

从概率和用户感知出发 实现高可用架构

史海峰@当当





本宝宝已崩溃

这日子没法过了



硬件故障·网络故障·能源故障
自然灾害·战争·外部不可抗力

挖掘机

天灾





人为失误·应用BUG·设计缺陷
性能瓶颈·资源不足
安全攻击

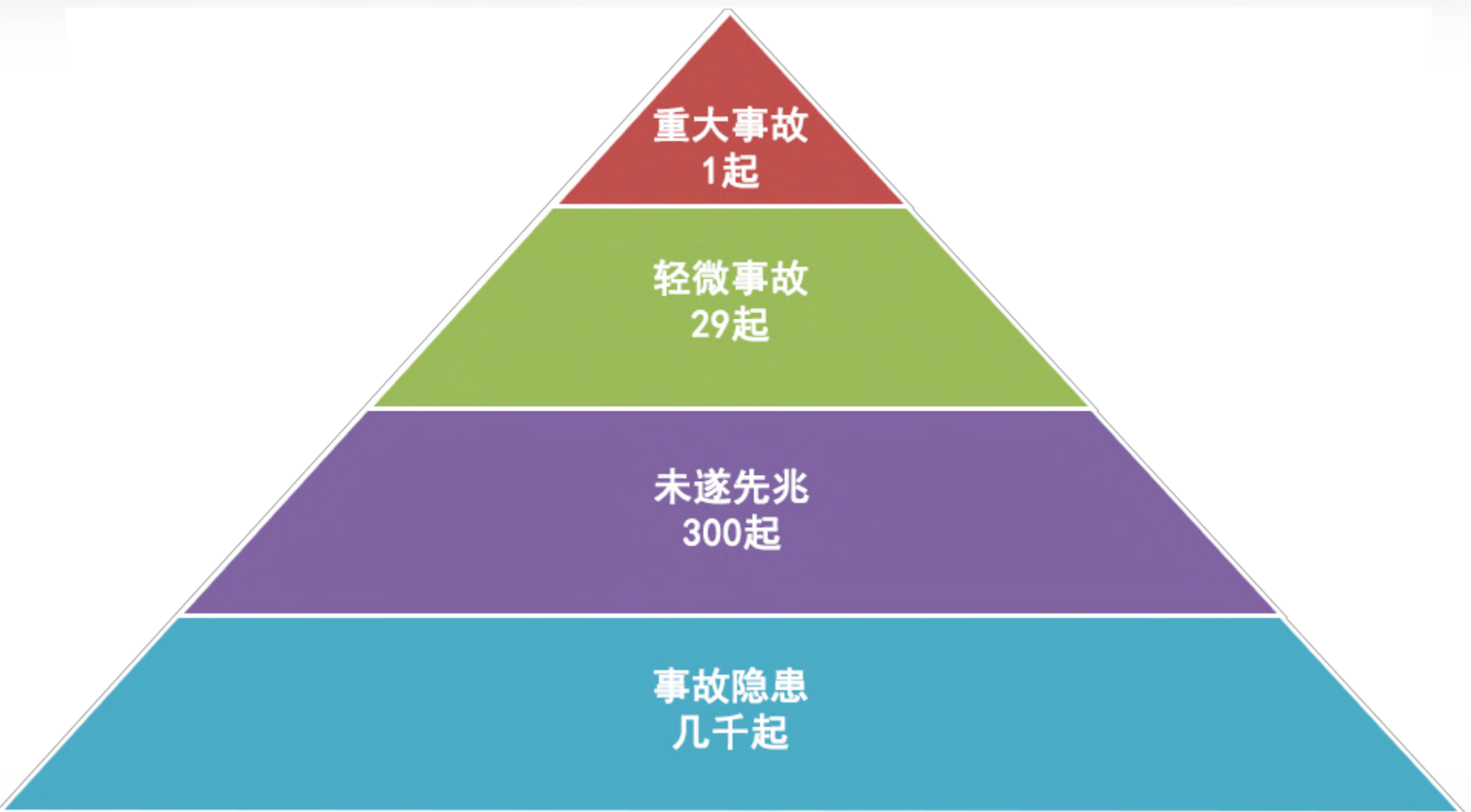
人祸

Terrorist Action

灵异事件·黑天鹅效应·未解之谜



未知

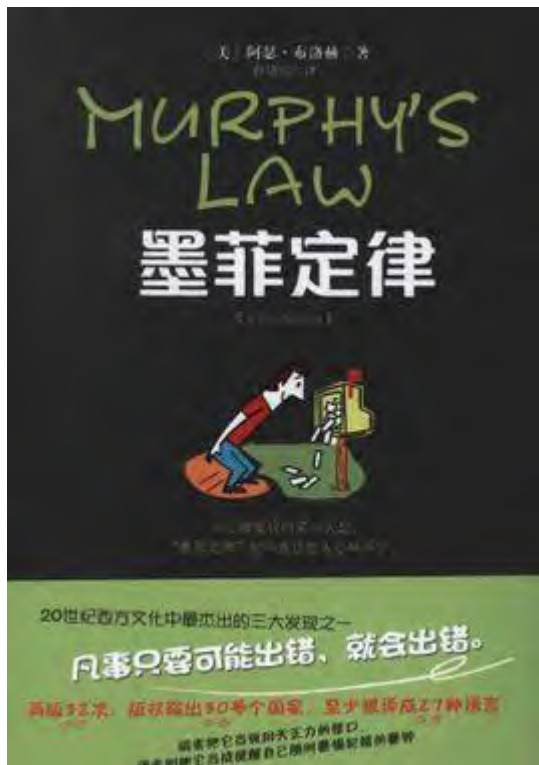


高可用不是万无一失



单个应用节点不可用概率？
单系统部分业务影响概率？
集群不可用故障概率？
核心业务不可用概率？

越复杂的系统越难以评估



人算不如天算

什么是高可用

| 互联网应用架构实战峰会

降低故障
出现概率



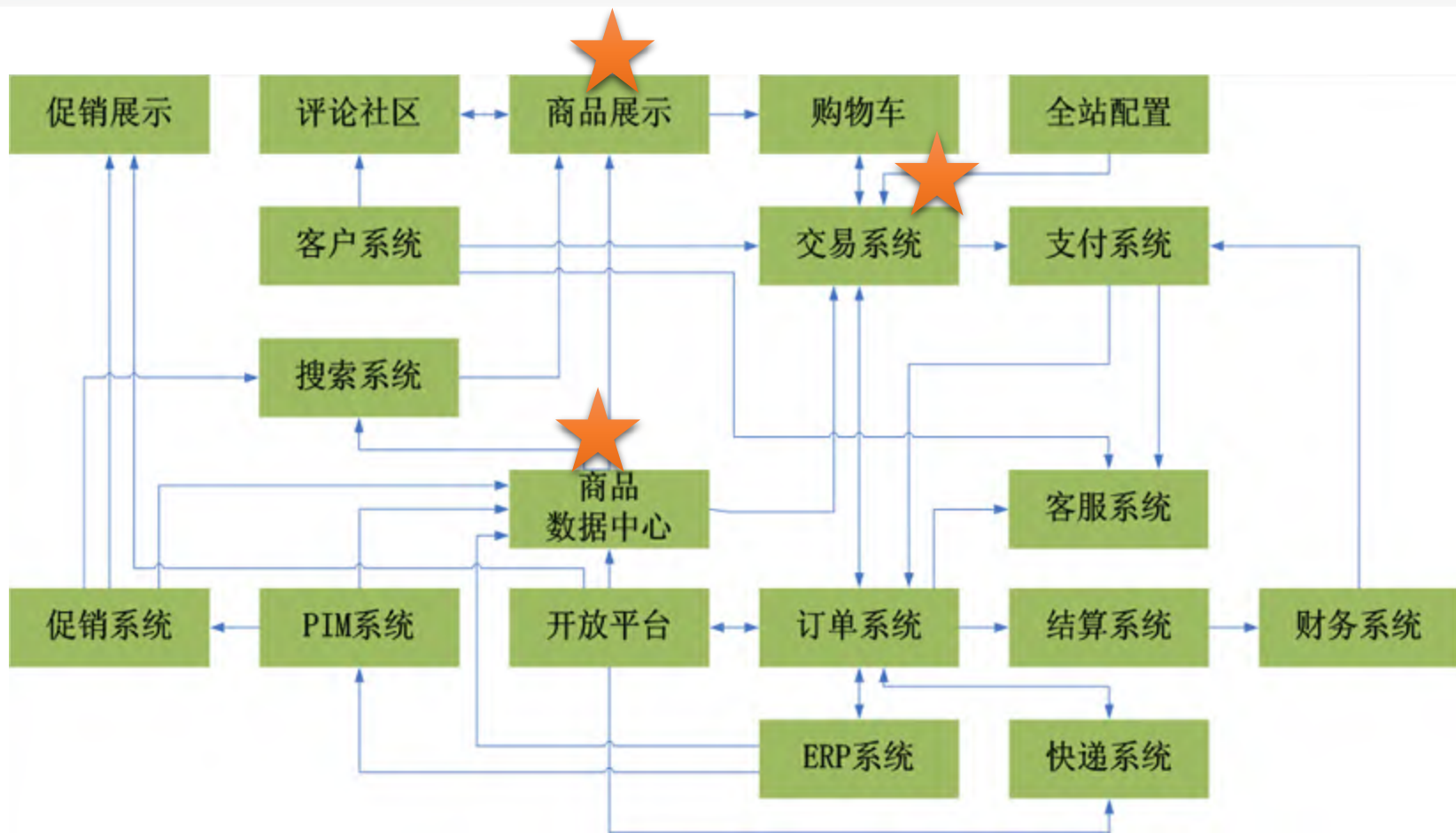
缩小故障
影响范围

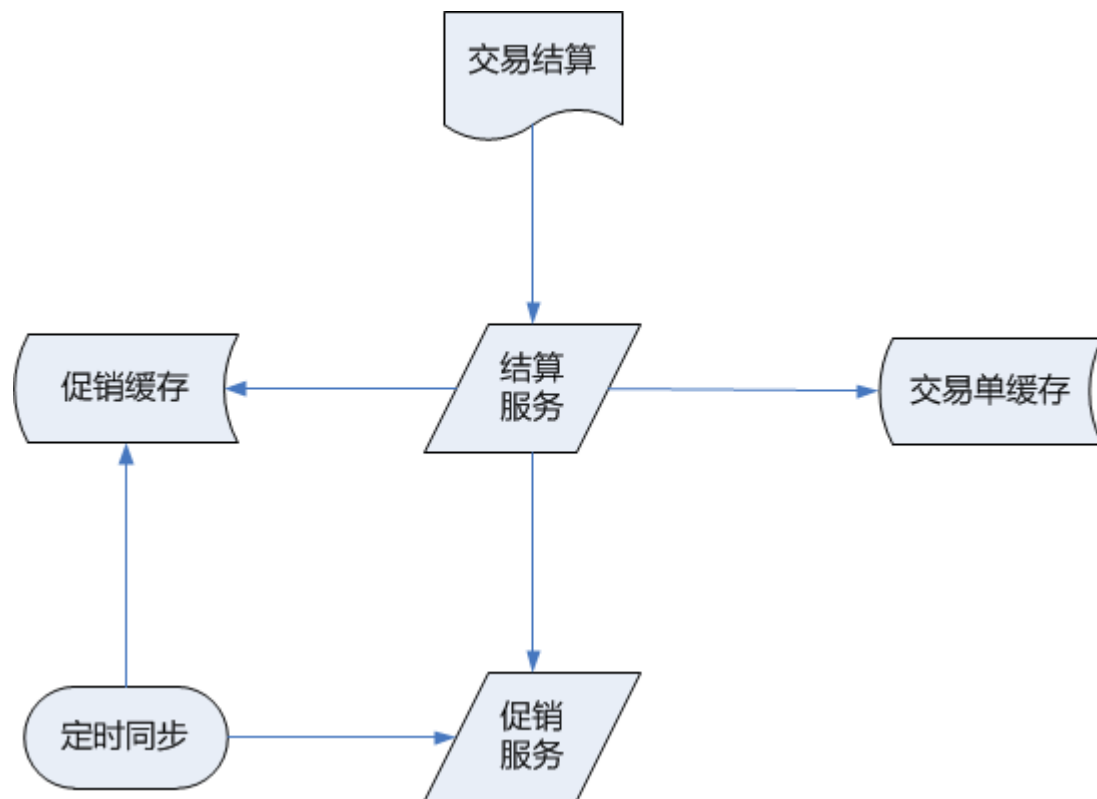


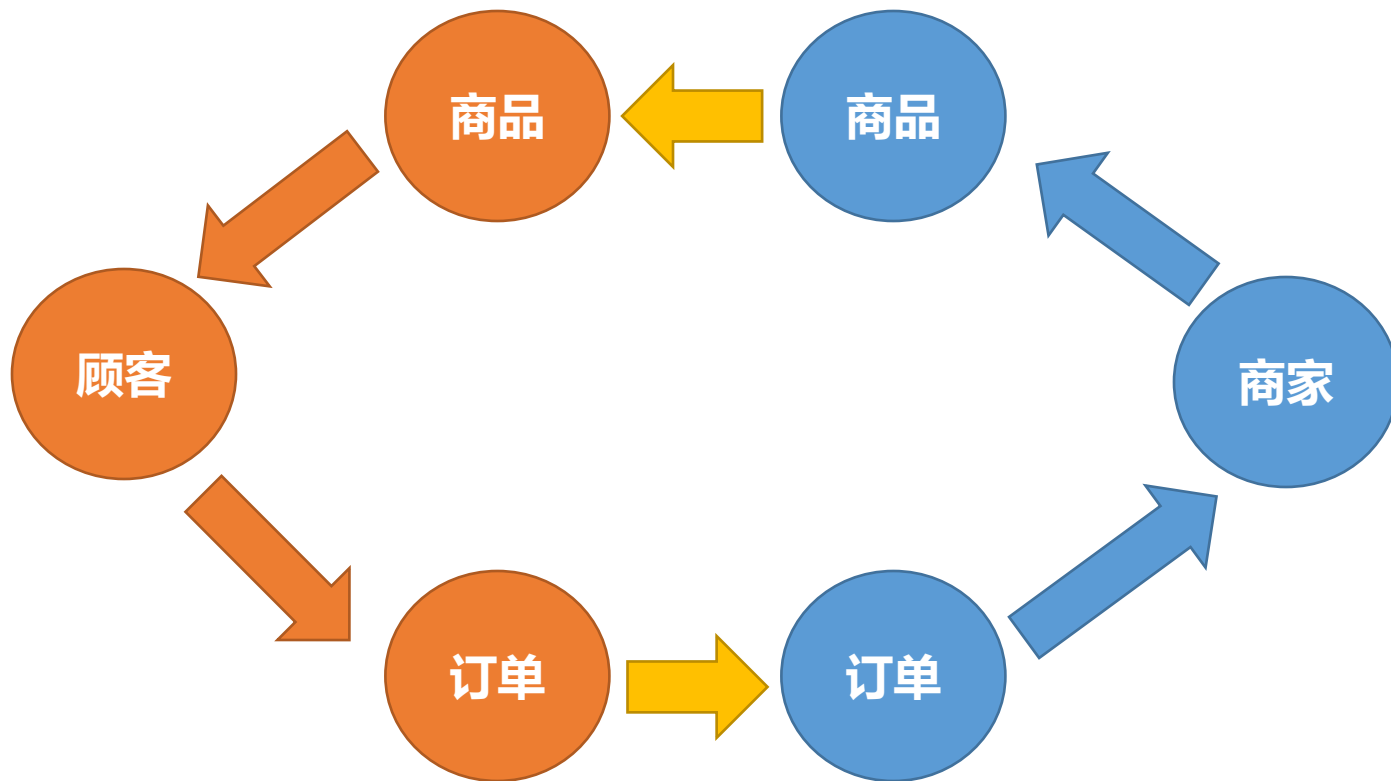
出现故障
快速恢复

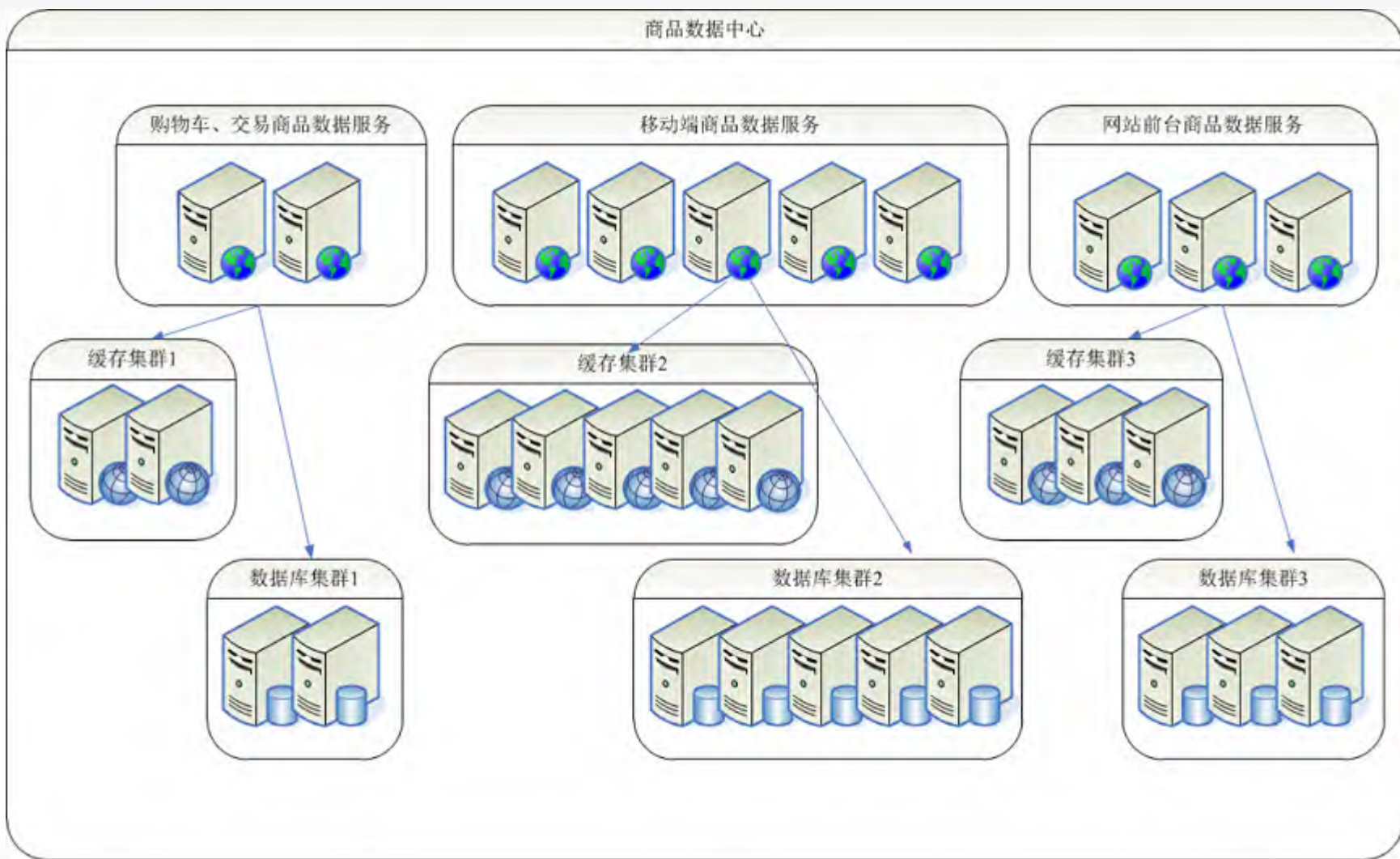


终极目标
用户第一





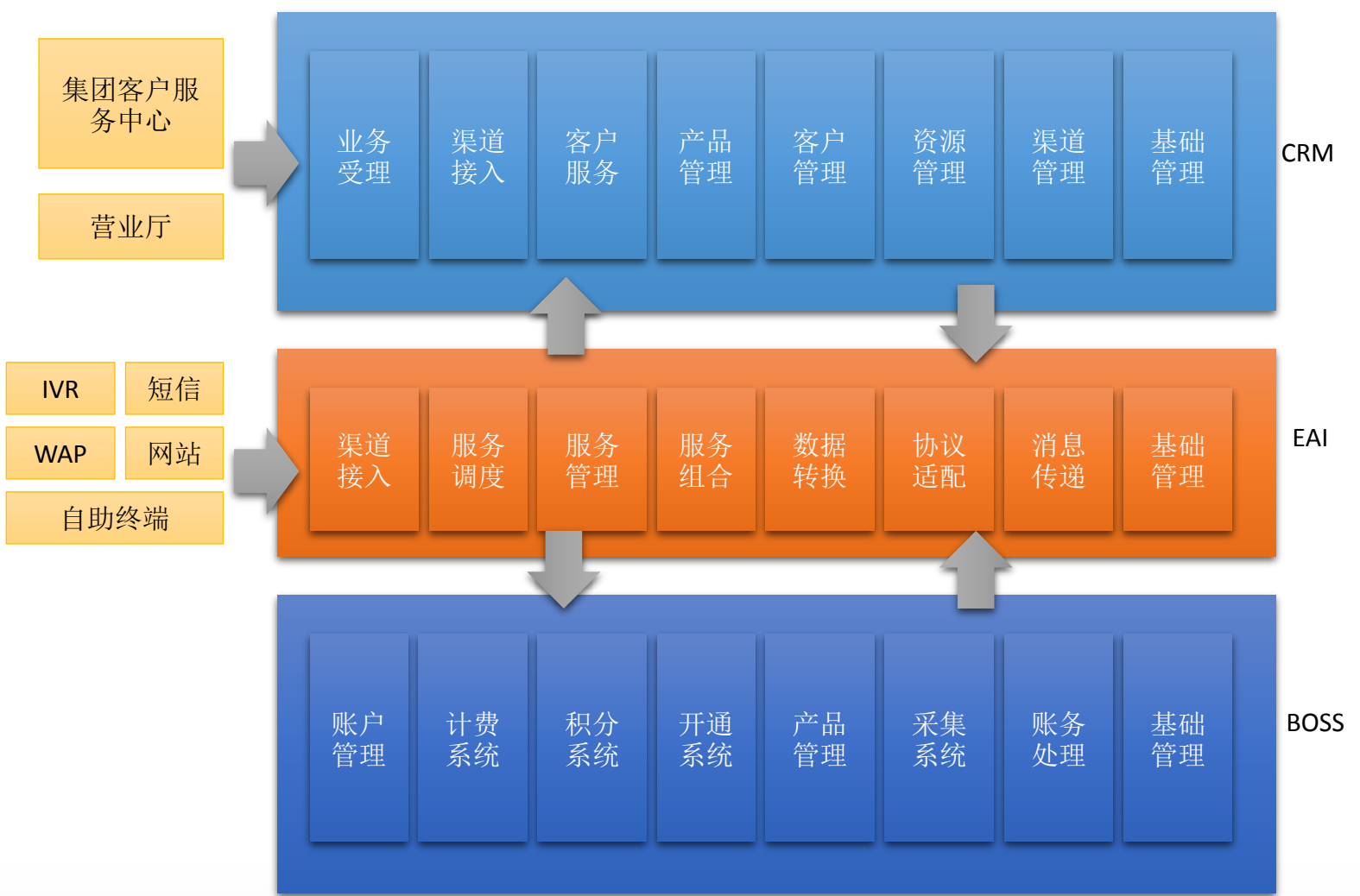






限流!
限流!
限流!

某移动自助渠道高可用架构设计 | 互联网应用架构实战峰会





业务量占总量70%
月初月末存在高峰期
每天业务量存在高峰期
长时间无反应应用户习惯
重发
重发无果部分用户会电
话投诉



异步

• 采用异步队列方式进行缓冲

细分

• 根据业务类型优先级进一步细分资源池

提示

• 判断有较多请求未处理，反馈已受理，业务繁忙，请耐心等待

排重

• 判断有相同业务请求未处理，反馈上一条在处理中，请耐心等待

限流

• 设置每个队列长度限制，超出反馈业务繁忙，请稍后再试

