



Fluid 0.9.0 规划

<https://github.com/fluid-cloudnative/fluid>

Fluid简介

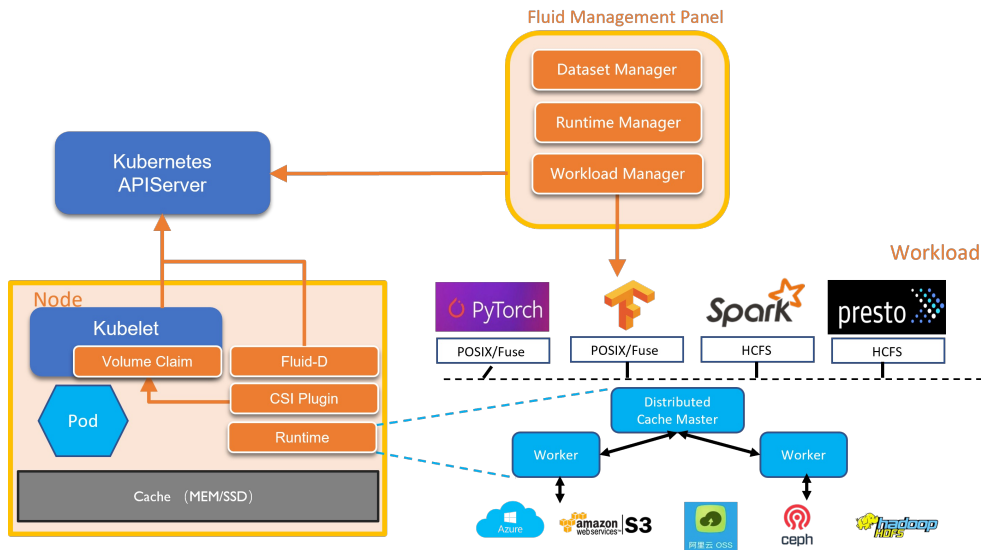


Fluid即流体

通过自定义资源（CRD）提供数据层的抽象，让数据在Kubernetes里自由的流动，更好地为数据密集型应用服务。

Fluid定位

服务于云原生AI和大数据应用的数据调度系统，关心数据本身和使用数据的应用生命周期，定义标准的K8s上数据使用流程，特别适用于计算存储异地部署的混合云场景



挑战

- 多种存储复杂度高
- 数据访问慢
- 低效的负载调度
- 云原生环境复杂多样
- 大量数据集的管理成本高昂和资源浪费

趋势

- 弹性资源可编程
- 硬件环境升级
- 云原生生态变为业界标准

现状

- 2021.4进入cncf sandbox
- OSCAR 尖峰开源项目和开源社区
- 15+ 生产使用

1. 将数据加速能力
在云原生环境中开
箱可用

2. 成为更多数据访
问场景的关键路径

3. 云原生数据场景
下的基础设施

跨越边界

跨越namespace，跨越集群数据访问



简单接入

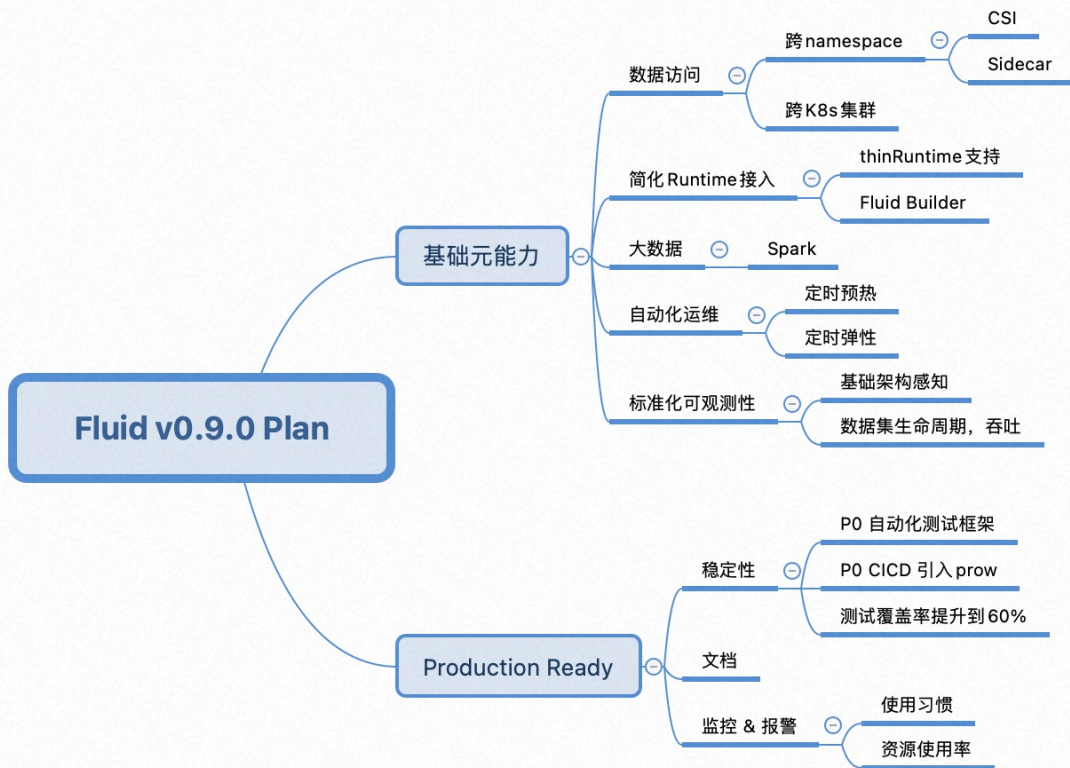
简化数据客户端接入，降低CSI的维护成本



大数据场景

支持HCFS协议访问数据，提供数据亲和调度和依赖注入

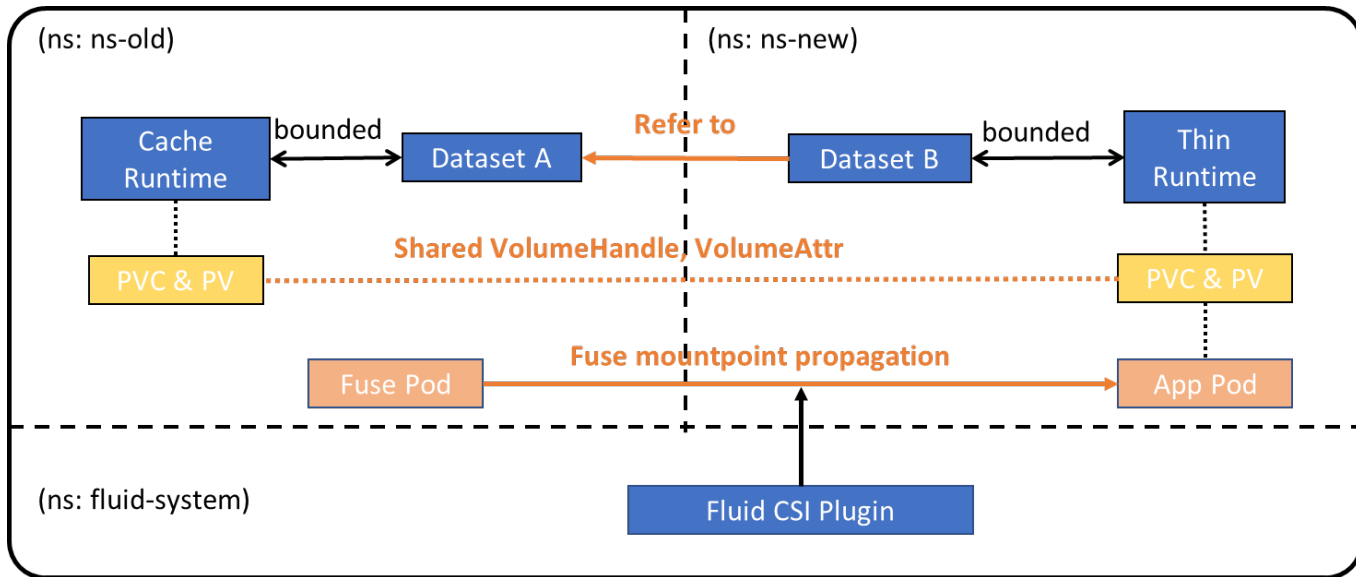




- **数据访问**
 - 跨namespace访问 (P0)
 - 跨Kubernetes集群访问 (P1)
- **存储客户端简单接入**
 - 支持0代码的存储客户端接入 (P0)
 - 无缓存或者本地缓存的场景
 - 支持分布式缓存runtime生成的脚手架
- **大数据场景**
 - 大数据负载 (Spark) 概念验证POC (P0)
- **智能运维**
 - 定时元数据/数据预热 (P0)
 - 定时Runtime 扩缩容 (P1)
- **产品化建设**
 - 文档建设 (P0)
 - 稳定性建设 (P0)

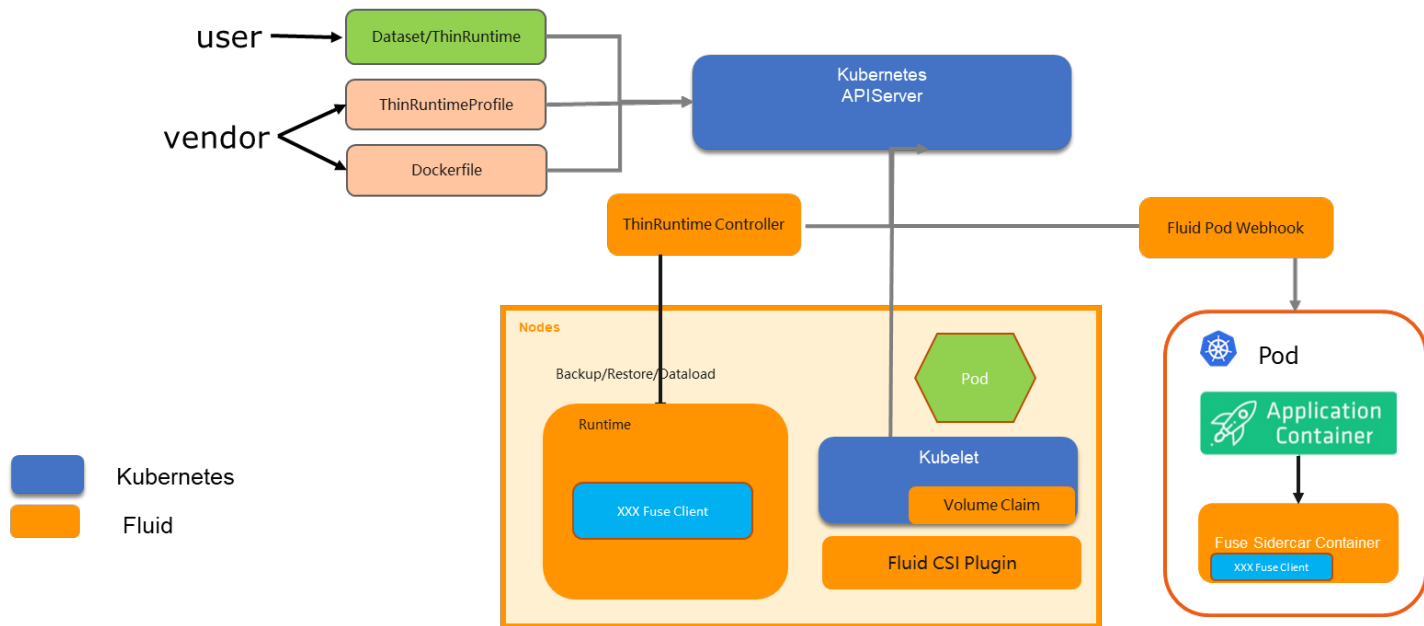
跨namespace数据访问

- CSI场景，涉及fuse客户端共享
- Serverless场景（Sidecar）
- 生命周期管理



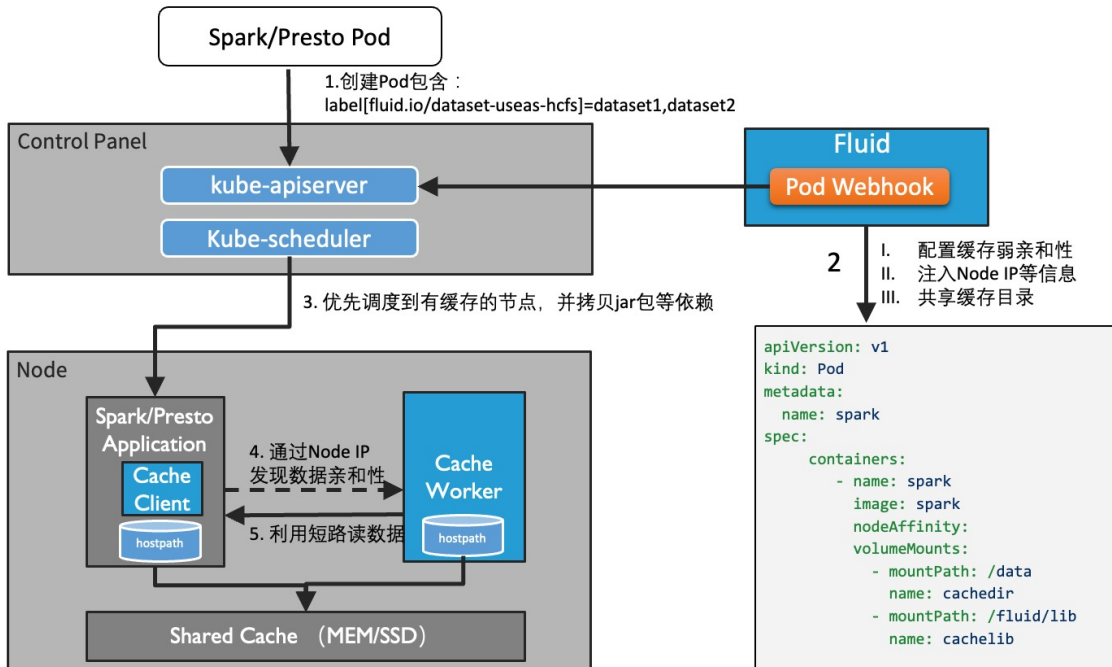
低代码的存储客户端接入Fluid

- 只需容器镜像维护, 不希望维护CSI Plugin
- 在Serverless场景下运行自定义的客户端



大数据负载（Spark）

- 感知Application Pod使用的数据集
- 结合数据亲和性信息进行工作负载的调度
- 支持缓存启用短路读/写特性



01

文档优化

- 结构调整，一站式的地图入口避免跳来跳去
- 英文文档完善

02

自动化能力

- 功能测试自动化覆盖率提升到60%
- 提供端到端测试



Thank You!