基于Flink on Kubernetes 的大数据平台

张凯　　·　阿里云 / 高级技术专家
任春德　·　阿里云 / 高级技术专家

Apache Flink

CONTENT
目录 >>

Apache Flink

01

K8S简介

Evolution of Container and Cloud Native technology

An open cloud-native world is coming

领域创新　（进行时）

Kubeflow

标准化

成熟

containerd

gVisor

OPEN CONTAINER INITIATIVE

Windows Server 2016

2018

katacontainers

发展

2013

2015

2017

萌芽

2011

LXC

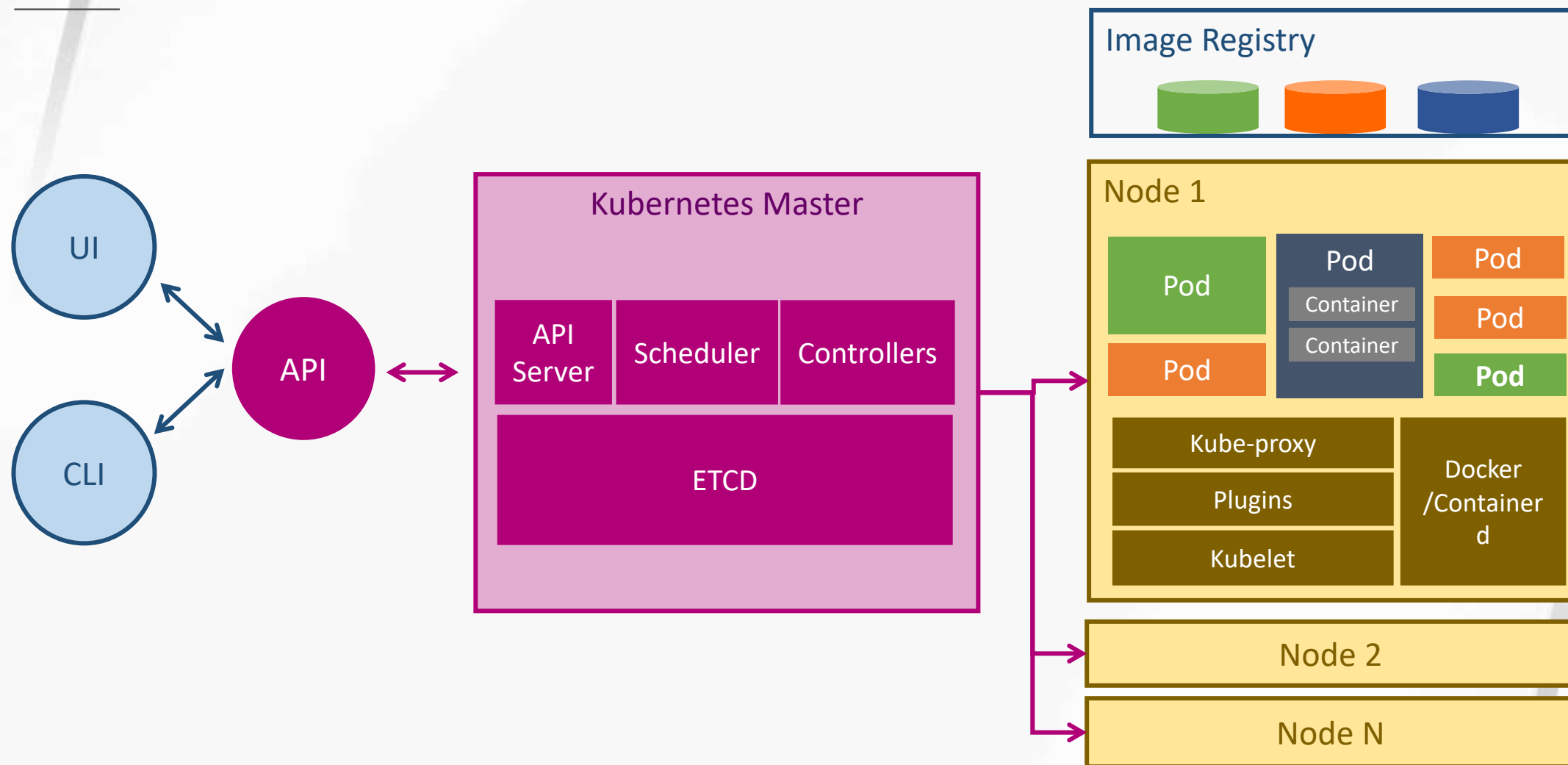CLOUD FOUNDRY

docker SWARM

2000　　　2006

Linux VServer
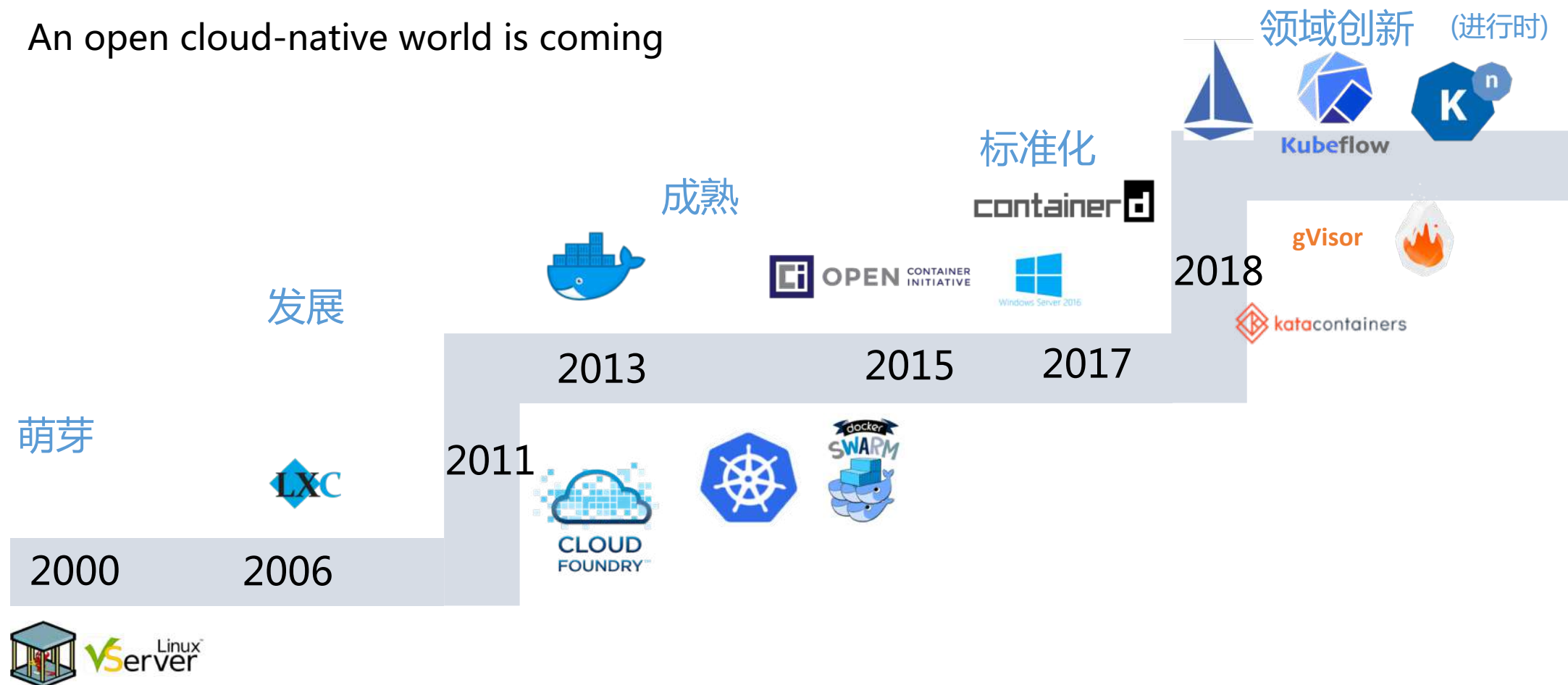
What is Kubernetes?

- Kubernetes is an open source container orchestration engine for automating deployment, scaling, and management of containerized applications.

  – Resource management

  – Container orchestration

  – Application management

  – Operation automation
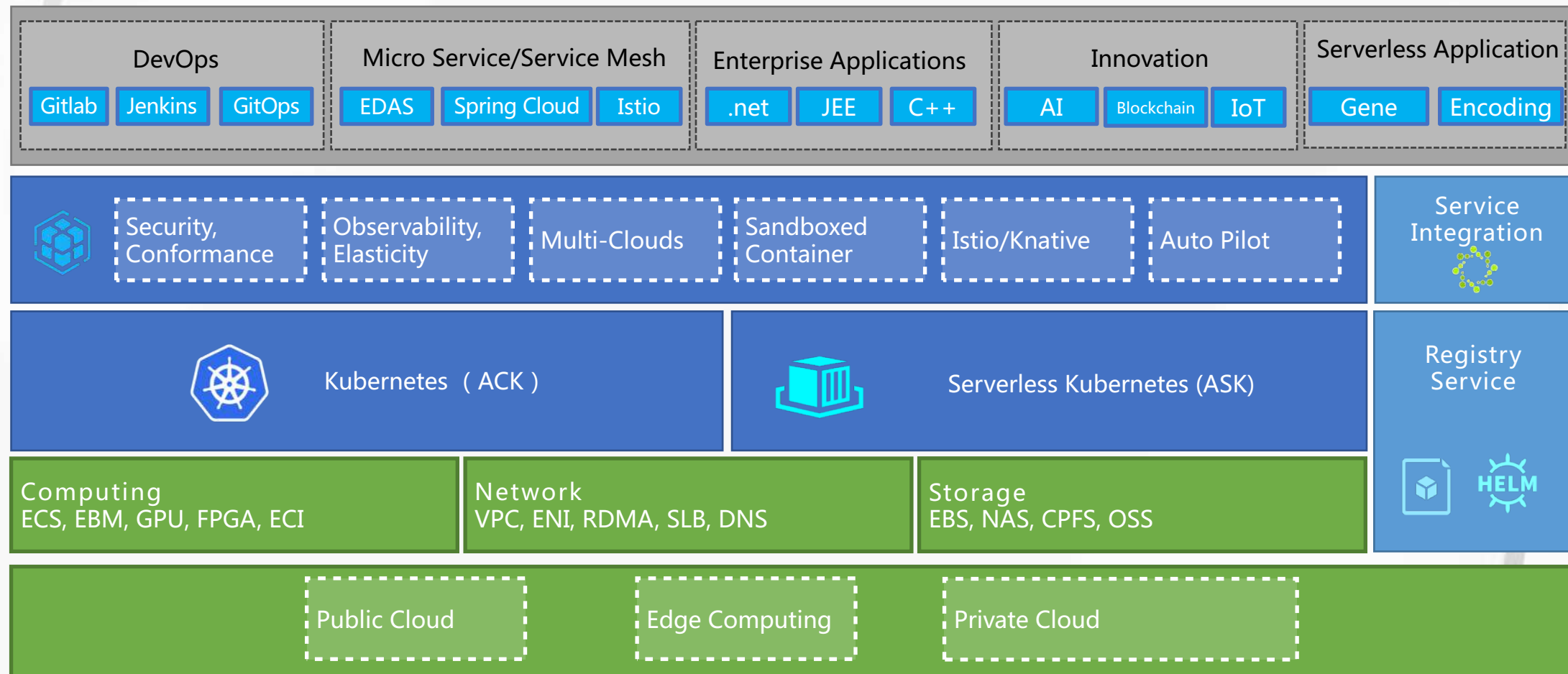
  – Cloud native application platform
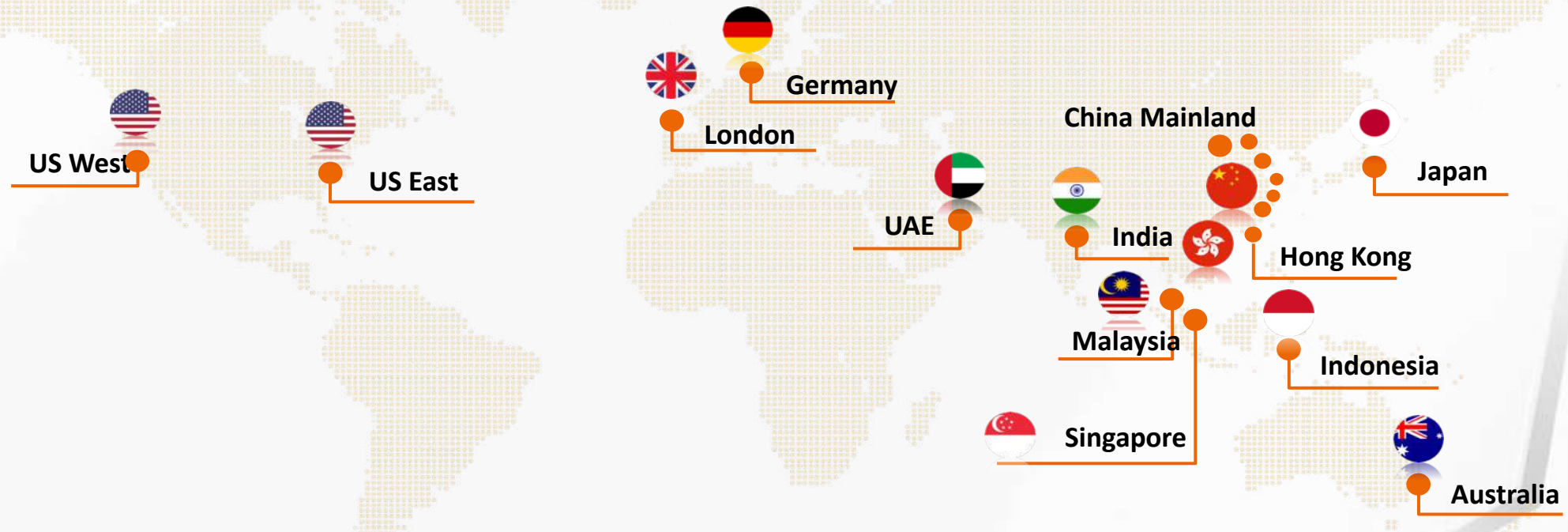
Kubernetes Architecture

ACK—Alibaba cloud Container service for Kubernetes

Apache Flink

DevOps	Micro Service/Service Mesh	Enterprise Applications	Innovation	Serverless Application
Gitlab Jenkins GitOps	EDAS Spring Cloud Istio	.net JEE C++	AI Blockchain IoT	Gene Encoding

Security, Conformance	Observability, Elasticity	Multi-Clouds	Sandboxed Container	Istio/Knative	Auto Pilot	Service Integration

Kubernetes（ACK）	Serverless Kubernetes (ASK)	Registry Service

Computing ECS, EBM, GPU, FPGA, ECI	Network VPC, ENI, RDMA, SLB, DNS	Storage EBS, NAS, CPFS, OSS	HELM

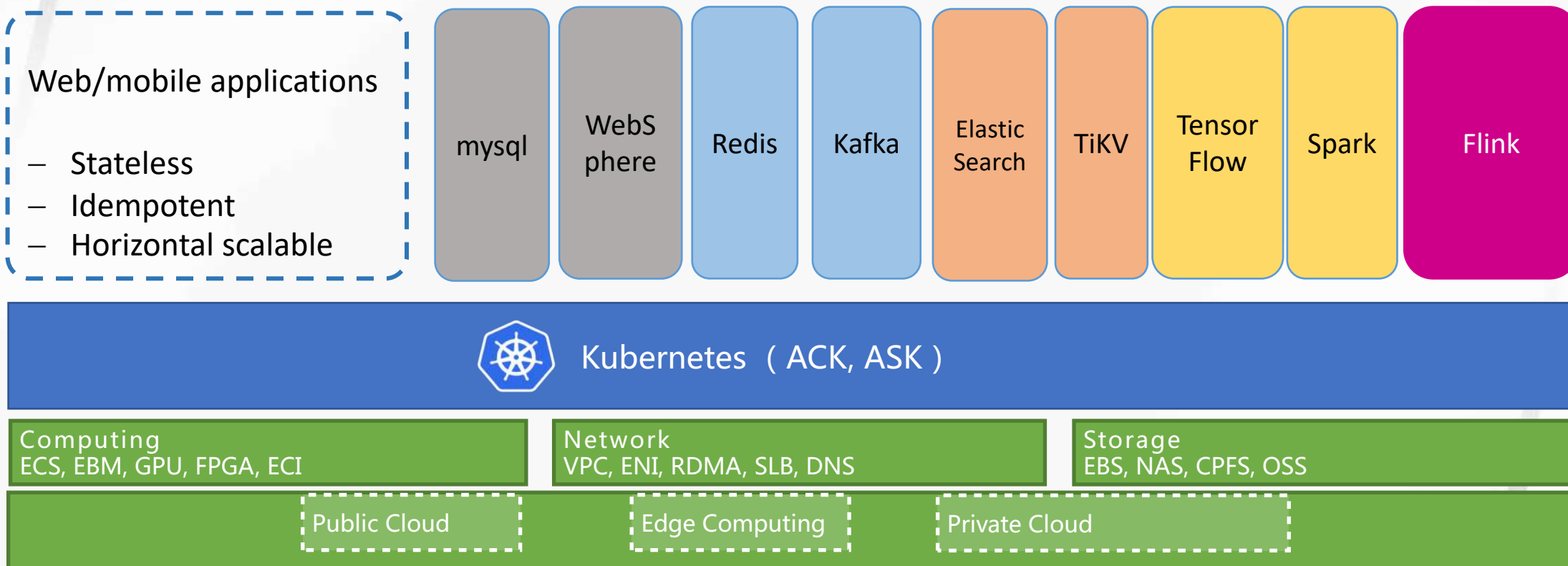| Public Cloud	Edge Computing	Private Cloud

ACK is available in 18 Regions

- IDC: market share of Alibaba Cloud is 43%, ranks top 1 in China
- Gartner: Alibaba Cloud led the Asia Pacific market for IaaS and IUS with 19.6% market share

More and more workloads run on Kubernetes

Web/mobile applications

- Stateless
- Idempotent
- Horizontal scalable

| mysql | WebSphere | Redis | Kafka | Elastic Search | TiKV | Tensor Flow | Spark | Flink |

Kubernetes （ACK, ASK）

Computing
ECS, EBM, GPU, FPGA, ECI

Network
VPC, ENI, RDMA, SLB, DNS

Storage
EBS, NAS, CPFS, OSS

Public Cloud   Edge Computing   Private Cloud

02

Flink on K8s

Flink on K8s – Standalone

Apache Flink

jobmanager-deployment.yaml

```yaml
apiVersion: extensions/v1beta1
kind: Deployment
metadata:
  name: flink-jobmanager
spec:
  replicas: 1
  template:
    metadata:
      labels:
        app: flink
        component: jobmanager
    spec:
      containers:
      - name: jobmanager
        image: flink:latest
        args:
        - jobmanager
```

taskmanager-deployment.yaml

```yaml
apiVersion: extensions/v1beta1
kind: Deployment
metadata:
  name: flink-taskmanager
spec:
  replicas: 2
  template:
    metadata:
      labels:
        app: flink
        component: taskmanager
    spec:
      containers:
      - name: taskmanager
        image: flink:latest
        args:
        - taskmanager
```
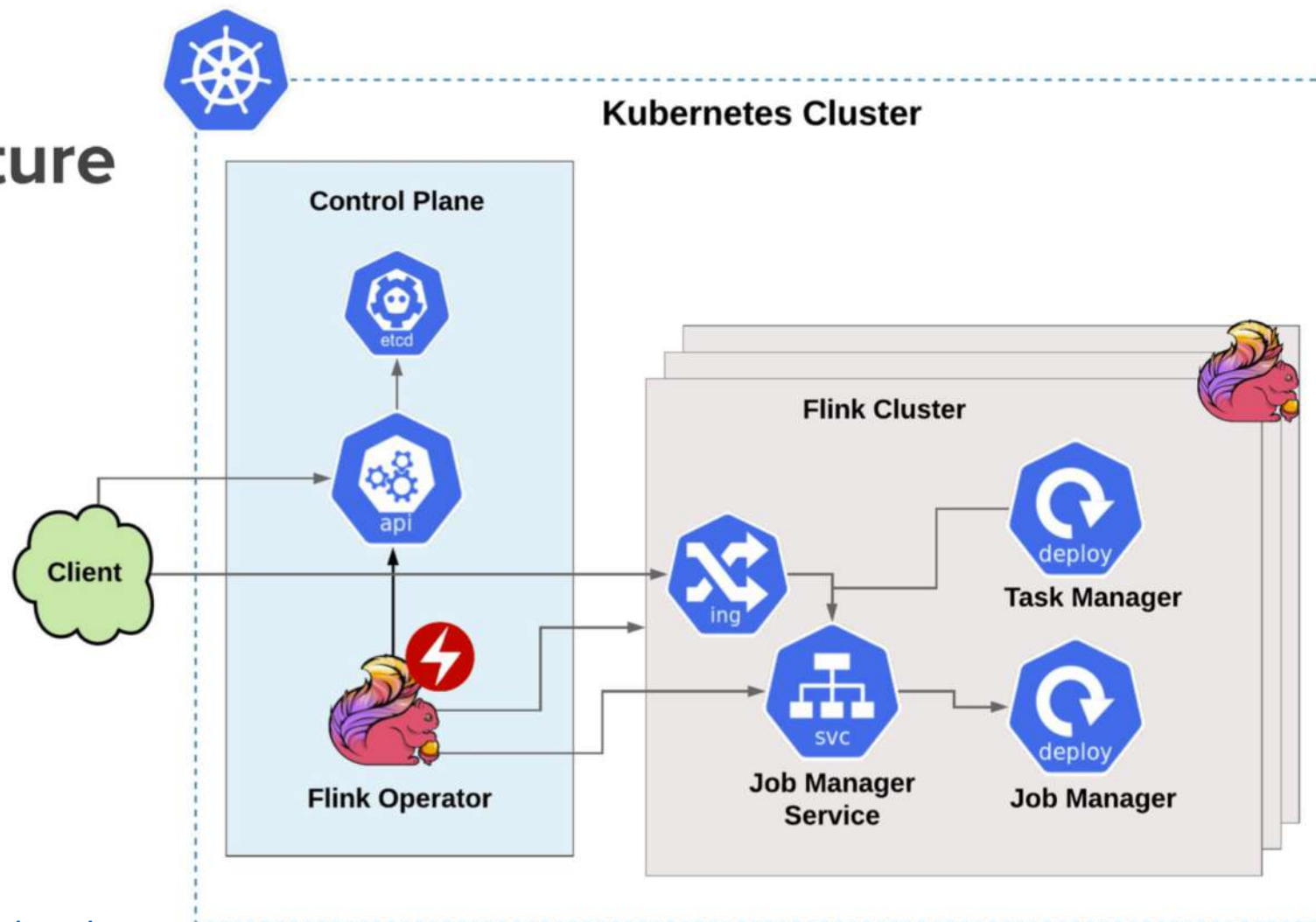
jobmanager-service.yaml

```yaml
apiVersion: v1
kind: Service
metadata:
  name: flink-jobmanager
spec:
  ports:
  - name: rpc
    port: 6123
  - name: blob
    port: 6124
  - name: query
    port: 6125
  - name: ui
    port: 8081
  selector:
    app: flink
    component: jobmanager
```
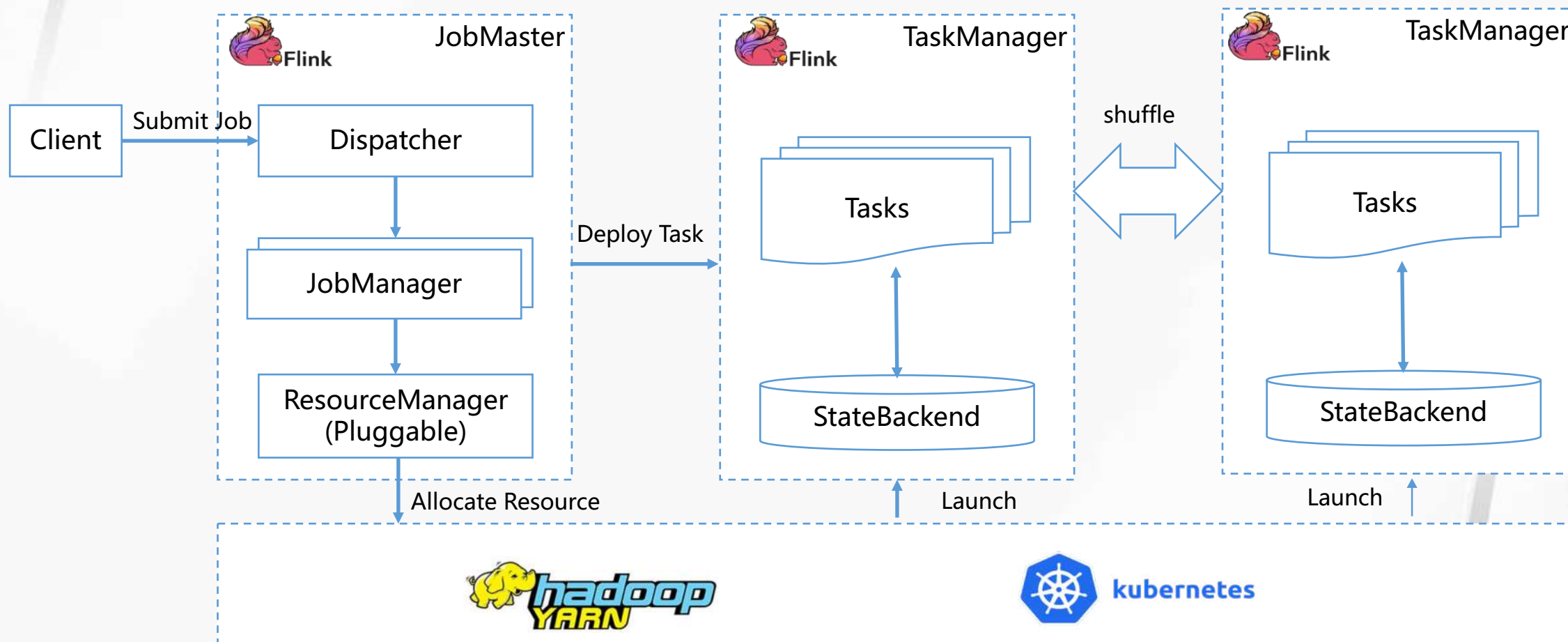
Helm Chart： https://github.com/docker-flink/examples

Flink on K8s – Operator
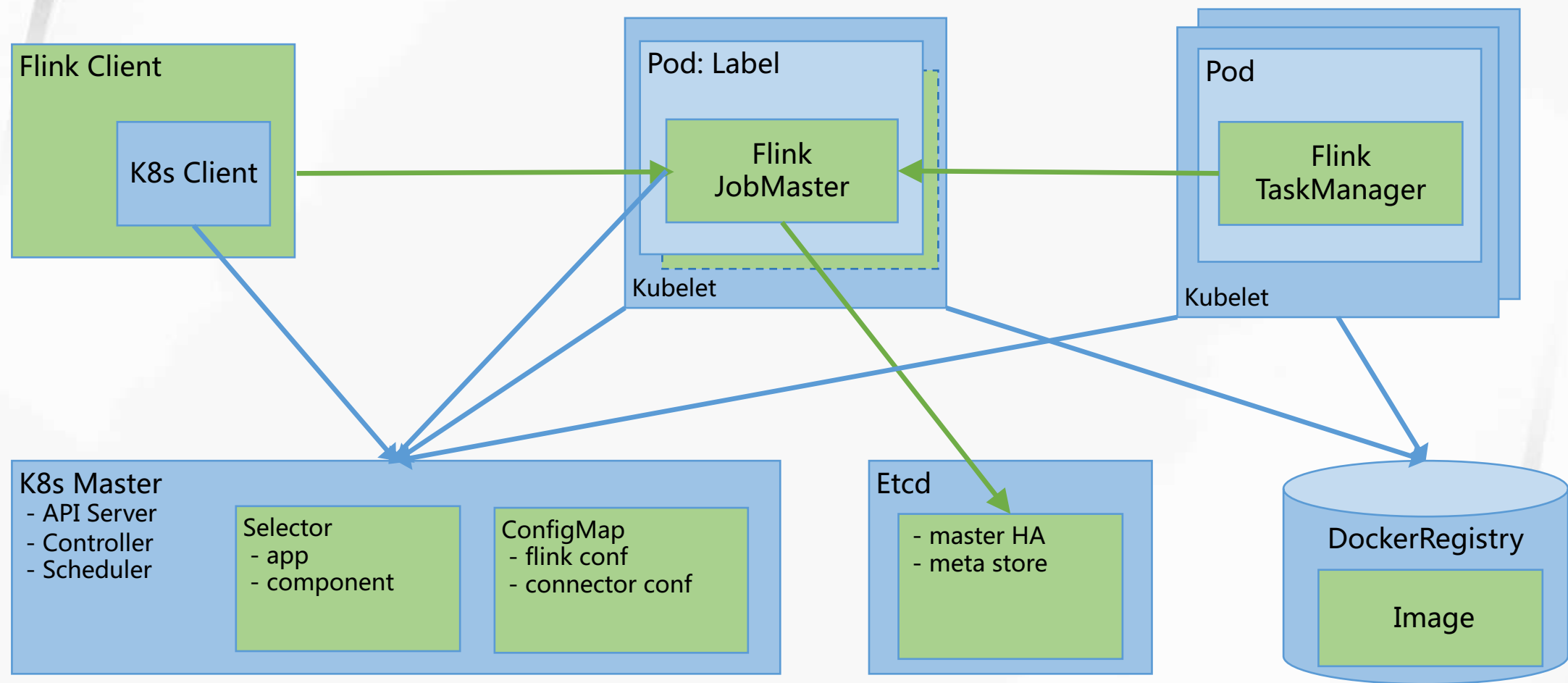

Operator：https://github.com/lyft/flinkk8soperator

Flink on K8s — Natively
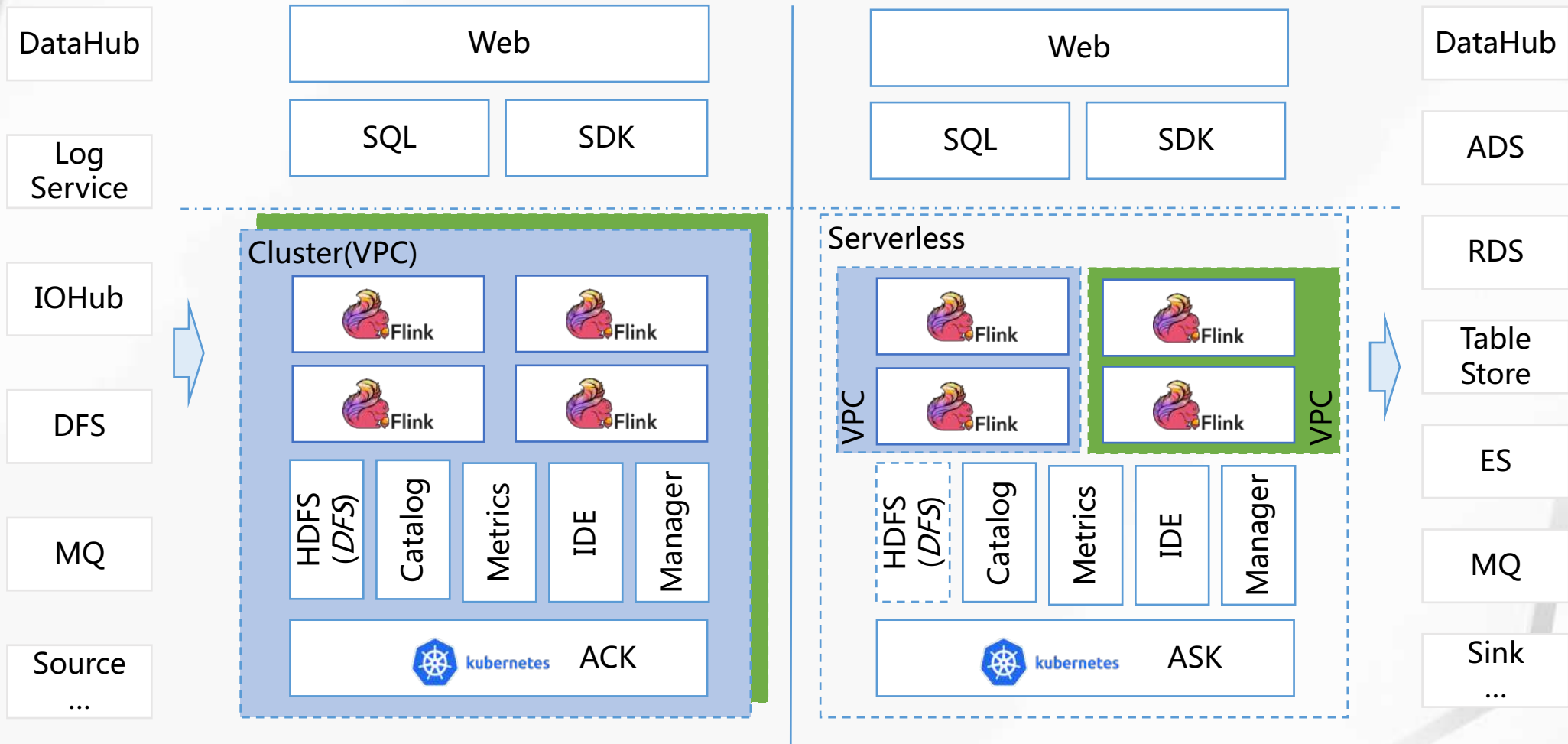
Flink on K8s 一 Natively

03

大数据平台

大数据平台 － on YARN

大数据平台

大数据平台

Apache Flink

大数据平台

Apache Flink

```sql
 9        item_id bigint,
10        catelog_id int,
11        behavior varchar,
12        ts bigint,
13        rowtime as to_timestamp(ts*1000),
14        watermark for rowtime as withOffset(rowtime, 1000)
15    ) with (
16      type='csv',
17      path='hdfs:///example/UserBehavior.csv'
18    );
19
20    create table print_sink (
21        item_id bigint,
22        window_end timestamp,
23        pv bigint,
24        row_num bigint
25    ) with (
26      type='print'
27    );
28
29
30    create view item_pv as
31    select
32      item_id,
33      HOP_END(rowtime, INTERVAL '1' HOUR, INTERVAL '5' MINUTE) window_end,
34      COUNT(*) pv
35    from user_behavior
36    group by item_id, HOP(rowtime, INTERVAL '1' HOUR, INTERVAL '5' MINUTE);
```

Create   Save As...   Save   ← Undo   → Redo   Q Find   Format   FullScreen   SETTING

SQL HotItem ×

VERVERICA

Trial License

2

CPU

⊗ 0   All files are up-to-date (2019-06-21, 10:25:59)   31 : 8   BLINK   SQL
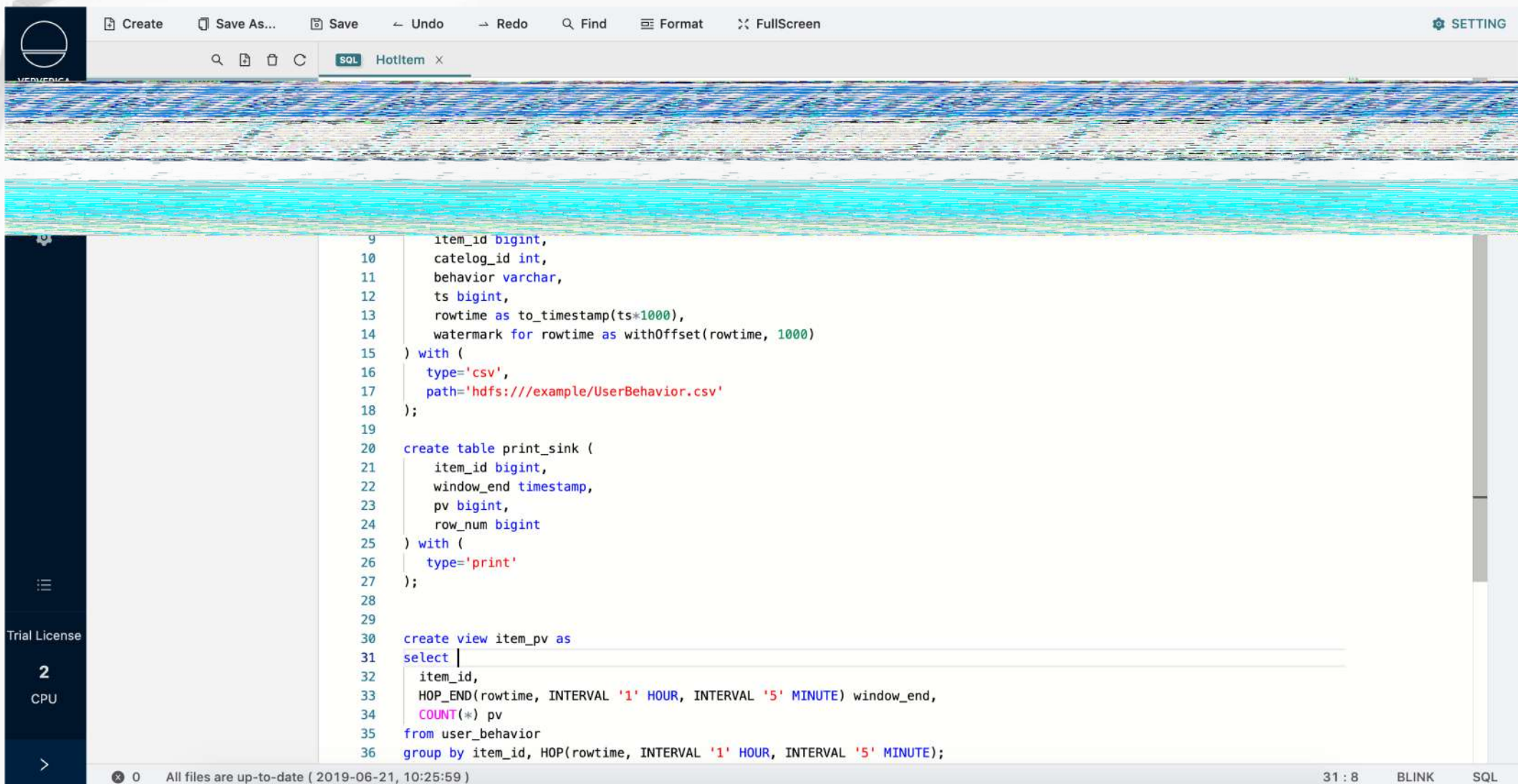
04

展望

展望

- Batch on K8s
  - Scheduling Throughput
  - Isolation
  - Shuffle Service

- Hybrid Workloads on K8s
  - Online Service
  - Streaming Job
  - Batch Job

Apache Flink

谢谢！
Q&A

【2群】Apache Flink C...

扫一扫群二维码，立刻加入该群。

How Kubernetes works?