

# Matrix

吕毅@百度基础架构部

# About Myself

- 天津大学
- 2007 加入百度
- 2009 Noah监控&机器管理
- 2012 Matrix
- 微博：毅LV(Lvyi@baidu.com)



# OutLines

- 主要问题
- 解决思路和挑战
- 关键问题
- 现状和未来

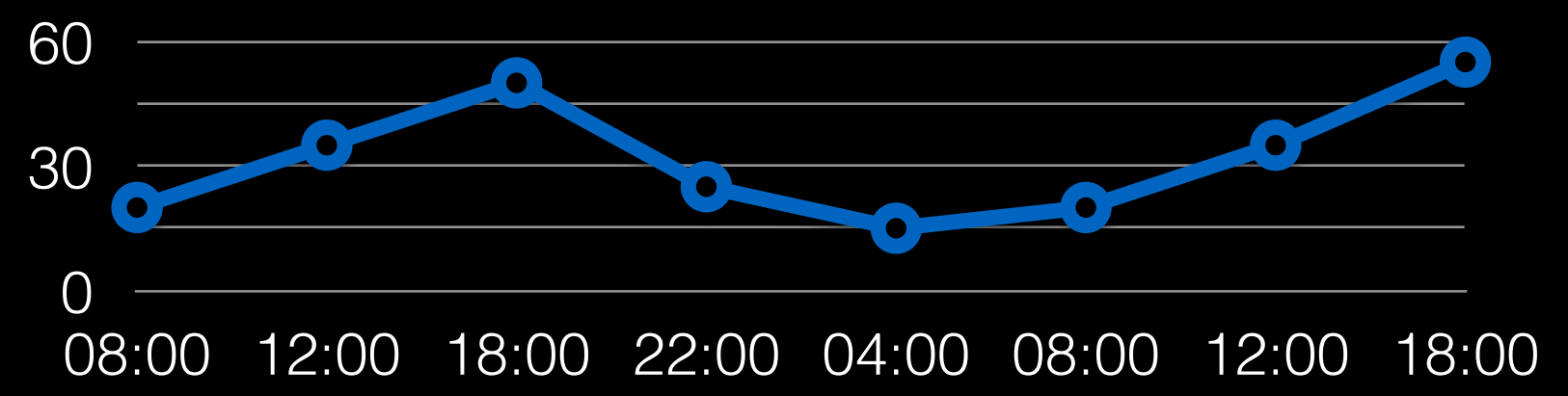
- 比肩Google
- 国内领先

# 主要问题

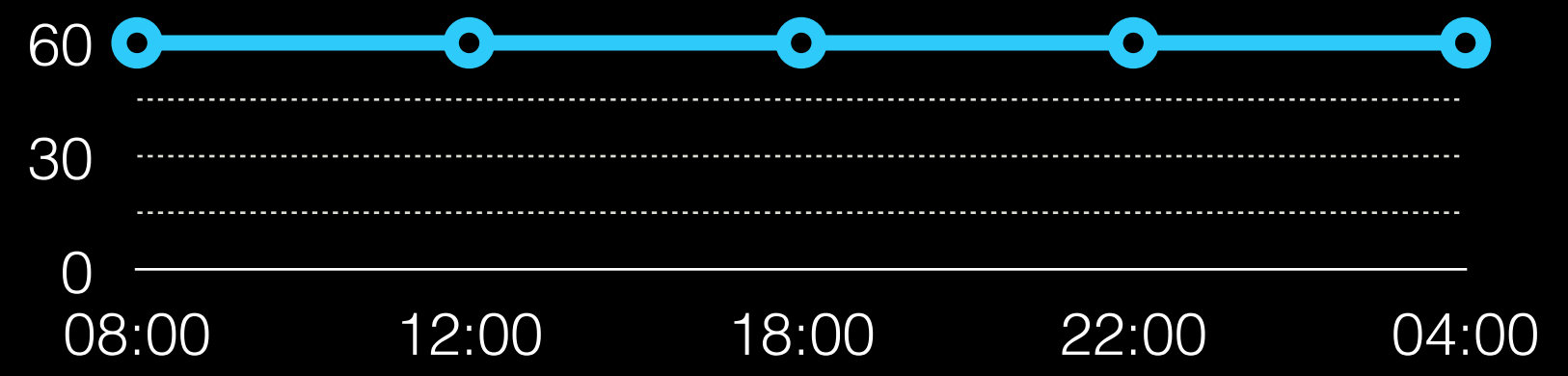
- 资源利用不均衡
- 故障处理效率低
- 预算/调配复杂

# 问题1: 资源利用不均衡

- 在线业务



- 离线计算



- Buffer资源

# 问题2: 故障处理效率低

- 故障规模大，种类多
- 耦合多，维修困难

# 问题3: 预算调配复杂

- 独立业务预算
- 业务独享机器, 备机
- 独立物理资产归属



# 解决思路

- 混部：在线和离线计算混部 —>资源利用率
- 解耦：应用和物理资源解耦 —>故障处理
- 统一：机器资源统一管理调度—>故障处理&预算

# 挑战

- 混部
  - 混部/解耦带来的技术问题
  - 互相影响：container技术
  - 服务发现：naming机制
- 降低业务迁移成本
  - 旧自动化系统&调度器低成本迁移
- 应用问题
  - 设计新预算模式，新运维模式
  - 业务分级和全局序

# 架构

百度业务

搜索

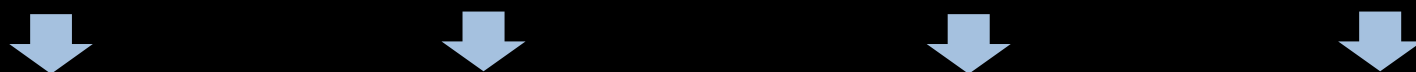
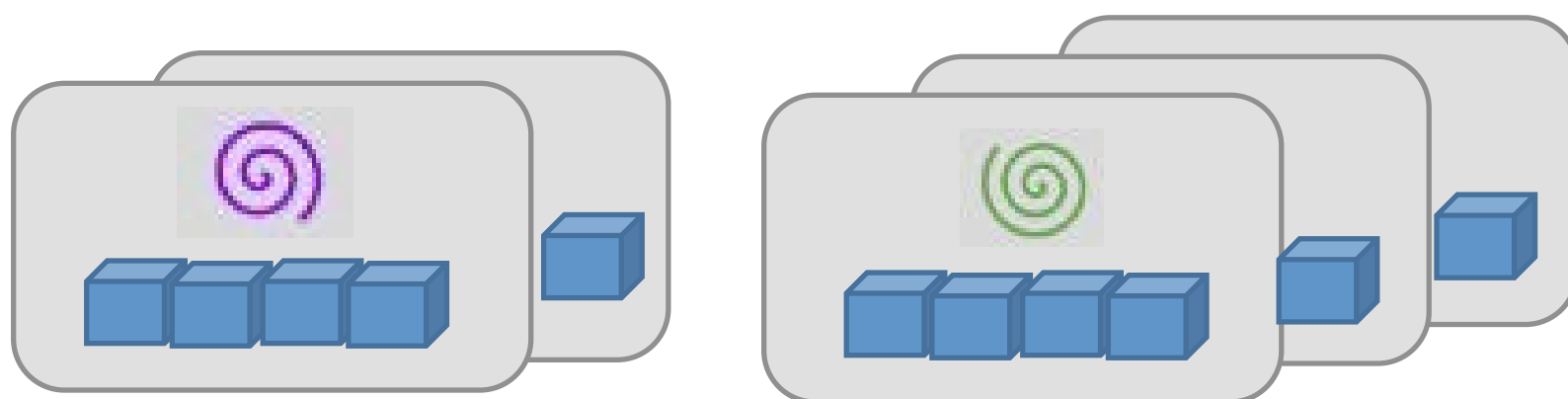
社区

移动

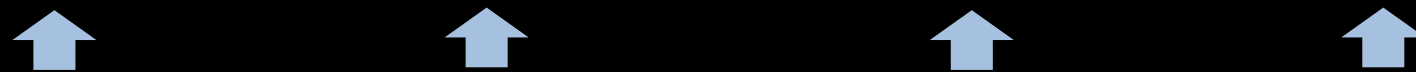
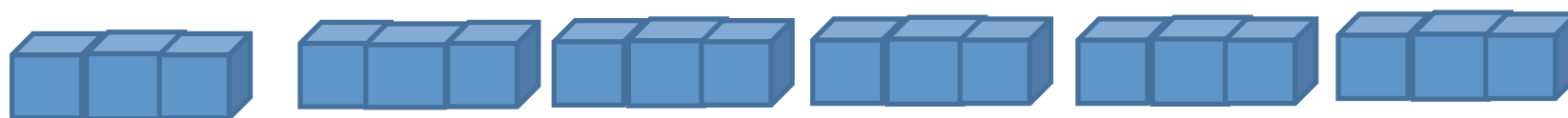
支付

团购

业务平台  
调度/部署



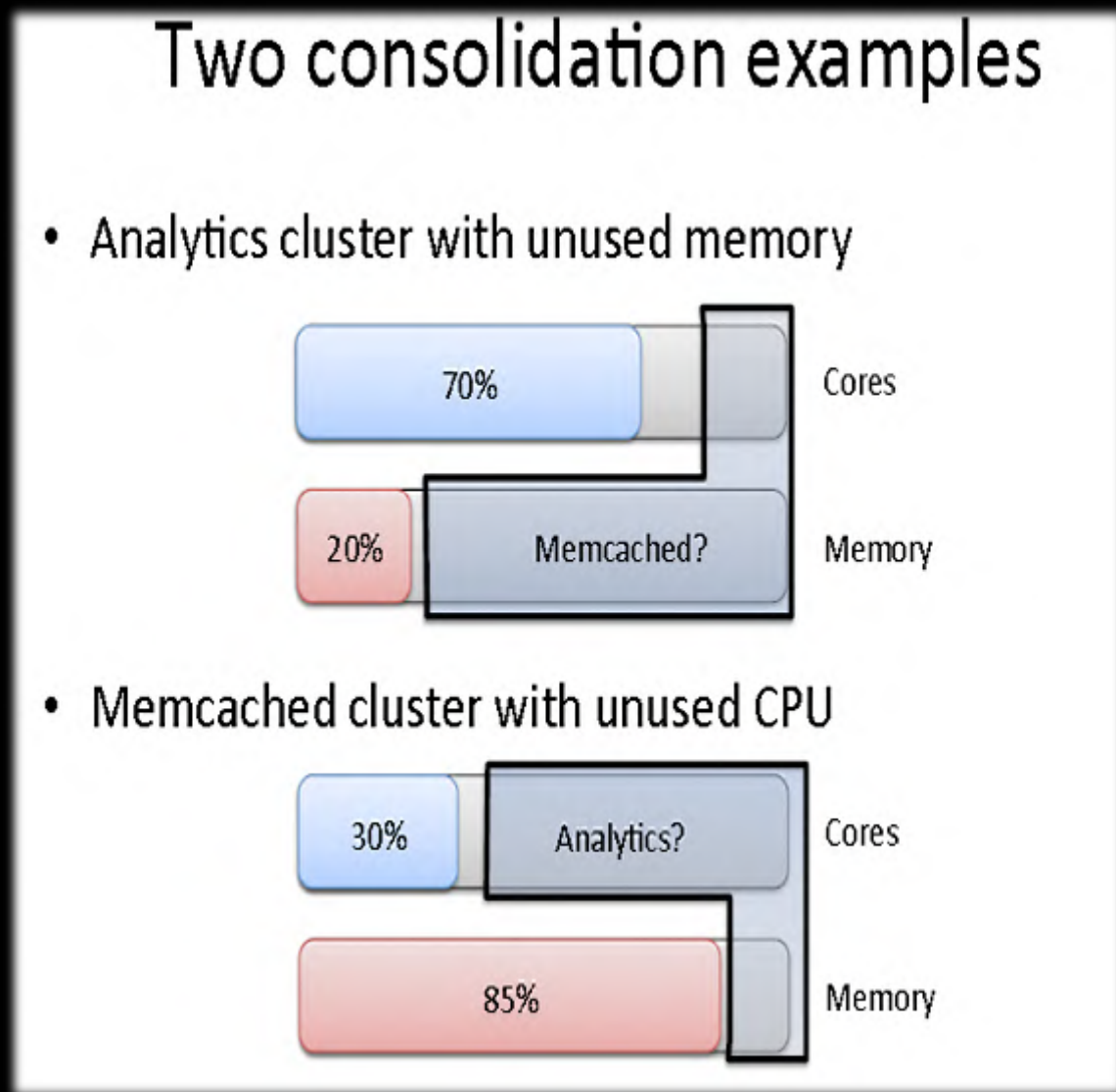
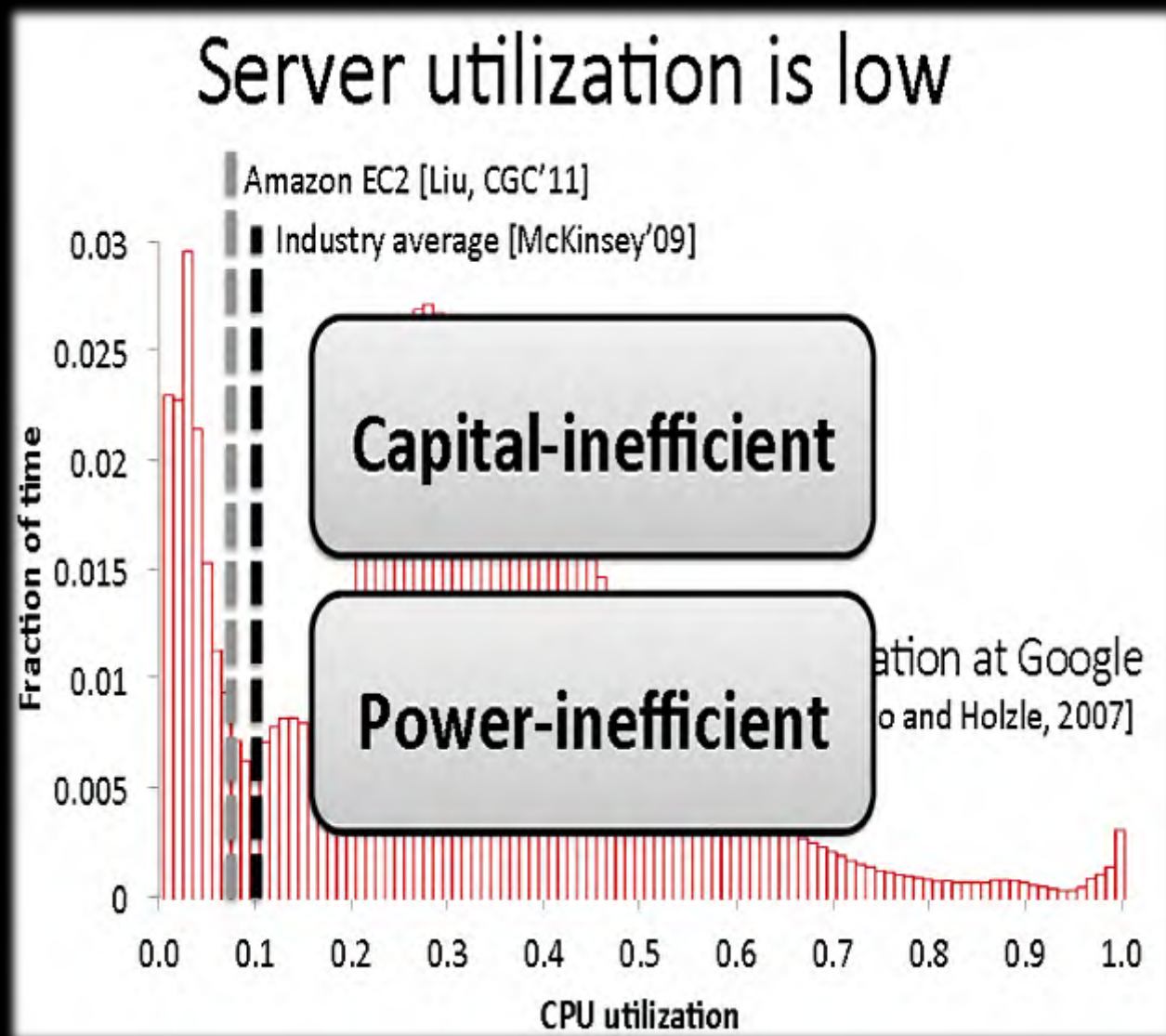
Matrix



服务器

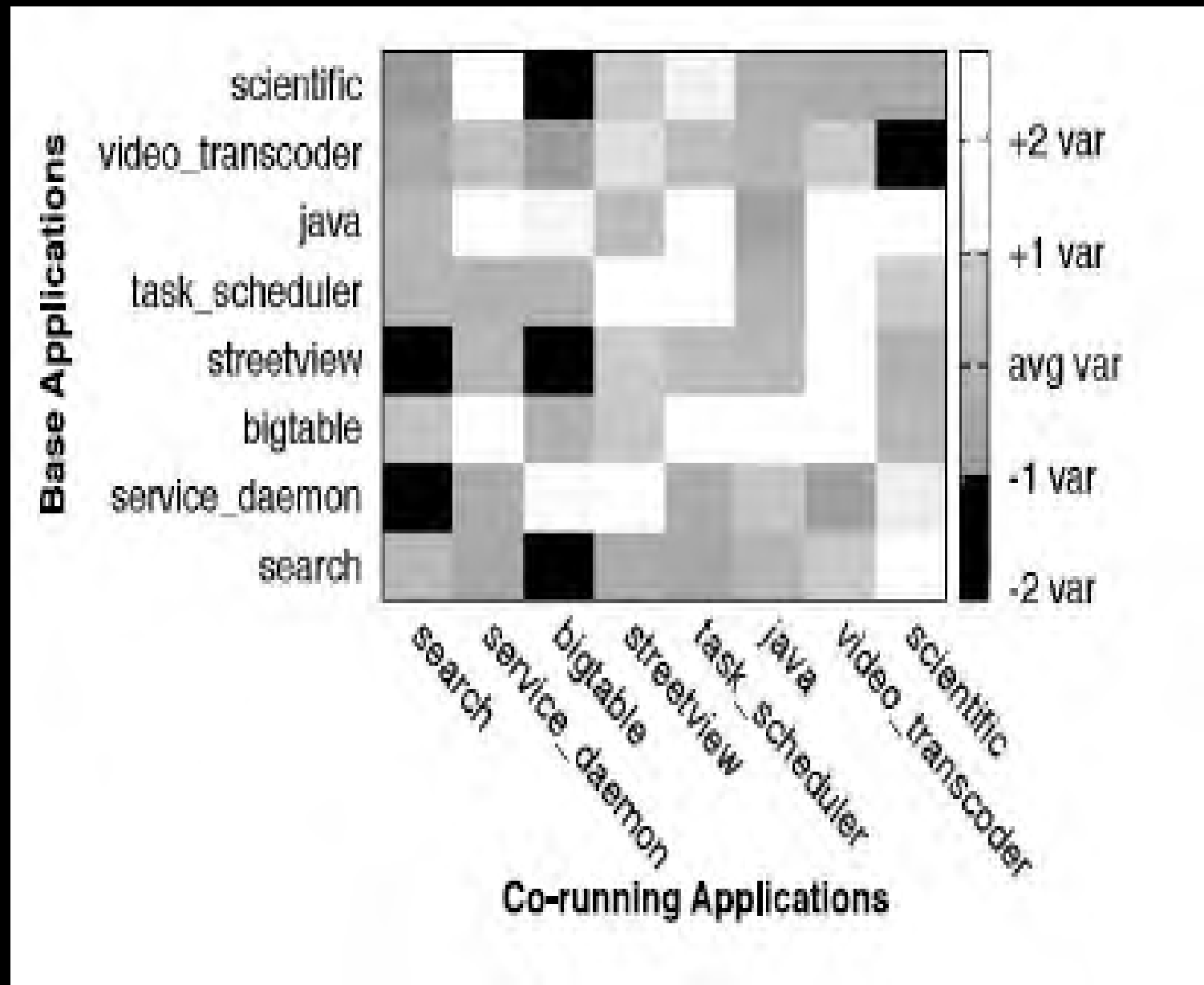


# 核心问题1：服务质量vs利用率



合理的利用率

# 核心问题2：哪些业务适合混部



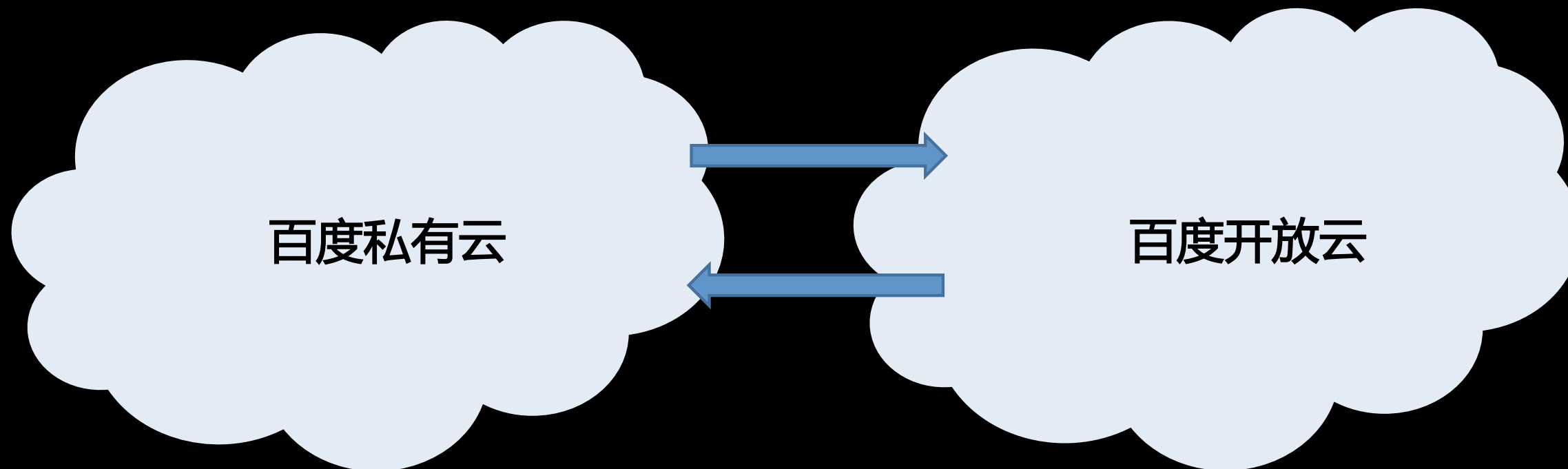
合适的业务

# 当前效果

- 公司内部业务的基石
  - 管理机器规模：几十万
  - 单机群规模：万
  - 支持所有计算：Map-Reduce、MPI，流式计算等
  - 支持所有核心在线业务：网页搜索，广告，社区等
  - 完美支持在离线混部

# 收益

- 取得阶段性收益
  - 年节省服务器：万
  - 自动化：硬件故障全自动
  - 利用率：提高总体利用率



MATRIX集群操作系统

Physical Infrastructure(服务器, IDC, 网络)

平台的平台



谢谢大家!

*-Lvyi@baidu.com*