

分布式数据库在云环境下的实践



李淳竹

PingCAP·Cloud 研发工程师



目录

CONTENTS

PART 01 TiDB 架构简介

PART 02 TiDB Operator 架构

PART 03 TiDB Operator 原理

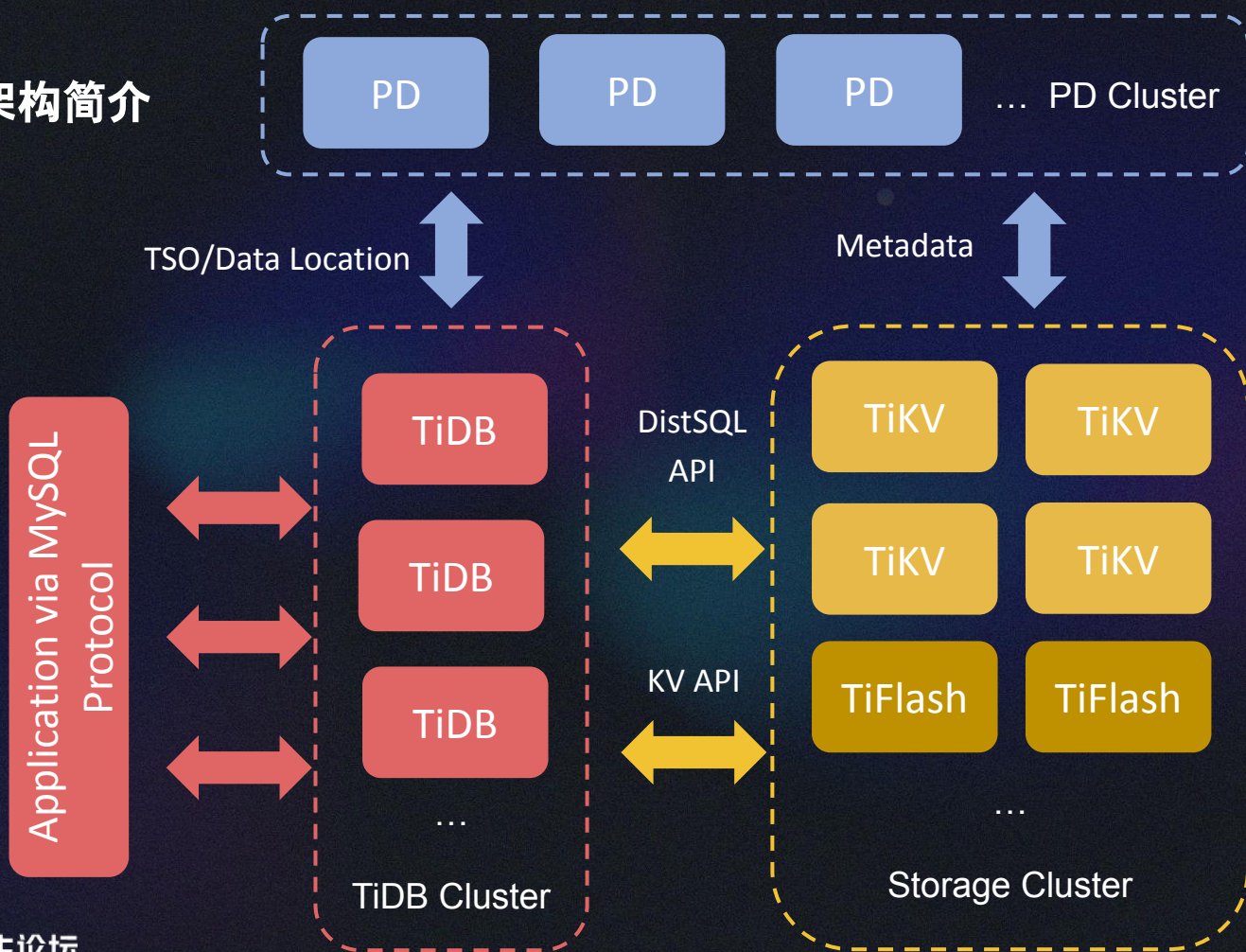
PART 04 TiDB Operator 后续演进

TiDB 架构简介

网易数字 + 云原生论坛



TiDB 架构简介





TiDB 架构简介

我们想要

TiDB 跟我的微服务一样：

- 声明式管理
- 自动运维
- 弹性资源配置

以及最重要的

- 稳定性
- 稳定性
- 还是稳定性！



TiDB 架构简介

TiDB 能适应动态化的 Kubernetes 环境

- 多副本容错
- 水平伸缩
- 良好的可观测性



TiDB



Kubernetes



TiDB 架构简介

Kubernetes 也能高效服务 TiDB

- 声明式 API 简化集群管理
- 弹性资源池简化扩缩容与故障转移



TiDB



Kubernetes



TiDB 架构简介

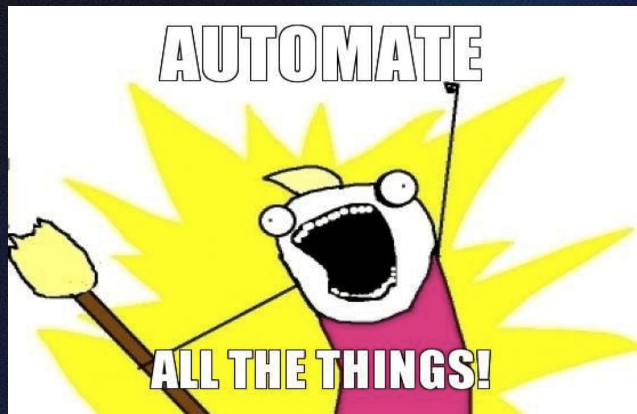
TiDB 和 Kubernetes 都遵循同样的云原生最佳实践原则

因此，假如你熟悉 TiDB 和 Kubernetes 的知识，很容易便能编写一组 Kubernetes 的资源定义文件来部署一个 TiDB 集群。



TiDB 架构简介

还不够！



TiDB Operator 架构

网易数字 + 云原生论坛

为什么需要 TiDB Operator



- Kubernetes 有 pod, deployment, statefulset, etc.
- 但 Kubernetes 并不知道如何管理一个 TiDB 集群的生命周期

我们希望 TiDB Operator 做什么

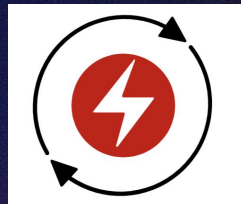
完整管理 TiDB 集群 lifecycle

- 部署
- 升级
- 扩缩容
- 故障恢复
- 数据备份/恢复/导入/导出/迁移

TiDB

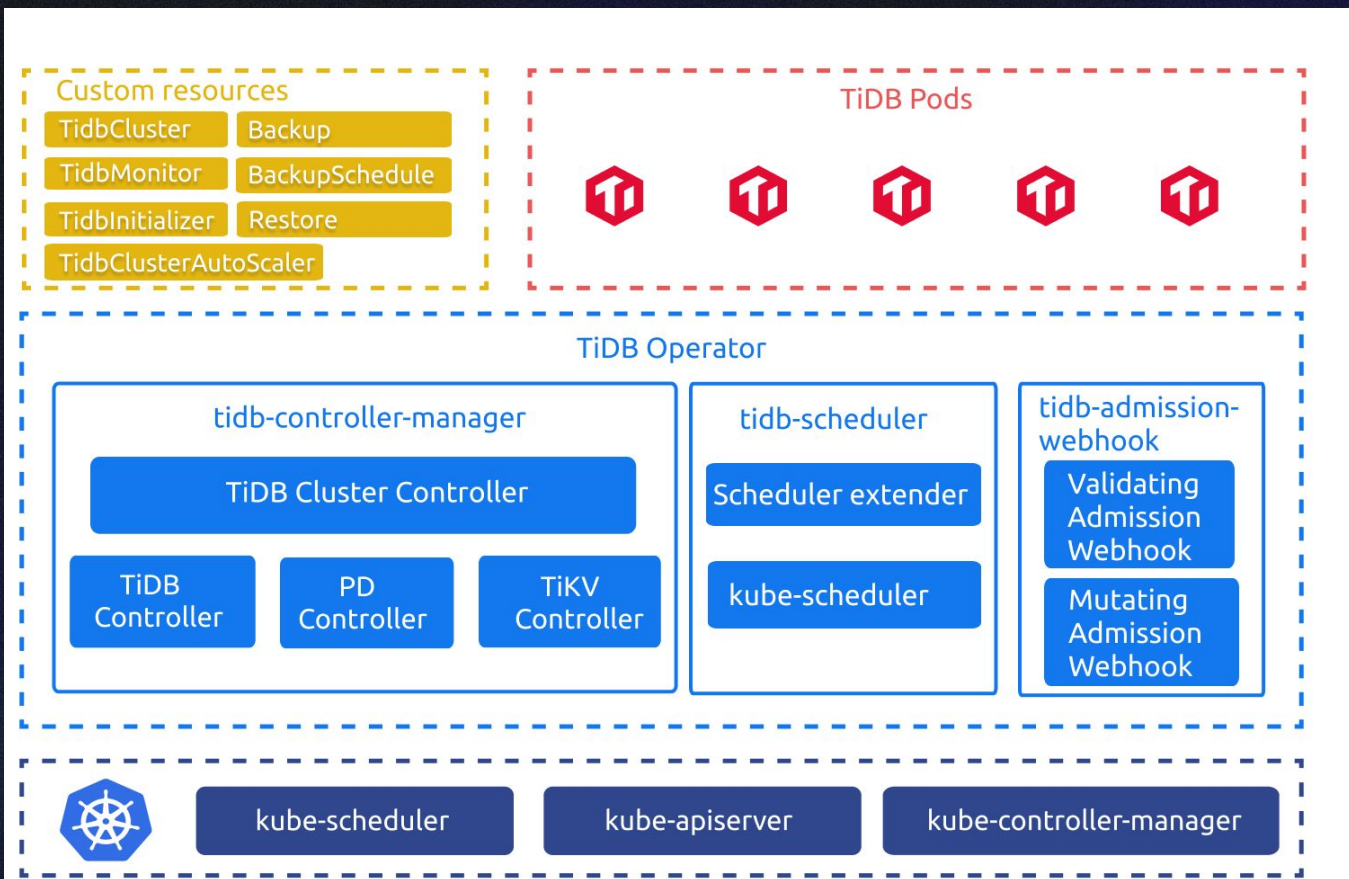


TiDB Operator





TiDB Operator 架构图





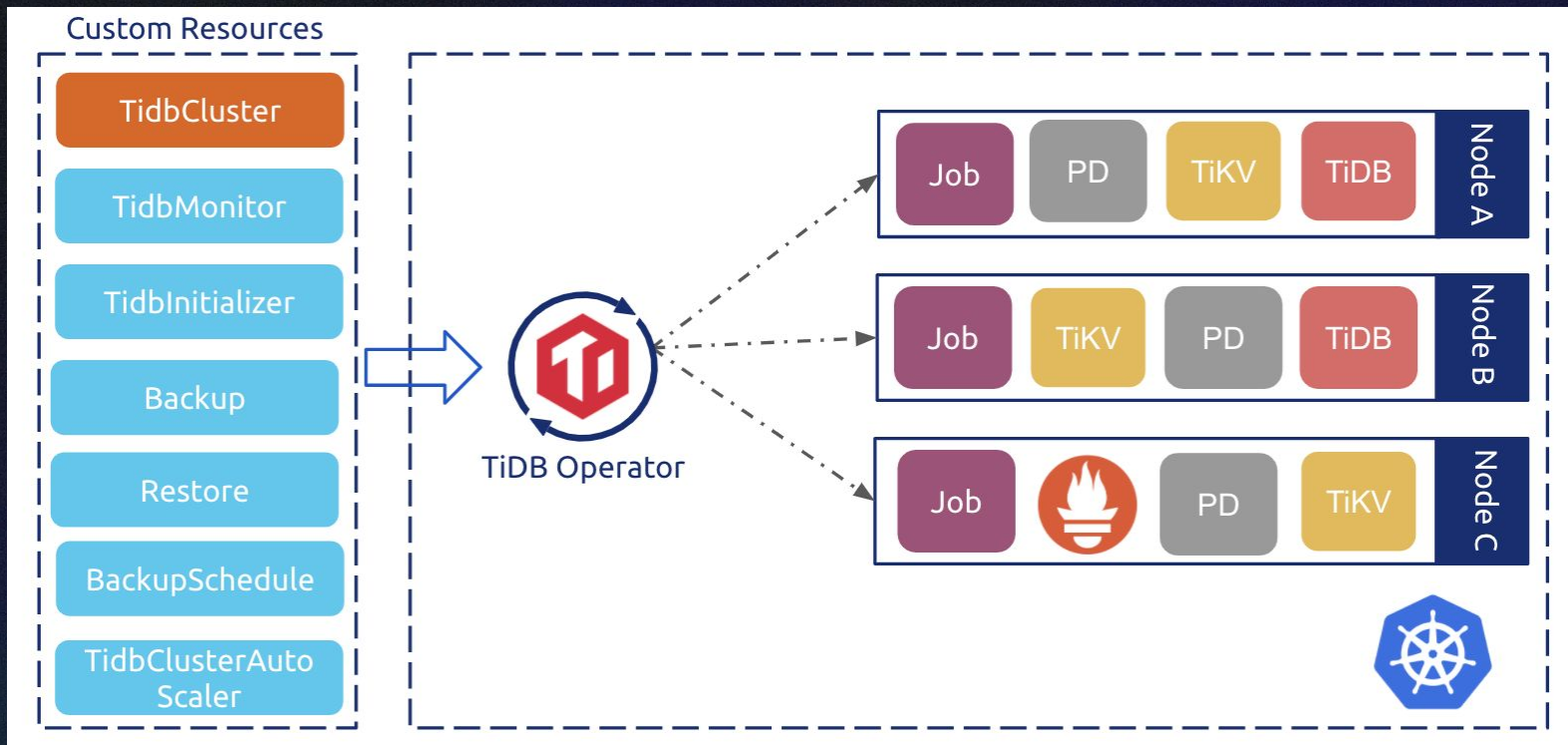
自定义类型

```
apiVersion: pingcap.com/v1alpha1
kind: TidbCluster
metadata:
  name: awesome-cluster
spec:
  version: v4.0.7
  pvReclaimPolicy: Retain
  pd:
    baseImage: pingcap/pd
    replicas: 3
    requests:
      storage: "10Gi"
  tikv:
    baseImage: pingcap/tikv
    replicas: 3
    requests:
      storage: "500Gi"
  tidb:
    baseImage: pingcap/tidb
    replicas: 3
    service:
      type: ClusterIP
```

```
status:
  auto-scaler: null
  clusterID: "6884125266345531046"
  conditions:
  - lastTransitionTime: 2020-10-19T01:52:07Z
    lastUpdateTime: 2020-10-19T01:52:07Z
    message: TiDB cluster is fully up and running
    reason: Ready
    status: "True"
    type: Ready
  pd:
    image: pingcap/pd:v4.0.7
    leader:
      clientURL: http://basic-pd-0.basic-pd-peer.demo.svc:2379
      health: true
      id: "3786817748295515183"
      lastTransitionTime: 2020-10-17T17:16:15Z
      name: basic-pd-0
    members:
      basic-pd-0:
        clientURL: http://basic-pd-0.basic-pd-peer.demo.svc:2379
        health: true
        id: "3786817748295515183"
        lastTransitionTime: 2020-10-17T17:16:15Z
        name: basic-pd-0
  phase: Normal
```

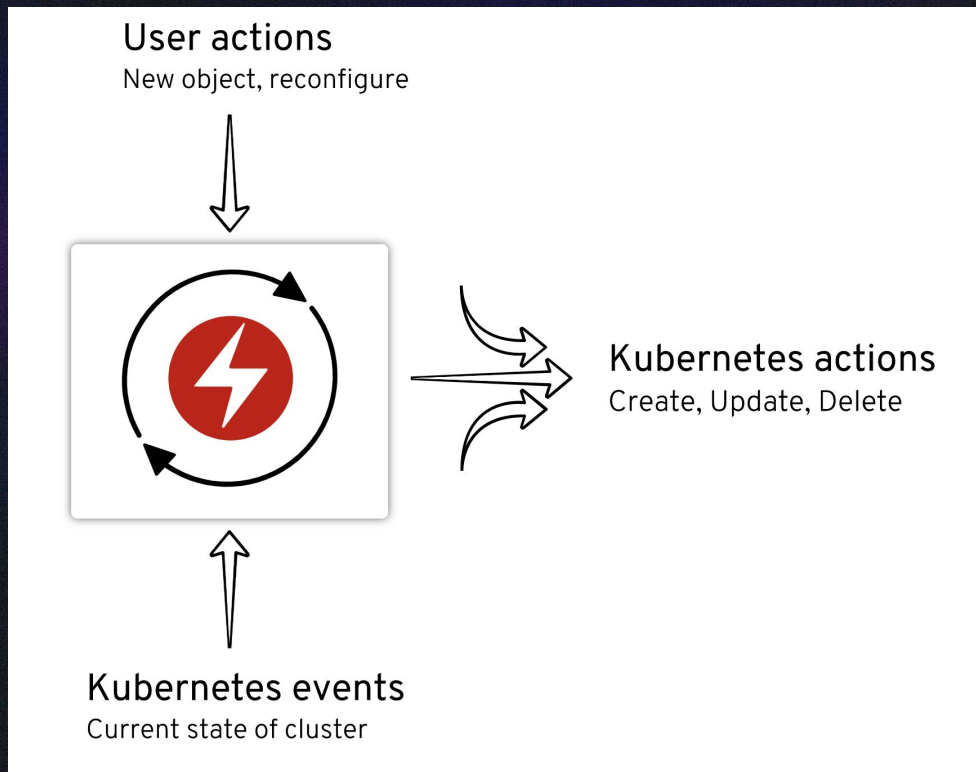
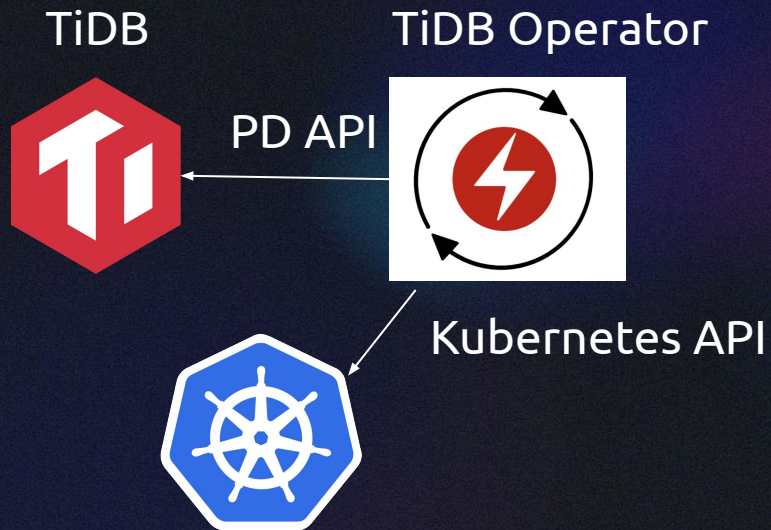



自定义控制器

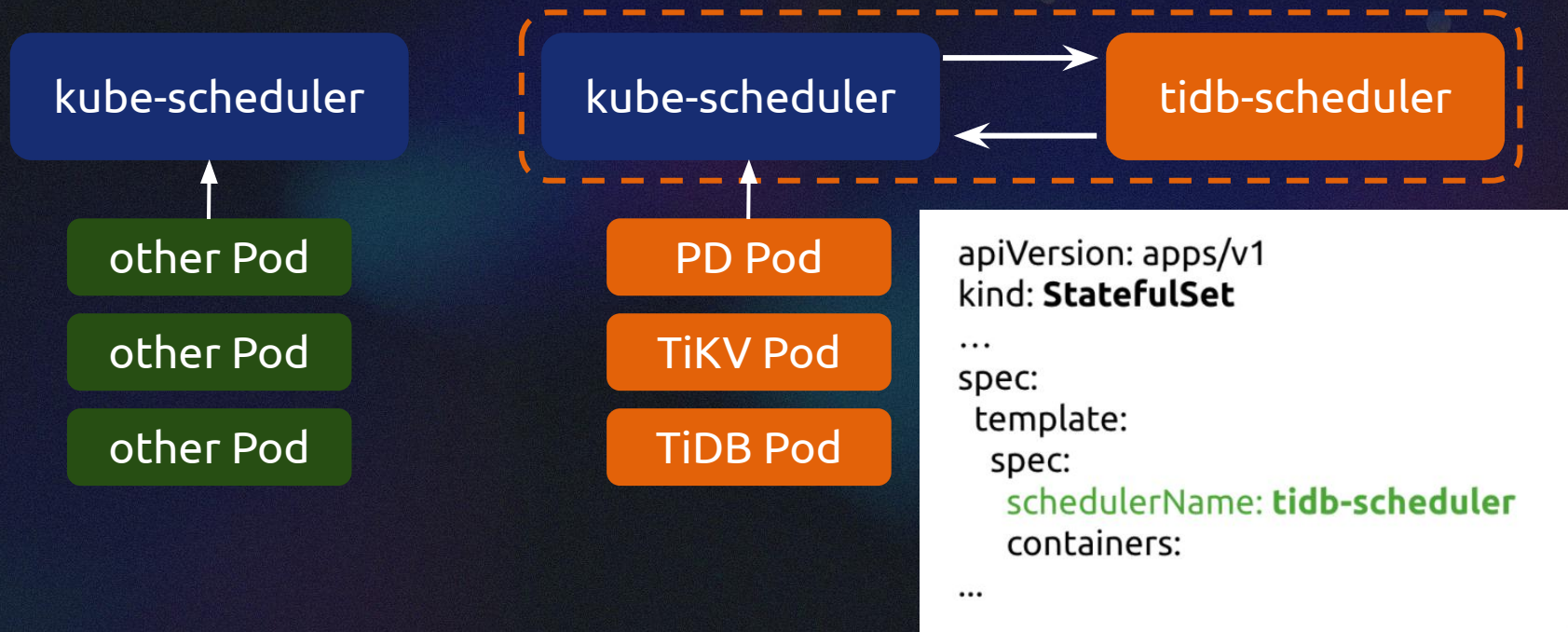




自定义控制器



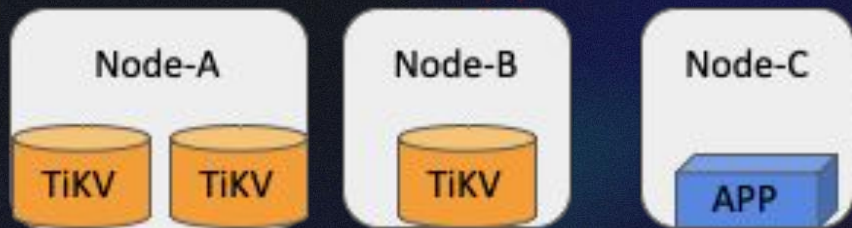
扩展调度器



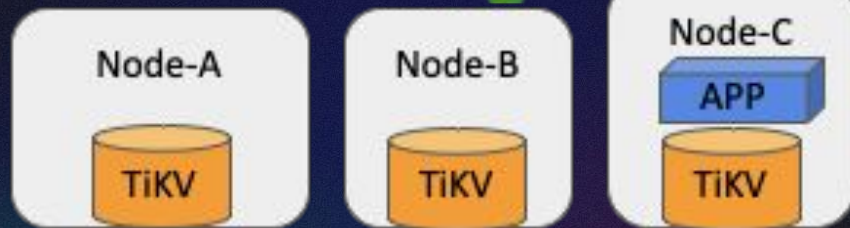


扩展调度器

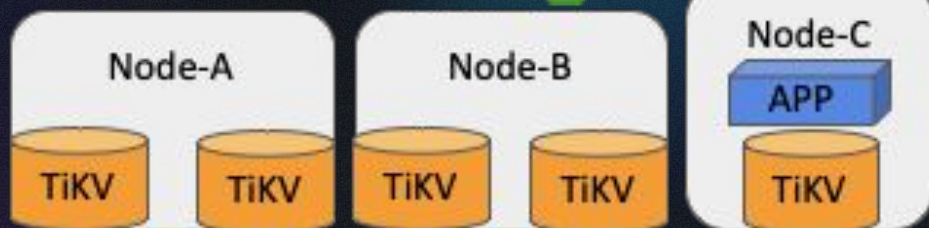
3 TiKV ✗



3 TiKV ✓



> 3 TiKV ✓



TiDB Operator 原理

网易数字 + 云原生论坛

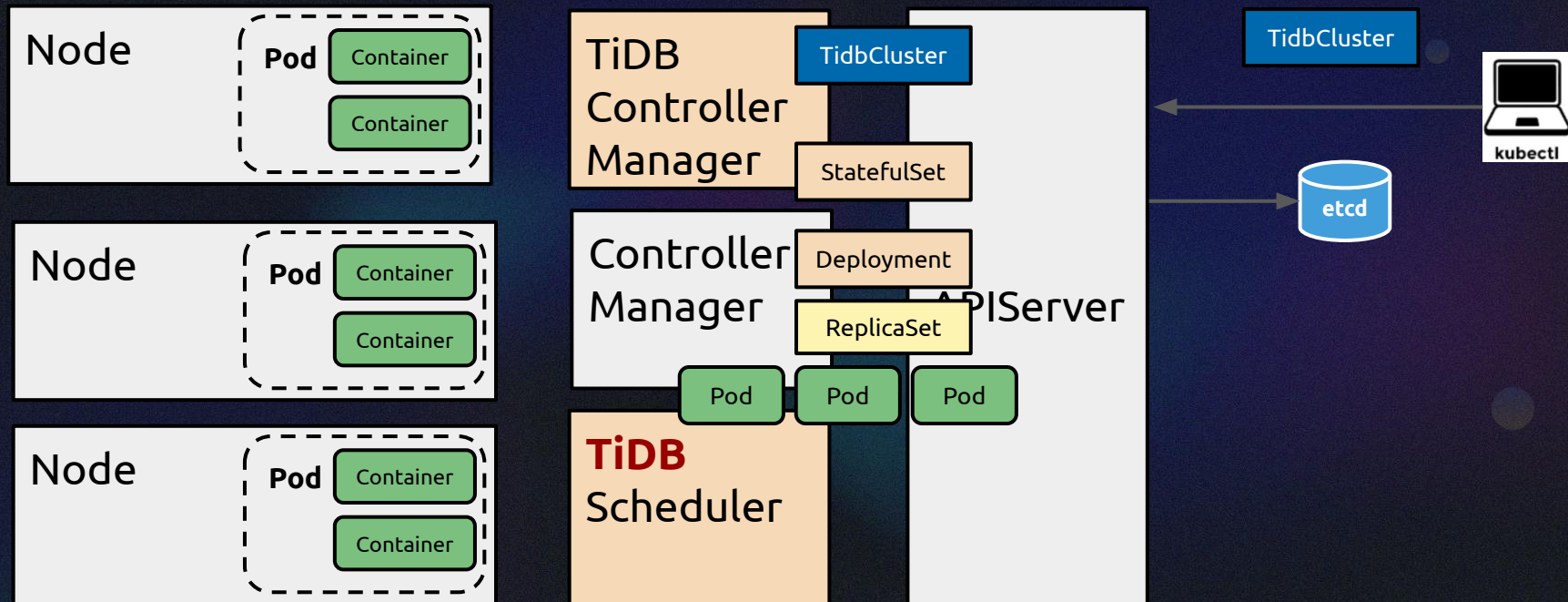
部署集群

用户侧

- 编写 yaml 文件定制集群
- `kubectl apply -f tidb.yaml`



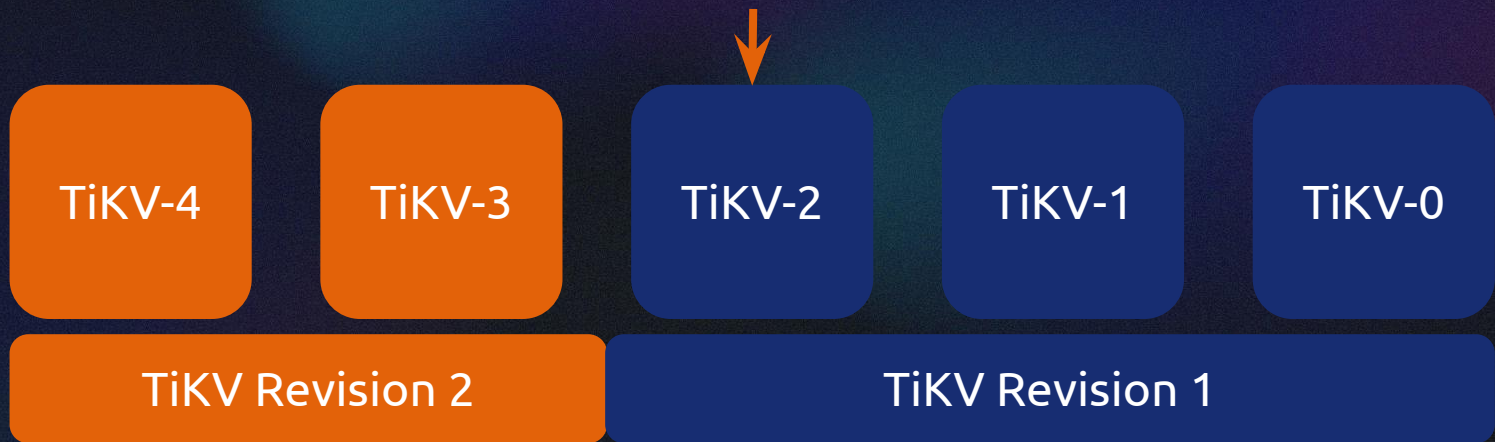
部署集群





升级集群/更新集群配置

- 先调用 PD 迁移 region leader, 再重建容器进行升级





扩容/缩容集群

缩容集群

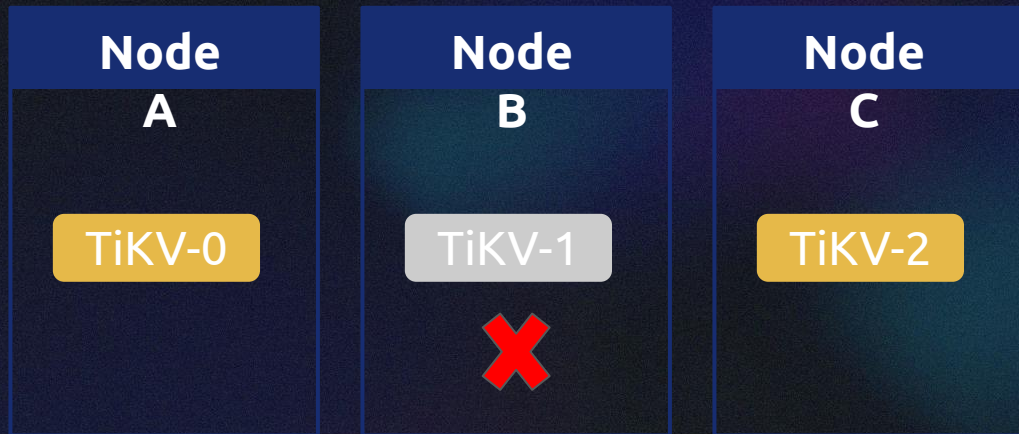
- 标记 PVC 待删除
- 调用 PD 迁移 region leader
- 通过 PD API 删除 TiKV
- Scale in statefulset

扩容集群

- 删除标记的 PVC
- Scale out statefulset



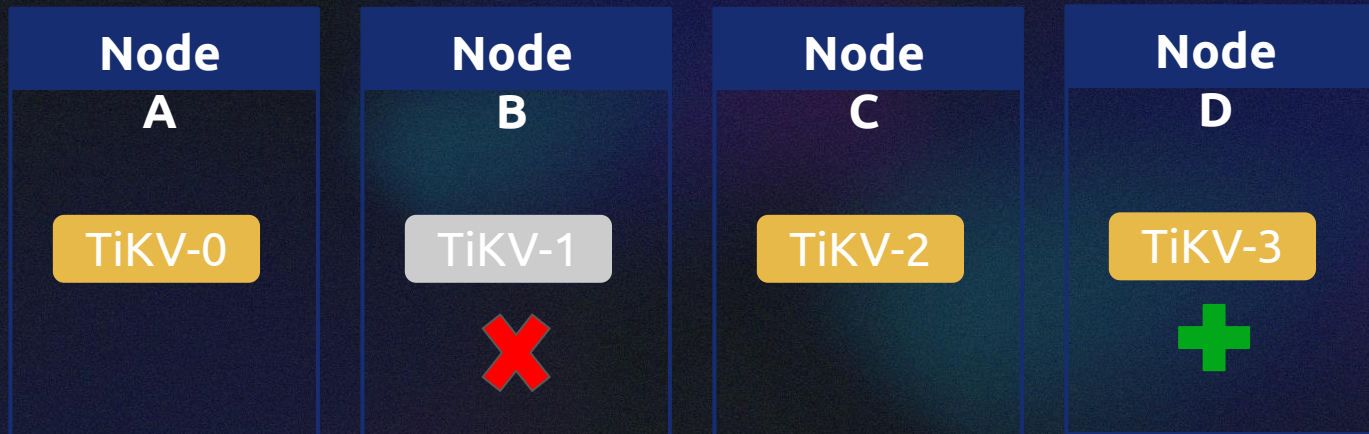
故障转移



TiDB Operator 的控制器能够通过 PD 中的状态信息判断 TiKV store 是否异常



故障转移



补充新的
TiKV



数据备份/恢复/导入/导出

Backup

Restore

BackupSchedule

- 定制化 CR, 减少用户配置部署负担
- 创建备份恢复 Job
- 统计信息到 CR Status
- 支持定时备份, 数据更安全



性能问题

- 支持独享节点 (Dedicated) 部署与混部, 可以按照优先级权衡性能与成本;
- 支持通过 HostNetwork 部署集群, 消除二层网络开销;



稳定性问题

- K8S Master 故障: 不影响集群
- K8S Node 故障: 自动故障转移
- K8S 全集群故障: 保留存储, 确保不丢数据
- 灾难恢复: 周期性备份

TiDB Operator 后续演进

网易数字 + 云原生论坛



更灵活、更高可用的部署方案

TiDB Operator 支持配置异构部署

- 支持部署异构的 TiDB 集群, 异构集群间的配置不同

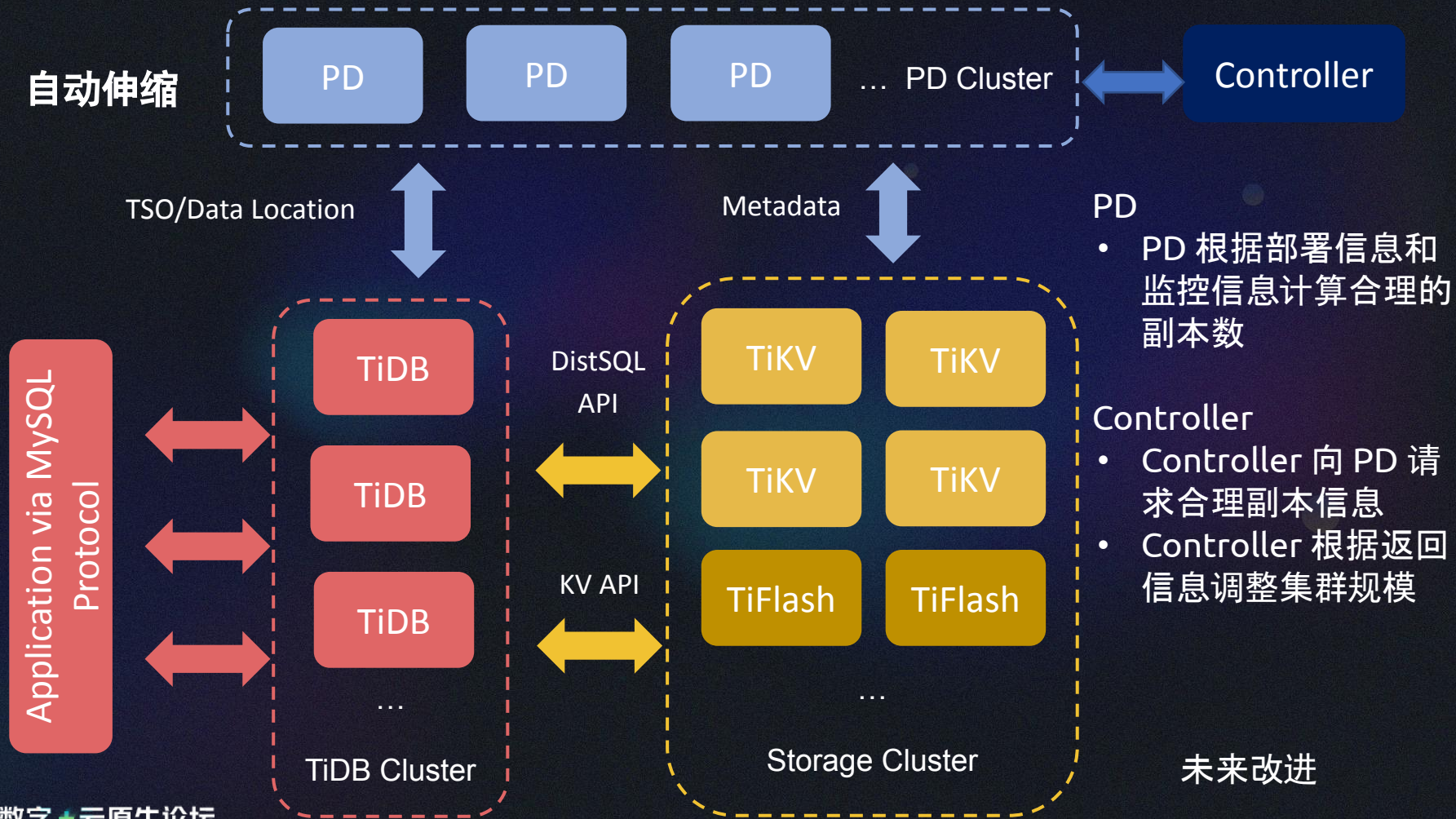
TiDB Operator 支持跨 Kubernetes 部署

- 使 Operator 能够跨多个部署在不同区域数据中心的 Kubernetes 集群部署 TiDB 集群, 以实现区域或数据中心级别的HA。

节省硬件开销, 部署配置更加灵活

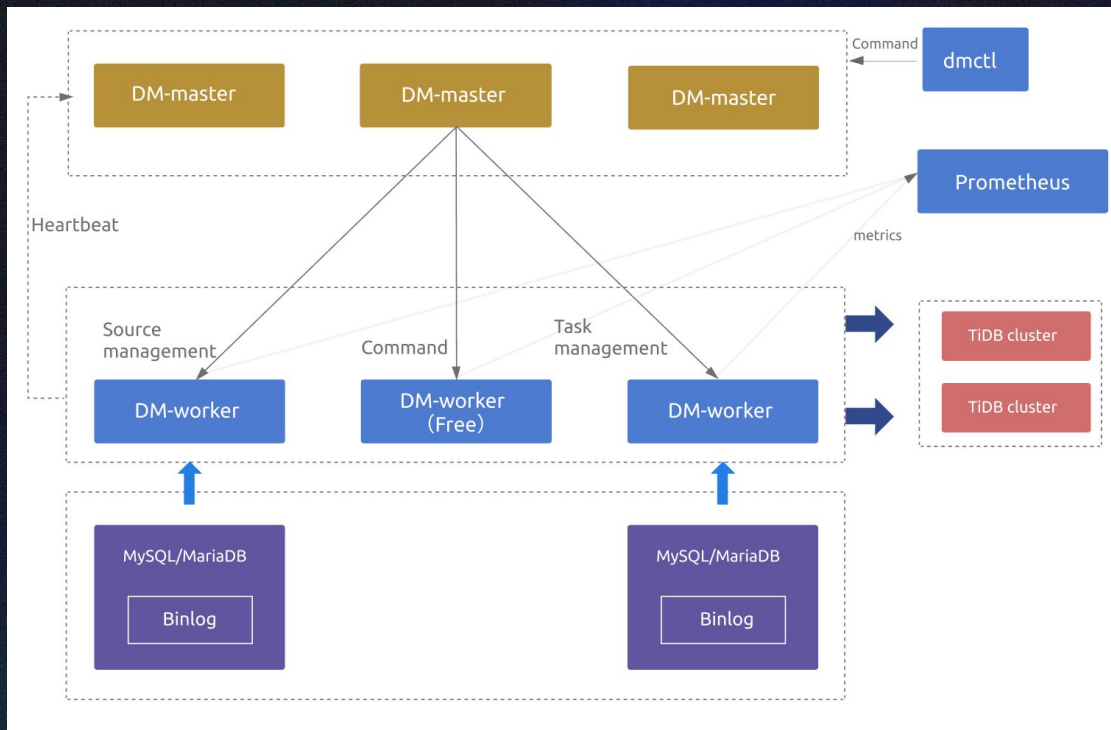


自动伸缩





支持部署数据迁移工具 TiDB DM



支持部署数据迁移工具
TiDB DM 到 Kubernetes
集群, 帮助用户方便、高可
用地迁移例如 MySQL 的
数据到 TiDB



后记

更多 TiDB Operator 的设计、讨论、使用：

- <https://docs.pingcap.com/zh/tidb-in-kubernetes/stable>
- <https://github.com/pingcap/tidb-operator>

也欢迎移步 TiDB sig-k8s channel 与我们讨论：

- <https://tidbcommunity.slack.com/archives/CHD0HA3LZ>



THANK YOU

网易数字+云原生论坛