

数据集Operation重构设计

仇伶俐（Fluid项目开发者和使用者）

中国电信

qiulingwei.js@chinatelecom.cn

目录



1

背景简介

2

现存问题

3

架构规划

4

详细设计

Fluid 是面向云原生**数据集**的编排与加速系统

Fluid 通过不同 CRD 向用户提供了不同**数据集操作能力**

通过 DataLoad 向用户提供了数据预热能力

通过 DataBackup 向用户提供了元数据备份能力

还计划引入其它 CRD 支持更多数据集操作能力

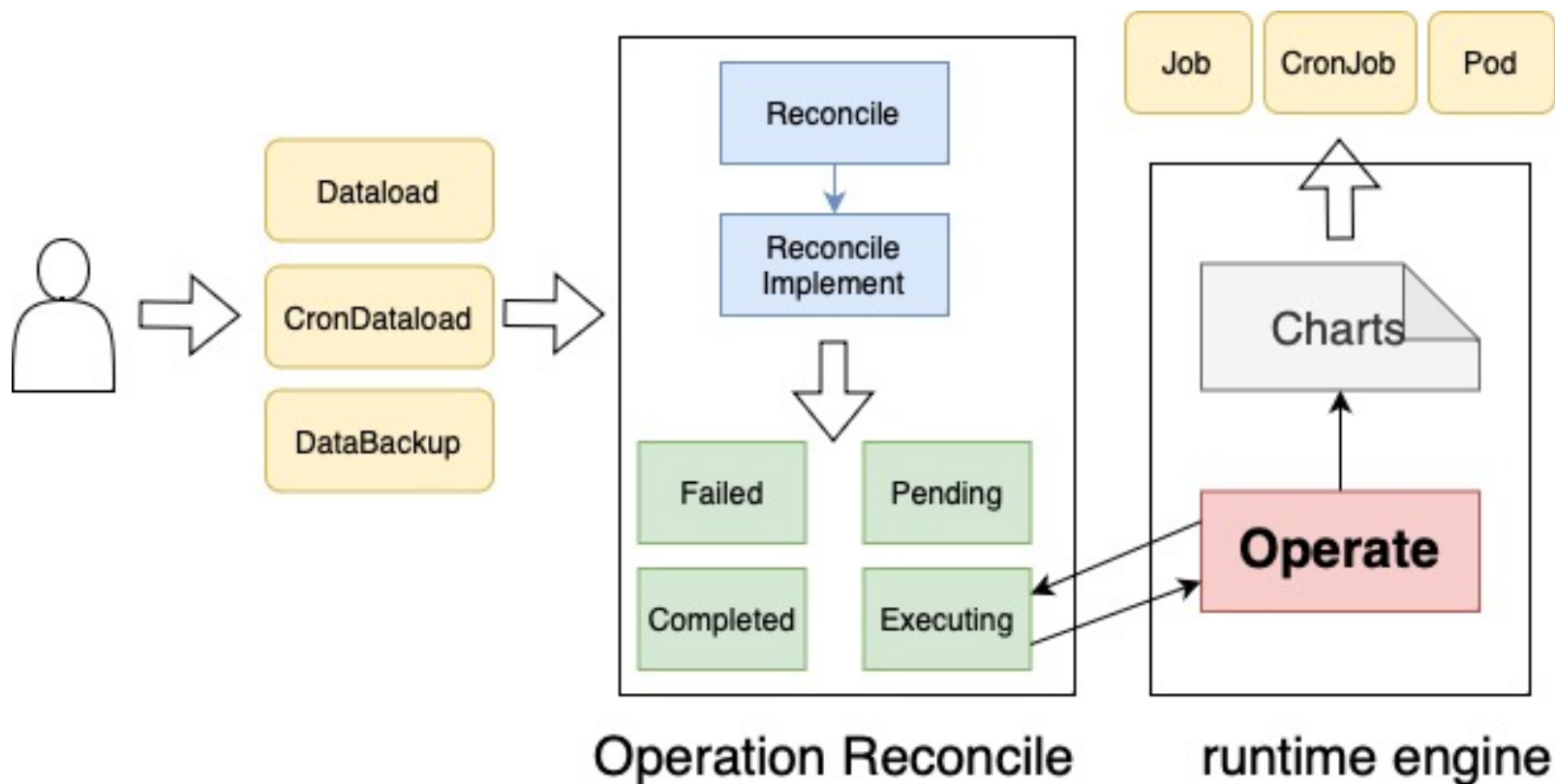
目前没有为各类 Operation 提供统一的调用入口

Fluid 中的各类 Runtime , 需要为每种 Operation 都实现专属方法

每次新增支持的 Operation 类型 , 也需要同步修改所有 Runtime 的实现

面向Operation的**统一抽象**：

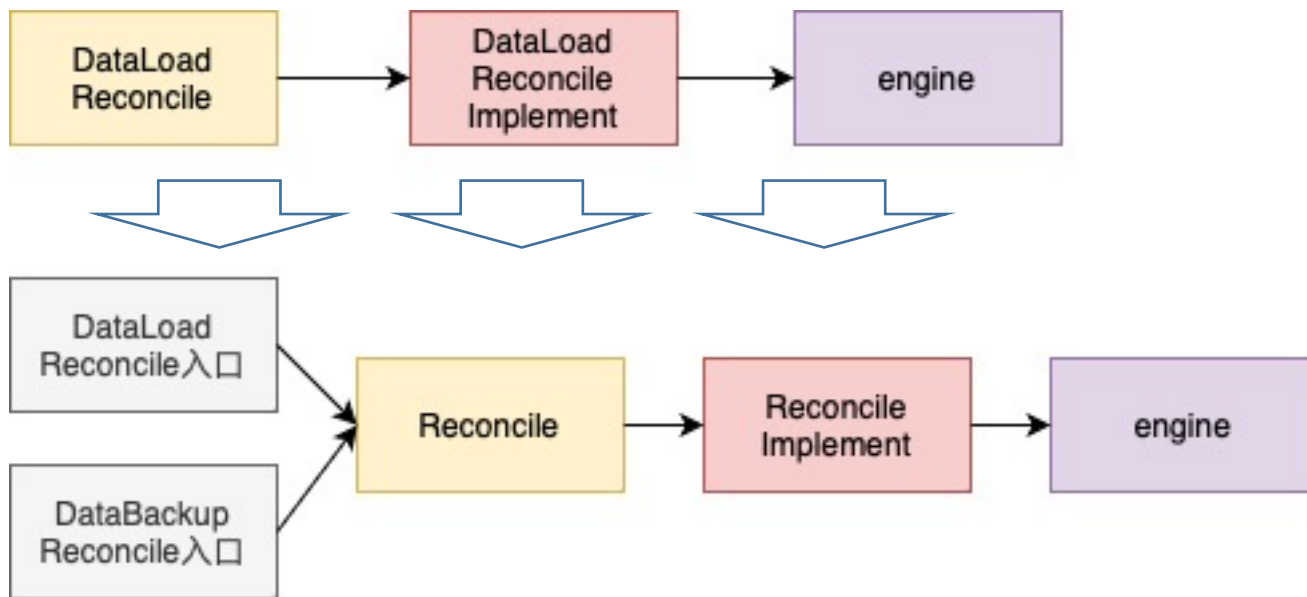
- CR对象Reconcile逻辑
- Operation生命周期
- Charts执行方式



统一抽象CR对象Reconcile逻辑



每个CR对象有独立的Reconcile逻辑



每个CR对象有只有一个Reconcile入口

根据Namespace+Name获取CR对象后，进入统一的Reconcile逻辑

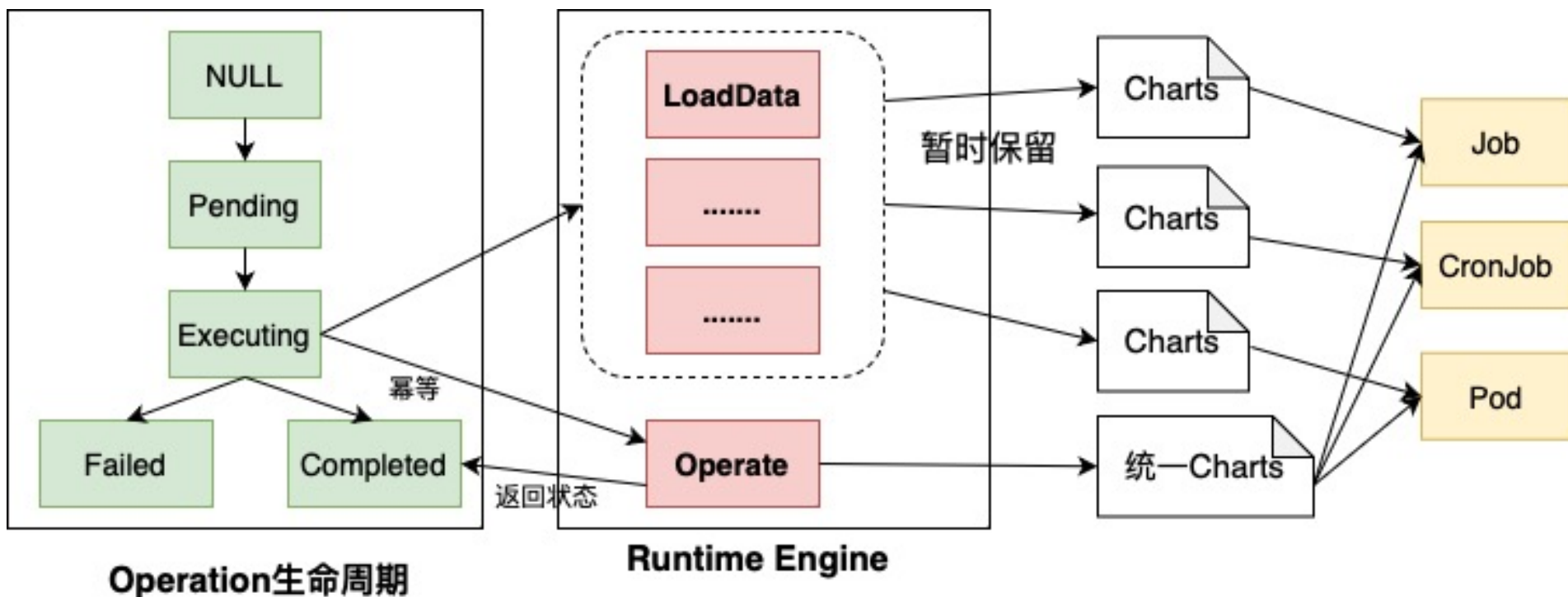
统一抽象Operation生命周期和Charts



engine.Operate方法接收传入的CR对象，随后进行类型断言

engine.Operate方法应当是幂等的

engine.Operate方法返回Operation的当前执行状态



- 现有数据预热能力优化

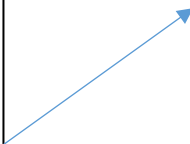
Dataload CRD保留了Options字段，可用于传递参数

- 引入定时数据预热能力

CronDataload CRD设计：

```
spec:
  suspend: false
  schedule: "*/5 * * * *"
  startingDeadlinesconds
  dataloadTemplate:
    spec:
      ...
  successFulJobsHistoryLimit: 3
  failedJobsHistoryLimit: 1
Status:
  active: ...
  lastScheduleTime: ...
  lastSuccessfulTime: ...
```

根据CronDataLoad
创建和维护CronJob



相关PR已在开发中
完成后欢迎Review!