



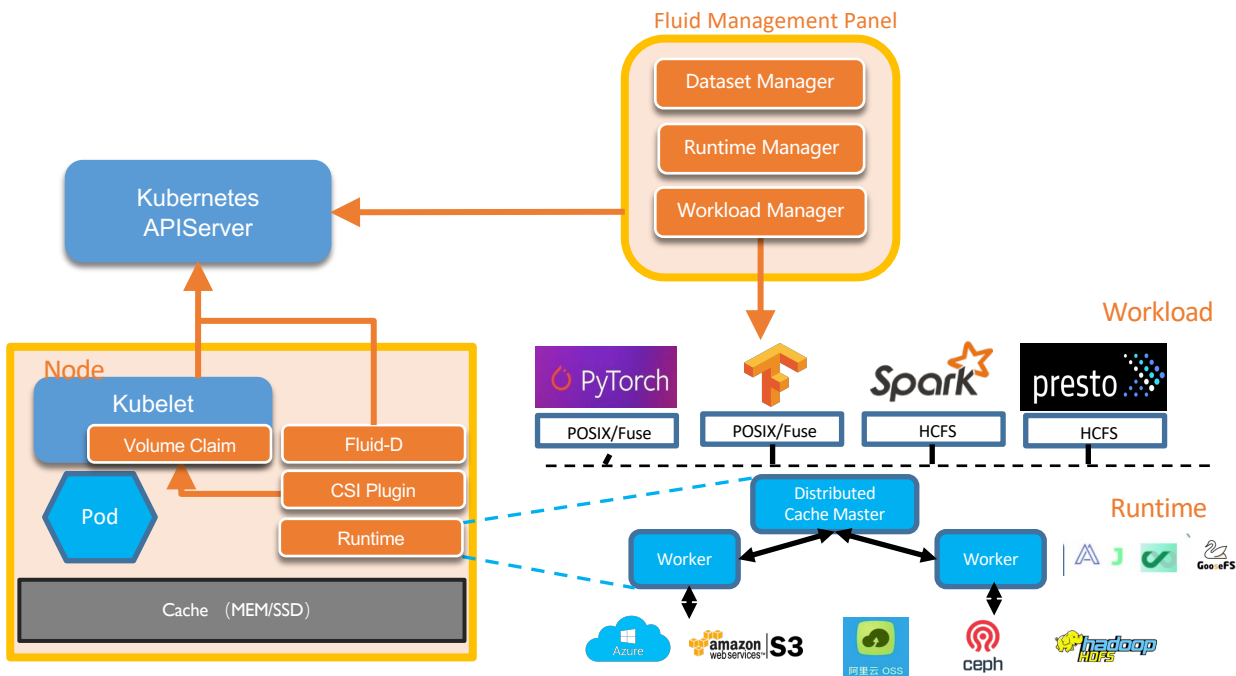
Fluid 0.8.0 规划

Fluid简介 (<https://github.com/fluid-cloudnative/fluid>)



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**，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



计算，内存，磁盘，
网络，缓存作为资源，
可观测和管理

资源感知



本地/近地缓存的亲和性/
协同调度和管理

应用调度/管理



AI, Big Data,
Serverless,
边缘，混合云

场景融合

数据流动

数据可以在计算集群中调度，
排布和迁移



数据弹性

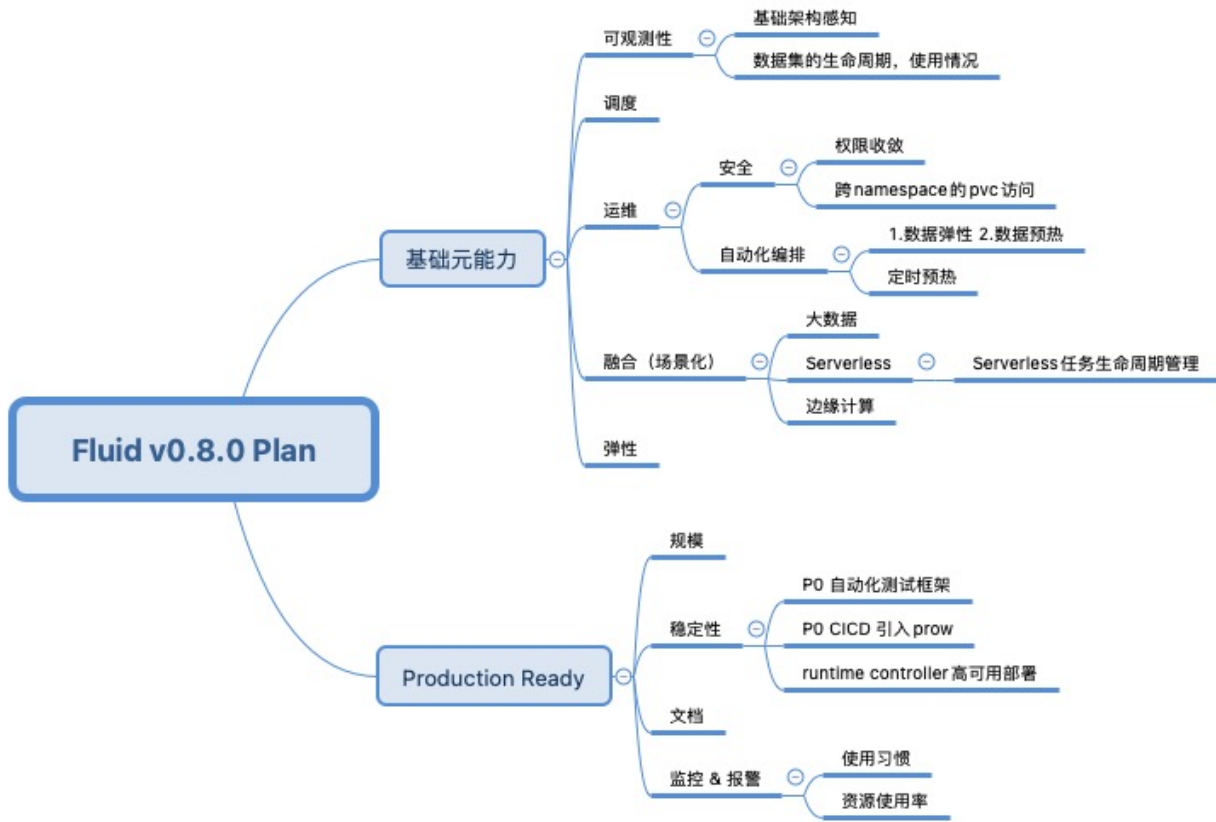
容量、吞吐随时按需扩缩，
Cache Runtime可扩展



智能运维

数据预热，备份，恢复，
加密；数据使用模式分析；
即插即用；





➤ 场景融合

- Serverless场景可以在生产环境运行(P0)
- 大数据负载(Presto, Spark) 概念验证POC(P0)
- 边缘场景概念验证POC(P1)

➤ 资源感知和管理

- 将内存，磁盘作为调度资源进行管理和调度 (P0)
- 对于分配的资源进行隔离和限制 (P1)

➤ 智能运维

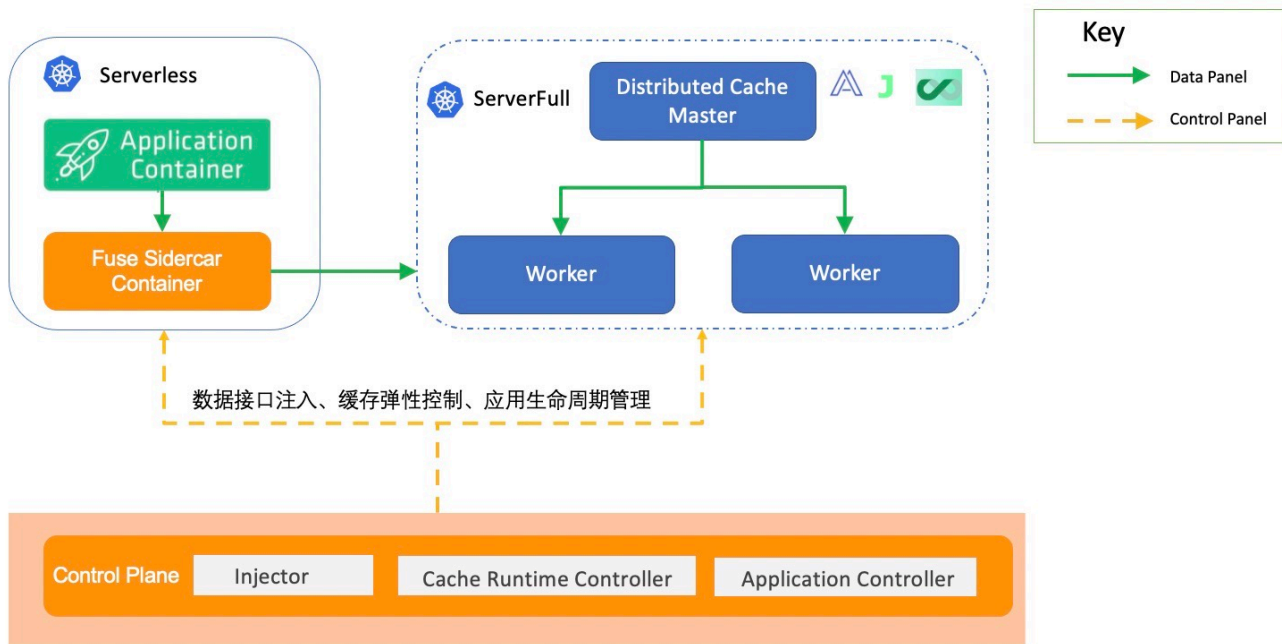
- 定时元数据/数据预热 (P1)
- 跨namespace的数据集赋权 (P1)
- 支持容器网络(P1)

➤ Production Ready

- 文档建设(P0)
- 稳定性建设(P0)

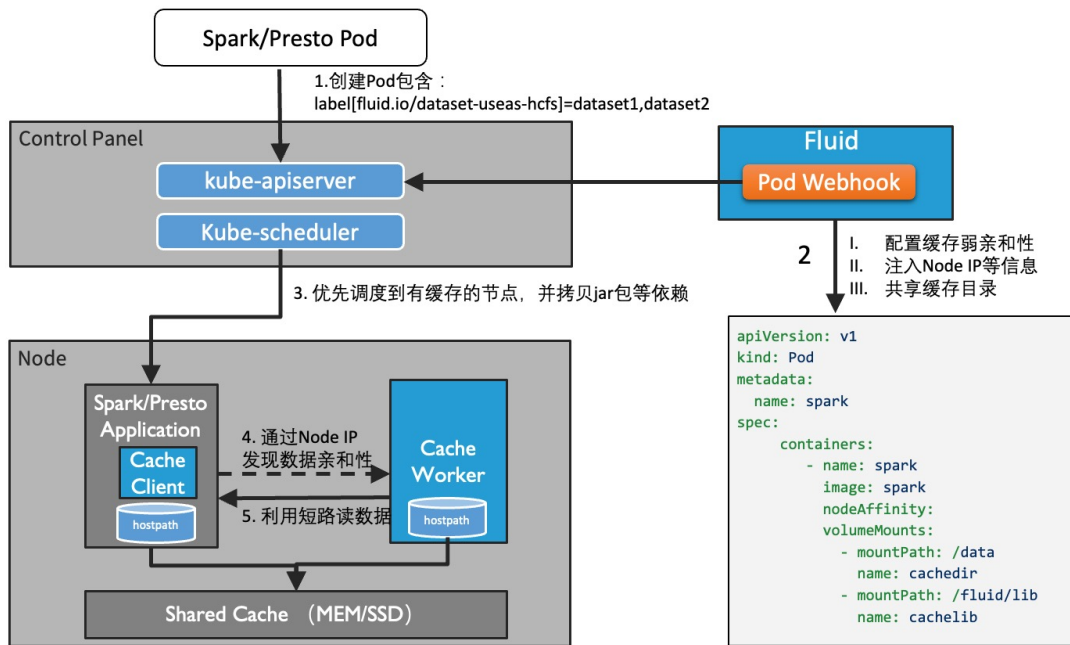
功能细节

- Serverless场景可以在生产环境运行
 - 安全问题（移除Fuse的privileged权限）
 - 应用全生命周期支持（任务类型应用，应用镜像预热）
 - 规模化问题



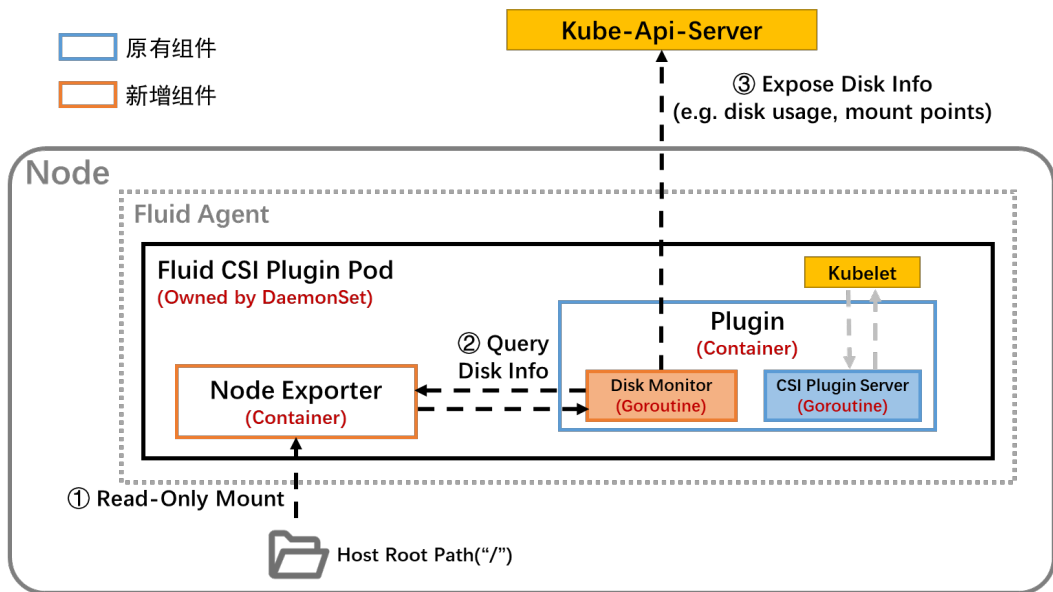
功能细节

- 大数据负载(Presto, Spark) 概念验证POC
 - 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



功能细节

- 将内存，磁盘作为资源进行管理和调度
 - 通过Node Exporter实现资源发现
 - 通过Extended Resource实现资源调度
 - 将目录和设备进行映射，对于最终用户隐藏调度复杂性



➤ 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

➤ 自动化能力

- 功能测试自动化覆盖率提升到30%