

Managing Apache Flink on Kubernetes - FlinkK8sOperator

Anand Swaminathan

Nov 19, 2019

 @anand12100

 @anand12100



Agenda

1

Streaming @Lyft

2

Background

Summary of Lyft's legacy Flink Deployment

3

Solution

Flink Kubernetes Operator

4

Flink Operator State Machine

5

Demo

6

Current Status

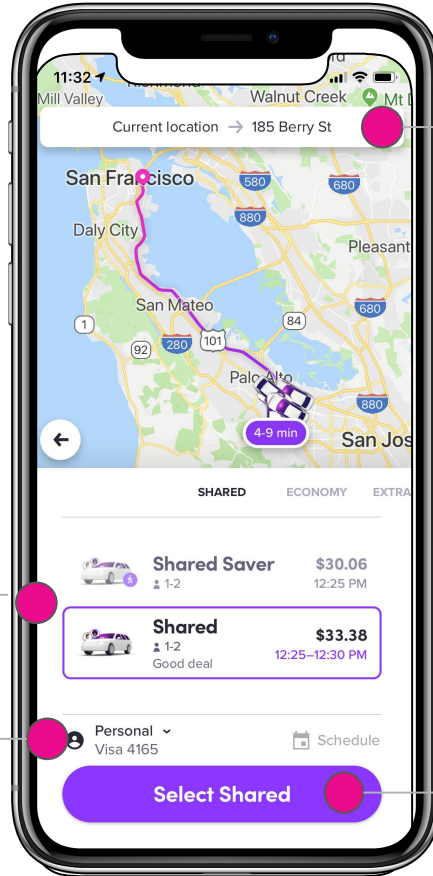
Streaming @Lyft

Pricing

Dynamic Pricing
Supply/Demand curve

Fraud

Behaviour Fingerprinting
Monetary Impact
Imperative to act fast



Core Experience

Top Destinations

User Experience

Notifications
ETA Prediction
Coupons

Streaming @Lyft

Apache Flink

- Great Functional API support
- Streaming SQL support
- Complete implementation of Apache Beam
- Support for late data handling
- Event time processing
- **Downside:** Deployment model



Streaming @Lyft

Abstraction Levels

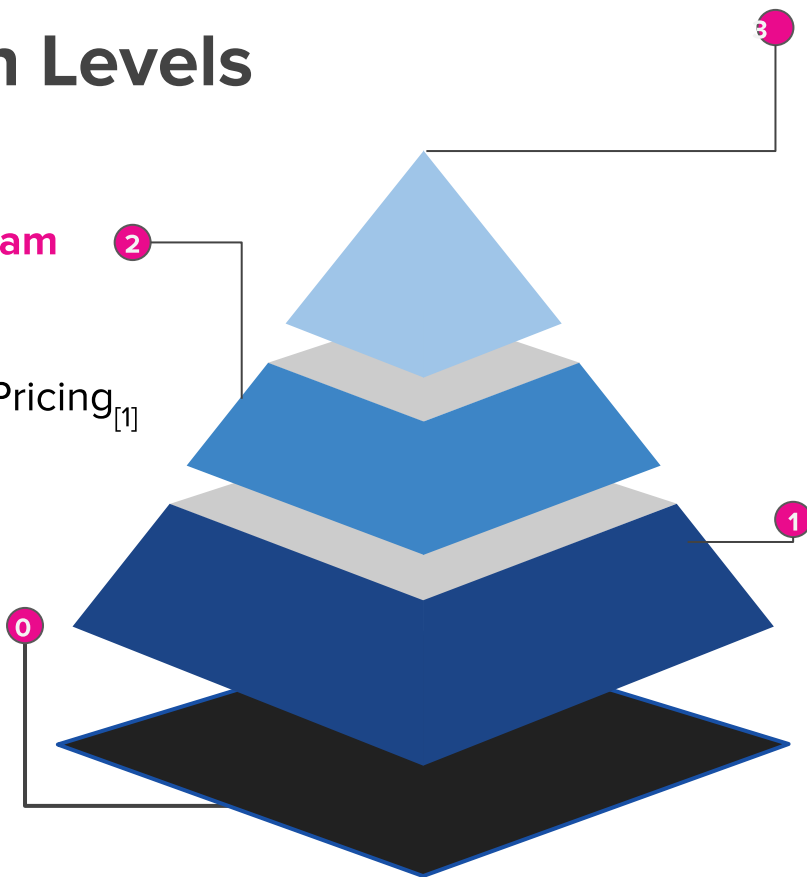
Flink + Python ⇒ Beam

API: Python

Users: ML Engineers

Use cases: Dynamic Pricing_[1]

**Streaming
Platform**



Dryft, (Flyte, Airflow)

API: Streaming SQL,
Workflows

Users: Research Scientists,
Data Engineers

Use cases: Fraud
detection, coupons

Flink jobs

API: Java

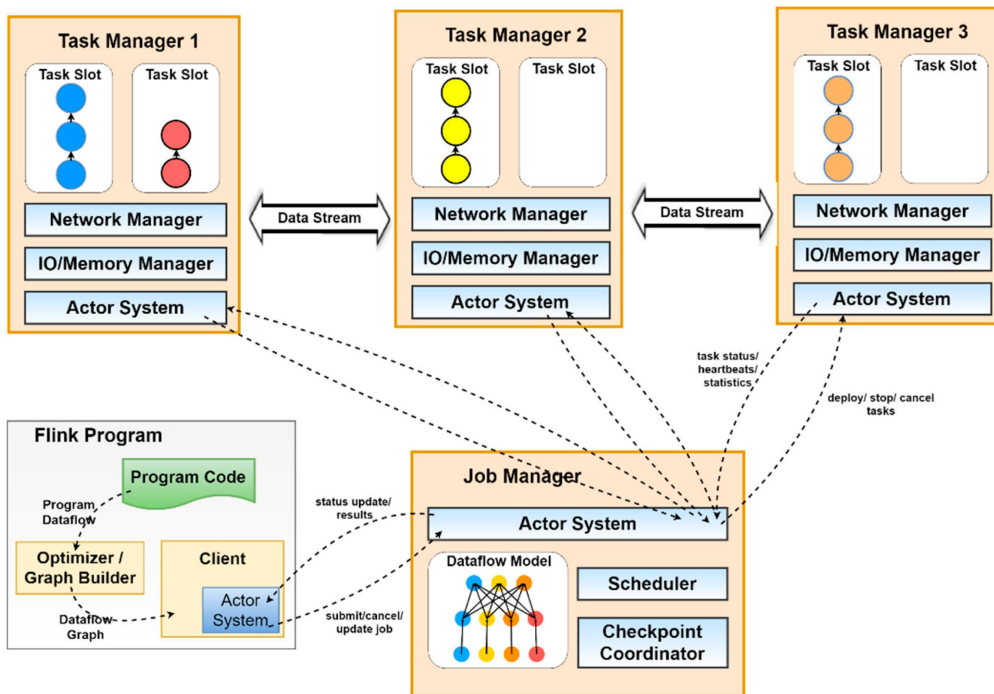
Users: Engineers

Use cases: Data pipelines,
ETA, Mapping

[1] [Streaming your Lyft ride prices](#)

Background

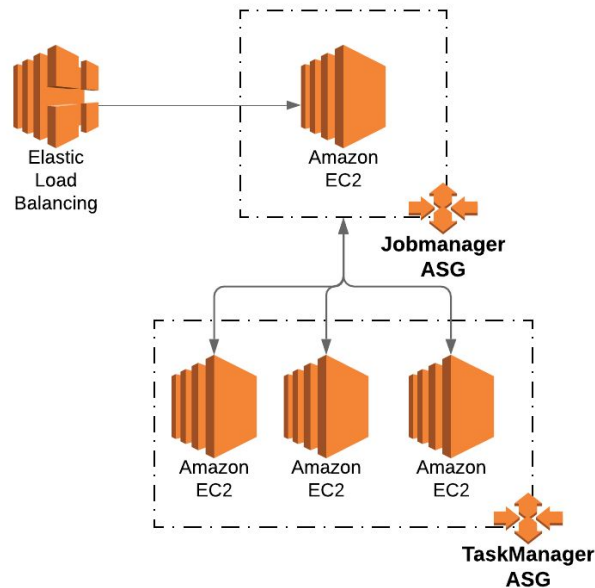
Apache Flink Architecture



Background

Legacy Deployment @Lyft

- ***Saltstack*** deployment
- **AWS Resources:**
 - Jobmanager ASG
 - Taskmanager ASG
 - Elastic load balancer for UI
- **Deploy the binaries to the machines**
 - Run script after the deployment finishes
- **Separate Zookeeper cluster for HA**



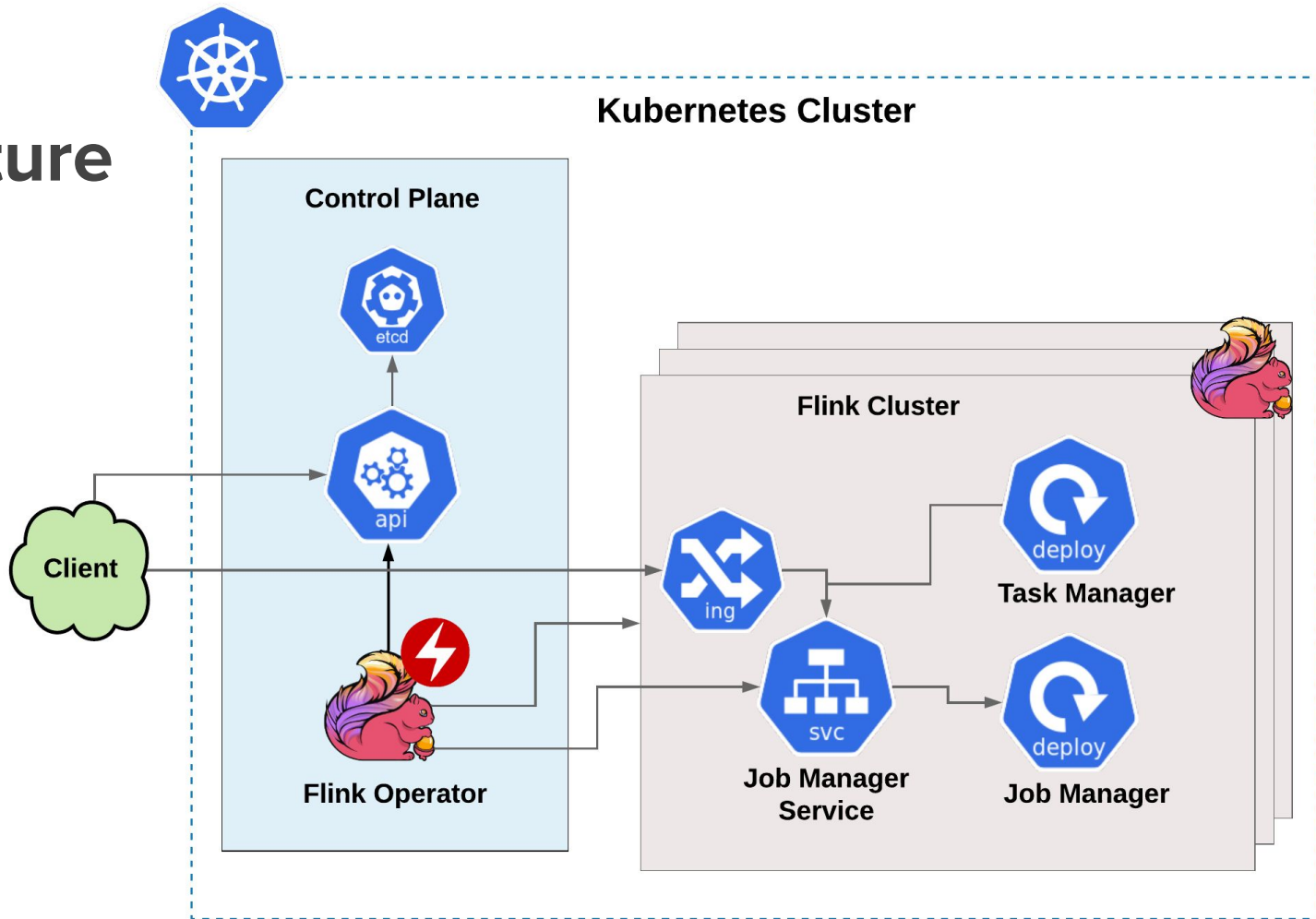
Background

Problems

- **Single flink job requires multiple Autoscaling groups in AWS**
 - Separate ASG for Staging, Production
- **Run multiple flink jobs in a single Flink cluster**
 - Hard to debug logs
 - Interference, Memory leak
- **Manual error prone process of updating jobs**
- **10+ minute replacement time for node failures**
 - Scaling during traffic also bound on Node startup times
- **No automated rollback support**

Introducing Flink-k8s-operator

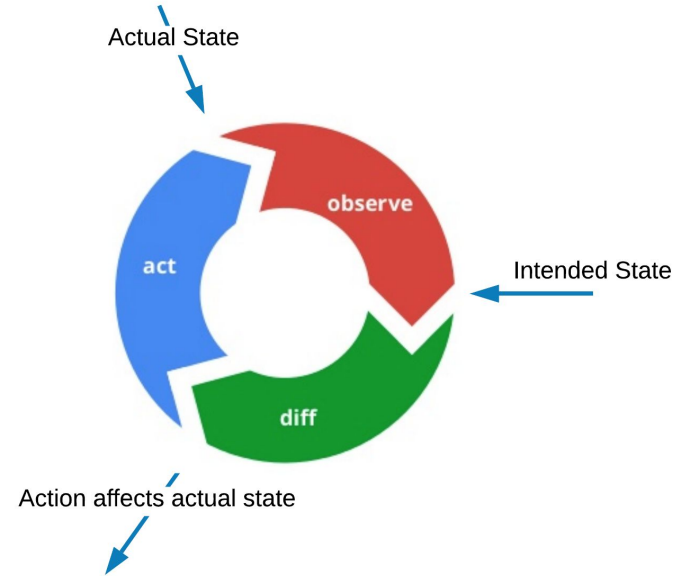
Solution Architecture



Solution

Control Loops

- **Control loops** are fundamental building block industrial control systems
- **Desired State** refers to the intended state as requested
- **Current/Observed State** is the state of the system as **observed** by the controller
- Controller runs control loops
- Drive **Current State -> Desired State**
- Flink operator employs this principle as well



Solution

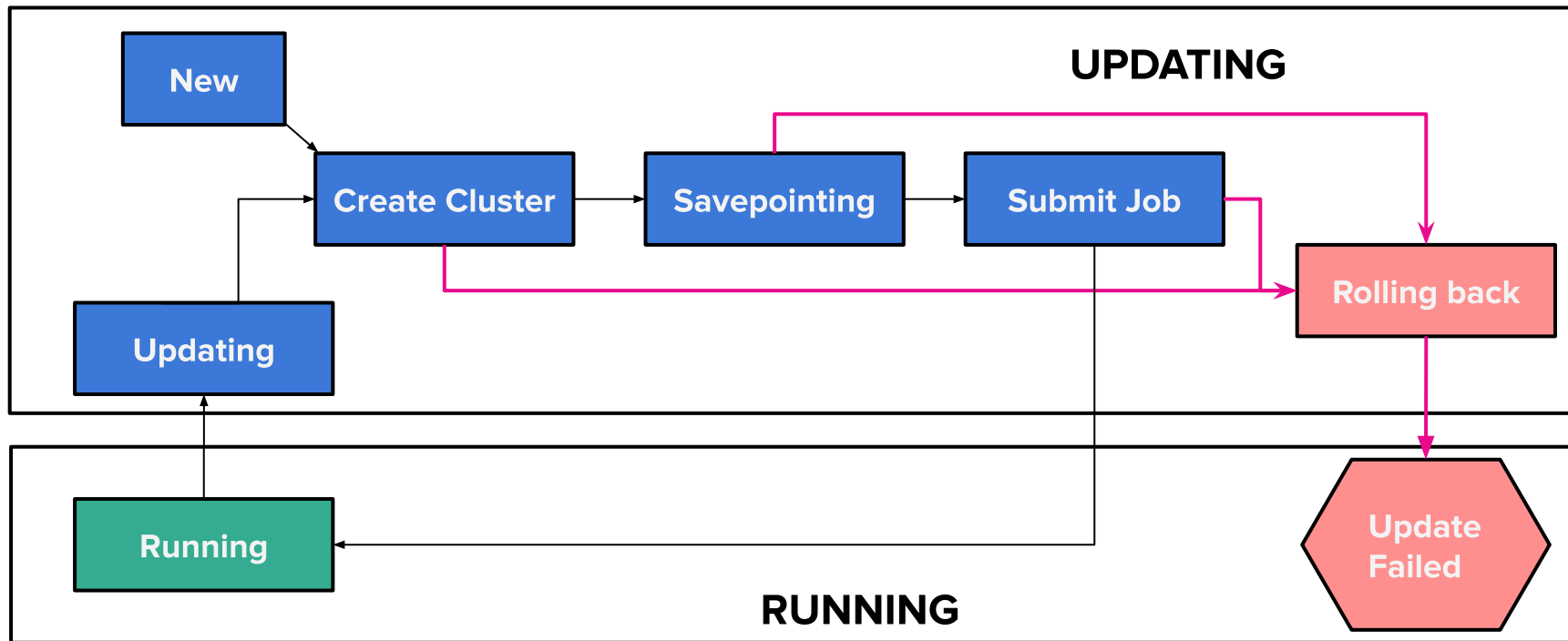
Flink Application Custom Resource

- Each custom resource corresponds to a Flink application
- Each Flink application runs a single Flink job
- Docker image should be runnable
- *Spec*: Desired state indicated by the user
- *Status*: Current state updated by the operator

```
apiVersion: v1beta1
kind: FlinkApplication
metadata:
  name: wordcount-operator-example
spec:
  flinkConfig:
    state.backend.fs.checkpointdir: file:///flink/checkpoints
    state.checkpoints.dir: file:///flink/externalized-checkpoints
  flinkVersion: "1.8"
  image: docker.io/lyft/wordcount-operator-example:3b0347b
  jarName: wordcount-operator-example-1.0.0-SNAPSHOT.jar
  jobManagerConfig:
    resources:
      requests:
        cpu: 200m
        memory: 200Mi
  parallelism: 3
status:
  clusterStatus:
    availableTaskSlots: 4
    healthyTaskManagers: 2
  jobStatus:
    completedCheckpointCount: 0
    health: Green
    lastSeenError: null
  phase: Running
```

Solution

State Machine



Solution

Impact

- **Improved stability and recoverability**
 - Separate Flink cluster for each Flink Job
 - Faster scaling
 - Task manager failures recovers in seconds
- **Reduction in common failures**
 - Less downtime during failed deploys
- **Easier configuration and tooling**
 - Wrappers around the *FlinkApplication* custom resource
- **More flexible deployment strategies**
 - Single, Dual, Green-Blue

Demo

Current Status

- **Project status: Beta**
- **Open Source:** <https://github.com/lyft/flinkk8soperator>
 - 5 external contributors
 - 20 external contributions
- **Already in Production @Lyft**
 - Running 60 Flink applications
- **External users**
 - The Trade Desk, Lightbend, McAfee
- **Join our slack channel:** http://go.lyft.com/flinkoperator_slack

Join us for some local beer, wine, and tacos!

Lyft Happy Hour

Date: Tuesday, Nov 19

Time: 7pm-10pm

Where: Thorn Barrio Logan (1745 National Avenue, San Diego, CA 92113)

RSVP: <https://lyft-kubecon.splashthat.com/> (you can also register at the door)

