

百度大规模在离线混部系统演进

张慕华

百度-基础架构部-资深研发工程师

极客邦科技 会议推荐2019

5月

QCon 北京

全球软件开发大会

大会: 5月6-8日
培训: 5月9-10日

QCon 广州

全球软件开发大会

培训: 5月25-26日
大会: 5月27-28日

6月

GTLC
GLOBAL
TECH LEADERSHIP
CONFERENCE

上海

技术领导力峰会

时间: 6月14-15日

GMTC 北京

全球大前端技术大会

大会: 6月20-21日
培训: 6月22-23日

7月

ArchSummit 深圳

全球架构师峰会

大会: 7月12-13日
培训: 7月14-15日

10月

QCon 上海

全球软件开发大会

大会: 10月17-19日
培训: 10月20-21日

11月

GMTC 深圳

全球大前端技术大会

大会: 11月8-9日
培训: 11月10-11日

AiCon 北京

全球人工智能与机器学习大会

大会: 11月21-22日
培训: 11月23-24日

12月

ArchSummit 北京

全球架构师峰会

大会: 12月6-7日
培训: 12月8-9日

自我介绍

团队

- 2013年加入百度
- 基础架构部IaaS团队

工作方向

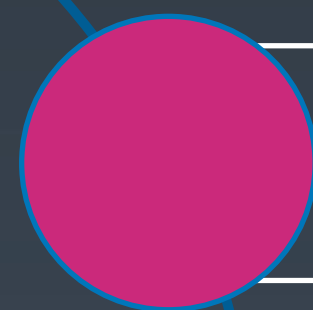
- 2013-2015: MapReduce、大数据分布式计算
- 2015-Now: 集群调度、容器、资源隔离、离线PaaS

负责系统

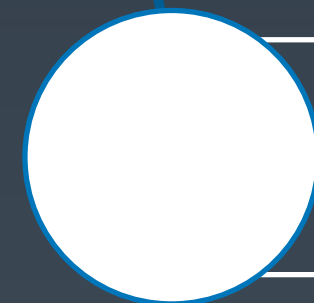
- ClusterOS: Matrix
- 离线调度系统: Normandy
- 机器管理: Ultron



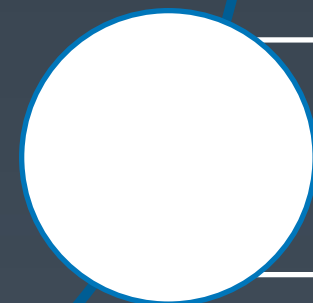
目录



一、背景：为什么要做混部



二、方案：混部架构和核心技术



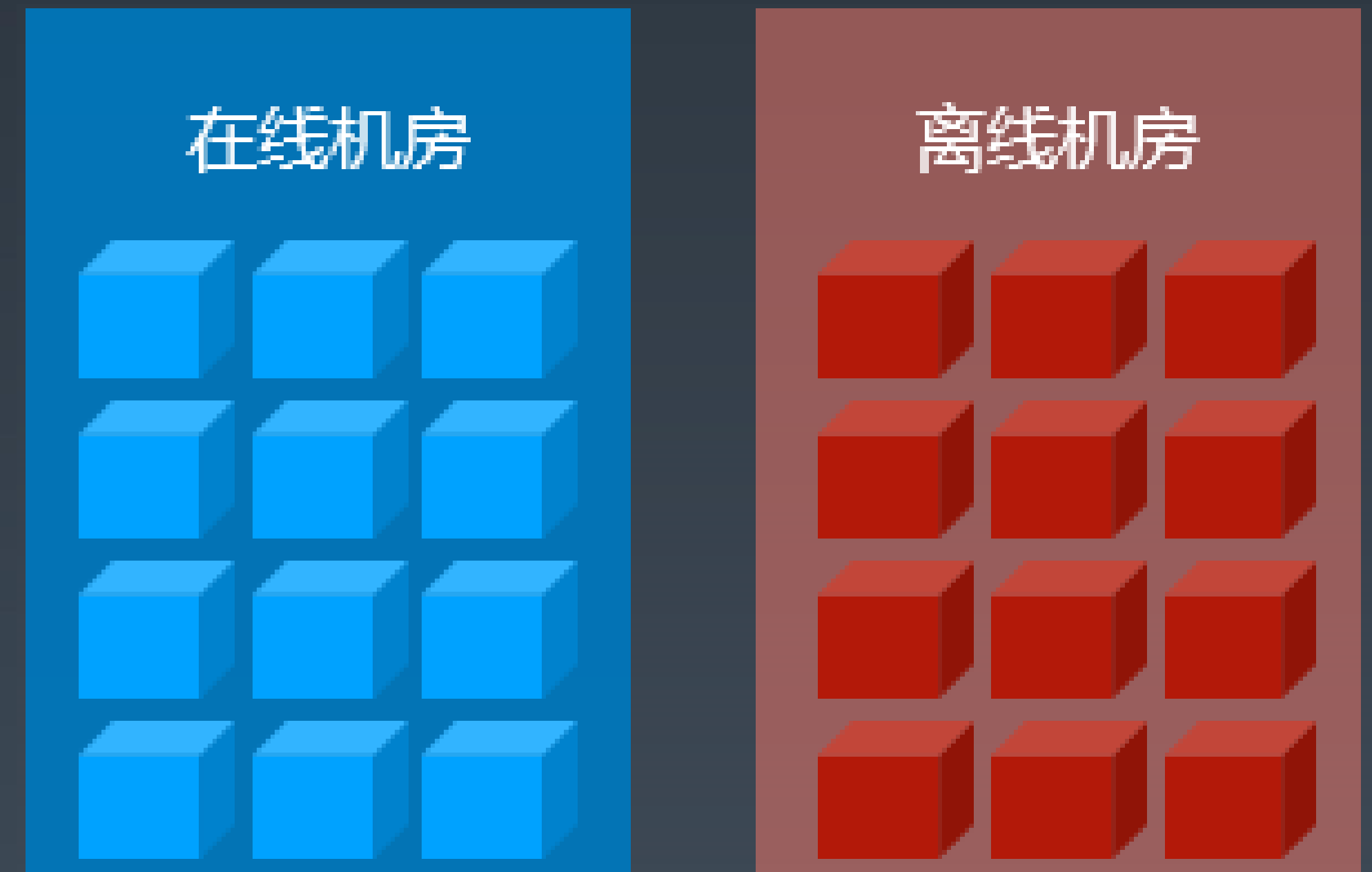
三、展望：收益和未来展望

从成本说起..

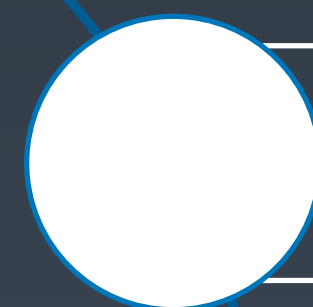
- 公司从什么阶段开始关注成本
- **规模化、平稳发展期**
- 公司的成本构成
- **研发、运营、服务器**
- 提升资源利用率是降低服务器成本的重要手段！！

资源利用率现状

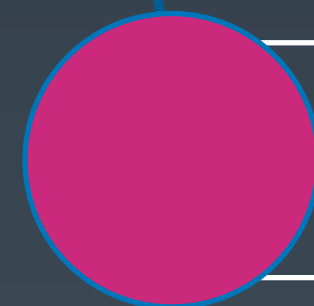
- IDC规划：
- 在线机房、离线机房分离规划
- 资源利用率的矛盾：
- 1) 离线利用率高，资源不足
- 2) 在线利用率低，峰值高但不够弹性
- 现状分析：
- 1) 在线天然利用率不足，只有离线、在线统一规划集群才能充分利用资源
- 2) 通过实施在离线混部：提升利用率、减少离线、在线服务器采购成本



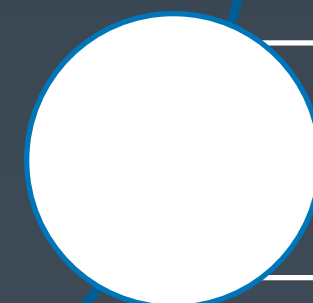
目录



一、背景：为什么要做混部



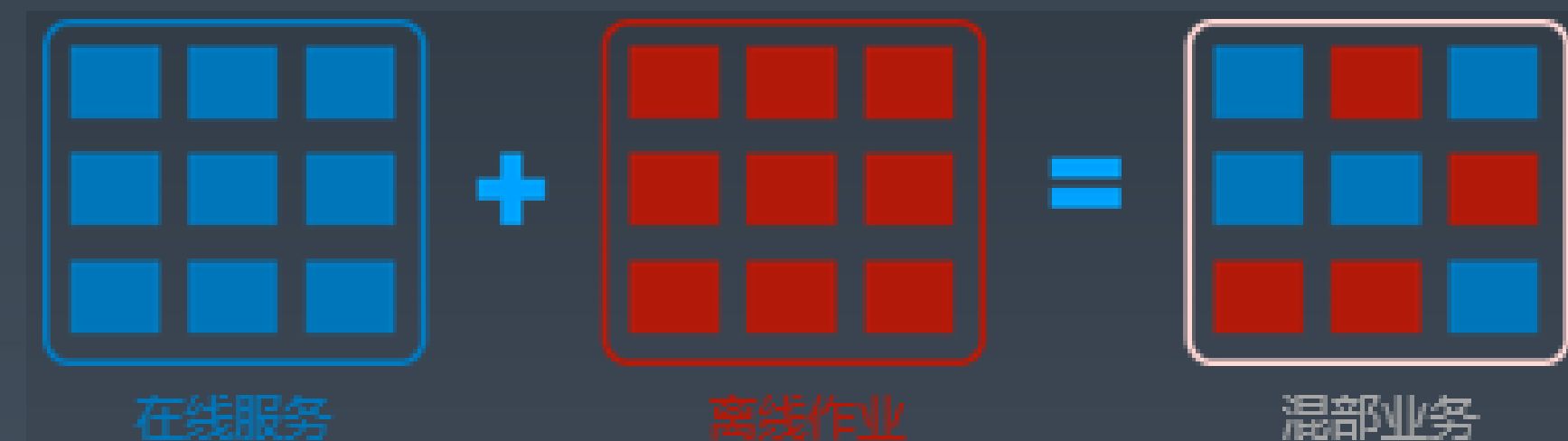
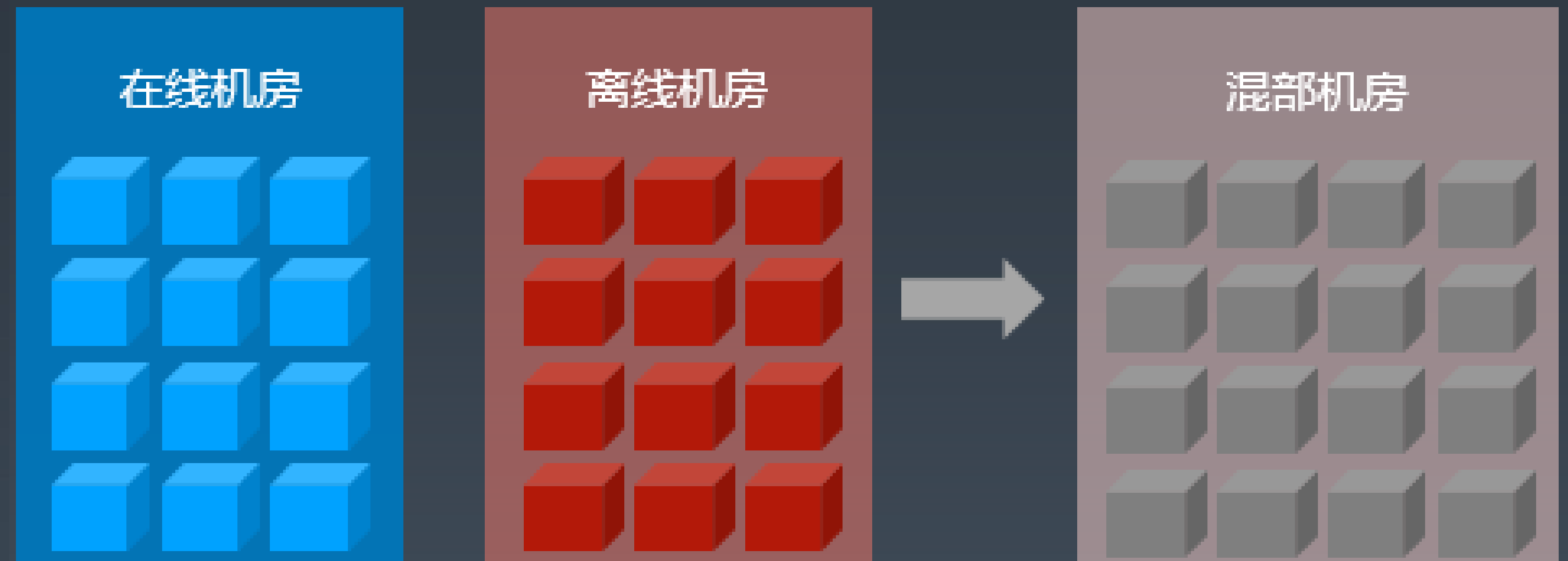
二、方案：混部架构和核心技术



三、展望：收益和未来展望

在离线混部思路

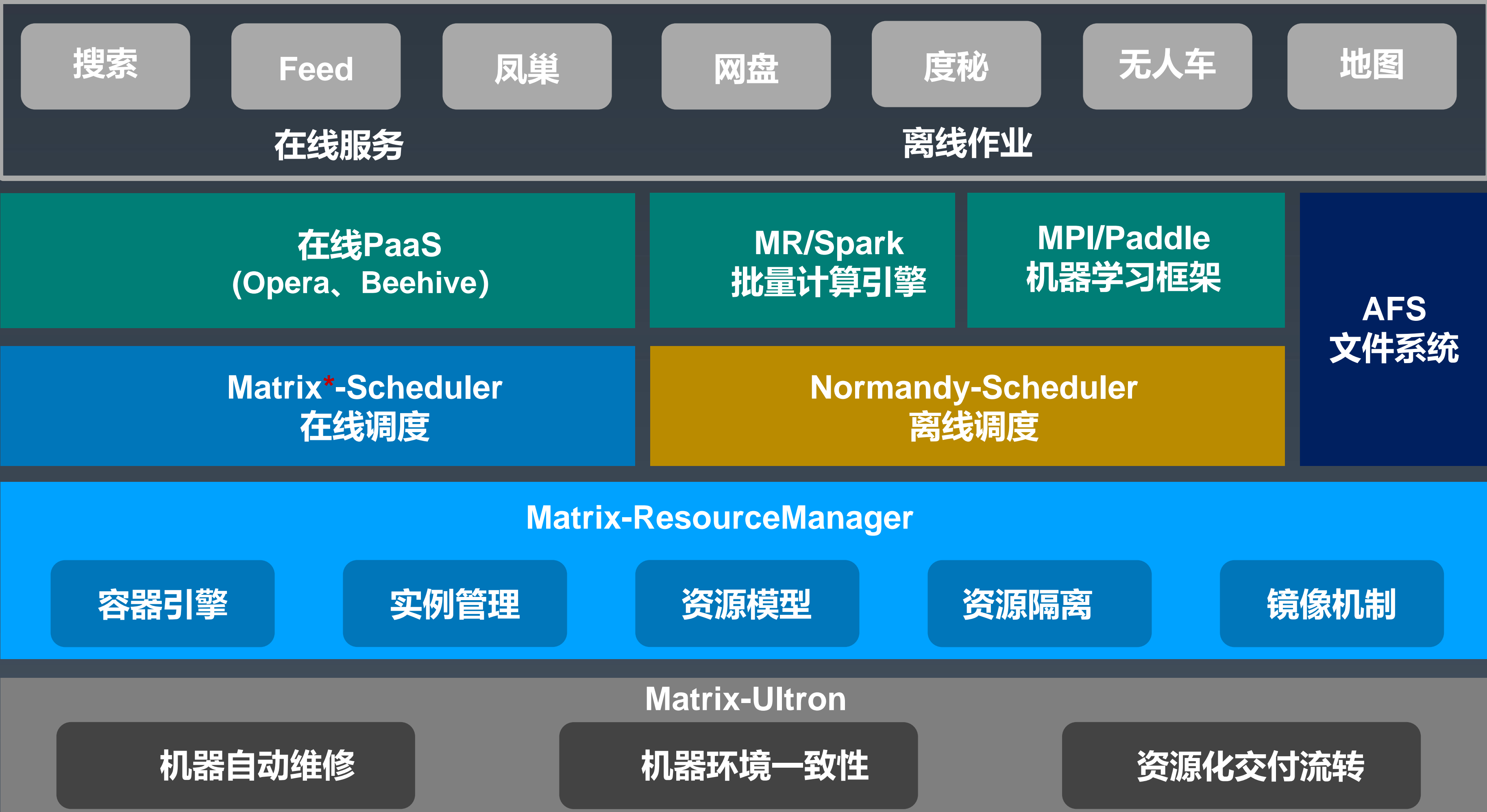
- 定义
 - 1) 机房统一规划
 - 2) 业务混合运行
- 难点
 - 1) 保证在线不受到影响
 - 2) 保证离线运行良好
 - 3) 挖潜更多的混部资源



百度混部演进历程

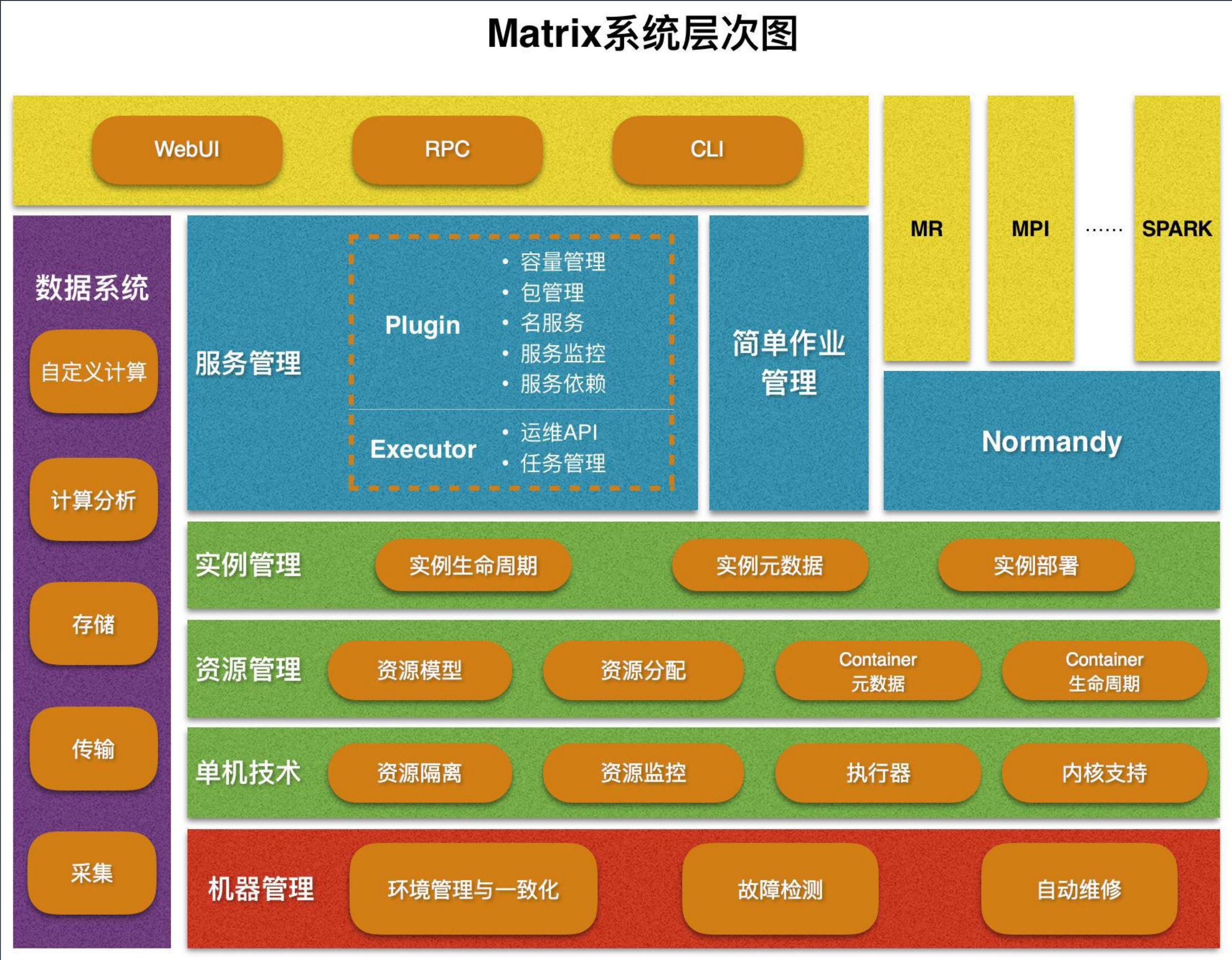


混部技术栈



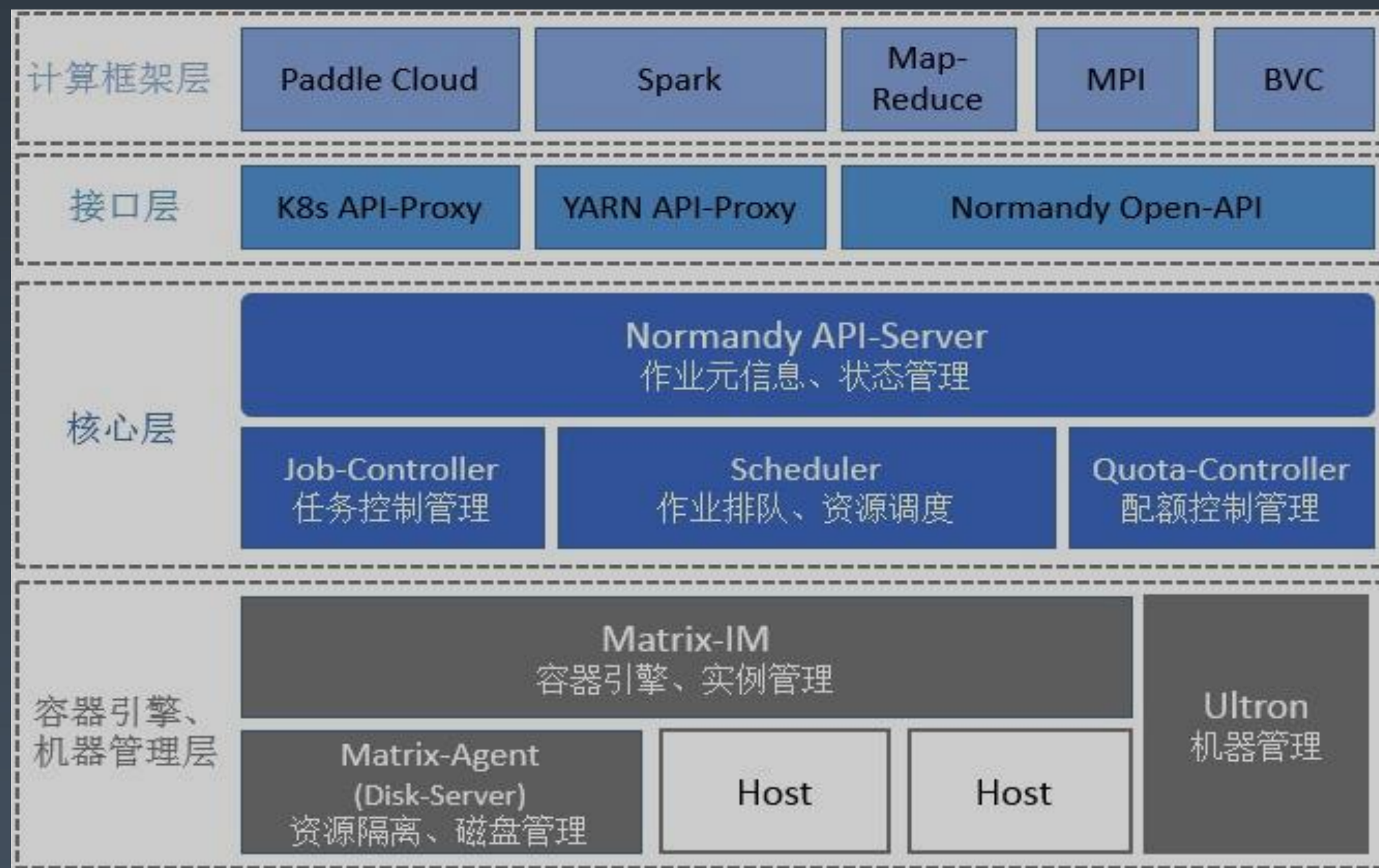
Matrix

- 定位：百度的Borg
 - 1) 2012年启动
 - 2) 容器化基础设施
 - 3) 混部隔离的基础设施
- 设计：
 - 1) 容器引擎
 - 2) 镜像管理
 - 3) 实例管理
 - 4) 资源调度
 - 5) 隔离技术



Normandy

- 定位：离线调度系统
- 1) 离线计算混部
- 2) 混部的离线资源调度
- 3) Plugins on Matrix
- 设计：
- 1) 任务管理
- 2) 作业排队
- 3) 资源调度
- 4) 等等



混部的理论基础

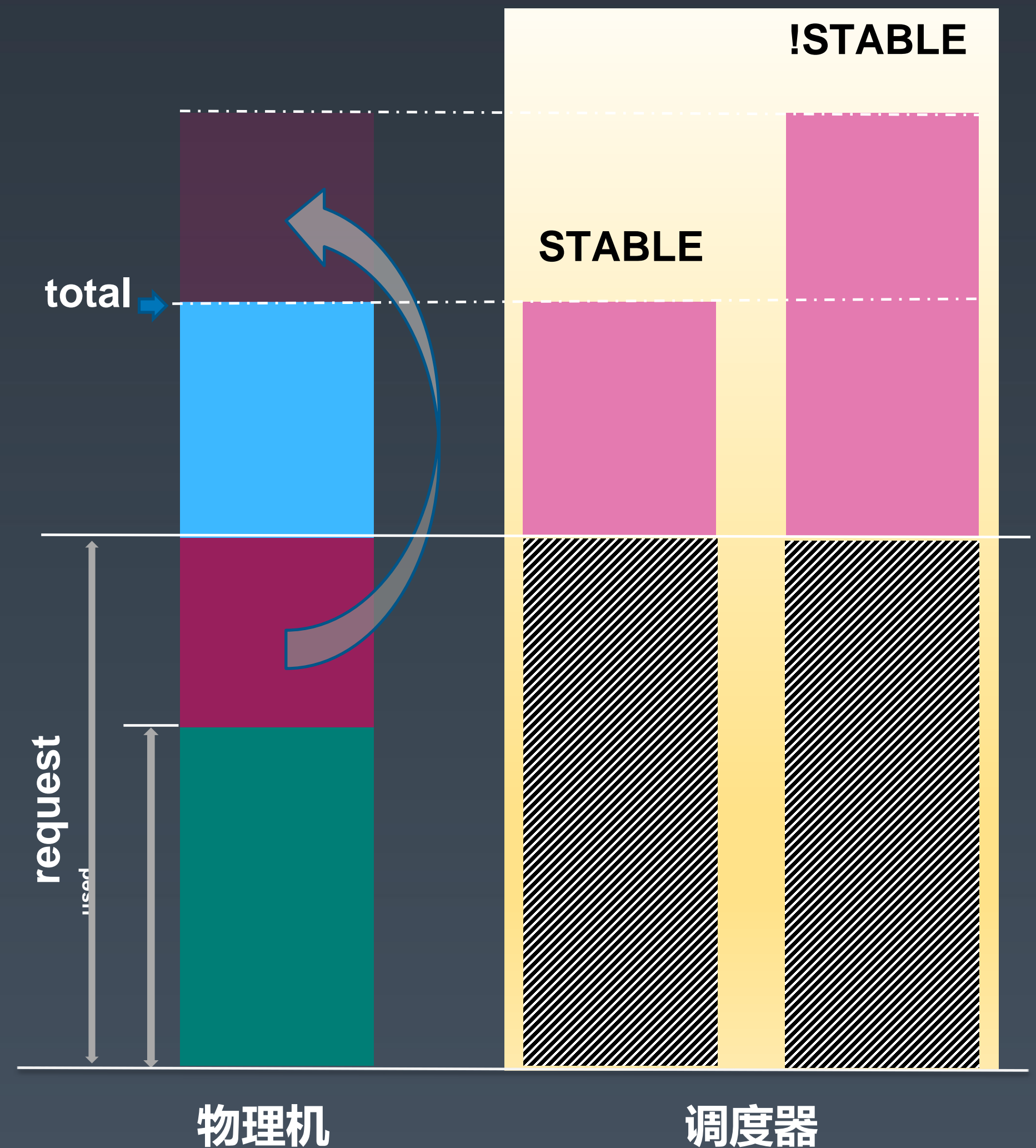
- 混部原则：一套在线、离线在单机运行的规则规范。
- 单机资源模型：优先级和超发模型

优先级

- $STABLE > NORMAL > BESTEFFORT$
- 单机：可否抢占
- 集群：是否有Quota

超发

- 用户行为： $request - used > 0$
- 超发： $total' = total + \sum(request - used)$
- 约束： $\sum(STABLE) < total$



混部的核心思路

单机隔离

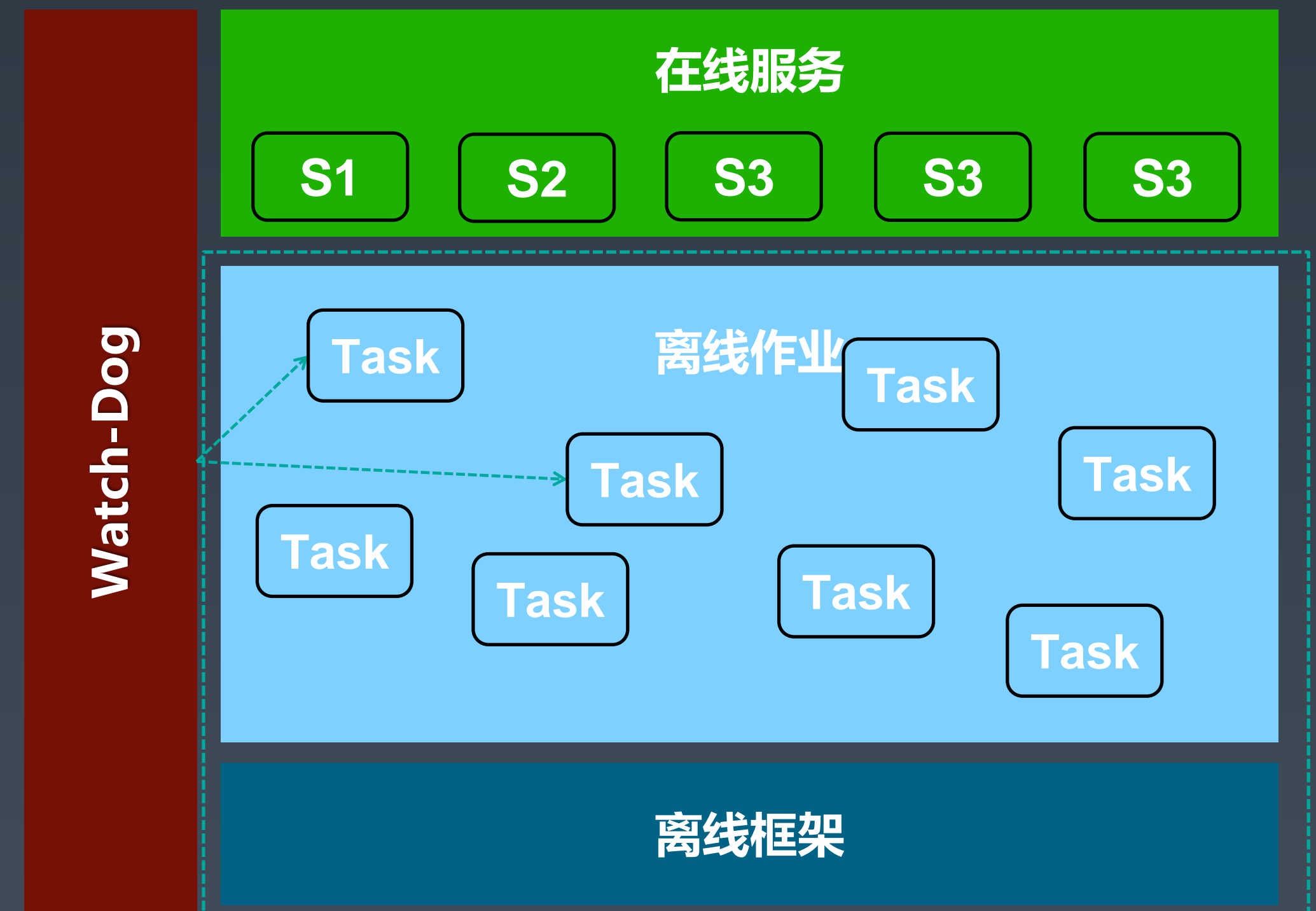
响应及时，快速避让！

集群调度

响应慢，但是有规划的调度

隔离：保证在线不受影响

- 基本思路：
 - 1) 内核和用户态管控
 - 2) 离线“大框”模型
- CPU：大框绑核
- 内存：大框硬限
- IO：隔离磁盘、计算框架优化
- 网络：QoS、Transkeeper
- 兜底策略：单机SLA



隔离：更加精细的管控策略

- **CPU:**

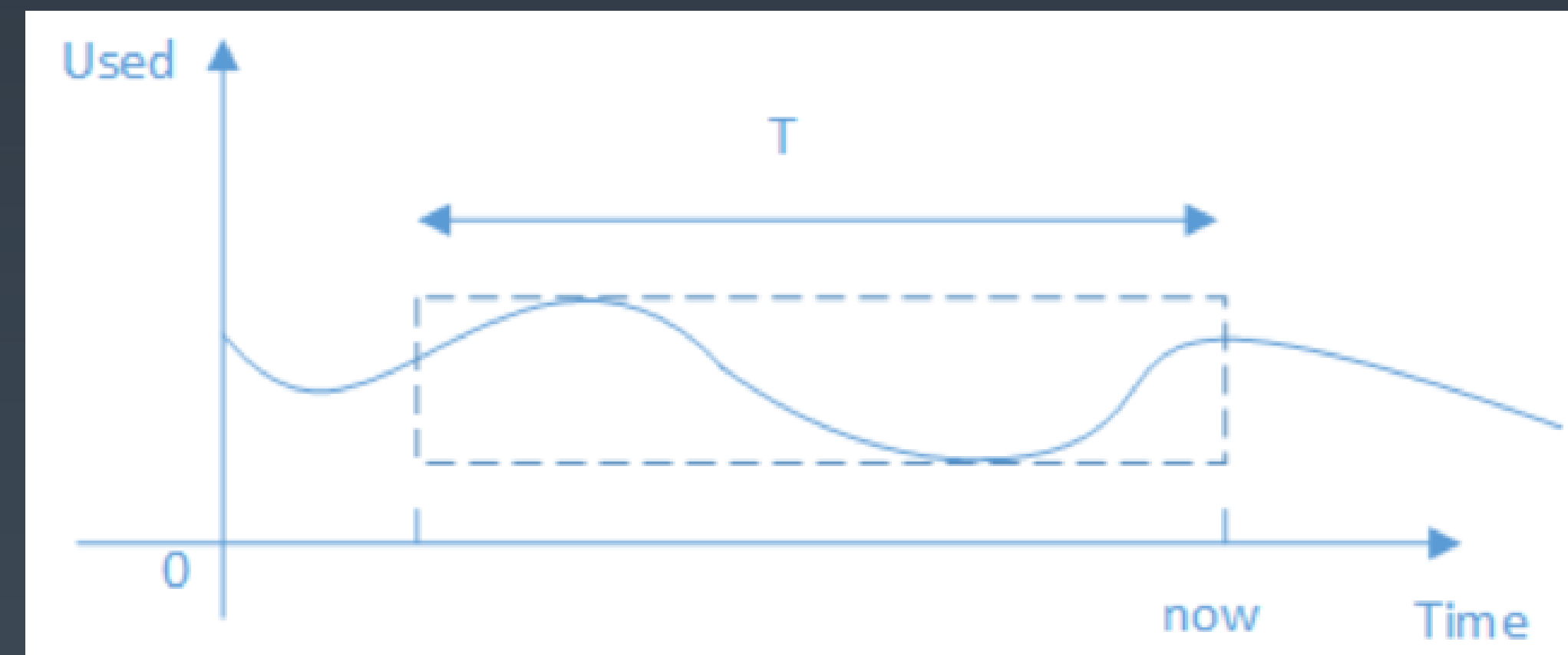
- 1) 离线大框绑核：根据单机reclaim资源量；离线内共享；
- 2) core-aware：快退避，慢启动，避免HT干扰；
- 3) HT干扰规避：自动从迁移离线作业，避免和在线服务干扰；
- 4) L3-cache：CAT隔离；
- 5) CPI干扰抑制：检测、干扰源识别、避让；

- **内存:**

- 1) 离线大框嵌套：根据单机reclaim资源量；离线嵌套内存Cgroup；
- 2) 动态内存伸缩

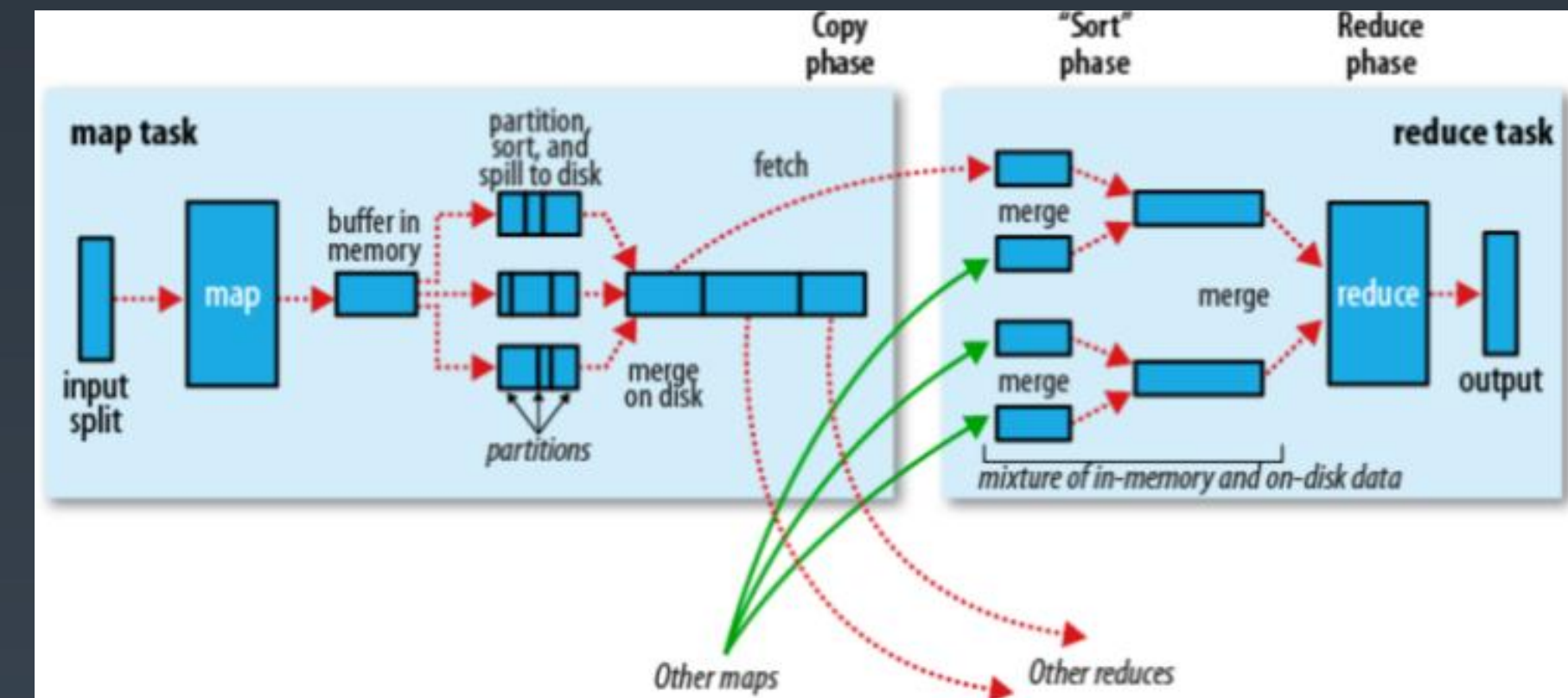
避让：保证离线运行良好

- 精细的单机避让
 - 1) 可压缩资源优先压缩;
 - 2) 不可压缩资源避让足量即可;
- 调度器优化
 - 1) $used = \text{Max}(\text{used in last } T)$
 - 2) 建立学习反馈闭环



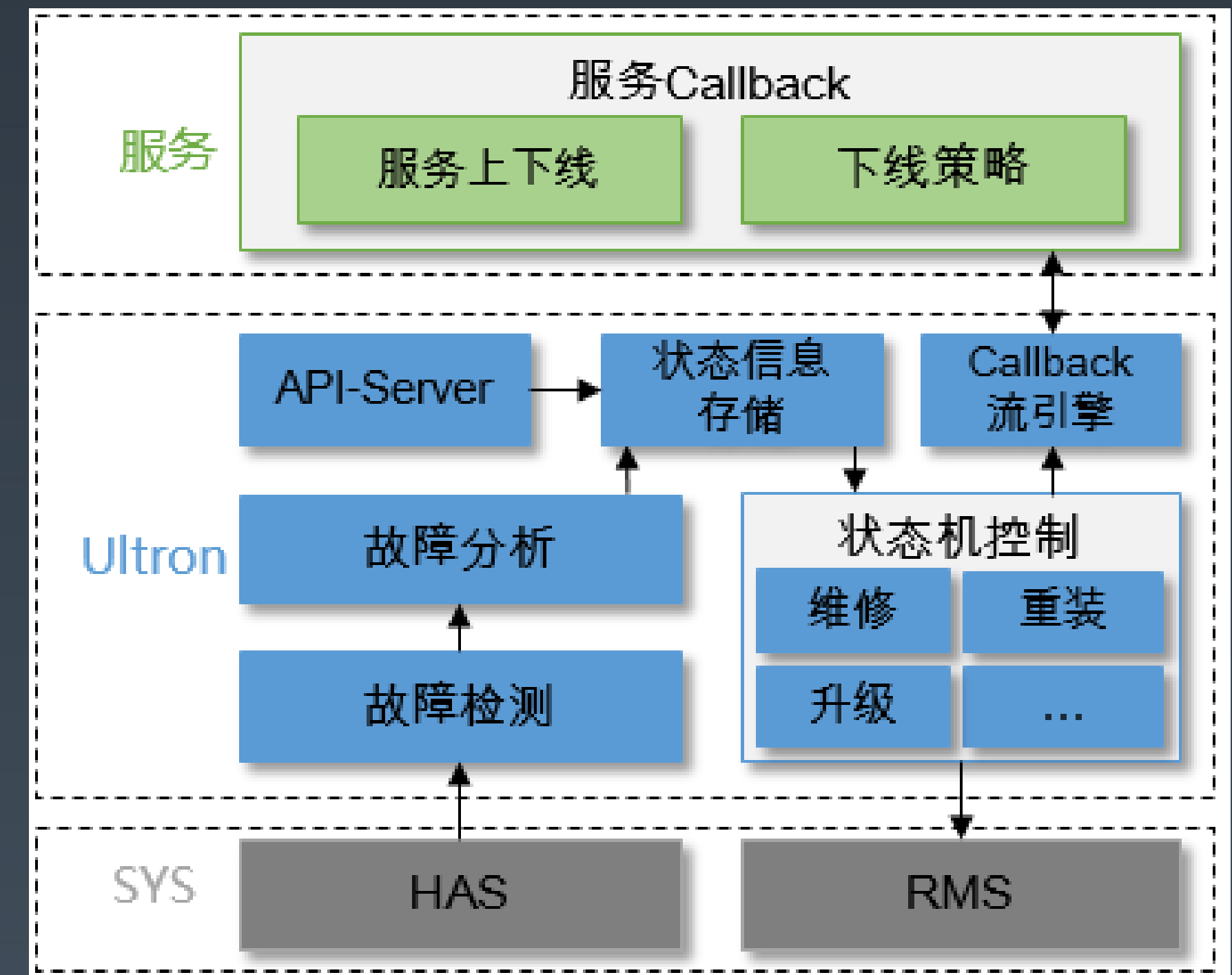
计算框架(MR/Spark)的优化

- 避免本地磁盘IO：
 - 1) Shuffle由Pull变成Push
 - 2) 通过FS（全内存）交换
- 减少重算：
 - 1) Map端节点挂掉重算
 - 2) 做好Backup

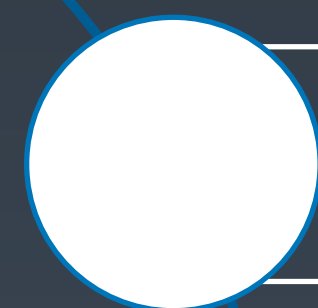


支撑系统建设

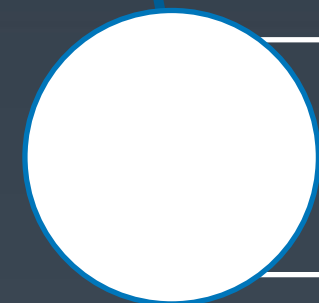
- 技术债：
 - 1) 混部框架自动化入退场
 - 2) 机器环境一致性：网卡多队列、内核参数等等
 - 3) 资源化交付：资源订单至Quota交付，业务看不见机器



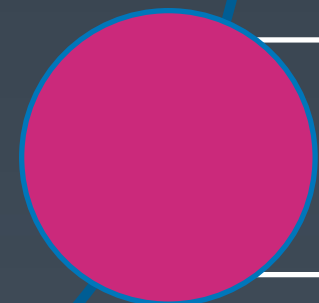
目录



一、背景：为什么要做混部



二、方案：混部架构和核心技术



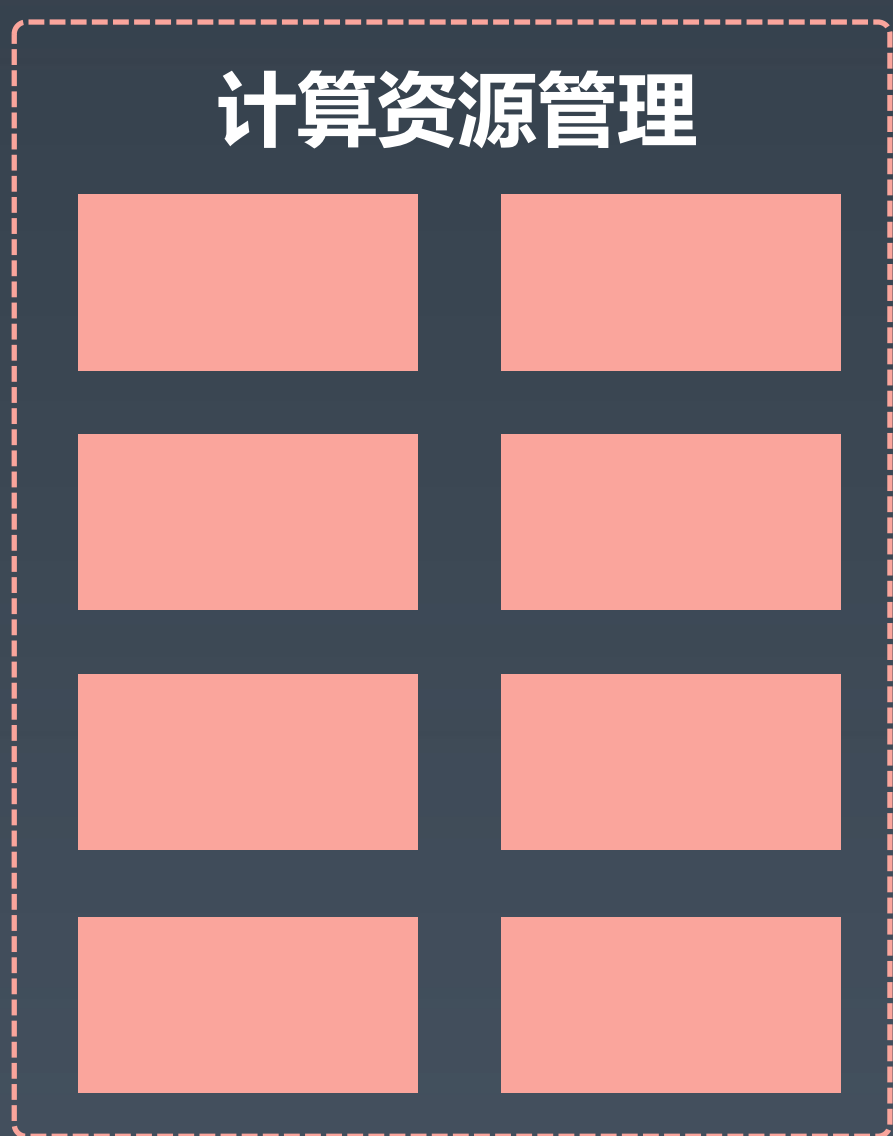
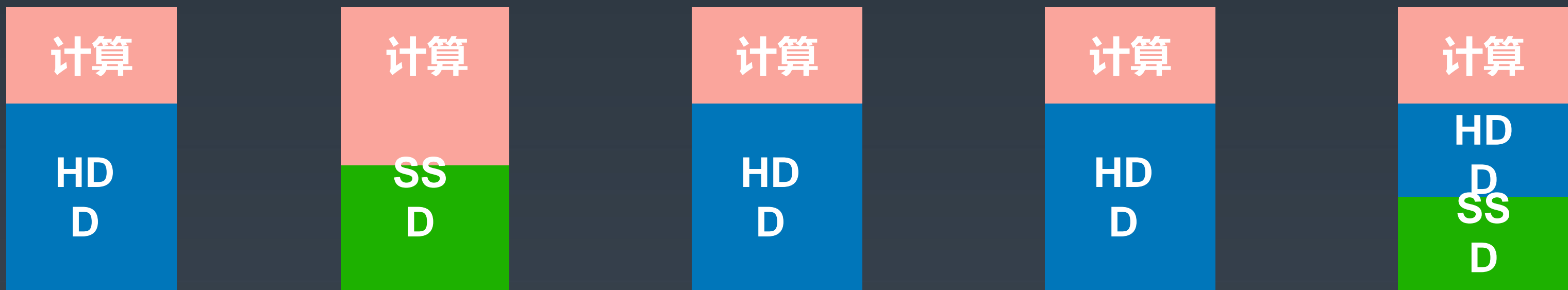
三、展望：收益和未来展望

混部收益

混部集群CPU利用率大幅提升**40-80%**!

离线预算节省服务器**30000+**台!

存储计算分离



问题

1. 磁盘调度是一个“超维”问题
2. 磁盘利用率低



效果

1. 磁盘管理调度扁平化
2. 提升磁盘管理效率



无差别混部

离线间混部

在线间混部

在离线混部



无差别混部

TGO 鲲鹏会

汇聚全球科技领导者的高端社群

🏢 全球12大城市

👤 850+ 高端科技领导者

使命
Mission

为社会输送更多优秀的
科技领导者

愿景
Vision

构建全球领先的有技术背景
优秀人才的学习成长平台



扫描二维码，了解更多内容



全球技术领导力峰会

Geekbang 极客邦科技 | TGO 鲲鹏会

500+ 高端科技领导者与你一起探讨 技术、管理与商业那些事儿



🕒 2019年6月14-15日 | 📍 上海圣诺亚皇冠假日酒店



扫码了解更多信息

THANKS!

QCon  th