

算法服务平台

调度器的功能和设计思想

何龙龙

2020-04-22

调度的定义

- 资源角度

调度是为了高效利用**有限资源**将任务分配给资源的方法

1. 单处理器的多任务处理
2. 凌晨定时跑批任务

特点：资源多指计算资源，且可灵活切割。任务多指批处理，基于数据属性的聚合任务。

调度的定义

- 任务角度

调度是为了**完成各种任务**用于分配任务所需资源的方法。

1. 负载均衡（F1, LVS, 网关, 分布式应用等）
2. 供应链的调配, 物流的调度

特点：资源包括数据、物料、被上层所依赖的抽象等， 稳定变化小。任务是偏业务功能的， 基于功能的支持任务

调度器

- 定义

进行调度工作的程序叫做调度器

- 分类

- 抢占调度器：面向资源，对任务的调度
- 协同调度器：面向任务，对资源的调度

I/O调度器，作业调度器，网络调度器

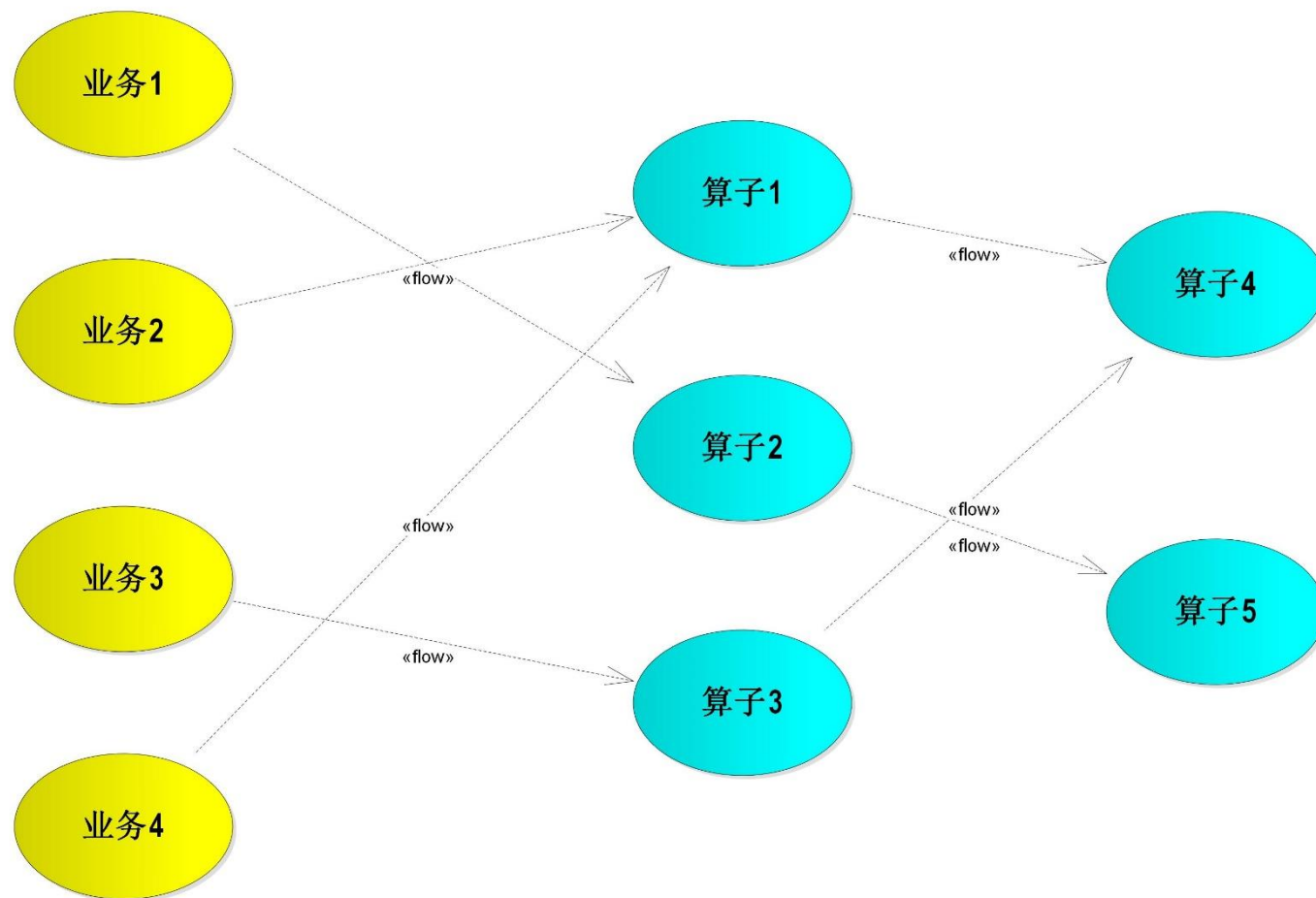
调度器

- 区别
 - 调度任务还是调度资源
 - 任务有优先级还是资源有优先级
 - 任务稳定还是资源稳定
- 目标
 - 吞吐率最大化
 - 响应时间最小化
 - 最低延迟
 - 最大化公平

算法服务平台调度器

- 业务场景

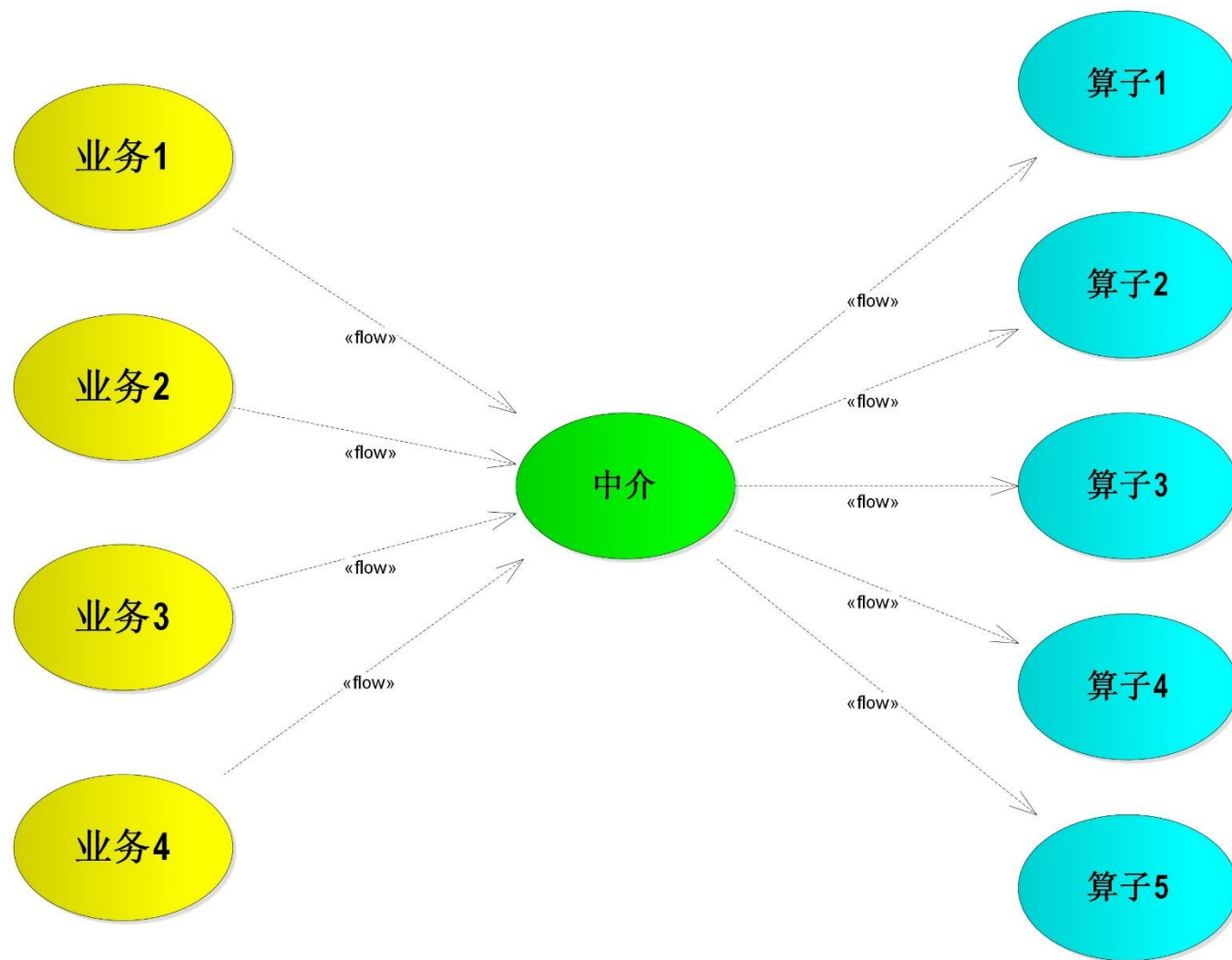
- 算子响应业务慢（2周到数月）
- 业务与算子紧耦合
- 调用关系混乱
- 业务接入算子难度高



算法服务平台调度器

- 设计思想

- 中介模式解决了紧耦合和调用关系混乱的问题
- 桥接模式抽象统一接口，将实现与抽象分离
- 通过抽象工厂实现依赖注入
- 算法平台对调用进行控制反转
- DAG调度器满足依赖调用



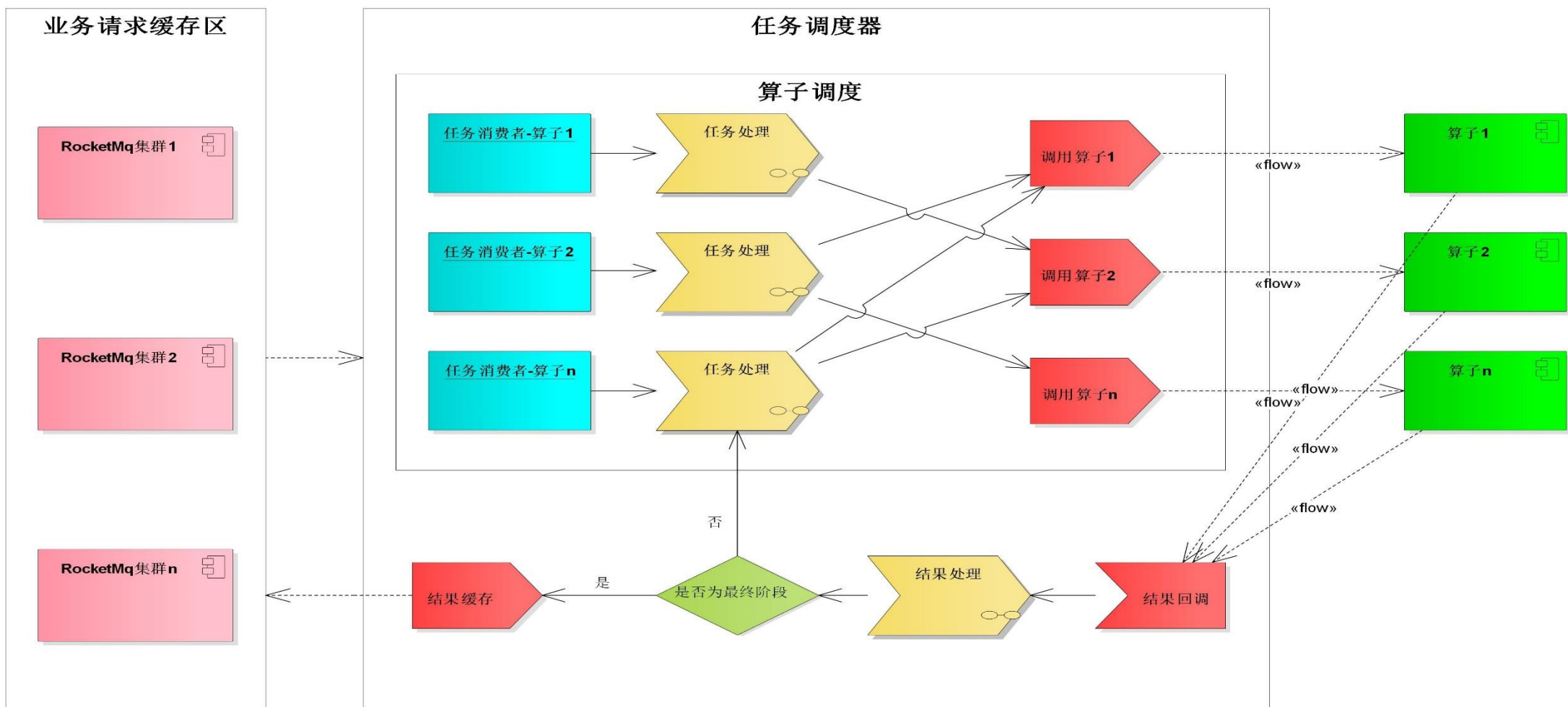
调度器的最终功能

AP-SCHEDULER 算法服务平台用于调度资源给算子，具有跨平台、可编排、高可用、一致性、异步并行、动态扩展、实时监控、协同调度、dag调度。

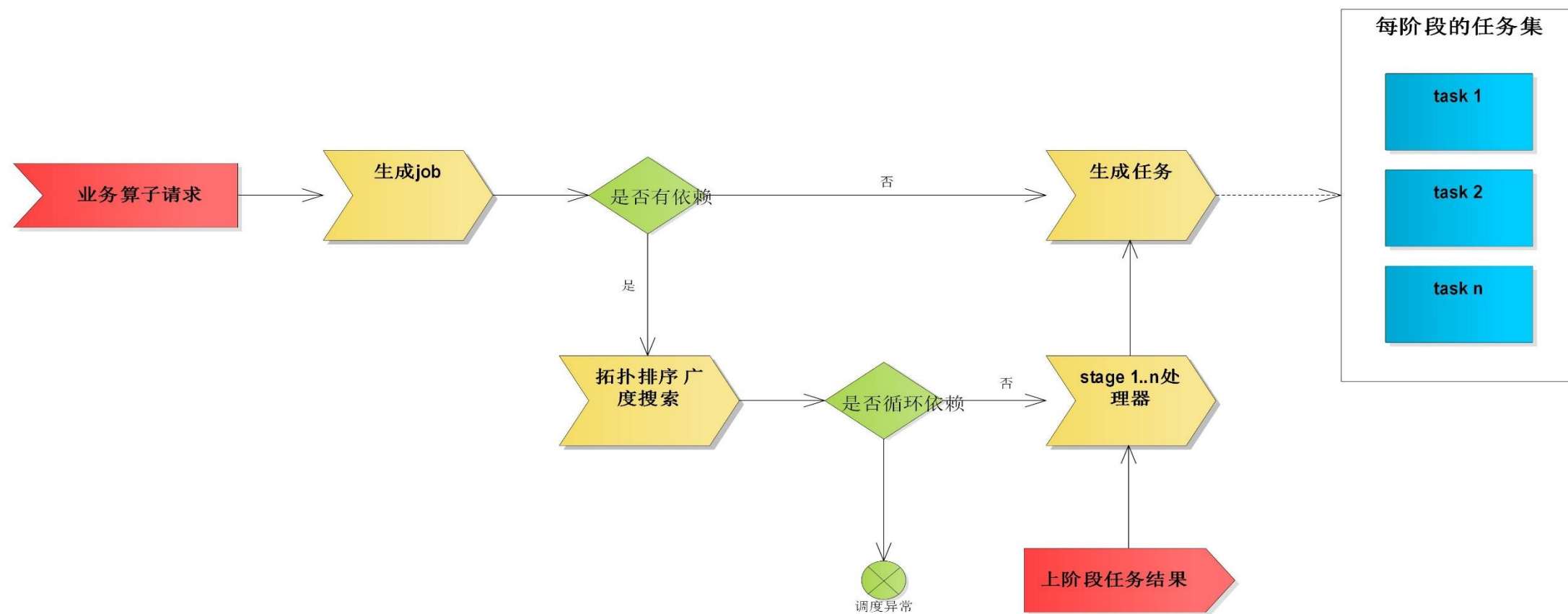
- 基于抽象接口面向算子对数据资源的调度
- 及时响应（毫秒级）--在线流式打标能力
- 资源的优先级
- 算子能力的复用
 - 组合
 - 继承

核心目标：持续提升算子对业务的整体响应性

调度器设计



任务处理



算子能力复用

- 抽象接口提供了灵活组合的能力
- DAG调度器，支撑依赖调用
- 算子基于已有能力的创新

