淘宝网
Taobao.com

了解应用服务器

核心系统数据库组 余锋
http://yufeng.info
@淘宝褚霸
2012-08-15

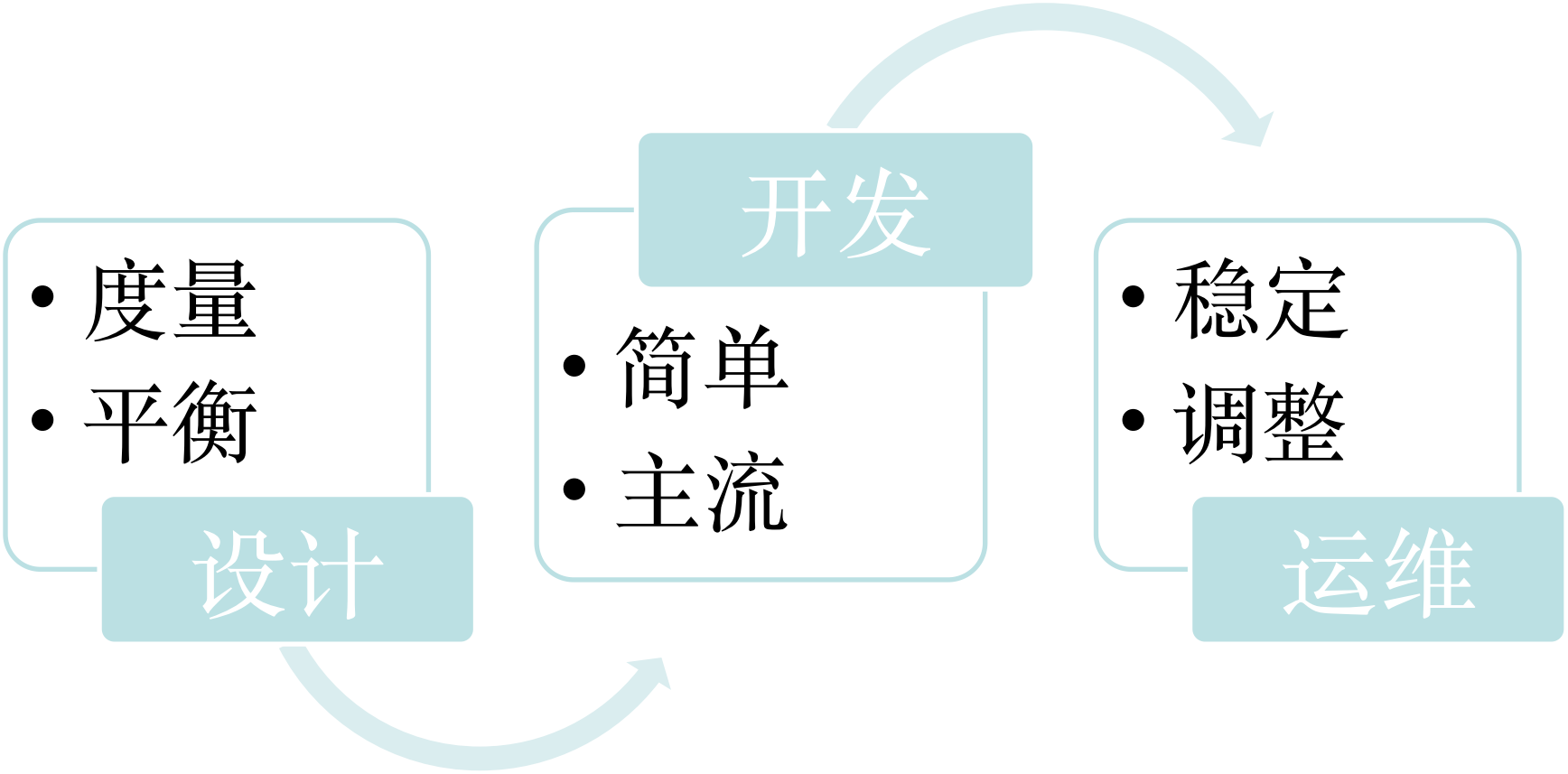
淘宝网 Taobao.com

C10K问题

- http://www.kegel.com/c10k.html
- 提出时间是2001年，10年过去了
- 挑战还在:
  - 用户对服务响应时间和可靠性要求越来越高。
  - 没有革命性的技术改进，算法和操作系统和库变化不大。
  - 硬件，操作系统，库，平台，应用的层次越来越深。

淘宝网 Taobao.com

应用服务器关键



淘宝网 Taobao.com

硬件的选择

- 切合业务的需求
- 系统平衡
- 可靠性
- 性价比

淘宝网 Taobao.com

平台的选择

- 白盒子还是黑盒子
- 最大程度的挖掘硬件与系统的潜力
- 应用系统最大的框
- 平台的生命力

淘宝网 Taobao.com

并行化

- 理论依据
  - 新硬件趋向并行化
  - 软件需要提高并行度
- 解决
  - CPU计算力过剩
  - 适应NUMA架构下大内存
  - 设备IO能力过剩
- 多实例还是虚拟化

淘宝网 Taobao.com

进程模型的选择

- 单线程多进程
  - 极大避免进程上下文切换的影响
  - 编程模型复杂
- 单进程多线程
  - 编程模型简单
  - 如有可能不要尝试多线程，实在不行用原语高级点的库
  - 很难调试 做好诊断设施

淘宝网 Taobao.com

通讯模型的选择

- Don't communicate by sharing memory
- share memory by communicating
- 进程间通信（IPC）

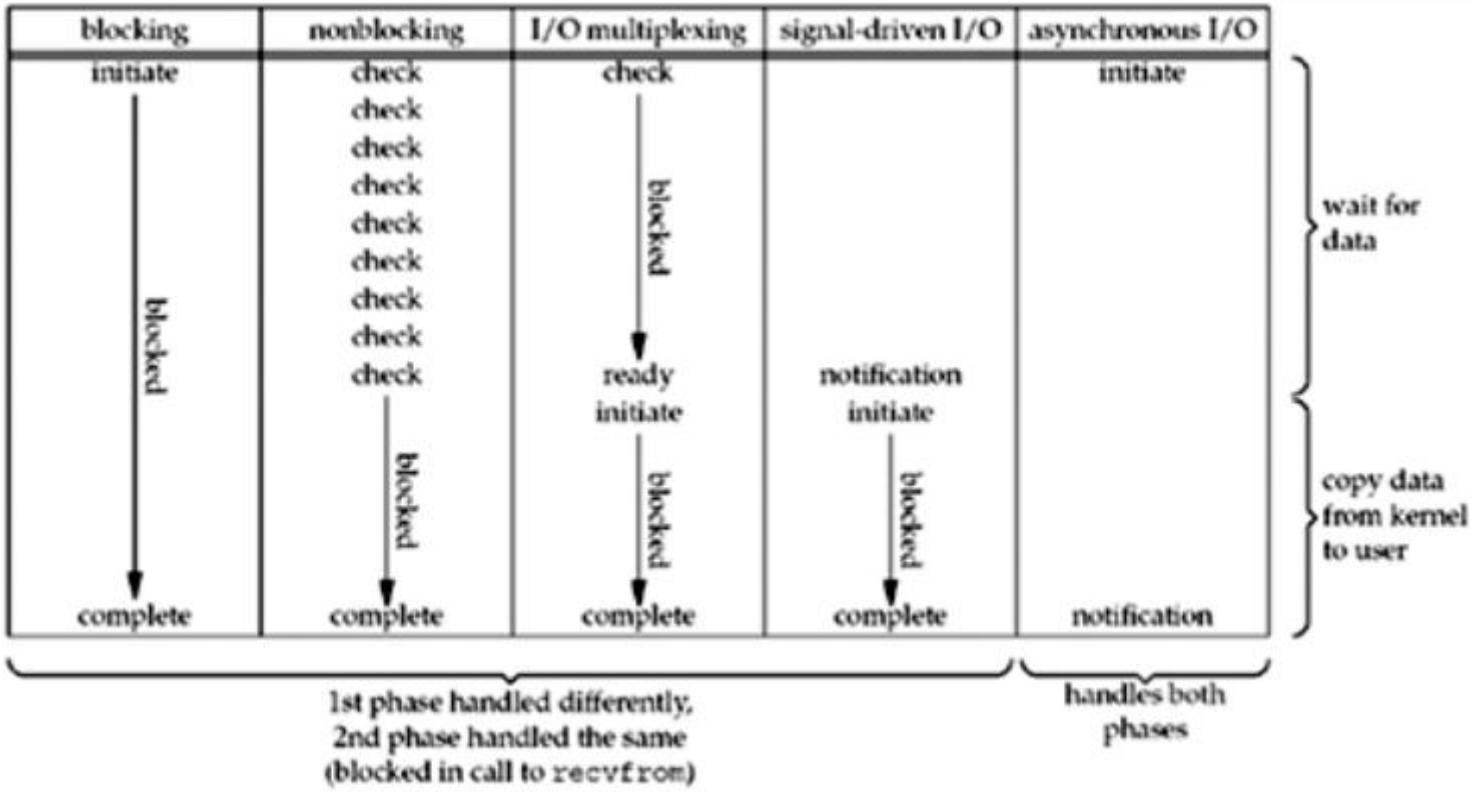
淘宝网 Taobao.com

IO模型的选择

- 阻塞I/O
- 非阻塞I/O
- I/O多路复用
- 事件驱动 I/O (SIGIO)
- 异步I/O (POSIX aio_函数和native aio)

淘宝网 Taobao.com

IO模型的差别



淘宝网 Taobao.com

语言的选择

- 语言成熟度
  - 开发人员成本
  - 社区经验
- 涉及面广
  - 习惯
  - 性能
  - 库
  - 运行期
  - 维护期

淘宝网 Taobao.com

业务逻辑

- 整个业务就是一个大的状态变迁图
- 一个外部对象对于一组状态机
- 状态的变化是消息引起的
- 消息是可以跟踪的

淘宝网 Taobao.com

库的选择

- 语言的延伸
- 简单够用，业界主流时间验证过
- 抵制重复造轮子
- 一次做一件事，做深做透

淘宝网 Taobao.com

协议选择

- 业界主流，方便对接
- 文本协议
- 二进制协议
- 基于规则自动解析
- 手动解析

淘宝网 Taobao.com

基础数据结构的选择

- 简单就是美
- 压缩数据集，避免数据搬动
- 除非必要不要用非常复杂的数据结构
- 数组、 链表解决大部分事情
- 数组方便利用硬件cache
- 不需要为未来的东西付出代价

淘宝网 Taobao.com

可扩展设计

- 方便业务扩展
- 方便团队参与
- 接口要足够简单
- 可以考虑用脚本
- 形成生态系统

淘宝网 Taobao.com

可配置

- 适应经常性的变化和个性化需求
- Unix .conf或者.ini格式
- 可以用类似Lua等高级语言来描述

淘宝网 Taobao.com

稳定性

- 稳定压倒一切
- 资源预分配
- 不可以发生泄漏，堆积
- 不可使用过多资源
- 避免频繁的分配/释放内存（内存碎片）

淘宝网 Taobao.com

制约性能的杀手

- 数据拷贝
  - 4K数据拷贝的开销1us
- 上下文切换
  - 考虑到缓存践踏，开销20-30us
- 内存分配
  - 不可再生资源，倒腾代价巨大
- 锁竞争
  - 最快的锁也要0.1us
- 昂贵的系统调用

淘宝网 Taobao.com

性能可微调

- 提供简单的接口动态改变服务器关键数据
- 自动根据系统的资源情况来决定默认值
- 有统计模块做微调依据

淘宝网 Taobao.com

移植性

- 平台移植
- 硬件移植
- 最小化特性
- ANSI C代码，最大可移植性

淘宝网 Taobao.com

工程实践的考虑

- 严谨的错误处理
- 为易诊断而编码
- 完善的日志系统
- 健康监控

淘宝网 Taobao.com

总结

- 充分利用软硬件特性
- 简单就是美
- 不造轮子
- 开放融入主流社区
- 稳定压倒一切
- 为运维而准备

提问时间

谢谢大家！