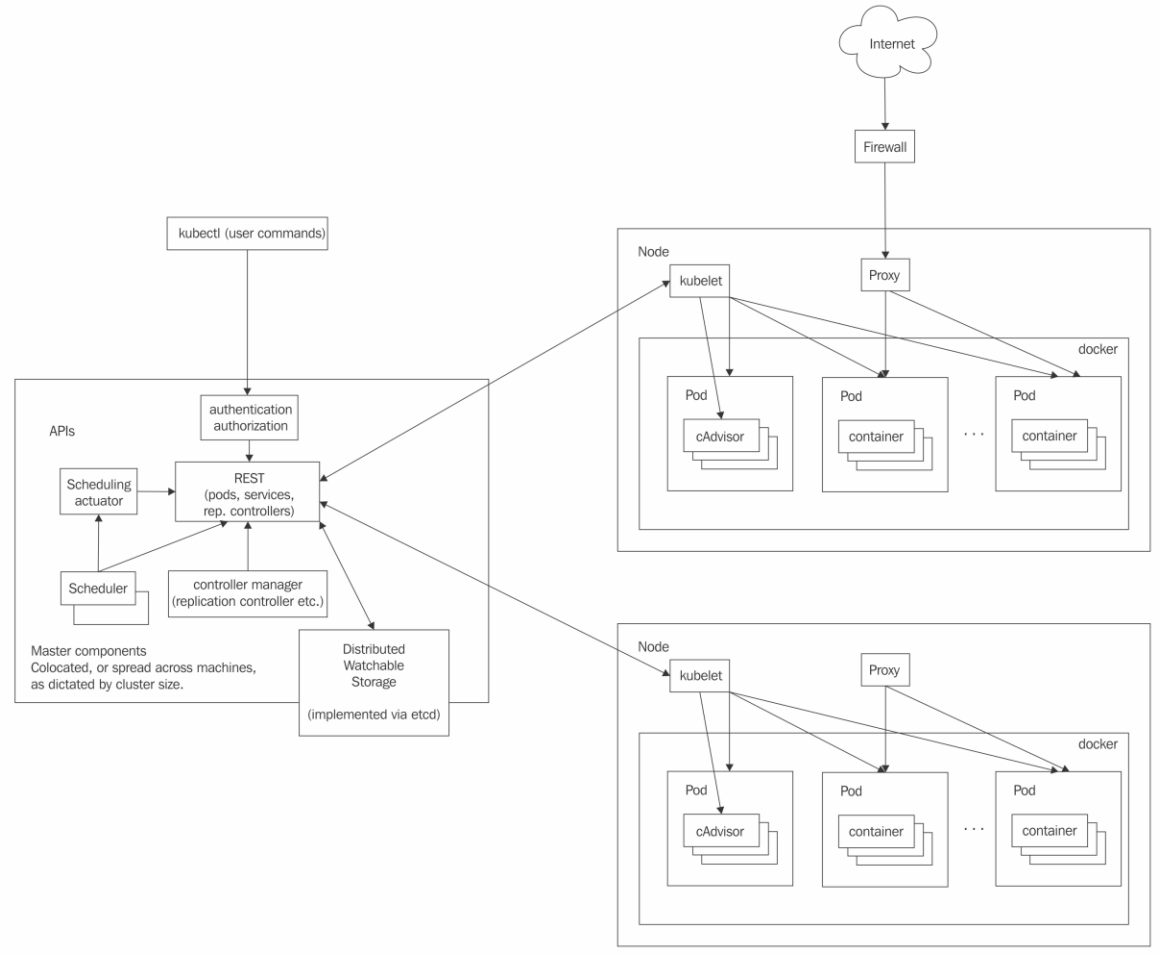
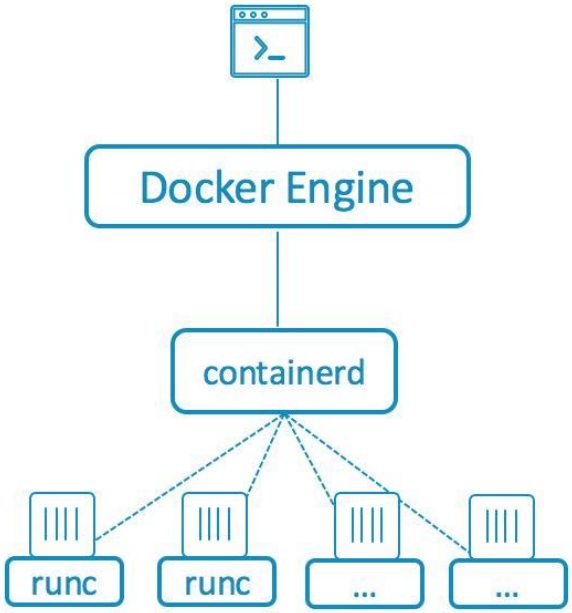
Chapter 1: Understanding Kubernetes Architecture



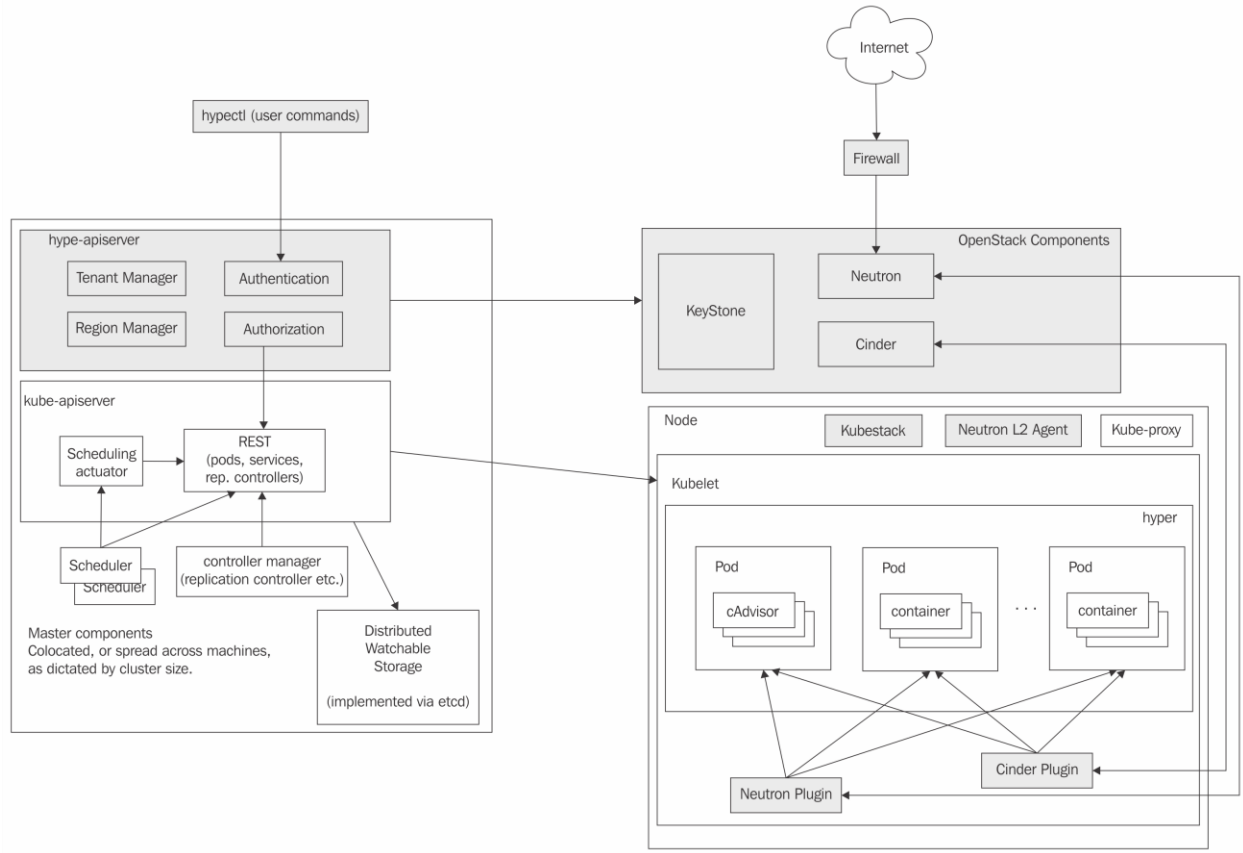


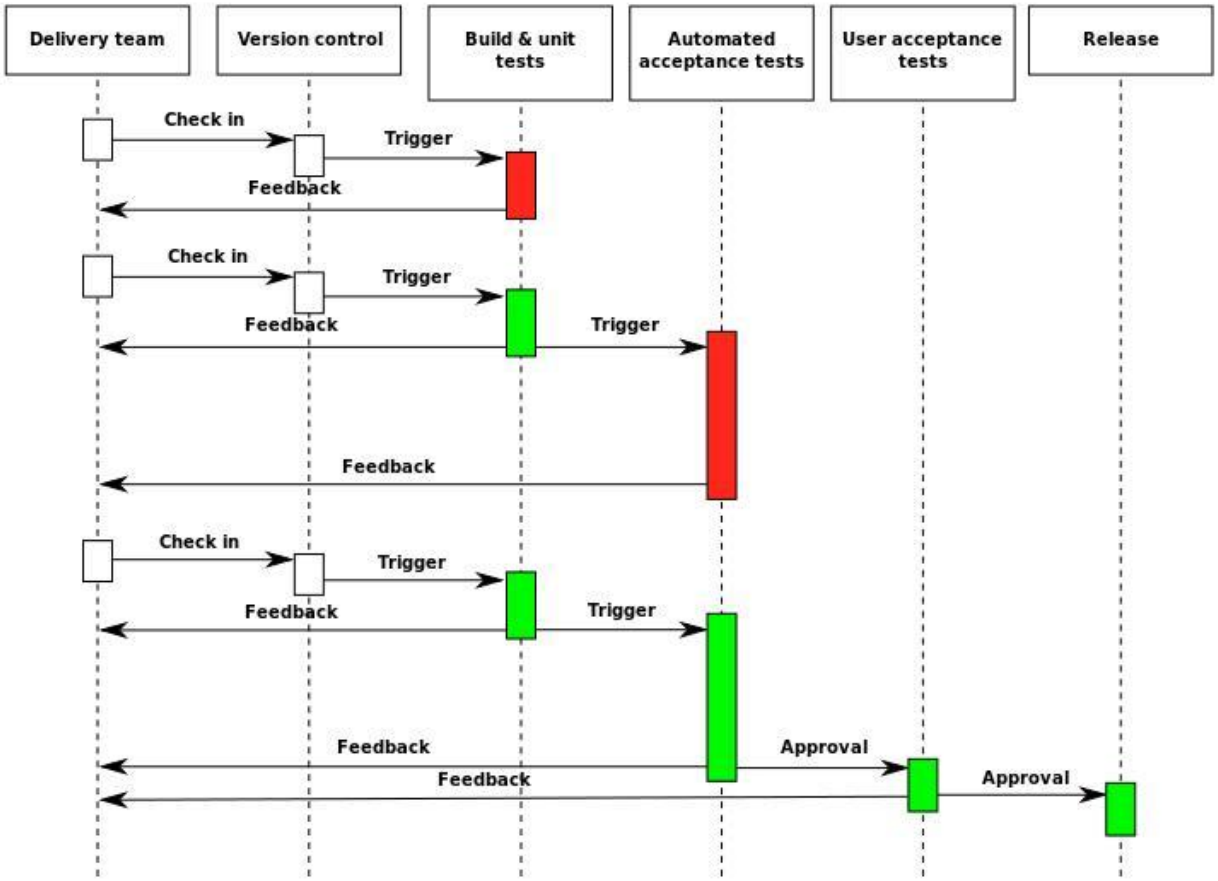
Same Docker UI and commands

User interacts with the Docker Engine

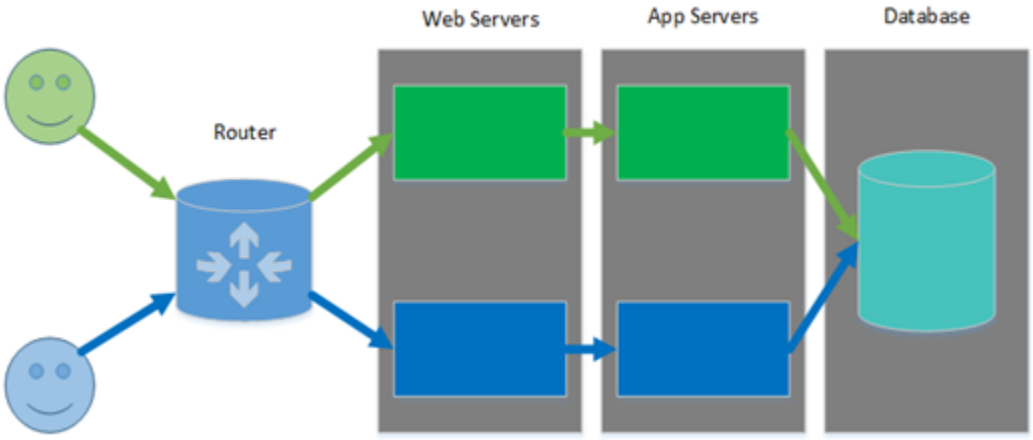
Engine communicates with containerd

containerd spins up runc or other OCI compliant runtime to run containers

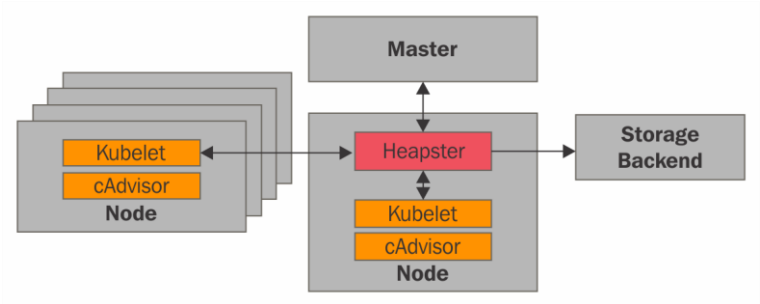



Delivery team
Version control
Build & unit tests
Automated acceptance tests
User acceptance tests
Release
Check in
Trigger
Feedback
Check in
Trigger
Feedback
Trigger
Feedback
Check in
Trigger
Feedback
Trigger
Feedback
Approval
Feedback
Approval

Chapter 2: Creating Kubernetes Clusters



Chapter 3: Monitoring, Logging, and Troubleshooting



Usage

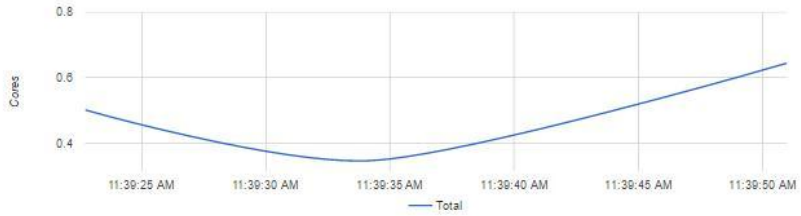
Overview

Processes

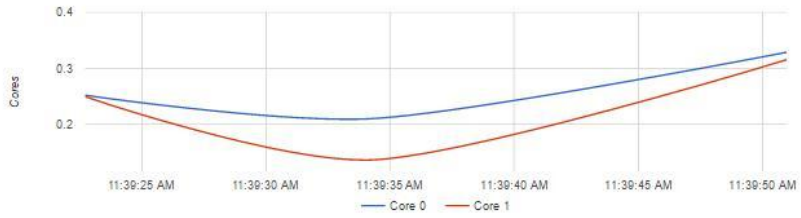
No processes found

CPU

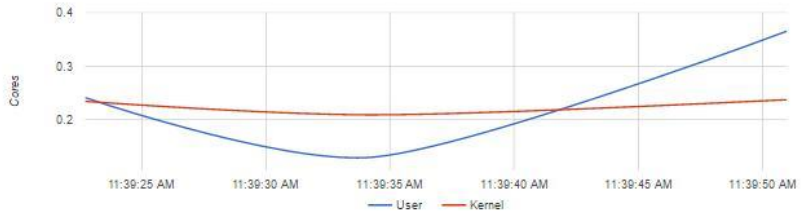
Total Usage

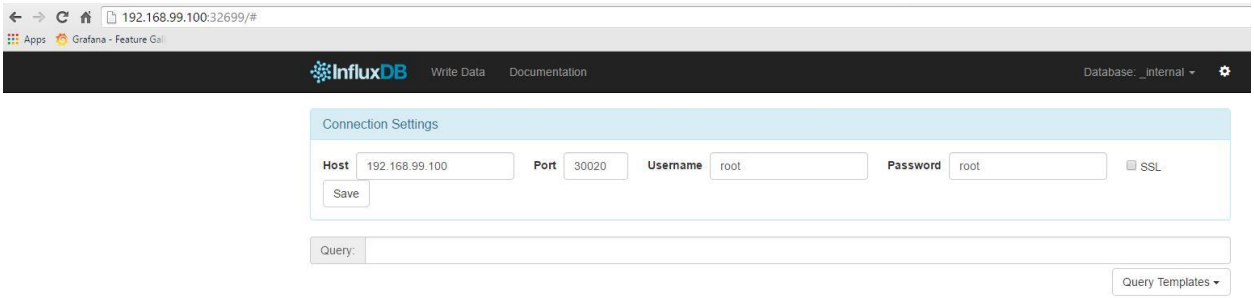


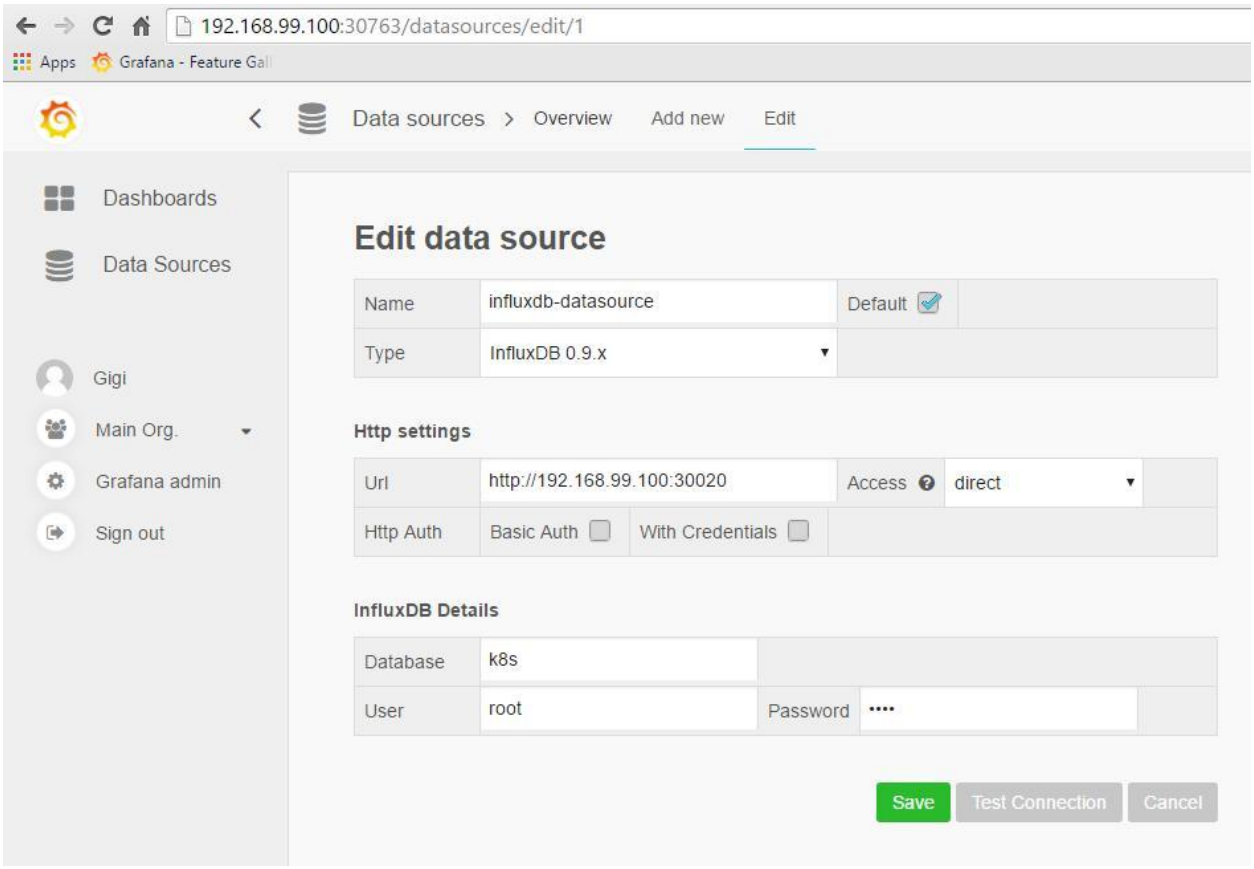
Usage per Core

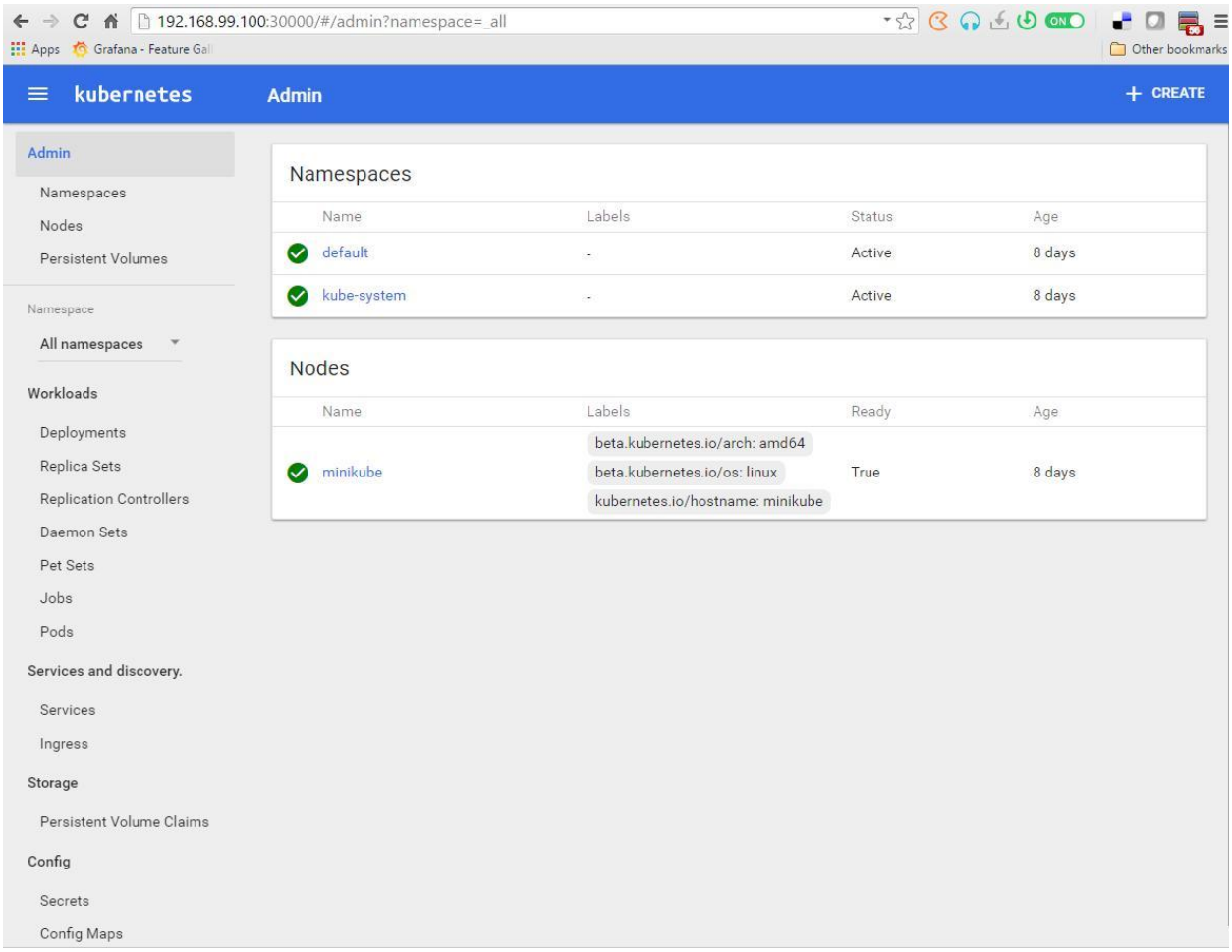


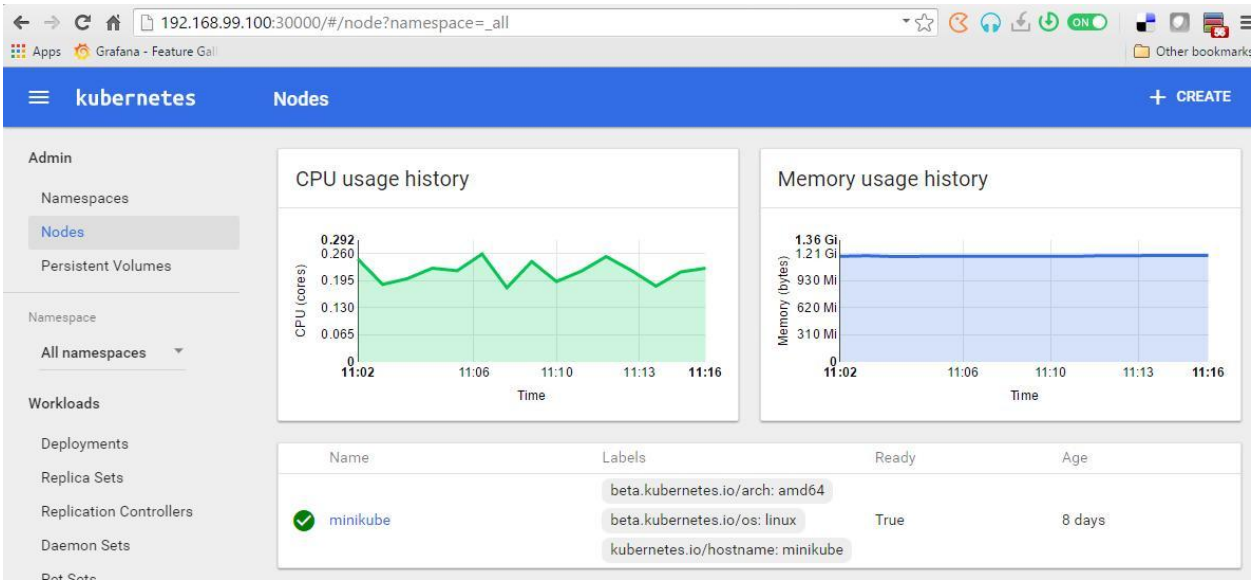
Usage Breakdown

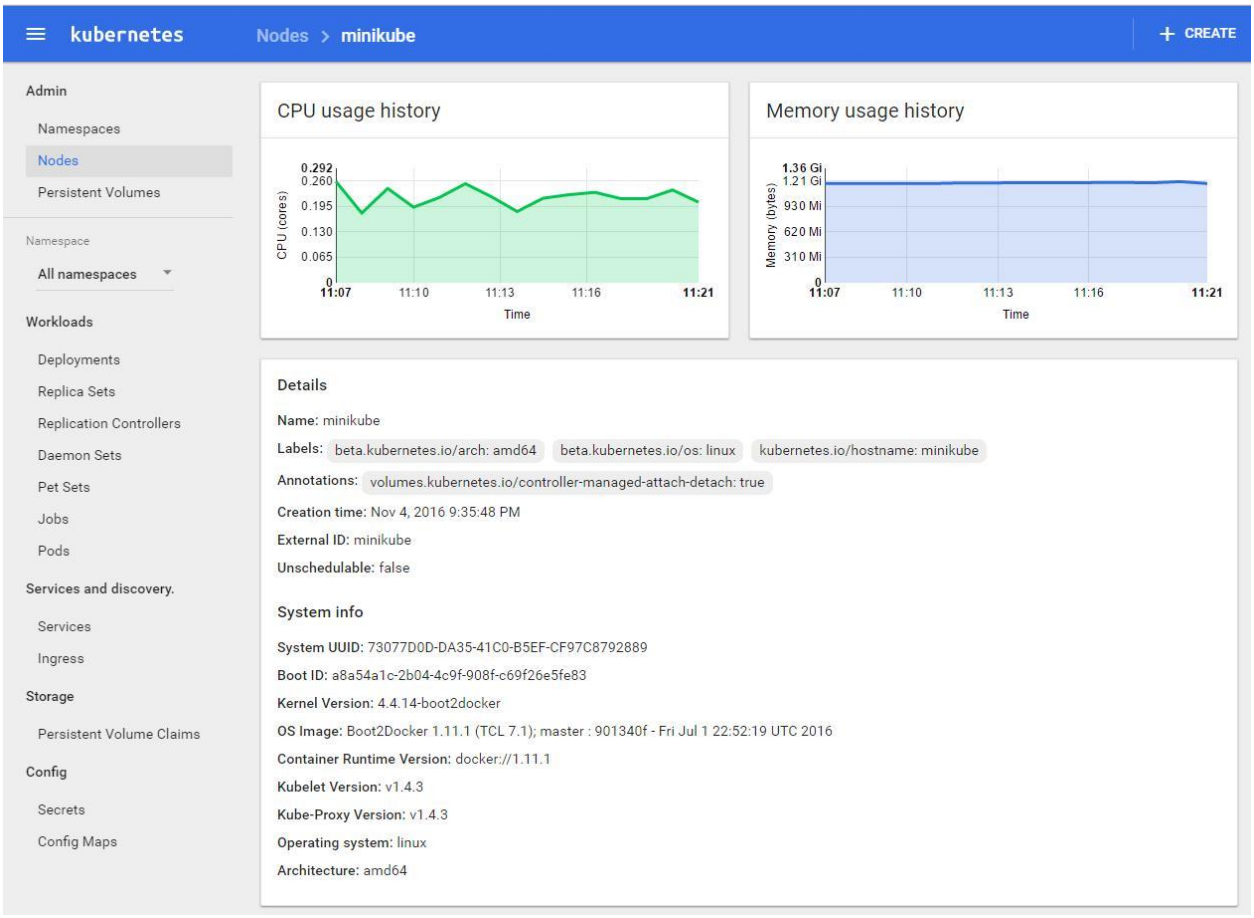


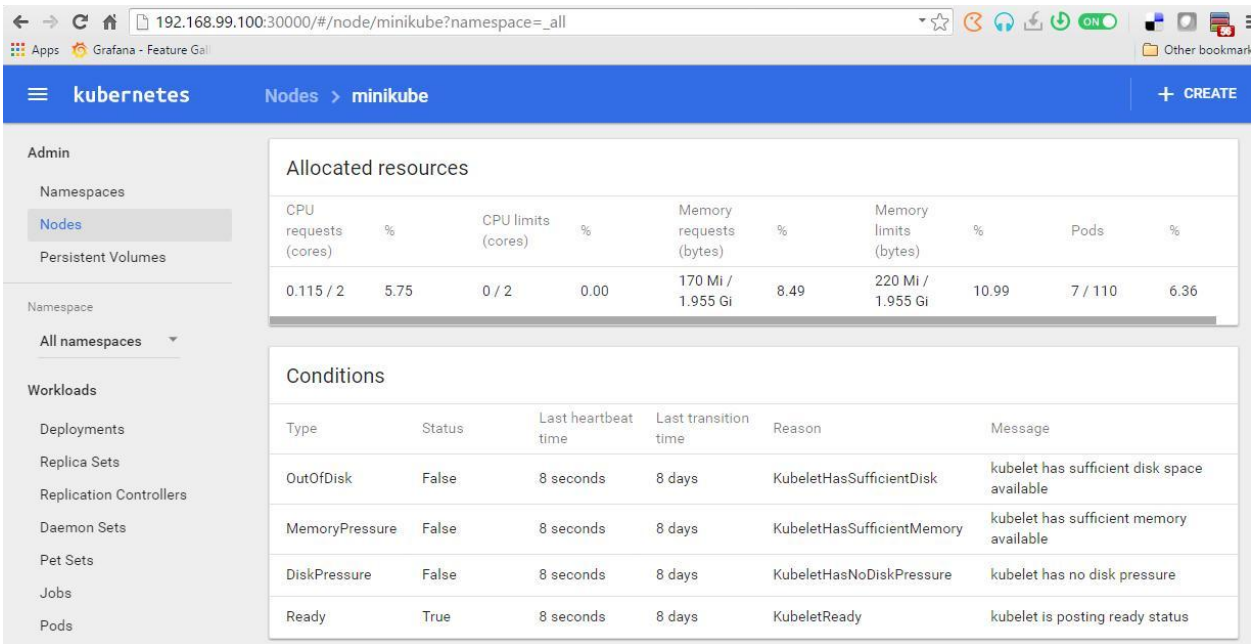


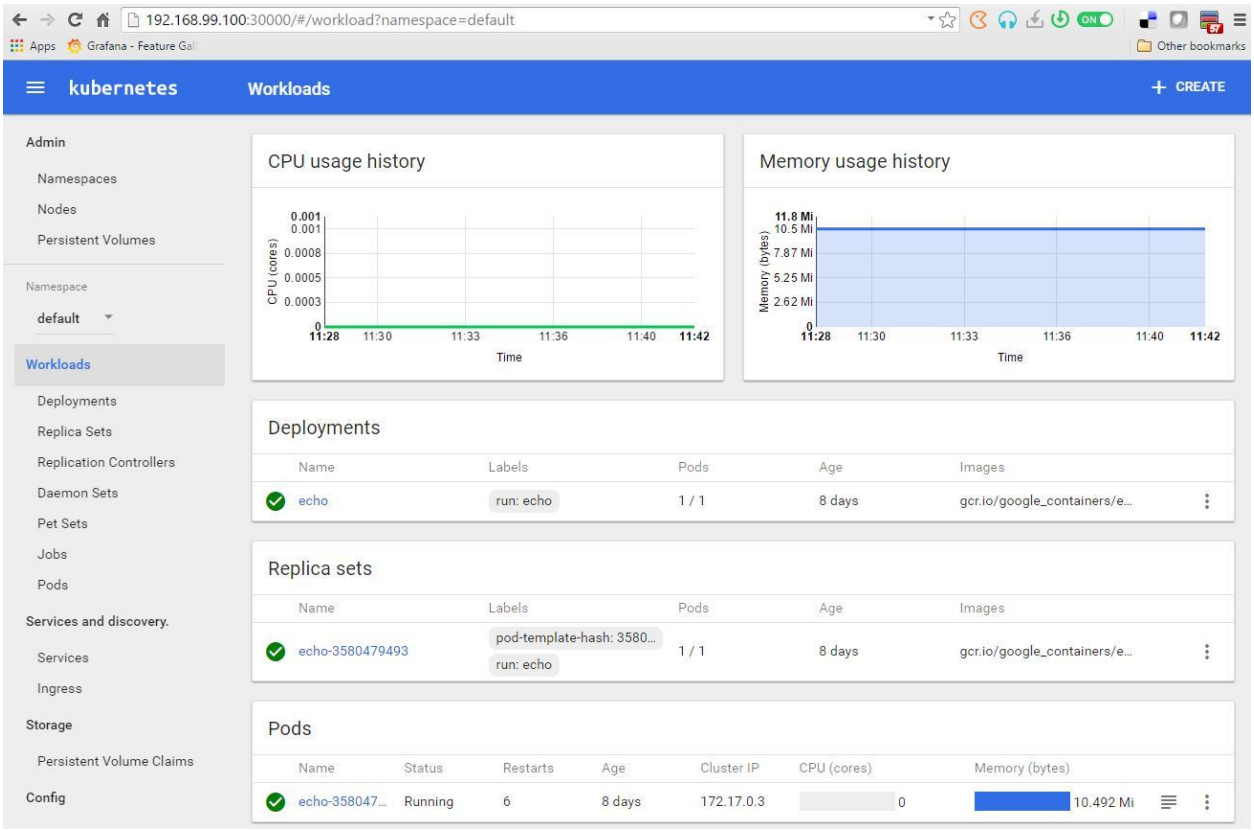

192.168.99.100:30763/datasources/edit/1
Apps
Grafana - Feature Gal
Data sources
Overview
Add new
Edit
Dashboards
Data Sources
Gigi
Main Org.
Grafana admin
Sign out
Edit data source
Name
influxdb-datasource
Default
Type
InfluxDB 0.9.x
Http settings
Url
http://192.168.99.100:30020
Access
direct
Http Auth
Basic Auth
With Credentials
InfluxDB Details
Database
k8s
User
root
Password
Save
Test Connection
Cancel

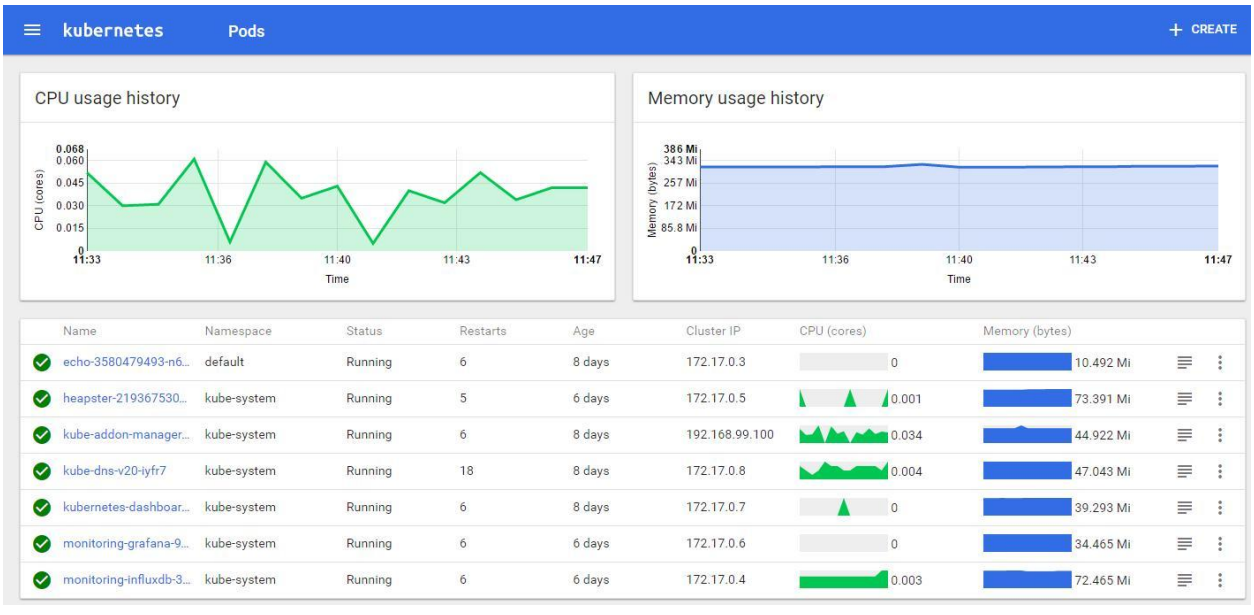

192.168.99.100:30000/#/admin?namespace=_all
kubernetes
Admin
+ CREATE
Admin
Namespaces
Nodes
Persistent Volumes
Namespace
All namespaces
Workloads
Deployments
Replica Sets
Replication Controllers
Daemon Sets
Pet Sets
Jobs
Pods
Services and discovery.
Services
Ingress
Storage
Persistent Volume Claims
Config
Secrets
Config Maps
Namespaces
Name
Labels
Status
Age
default
-
Active
8 days
kube-system
-
Active
8 days
Nodes
Name
Labels
Ready
Age
minikube
beta.kubernetes.io/arch: amd64
beta.kubernetes.io/os: linux
kubernetes.io/hostname: minikube
True
8 days

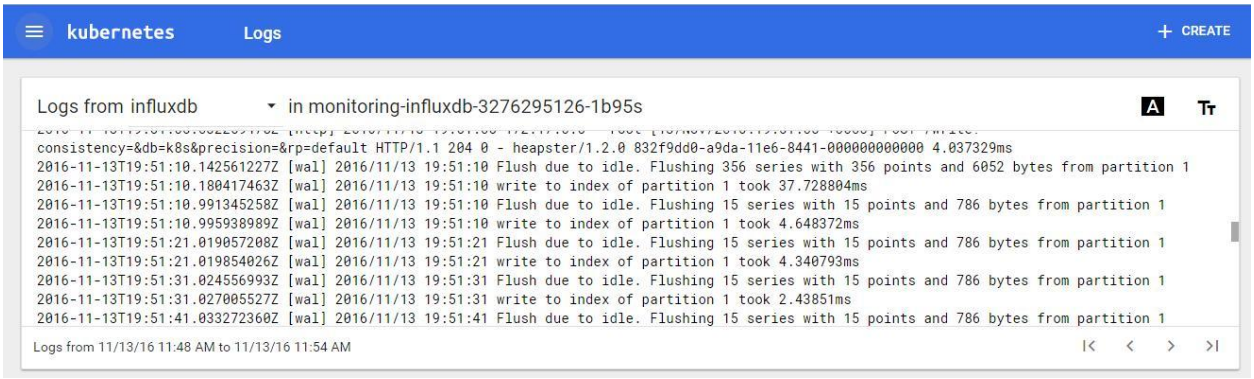

192.168.99.100:30000/#/node?namespace=_all
kubernetes
Nodes
+ CREATE
Admin
Namespaces
Nodes
Persistent Volumes
Namespace
All namespaces
Workloads
Deployments
Replica Sets
Replication Controllers
Daemon Sets
CPU usage history
CPU (cores)
0.292
0.260
0.195
0.130
0.065
0
11:02
11:06
11:10
11:13
11:16
Time
Memory usage history
Memory (bytes)
1.36 Gi
1.21 Gi
930 Mi
620 Mi
310 Mi
0
11:02
11:06
11:10
11:13
11:16
Time
Name
Labels
Ready
Age
minikube
beta.kubernetes.io/arch: amd64
beta.kubernetes.io/os: linux
kubernetes.io/hostname: minikube
True
8 days

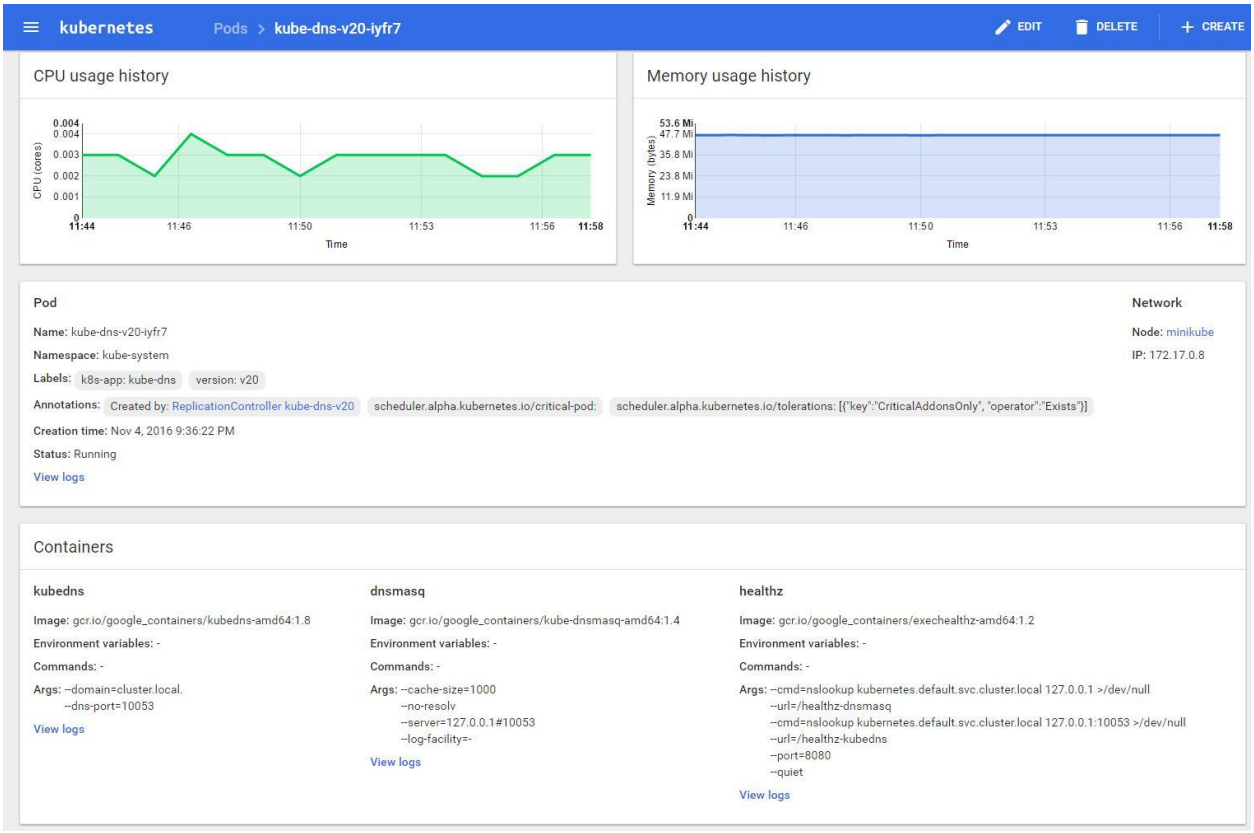

kubernetes
Nodes > minikube
+ CREATE
Admin
Namespaces
Nodes
Persistent Volumes
Namespace
All namespaces
Workloads
Deployments
Replica Sets
Replication Controllers
Daemon Sets
Pet Sets
Jobs
Pods
Services and discovery.
Services
Ingress
Storage
Persistent Volume Claims
Config
Secrets
Config Maps
CPU usage history
CPU (cores)
0.292
0.260
0.195
0.130
0.065
0
11:07
11:10
11:13
11:16
11:21
Time
Memory usage history
Memory (bytes)
1.36 Gi
1.21 Gi
930 Mi
620 Mi
310 Mi
0
11:07
11:10
11:13
11:16
11:21
Time
Details
Name: minikube
Labels: beta.kubernetes.io/arch: amd64 beta.kubernetes.io/os: linux kubernetes.io/hostname: minikube
Annotations: volumes.kubernetes.io/controller-managed-attach-detach: true
Creation time: Nov 4, 2016 9:35:48 PM
External ID: minikube
Unschedulable: false
System info
System UUID: 73077D0D-DA35-41C0-B5EF-CF97C8792889
Boot ID: a8a54a1c-2b04-4c9f-908f-c69f26e5fe83
Kernel Version: 4.4.14-boot2docker
OS Image: Boot2Docker 1.11.1 (TCL 7.1); master : 901340f - Fri Jul 1 22:52:19 UTC 2016
Container Runtime Version: docker://1.11.1
Kubelet Version: v1.4.3
Kube-Proxy Version: v1.4.3
Operating system: linux
Architecture: amd64

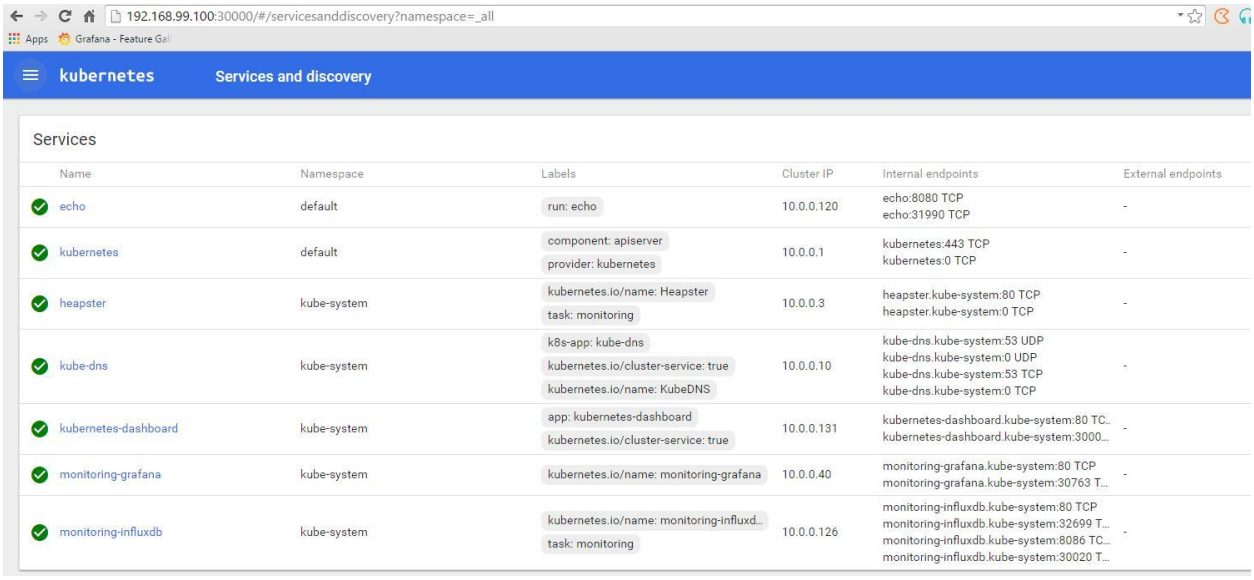

192.168.99.100:30000/#/node/minikube?namespace=_all
Apps
Grafana - Feature Gal
Other bookmark
kubernetes
Nodes > minikube
+ CREATE
Admin
Namespaces
Nodes
Persistent Volumes
Namespace
All namespaces
Workloads
Deployments
Replica Sets
Replication Controllers
Daemon Sets
Pet Sets
Jobs
Pods
Allocated resources
CPU requests (cores) | % | CPU limits (cores) | % | Memory requests (bytes) | % | Memory limits (bytes) | % | Pods | %
0.115 / 2 | 5.75 | 0 / 2 | 0.00 | 170 Mi / 1.955 Gi | 8.49 | 220 Mi / 1.955 Gi | 10.99 | 7 / 110 | 6.36
Conditions
Type | Status | Last heartbeat time | Last transition time | Reason | Message
OutOfDisk | False | 8 seconds | 8 days | KubeletHasSufficientDisk | kubelet has sufficient disk space available
MemoryPressure | False | 8 seconds | 8 days | KubeletHasSufficientMemory | kubelet has sufficient memory available
DiskPressure | False | 8 seconds | 8 days | KubeletHasNoDiskPressure | kubelet has no disk pressure
Ready | True | 8 seconds | 8 days | KubeletReady | kubelet is posting ready status

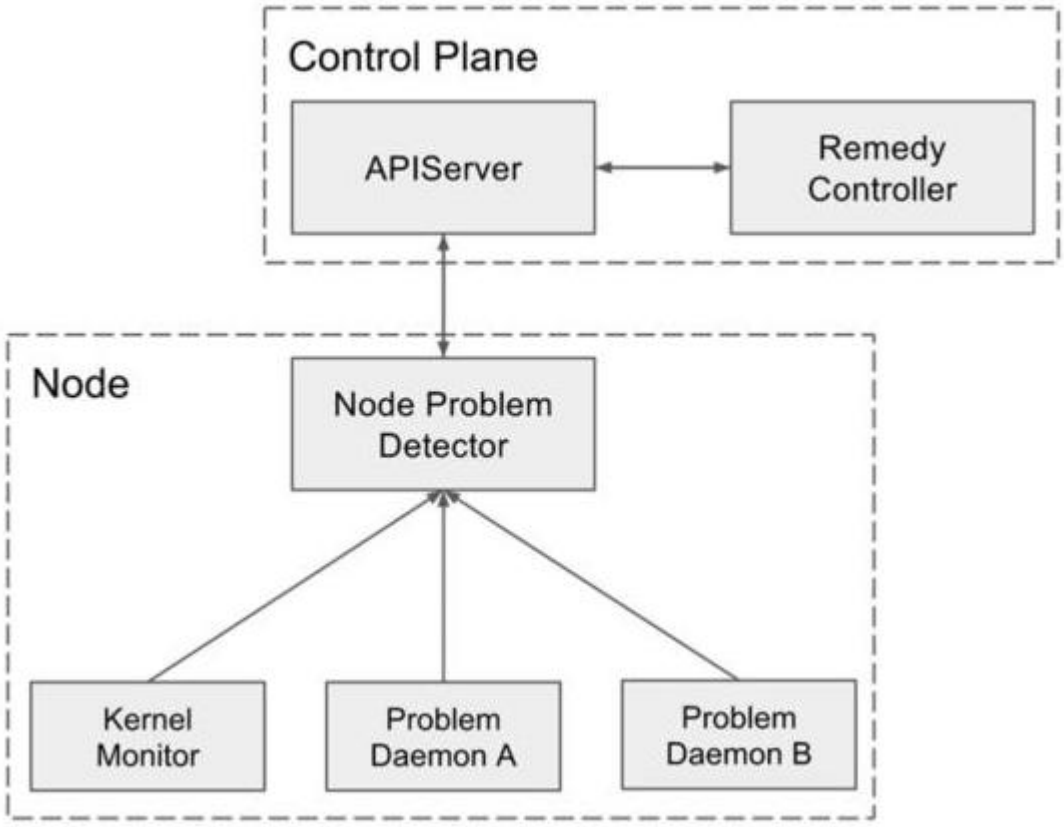

192.168.99.100:30000/#/workload?namespace=default
kubernetes
Workloads
+ CREATE
Admin
Namespaces
Nodes
Persistent Volumes
Namespace
default
Workloads
Deployments
Replica Sets
Replication Controllers
Daemon Sets
Pet Sets
Jobs
Pods
Services and discovery.
Services
Ingress
Storage
Persistent Volume Claims
Config
CPU usage history
CPU (cores)
Time
Memory usage history
Memory (bytes)
Time
Deployments
Name Labels Pods Age Images
echo run: echo 1 / 1 8 days gcr.io/google_containers/e...
Replica sets
Name Labels Pods Age Images
echo-3580479493 pod-template-hash: 3580... run: echo 1 / 1 8 days gcr.io/google_containers/e...
Pods
Name Status Restarts Age Cluster IP CPU (cores) Memory (bytes)
echo-358047... Running 6 8 days 172.17.0.3 0 10.492 Mi

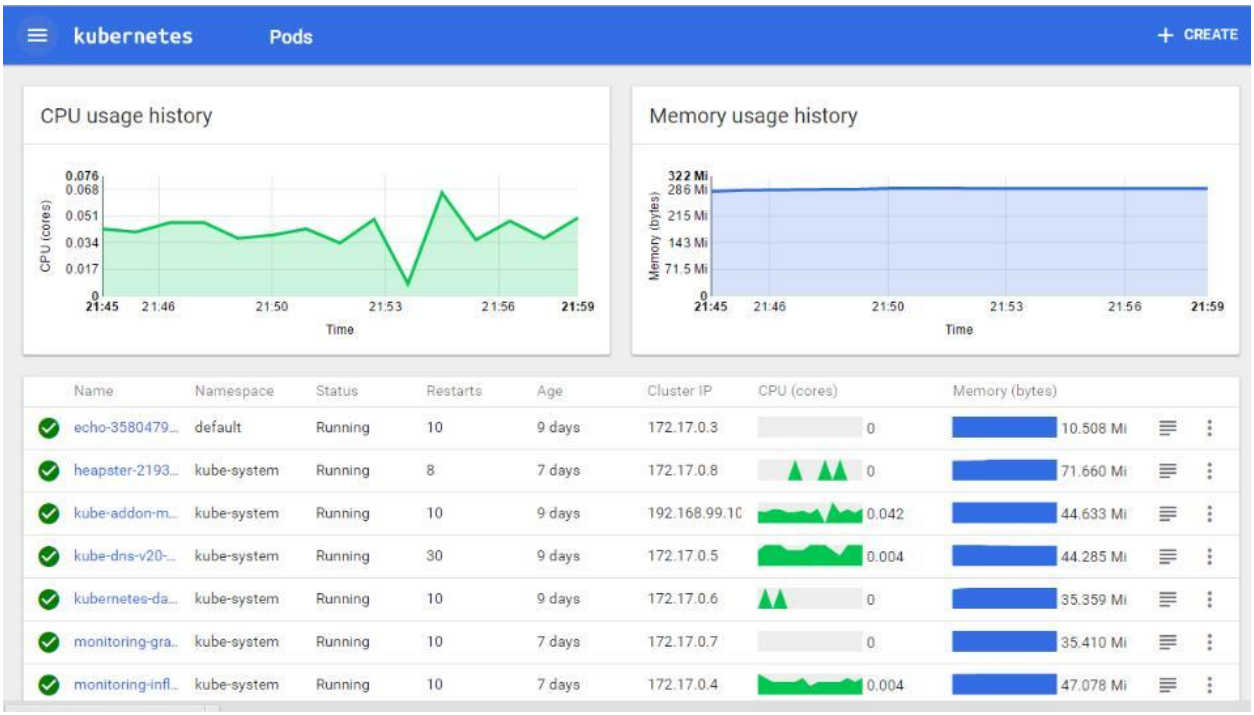

kubernetes
Pods
+ CREATE
CPU usage history
CPU (cores)
Time
Memory usage history
Memory (bytes)
Time
Name Namespace Status Restarts Age Cluster IP CPU (cores) Memory (bytes)
echo-3580479493-n6... default Running 6 8 days 172.17.0.3 0 10.492 Mi
heapster-219367530... kube-system Running 5 6 days 172.17.0.5 0.001 73.391 Mi
kube-addon-manager... kube-system Running 6 8 days 192.168.99.100 0.034 44.922 Mi
kube-dns-v20-iyfr7 kube-system Running 18 8 days 172.17.0.8 0.004 47.043 Mi
kubernetes-dashboar... kube-system Running 6 8 days 172.17.0.7 0 39.293 Mi
monitoring-grafana-9... kube-system Running 6 6 days 172.17.0.6 0 34.465 Mi
monitoring-influxdb-3... kube-system Running 6 6 days 172.17.0.4 0.003 72.465 Mi


kubernetes
Logs
CREATE
Logs from influxdb
in monitoring-influxdb-3276295126-1b95s
consistency=&db=k8s&precision=&rp=default HTTP/1.1 204 0 - heapster/1.2.0 832f9dd0-a9da-11e6-8441-000000000000 4.037329ms
2016-11-13T19:51:10.142561227Z [wal] 2016/11/13 19:51:10 Flush due to idle. Flushing 356 series with 356 points and 6052 bytes from partition 1
2016-11-13T19:51:10.180417463Z [wal] 2016/11/13 19:51:10 write to index of partition 1 took 37.728804ms
2016-11-13T19:51:10.991345258Z [wal] 2016/11/13 19:51:10 Flush due to idle. Flushing 15 series with 15 points and 786 bytes from partition 1
2016-11-13T19:51:10.995938989Z [wal] 2016/11/13 19:51:10 write to index of partition 1 took 4.648372ms
2016-11-13T19:51:21.019057208Z [wal] 2016/11/13 19:51:21 Flush due to idle. Flushing 15 series with 15 points and 786 bytes from partition 1
2016-11-13T19:51:21.019854026Z [wal] 2016/11/13 19:51:21 write to index of partition 1 took 4.340793ms
2016-11-13T19:51:31.024556993Z [wal] 2016/11/13 19:51:31 Flush due to idle. Flushing 15 series with 15 points and 786 bytes from partition 1
2016-11-13T19:51:31.027005527Z [wal] 2016/11/13 19:51:31 write to index of partition 1 took 2.43851ms
2016-11-13T19:51:41.033272360Z [wal] 2016/11/13 19:51:41 Flush due to idle. Flushing 15 series with 15 points and 786 bytes from partition 1
Logs from 11/13/16 11:48 AM to 11/13/16 11:54 AM

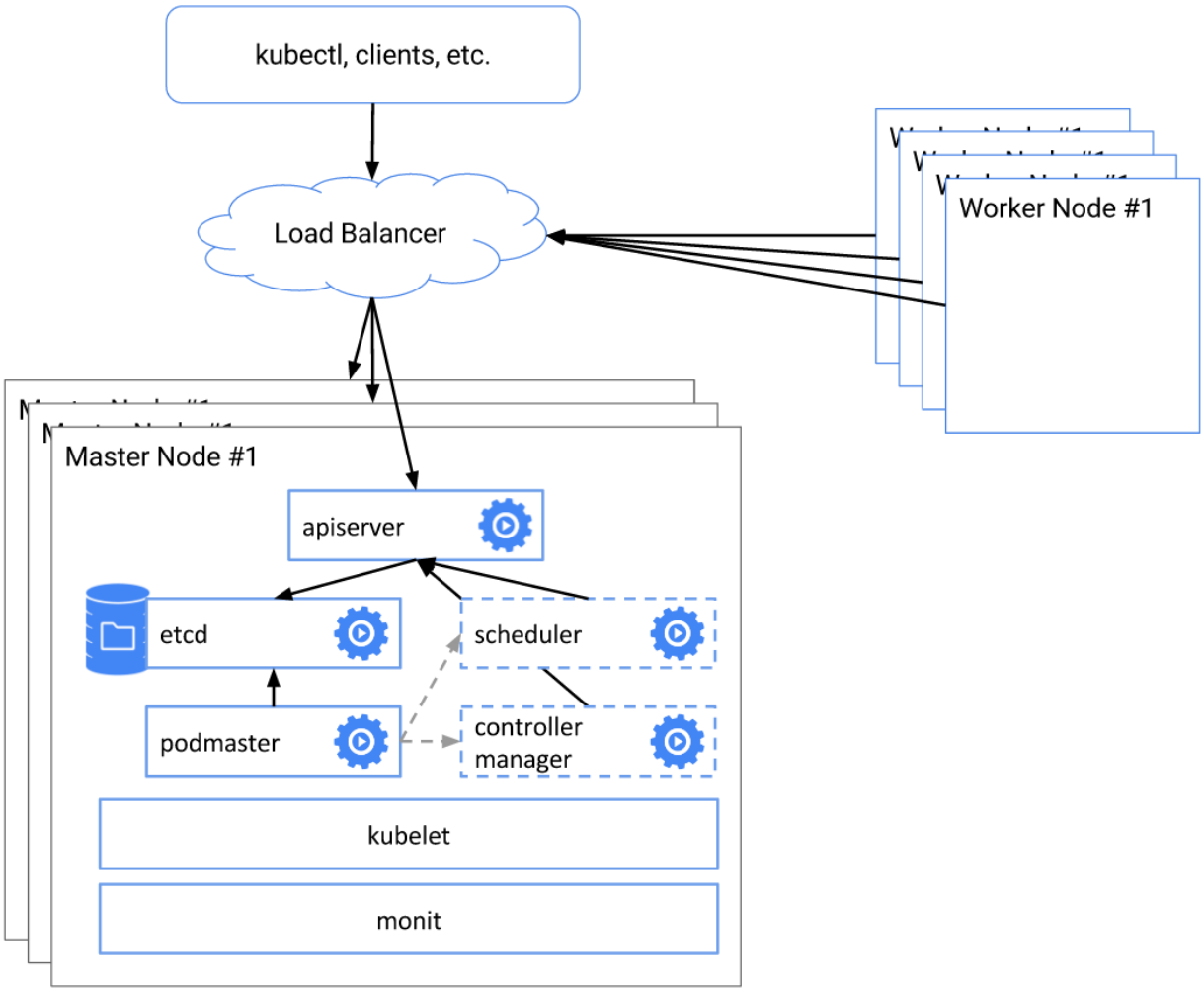

kubernetes
Pods > kube-dns-v20-iyfr7
EDIT
DELETE
CREATE
CPU usage history
CPU (cores)
0.004
0.003
0.002
0.001
0
11:44
11:46
11:50
11:53
11:56
11:58
Time
Memory usage history
Memory (bytes)
53.6 Mi
47.7 Mi
35.8 Mi
23.8 Mi
11.9 Mi
0
Time
Pod
Name: kube-dns-v20-iyfr7
Namespace: kube-system
Labels: k8s-app: kube-dns version: v20
Annotations: Created by: ReplicationController kube-dns-v20 scheduler.alpha.kubernetes.io/critical-pod: scheduler.alpha.kubernetes.io/tolerations: [{"key":"CriticalAddonsOnly", "operator":"Exists"}]
Creation time: Nov 4, 2016 9:36:22 PM
Status: Running
View logs
Network
Node: minikube
IP: 172.17.0.8
Containers
kubedns
Image: gcr.io/google_containers/kubedns-amd64:1.8
Environment variables: -
Commands: -
Args: --domain=cluster.local.
--dns-port=10053
View logs
dnsmasq
Image: gcr.io/google_containers/kube-dnsmasq-amd64:1.4
Environment variables: -
Commands: -
Args: --cache-size=1000
--no-resolv
--server=127.0.0.1#10053
--log-facility=-
View logs
healthz
Image: gcr.io/google_containers/exechealthz-amd64:1.2
Environment variables: -
Commands: -
Args: --cmd=nslookup kubernetes.default.svc.cluster.local 127.0.0.1 >/dev/null
--url=/healthz-dnsmasq
--cmd=nslookup kubernetes.default.svc.cluster.local 127.0.0.1:10053 >/dev/null
--url=/healthz-kubedns
--port=8080
--quiet
View logs

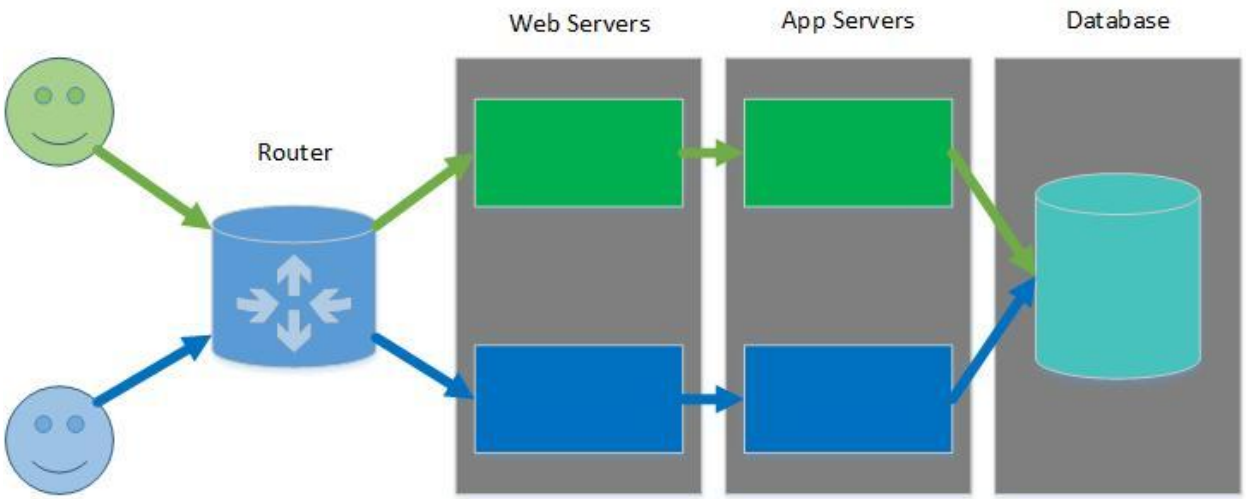

192.168.99.100:30000/#/servicesanddiscovery?namespace=_all
Apps
Grafana - Feature Gal
kubernetes
Services and discovery
Services
Name
Namespace
Labels
Cluster IP
Internal endpoints
External endpoints
echo
default
run: echo
10.0.0.120
echo:8080 TCP
echo:31990 TCP
-
kubernetes
default
component: apiserver
provider: kubernetes
10.0.0.1
kubernetes:443 TCP
kubernetes:0 TCP
-
heapster
kube-system
kubernetes.io/name: Heapster
task: monitoring
10.0.0.3
heapster.kube-system:80 TCP
heapster.kube-system:0 TCP
-
kube-dns
kube-system
k8s-app: kube-dns
kubernetes.io/cluster-service: true
kubernetes.io/name: KubeDNS
10.0.0.10
kube-dns.kube-system:53 UDP
kube-dns.kube-system:0 UDP
kube-dns.kube-system:53 TCP
kube-dns.kube-system:0 TCP
-
kubernetes-dashboard
kube-system
app: kubernetes-dashboard
kubernetes.io/cluster-service: true
10.0.0.131
kubernetes-dashboard.kube-system:80 TC..
kubernetes-dashboard.kube-system:3000...
-
monitoring-grafana
kube-system
kubernetes.io/name: monitoring-grafana
10.0.0.40
monitoring-grafana.kube-system:80 TCP
monitoring-grafana.kube-system:30763 T...
-
monitoring-influxdb
kube-system
kubernetes.io/name: monitoring-influxd...
task: monitoring
10.0.0.126
monitoring-influxdb.kube-system:80 TCP
monitoring-influxdb.kube-system:32699 T...
monitoring-influxdb.kube-system:8086 TC...
monitoring-influxdb.kube-system:30020 T...
-


Control Plane
APIServer
Remedy Controller
Node
Node Problem Detector
Kernel Monitor
Problem Daemon A
Problem Daemon B

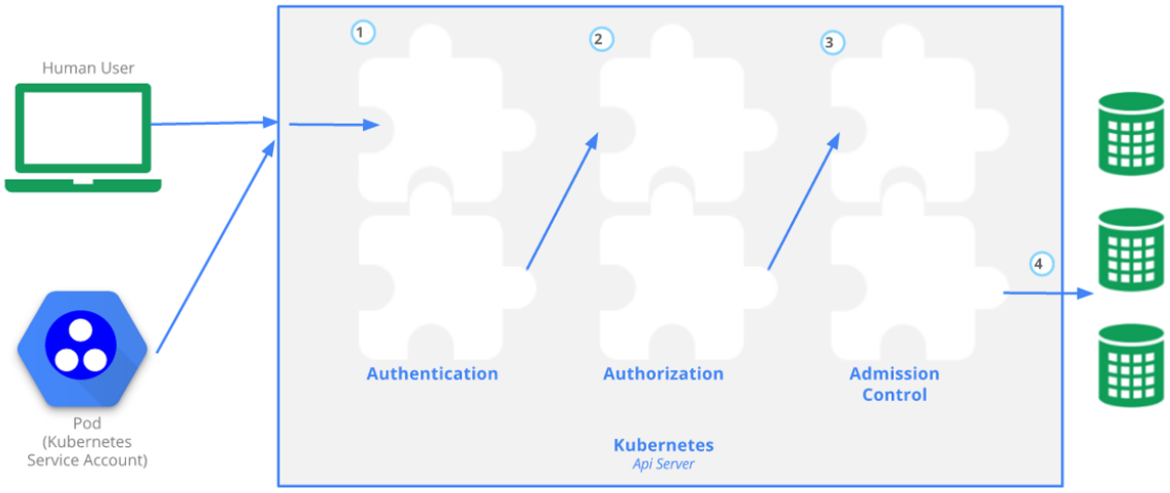

kubernetes
Pods
+ CREATE
CPU usage history
CPU (cores)
0.076
0.068
0.051
0.034
0.017
0
21:45
21:46
21:50
21:53
21:56
21:59
Time
Memory usage history
Memory (bytes)
322 Mi
286 Mi
215 Mi
143 Mi
71.5 Mi
0
21:45
21:46
21:50
21:53
21:56
21:59
Time
Name
Namespace
Status
Restarts
Age
Cluster IP
CPU (cores)
Memory (bytes)
echo-3580479... default Running 10 9 days 172.17.0.3 0 10.508 Mi
heapster-2193... kube-system Running 8 7 days 172.17.0.8 0 71.660 Mi
kube-addon-m... kube-system Running 10 9 days 192.168.99.10 0.042 44.633 Mi
kube-dns-v20-... kube-system Running 30 9 days 172.17.0.5 0.004 44.285 Mi
kubernetes-da... kube-system Running 10 9 days 172.17.0.6 0 35.359 Mi
monitoring-gra... kube-system Running 10 7 days 172.17.0.7 0 35.410 Mi
monitoring-infl... kube-system Running 10 7 days 172.17.0.4 0.004 47.078 Mi

Chapter 4: High Availability and Reliability

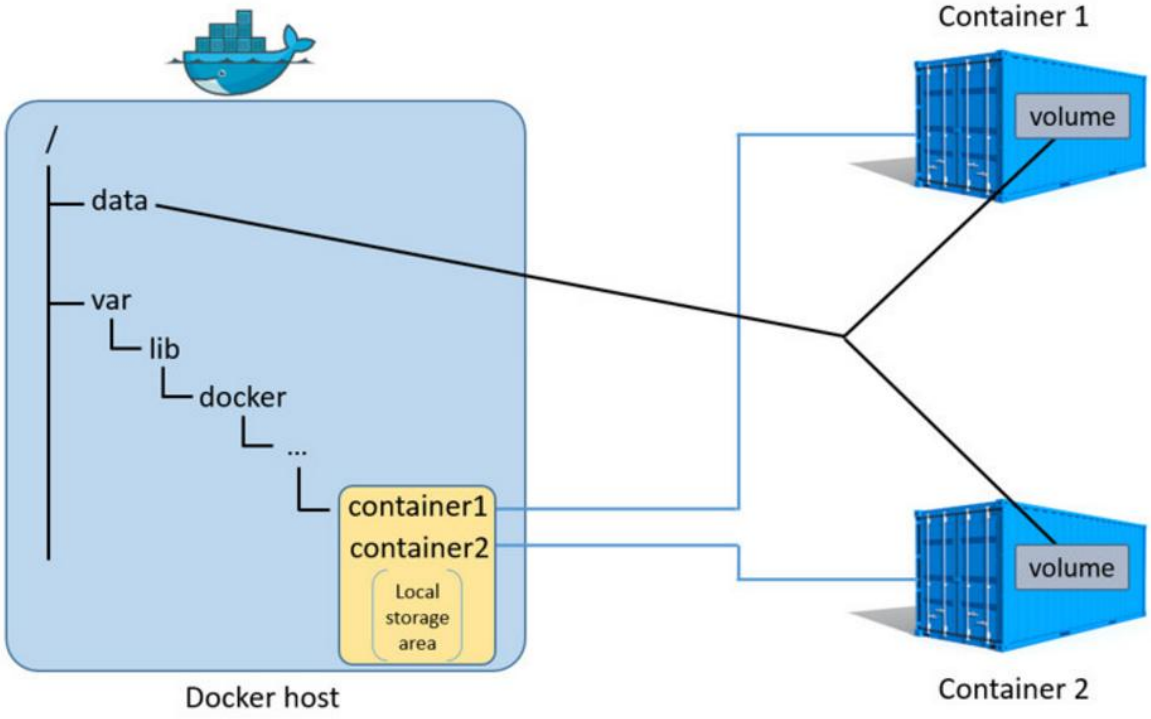


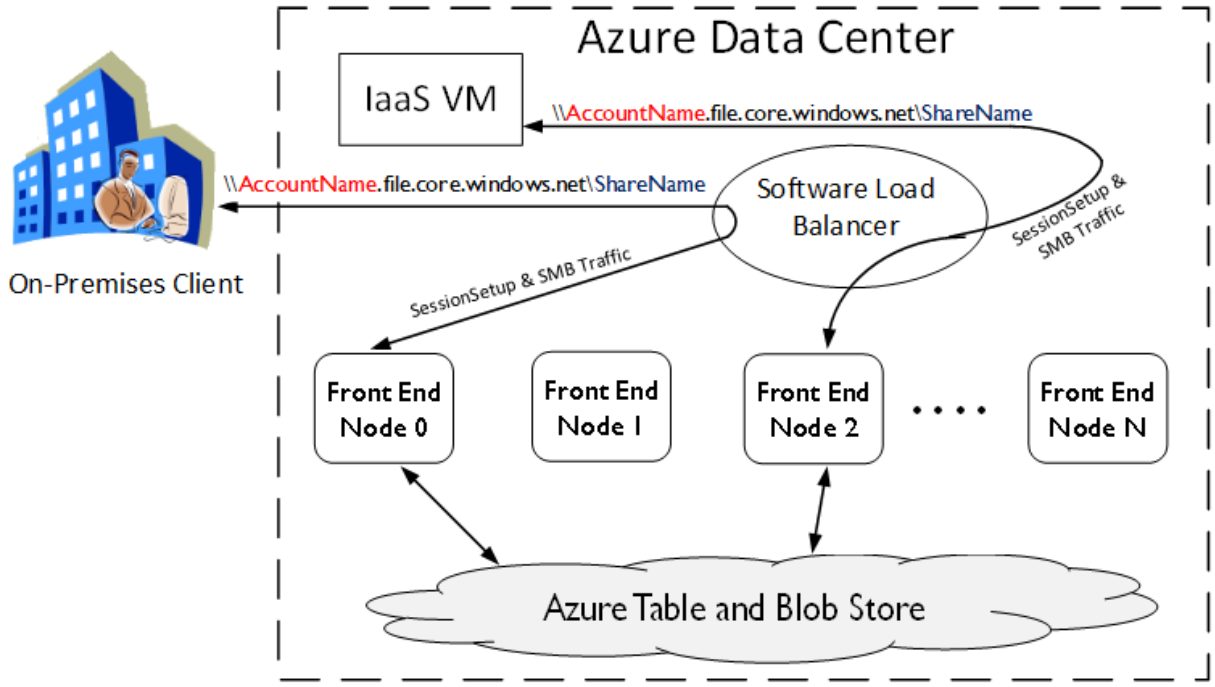

Web Servers
App Servers
Database
Router

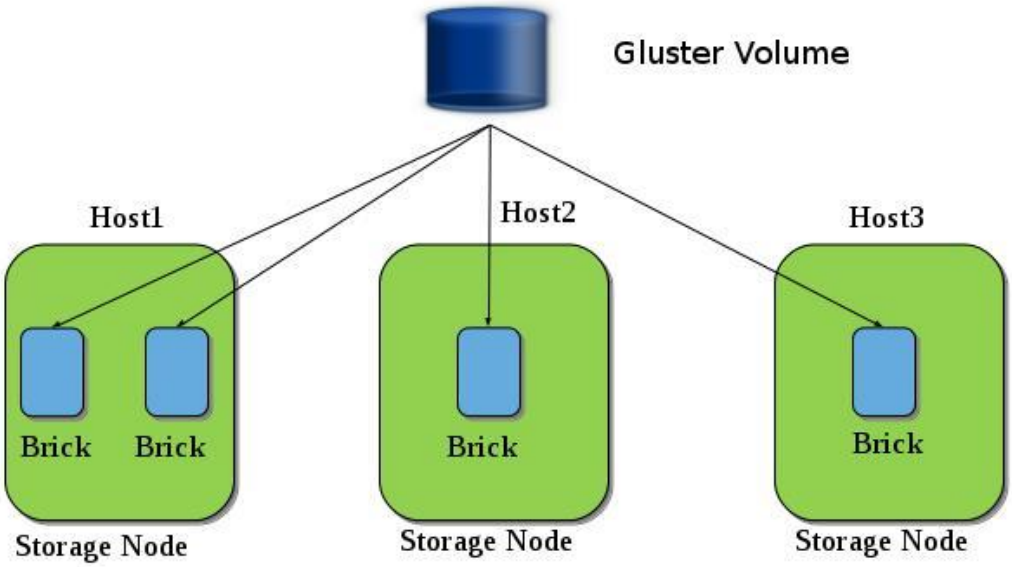
Chapter 5: Configuring Kubernetes Security, Limits, and Accounts

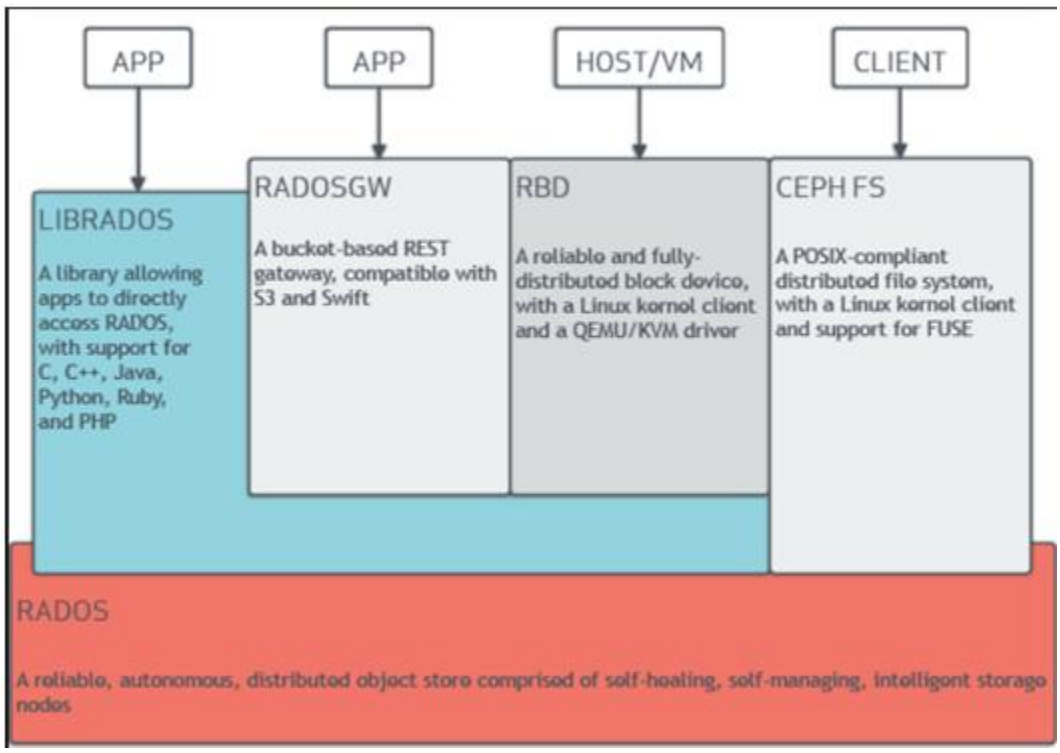


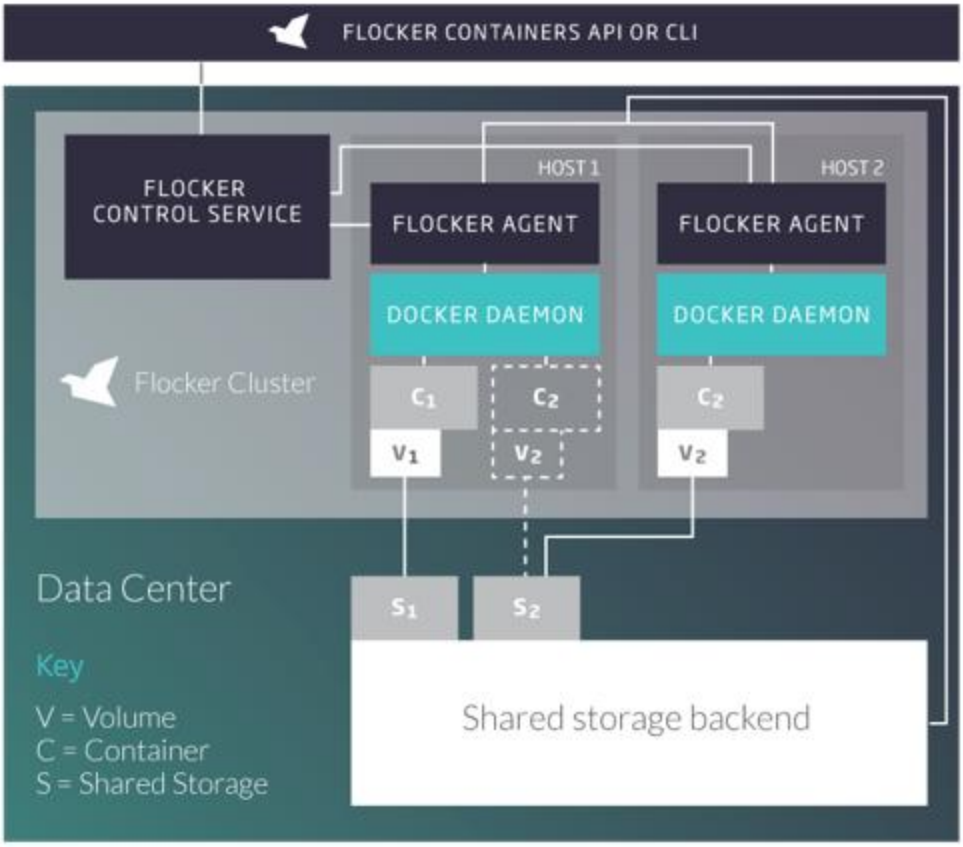
Chapter 7: Handling Kubernetes Storage

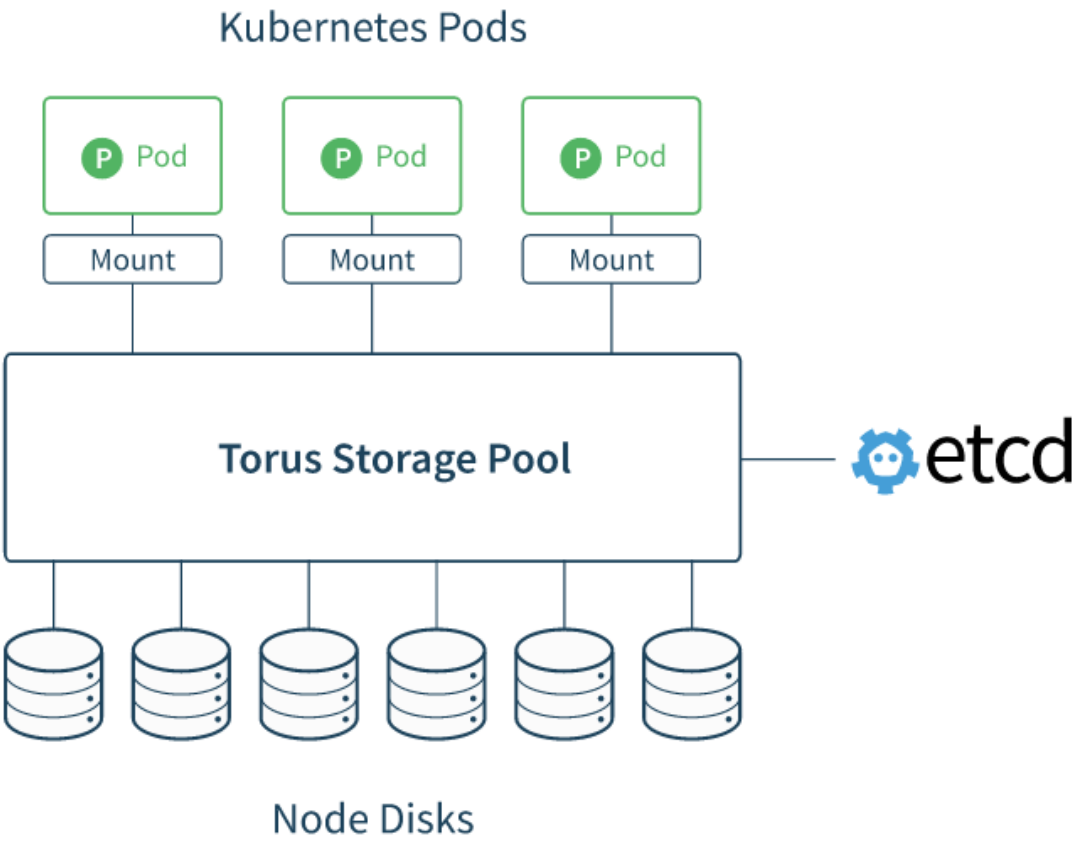




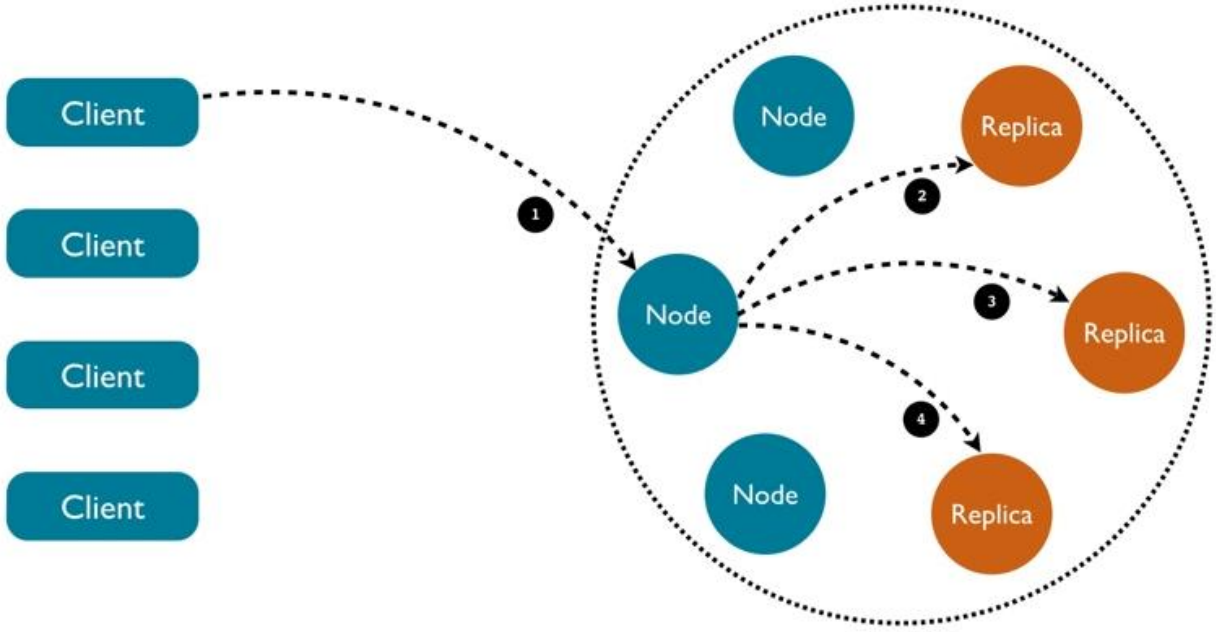

Gluster Volume
Host1
Host2
Host3
Brick
Brick
Brick
Brick
Storage Node
Storage Node
Storage Node


APP
APP
HOST/VM
CLIENT
LIBRADOS
A library allowing apps to directly access RADOS, with support for C, C++, Java, Python, Ruby, and PHP
RADOSGW
A bucket-based REST gateway, compatible with S3 and Swift
RBD
A reliable and fully-distributed block device, with a Linux kernel client and a QEMU/KVM driver
CEPH FS
A POSIX-compliant distributed file system, with a Linux kernel client and support for FUSE
RADOS
A reliable, autonomous, distributed object store comprised of self-healing, self-managing, intelligent storage nodes

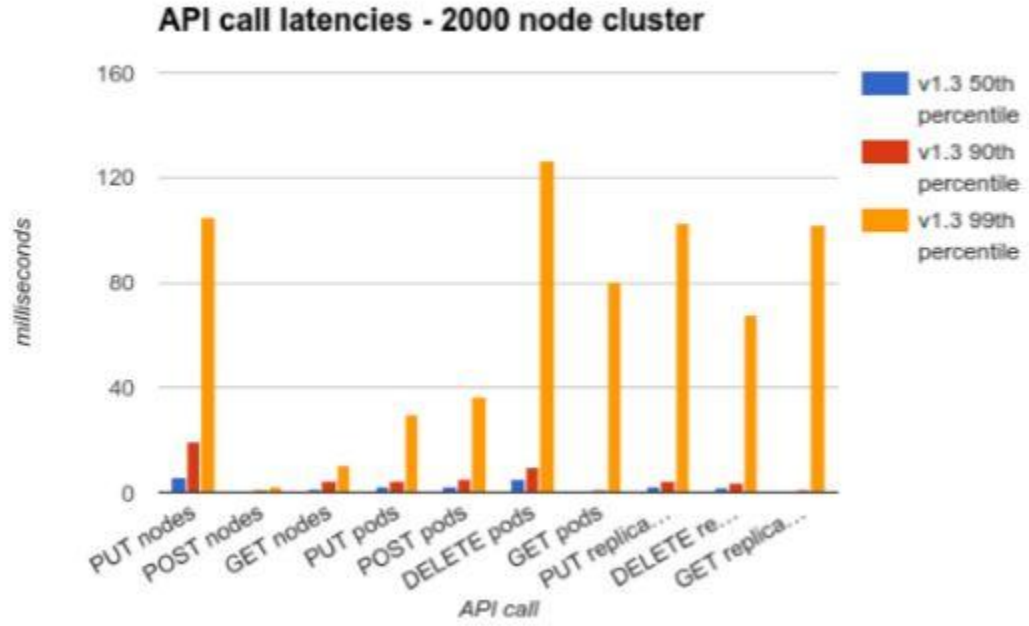

FLOCKER CONTAINERS API OR CLI
FLOCKER CONTROL SERVICE
HOST 1
FLOCKER AGENT
DOCKER DAEMON
HOST 2
FLOCKER AGENT
DOCKER DAEMON
Flocker Cluster
C1
C2
V1
V2
C2
V2
Data Center
S1
S2
Shared storage backend
Key
V = Volume
C = Container
S = Shared Storage

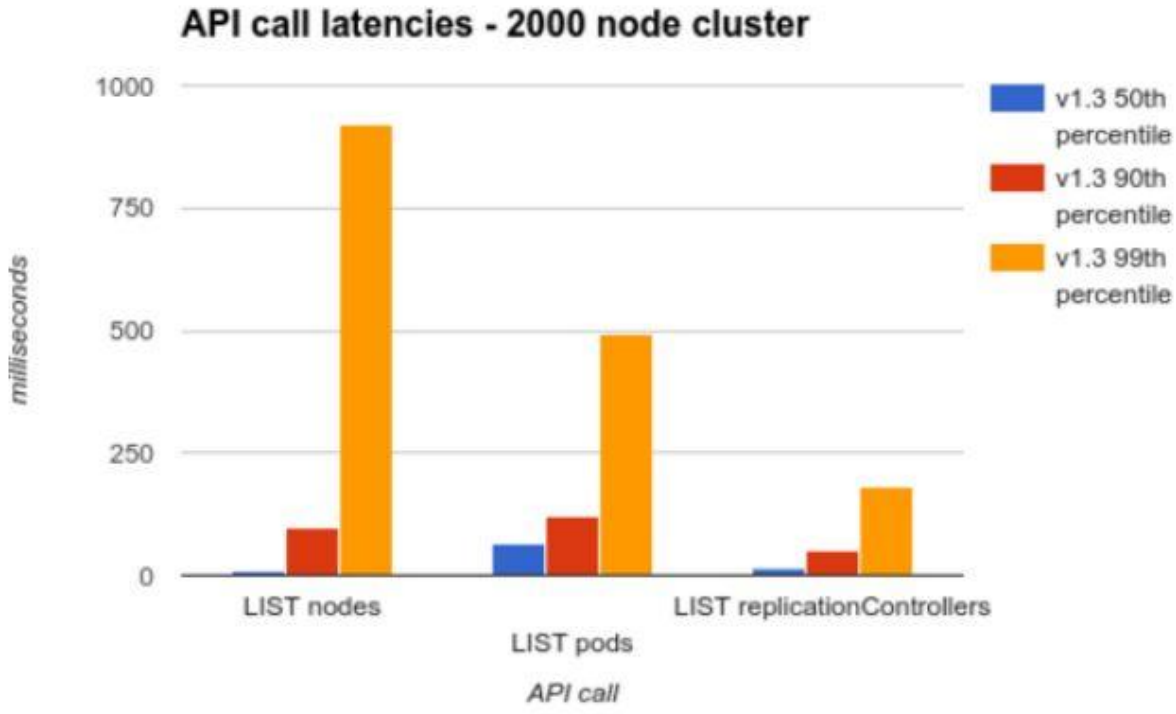

Kubernetes Pods
Pod
Pod
Pod
Mount
Mount
Mount
Torus Storage Pool
etcd
Node Disks

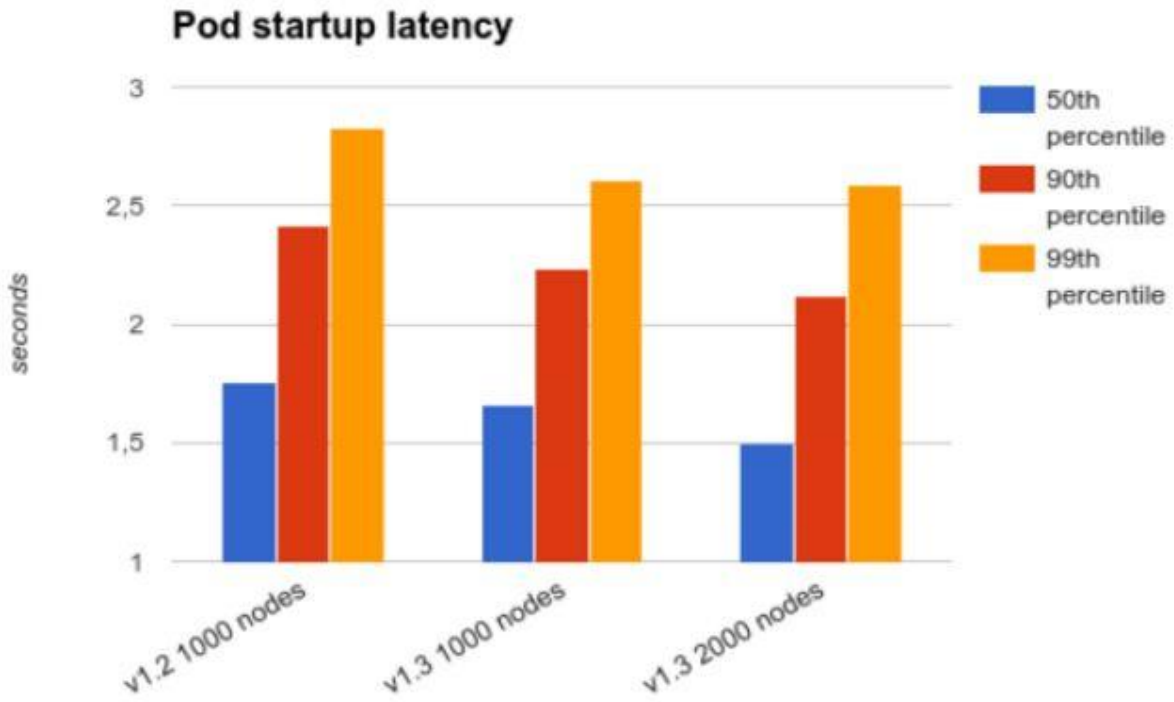
Chapter 8: Running Stateful Applications with Kubernetes

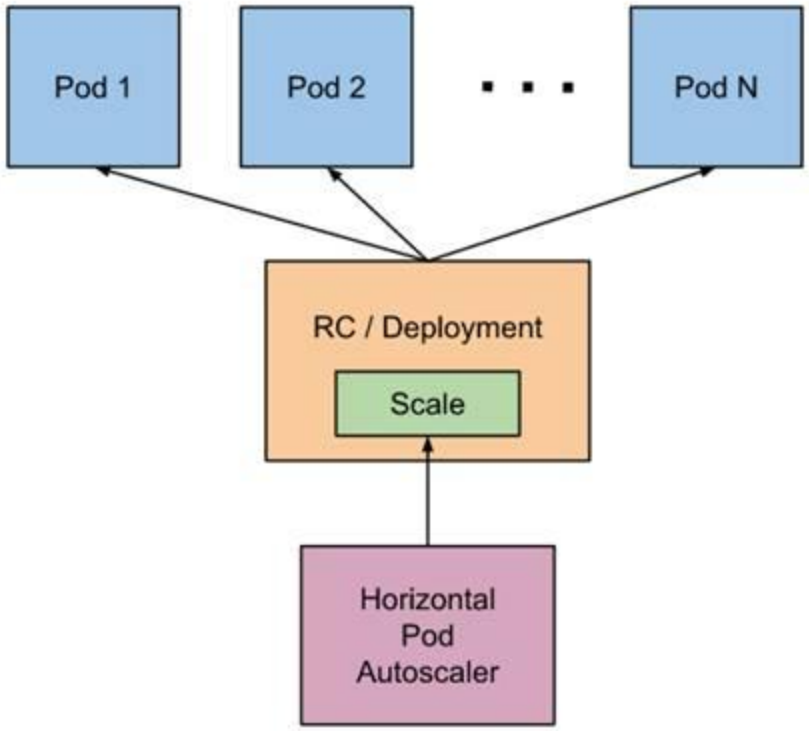


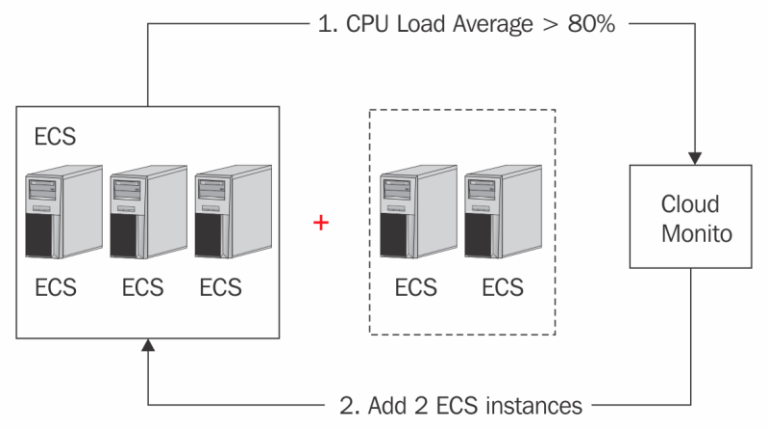
Chapter 9: Rolling Updates, Scalability, and Quotas

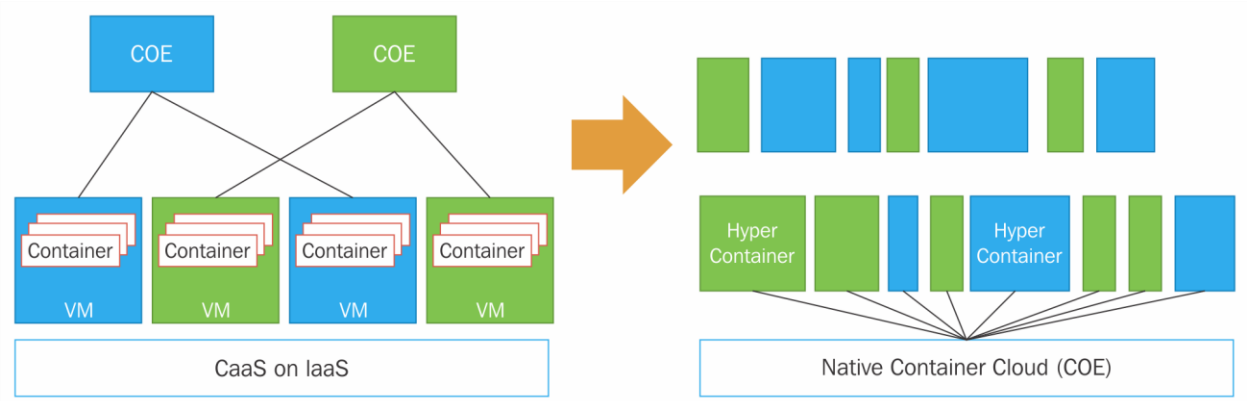


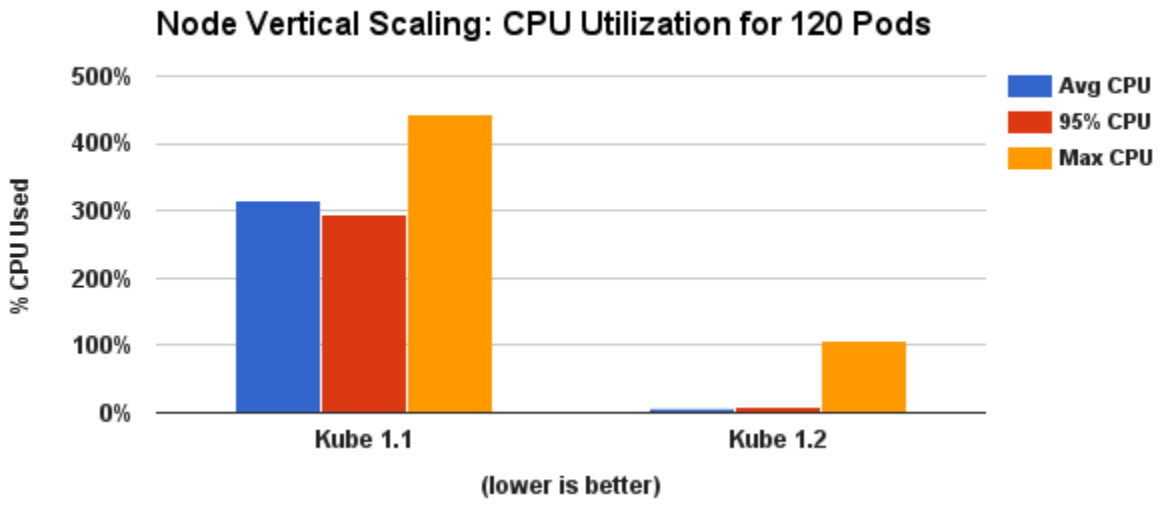



Pod startup latency
seconds
3
2,5
2
1,5
1
v1.2 1000 nodes
v1.3 1000 nodes
v1.3 2000 nodes
50th percentile
90th percentile
99th percentile

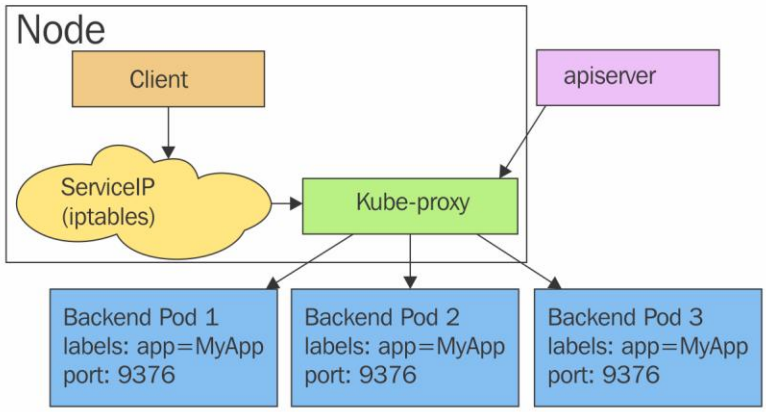

Pod 1
Pod 2
Pod N
RC / Deployment
Scale
Horizontal Pod Autoscaler

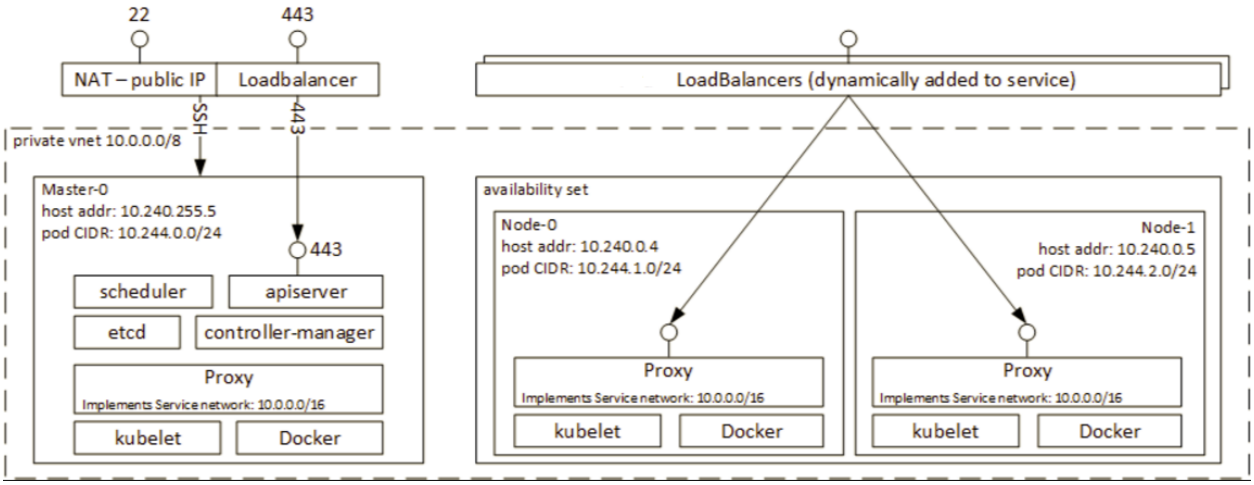

1. CPU Load Average > 80%
ECS
ECS ECS ECS
+
ECS ECS
Cloud Monito
2. Add 2 ECS instances

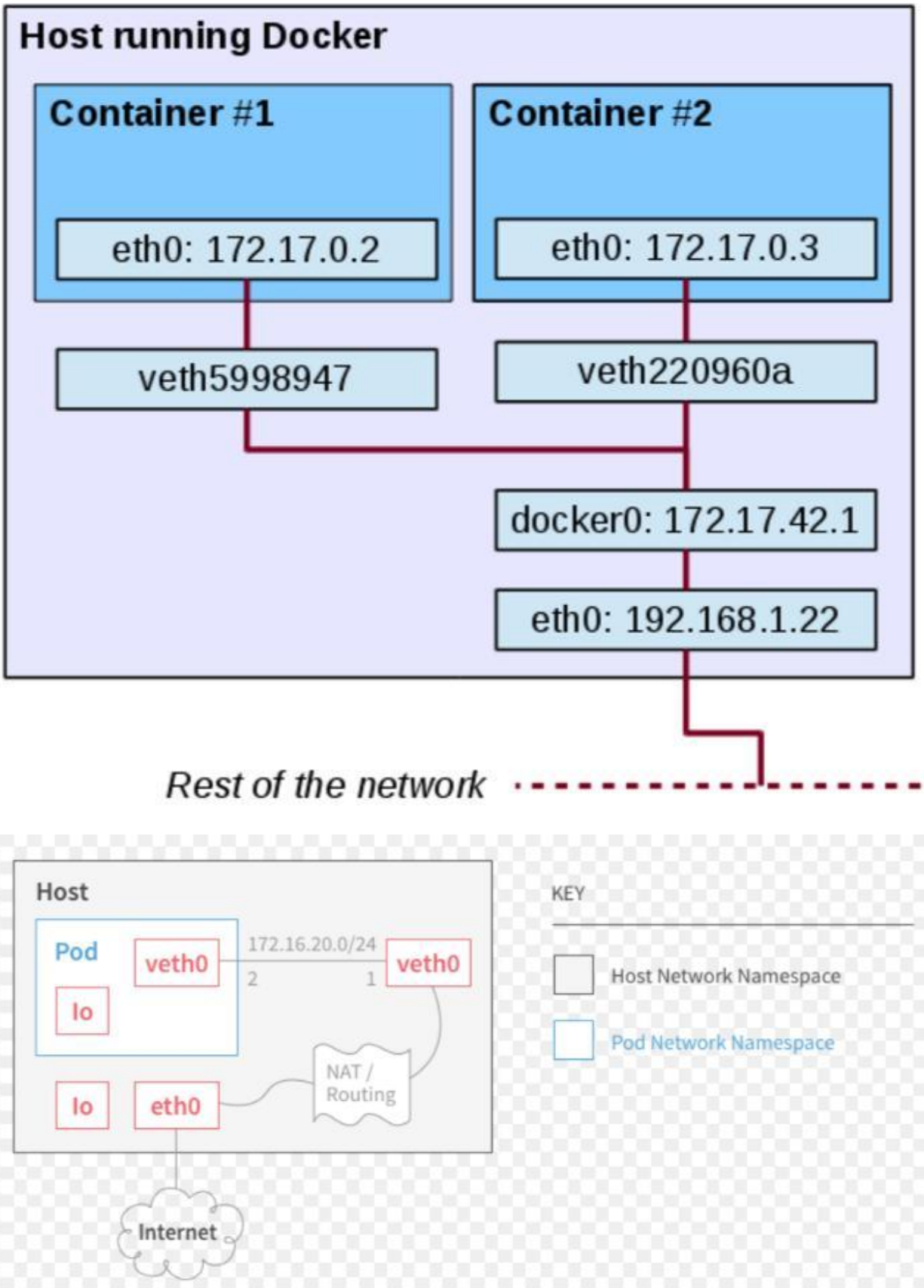

COE
COE
Container
Container
Container
Container
VM
VM
VM
VM
CaaS on IaaS
Hyper Container
Hyper Container
Native Container Cloud (COE)

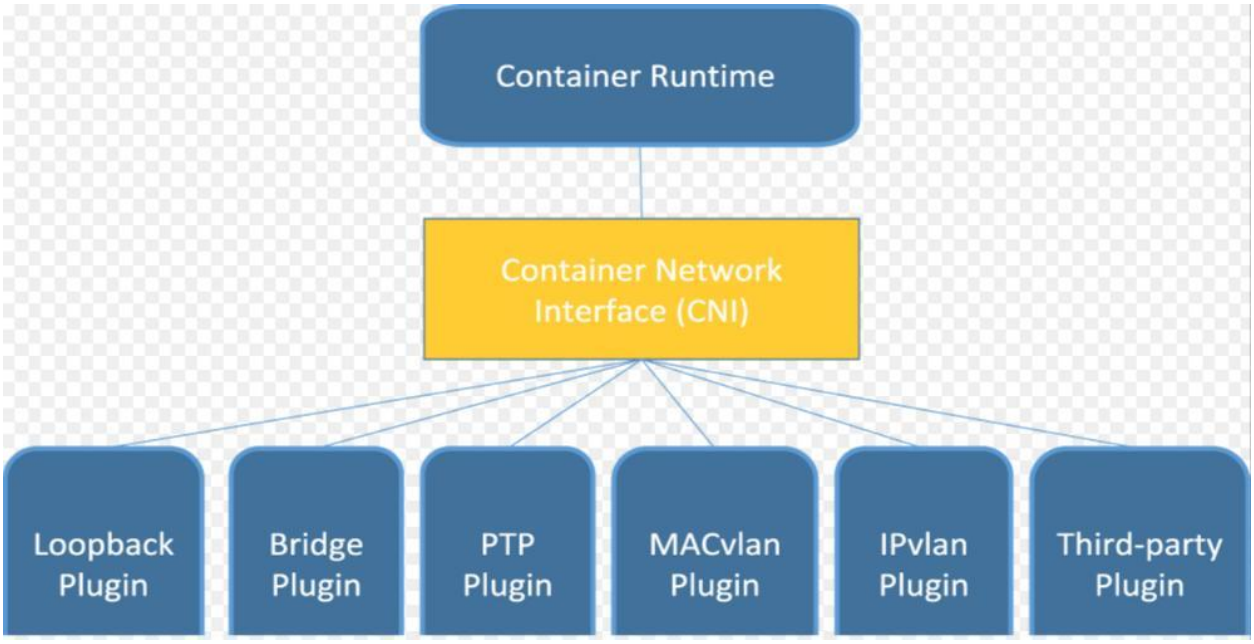

Node Vertical Scaling: CPU Utilization for 120 Pods
% CPU Used
500%
400%
300%
200%
100%
0%
Kube 1.1
Kube 1.2
(lower is better)
Avg CPU
95% CPU
Max CPU

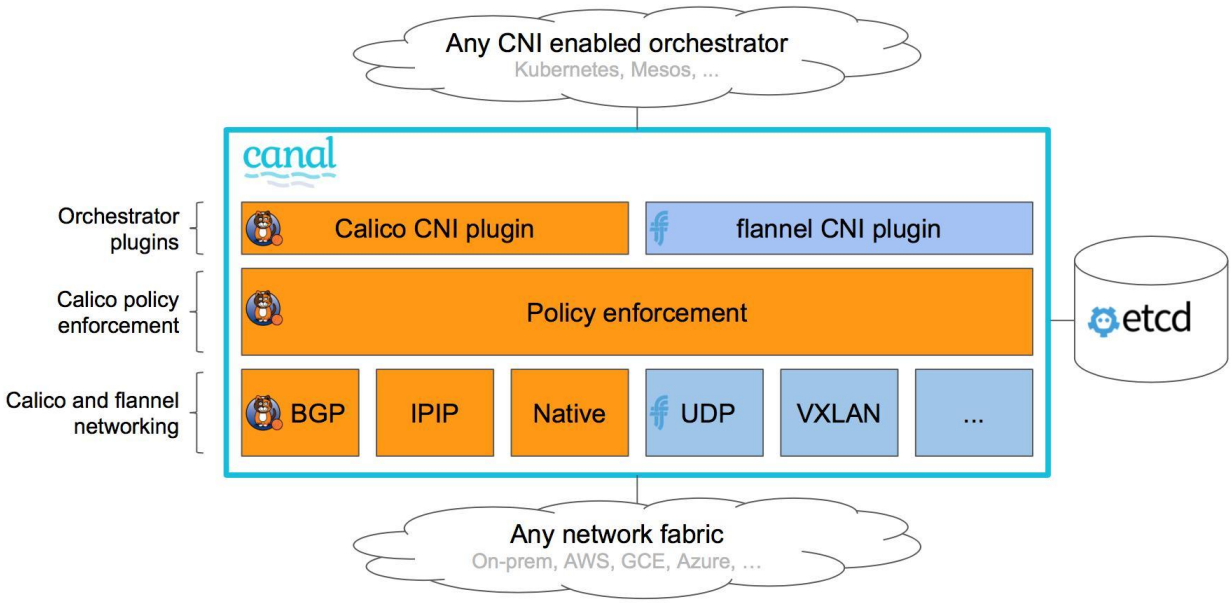
Chapter 10: Advanced Kubernetes Networking

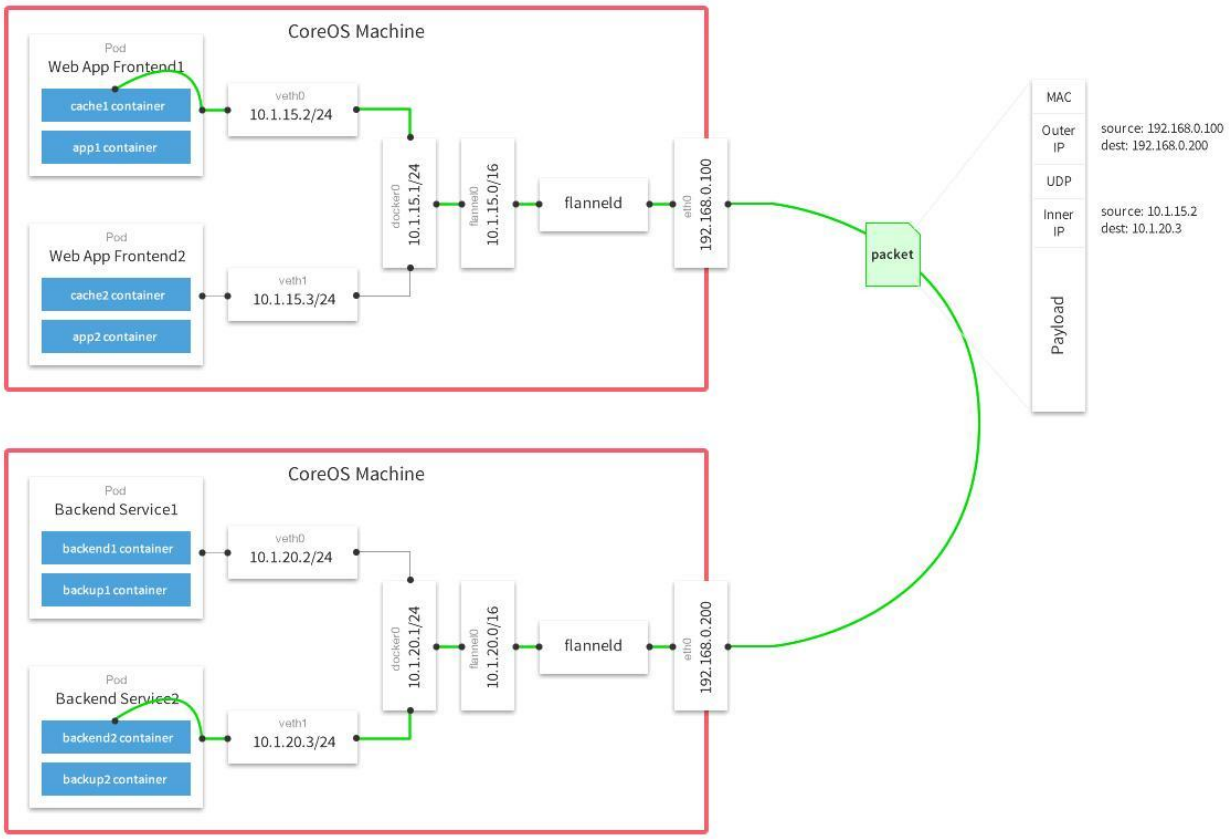





Host running Docker
Container #1
eth0: 172.17.0.2
Container #2
eth0: 172.17.0.3
veth5998947
veth220960a
docker0: 172.17.42.1
eth0: 192.168.1.22
Rest of the network
Host
Pod
veth0
lo
172.16.20.0/24
2
1
veth0
NAT / Routing
lo
eth0
Internet
KEY
Host Network Namespace
Pod Network Namespace

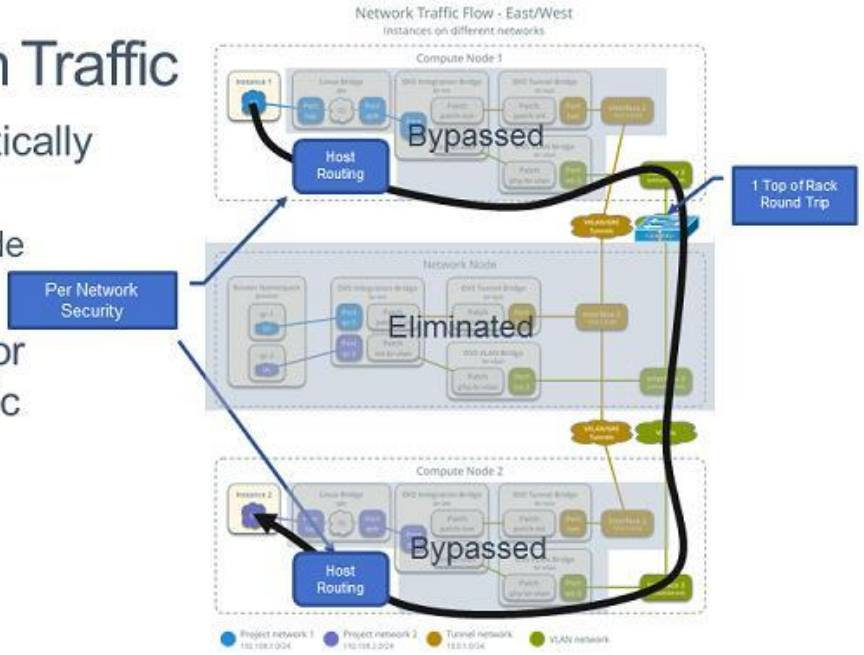

Container Runtime
Container Network Interface (CNI)
Loopback Plugin
Bridge Plugin
PTP Plugin
MACvlan Plugin
IPvlan Plugin
Third-party Plugin

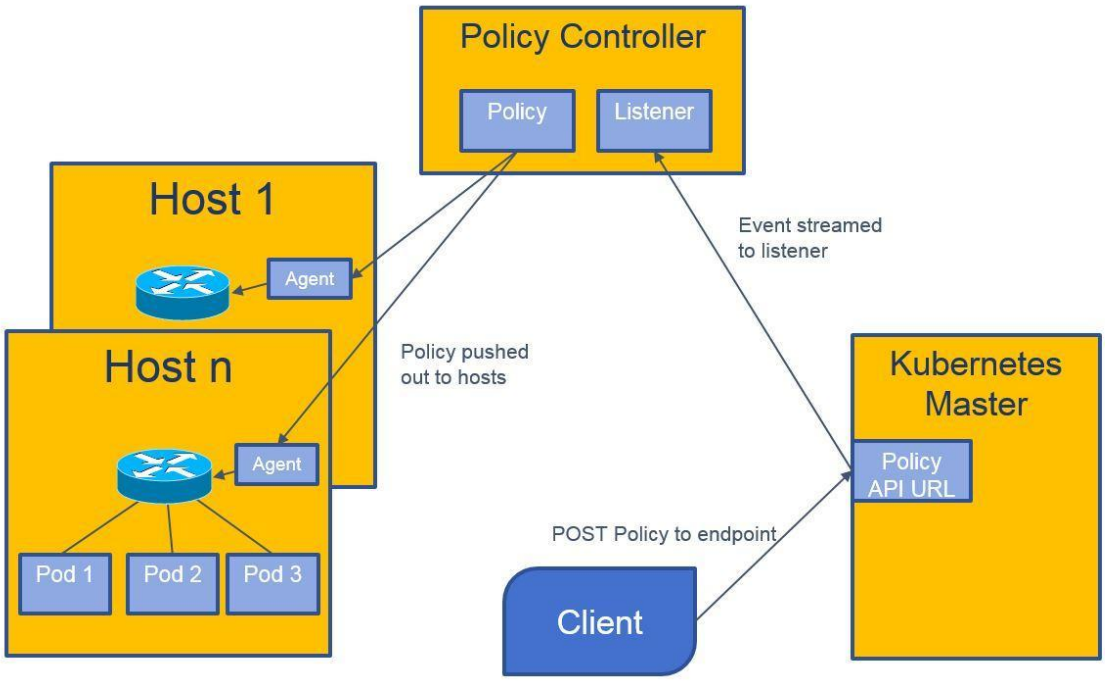

Any CNI enabled orchestrator
Kubernetes, Mesos, ...
canal
Orchestrator plugins
Calico CNI plugin
flannel CNI plugin
Calico policy enforcement
Policy enforcement
etcd
Calico and flannel networking
BGP
IPIP
Native
UDP
VXLAN
...
Any network fabric
On-prem, AWS, GCE, Azure, ...

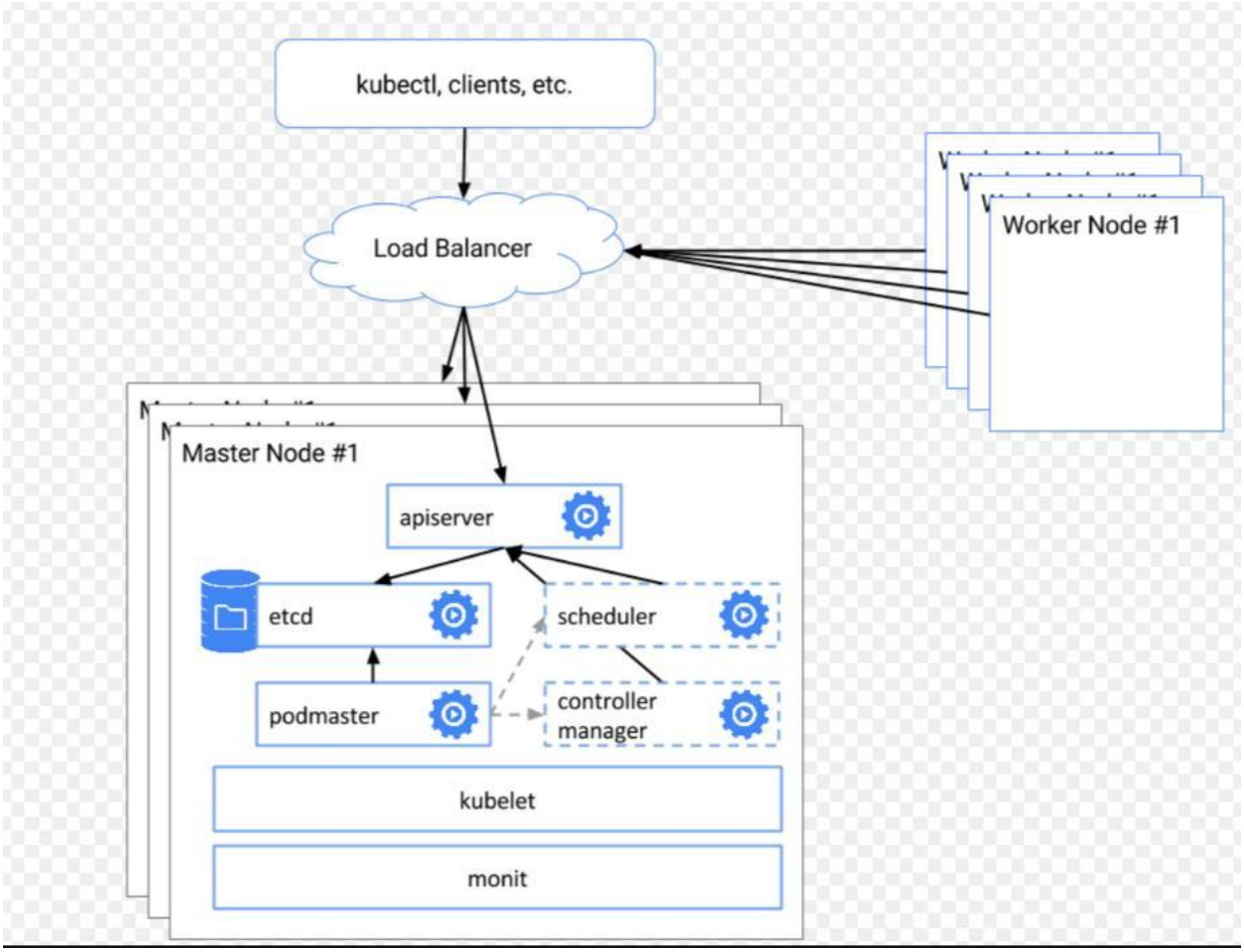


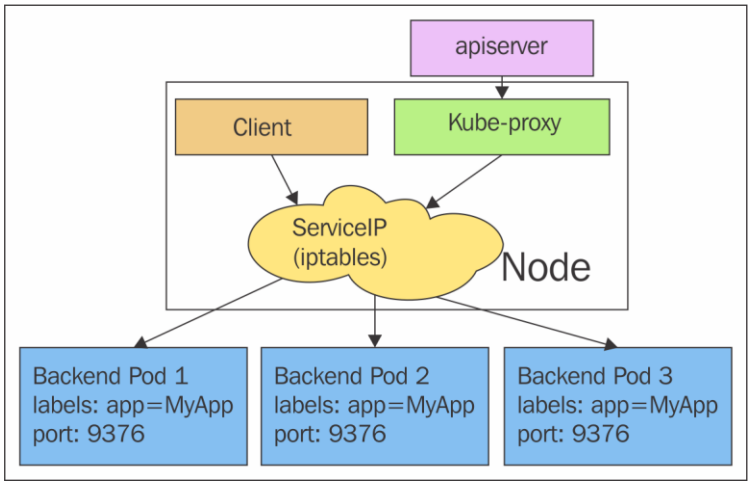
North/South Traffic

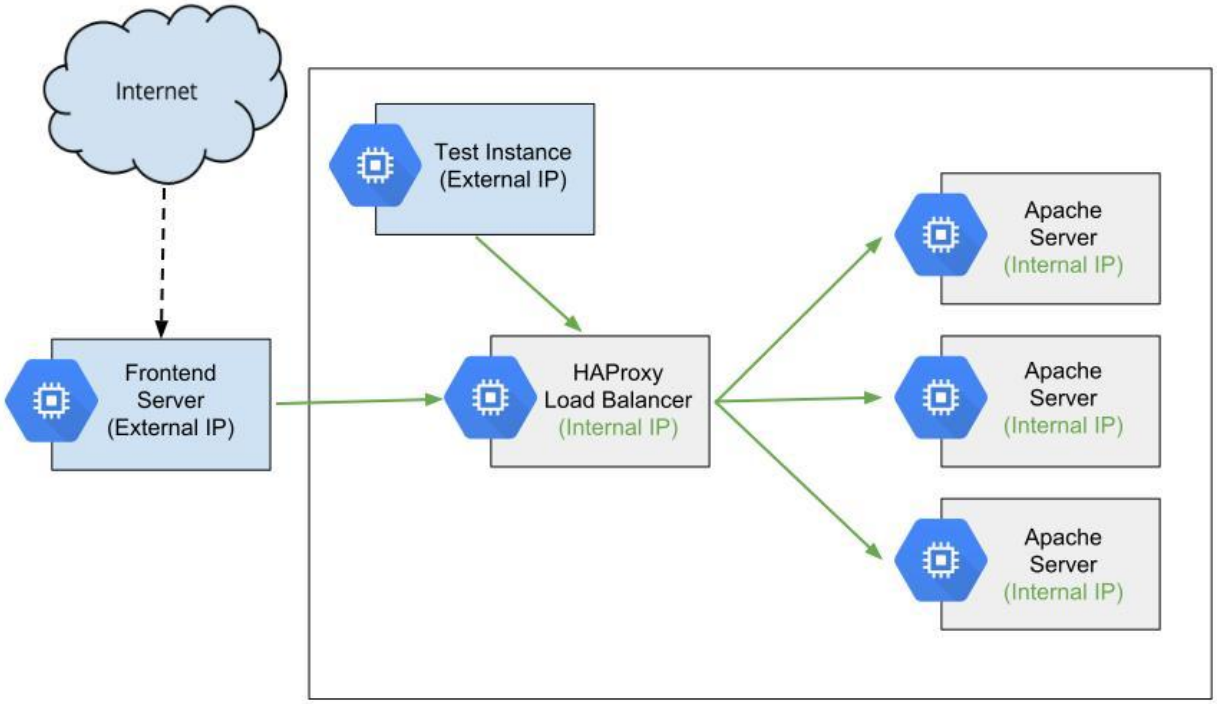
- Latency dramatically reduced
  - No Network node
  - No encap
- Identical path for East/West traffic



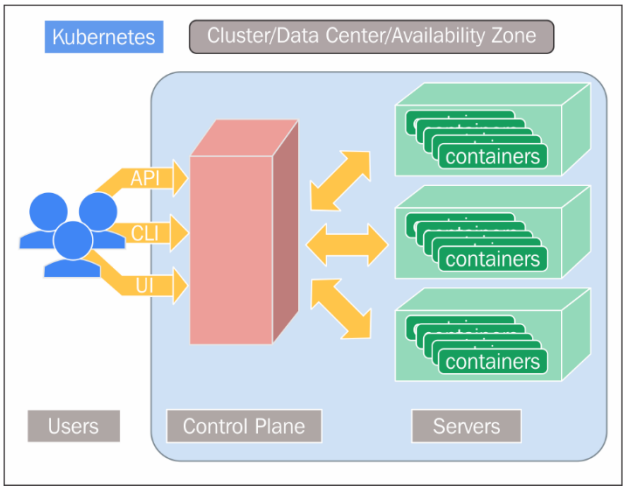

Policy Controller
Policy
Listener
Host 1
Agent
Host n
Agent
Pod 1
Pod 2
Pod 3
Event streamed
to listener
Policy pushed
out to hosts
Kubernetes
Master
Policy
API URL
POST Policy to endpoint
Client

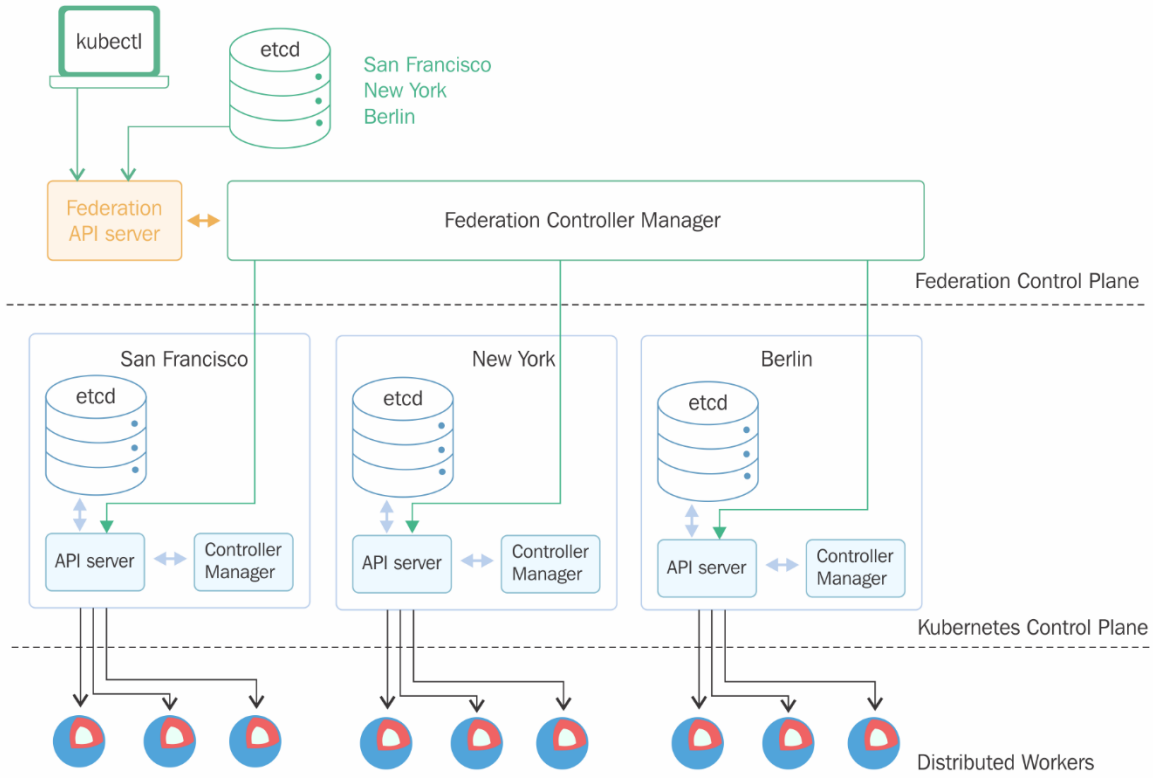

kubectl, clients, etc.
Load Balancer
Worker Node #1
Master Node #1
apiserver
etcd
scheduler
podmaster
controller
manager
kubelet
monit


apiserver
Client
Kube-proxy
ServiceIP
(iptables)
Node
Backend Pod 1
labels: app=MyApp
port: 9376
Backend Pod 2
labels: app=MyApp
port: 9376
Backend Pod 3
labels: app=MyApp
port: 9376

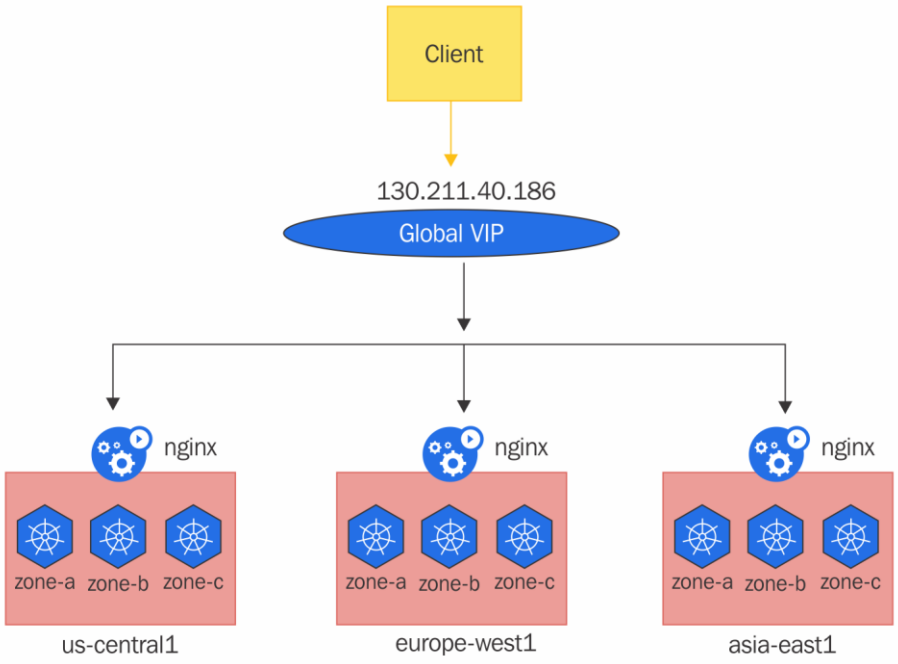

Internet
Test Instance
(External IP)
Frontend
Server
(External IP)
HAProxy
Load Balancer
(Internal IP)
Apache
Server
(Internal IP)
Apache
Server
(Internal IP)
Apache
Server
(Internal IP)

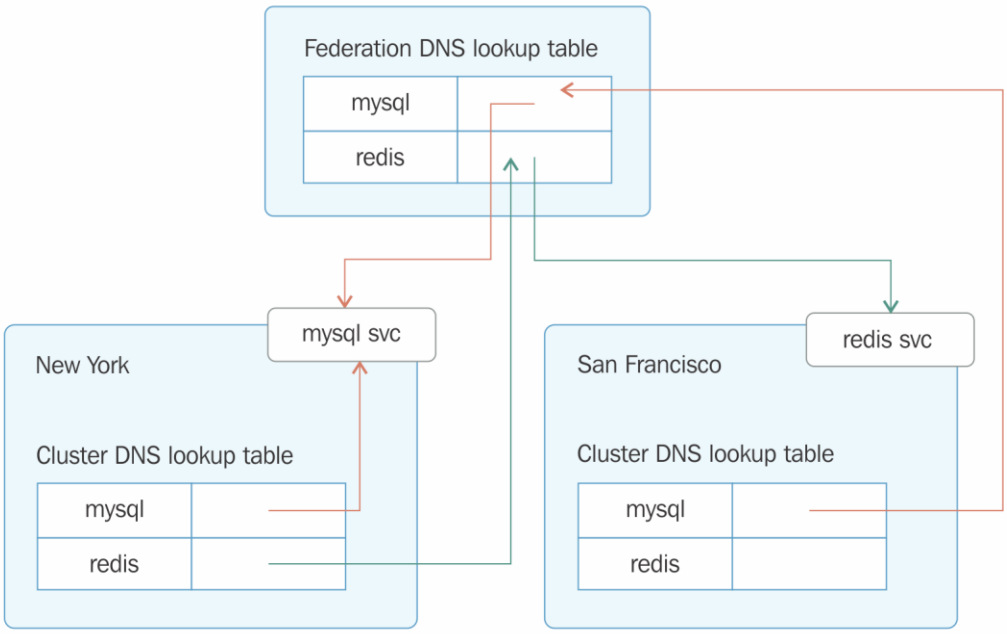
Chapter 11: Running Kubernetes on Multiple Clouds and Cluster Federation



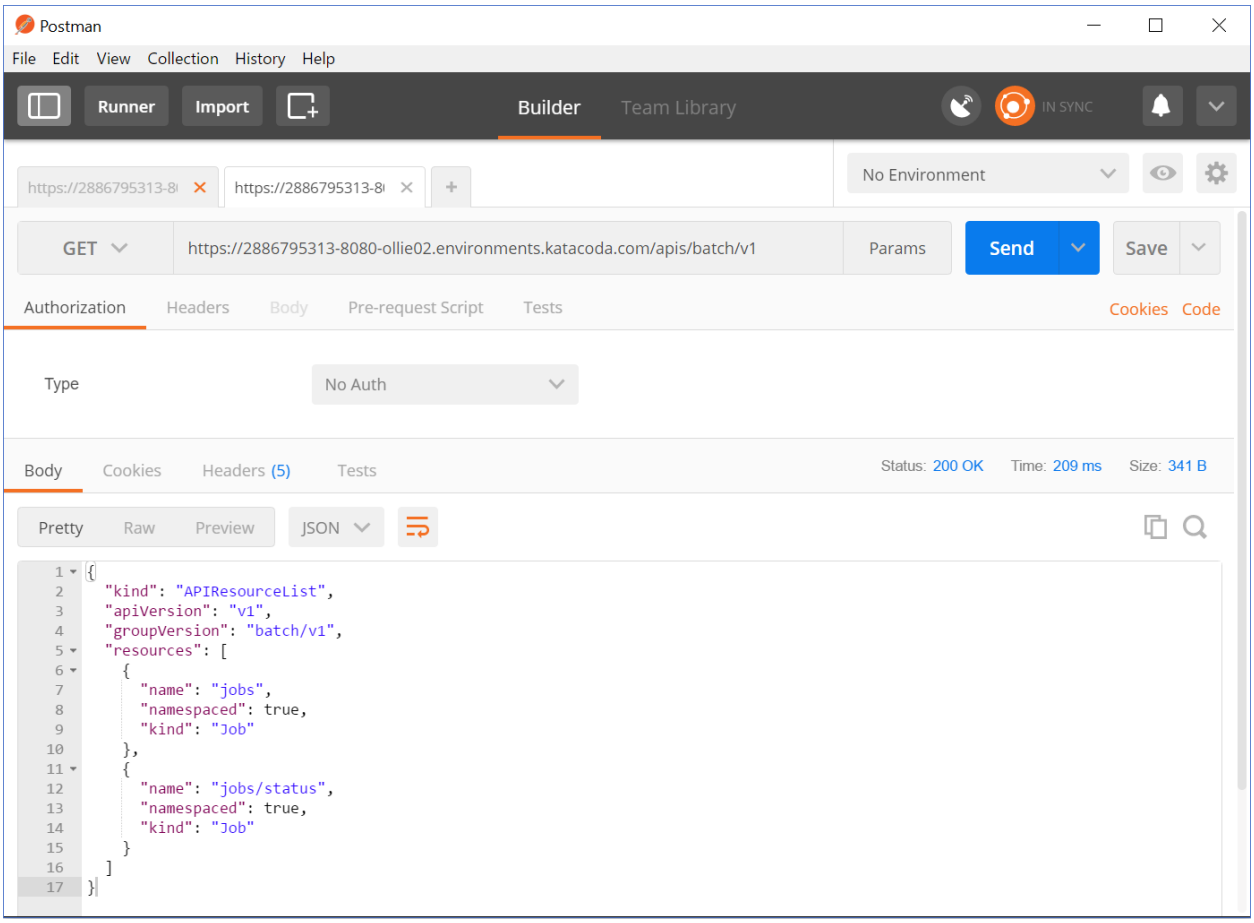



observe
act
diff

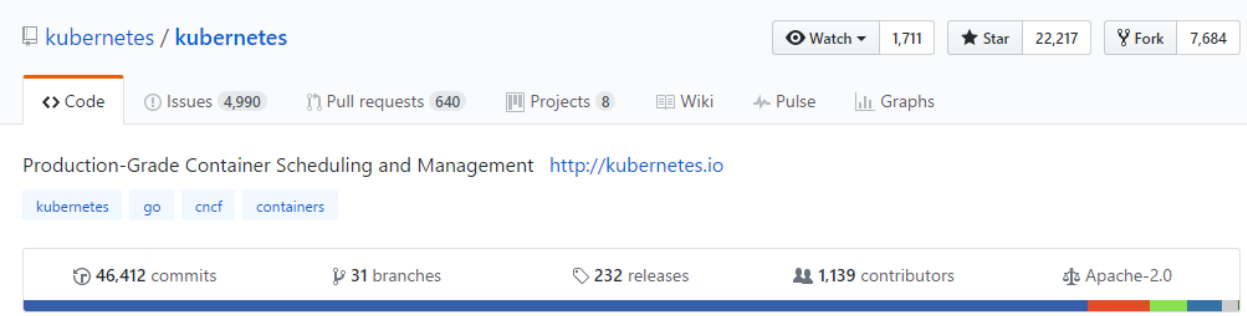

Client
130.211.40.186
Global VIP
nginx
zone-a
zone-b
zone-c
us-central1
nginx
zone-a
zone-b
zone-c
europe-west1
nginx
zone-a
zone-b
zone-c
asia-east1


Federation DNS lookup table
mysql
redis
mysql svc
redis svc
New York
Cluster DNS lookup table
mysql
redis
San Francisco
Cluster DNS lookup table
mysql
redis

Chapter 12: Customizing Kubernetes - APIs and Plugins



Chapter 14: The Future of Kubernetes



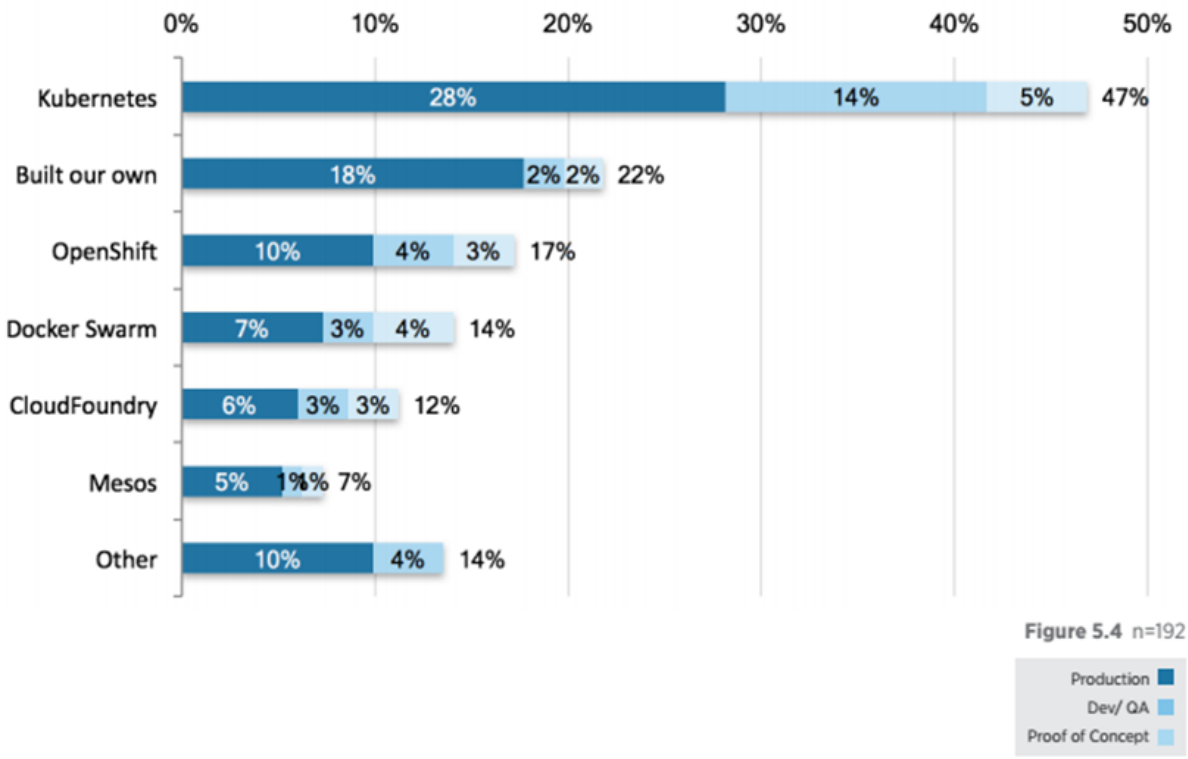


Figure 5.4 n=192