RESILIENCE

REALIZED

KubeCon | CloudNativeCon

North America 2021

Fluid v0.7.0

Fluid Theme

- Data and Workload orchestration
- How to use cache better for Big Data and AI Application
- Scalable，Portable，Manageable Dataset
- Different runtime with the same concept model
- Easy to use for Hybrid Cloud and Edge

V0.7.0 Overview

- Orchestration
  - Dataset Scheduling
  - Task Scheduling
  - Collaboration of task and dataset scheduling
- Data Driven Application
  - Machine Learning (TensorFlow, PyTorch) workloads through PVC(POSIX API)
  - Big Data(Presto, Spark) workloads through ConfigMap(HCFS API)
- Automated Operation
  - CronJob for metadata/data load
- Runtime Life Cycle management
  - HA Improvement
  - Refactor CSI Plugin for stable architecture
  - Support Runtime in Container Network for underlayer network
  - Worker/Fuse in the same Linux Process
- Agent
  - Report disk number and capacity , Page cache and so on

Enhancement

New

Feature details

- Co-orchestrate workload and dataset
  - Submit Task and dataset at the same time
  - Prefer the node with cache capability when the resource requirement of the workload is meeting
  - Then schedule the cache workers to the nodes which already have workload on it

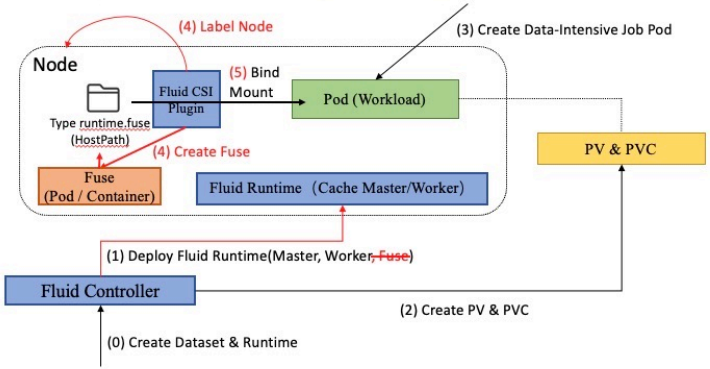
- Big Data(Presto, Spark) workloads through ConfigMap(HCFS API)
  - Workload should be scheduled with data locality
  - . The cache directory should be injected into workload Pod
  - Short circuit read should be enabled with considering workload pod in container Network

- Support Runtime in Container Network for underlayer network
  - Master/Worker in Container Network, Fuse in HostNetwork
  - Runtime should also support this mechanism (Node ip/Pod ip)

Feature details

Refactor CSI Plugin

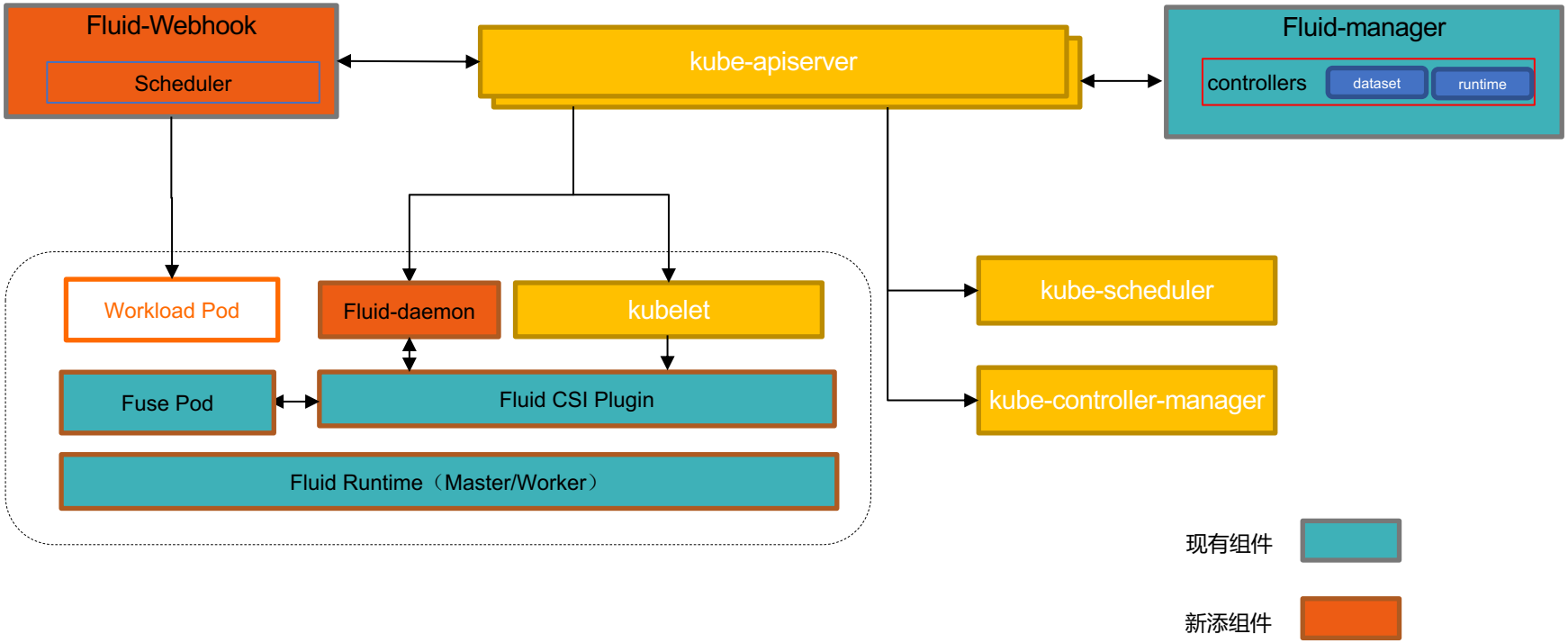
- Lazy Pod initialization
- Avoid Pod eviction and OOM Killer
- Make Scheduler aware of reusable fuse cache

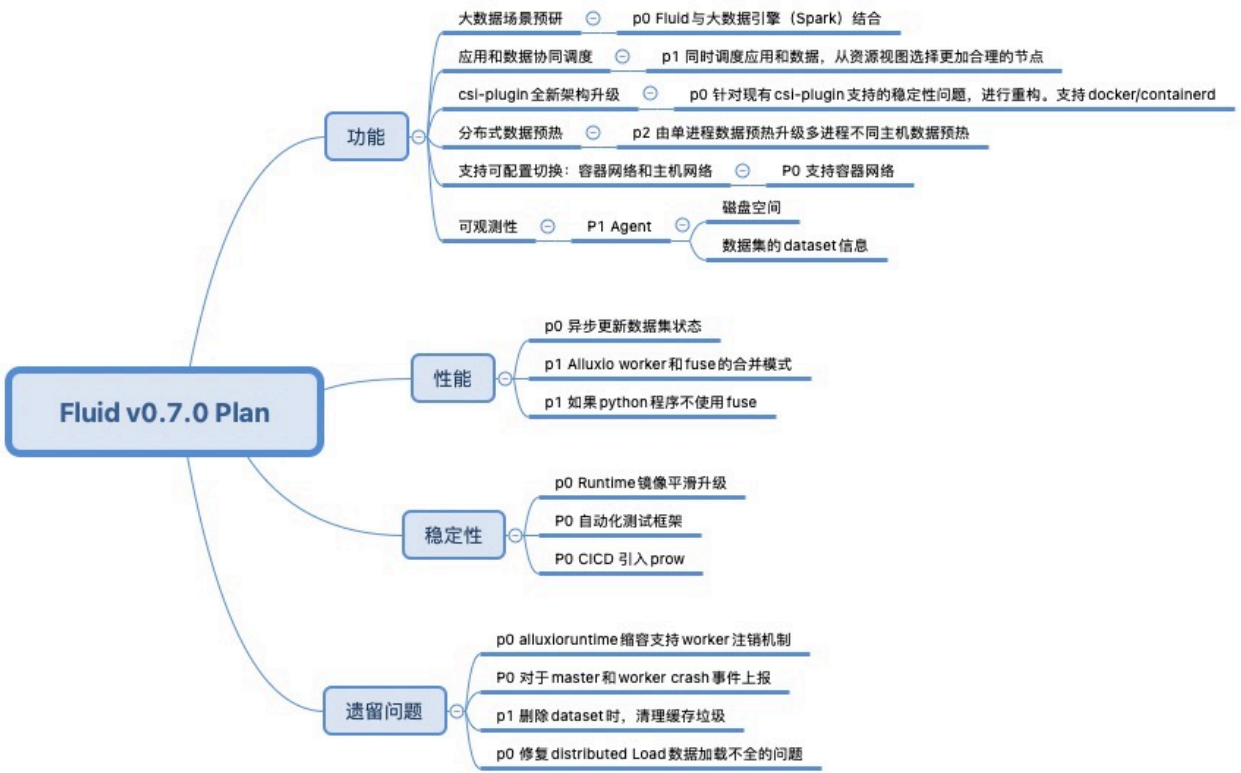


Agent

- Report disk capacity
- Map Disk and path of tiered storage
- Make Dataset scheduling with disk capacity info

Architecture




Fluid v0.7.0 Plan
功能
大数据场景预研
p0 Fluid与大数据引擎（Spark）结合
应用和数据协同调度
p1 同时调度应用和数据，从资源视图选择更加合理的节点
csi-plugin全新架构升级
p0 针对现有csi-plugin支持的稳定性问题，进行重构。支持docker/containerd
分布式数据预热
p2 由单进程数据预热升级多进程不同主机数据预热
支持可配置切换：容器网络和主机网络
P0 支持容器网络
可观测性
P1 Agent
磁盘空间
数据集的dataset信息
性能
p0 异步更新数据集状态
p1 Alluxio worker和fuse的合并模式
p1 如果python程序不使用fuse
稳定性
p0 Runtime镜像平滑升级
P0 自动化测试框架
P0 CICD 引入prow
遗留问题
p0 alluxioruntime缩容支持worker注销机制
P0 对于master和worker crash事件上报
p1 删除dataset时，清理缓存垃圾
p0 修复distributed Load数据加载不全的问题