CHINA DATA ANALYST SUMMIT

CDA 数据分析师 www.cda.cn

面向海量结构化数据管理的高可用分布式数据库系统

演讲人：宫学庆
xqgong@obase.com.cn

跨界互联 数聚未来

第四届中国数据分析师行业峰会

CHINA DATA ANALYST SUMMIT

北京 中国大饭店 2017.07

OBASE分布式数据库

OBASE

- **采用分布式架构，数据多副本冗余存储**
  - 基于Raft协议实现集群的高可用，支持跨机房多活部署；
  - 无需用户定义分库分表，数据表自动按主键划分为多个子表；
  - 数据冗余存储在集群中，能够自动迁移；
- **面向OLTP应用，支持Read Commited事务隔离级别**
  - 通过内存事务引擎实现集中式写事务，避免分布式事务；
  - 支持可扩展的数据存储和分布式读事务；
- **支持SQL查询，SQL语法遵循SQL92标准；**

SQL语言支持

数据类型

- Bigint
- Int
- Varchar
- Bool
- Datetime
- Timestamp
- Date/Time
- Double/Real
- Decimal

函数

- 类型转换
- 字符串处理
- 日期时间处理
- 聚集函数
- 窗口函数
- ……

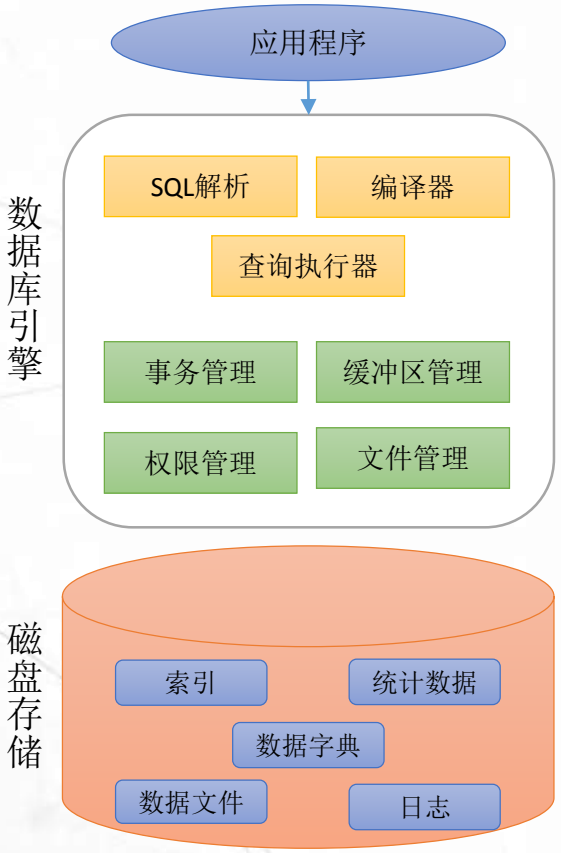
DDL&DML

- Database
- Table
- Index
- Sequence
- Insert
- Replace
- Update
- Delete
- Select
- ……

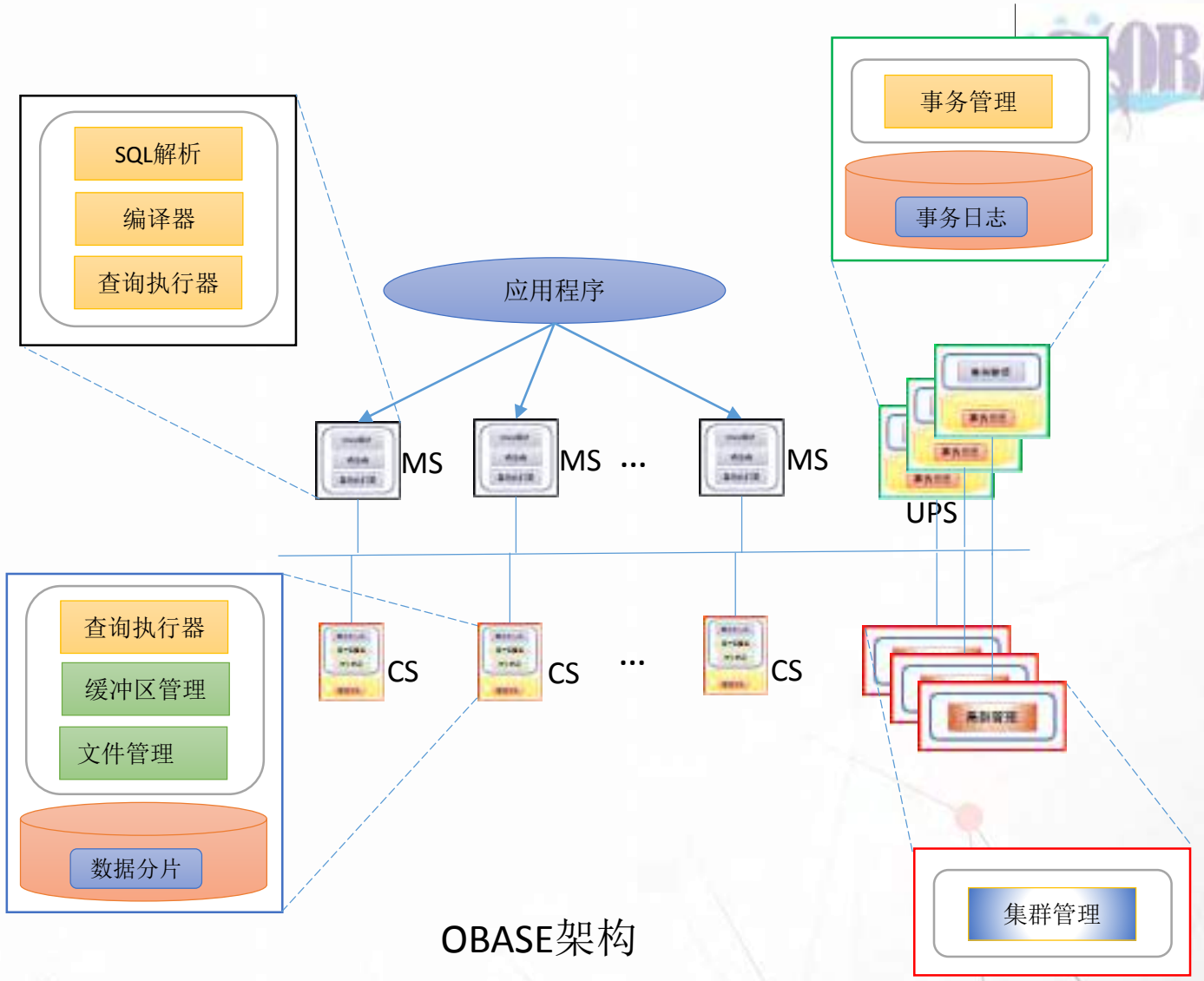
事务&管理

- Start transaction / Begin
- Commit
- Rollback
- Create user
- Drop user
- Grant
- Revoke
- Prepare
- Explain

基本原理



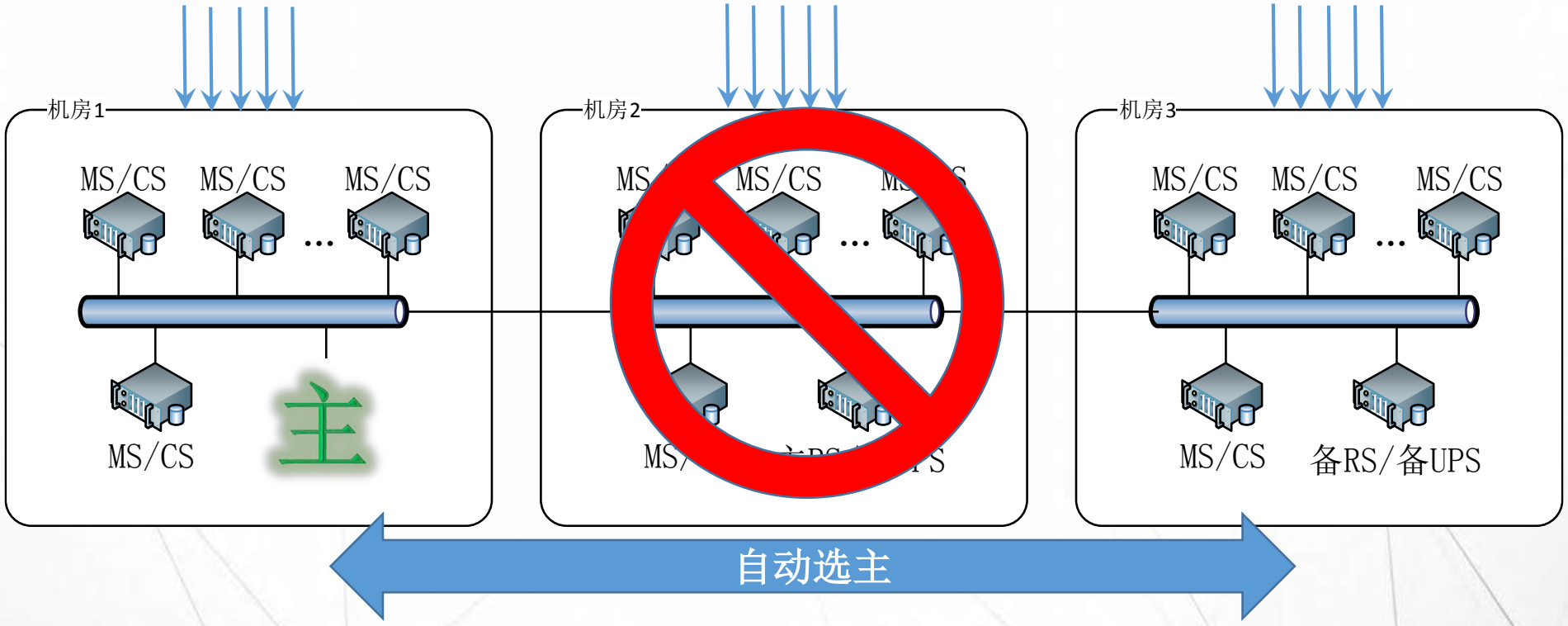
传统RDBMS架构



OBASE架构

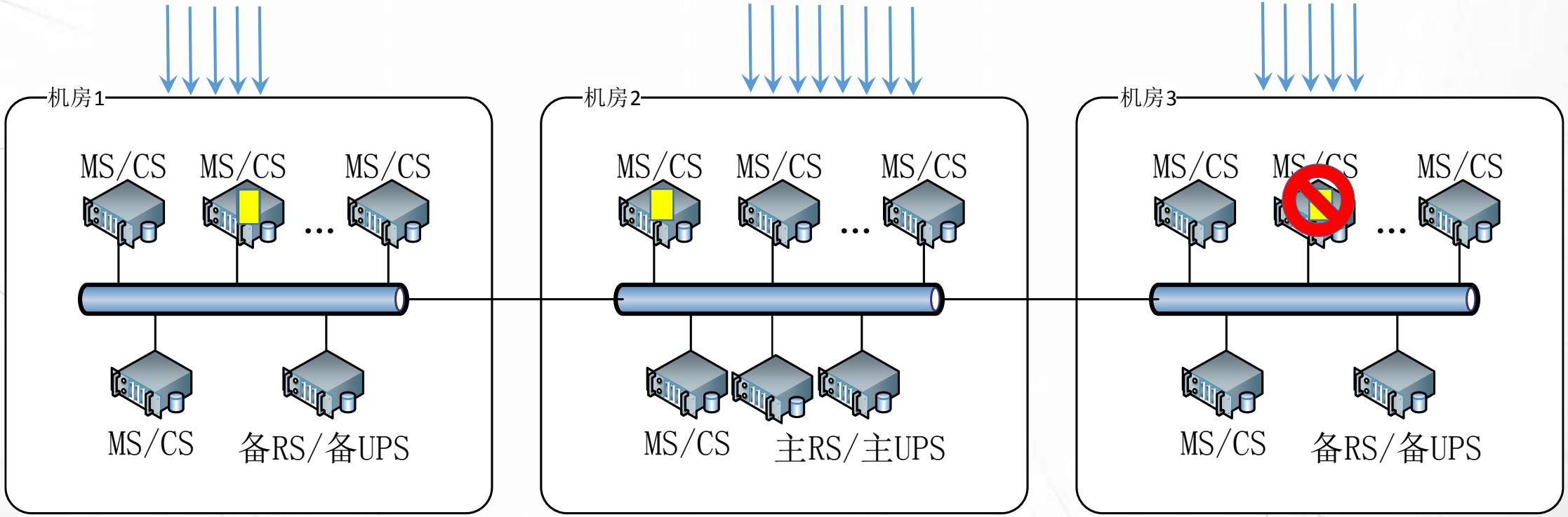
多活部署&跨机房容灾

- Root Server：实现数据分布、副本管理、集群管理等功能，自动选主；
- Update Server：内存数据库引擎，实现事务处理，存储增量数据，由主RS选主；
- Chunk Server：分布式数据存储服务，存储基准数据；
- Merge Server：查询处理服务，实现SQL解析、生成并执行查询计划；



在线扩容&自动容错

- 随着业务负载的增加，可以随时向集群中添加MS/CS，分担负载；
- 系统会自动进行数据分片的迁移，以实现负载均衡；
- 当节点出现故障时，系统自动计算数据分片的副本数，根据需要生成新副本；

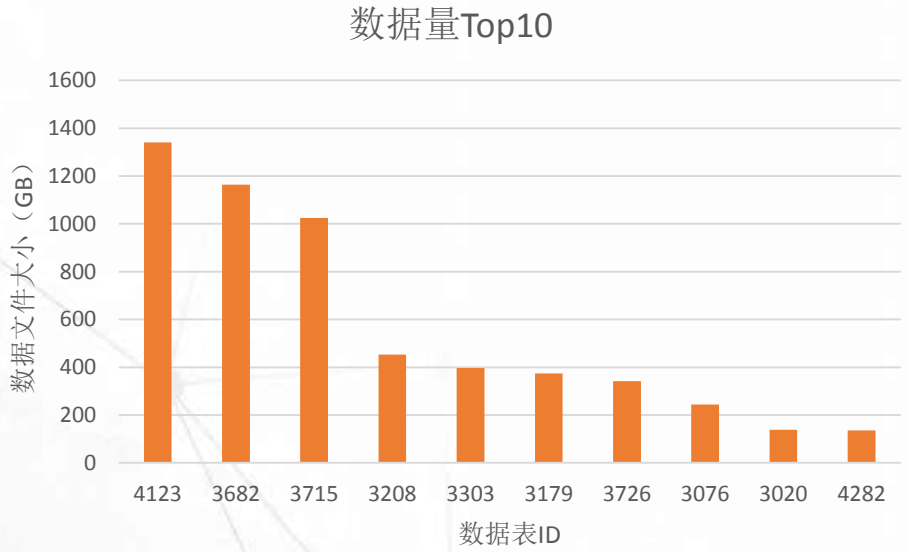


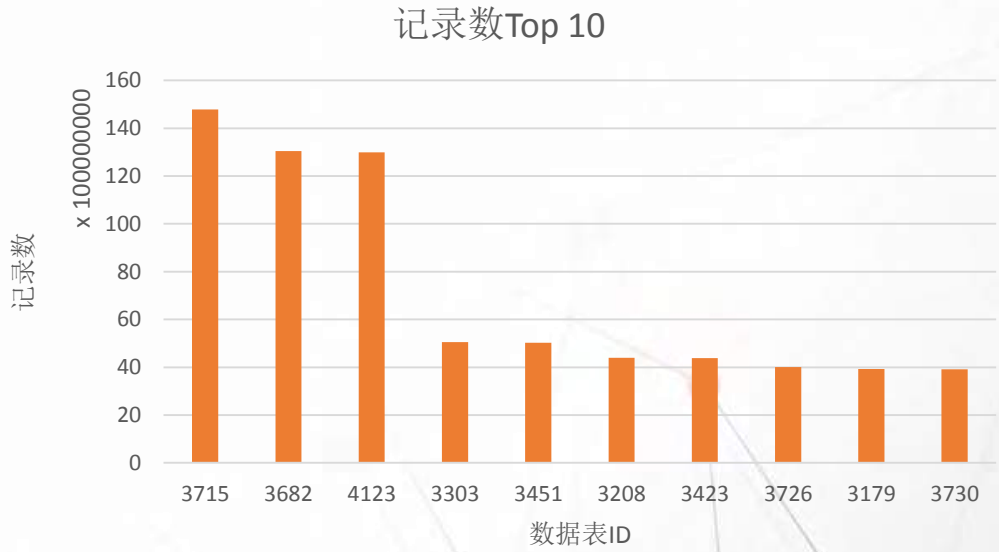
数据安全性

- **数据文件在集群内多副本冗余存储**
  - 缺省为3个副本，自动做负载均衡；
  - 在<3台同时损坏的场景下不丢数据、不停服务；
  - 机房断电后，重启时通过自动回放日志恢复内存数据；
- **管理节点基于Raft协议实现自动选主**
  - 与传统数据库采用相同的WAL日志策略；
  - 主机日志同步到半数以上备机时才可以提交事务；
- **支持在线灰度升级和扩容**
- **支持在线更换磁盘**

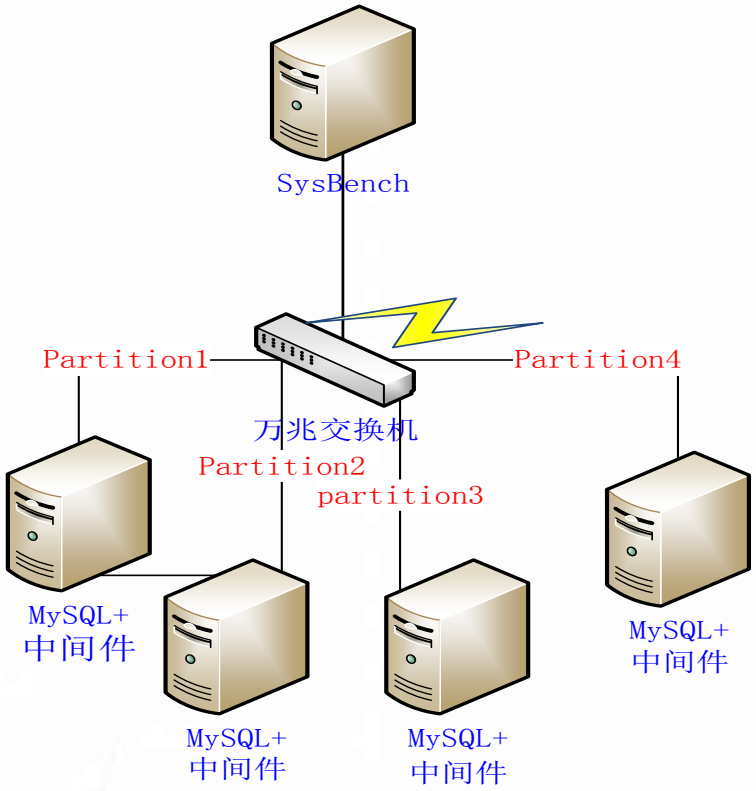
管理海量结构化数据

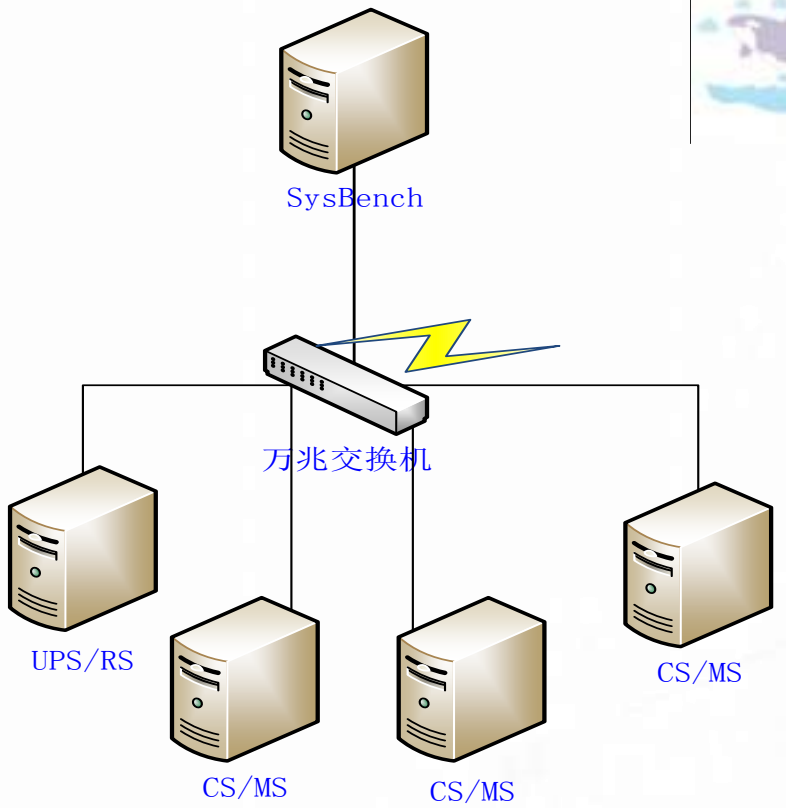
- 理论上单表记录数没有上限；
- XX银行历史库系统2014年10月上线；
  - 同城异地双活部署（单集群21台x86服务器）
  - 40TB，单表最大140+亿记录/1.3+TB





SYSBENCH性能测试

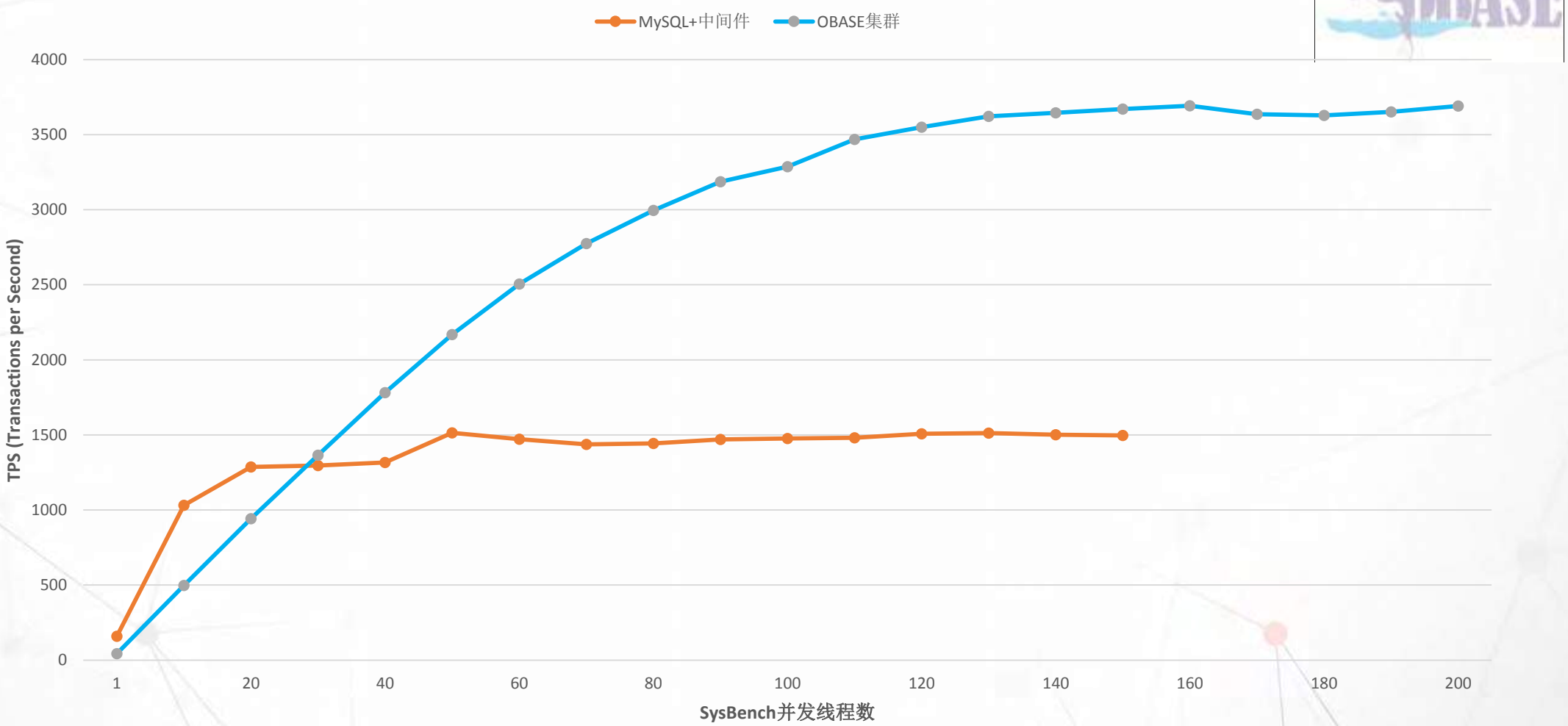




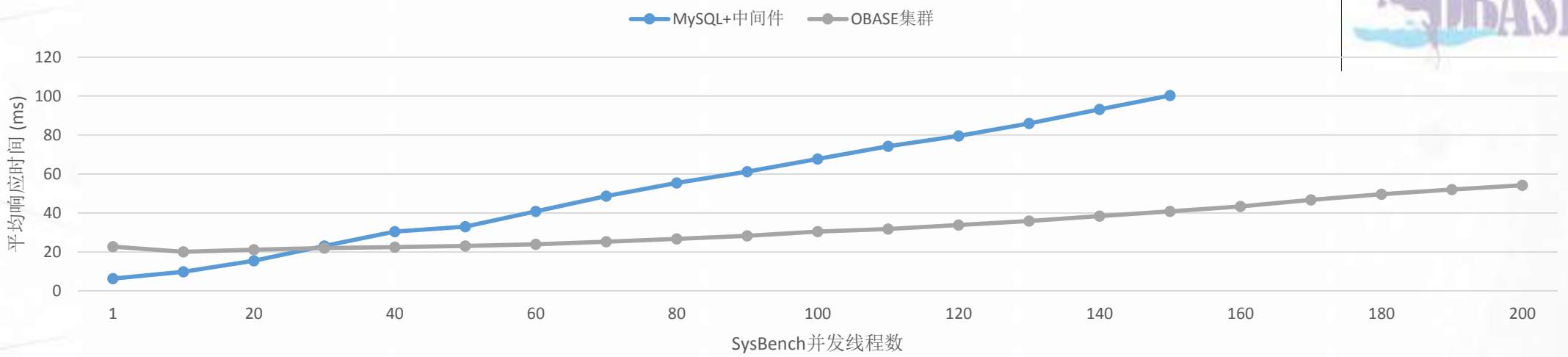
环境配置

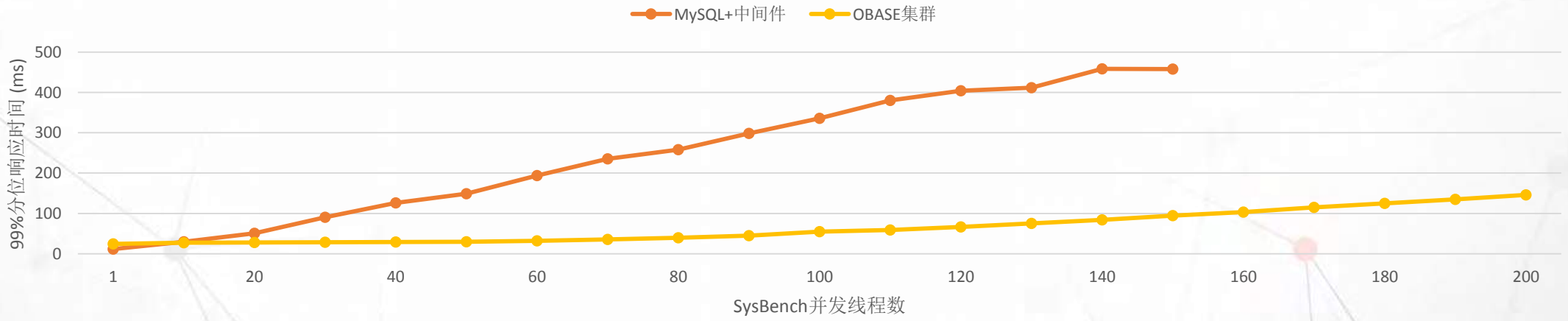
- CPU：E5-2650v3*2　　内存：512GB
- SSD：600GB SATA SSD*6SAS　　磁盘：600GB 10K *2/阵列卡
- 数据集：10,000,000记录/表*10

30分钟TPS对比


MySQL+中间件
OBASE集群
TPS (Transactions per Second)
4000
3500
3000
2500
2000
1500
1000
500
0
1
20
40
60
80
100
120
140
160
180
200
SysBench并发线程数

30分钟响应时间对比


MySQL+中间件
OBASE集群
平均响应时间 (ms)
120
100
80
60
40
20
0
1
20
40
60
80
100
120
140
160
180
200
SysBench并发线程数

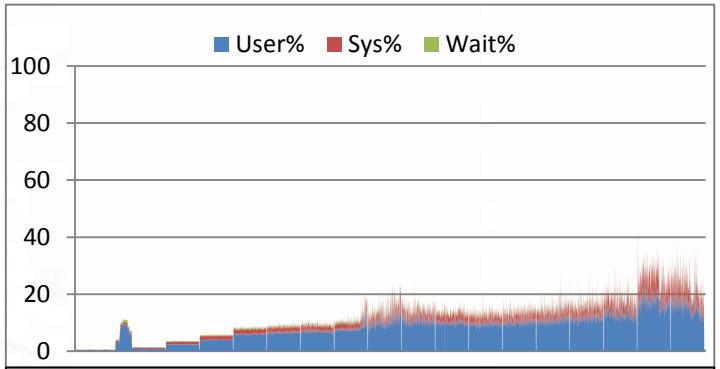

MySQL+中间件
OBASE集群
99%分位响应时间 (ms)
500
400
300
200
100
0
1
20
40
60
80
100
120
140
160
180
200
SysBench并发线程数

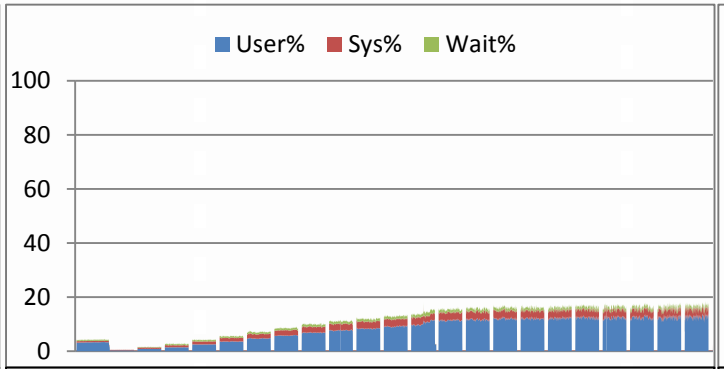
系统资源消耗分析

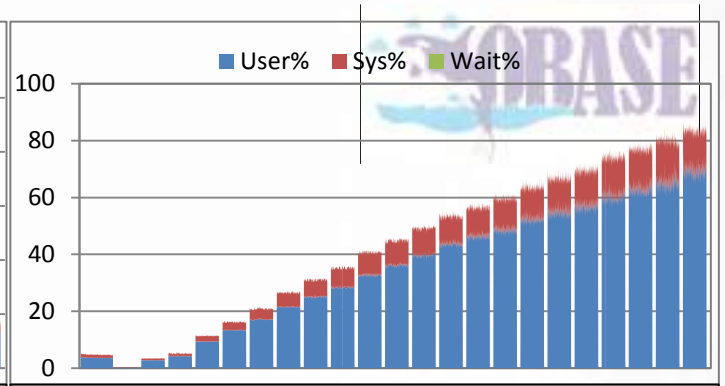
CPU使用率

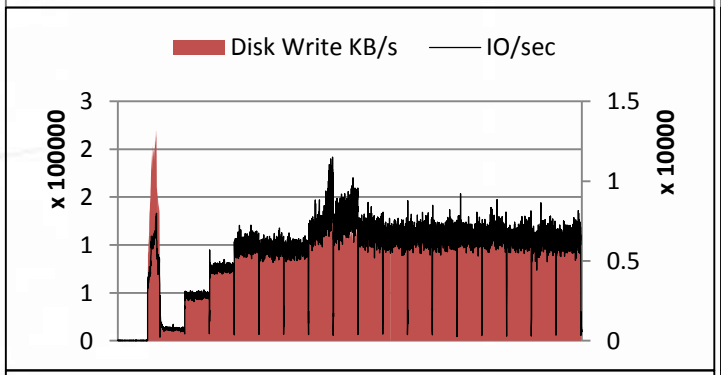
SSD磁盘IO

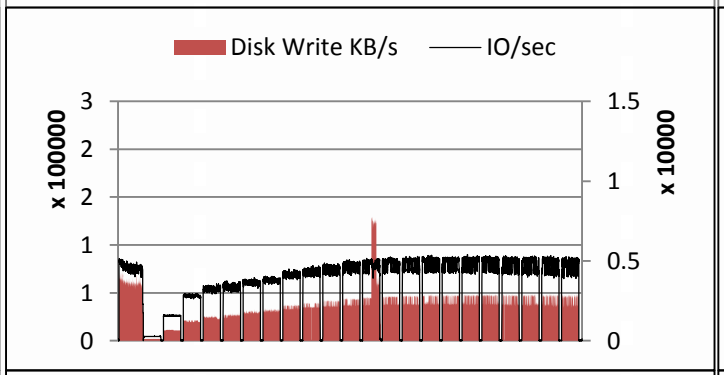
网络IO

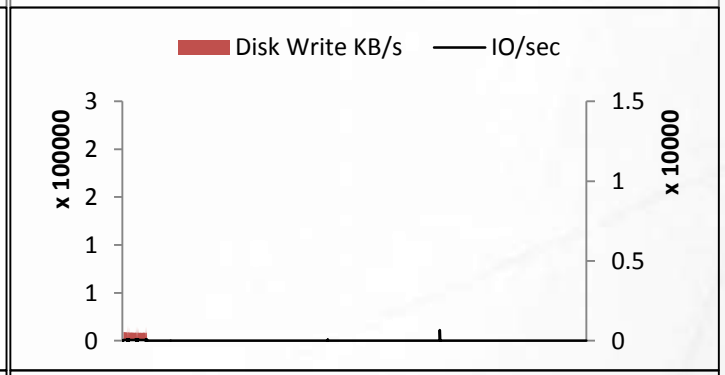


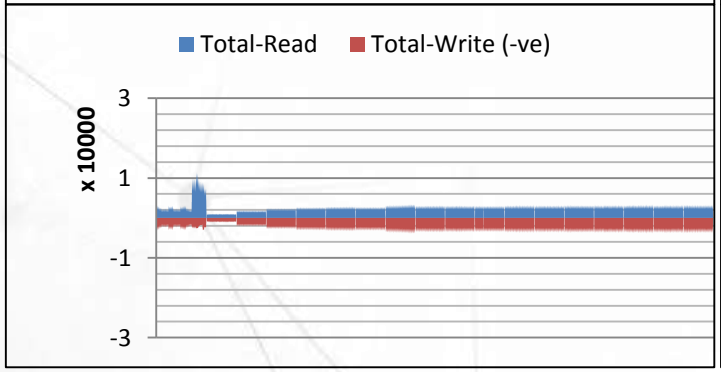


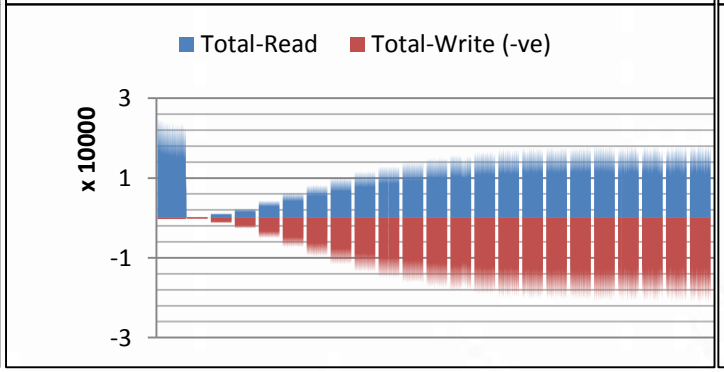










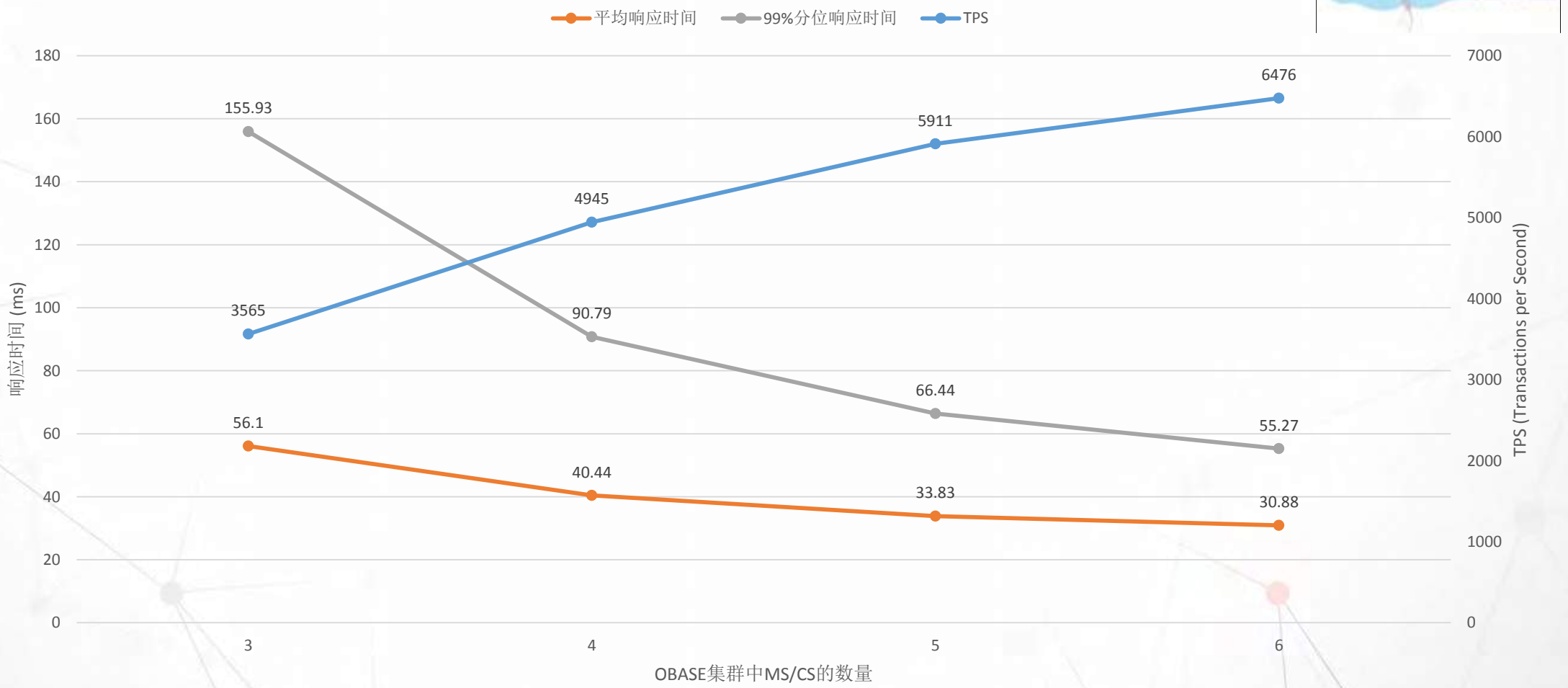


MySQL节点

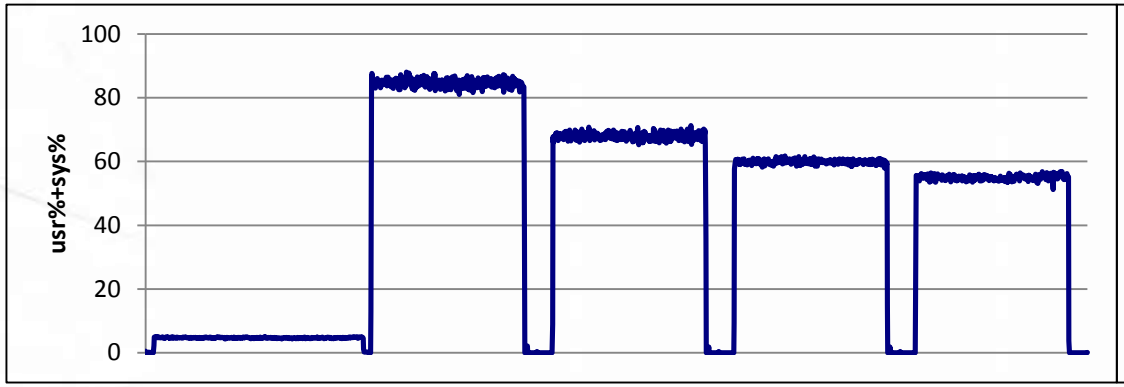
UPS/RS节点

MS/CS节点

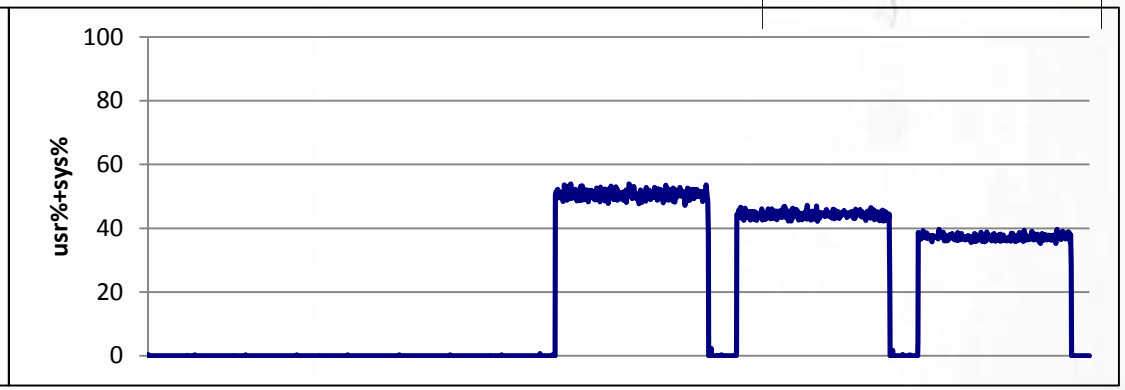
OBASE集群的扩展性



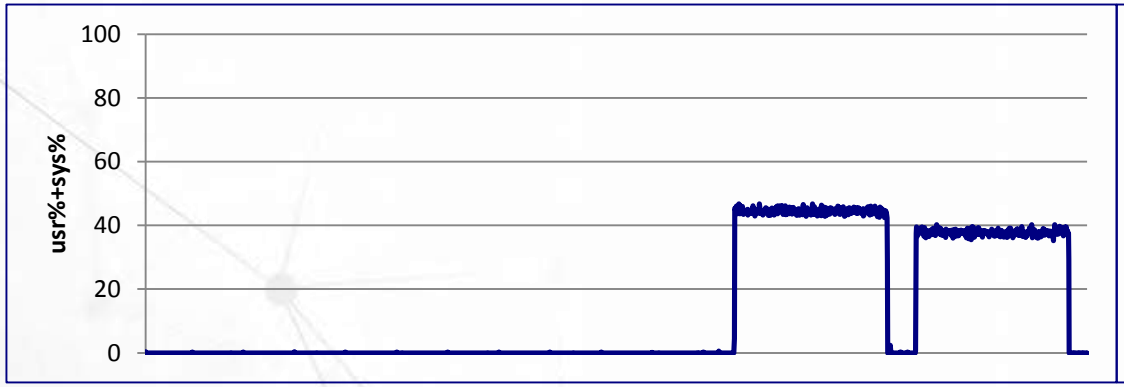
CPU使用率



原有MS/CS节点



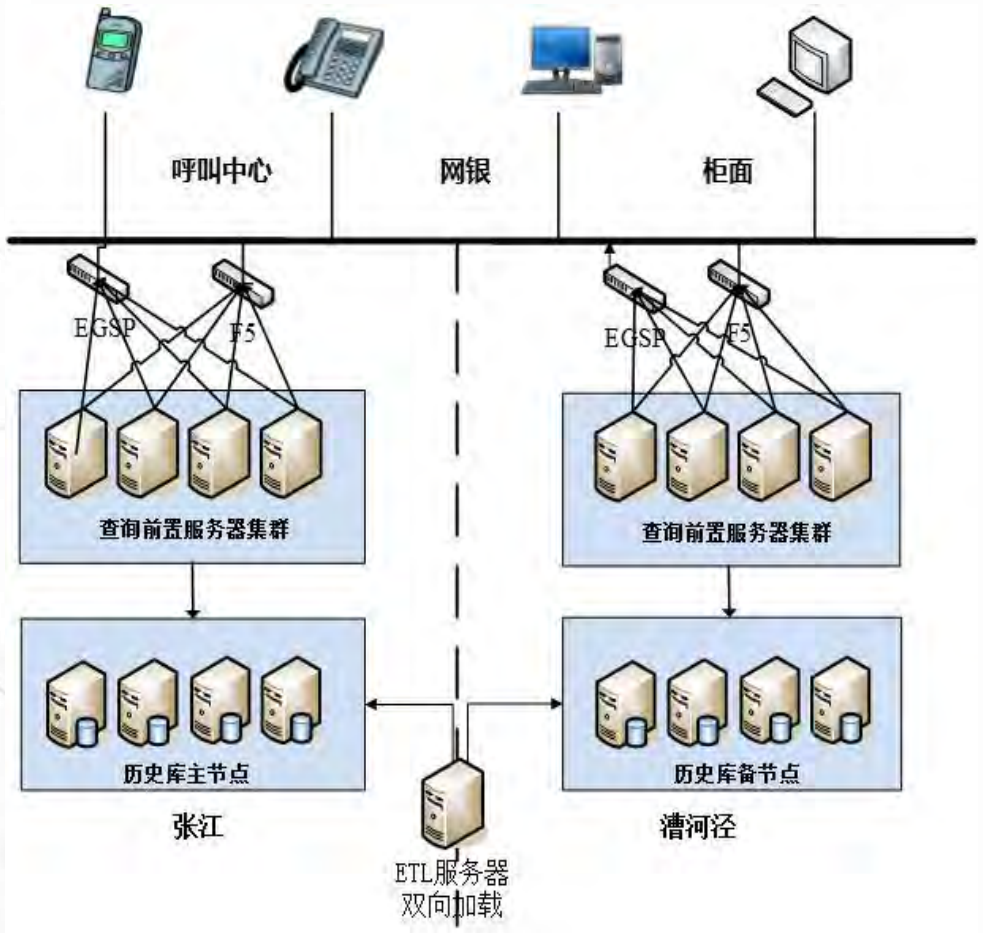
新增第1个MS/CS节点



新增第2个MS/CS节点

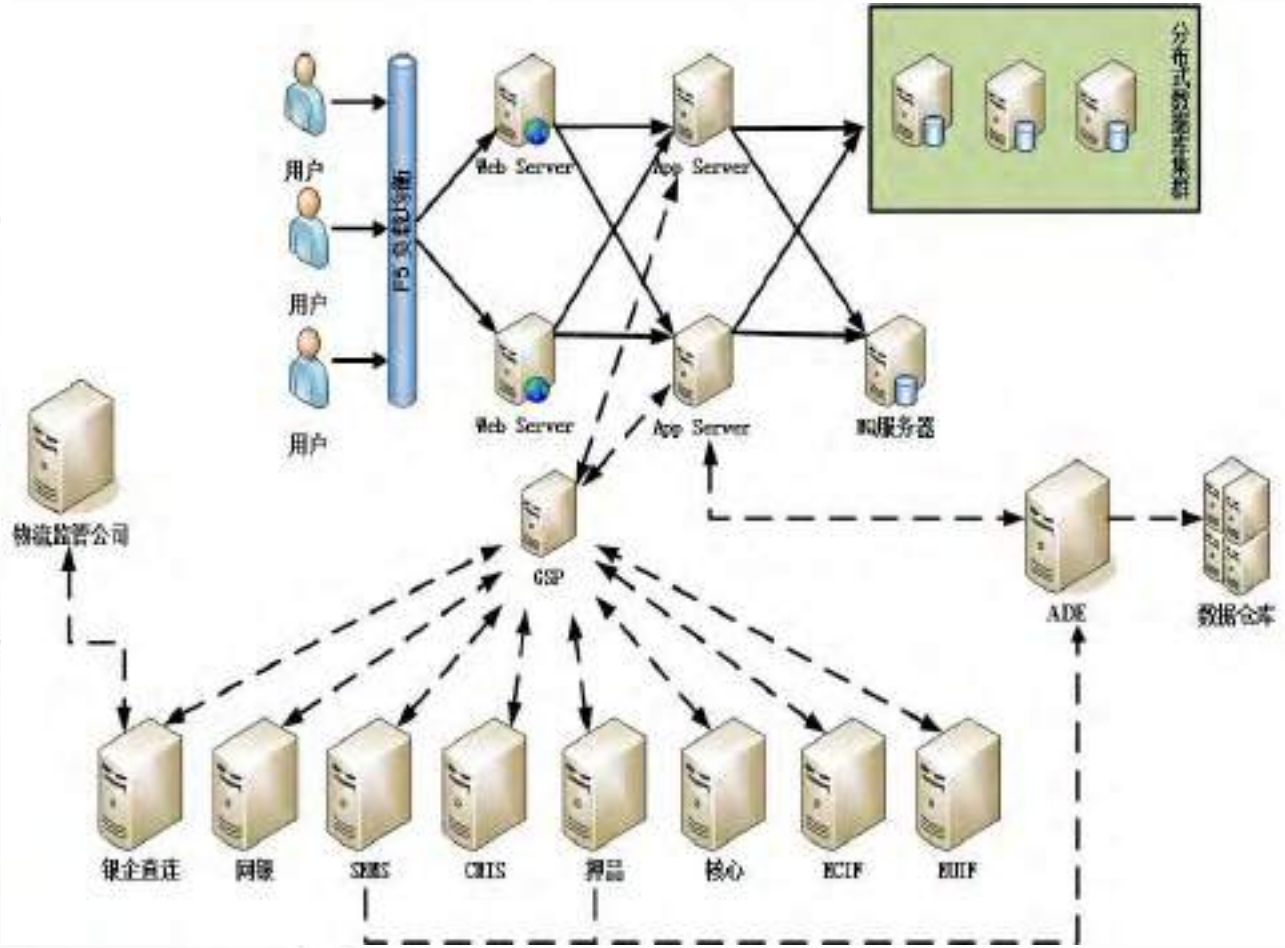
新增第3个MS/CS节点

金融历史库系统



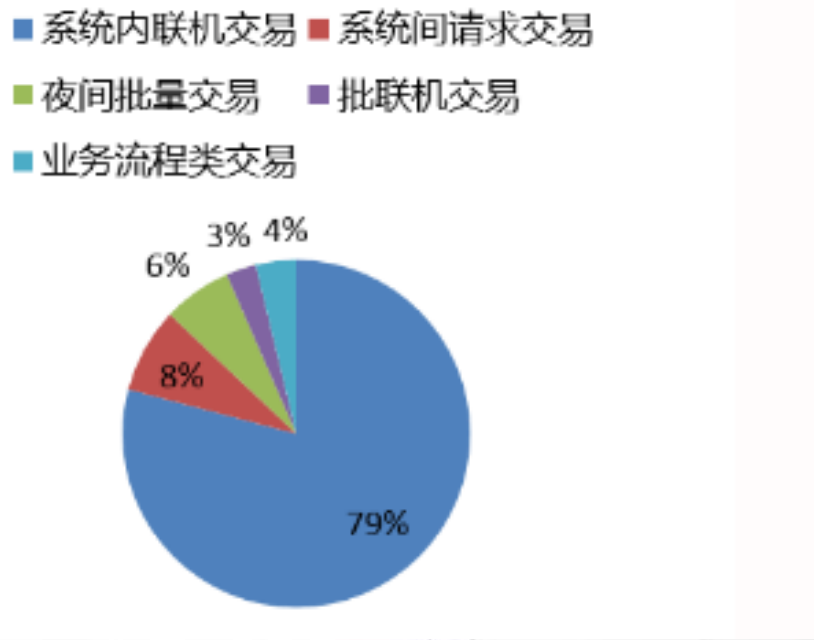
- 在生产环境中稳定运行30个月
- 日均交易量300万，数据量150TB
- 最大单表超过140亿条记录
- 大并发查询200笔/秒的情况下，平均响应时间小于300ms

新型金融资产管理系统

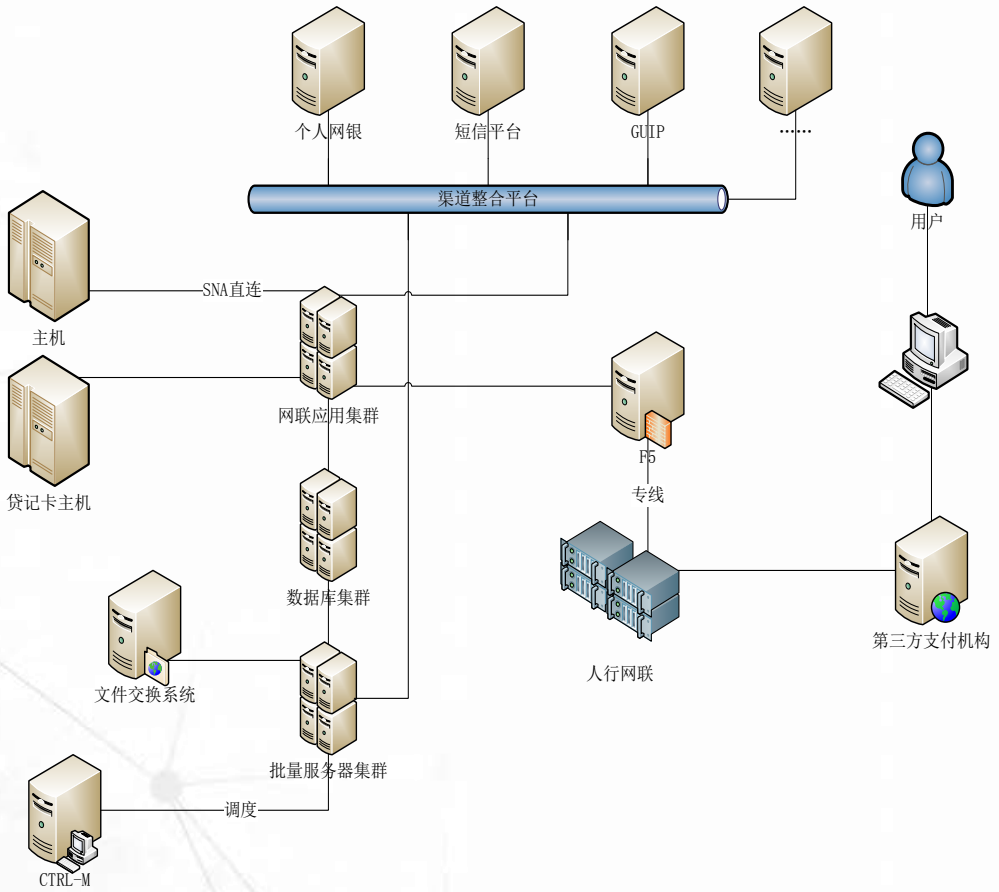


系统各模块功能菜单分类（共计375个）

系统交易分类（共1249个）

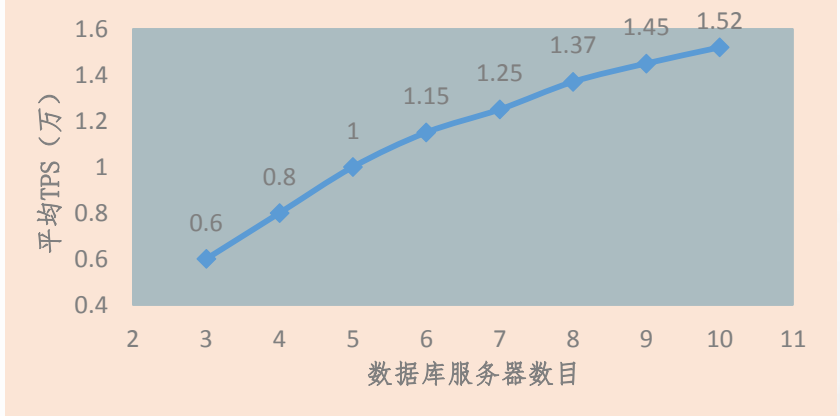


网联支付系统



项目性能要求
• 高峰处理能力：12000tps

obase联机压力关键支付交易测试	
服务器硬件	20核/CPU 2.3GHZ 内存 500GB
应用服务器数量	5台
压测交易	支付交易（8条SQL语句，1个新增，2个更新，5个查询）
初始数据描述	3.5亿交易流水，5000万签约协议



经济效益情况

经济效益	新/老历史库	节省
软硬件成本（万元）	1196 / 4908	3712
运维成本（万元/年）	10/100	90

经济效益	新/老供应链	节省
软硬件成本（万元）	300 / 500	200
运维成本（万元/年）	10 / 30	20

OBASE数据库的特点

- 支持海量结构化数据存储和管理；
- 支持高并发的事务处理；
- 基于PC服务器集群，拥有成本低；
- 分布式架构，具有良好的扩展性和可用性；
- 原生开发团队，自主可控，原生技术支持；
- 采用关系模型，支持SQL语言，开发和使用简单；
- 多活部署、在线扩容、容灾，易于运维；
- 金融行业应用案例；

上海丛云信息科技有限公司（简称“丛云科技”）2015年3月注册于上海张江高科技园区，是一家专注于新型数据库系统研发的技术型企业。公司以国内高校的科研力量为依托，面向大数据量和高负载的企业级应用需求，致力于自主可控分布式数据库系统的研发和应用。目前，公司所研发的数据库系统已经在大型国有银行、大型物流企业中得到了应用，成功替代了国外厂商的商业数据库产品。

contact@obase.com.cn