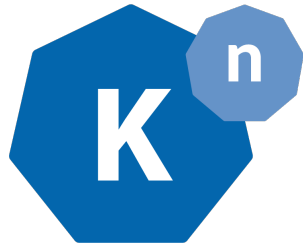
Inside Knative Serving

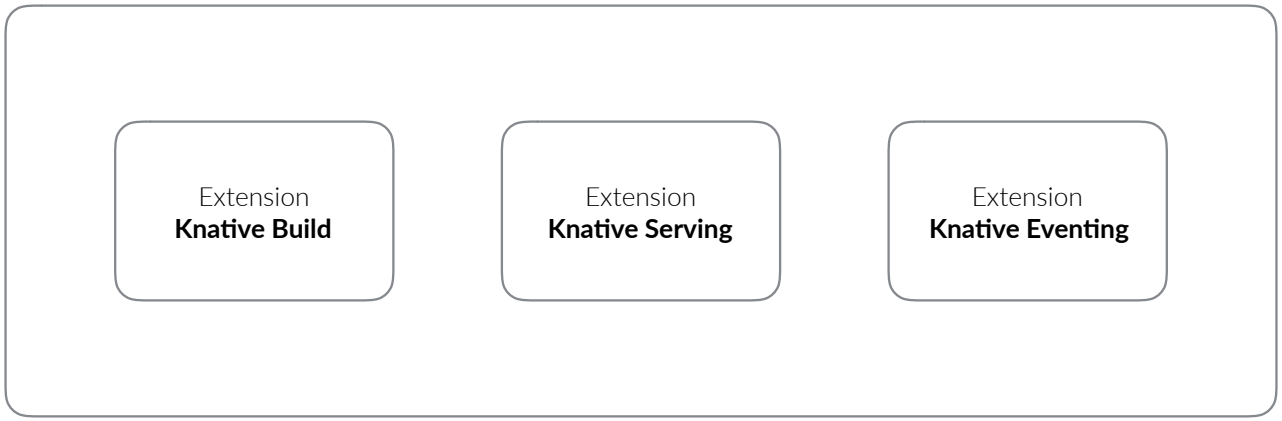
KubeCon Shanghai 2019



Introduction

介绍

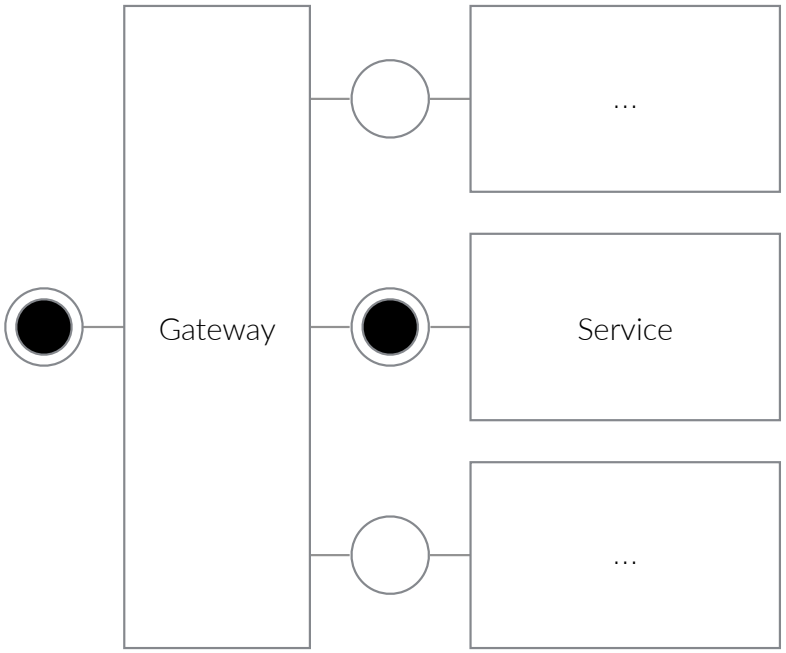
Extension
Knative


Extension
Knative Build
Extension
Knative Serving
Extension
Knative Eventing

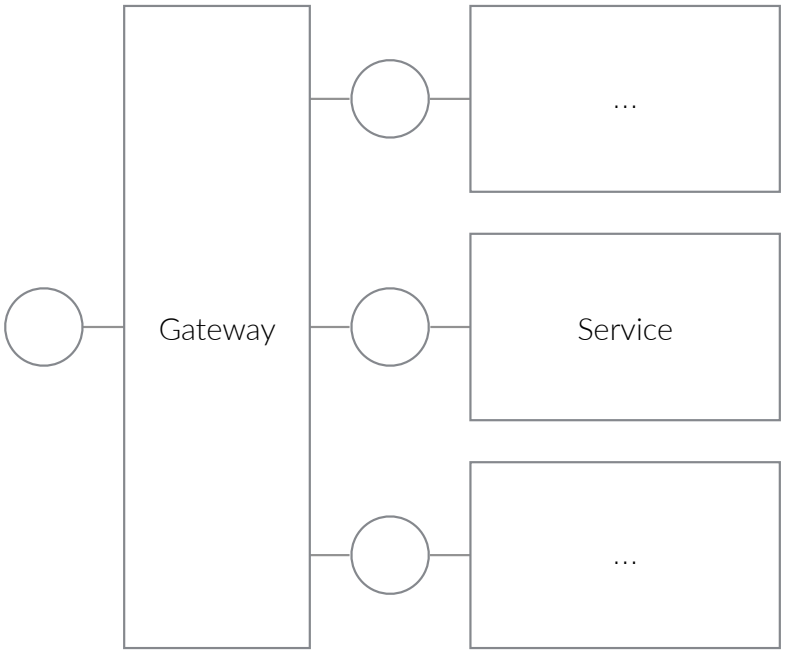
Lifecycle Management

Overview

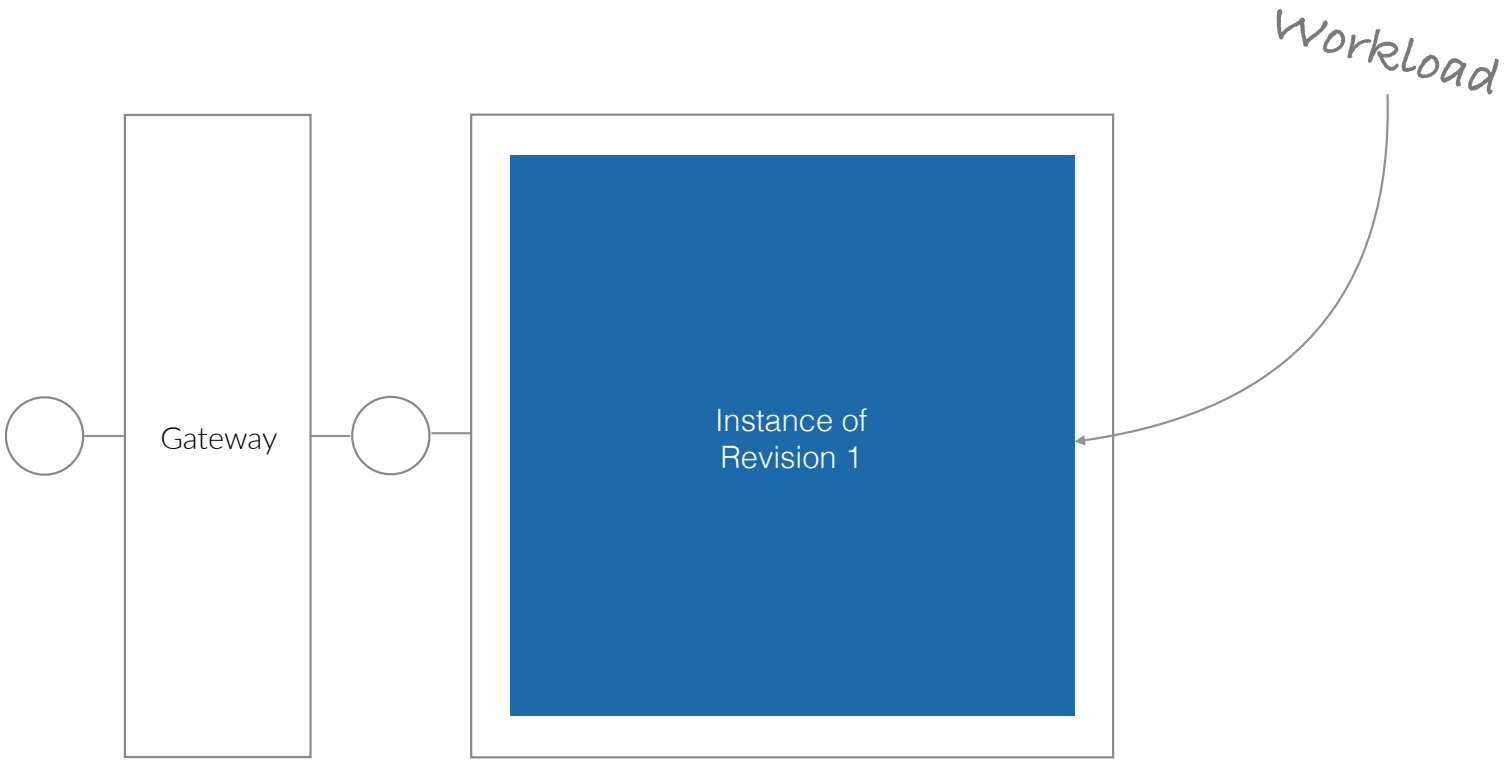
Architectural Perspective



Operational Perspective

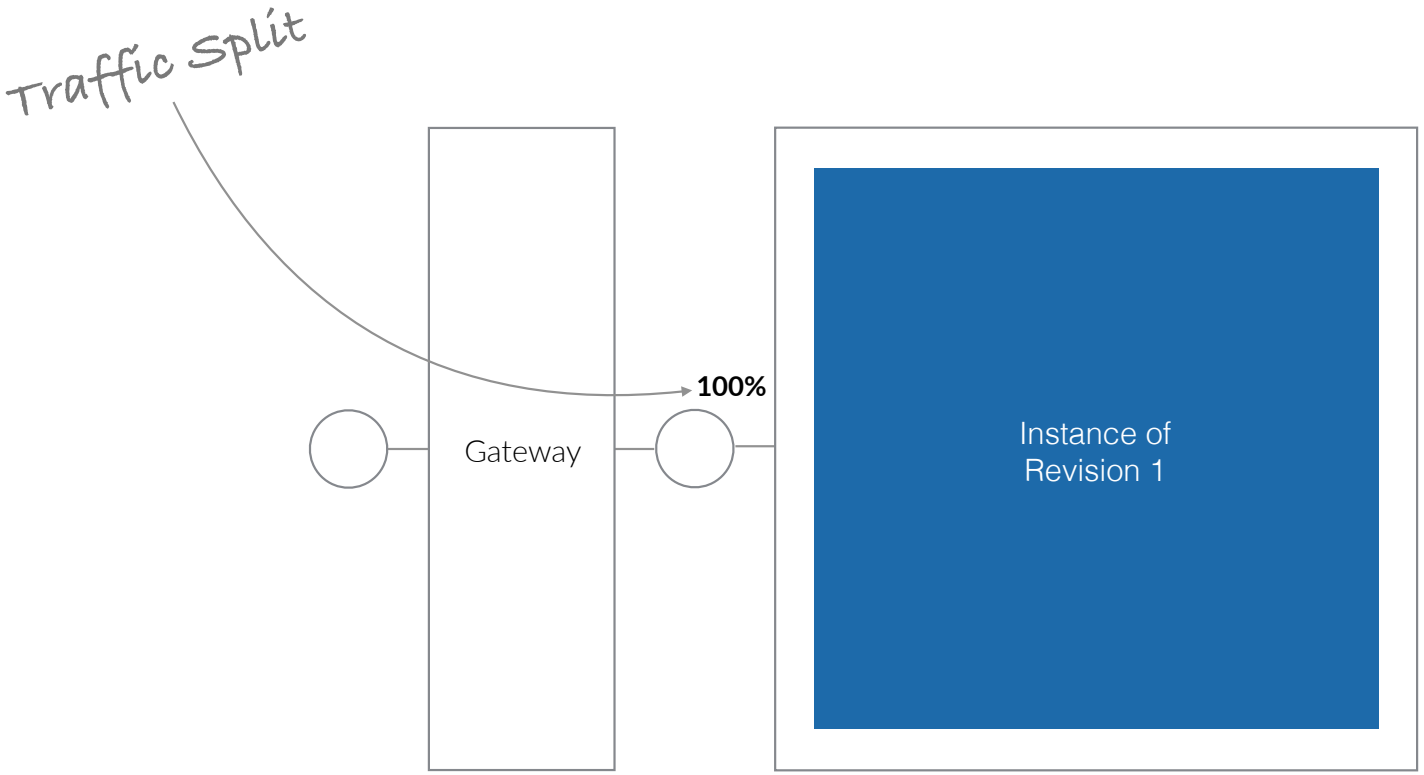


Operational Perspective



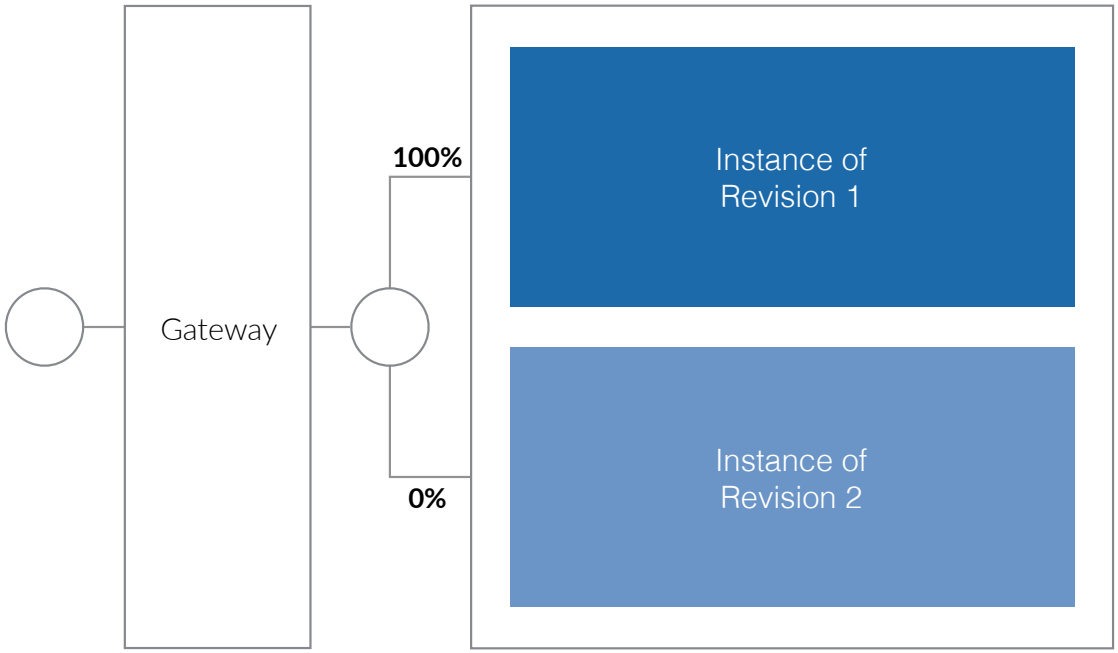
create(Workload Specification for Revision 1)

Operational Perspective



create(Traffic Split Specification (Revision 1, 100%))

Operational Perspective

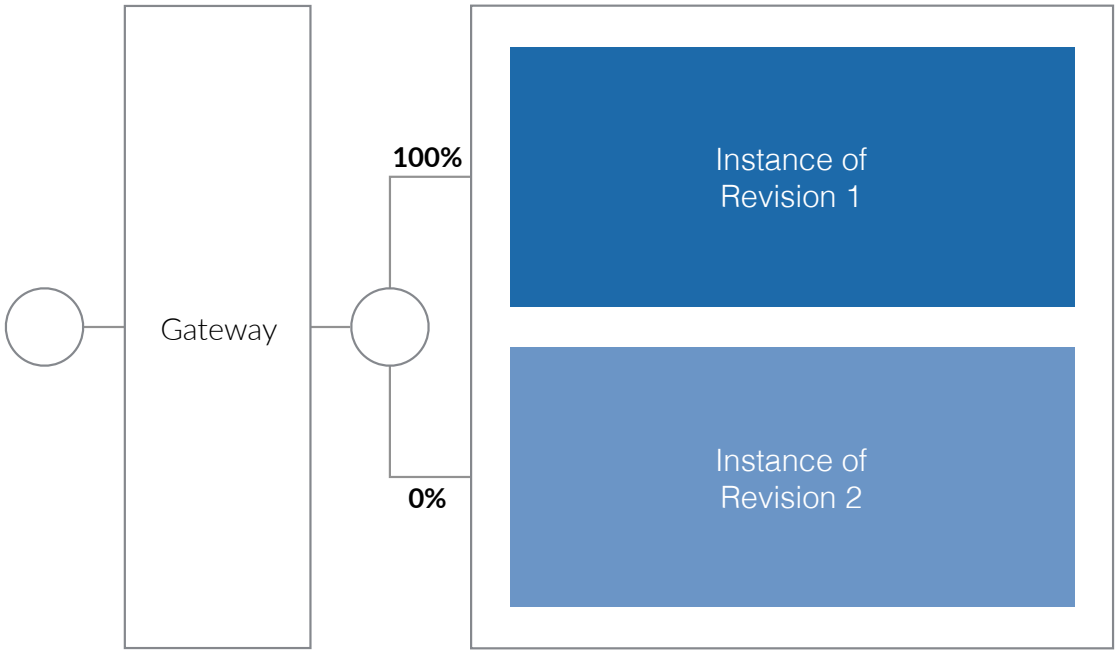


create(Workload Specification for Revision 2)

Operational Perspective

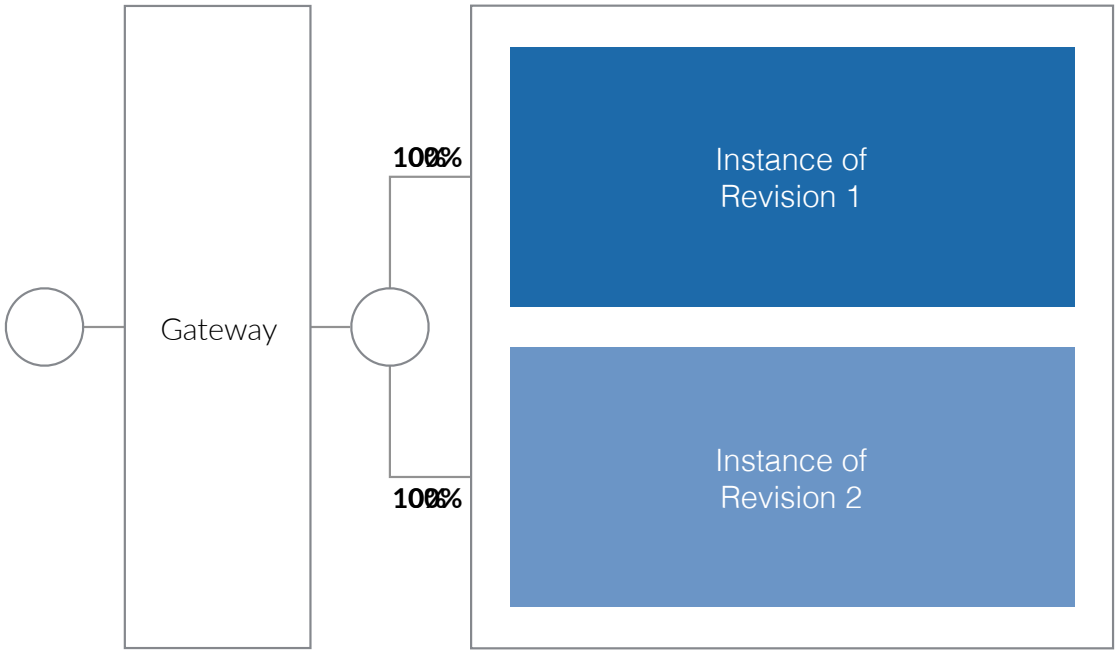
Immediate Rollout

Ready. Set. Go.



Operational Perspective

Immediate Rollout

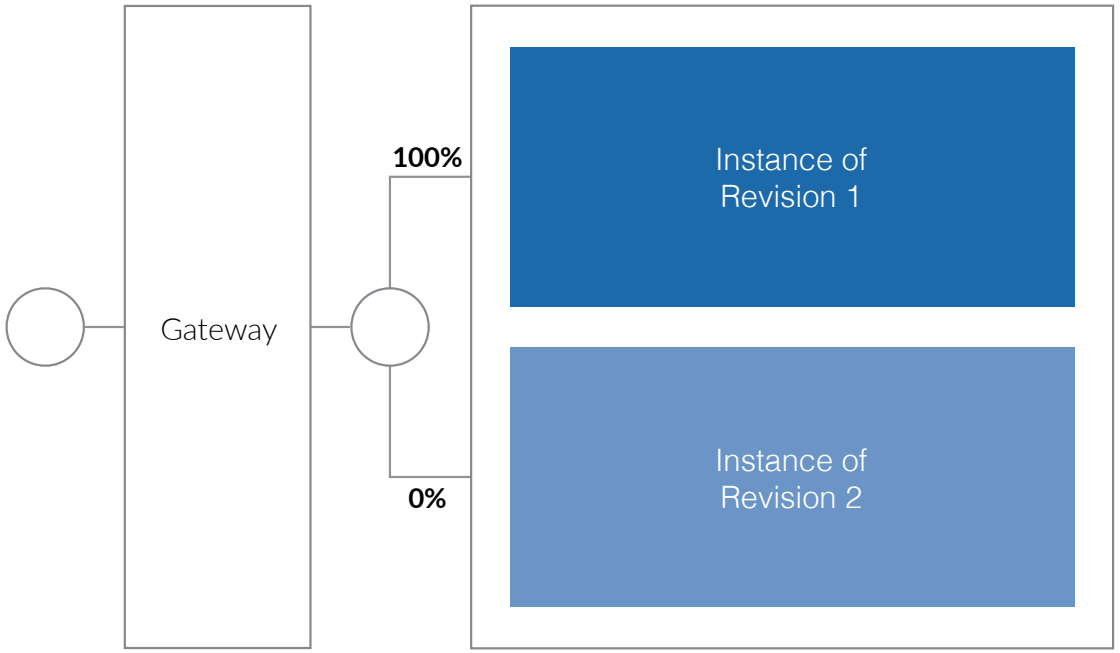


update(Traffic Split Specification (Revision 1, 0%), (Revision 2, 100%))

Operational Perspective

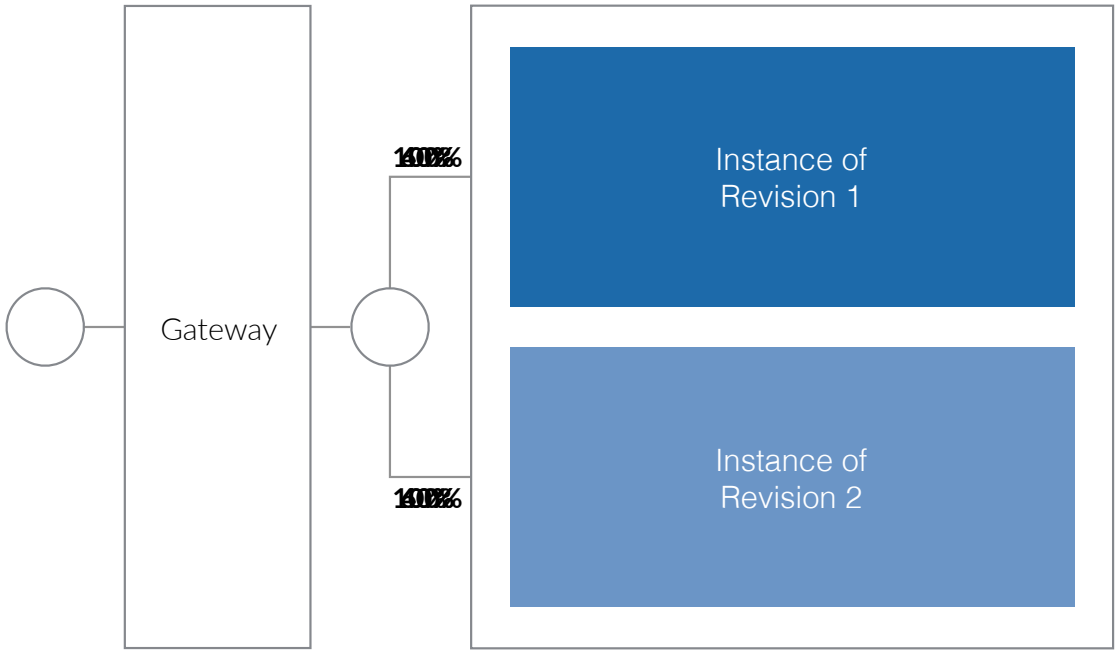
Gradual Rollout

Ready. Set. Go.



Operational Perspective

Gradual Rollout



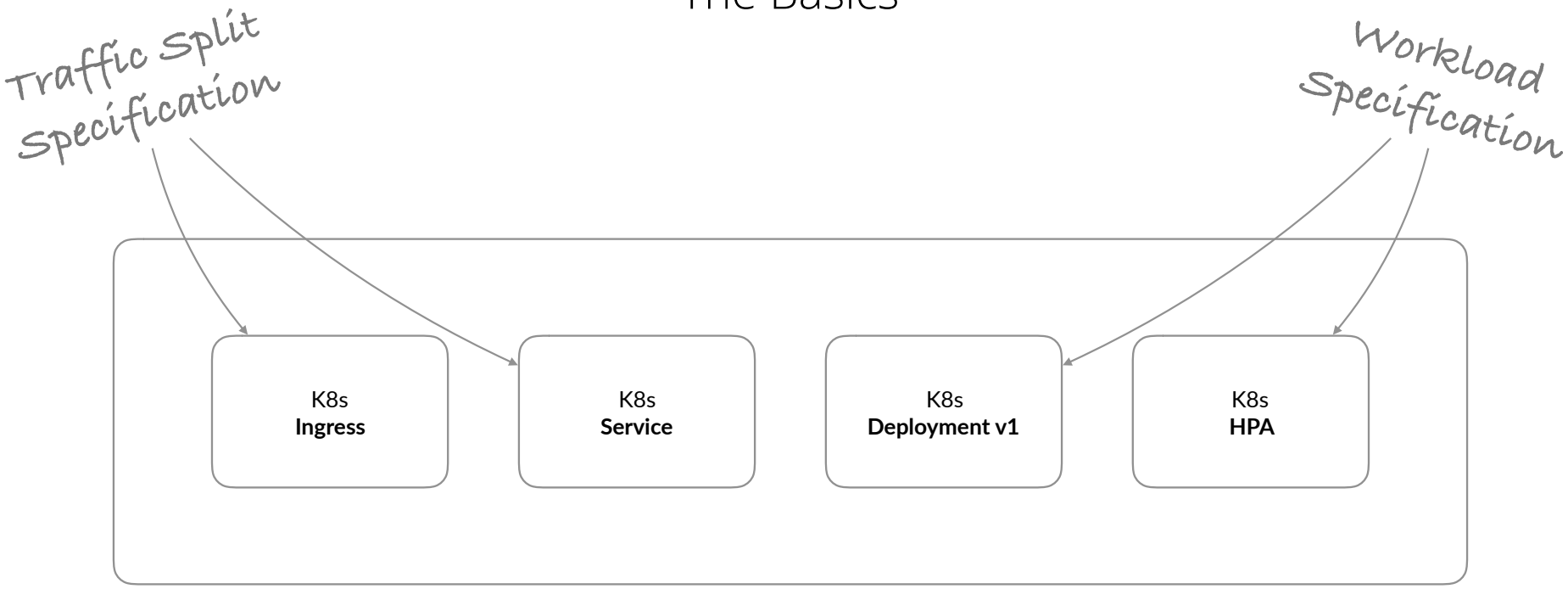
update(Traffic Split Specification (Revision 1, 60%), (Revision 2, 40%))

Lifecycle Management

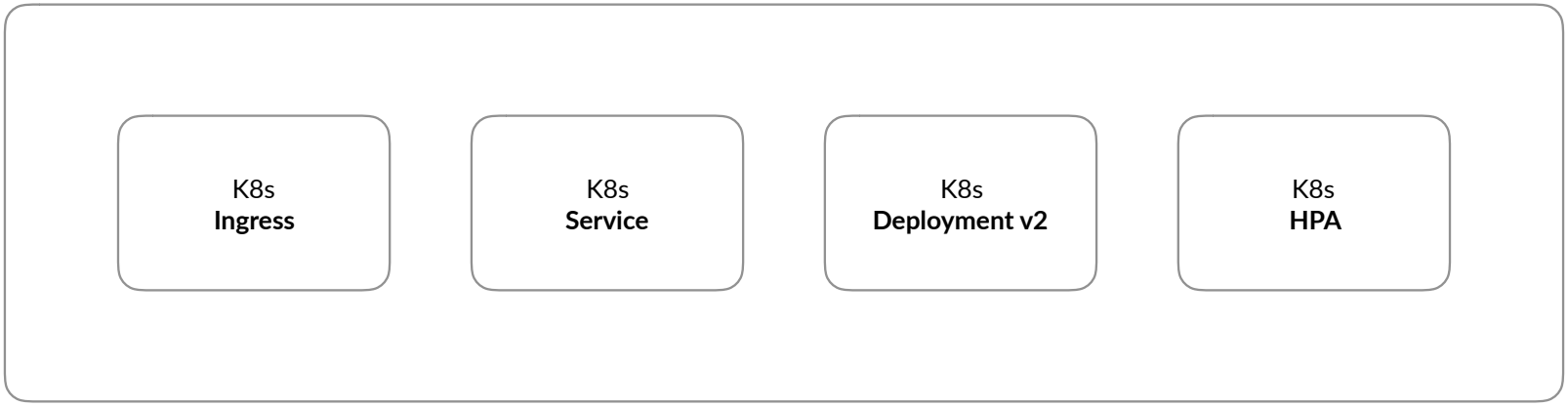
Manual Operations

Microservice

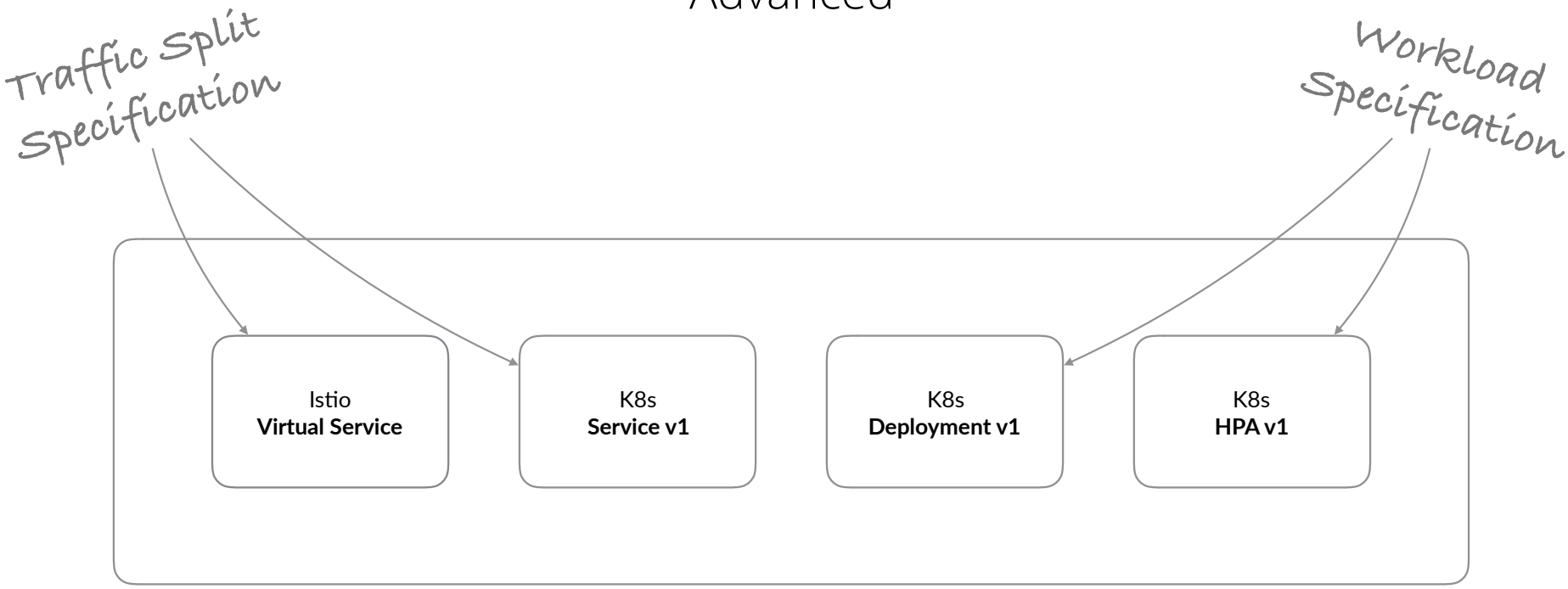
The Basics


Traffic Split Specification
Workload Specification
K8s
Ingress
K8s
Service
K8s
Deployment v1
K8s
HPA

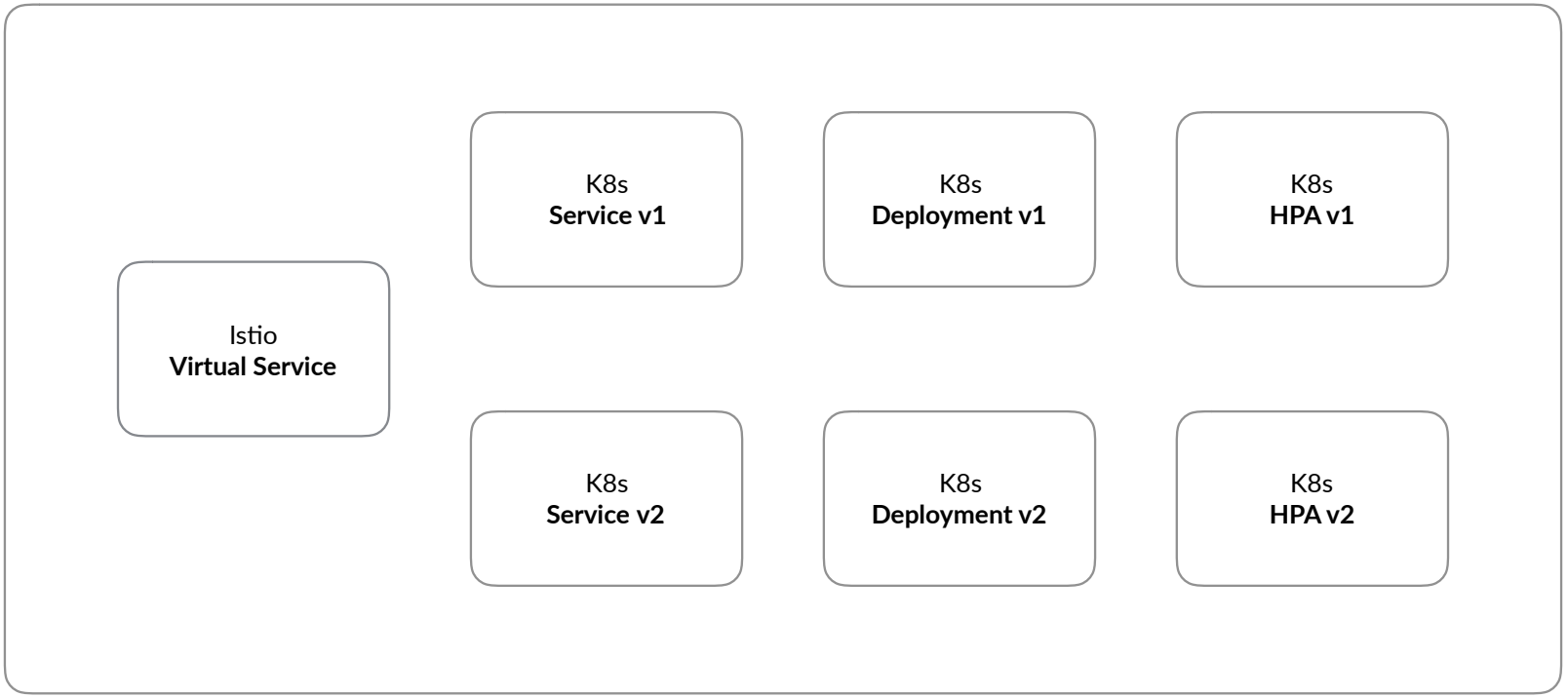
The Basics



Advanced


Traffic Split Specification
Workload Specification
Istio
Virtual Service
K8s
Service v1
K8s
Deployment v1
K8s
HPA v1

Advanced


Istio
Virtual Service
K8s
Service v1
K8s
Deployment v1
K8s
HPA v1
K8s
Service v2
K8s
Deployment v2
K8s
HPA v2

Manual Release

Initial Revision

Create
Deployment v1

Create
HPA v1

Create
Service v1

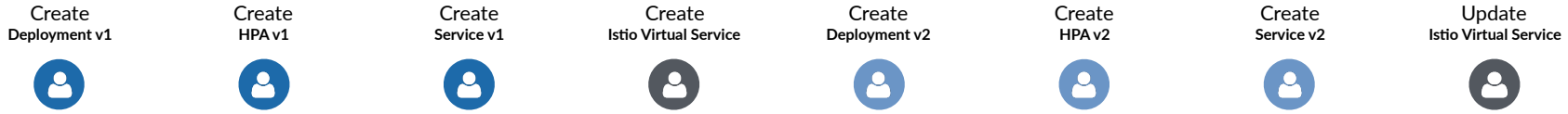
Create
Istio Virtual Service

create(Workload Specification for Revision 1)

create(Traffic Split Specification (Revision 1, 100%))

Manual Release

Subsequent Revision

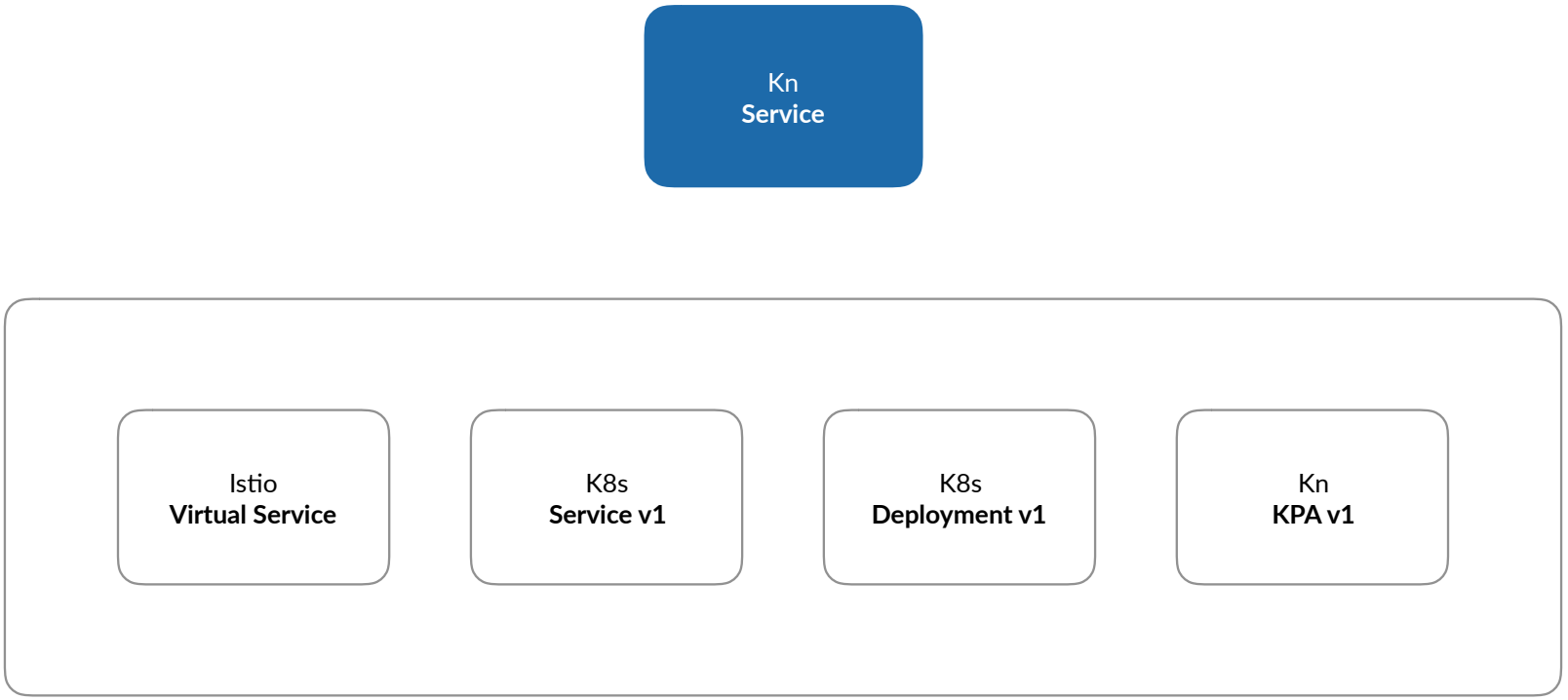


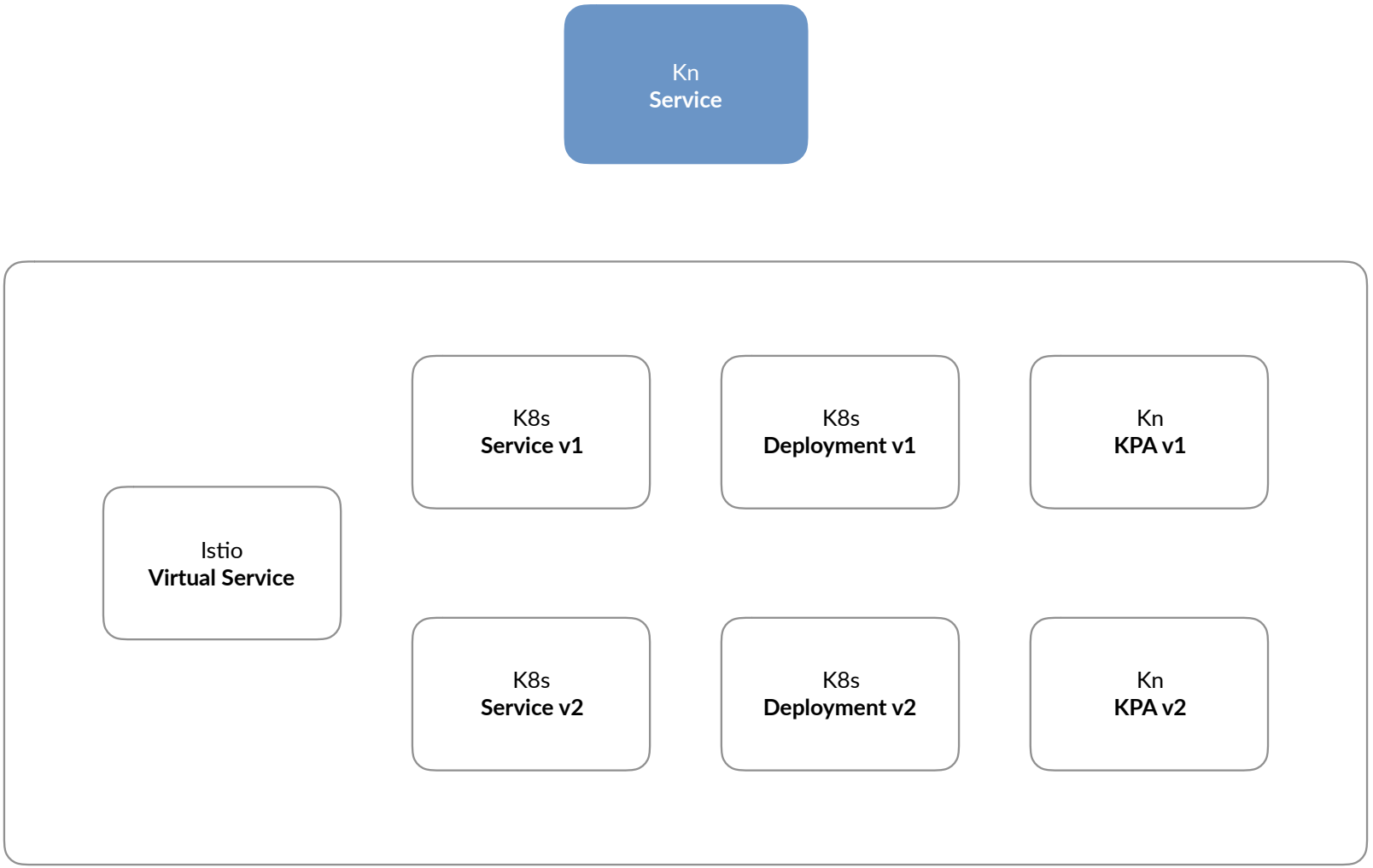
Meet
Knative Serving

Lifecycle Management

Automated Operations

Kn
Service


Kn
Service
Istio
Virtual Service
K8s
Service v1
K8s
Deployment v1
Kn
KPA v1


Kn
Service
K8s
Service v1
K8s
Deployment v1
Kn
KPA v1
Istio
Virtual Service
K8s
Service v2
K8s
Deployment v2
Kn
KPA v2

Automated Release

Initial Revision

Create
Kn Service

Create
Deployment v1

Create
HPA v1

Create
Service v1

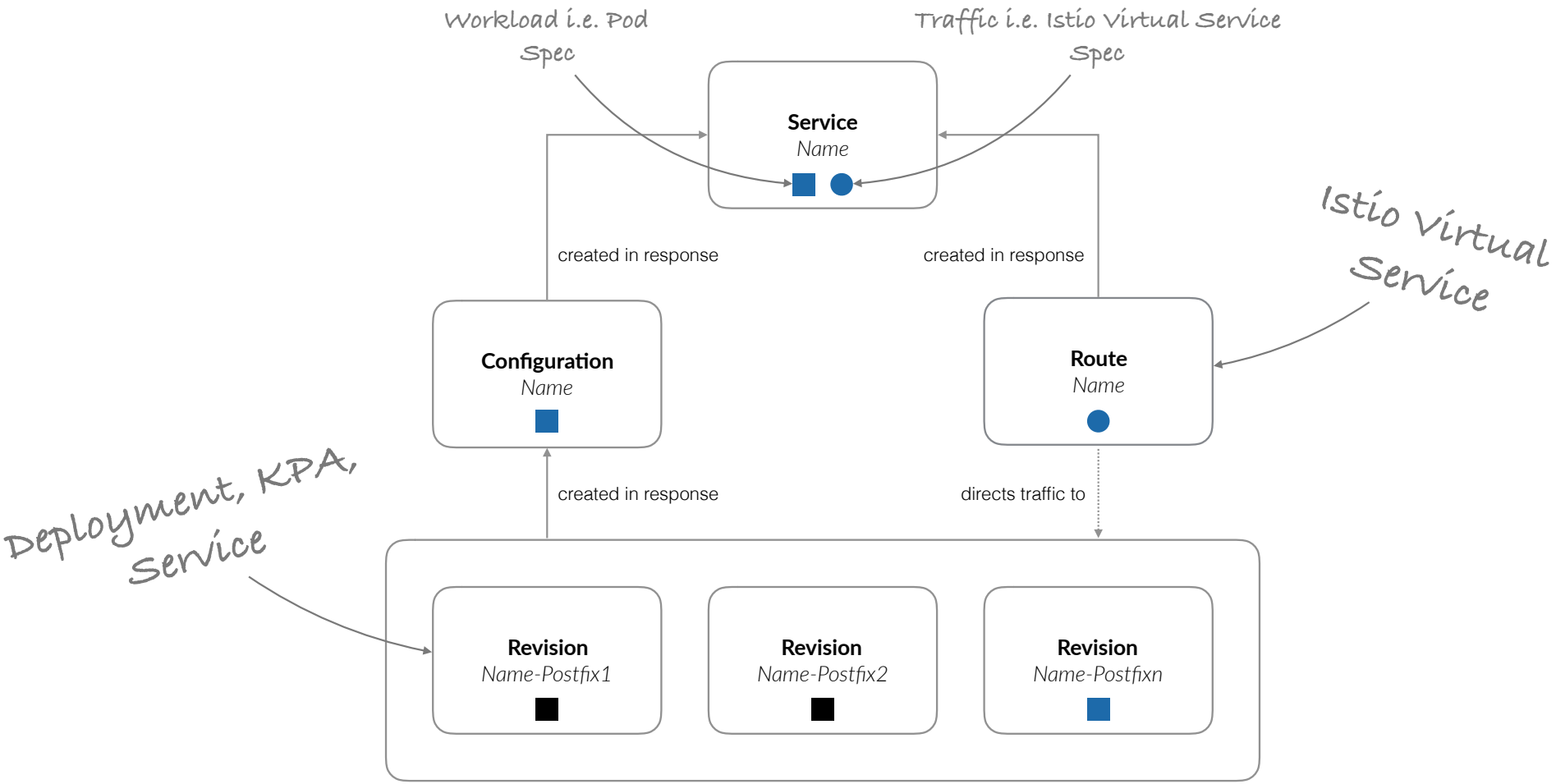
Create
Istio Virtual Service

Automated Release

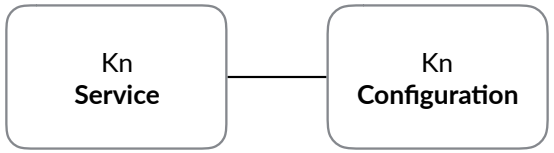
Subsequent Revision

Lifecycle Management

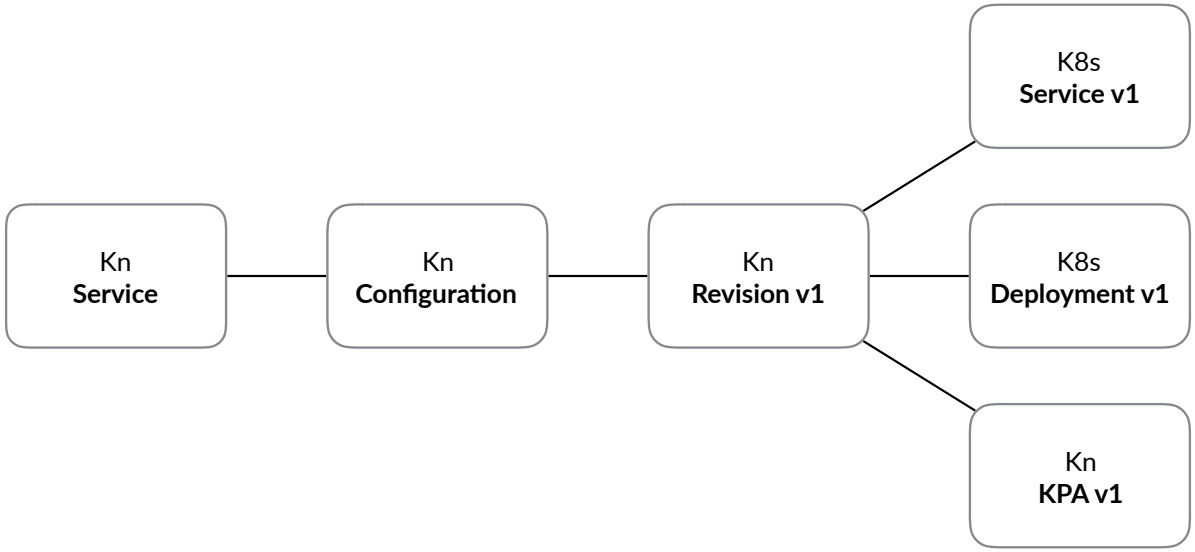
Knative Serving Internals

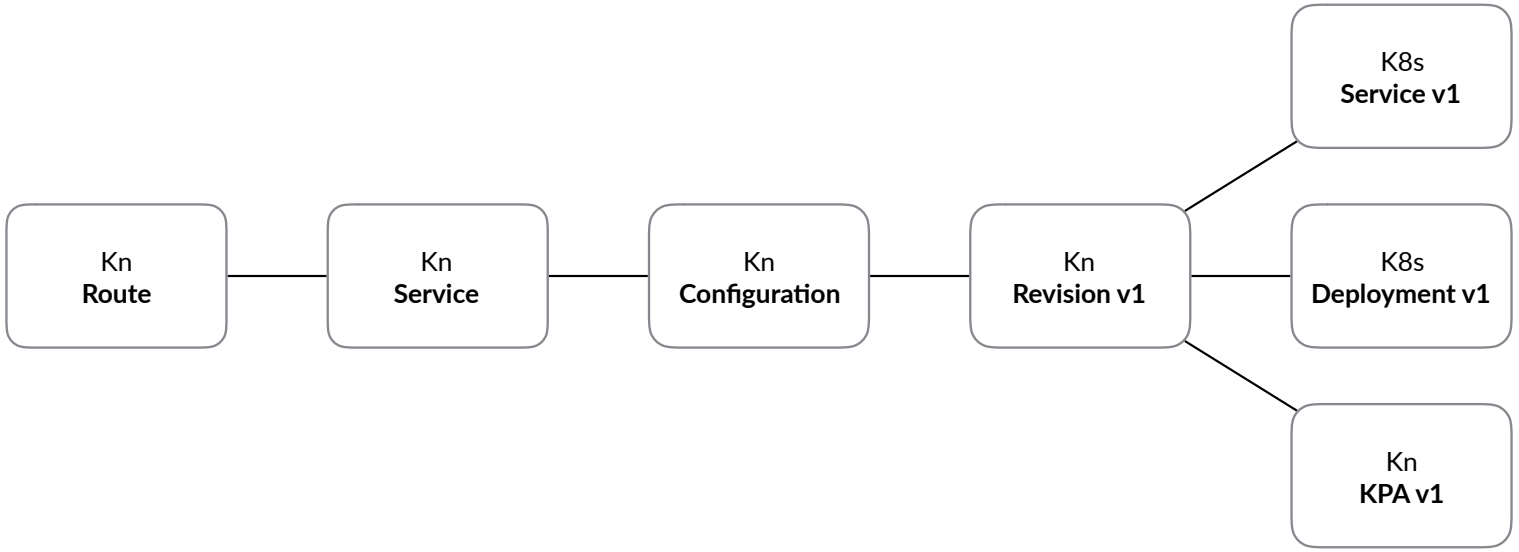

Workload i.e. Pod Spec
Traffic i.e. Istio Virtual Service Spec
Service
Name
Istio Virtual Service
created in response
created in response
Configuration
Name
Route
Name
Deployment, KPA, Service
created in response
directs traffic to
Revision
Name-Postfix1
Revision
Name-Postfix2
Revision
Name-Postfixn

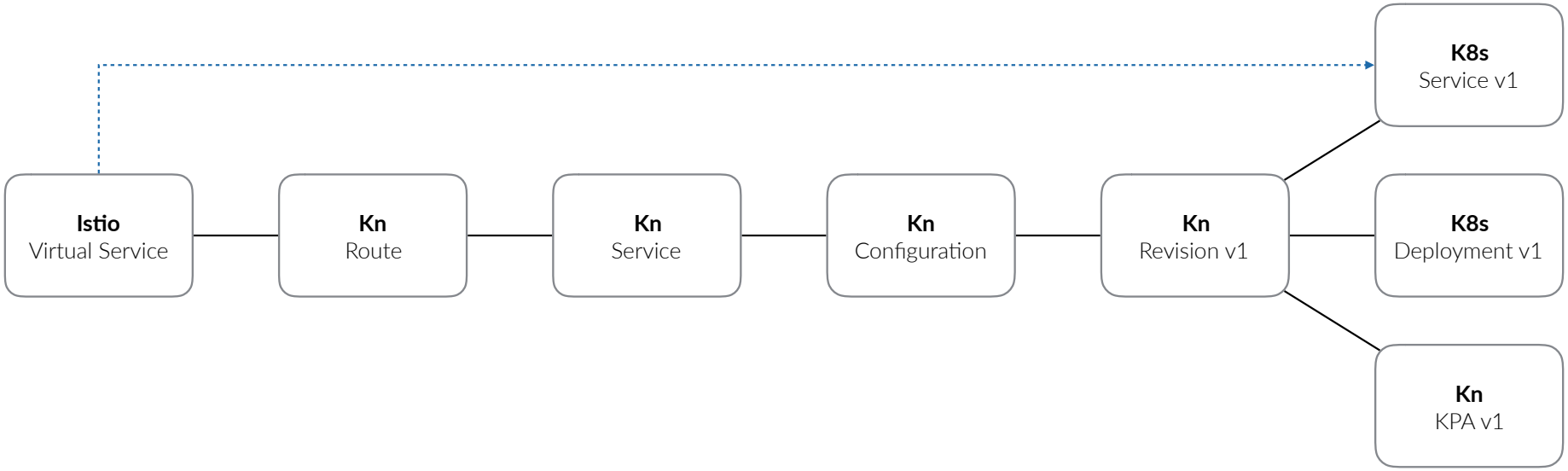
Kn
Service

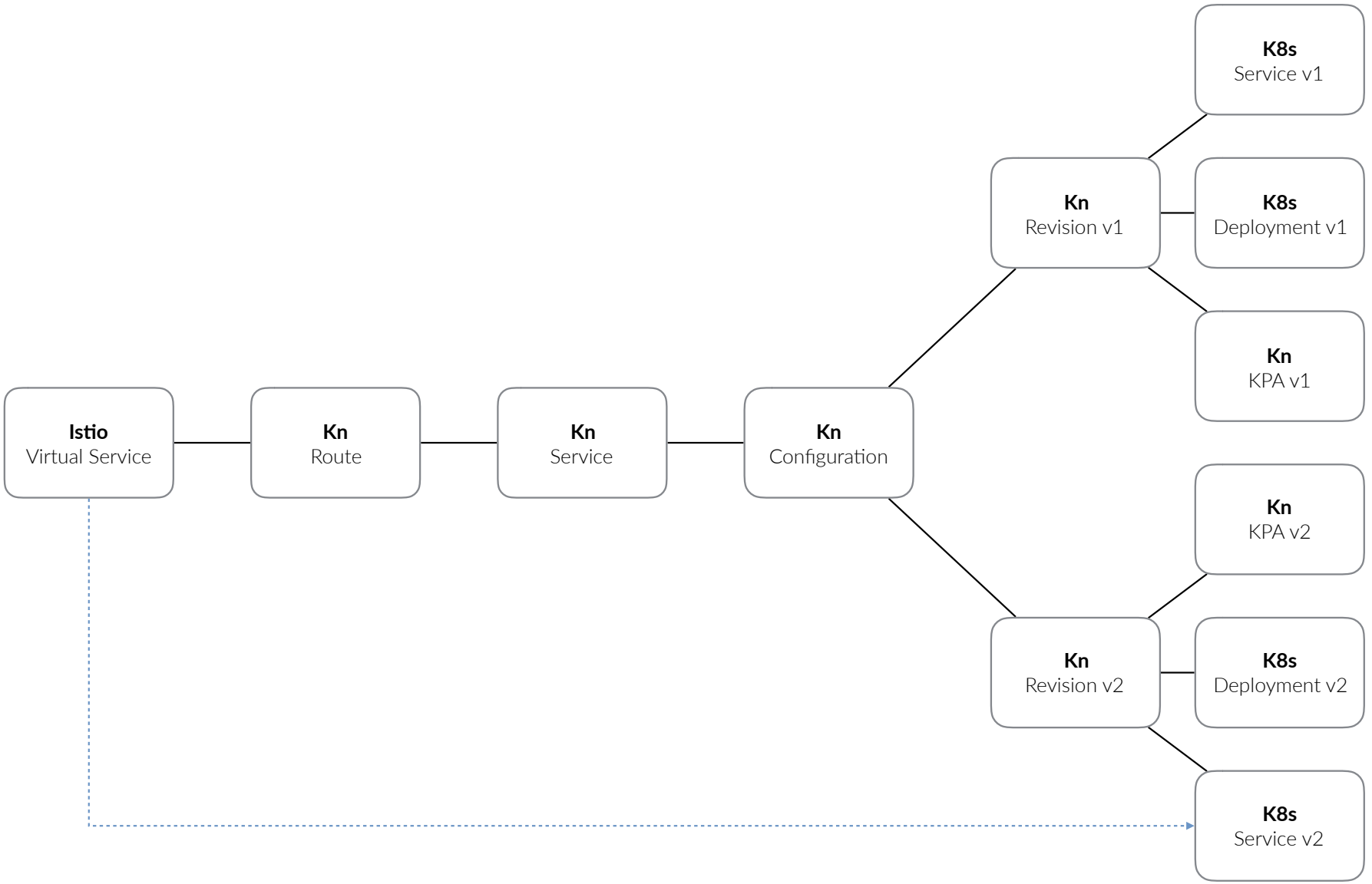

Kn
Service
Kn
Configuration

Kn
Service
Kn
Configuration
Kn
Revision v1


Kn
Service
Kn
Configuration
Kn
Revision v1
K8s
Service v1
K8s
Deployment v1
Kn
KPA v1

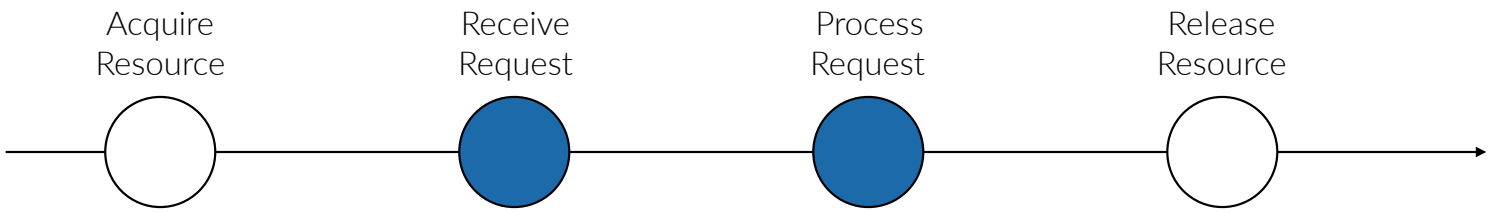

Kn
Route
Kn
Service
Kn
Configuration
Kn
Revision v1
K8s
Service v1
K8s
Deployment v1
Kn
KPA v1

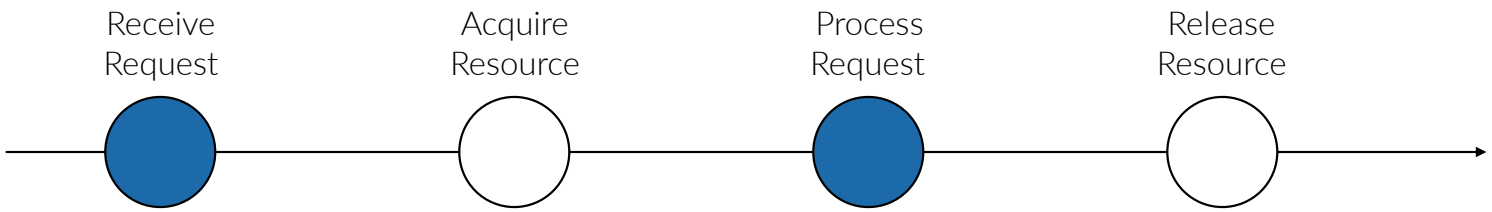

Istio
Virtual Service
Kn
Route
Kn
Service
Kn
Configuration
Kn
Revision v1
K8s
Service v1
K8s
Deployment v1
Kn
KPA v1

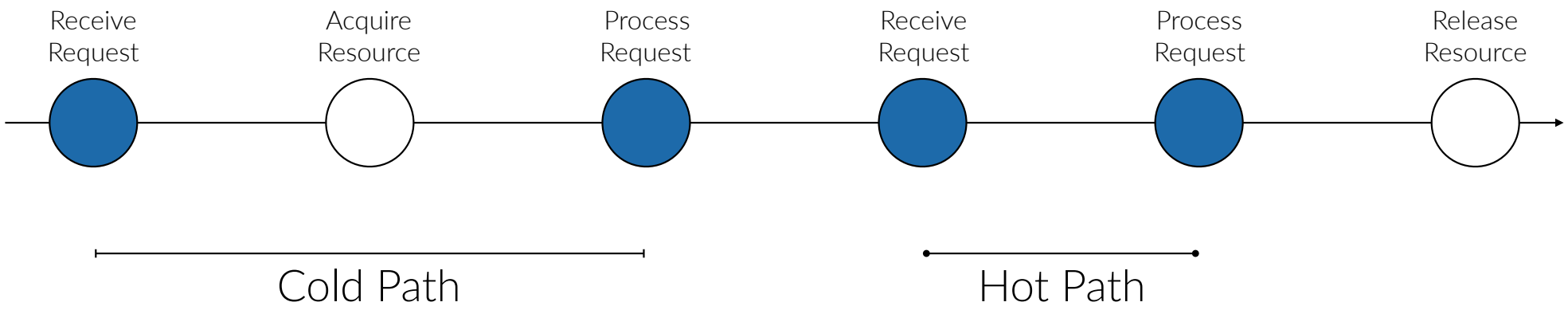

Istio
Virtual Service
Kn
Route
Kn
Service
Kn
Configuration
Kn
Revision v1
K8s
Service v1
K8s
Deployment v1
Kn
KPA v1
Kn
Revision v2
Kn
KPA v2
K8s
Deployment v2
K8s
Service v2

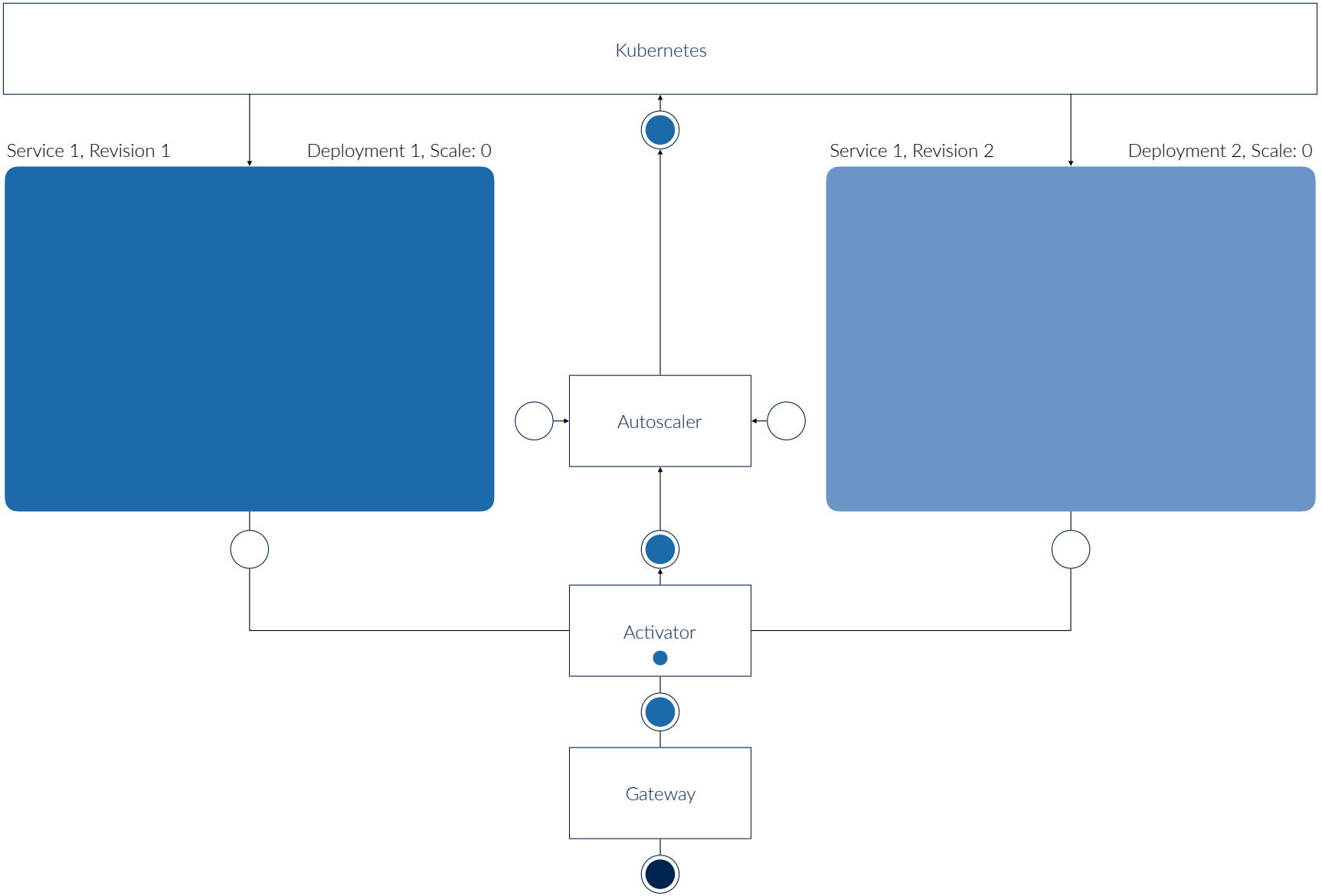
Resource Management

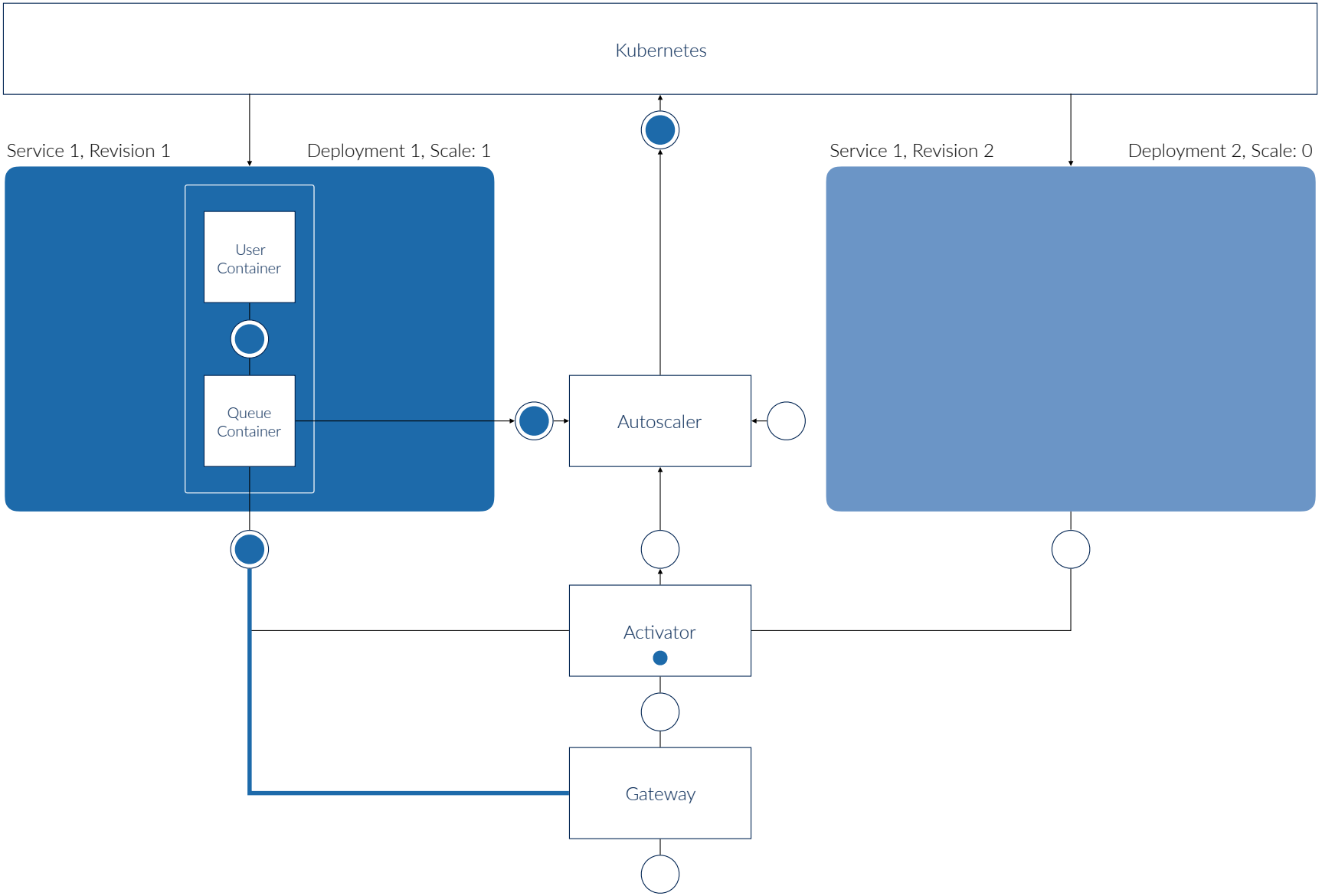
Knative Serving

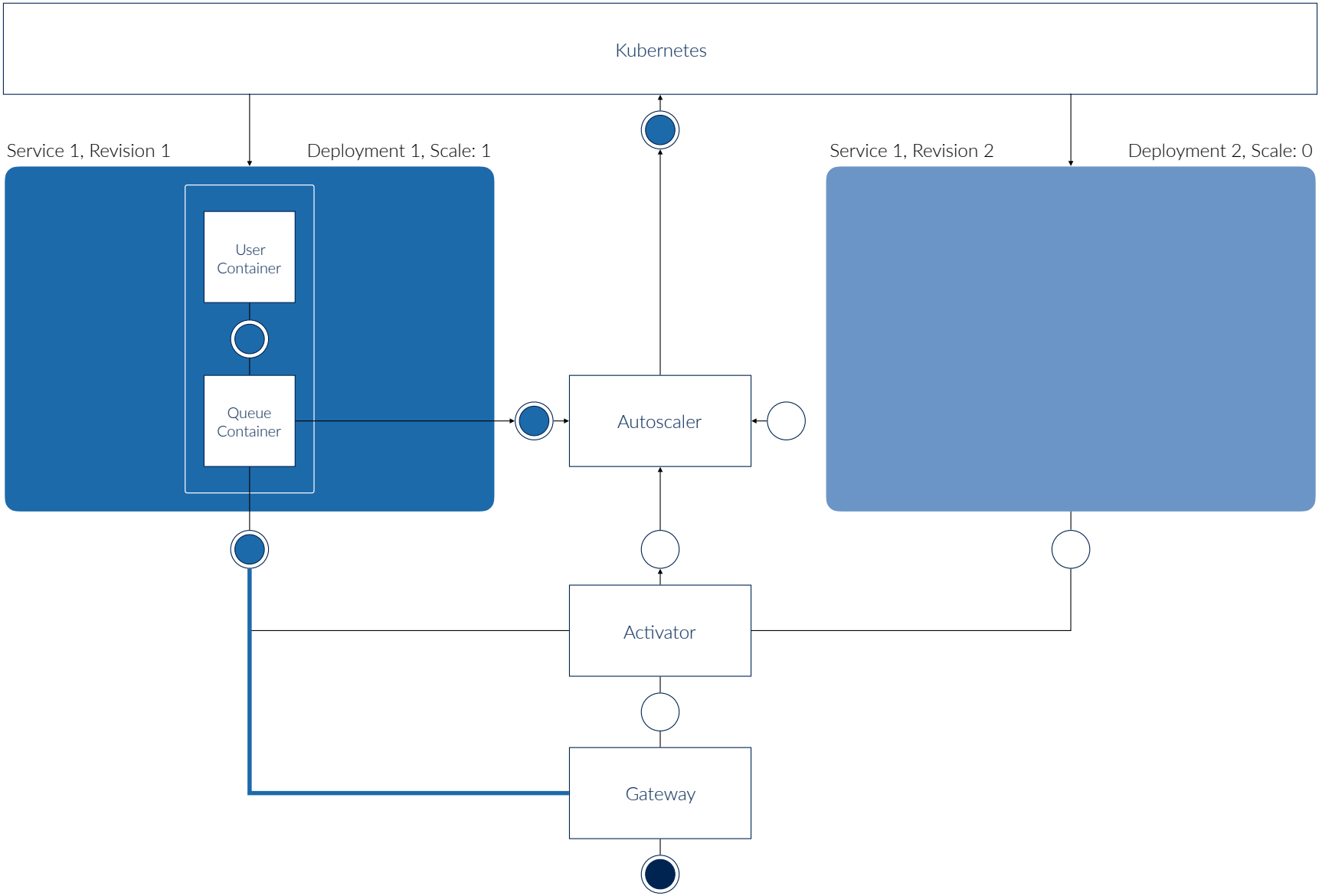

Acquire
Resource
Receive
Request
Process
Request
Release
Resource

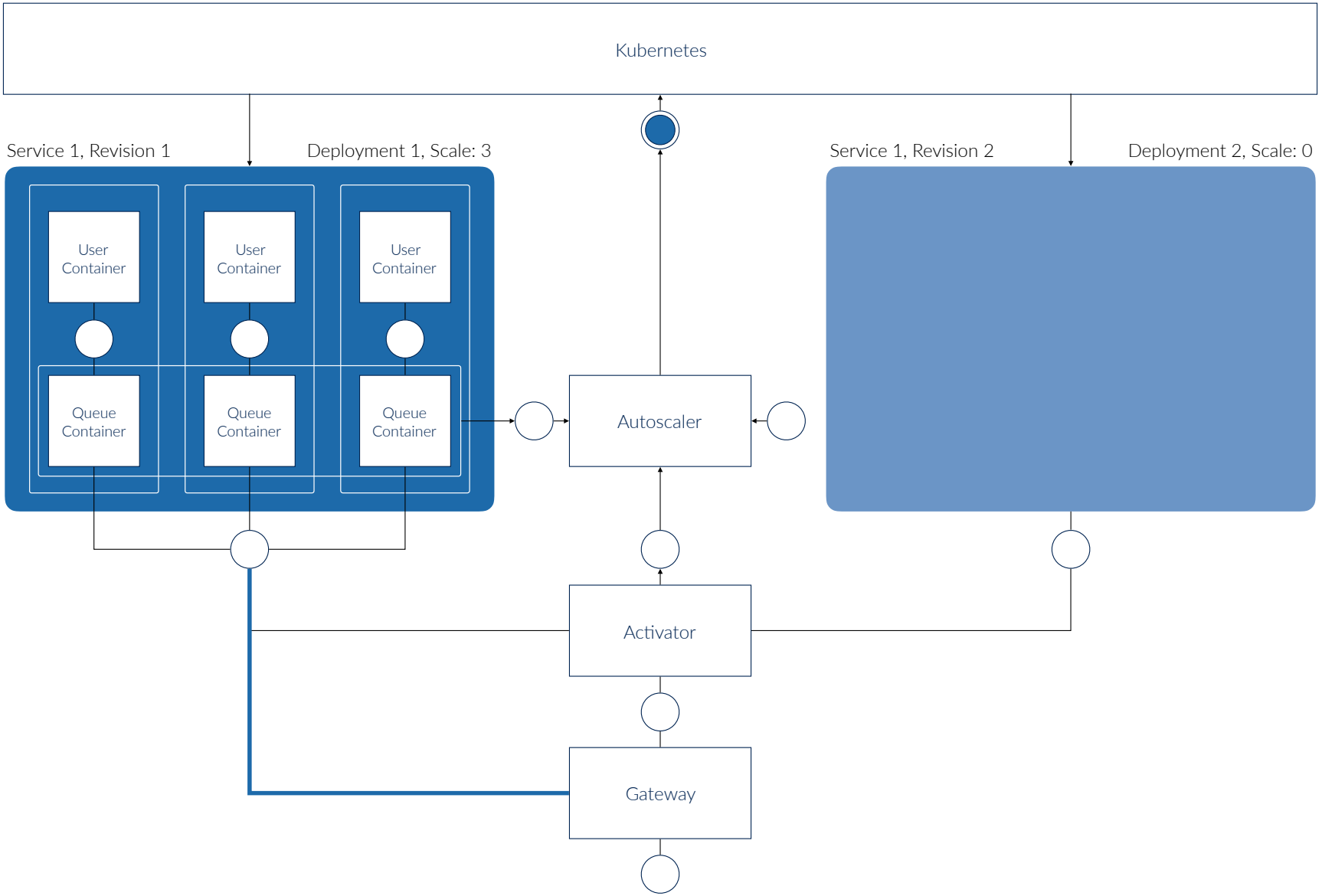

Receive
Request
Acquire
Resource
Process
Request
Release
Resource

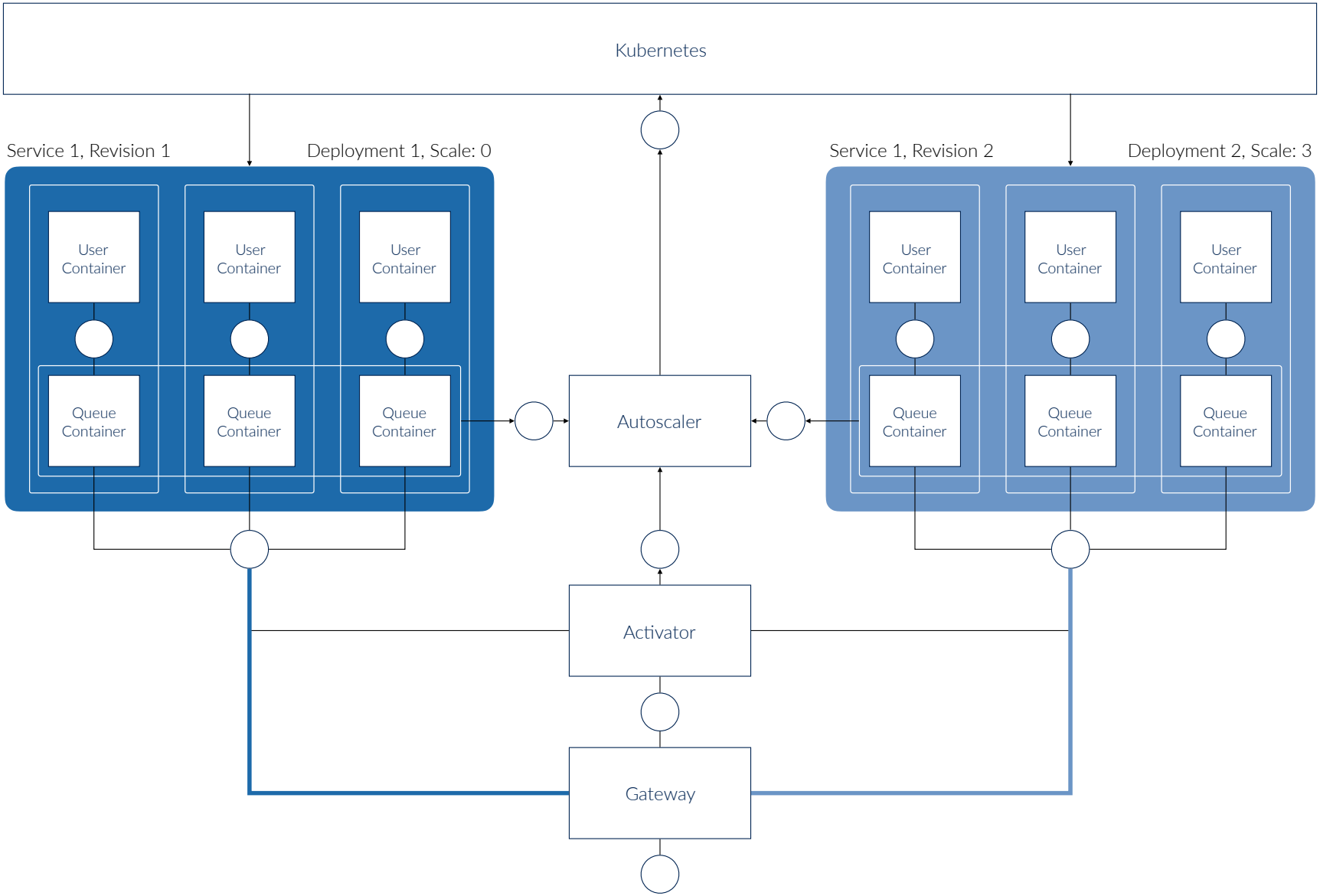

Receive Request
Acquire Resource
Process Request
Receive Request
Process Request
Release Resource
Cold Path
Hot Path


Kubernetes
Service 1, Revision 1
Deployment 1, Scale: 0
Service 1, Revision 2
Deployment 2, Scale: 0
Autoscaler
Activator
Gateway


Kubernetes
Service 1, Revision 1
Deployment 1, Scale: 1
User Container
Queue Container
Autoscaler
Service 1, Revision 2
Deployment 2, Scale: 0
Activator
Gateway


Kubernetes
Service 1, Revision 1
Deployment 1, Scale: 1
User Container
Queue Container
Autoscaler
Service 1, Revision 2
Deployment 2, Scale: 0
Activator
Gateway


Kubernetes
Service 1, Revision 1
Deployment 1, Scale: 3
Service 1, Revision 2
Deployment 2, Scale: 0
User Container
User Container
User Container
Queue Container
Queue Container
Queue Container
Autoscaler
Activator
Gateway


Kubernetes
Service 1, Revision 1
Deployment 1, Scale: 0
Service 1, Revision 2
Deployment 2, Scale: 3
User Container
User Container
User Container
User Container
User Container
User Container
Queue Container
Queue Container
Queue Container
Queue Container
Queue Container
Queue Container
Autoscaler
Activator
Gateway

Conclusion

结论

Thank You

谢谢