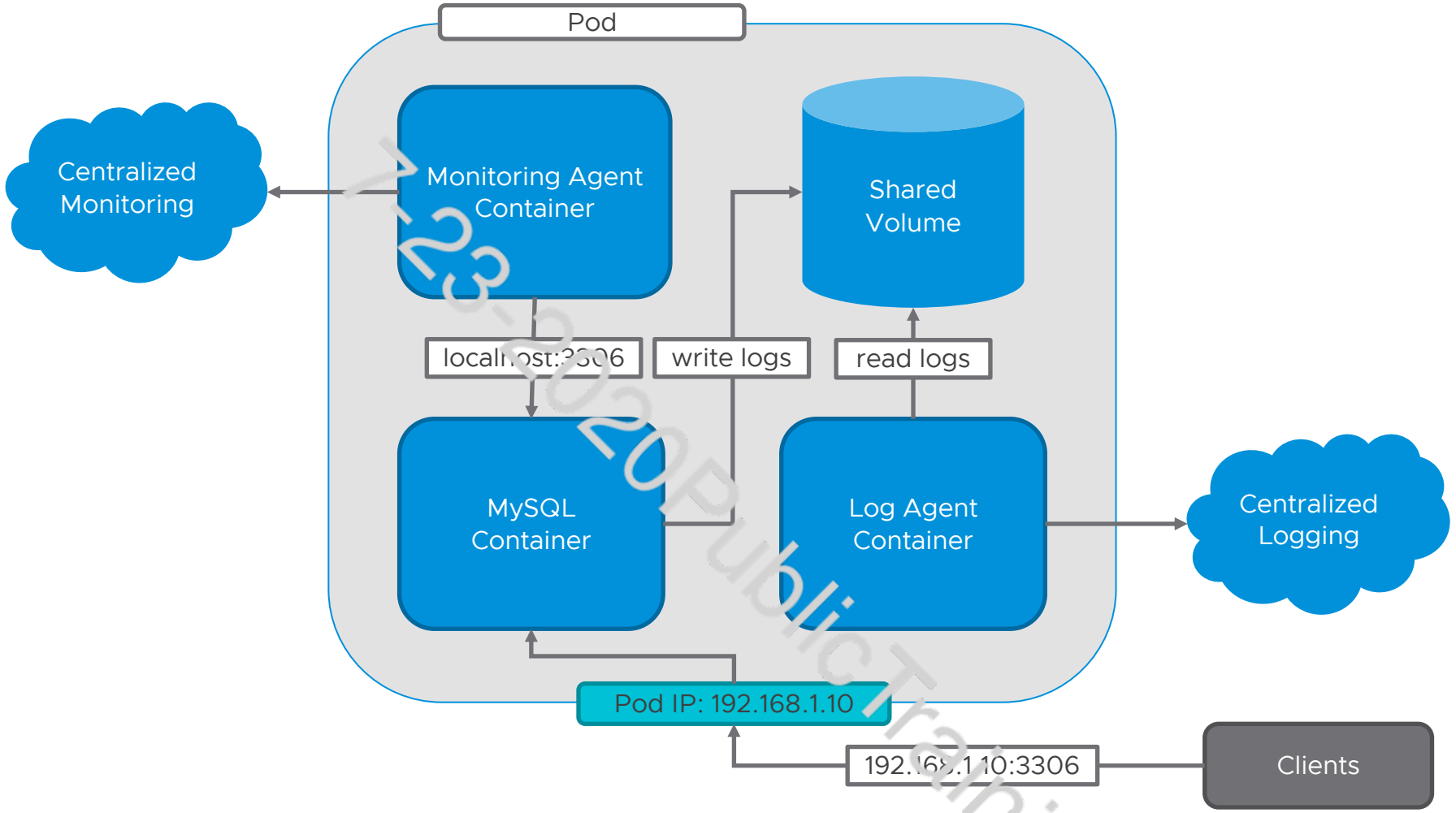
Pod Characteristics

- One or more containers per pod
  - the pod's existence/purpose is usually because of one particular container

- Assumptions
  - all containers for a pod will run on the same node
  - containers within a pod can talk to each other over localhost
  - containers can share volume resources

7-23-2020PublicTraining

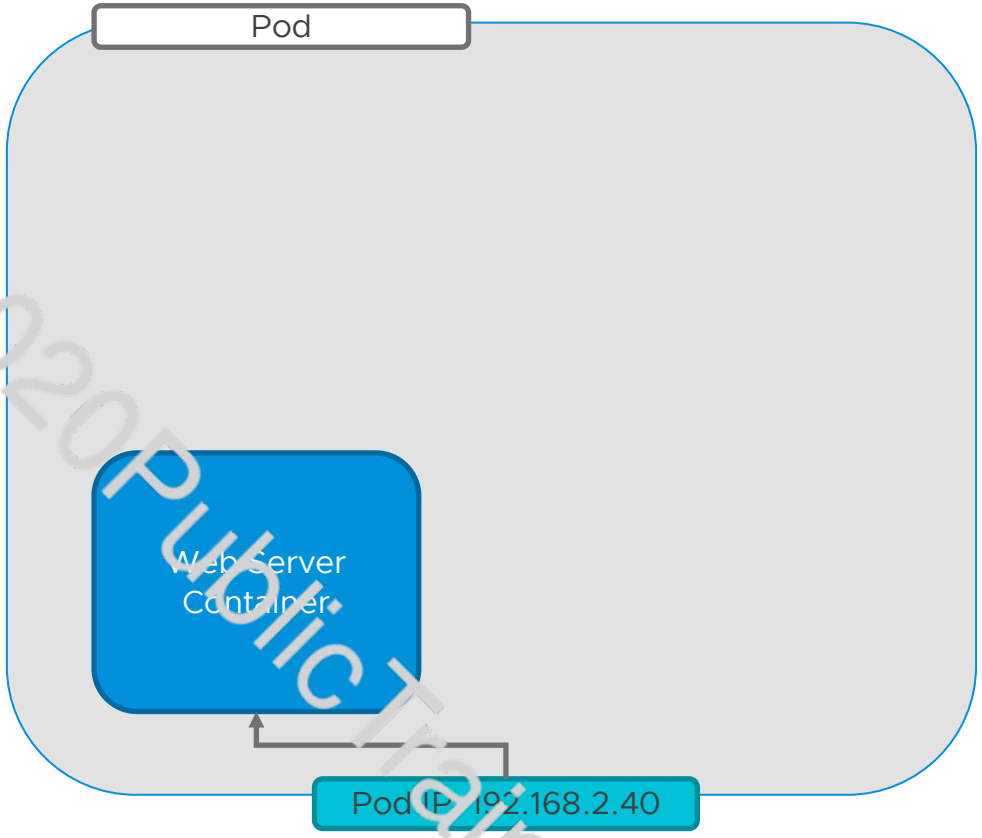
Containers in a Pod



V123-2020PublicTraining

Pod - Main Container

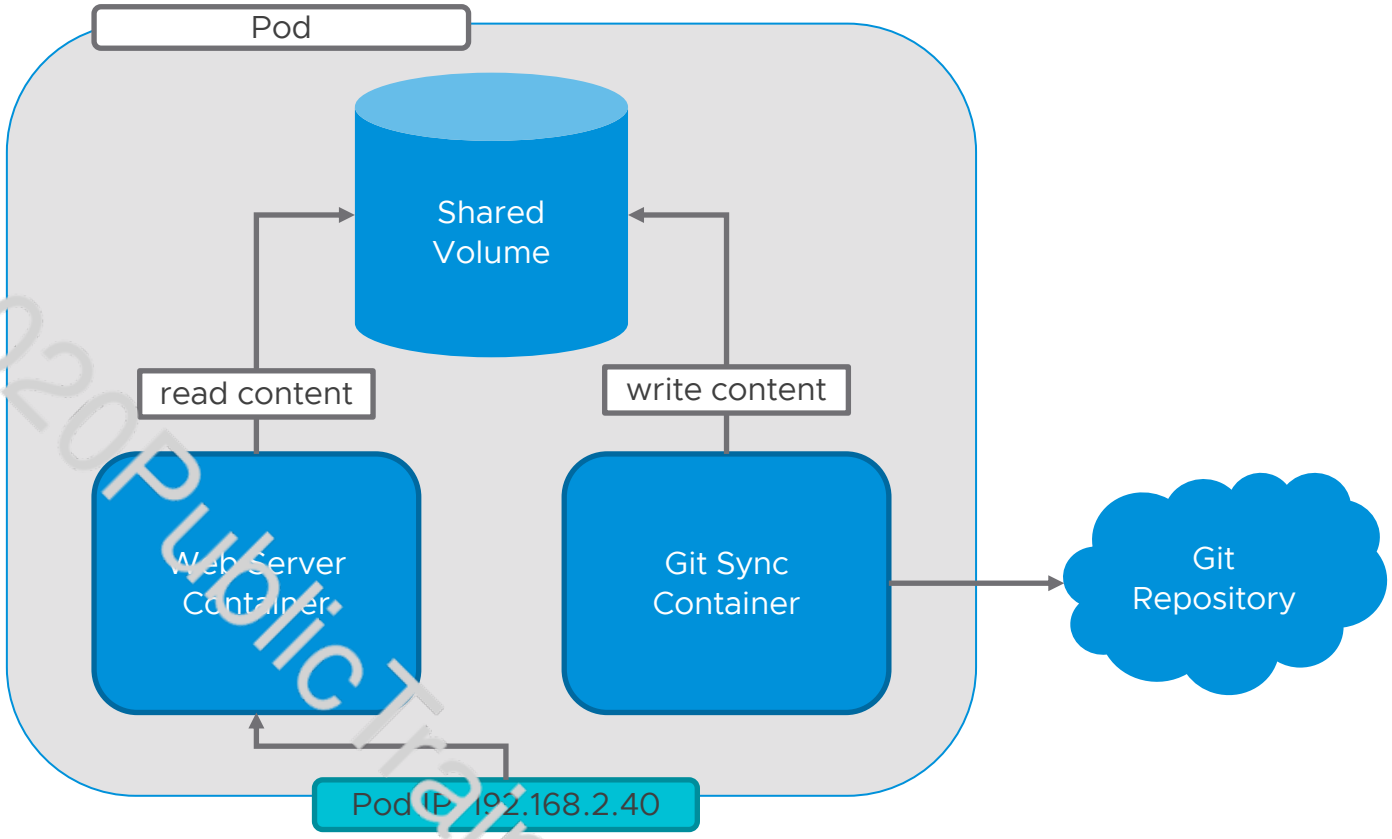
- main container is the reason a Pod exists



7-23-2020 Public Training

Pod - Sidecar Containers

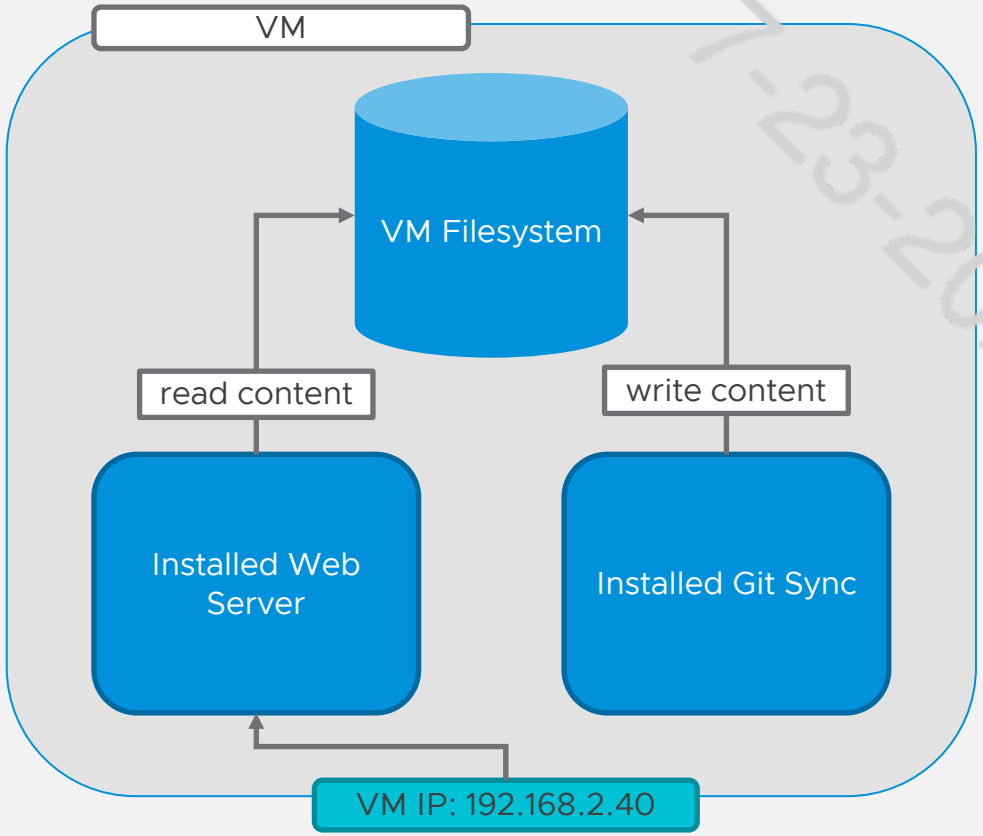
- sidecars help and assist functions
  - logic could be included in main container
  - separation allows for isolation and reuse
  - each container can have separate resource allocations



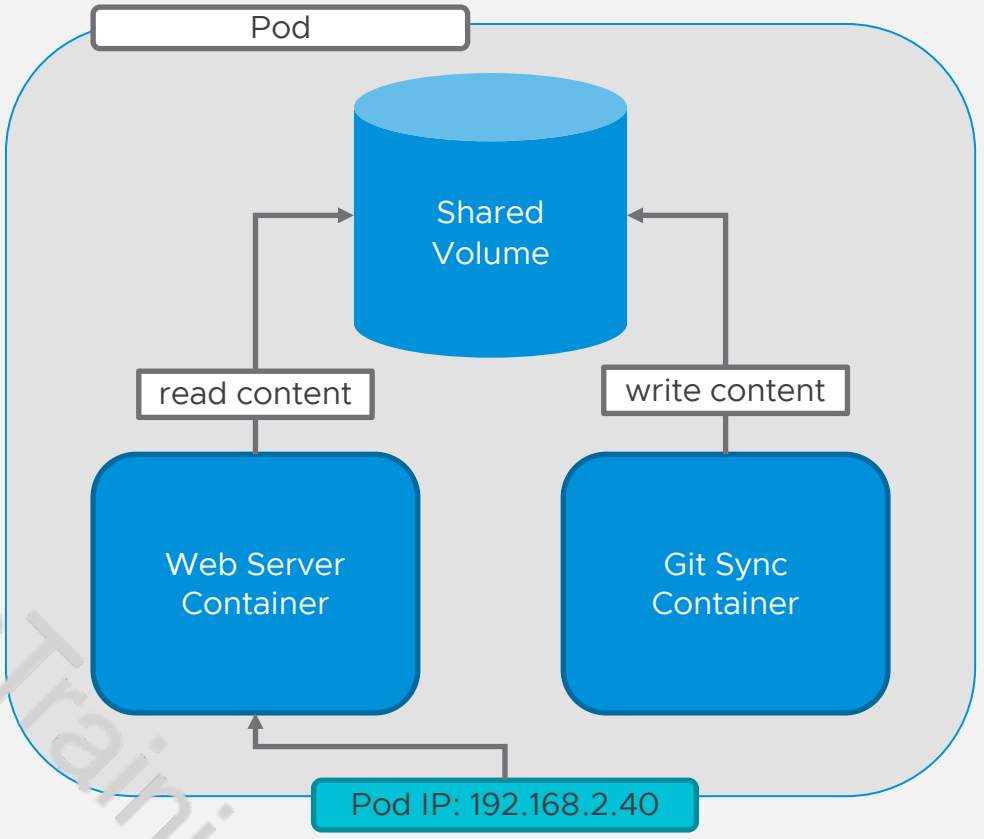
7-23-2020 Public Training

Similar Concepts

Virtualization



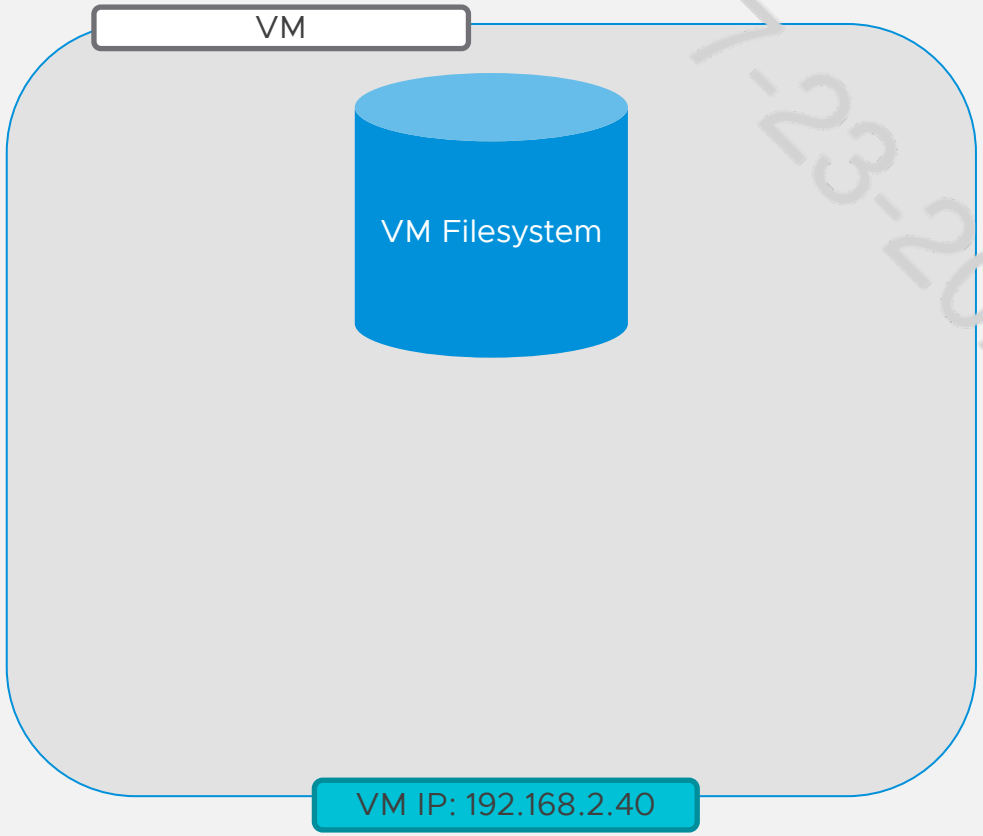
Kubernetes



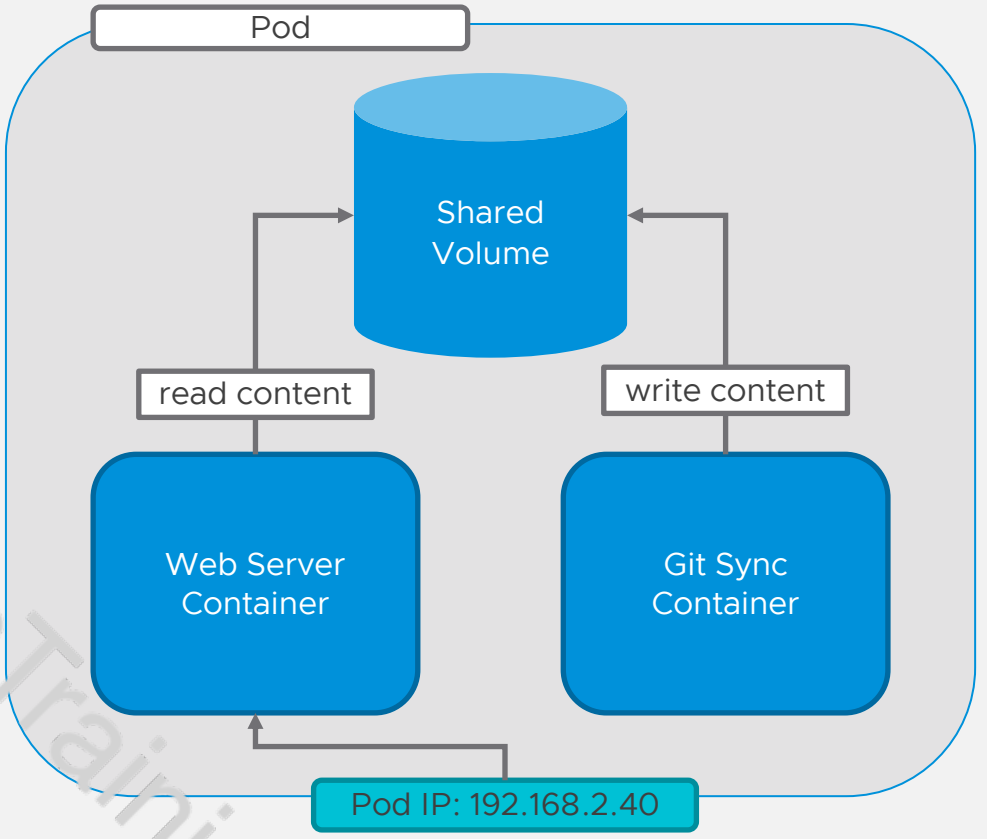
7-23-2020PublicTraining

Responsibility and Standardization

Virtualization



Kubernetes



7-23-2020PublicTraining

API - REST

- REST / JSON based
  - all internal and external components communicate via this API
  - explicitly versioned to allow breaking changes
  - provides abstraction layer on top of resource storage (etcd)

```
curl http://localhost:8001/api/v1/pods

{
  "kind": "PodList",
  "apiVersion": "v1",
  "metadata": {
    "selfLink": "/api/v1/pods",
    "resourceVersion": "3606680"
  },
  "items": [
```

7-23-2020PublicTraining

API - kubectl

- command line interface for the API
  - calls one or more REST API calls for each command line invocation
  - does contain some business logic not in the API

```
$ kubectl get pods

NAME                    READY   STATUS    RESTARTS   AGE
blog-56d5c9dbf7-6j67j   1/1     Running   0          19s
blog-56d5c9dbf7-dh7vb   1/1     Running   0          19s
blog-56d5c9dbf7-zv7d7   1/1     Running   0          19s
```

7-23-2020PublicTraining

Yet Another Markup Language (YAML)

```yaml
employeeNumber: 123
name:
  firstName: Jenny
  lastName: Smith
phones:
- label: work
  number: 555-555-5555
- label: mobile
  number: 267-867-5309
```


```json
{
  "employeeNumber": 123,
  "name": {
    "firstName": "Jenny",
    "lastName": "Smith"
  },
  "phones": [
    {
      "label": "work",
      "number": "555-555-555"
    },
    {
      "label": "mobile",
      "number": "267-867-5309"
    }
  ]
}
```


```xml
<employee id="123">
  <name>
      <firstName>Jenny</firstName>
      <lastName>Smith</lastName>
  </name>
  <phones>
    <phone label="work">
        <number>555-1212</number>
    </phone>
    <phone label="mobile">
        <number>867-5309</number>
    </phone>
  </phones>
</employee>
```

7-23-2020PublicTraining

API Resource Objects & YAML

- Often represented in YAML (can also use JSON)
- Represent API objects

```
apiVersion: v1
kind: Pod
metadata:
    name: mypod1
    labels:
      app: blog
spec:
    containers:
    - name: nginx
      image: nginx:1.13.1
    - name: www-syncer
      image: syncer:1.2
```

123-2020Public Training

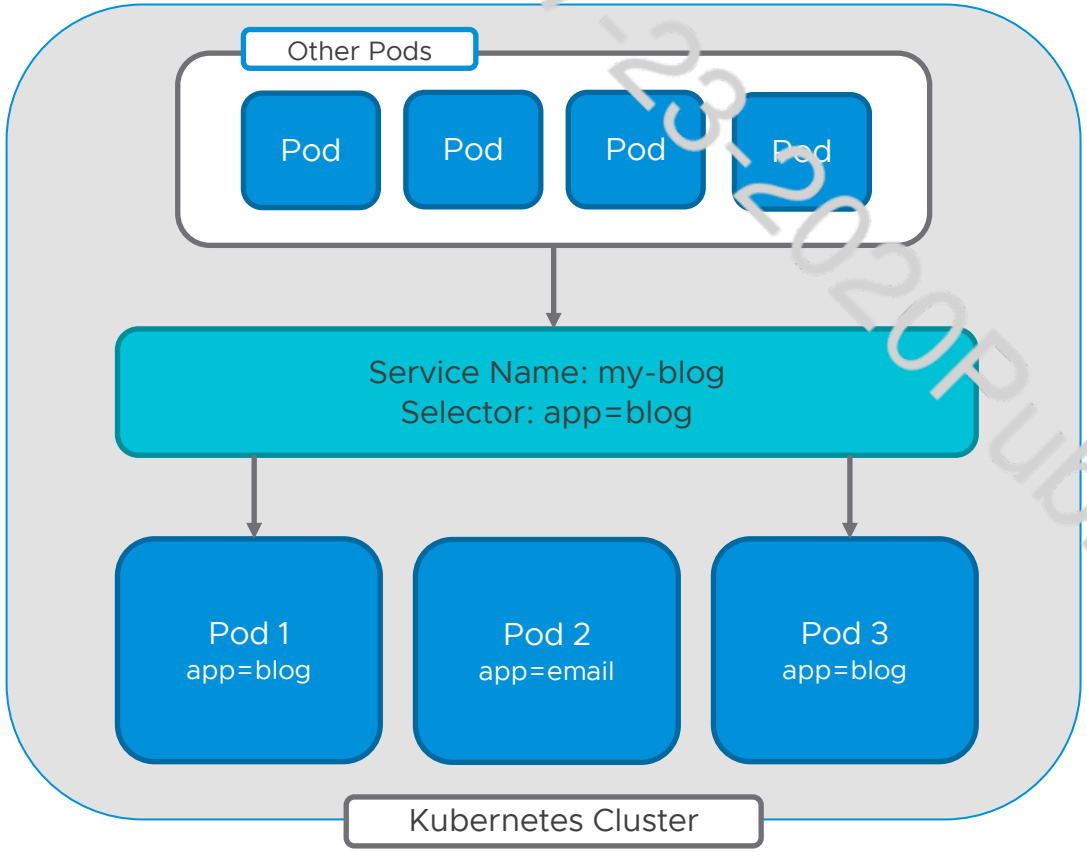
Creating Resources

- Declarative Approach (Preferred)
  - kubectl apply -f [<file> | <directory> | <url>]
    - Create new or update existing resource(s) from file(s)
- Imperative Approach
  - kubectl create - create new resource from a file
  - kubectl replace - update an existing resource from a file
  - kubectl edit - update existing resource using your default editor
  - kubectl patch - update existing resource by merging a code snippet

7-23-2020 Public Training

Services

- Load balancing for pods
- Use labels to determine target pods




```
apiVersion: v1
kind: Service
metadata:
  name: my-blog
  labels:
    app: blog
spec:
  selector:
    app: blog
  ports:
  - protocol: TCP
    port: 80
    targetPort: 80
```

7-23-2020PublicTraining

Pod Creation & Deployments

- What if I have a website that needs 20 NGINX servers?

7-23-2020PublicTraining

Pod Creation & Deployments

- What if I have a website that needs 20 NGINX servers?
  - 20 pod objects sent to Kubernetes
  - each object is exactly the same except for the name

- We need a higher level construct…

7-23-2020PublicTraining

Deployments

- Single object that will create other resources
- Pod spec is nested
- Leverages selectors

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: my-blog
spec:
  replicas: 3
  selector:
    matchLabels:
      app: blog
  template:
    metadata:
      labels:
        app: blog
    spec:
      containers:
      - name: nginx
        image: nginx:1.7.9
        ports:
        - containerPort: 80
```

7-23-2020Public Training

Labels

- Characteristics
  - map of key / value pairs
  - both organizational and functional (selectors on Services)
  - indexed and searchable
- Tips
  - avoid compound label values:
    - `app: blog-frontend` vs `app: blog / tier: frontend`
  - try to standardize on key/values across the cluster

LFS-2020-PublicTraining

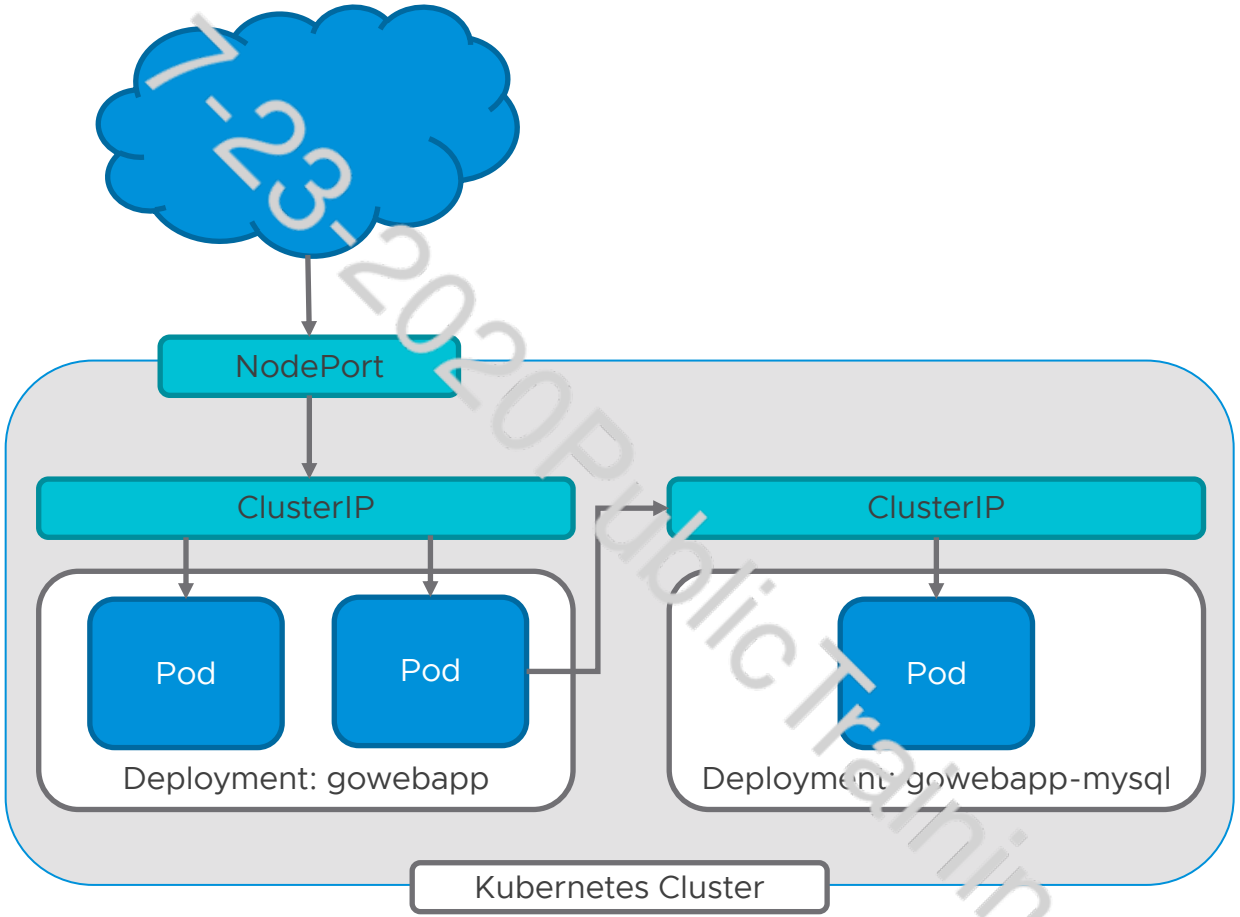
Kubernetes Documentation

- Kubernetes Docs
  - https://docs.kubernetes.io
- Kubernetes API Reference
  - https://kubernetes.io/docs/reference/
- kubectl Docs
  - https://kubectl.docs.kubernetes.io/

7-28-2020PublicTraining

What We're Building

Deploying Go Web App to Kubernetes



2020 Public Training

Lab 02

Using Kubectl

Getting familiar with kubectl

Deploy Applications

Write YAML files for Services

Write YAML files for Deployments

Deploy the Applications

7-23-2020PublicTraining

Kubernetes Architecture & Troubleshooting

Chapter 03

7-23-2020PublicTraining

Agenda

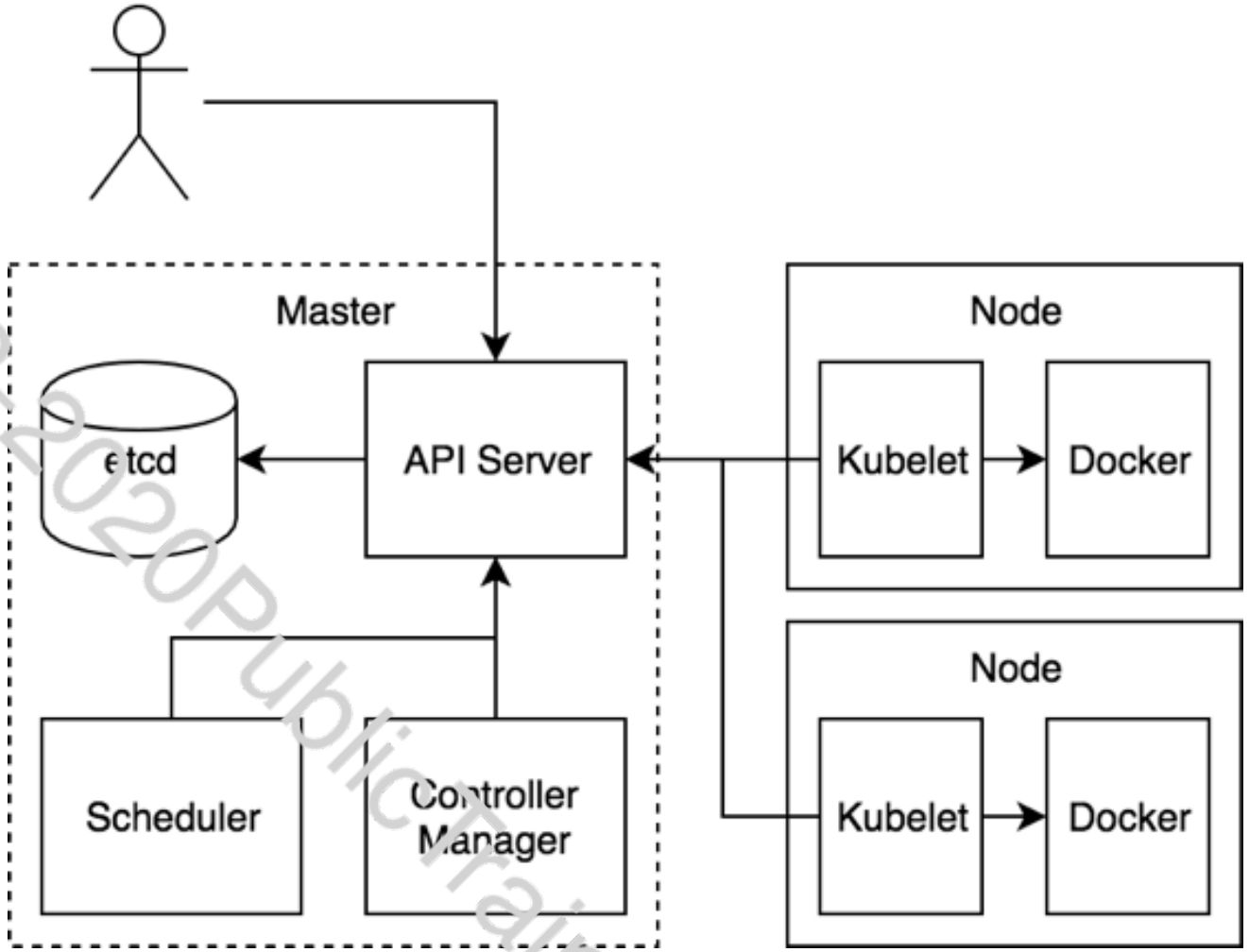
Part One

1. Introduction to Containers
2. Kubernetes Fundamentals
3. Kubernetes Architecture & Troubleshooting
4. Deployment Management
5. Pod & Container Configurations

7-23-2020 Public Training

Kubernetes Architecture

- Worker Nodes
  - kubelet
  - kubeproxy
  - container engine
- Control Plane
  - etcd
  - API server
  - Scheduler
  - Controller Manager



7-23-2020PublicTraining

Kubernetes Troubleshooting – Event Log

- The event log records events on the cluster
  - May contain helpful troubleshooting information
  - By default, it stores an hour of history

```
$ kubectl get events

LAST SEEN   TYPE     REASON                   KIND   MESSAGE
2m14s       Normal   NodeHasSufficientMemory  Node   Node master status is now: NodeHasSufficientMemory
2m14s       Normal   NodeHasNoDiskPressure    Node   Node master status is now: NodeHasNoDiskPressure
2m14s       Normal   NodeHasSufficientPID     Node   Node master status is now: NodeHasSufficientPID
2m14s       Normal   NodeAllocatableEnforced  Node   Updated Node Allocatable limit across pods
51s         Normal   RegisteredNode           Node   Node master event: Registered Node master in Controller
38s         Normal   Starting                 Node   Starting kube-proxy.
```

7-23-2020Public Training

Getting a list of objects

- Use `kubectl get` to retrieve a list of objects
  - Add `-o wide` for more details about each object

```
$ kubectl get pods –o wide

NAME                            READY   STATUS    RESTARTS   AGE     IP            NODE
gowebapp-54d74586cc-jqlmx       1/1     Running   0          3m5s    192.168.0.7   master
gowebapp-54d74586cc-sqpfk       1/1     Running   0          3m5s    192.168.0.6   master
gowebapp-mysql-9746475d-gdhtx   1/1     Running   0          3m35s   192.168.0.5   master

# When using –o wide with `kubectl get pods` the output will include the Pod’s IP address, and which
# node it’s running on.
```

7-23-2020PublicTraining

Getting more information about an object

- Use kubectl describe to view details about an object
  - The output includes event log entries generated by that object

```
$ kubectl describe pod gowebapp-54d74586cc-jqlmx

Name:               gowebapp-54d74586cc-jqlmx
Namespace:          default
Node:               master/172.31.30.8
Start Time:         Wed, 08 May 2019 15:17:46 +0000
Labels:             app=gowebapp
                    tier=frontend
Status:             Running
IP:                 192.168.0.7
Controlled By:      ReplicaSet/gowebapp-54d74586cc
```

7-28-2020PublicTraining

kubectl describe (continued)

```
Containers:
  gowebapp:
    Container ID:   docker://ae0321def949ed866fb0c303207eaa0d3188dd9830f648d52196418637bda13d
    Image:          localhost:5000/gowebapp:v1
    Image ID:       docker-pullable://localhost:5000/gowebapp@sha256:8227…
Events:
  Type    Reason     Age   From               Message
  ----    ------     ----  ----               -------
  Normal  Scheduled  11m   default-scheduler  Successfully assigned default/gowebapp-54d7458… to master
  Normal  Pulled     11m   kubelet, master    Container image "localhost:5000/gowebapp:v1" already present…
  Normal  Created    11m   kubelet, master    Created container
  Normal  Started    11m   kubelet, master    Started container
```

1-23-2020PublicTraining

Getting pod logs

- Use `kubectl logs` to retrieve log output from a container
  - Use `-c <container_name>` if the pod contains more than one container

```
$ kubectl logs gowebapp-54d74586cc-jqlmx

2019-05-08 03:17:48 PM Running HTTP :80
2019-05-08 03:41:09 PM 172.31.30.8:56701 GET /
2019-05-08 03:41:09 PM 172.31.30.8:56701 GET /static/css/bootstrap.min.css?1544384023
2019-05-08 03:41:09 PM 172.31.30.8:56703 GET /static/css/global.css?1544384023
2019-05-08 03:41:09 PM 172.31.30.8:56705 GET /static/js/underscore-min.js?1544384023
2019-05-08 03:41:09 PM 172.31.30.8:56707 GET /static/js/global.js?1544384023
2019-05-08 03:41:09 PM 172.31.30.8:56701 GET /static/favicons/favicon-196x196.png
2019-05-08 03:41:11 PM 172.31.30.8:56701 GET /login
2019-05-08 03:41:17 PM 172.31.30.8:56701 GET /register
```

7-23-2020 Public Training

Run commands inside a pod

- Use `kubectl exec` to run commands within a pod
  - Use `-c <container_name>` if the pod contains more than one container
  - can also start a shell within a container (if available)

```
$ kubectl exec gowebapp-54d74586cc-jqlmx -- ls -l /opt/gowebapp

total 12604
drwxr-xr-x 2 root root     4096 May  8 15:16 config
-rwxr-xr-x 1 root root 12890938 Dec  9 22:44 gowebapp
drwxr-xr-x 6 root root     4096 Dec  9 19:33 static
drwxr-xr-x 8 root root     4096 May  8 15:12 template
```

7-23-2020 Public Training

Lab 03

Kubernetes Troubleshooting

Explore Kubernetes Event Log

Explore Deployments and Services

Troubleshoot Deployments

Troubleshoot Services

7-23-2020PublicTraining

Deployment Management

Chapter 04

7-23-2020PublicTraining

Agenda

Part One

1. Introduction to Containers
2. Kubernetes Fundamentals
3. Kubernetes Architecture & Troubleshooting
4. Deployment Management
5. Pod & Container Configurations

7-23-2020 Public Training

Deployment Strategies

7-23-2020PublicTraining

Deployments - ReplicaSets

- One or more ReplicaSets
  - deployments handle the creation/modification of ReplicaSets
  - pods for a deployment span across all valid ReplicaSets
  - do not manually manage ReplicaSets that are owned by a Deployment

```
$ kubectl get deployment nginx -o wide

NAME    DESIRED CURRENT  UP-TO-DATE  AVAILABLE  AGE  CONTAINER(S)  IMAGE(S)
nginx   3       3        3           3          23m  nginx         nginx:1.13.1
```

1-23-2020PublicTraining

Deployment Strategies - Built-In

- RollingUpdate
  - default if not defined
  - new ReplicaSet is created, then scaled up as the old ReplicaSet is scaled down
- Recreate
  - removes all existing pods in the existing ReplicaSet first
  - then creates new pods in the new ReplicaSet

1-23-2020PublicTraining

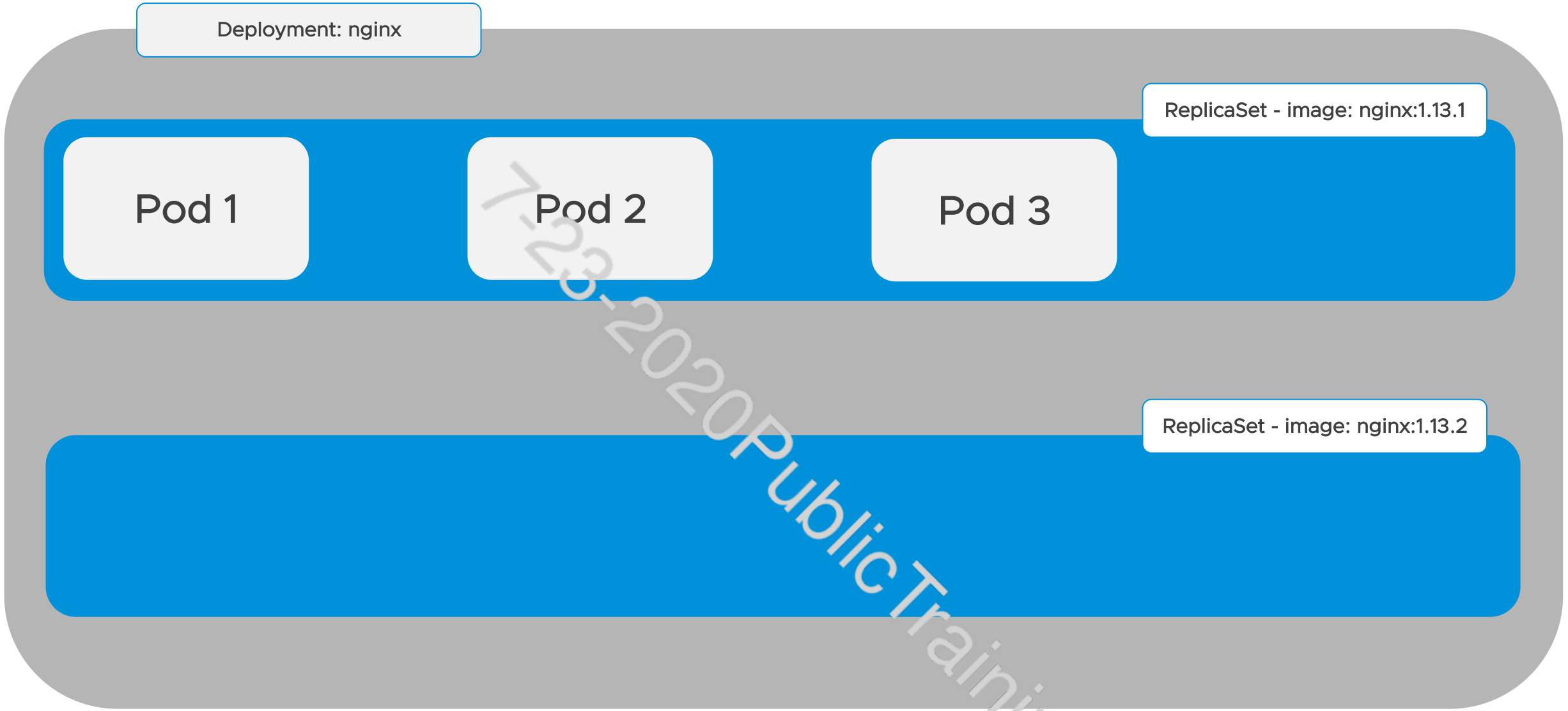
Deployment Strategies - RollingUpdate

7-23-2020PublicTraining

Deployment Strategies - RollingUpdate

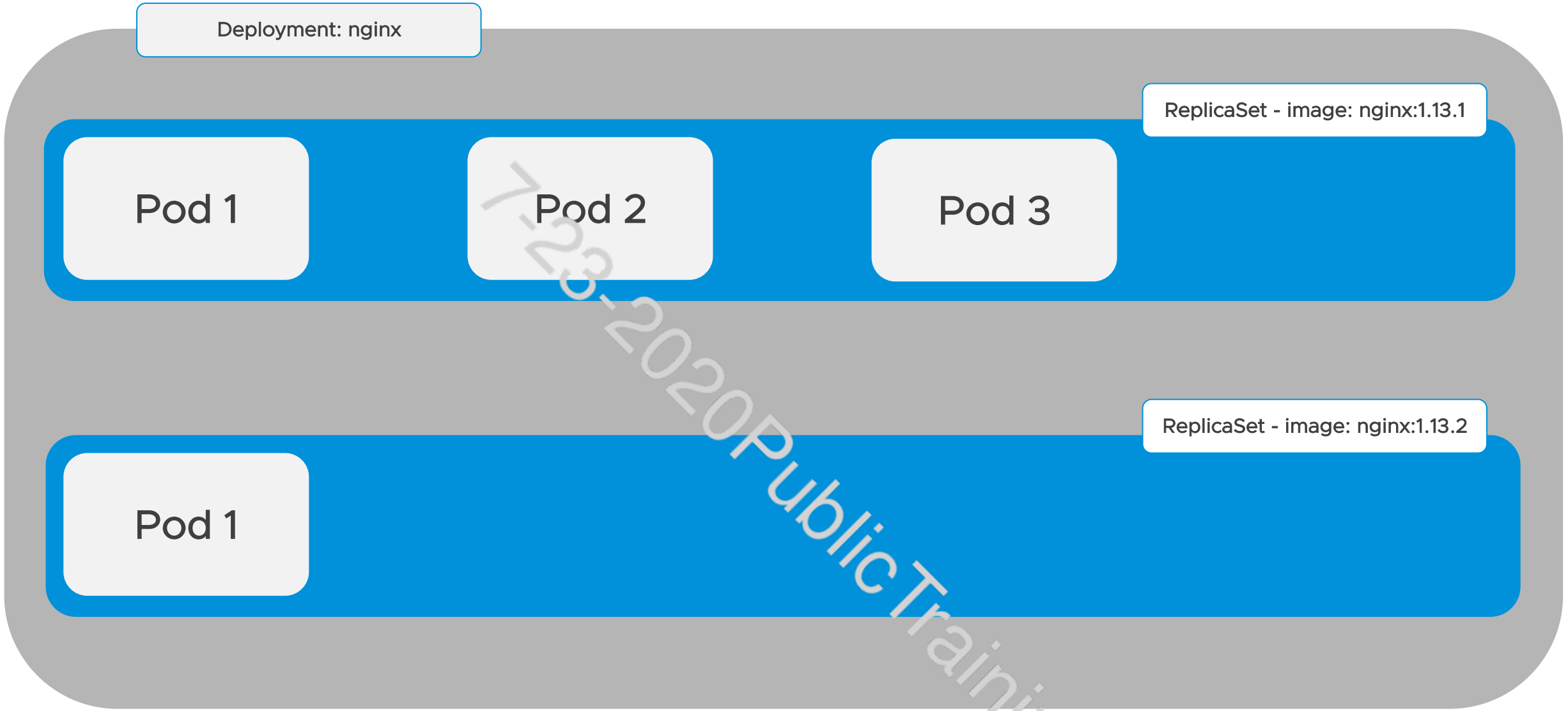
7-23-2020PublicTraining

Deployment Strategies - RollingUpdate



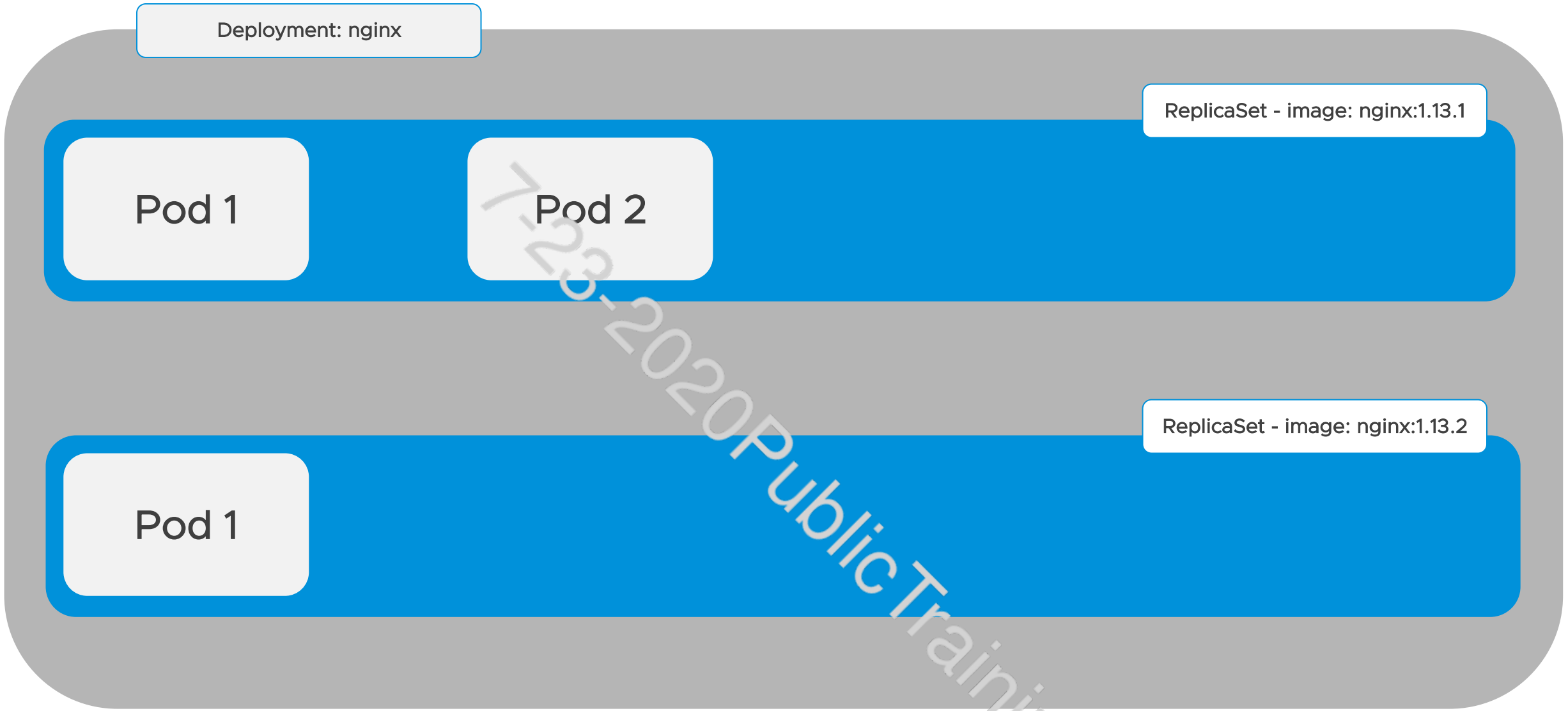
7-23-2020PublicTraining

Deployment Strategies - RollingUpdate



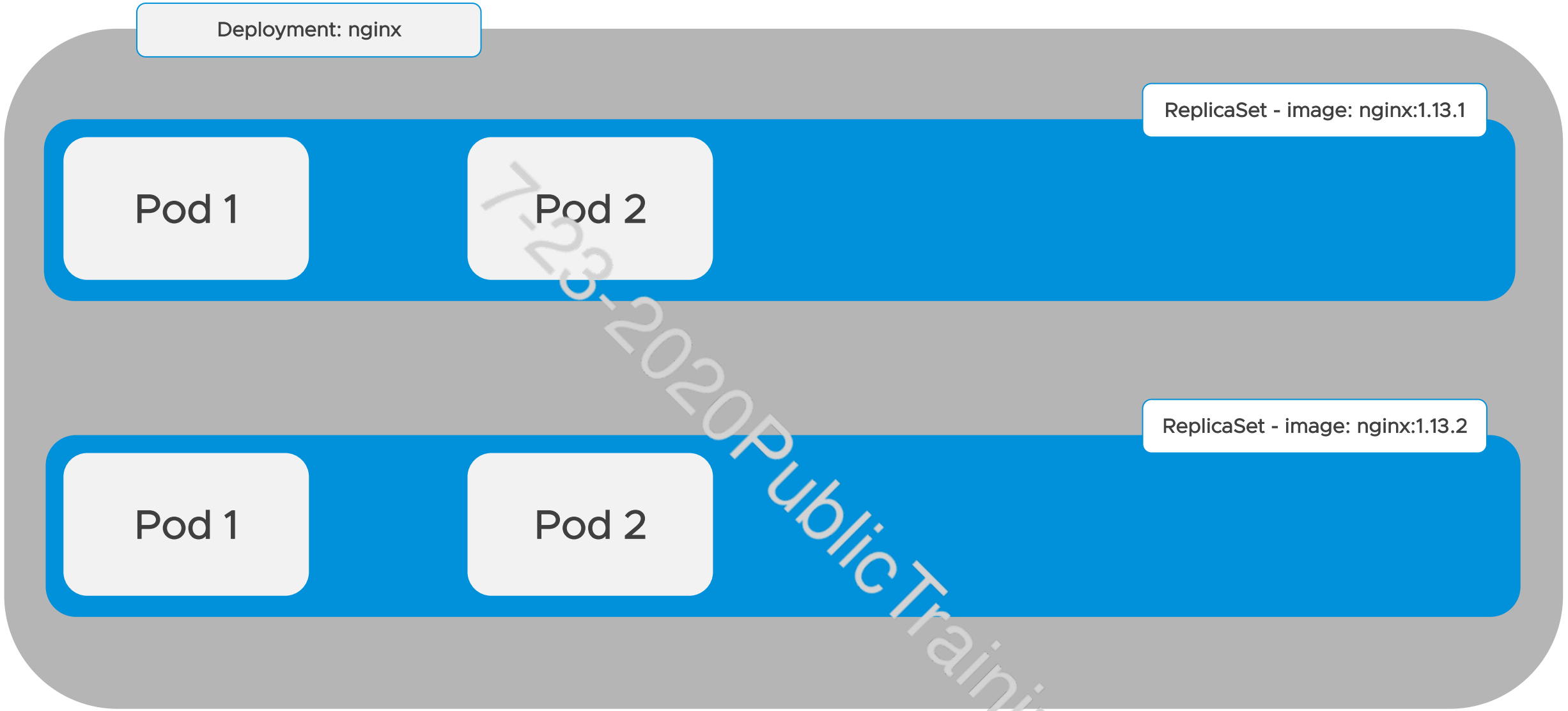
7-23-2020PublicTraining

Deployment Strategies - RollingUpdate



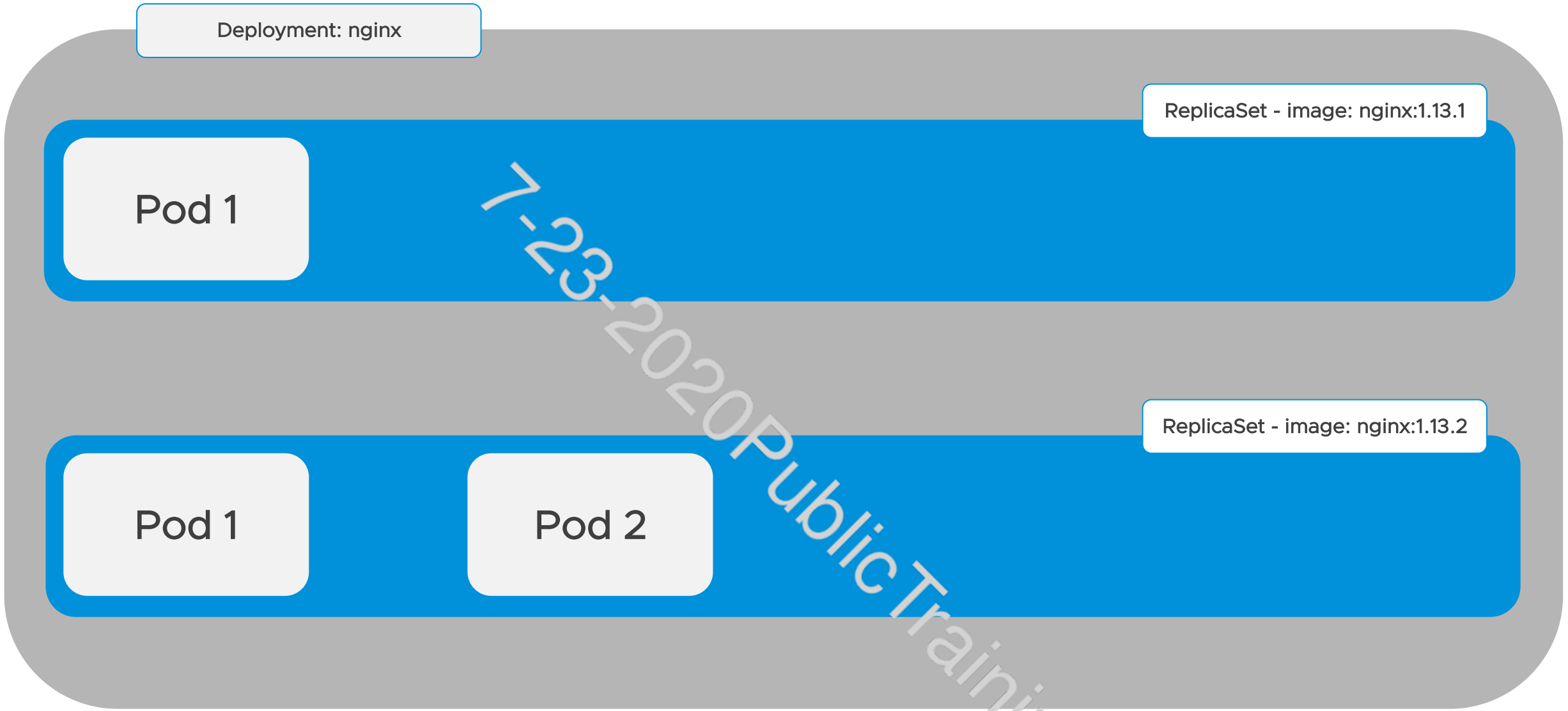
7-23-2020PublicTraining

Deployment Strategies - RollingUpdate



7-23-2020PublicTraining

Deployment Strategies - RollingUpdate



7-23-2020PublicTraining

Deployment Strategies - RollingUpdate

7-23-2020PublicTraining

Deployment Strategies - RollingUpdate

7-23-2020PublicTraining

Deployment Strategies - Recreate

7-23-2020PublicTraining

Deployment Strategies - Recreate

7-23-2020PublicTraining

Deployment Strategies - Recreate

7-23-2020PublicTraining

Deployment Strategies - Recreate

7-23-2020PublicTraining

Deployment Strategies - Recreate

7-23-2020PublicTraining

Deployment Strategies - Recreate

7-23-2020PublicTraining

Deployment Strategies - Recreate

7-23-2020PublicTraining

Deployment Strategies - Recreate

7-23-2020PublicTraining

Deployment Strategies - Others

- Canary
  - deploy new container to subset of traffic
  - involves two Deployments and label management between them

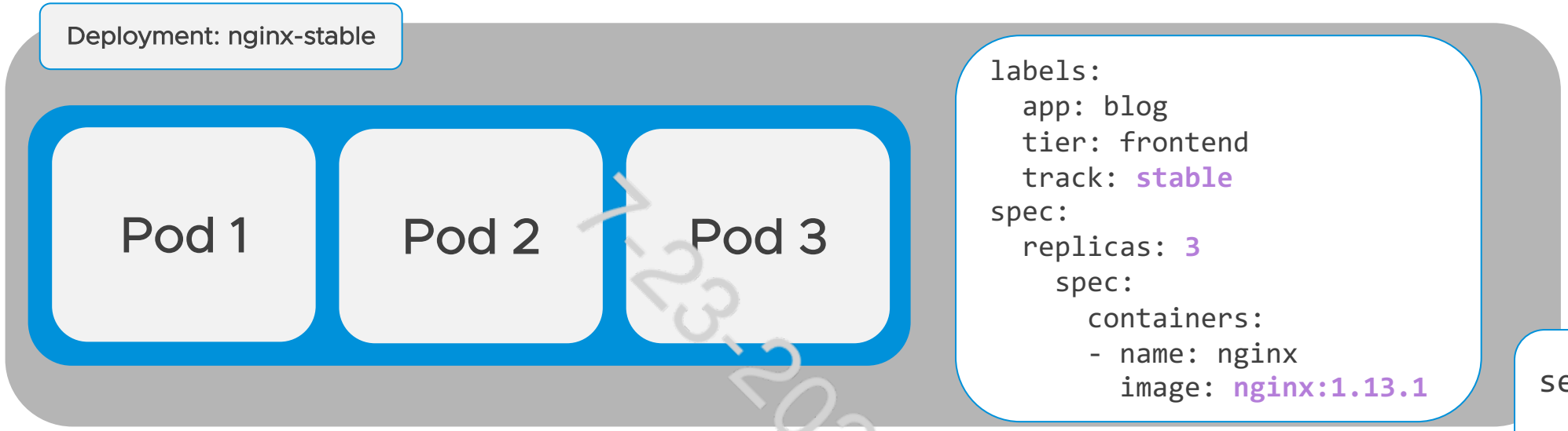
- Blue/Green
  - deploy all new containers, test, then "flip/switch" traffic to new containers
  - involves two Deployments and label management between them

1-23-2020PublicTraining

Deployment Strategies - Canary

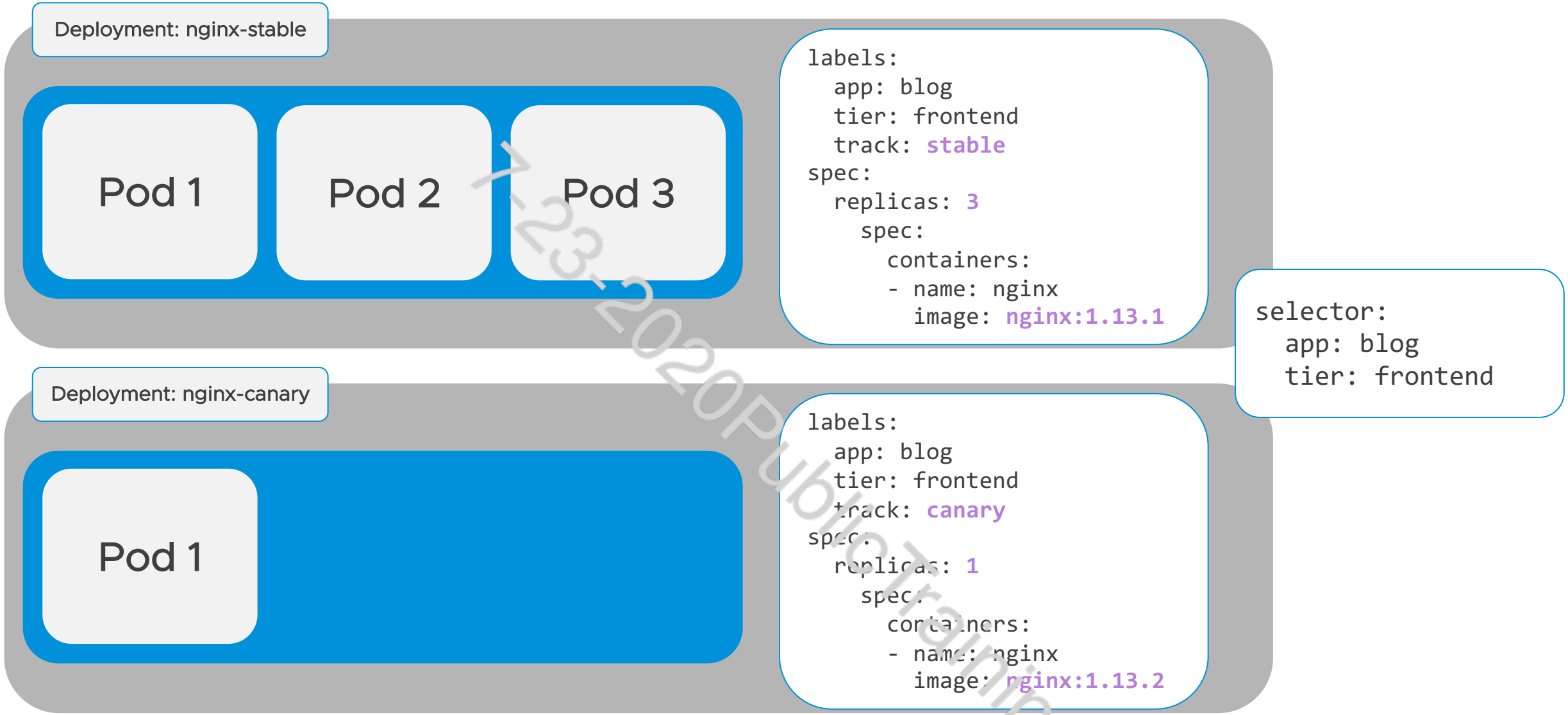
7-23-2020PublicTraining

Deployment Strategies - Canary



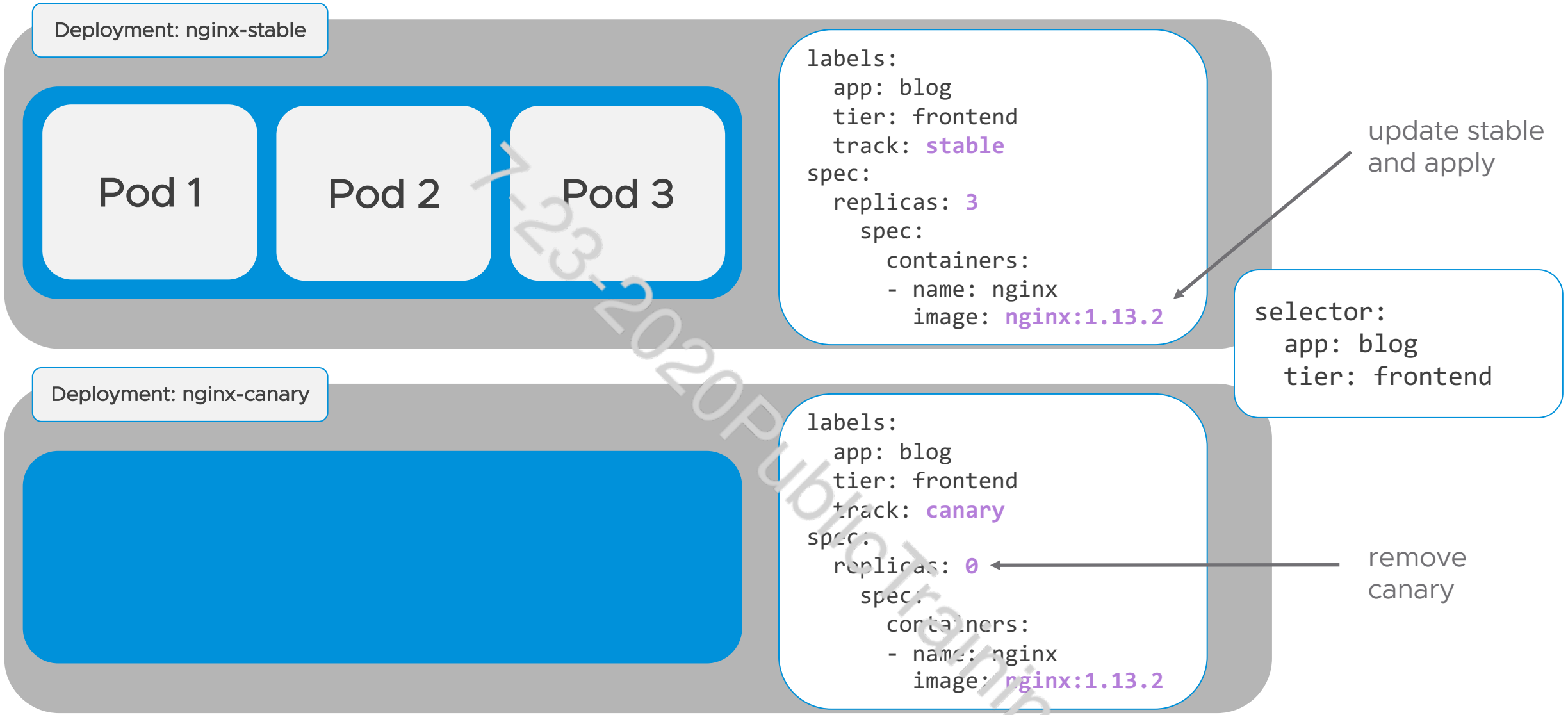
1-23-2020PublicTraining

Deployment Strategies - Canary



1-23-2020PublicTraining

Deployment Strategies - Canary

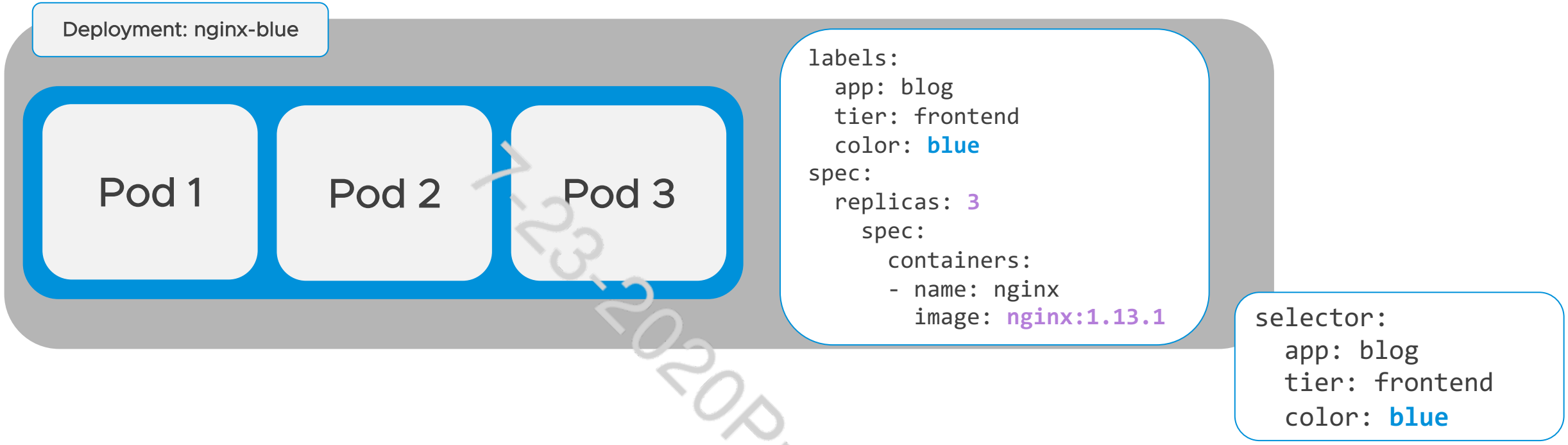


1-23-2020PublicTraining

Deployment Strategies - Blue/Green

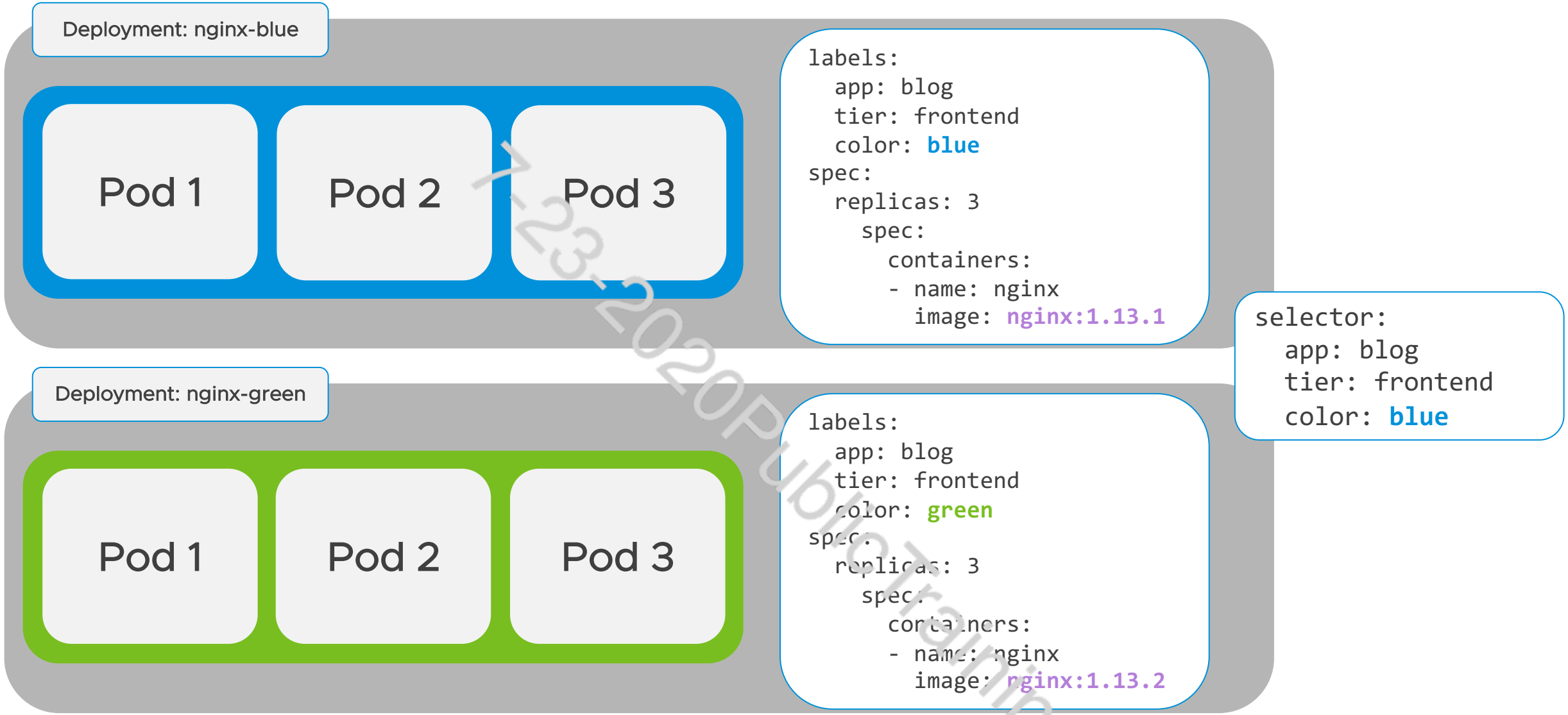
7-23-2020PublicTraining

Deployment Strategies - Blue/Green



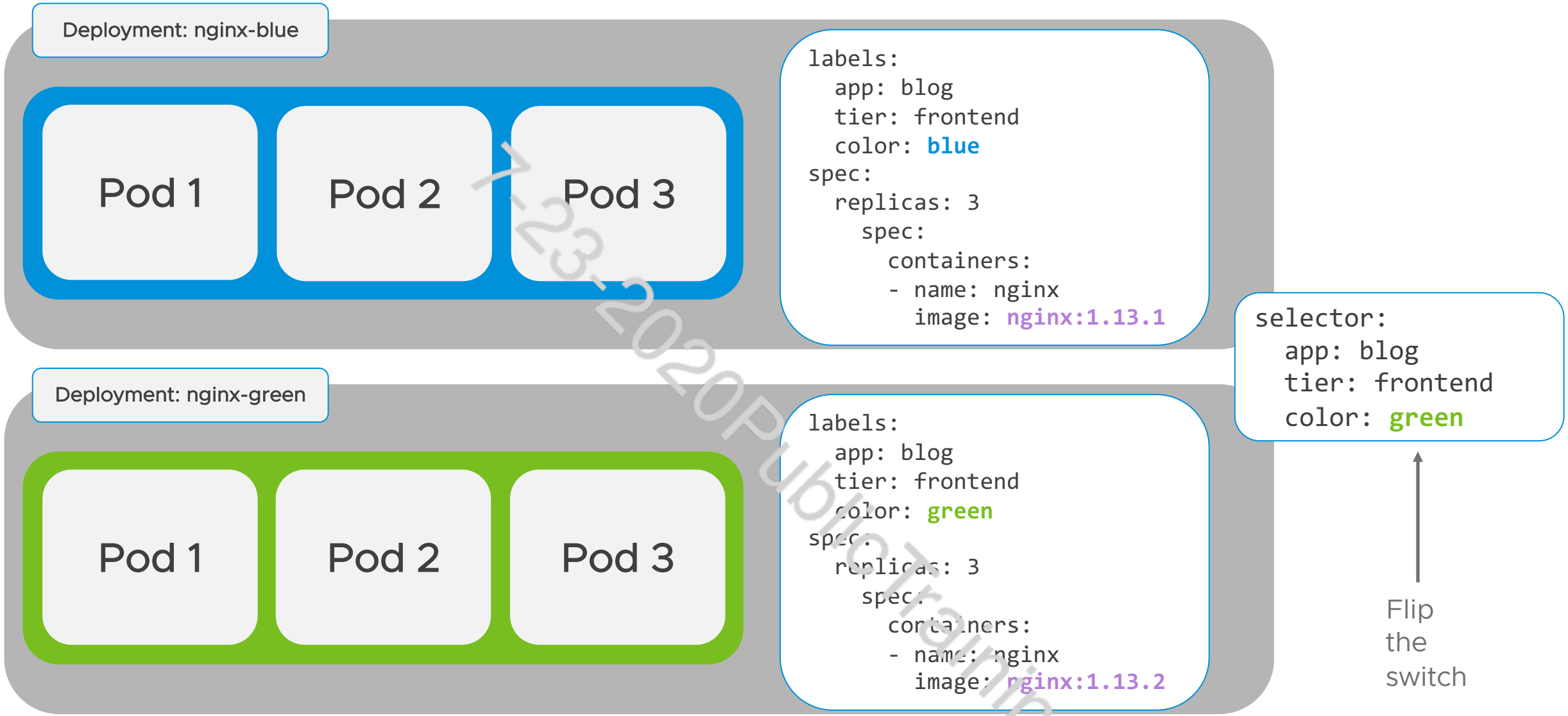
1-23-2020PublicTraining

Deployment Strategies - Blue/Green



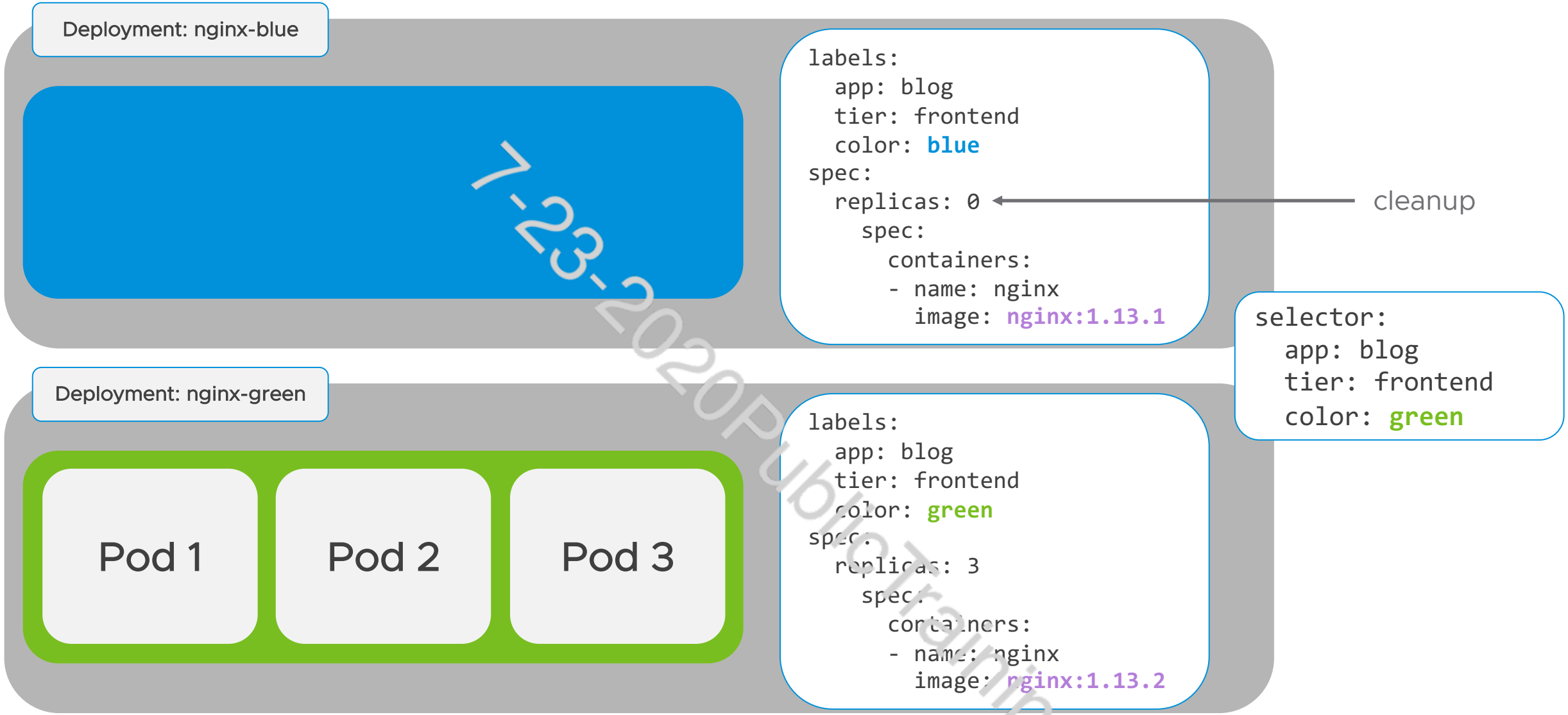
1-23-2020PublicTraining

Deployment Strategies - Blue/Green



1-23-2020PublicTraining

Deployment Strategies - Blue/Green



Public Training

Controlling Deployments

- Pausing
  - `kubectl rollout [ pause | resume ] deployment <name>`
- Rolling back to a previous version
  - Declarative (Preferred)
    - Modify the deployment's YAML file and re-apply
    - `kubectl apply -f deployment.yaml`
  - Imperative
    - `kubectl rollout undo deployment <name>`

7-28-2020 Public Training

General Optional Options

- progressDeadlineSeconds
  - max number of seconds you expect a deployment to ever take
- minReadySeconds
  - min seconds a newly created Pod should be ready without any of its containers crashing. defaults to 0
- revisionHistoryLimit
  - number of old ReplicaSets to retain to allow rollback. defaults to 10
- paused
  - pause a deployment; if set to true, then any changes to spec do not trigger deployment

723-2020PublicTraining

Lab 04

Controlling Deployments

Pause and control a rolling deployment

Canary Deployment

Setup and perform a Canary deployment

7-23-2020PublicTraining

Pod & Container Configurations

Chapter 05

7-23-2020PublicTraining

Agenda

Part One

1. Introduction to Containers
2. Kubernetes Fundamentals
3. Kubernetes Architecture & Troubleshooting
4. Deployment Management
5. Pod & Container Configurations

7-23-2020 Public Training

Probes

7-23-2020PublicTraining

Probes

- Liveness
  - whether the Container is running
  - upon failure - container will be restarted according to policy
- Readiness
  - whether the Container is ready to service requests
  - upon failure - pod is removed from Service so no requests are sent to it

1-23-2020PublicTraining

Probe Handlers

- Exec
  - run a command inside the container
  - success: return code from command is 0
- TCPSocket
  - TCP check
  - success: port is open and accepting connection
- HTTPGet
  - invoke HTTP GET against a URL
  - success: Any 2xx or 3xx HTTP response

1-23-2020PublicTraining

Example Uses

7-23-2020PublicTraining

Example Uses

- Cache initialization

7-23-2020PublicTraining

Example Uses

- Cache initialization
  - specify readiness probe once caches are hot

7-23-2020PublicTraining

Example Uses

- Cache initialization
  - specify readiness probe once caches are hot
- Ensure JVM started up successfully

7-28-2020PublicTraining

Example Uses

- Cache initialization
  - specify readiness probe once caches are hot
- Ensure JVM started up successfully
  - liveness probe against URL or looks for particular line in logs

7-28-2020PublicTraining

Example Uses

- Cache initialization
  - specify readiness probe once caches are hot
- Ensure JVM started up successfully
  - liveness probe against URL or looks for particular line in logs
- Container maintenance

7-28-2020PublicTraining

Example Uses

- Cache initialization
  - specify readiness probe once caches are hot
- Ensure JVM started up successfully
  - liveness probe against URL or looks for particular line in logs
- Container maintenance
  - container itself can take itself out of service by failing a readiness probe

7-28-2020PublicTraining

Probe Options

- initialDelaySeconds
  - how long to delay probing after container is started
- periodSeconds
  - how often to perform the probe; default = 10
- timeoutSeconds
  - default = 1

7-23-2020PublicTraining

Probe Options (Continued)

- successThreshold
  - minimum consecutive successes for the probe to be considered successful; default = 1
- failureThreshold
  - minimum consecutive successes for the probe to be considered failed; default = 3

4-23-2020PublicTraining

Resource Management

7-23-2020PublicTraining

Resources - CPU

- CPU Unit
  - 1 AWS vCPU
  - 1 GCP Core
  - 1 Azure vCore
  - 1 Hyperthread on a bare-metal Intel processor with Hyperthreading

- CPU values
  - 0.1 = = 100m
  - one hundred millicpu / 10% of a CPU unit

7-23-2020PublicTraining

Resources - Memory

- Memory Unit
  - E, P, T, G, M, K - 10 based
  - Ei, Pi, Ti, Gi, Mi, Ki - power of 2 based

- Memory values
  - `1Gi == 1.073741824 GB`
  - `1G  == 1.0 GB`

  - typically 10 based for storage and 2 based for memory

1-23-2020PublicTraining

Resource Requests

- Can be thought of as "minimum resource required" specification
- Helps Kubernetes more efficiently schedule pods
- Pods will be scheduled if the sum of the resource requests of the scheduled Containers is less than the capacity of the node

```
containers:
 - name: db
   image: mysql
   resources:
     requests:
       memory: "512Mi"
       cpu: "0.5"
```

1-26-2020PublicTraining

Resource Limits

- Protect against a runaway app
- If a Container exceeds its memory limit, it might be terminated
- If a Container exceeds its memory request, it is likely that its Pod will be evicted whenever the node runs out of memory

```
containers:
 - name: db
   image: mysql
   resources:
     limits:
       memory: "1Gi"
       cpu: "1"
```

7-28-2020PublicTraining

Lab 05

Probes

Add probes to the Go Web App Deployment

Resource Management

Add resource requests to Go Web App

Add resource limits to Go Web App

7-3-2020PublicTraining

Conclusion

7-23-2020PublicTraining

7-23-2020PublicTraining

Thank You

Confidential | ©2020 VMware, Inc.