

Runtime 整体架构

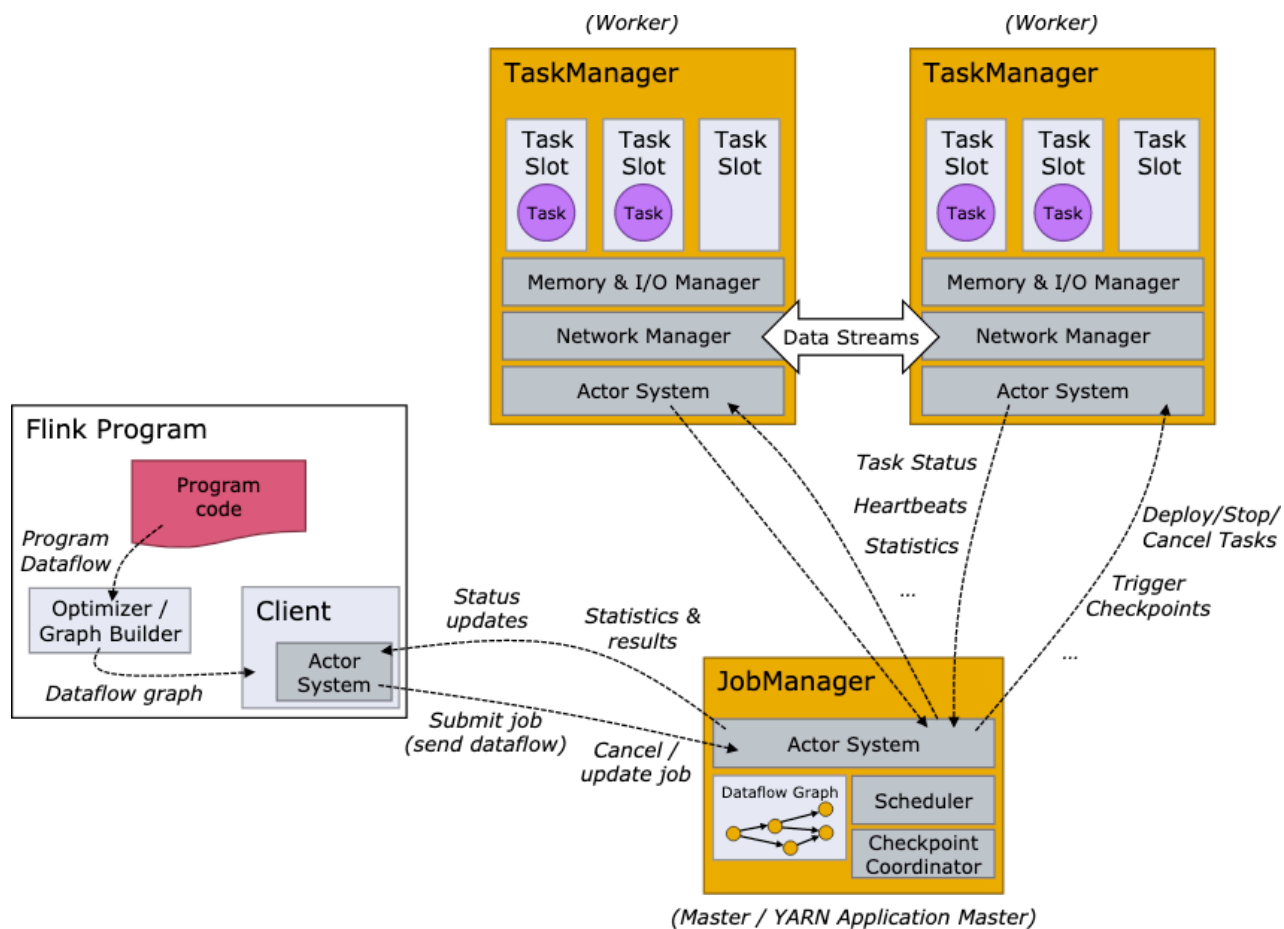


扫码试看/订阅

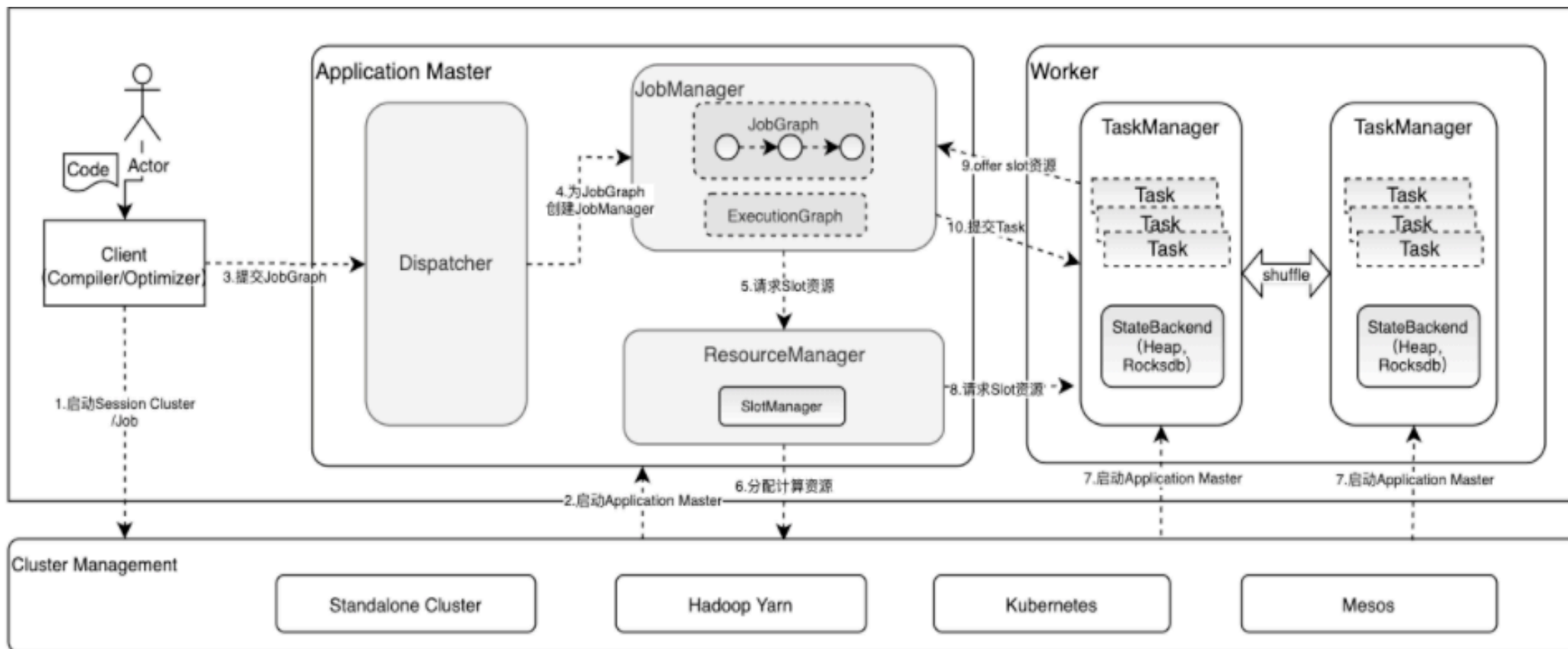
《Flink核心技术与实战》视频课程

Flink Runtime 整体架构

- 整体架构包含了三个部分：
 - JobManager (Master)
 - TaskManager (Worker)
 - Client

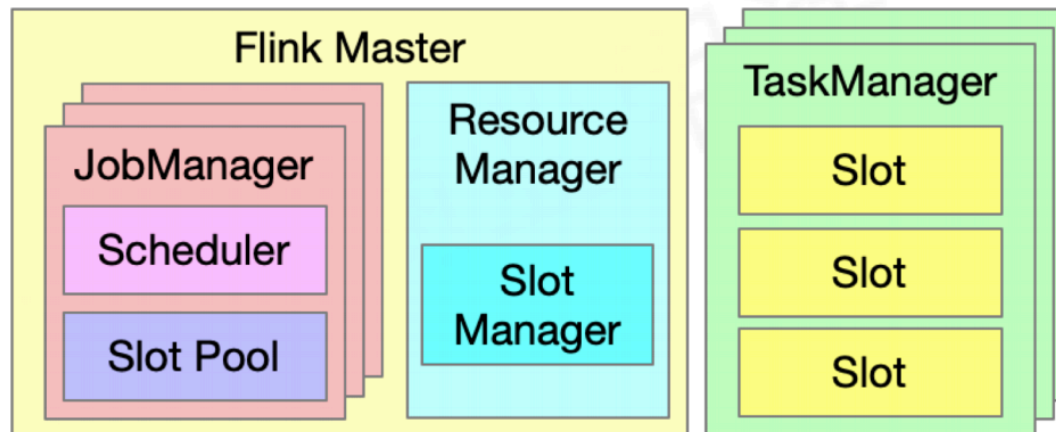


Runtime 整体架构设计



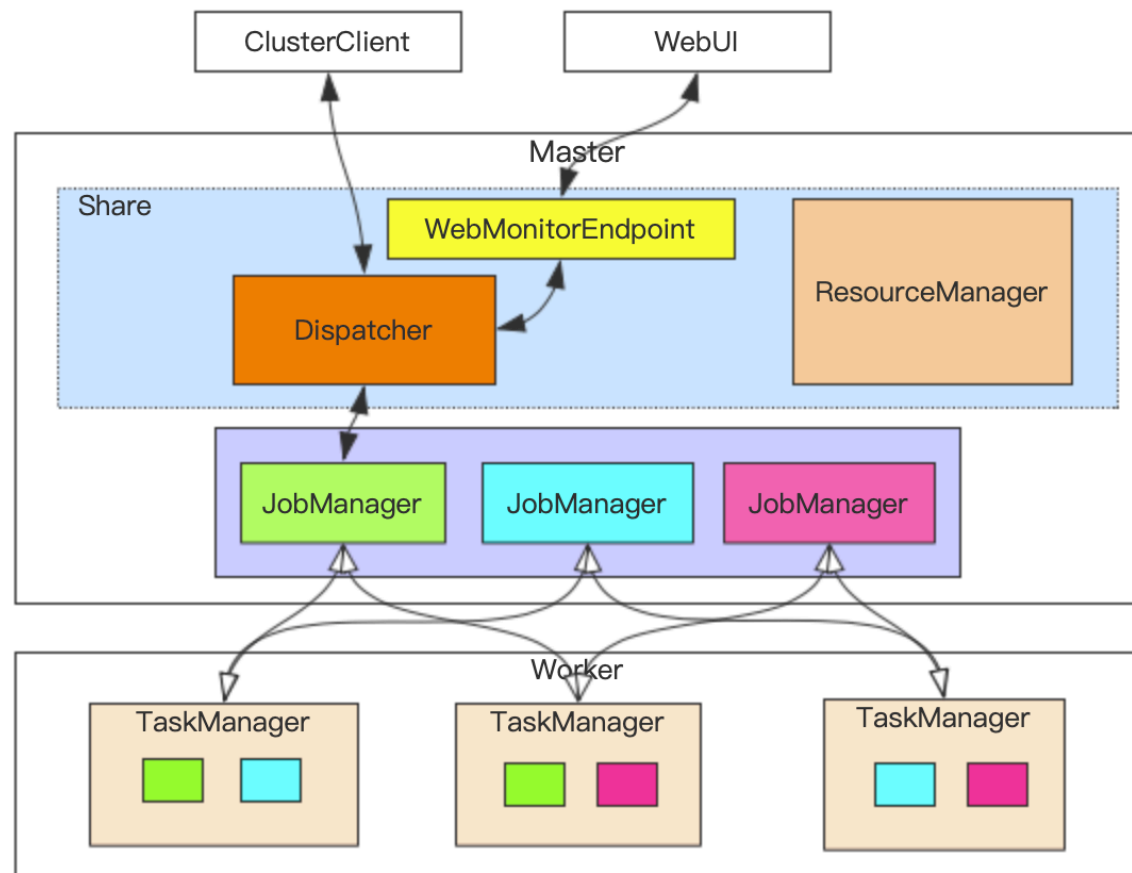
Runtime 核心组件

- Dispatcher
 - 集群 Job 的调度分发
 - 根据 JobGraph 启动 JobManager (JobMaster)
- ResourceManager
 - 集群层面资源管理
 - 适配不同的资源管理, eg Yarn, Kubernetes 等
 - 核心组件: SlotManager
- TaskManager
 - Slot 计算资源提供者
- JobManager
 - 负责管理一个具体的 Job
 - Scheduler
 - 调度和执行该 Job 的所有 Task
 - 发出 Slots 资源请求
 - Slot Pool



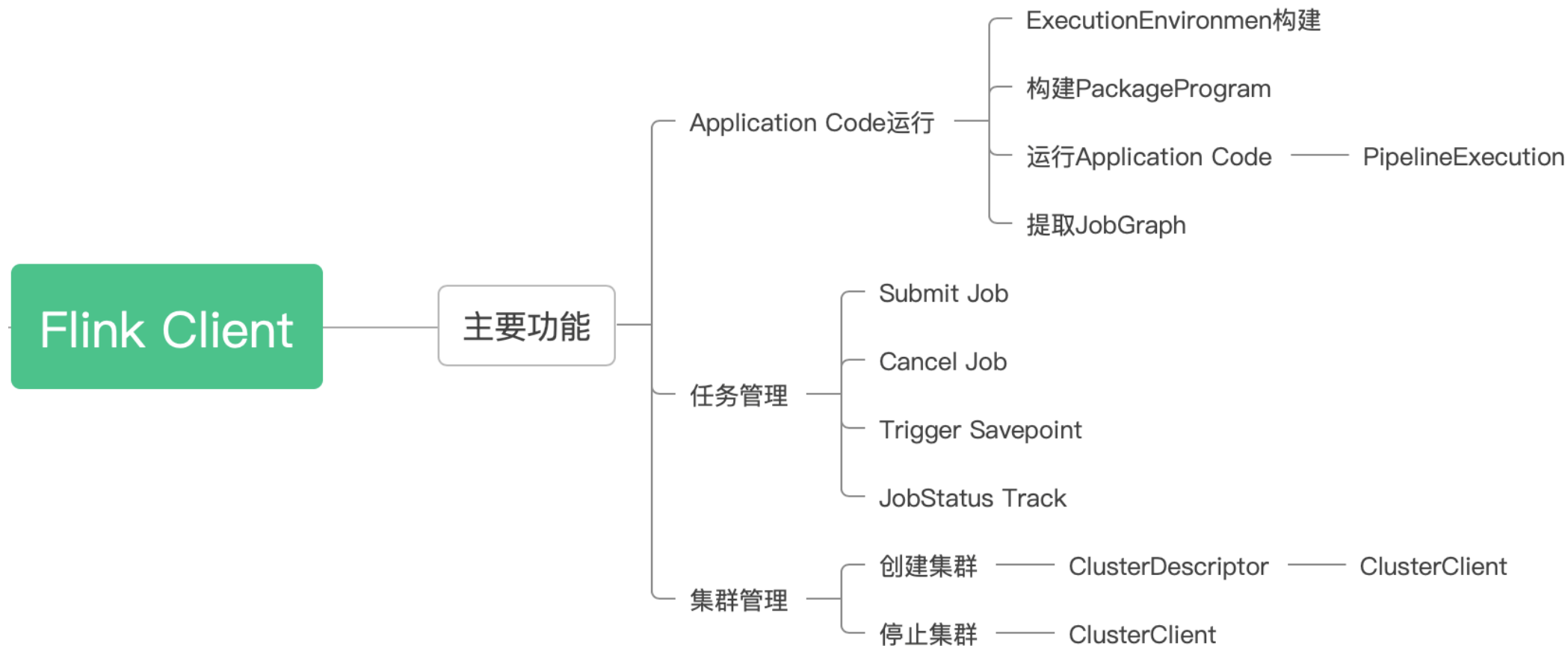
集群运行模式

- Session 模式
 - Runtime 集群组件共享
 - 资源复用
 - Runtime 中有多个 JobManager
- Per-Job 模式
 - Runtime 集群组件仅为单个 Job 服务
 - 资源相对独立
 - 不支持提交 JobGraph



Flink Client 实现原理

Flink Client 主要功能



Flink Client

主要组件

ContextEnvironment

执行环境

StreamExecutionEnvironment

ExecutionEnvironment

PipelineExecutor

Pipeline执行器

LocalExecutor — MiniCluster

AbstractJobClusterExecutor — YarnJobClusterExecutor

AbstractSessionClusterExecutor — RemoteExecutor

KubernetesSessionClusterExecutor

YarnSessionClusterExecutor

ClusterDescriptor

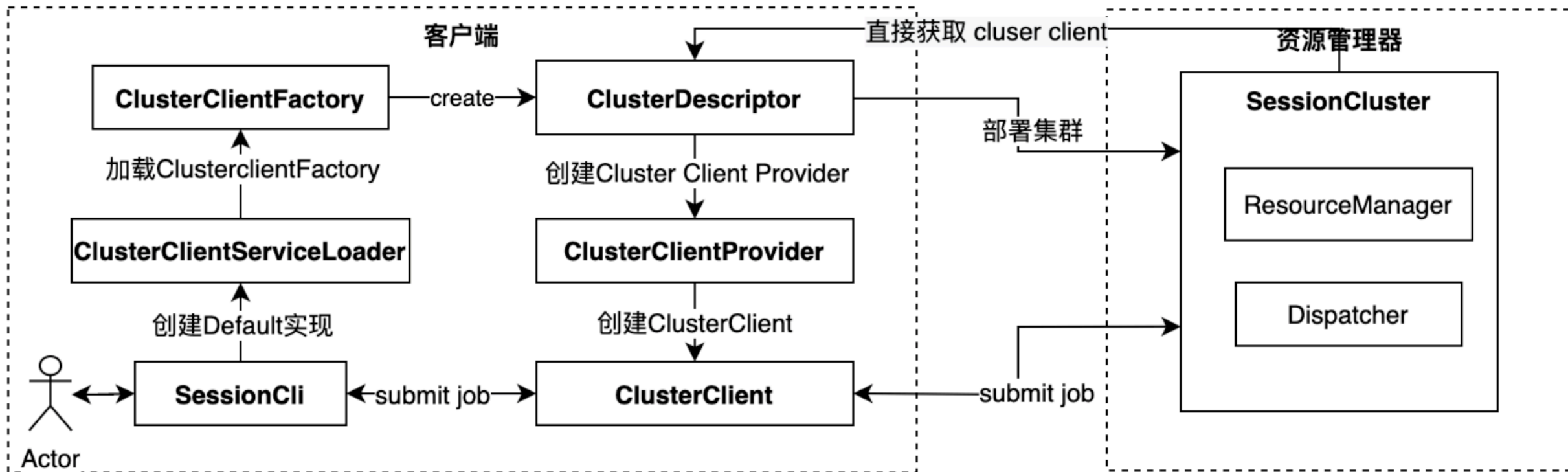
与具体资源管理器相连，创建对应集群

StandaloneClusterDescriptor

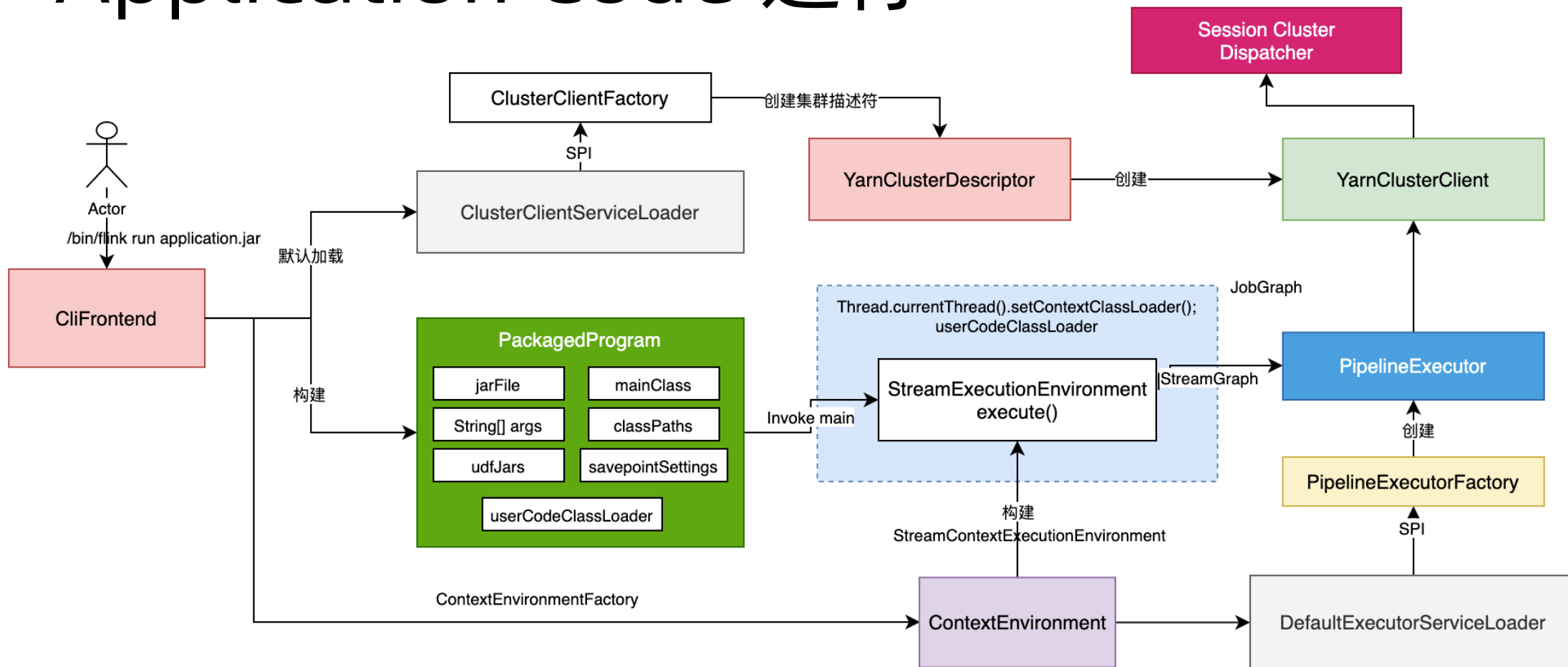
YarnClusterDescriptor

KubernetesClusterDescriptor

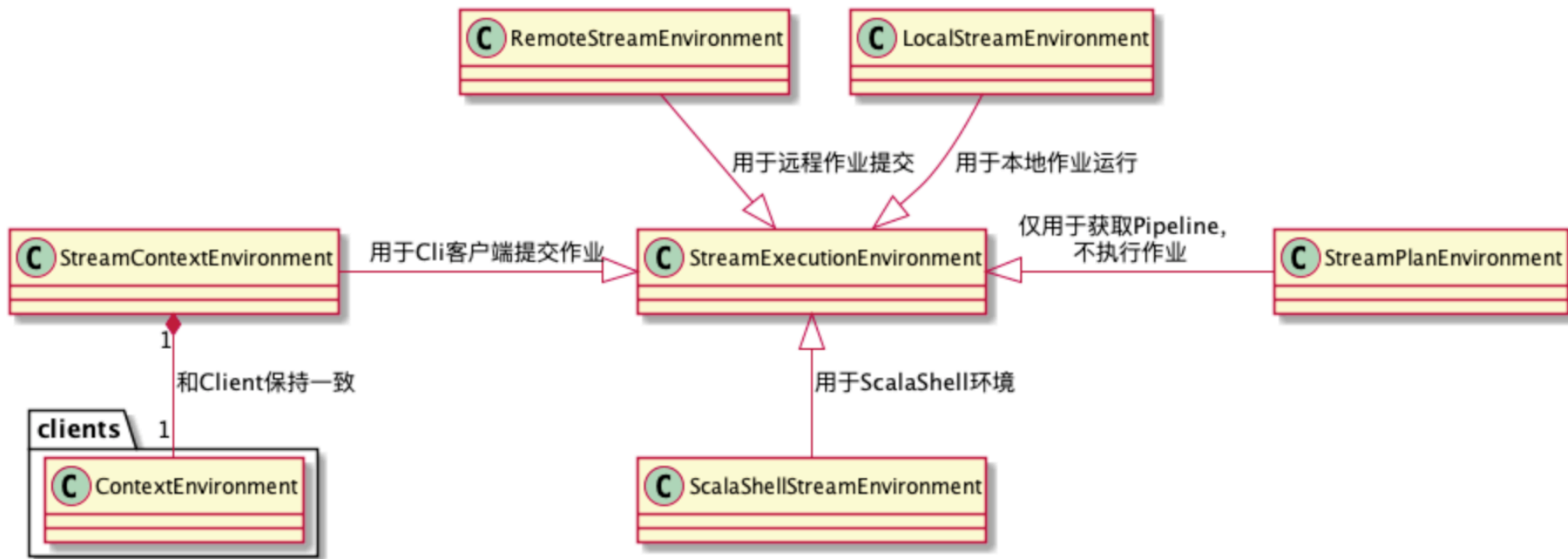
Session 集群创建流程



Application Code 运行



ExecutionEnvironment 分类



Talk is Cheap
Show me the code

ResourceManager

资源管理

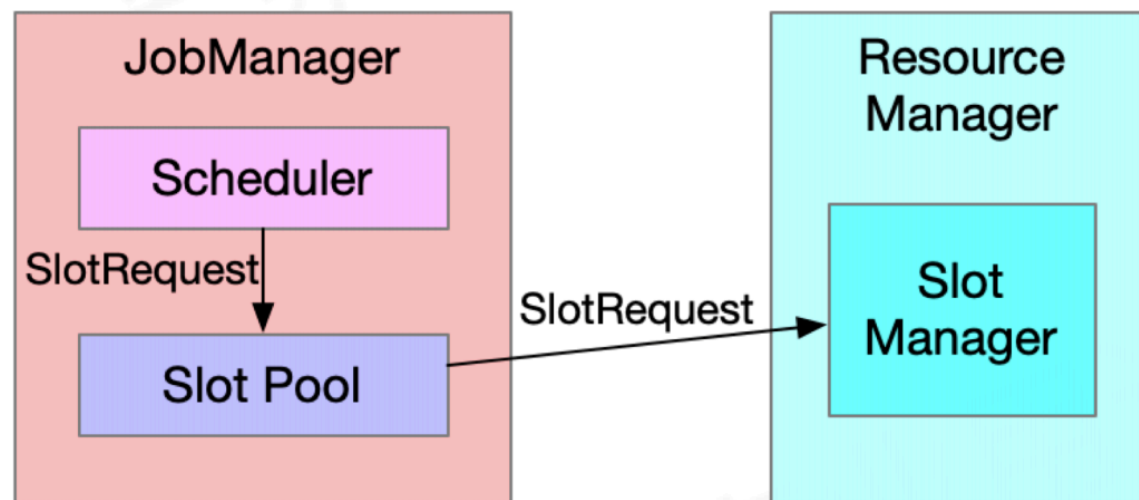
ResourceManager 资源管理

- 双层资源调度

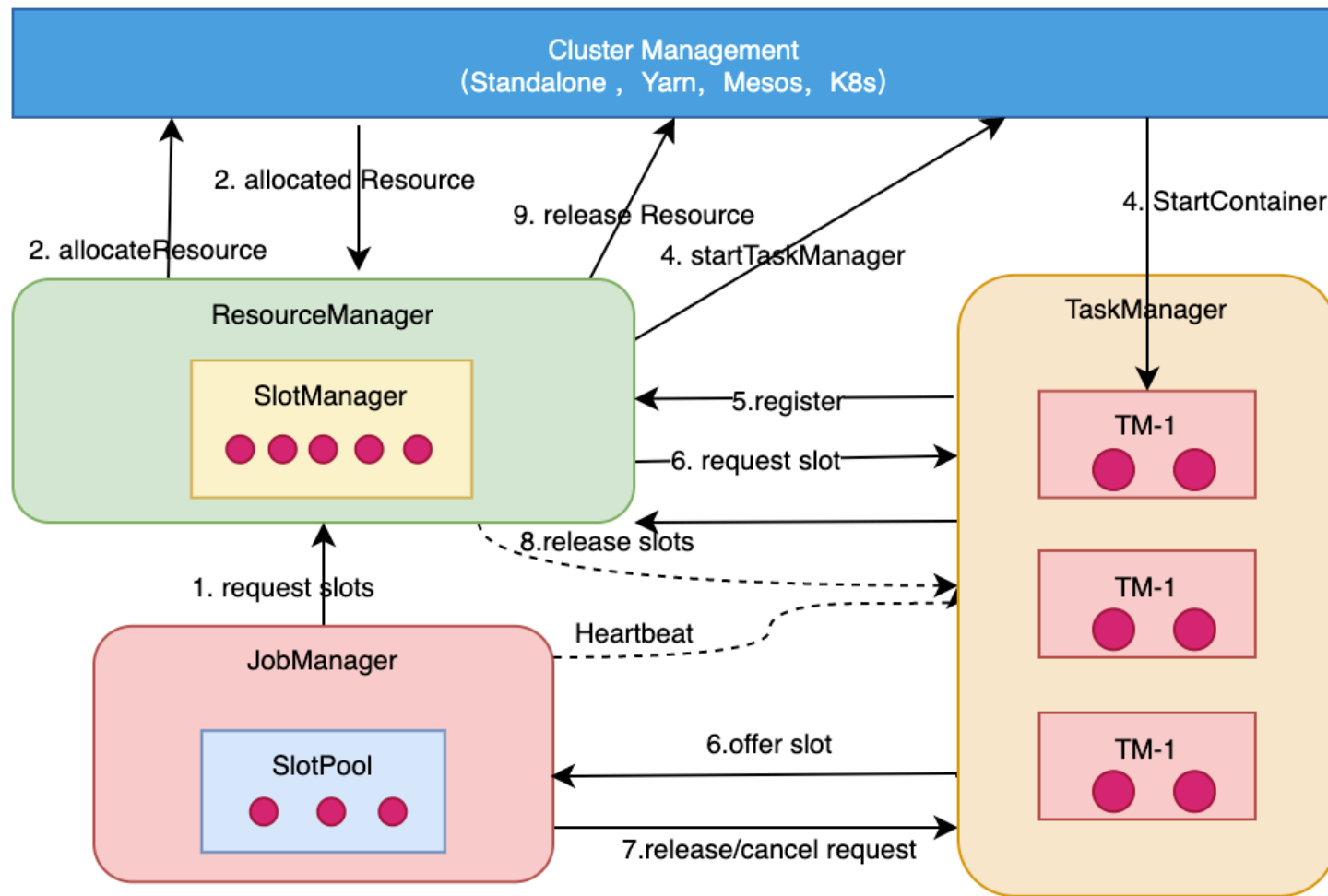
- Cluster->Job
 - SlotManager

- Job->Task

- SchedulerNG
- 单个Slot可以用于一个或多个Task执行
- 但相同的Task不能在一个Slot中运行

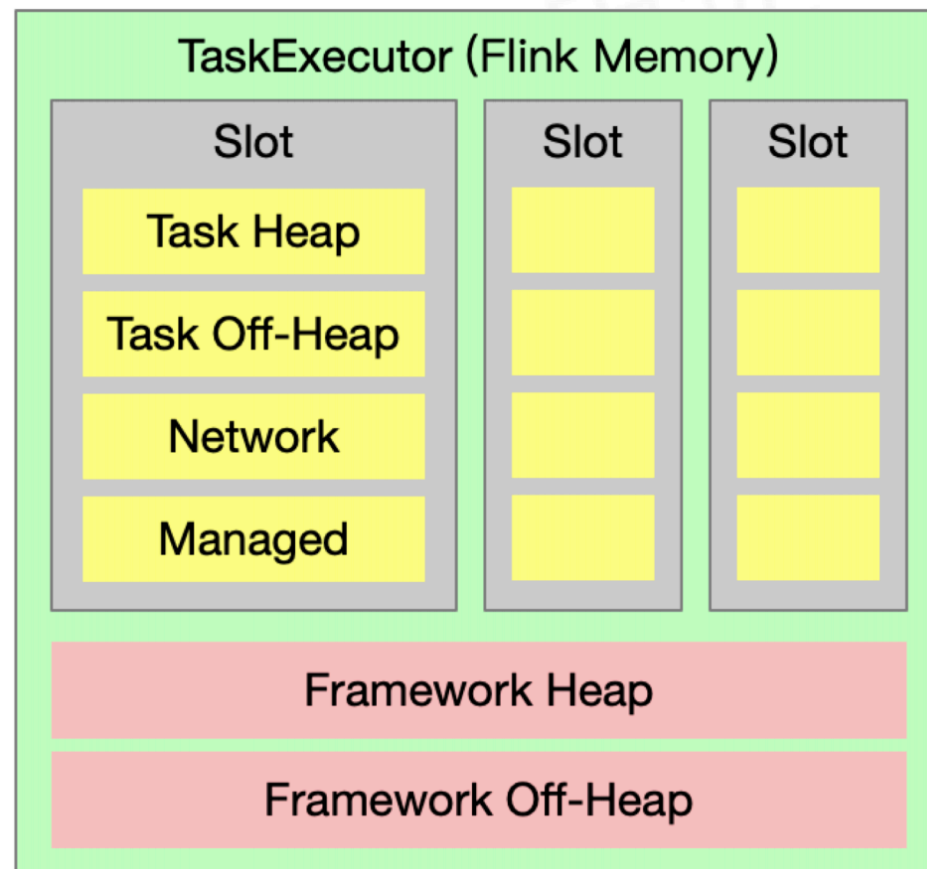


Slot 计算资源管理



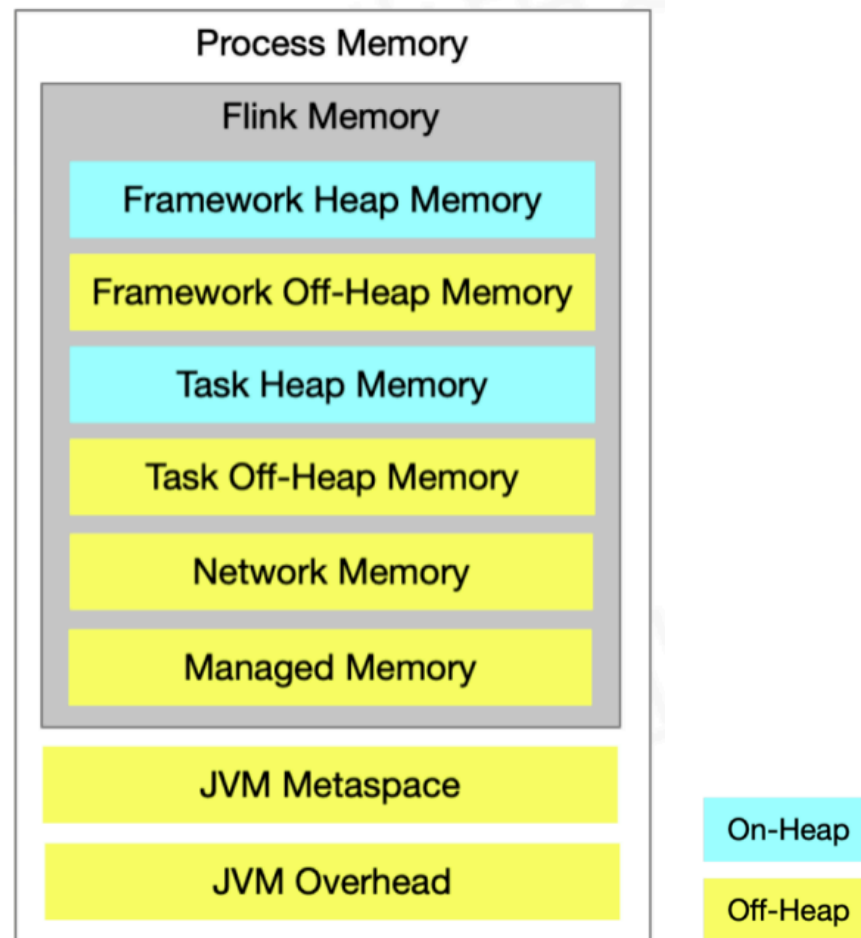
Slot 资源组成

- TM有固定数量的 Slot 资源
- Slot 数量由配置决定
- Slot 资源由 TM 资源及 Slot 数量决定
- 同一 TM 上的 Slot 之间无差别



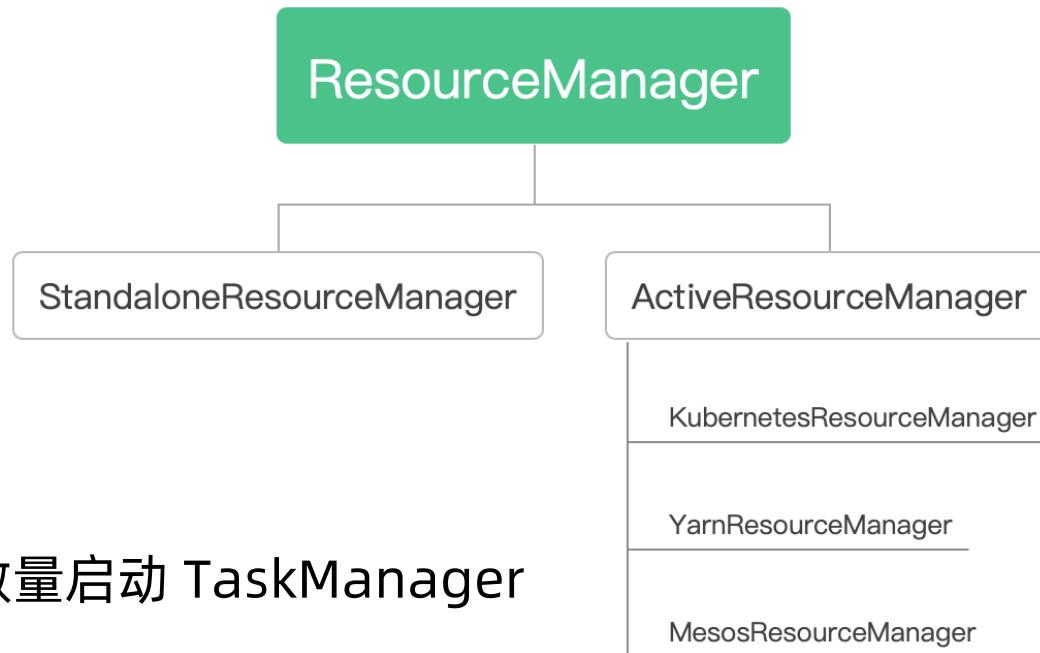
TaskManager 资源管理

- 资源类型
 - 内存
 - CPU
 - 其他拓展资源
 - GPU（待支持）

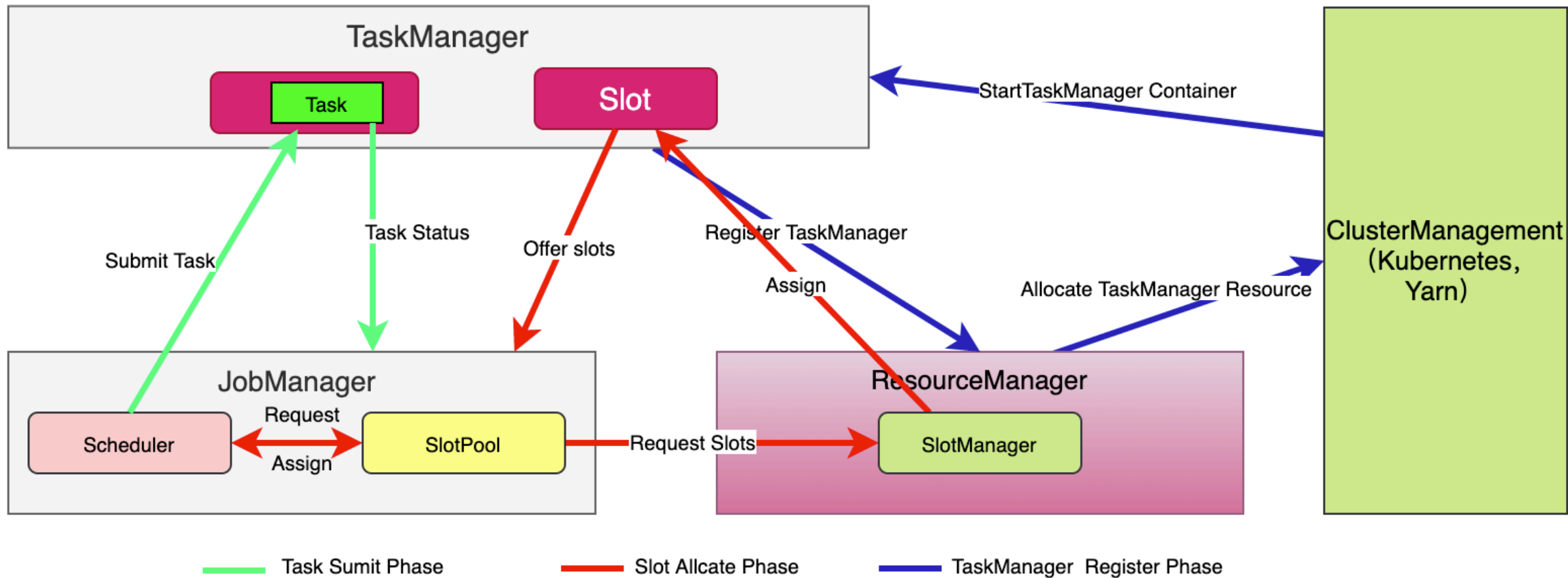


TaskManager 管理

- Standalone 部署模式
- ResourceManager 部署模式
 - Kubernetes, Yarn, Mesos
 - Slot 数量按需分配, 根据 Slot Request 请求数量启动 TaskManager
 - TaskManager 空闲一段时间后, 超时释放
 - On-Yarn 部署模式不再支持固定数量的 TaskManager

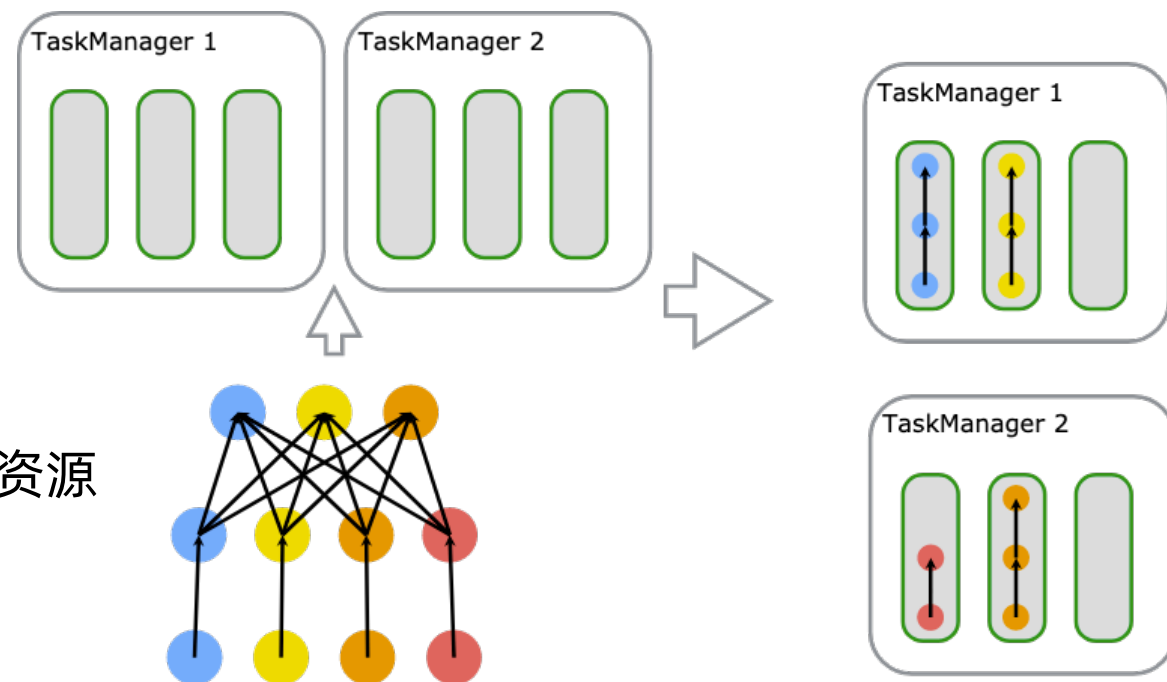


Job 资源调度



Task 调度执行

- SlotRequest:
 - Task + Slot \rightarrow Allocate TaskSlot
- Slot Sharing:
 - Slot Sharing Group 任务共享 Slot 计算资源
 - 单个 Slots 中相同任务只能有一个

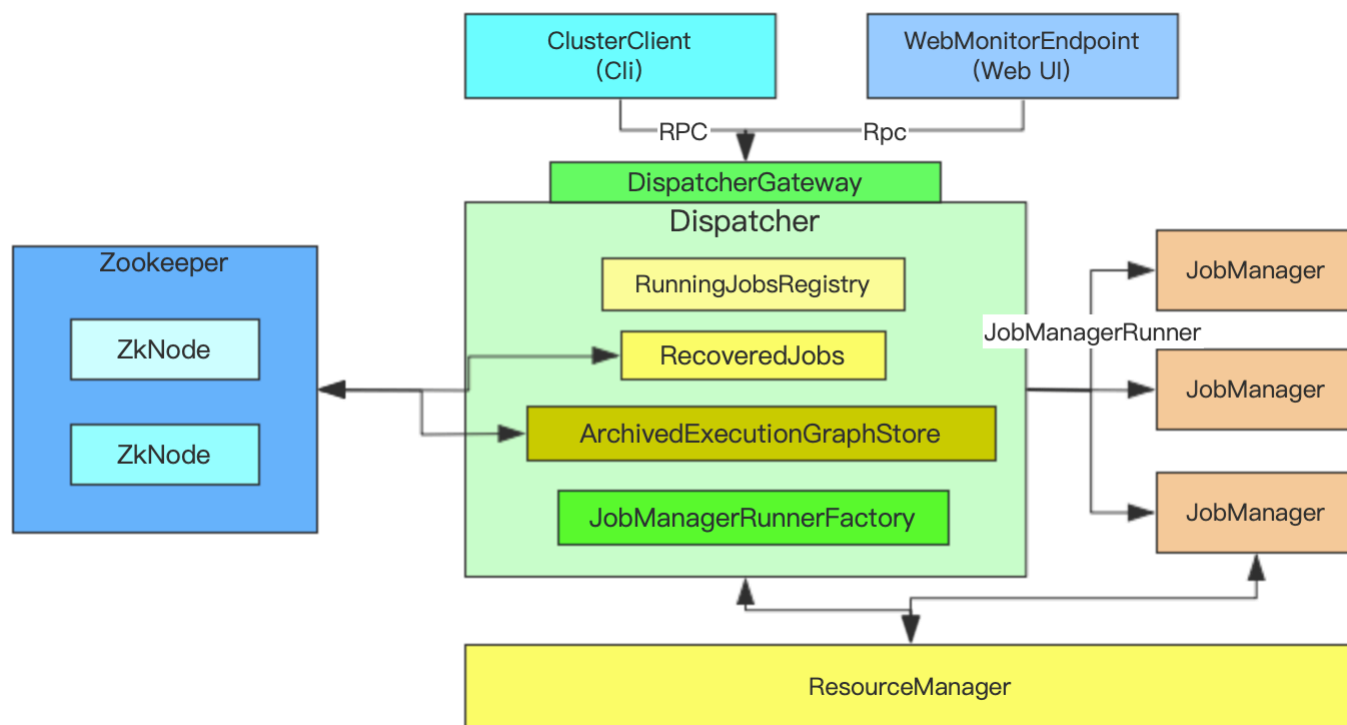


Talk is Cheap
Show me the code

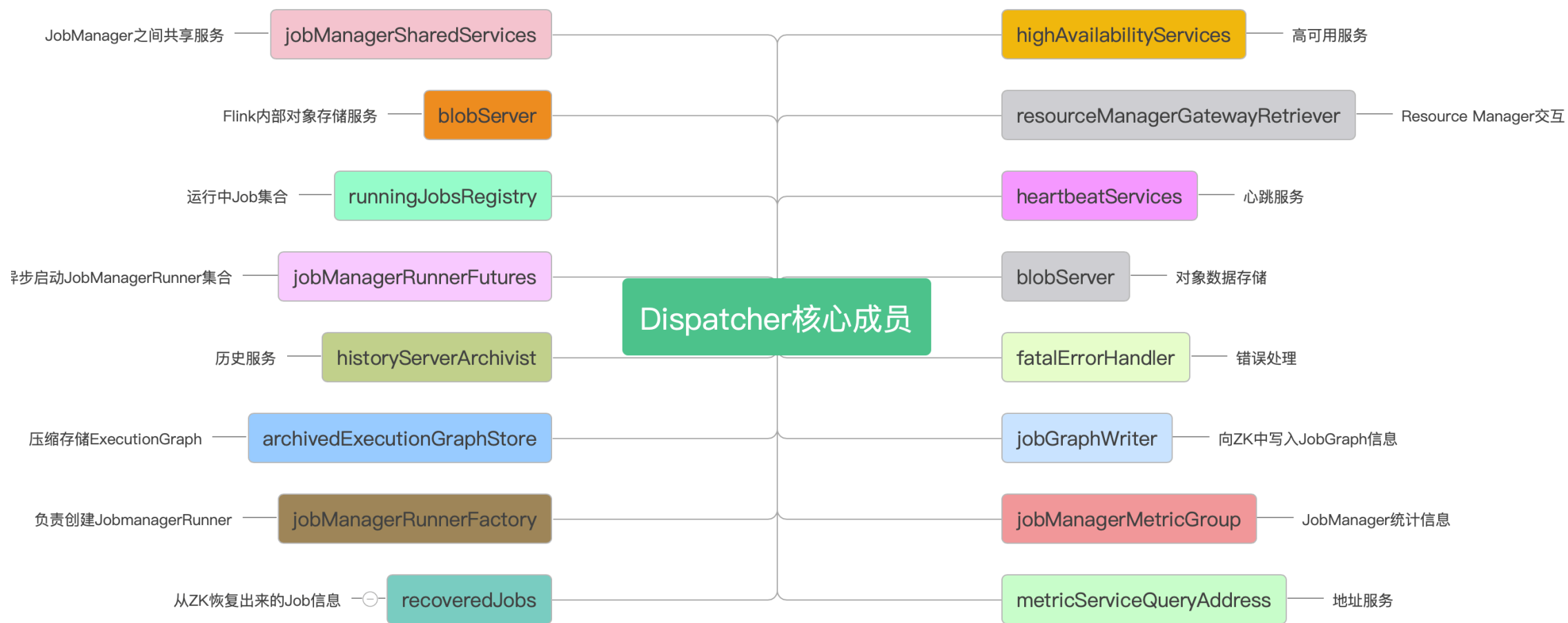
Dispatcher 任务分发器

Dispatcher 任务调度

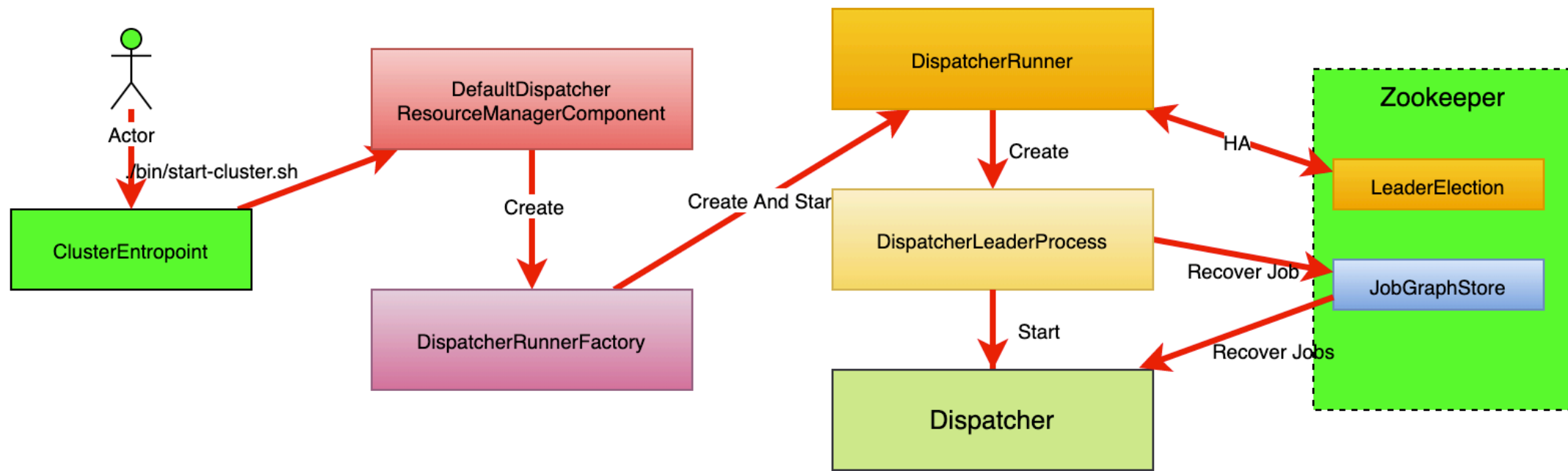
- 集群重要组件之一，主要负责 JobGraph 的接收
- 根据 JobGraph 启动 JobManager
- RpcEndpoint 服务
- 通过 DispatcherGateway Rpc 对外提供服务
- 从 ZK 中恢复 JobGraph (HA 模式)
- 保存整个集群的 Job 运行状态



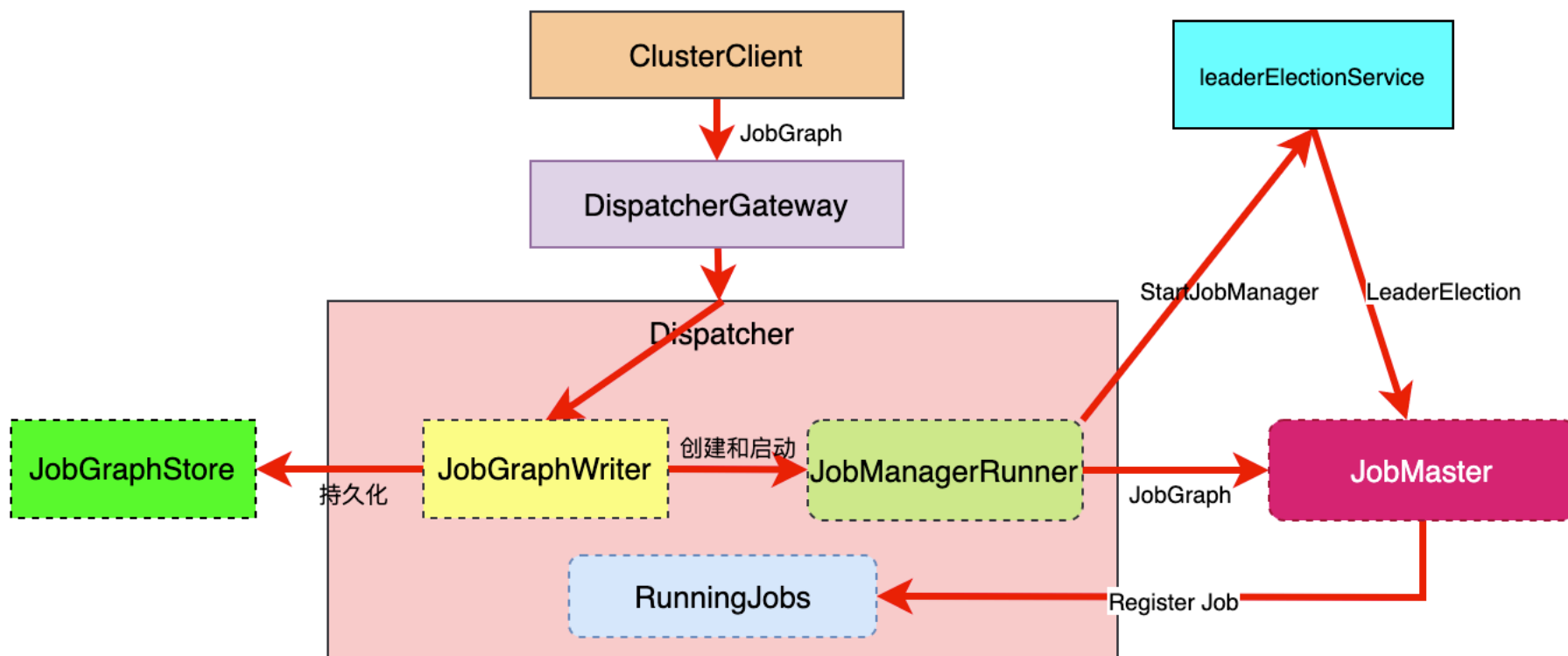
Dispatcher 组件核心成员



Dispatcher 启动流程



Dispatcher 接收任务



Talk is Cheap
Show me the code



扫码试看/订阅

《Flink核心技术与实战》视频课程