

阿里数据库架构变迁与展望

DTCC

2015中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015
大数据技术探索和价值发现

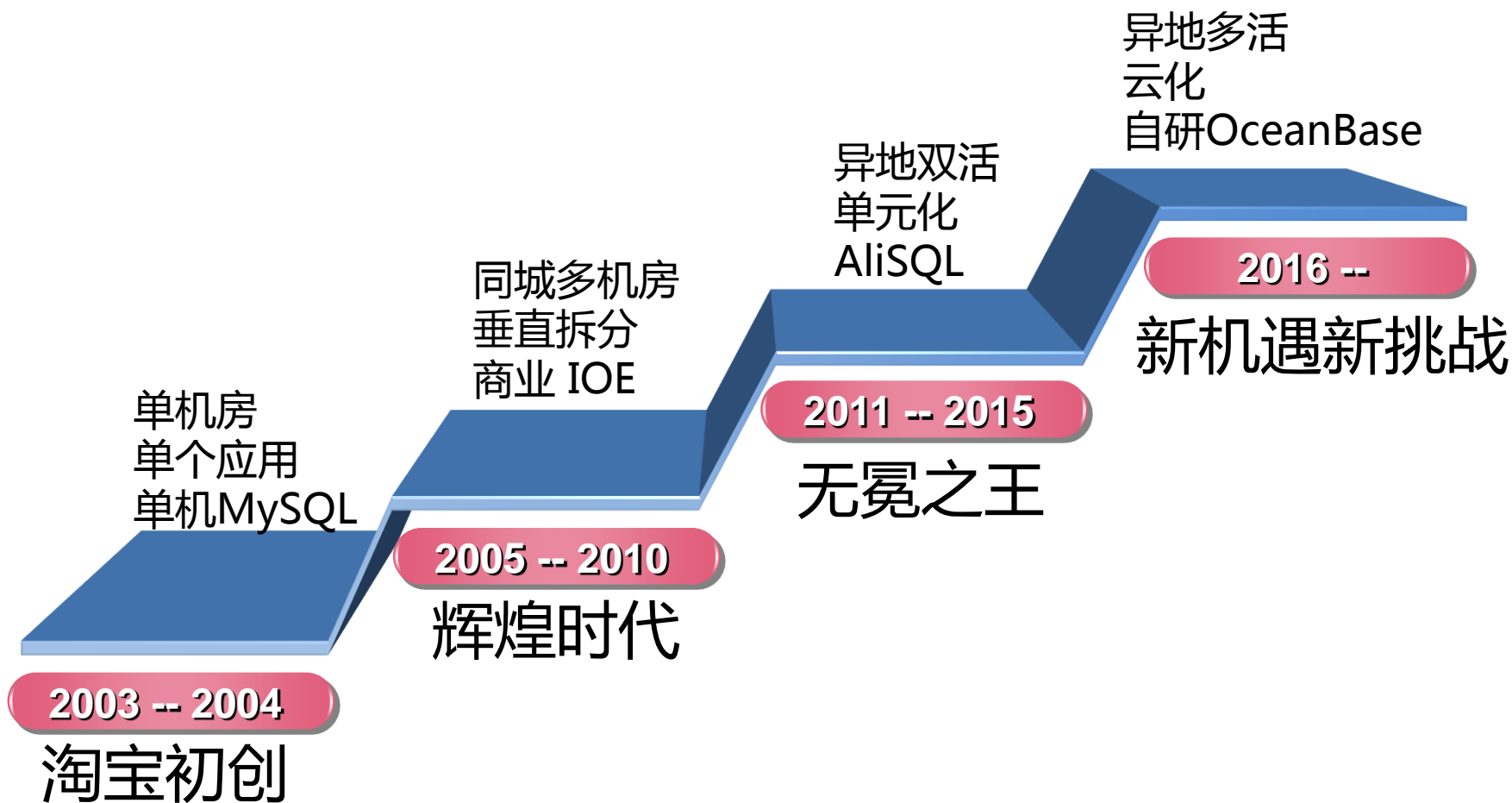


个人介绍

- 李圣陶
- 花名 刘昆，09年加入淘宝
- 阿里技术保障数据库专家
- 微博：@tao陶先生
- 邮箱：liukun@taobao.com



阿里数据库体系的四个时代



淘宝初创

The screenshot shows the early Taobao.com website. At the top left is the Taobao logo (a yellow four-petaled flower) and the text "淘宝网 生活的酷体验!". Below this is a navigation menu with four items: "卖家通道" (Seller Channel), "买家通道" (Buyer Channel), "淘宝者联盟" (Taobao Alliance), and "服务中心" (Service Center). To the right of the navigation menu is a search bar with a magnifying glass icon and the text "这里多的是你想要淘的“宝”!". Below the search bar are several product listings. On the left side of the main content area, there is a login section with fields for "会员登录名:" and "密码:", a "我来啦!" button, and a checkbox for "浏览时自动登陆:". At the bottom of the page, there is a footer with the text "首页 - 卖东西 - 注册会员 - 登录 - 帮助".

www.taobao.com

淘宝网
生活的酷体验!

卖家通道
买家通道
淘宝者联盟
服务中心

会员登录名:
密码:

 浏览时自动登陆:
我忘记了密码!

这里多的是你想要淘的“宝”!

热门的商品类目

索尼 DCR-V12000
日本最新科技! 索尼公司
荣誉出品! 全新, 没有使
用过!
一口价: ¥12000/架

宫廷时尚沙发
今年英国最流行的宫廷
沙发, 上百种款式!
开价: ¥20000/架

诺基亚 8390
全新瑞典诺基亚8390彩
屏手机, 有摄像头!
拍卖起价: ¥6000

ELLE
所有商品8折优惠!
现有多种最新款式,
欲购从速!

POLO 2003年夏季最新款
T恤, 现有红, 蓝, 黄, 紫
等多种颜色可供选择!
欢迎预定! 价格!

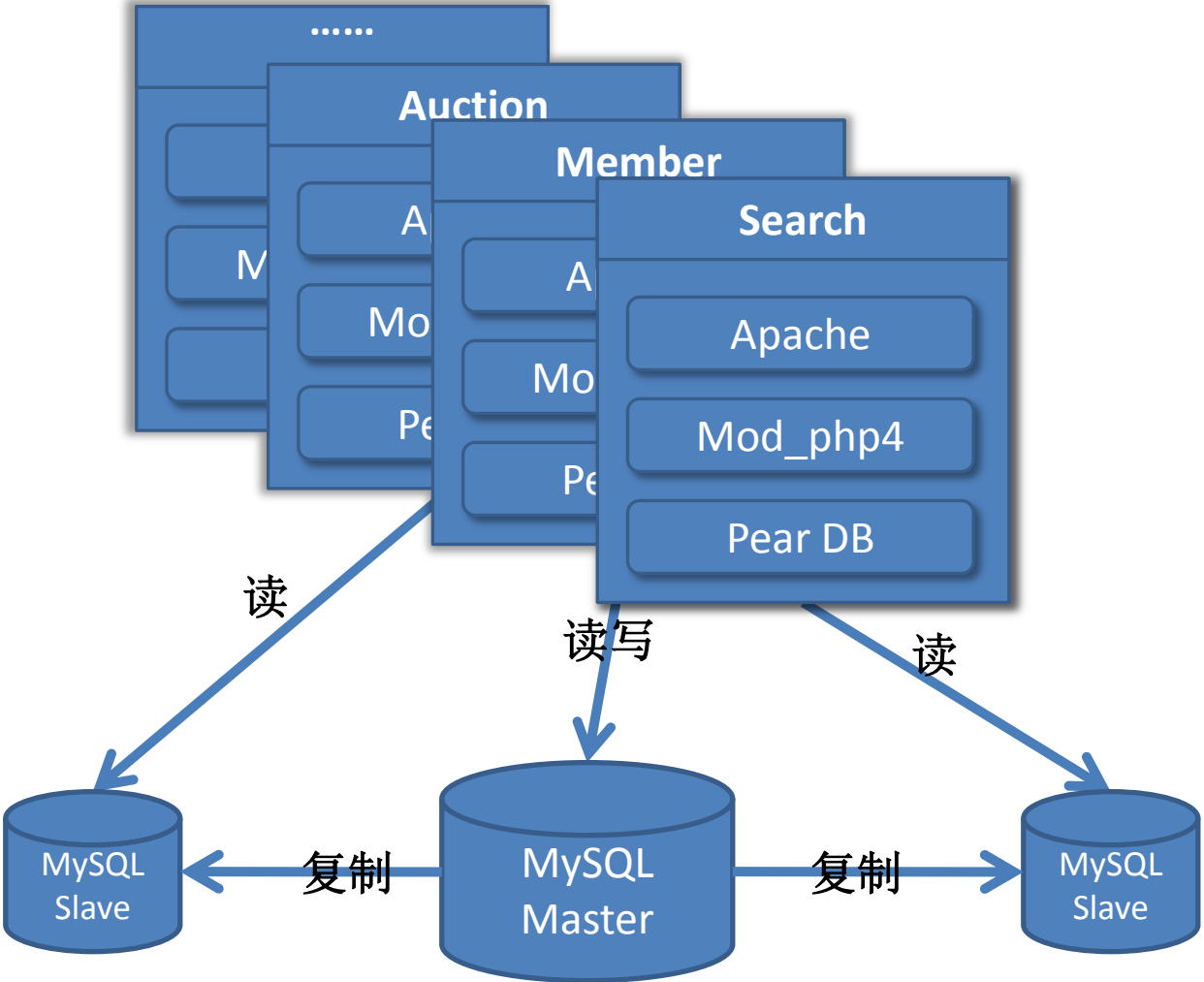
ELLE 1999年款登山军靴!
多种尺码, 黑色, 褐色可供
大家选择!
一口价: ¥1200

逛逛店铺一条街!

如何买卖? 去上一口价!
怎样拍货? 付费广告刊登?

首页 - 卖东西 - 注册会员 - 登录 - 帮助

淘宝初创



淘宝初创

问题:

- 单机MySQL数据库迅速达到瓶颈

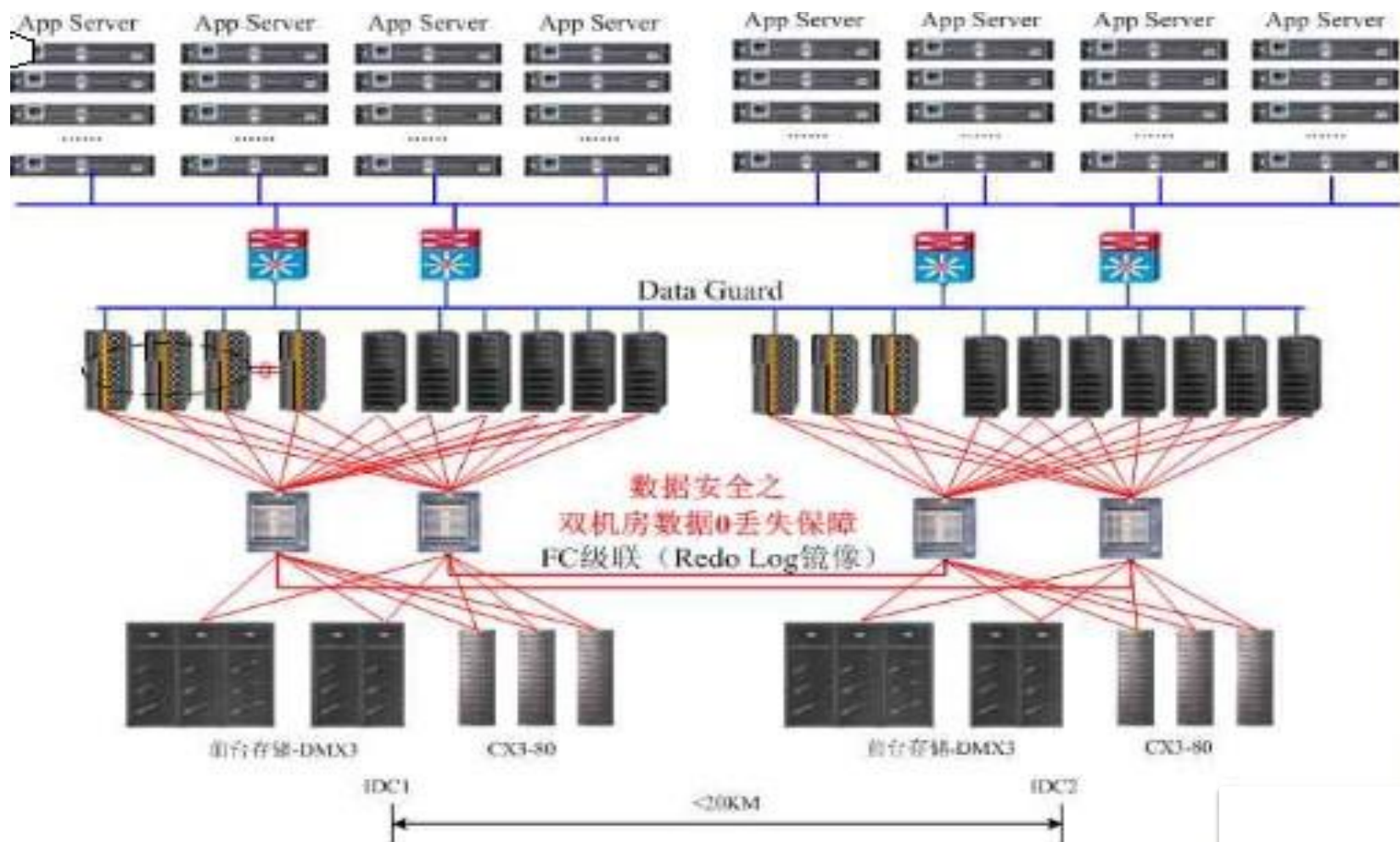
解法:

- MySQL迁移到Oracle，并逐步升级硬件，到小型机，高端存储，最终形成IOE 架构

效果:

- 支撑了淘宝 2004 到 2009 发展高峰

辉煌时代--IOE



辉煌时代--IOE

问题:

- 扩展性、可用性
 - 垂直扩展走到了极限
 - 集中式，稳定性的挑战
- 掌控力
 - 闭源的Oracle
 - 封闭的小机/存储

回首往事 - 从 IOE 架构走向 AliSQL 分布式架构

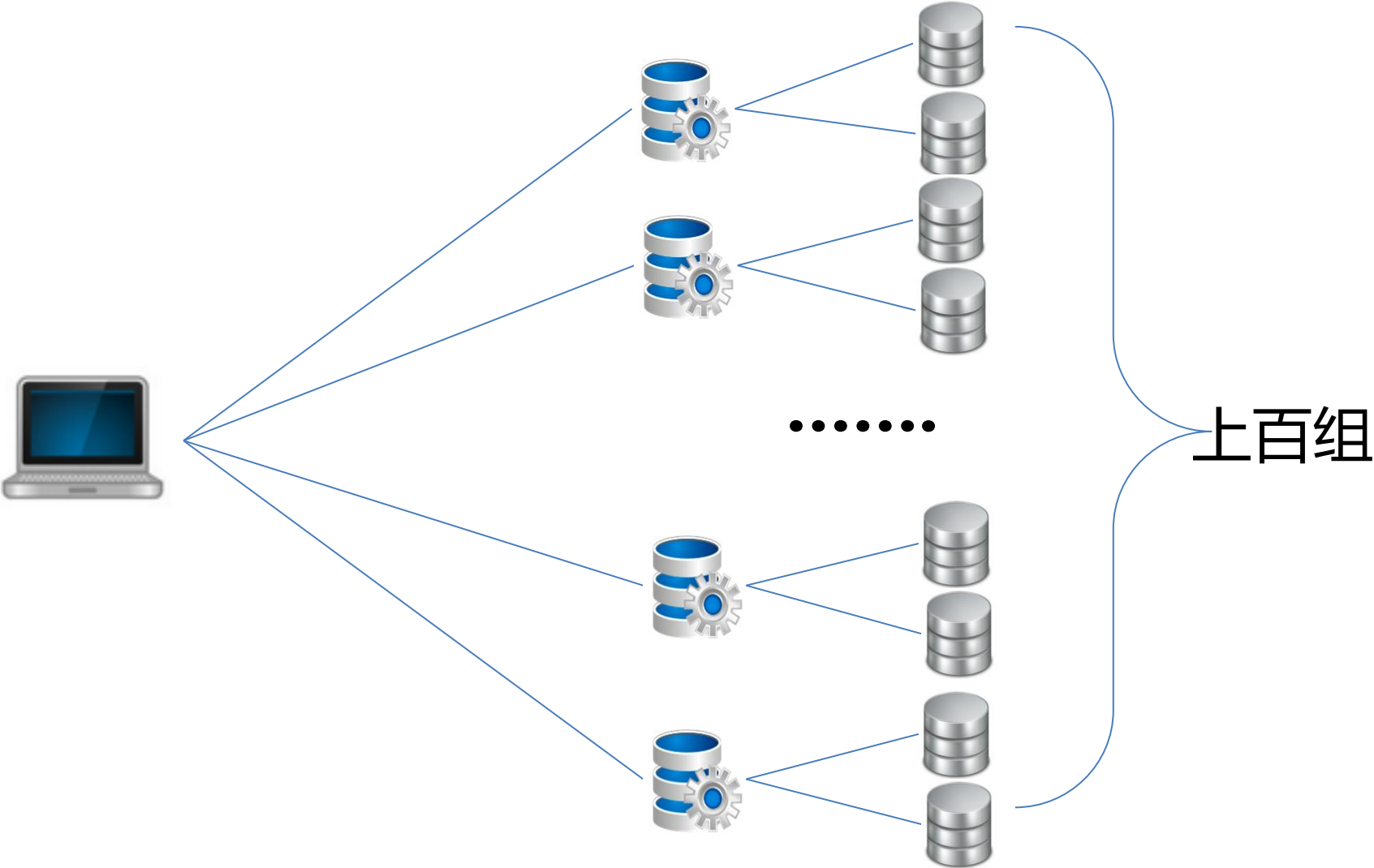
第一次推动中国数据库产业变革

➤ 获得无限掌控力

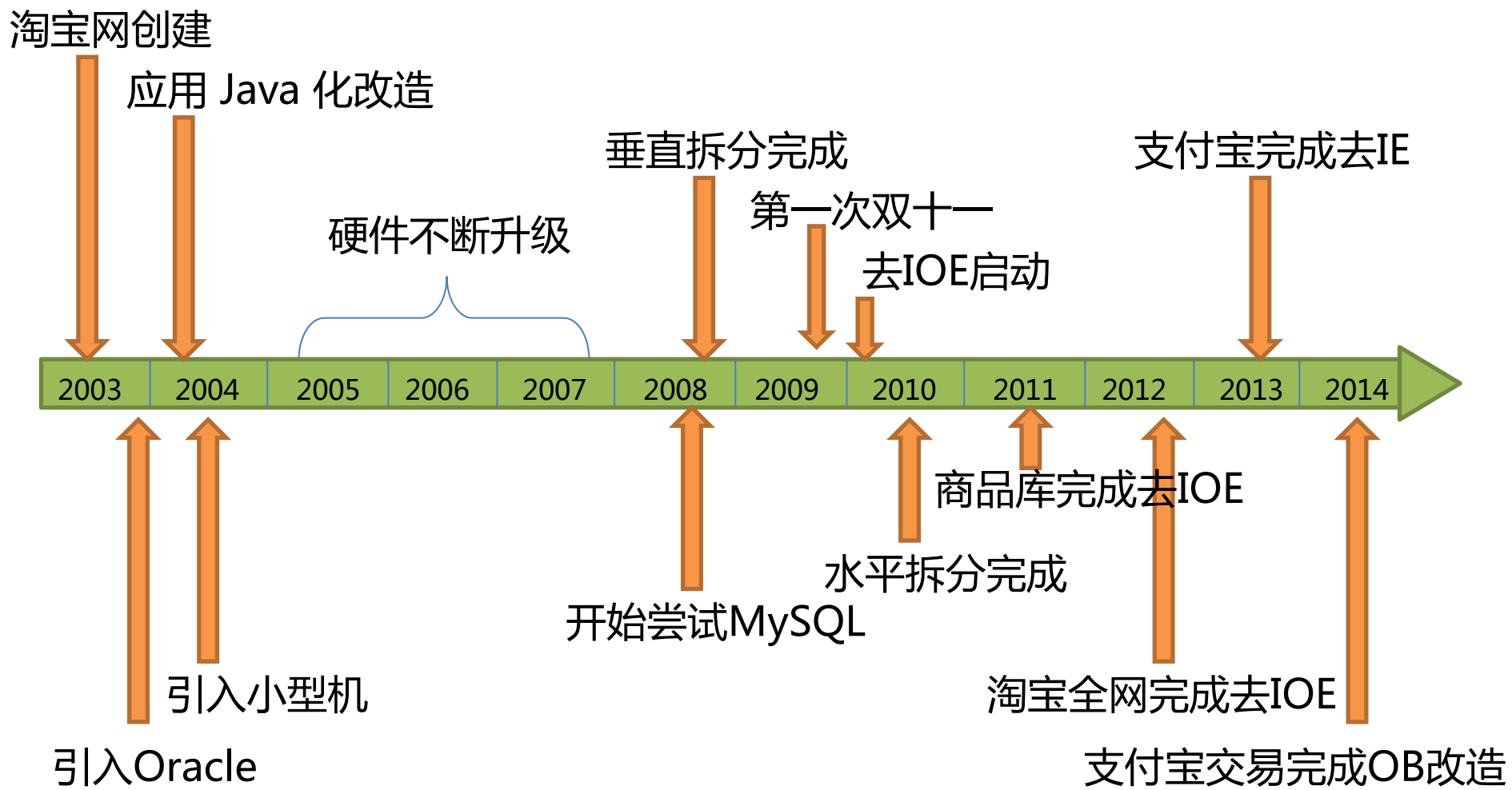
- 数据库限流 -- 第一次自己的命运自己掌握
- 热点更新优化 -- 定制优化热点商品减库存业务场景
- 线程池特性优化 -- 定制优化高连接数并发场景
- 专场7 数据库调优，阿里巴巴 章颖强

无限可能，尽在掌握

无冕之王-- AliSQL



12年历程回顾



新挑战 新机遇

➤ 全网架构

- 资源限制, 一个城市已经不能满足需求
- 容灾, 单地域机房风险

单元化

➤ AliSQL

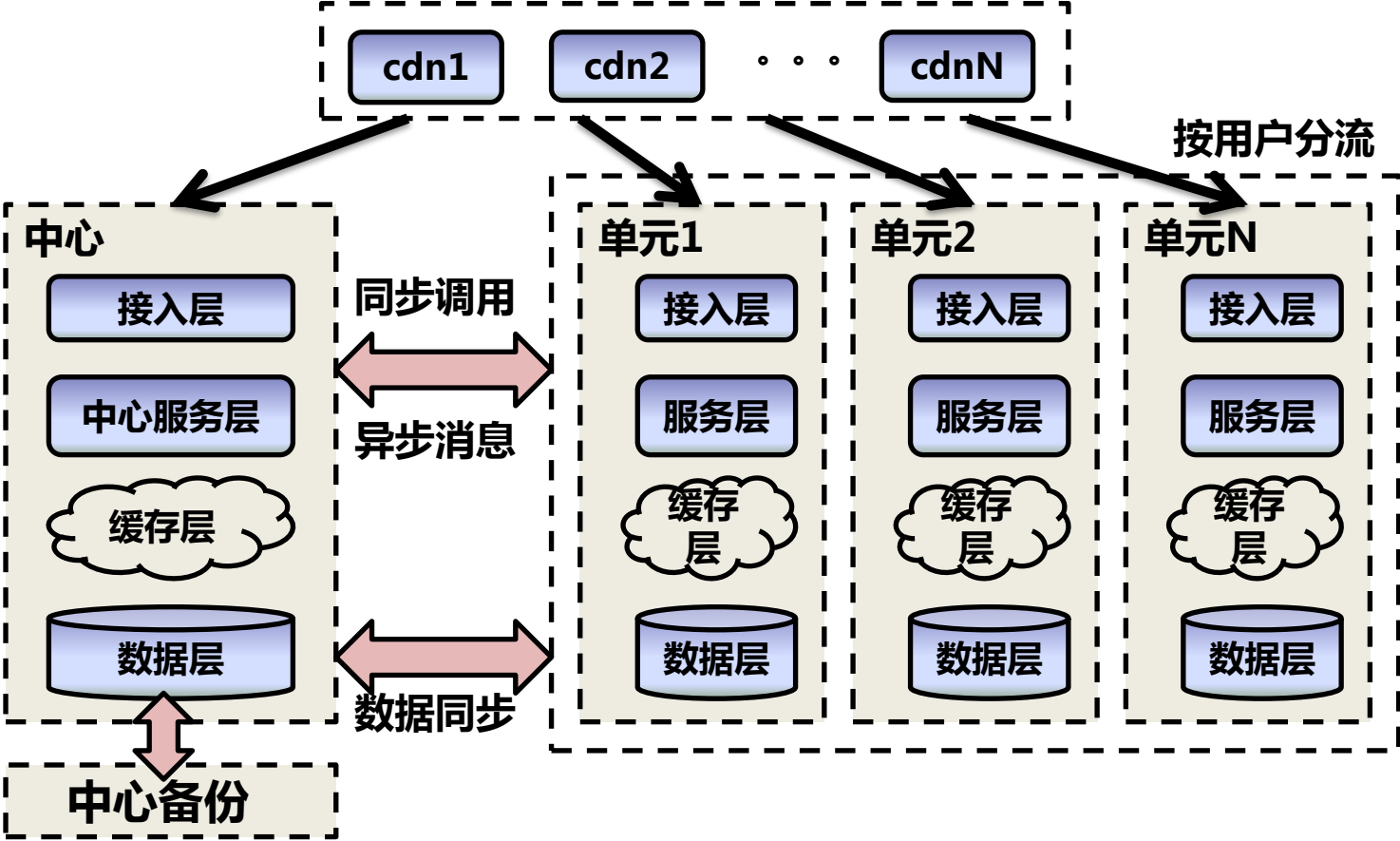
- 分表数量庞大
- 集群拆分接近极限

OceanBase

➤ 业务开发复杂度

- 路由, 关联, 聚合, 订正

新挑战 新机遇 -- 单元化



单元化

新挑战 新机遇 -- 单元化

对应用挑战的及解法

➤ 延迟同步调用的影响

- 同一机房0.2ms
- 同一城市1~5ms
- 跨城市5ms~100ms
- 几百次调用吞吐量下降
- 数据复制延迟问题

单元内封闭

新挑战 新机遇 -- 单元化

对数据库的挑战及解法

➤ 中心及各个单元间数据拆分原则

数据买家维度拆分

➤ 数据质量保障

● 数据多点写入风险

单元间DRC数据复制

● 数据复制一致性，正确性保障

● 专场13 阿里巴巴 钱在晨

新挑战 新机遇 -- 单元化

单元化效益

➤ 稳定性

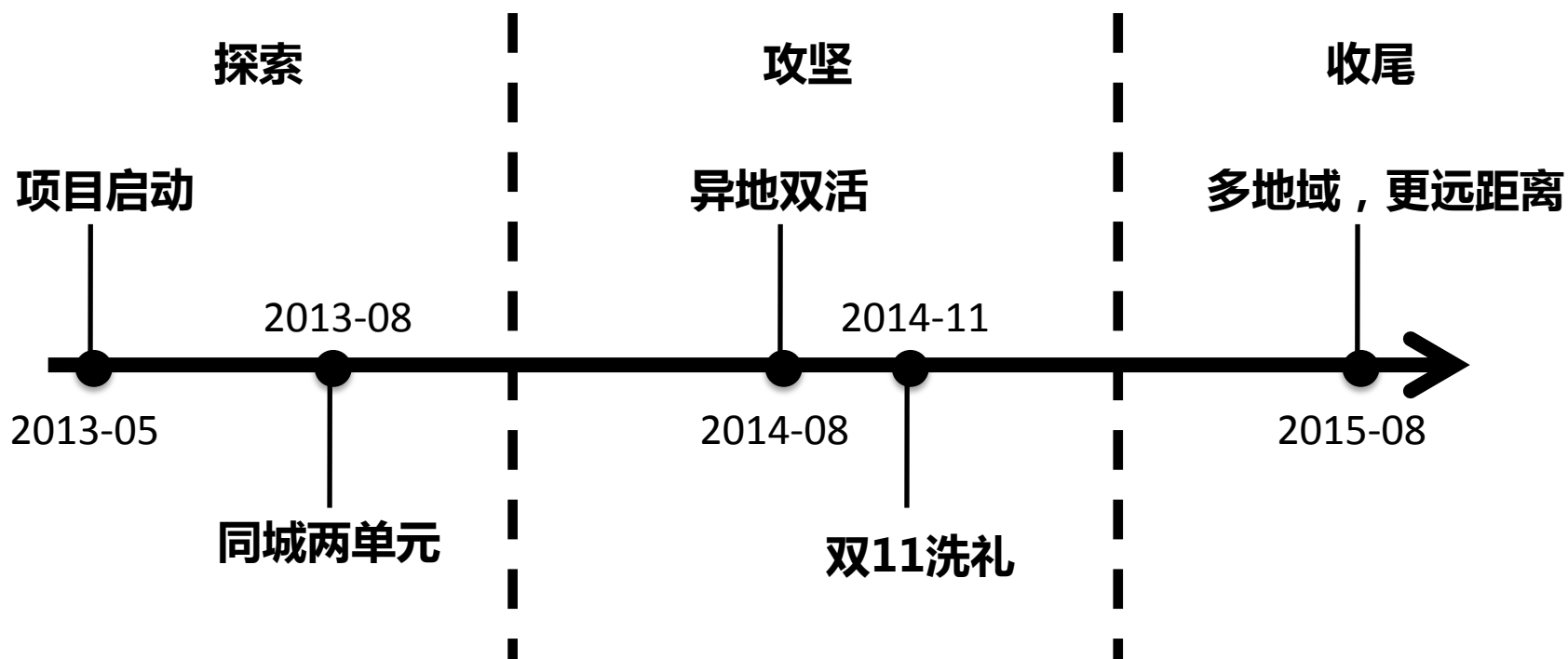
- 变更范围
- 故障恢复时间

➤ 伸缩能力

- 摆脱机房限制
- 伸缩规模再次增强

新挑战 新机遇 -- 单元化

横跨三年的项目



新挑战 新机遇

➤ 全网架构

- 资源限制, 一个城市已经不能满足需求
- 容灾, 单地域机房风险

单元化

➤ AliSQL

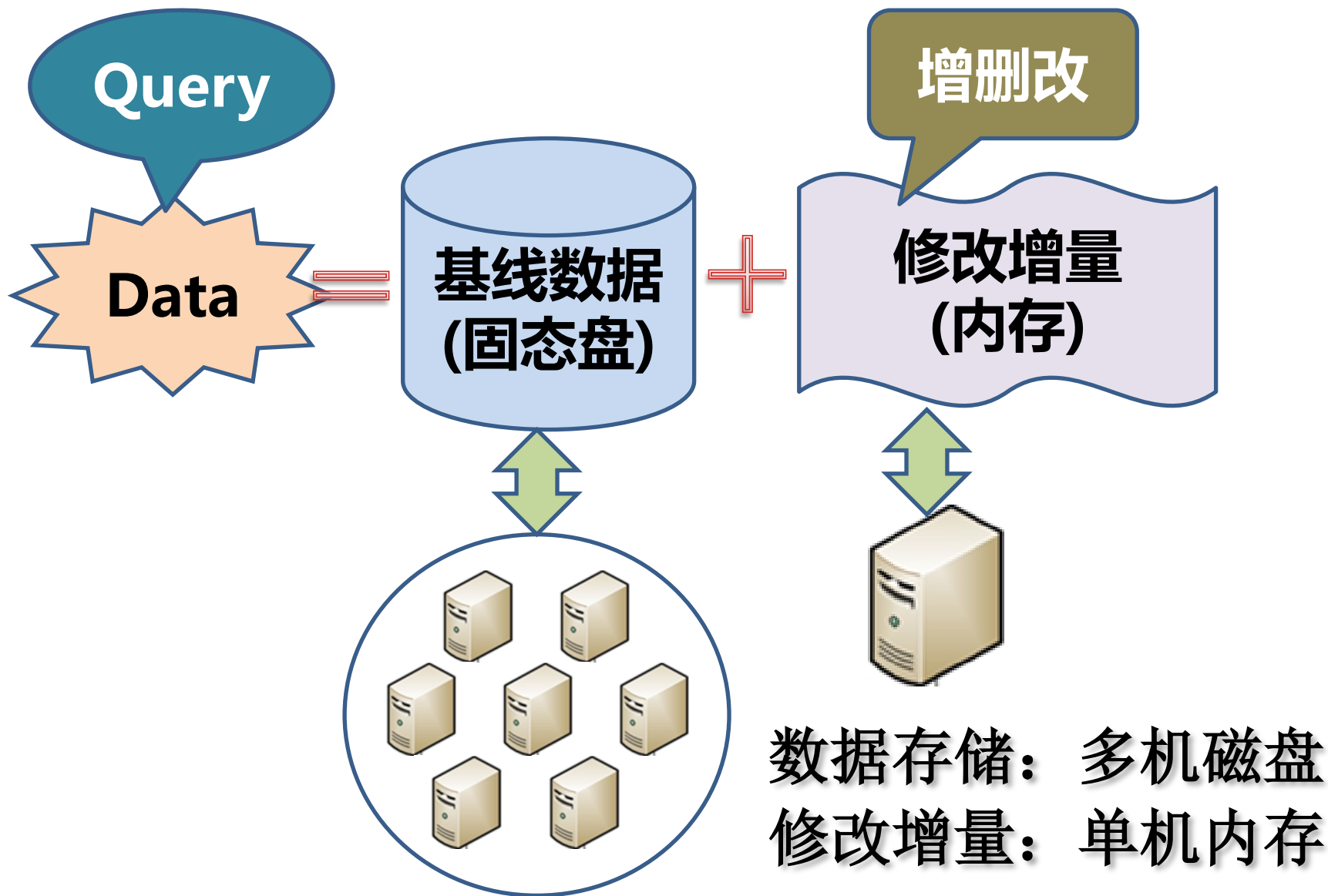
- 分表数量庞大
- 集群拆分接近极限

OceanBase

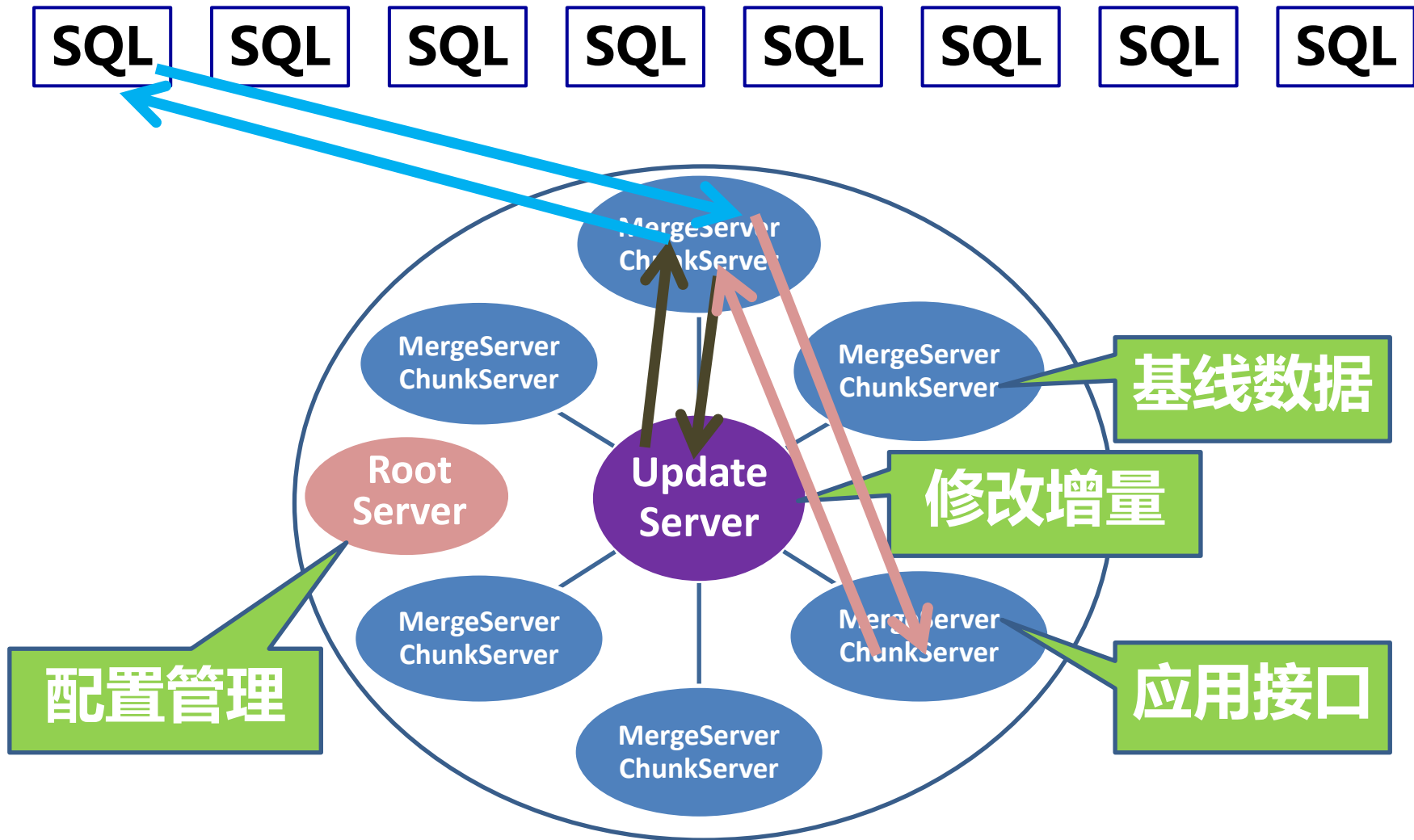
➤ 业务开发复杂度

- 路由, 关联, 聚合, 订正

新挑战 新机遇 -- OceanBase



新挑战 新机遇 -- OceanBase



新挑战 新机遇 -- OceanBase

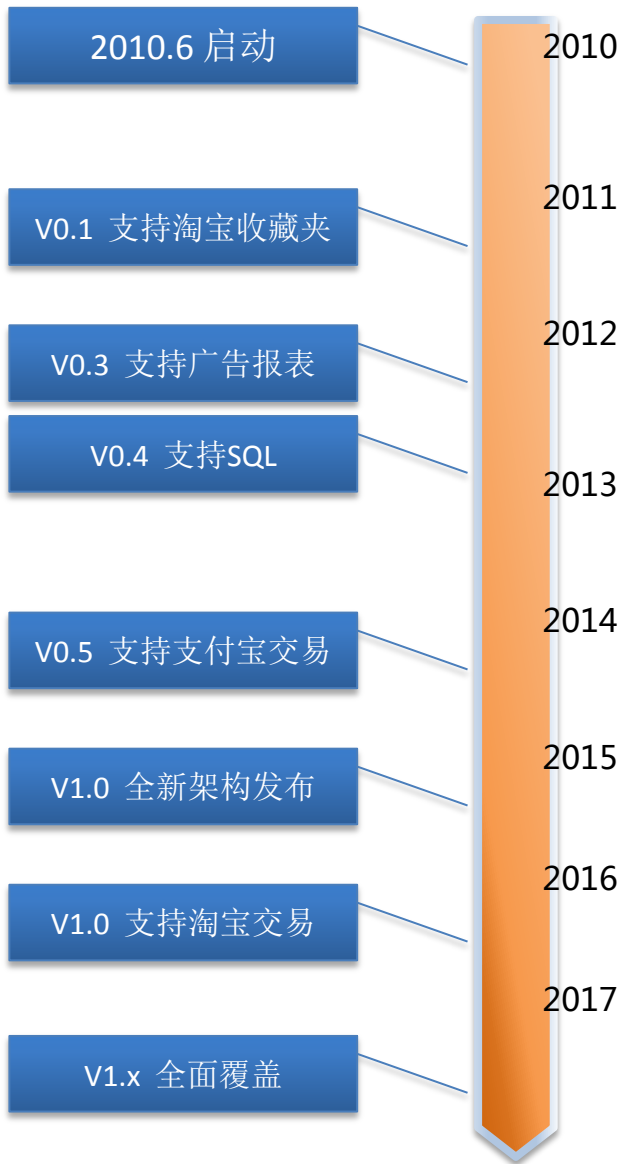
- 基于Paxos的高可用方案
 - 以不可靠部件提供可靠服
 - 小于半数的分区容忍性
 - 较高的可用性(最大35s不可用)
 - 强一致性

角色	异常描述	时间/影响
RootServer	宕机/程序异常退出	35S/不可用
UpdateServer	宕机/程序异常退出	25S左右/不可用
MergeServer	宕机/程序异常退出	1分钟/少量读超时
ChunkServer	宕机/程序异常退出	1分钟/少量读超时
主集群不可用	如出现机房断网情况	35s/不可用

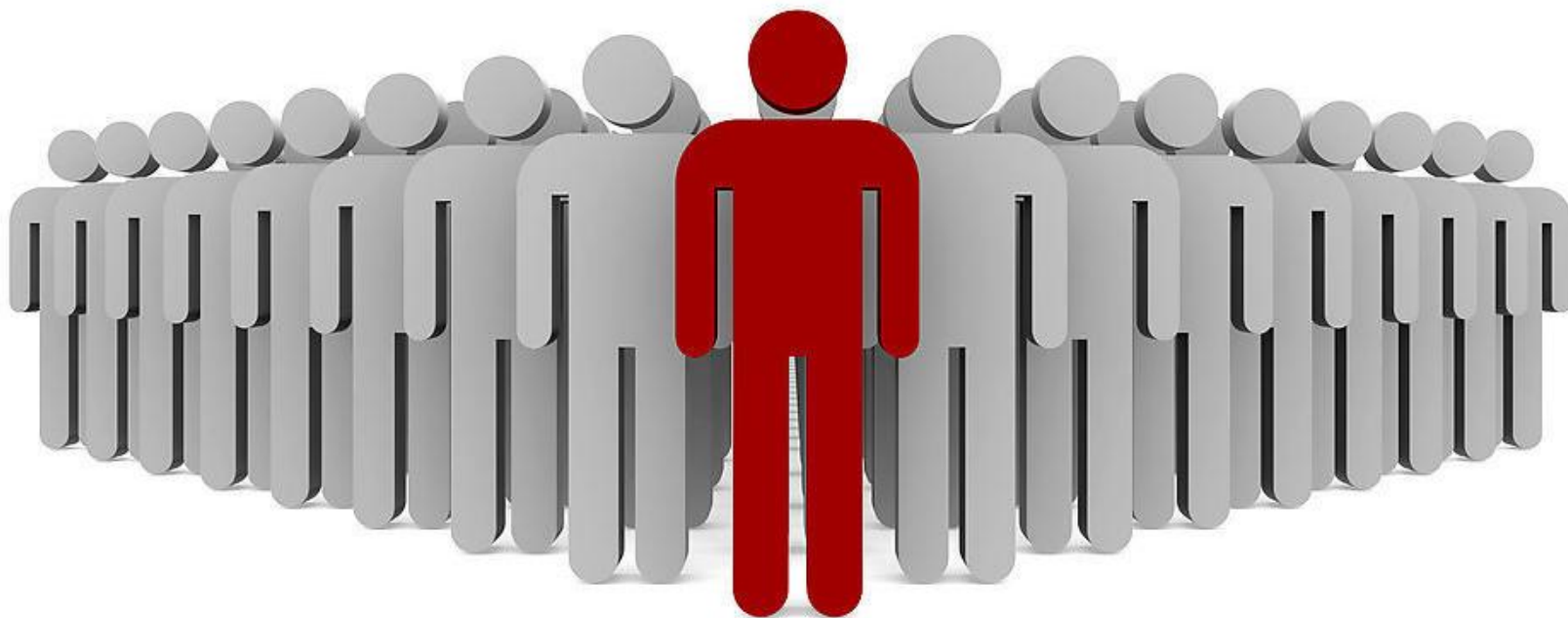
新挑战 新机遇 -- OceanBase



2014.2(v0.5)



欢迎你的加入 携手共创未来



IT168

ChinaUnif

ITPUB

THANKS