2016

数据库云企业最佳实践

CONCENTS

PART 01

以前的故事

传统运维之殇

睡个好觉不容易

高成本，低收益

52%

大量的硬件采购成本

36%

高额软件费用

20%

每年20%的硬件维保

传统运维之殇

告别复杂繁冗的架构

传统复杂的IT架构已经越来越难以管理，同时懂小型机技术、PC server技术、存储技术数据库技术、网络技术的技术人员凤毛菱角。而复杂结构之间的误操作甚至是错误却在所难免，可以说有时我们管理的不是一个系统，而是一枚定时炸弹。

告别低效的系统应用

一些系统未必需要高级资源，现实告诉我们由于缺乏有效的规划，导致大量的资源浪费，有些系统的资源占用率很低。

做好成本控制更省钱

成本控制，不仅从企业成本考虑，

保护好系统绝对安全

系统的安全性和稳定性是系统中最为重要的核心指标，没有安全保障，任何方法都是空谈。

理想的IT架构

更省钱

更简单

更好用

更安全

再见小型机！再见存储！

存储业的发展
使单盘的SSD和FLASH已经超越了传统磁盘阵列

随着芯片行业发展PC Server的CPU已经和小机已经不相上下

X86架构加高速Infiniband网络使系统架构更简便

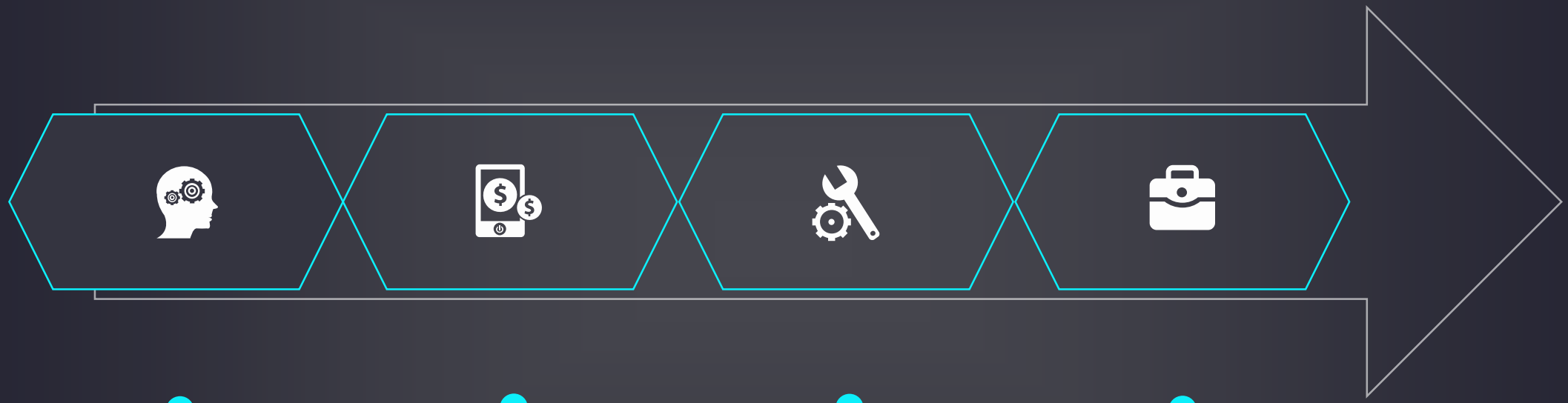
X86架构加
数据库双活
数据卫士保障
系统万无一失

再见！小机！再见！存储！

PART 02

现在的情形

我们的时间戳



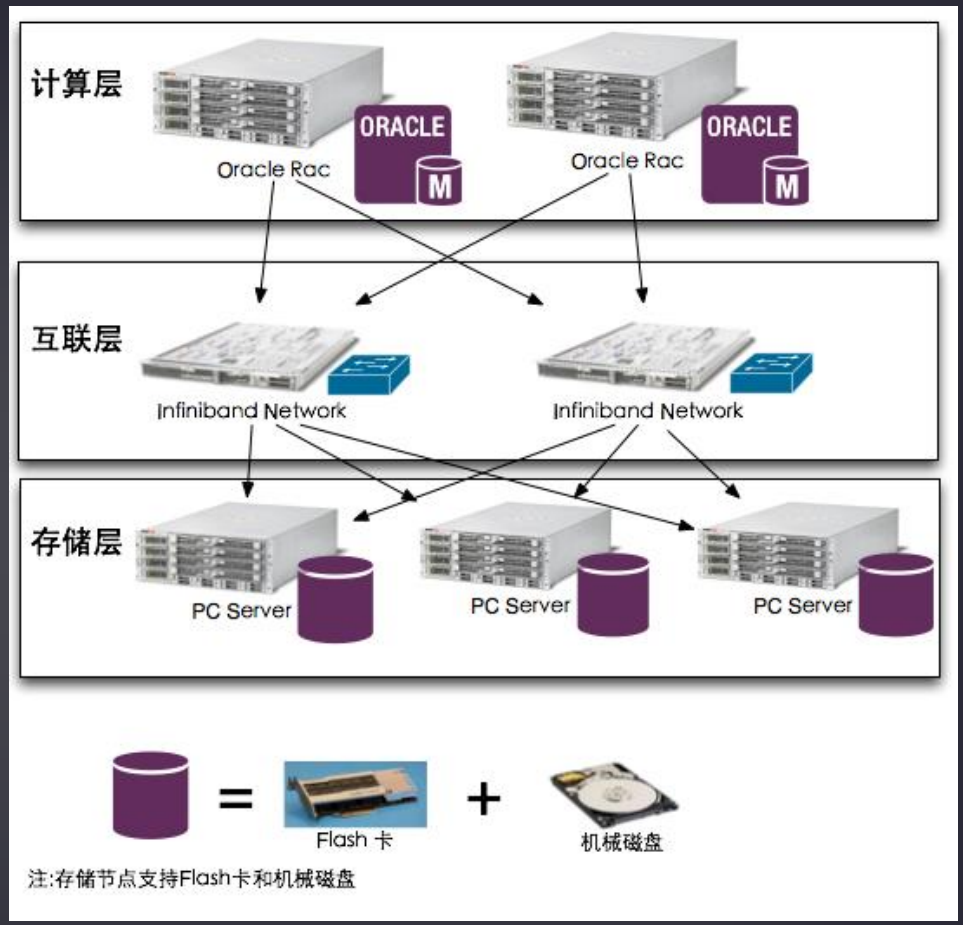
2013年7月，开展去IOE相关探索及技术研究，调研了阿里巴巴、oracle相关的产品

2013年12月，提出高性能数据库一体机/一体化解决方案

2014年2月，完成高性能数据库一体化解决方案验证及测试

2015年8月完成国网营销核心系统系统的迁移工作

我们的X86代替小型机方案



一体机解决方案的SWOT分析

S

高性能的IO吞吐能力

高性能Flash卡和SSD固态硬盘技术的大力发展以及高速存储网络的出现，使今天的PC Server有着恐怖的IO吞吐能力

W

X86服务器的不稳定性

X86 的架构的不稳定性是业界普遍批评PC Server的诟病，如何解决内核的不稳定性，是新架构能否实现的成败

O

国家战略大力推动自主研发

随着IT国产化运动的快速发展，拥有自主知识产权的国产IT产品将会在整个市场中占有很大的份额

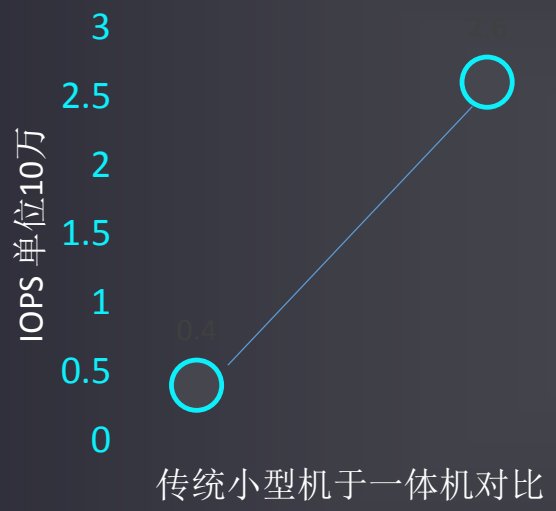
T

来自Oracle Exadata的威胁

Oracle公司强大的跨时代产品Exadata把Oracle数据库紧紧的握住。Oracle数据库的闭源致使Oracle自己的产品拥有超强的性能优势

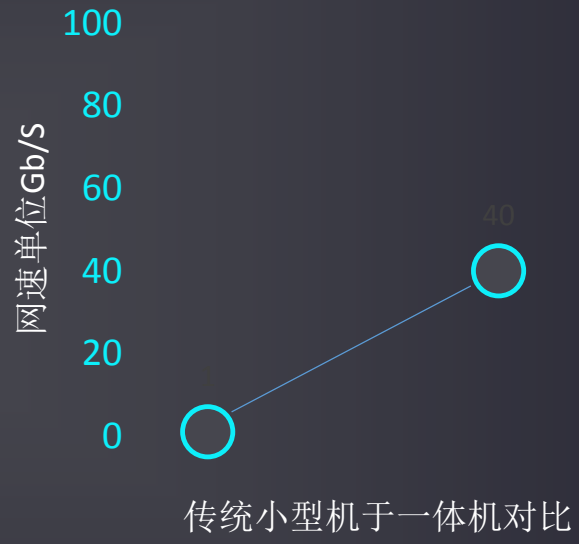
Strength 强大

IOPS的对比



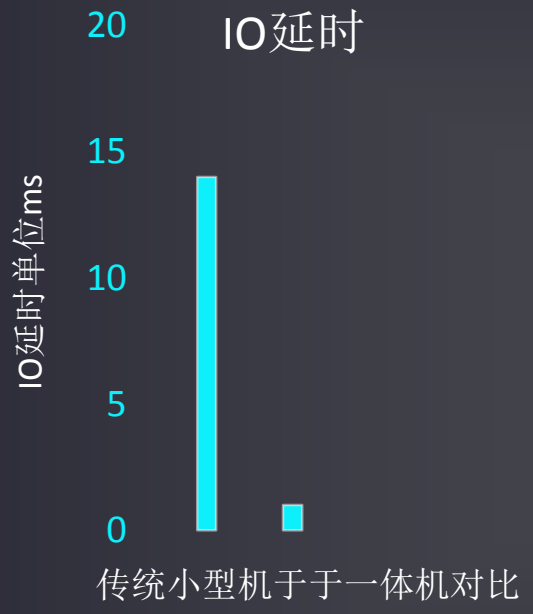
至少10倍的IO提升

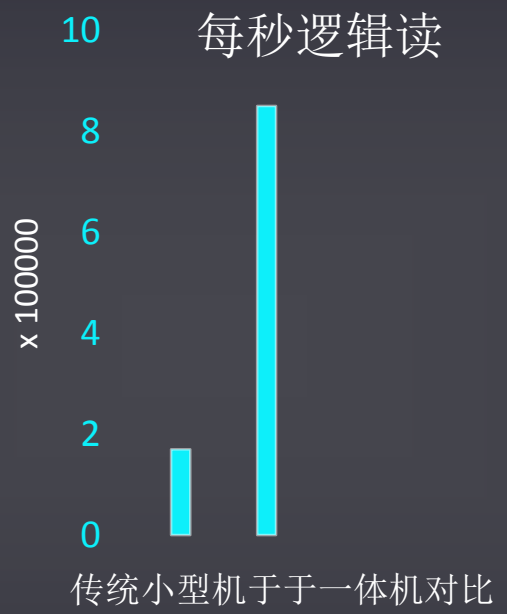
网络的对比



至少40倍的网络传输速度

Stength 强大





Oracle数据库各项指标的对比

Opportunities 机会

响应IT系统本土化

自主创新，自主研发出解决方案，解决生产过程中的实际问题，这一模式已经被市场认可

Weaknesses 缺点

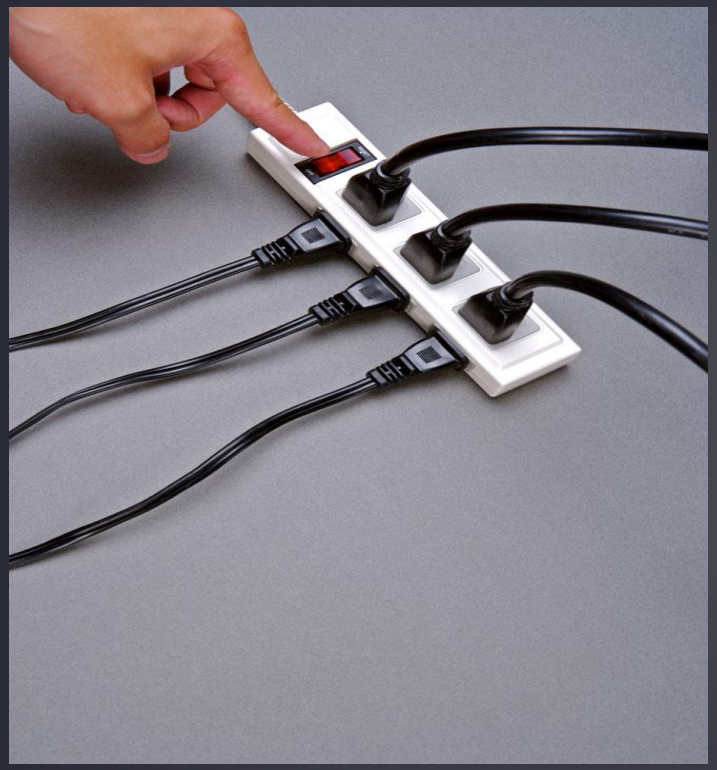
1-1%*1%*1%=99.999999%

99.999%

X86架构一体机解决方案的安全性

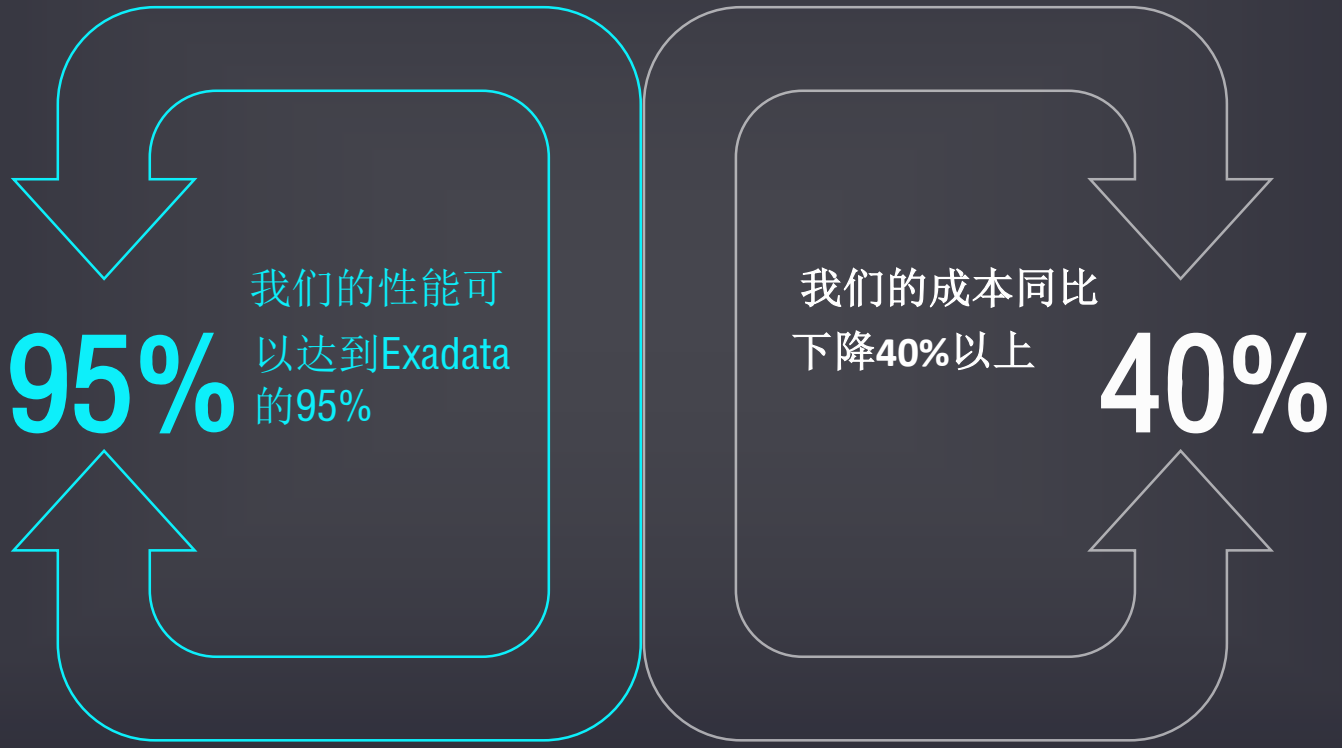
小型机架构的安全性

Weaknesses 缺点



2015年，凡是有客人参观龙江电力，我做的最多的事情就是拔主机电源，甚至直接替换服务器

Threats 威胁



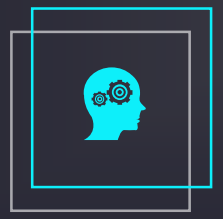
X86 PC Server代替小机存储方案

X86代替小型机方案更敏捷

X86代替小型机方案更高效

X86代替小型机方案更省钱

X86代替小型机方案更安全



X86代替小型机方案管理更简单

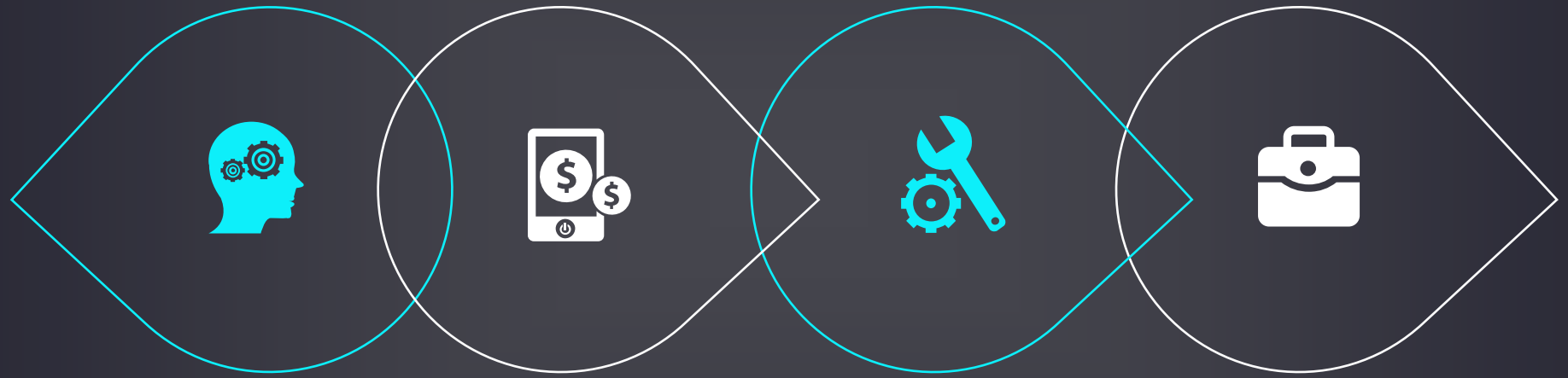
X86代替小型机方案更稳定

PART 03

填坑运动

迁移

跨平台、跨版本



大小端问题

迁移的稳定性

系统安全

迁移

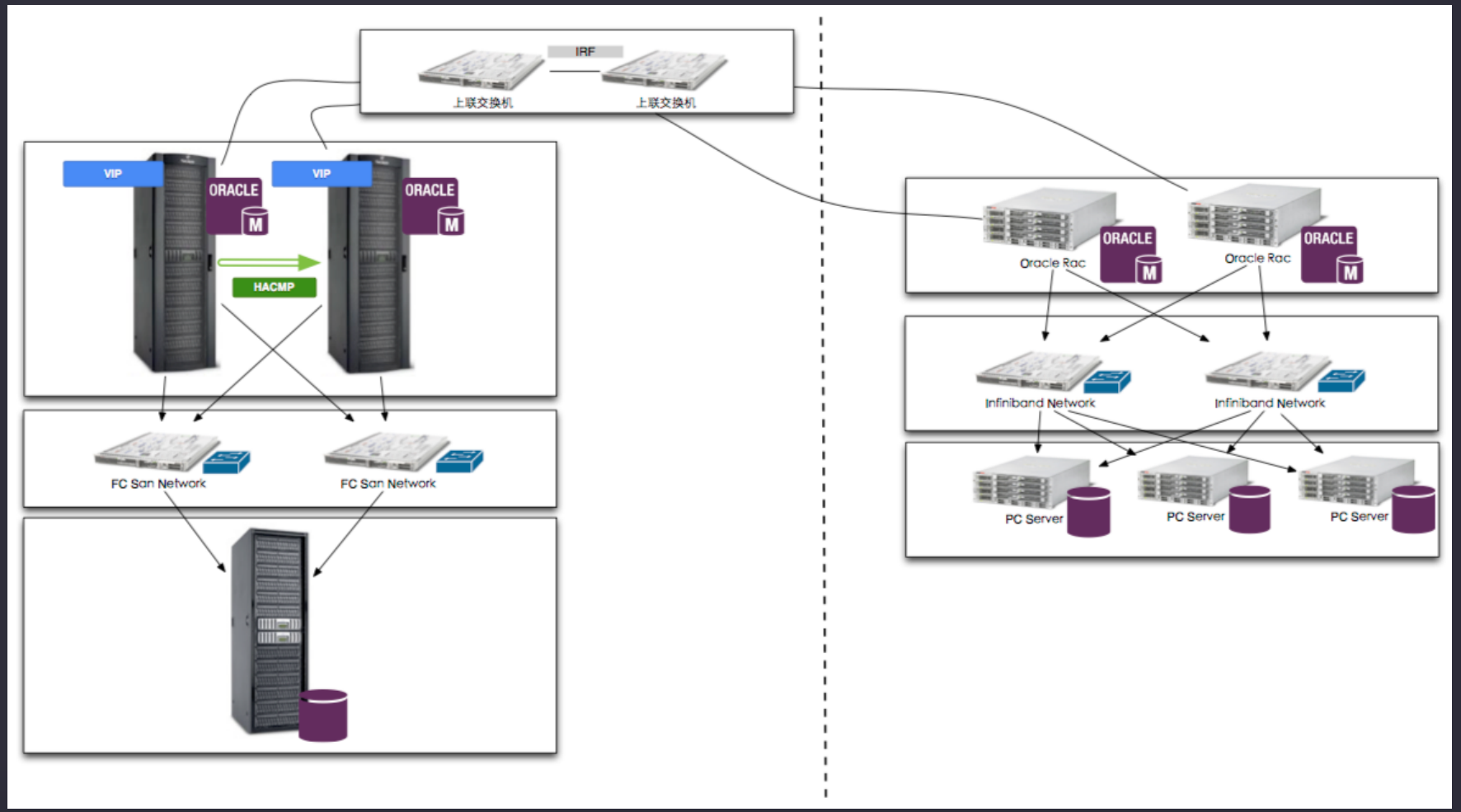
RMAN无法解决大小端问题

传输表空间完美解决问题

Goldengate 太过复杂，不能在有效的时间内完成任务

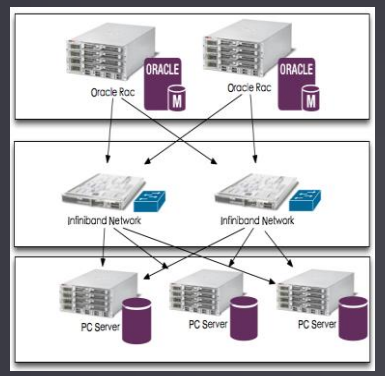
数据泵太慢，时间长

迁移

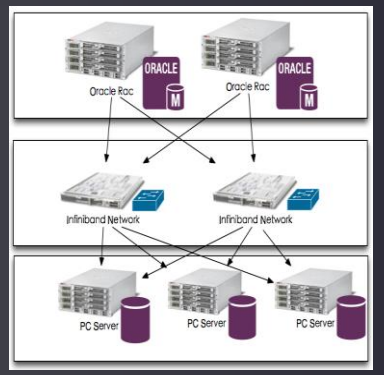

IRF
上联交换机
上联交换机
VIP
ORACLE
M
VIP
ORACLE
M
HACMP
FC San Network
FC San Network
Oracle Rac
ORACLE
M
Oracle Rac
ORACLE
M
Infiniband Network
Infiniband Network
PC Server
PC Server
PC Server

迁移

生产主库

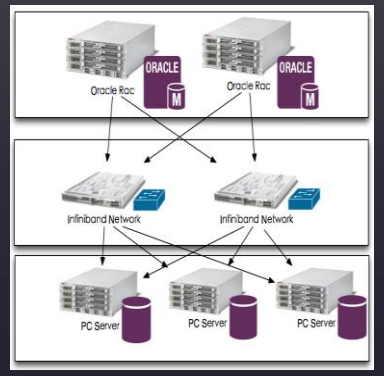


Active dataguard
实现数据同步



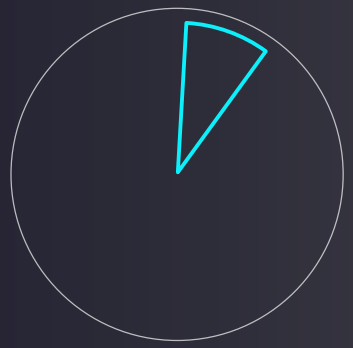
同城备库

Active dataguard
实现数据同步



北京容灾中心备库

对比



10%

系统IO等待时间降为原来的10%

15%

系统平均等待时间降为原来的15%

400%

系统吞吐能力提升了4倍

200%

单位时间事务处理量为原来的2倍

对比

同样一条SQL在执行计划不变的情况下的对比

6160S

在HP小机RX8640+存储架构下
跑了6160秒约合1小时40分钟

315S

在x86架构下跑了315秒约合5分钟

容量扩展

- 2U 服务器, 最多可以部署6张全高的PCI-E闪存卡, 如HP DL380 Gen9, 最高40TB的可用容量
- 3U 服务器, 最多可以部署11张全高的PCI-E闪存卡, 如Supermicro Gen X9DRX+-F 90TB裸容量, 最高80TB的可用容量
- 4U 服务器，最多可以部署9张全高的PCI-E闪存卡，如HP DL580 Gen8，最高50TB可用容量。

预算问题

- 纯flash卡，ssd架构
- 纯sas盘架构
- 混合架构

PART 04

云时代的来临

云时代的来临

Oracle的替代者

- MySQL
- PostgreSQL
- Mongodb

存储的替代者

- CEPH SDS

中间件的替代者

- 全栈技术的革命者 METEOR

Oracle 的替代者

Oracle数据库

商业数据库的老大，其性能和同步技术还是独步天下
一些企业盲目追寻去O，但是整个成本和代价实在很大，
而更重要的是系统的安全和性能受到很大质疑。

MySQL

SUN公司的遗孤，归入Oracle公司后
功能日益强大，但是其同步，容灾技术
尚需加强，而最关键的优化器技术导致
该数据库在运行大量复杂SQL的时候性能
很低，而真正运行好的数据库是DBA看出来的

PostgreSQL

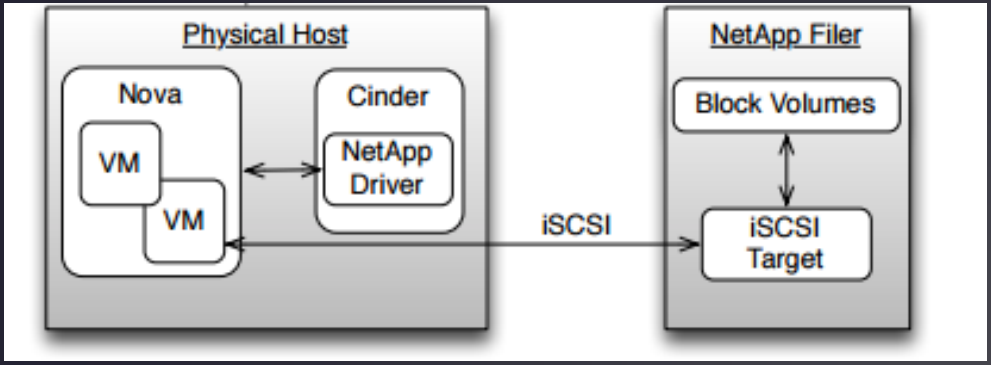
PostgreSQL的存储过程最像Oracle的PL-SQL，而且
相对来讲稳定性强，因此是企业代替Oracle的选择

Mongodb

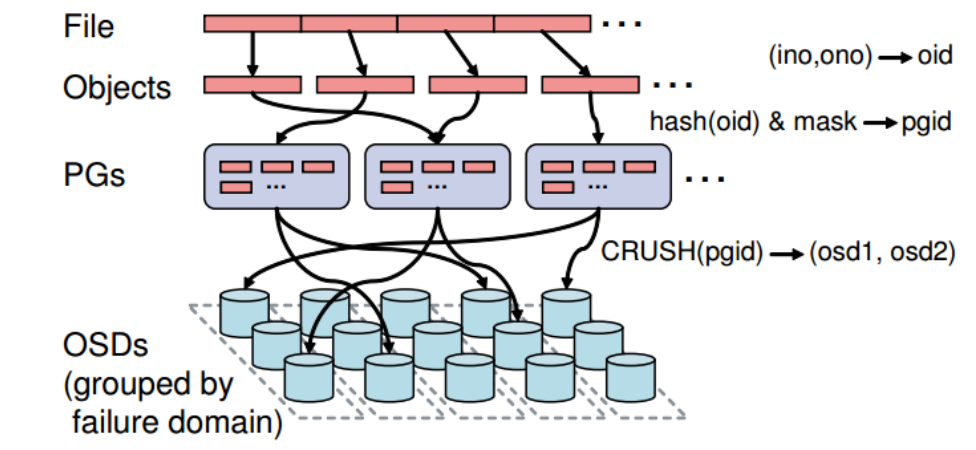
非关系型数据，库互联网
时代另外的一种选择

软件定义存储 CEPH

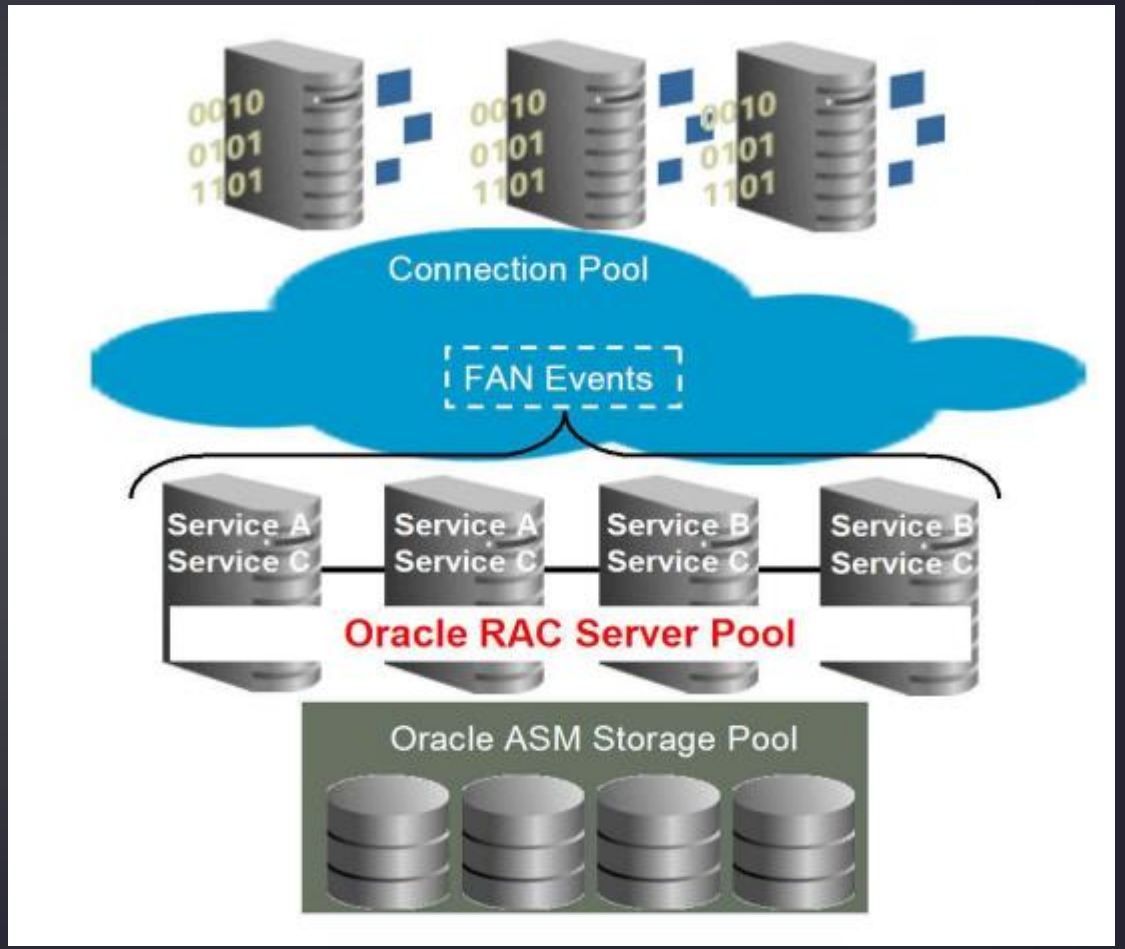
通过最新的40Gb以太网甚至100Gb以太网将ISCSI的延迟降低，消除网络瓶颈



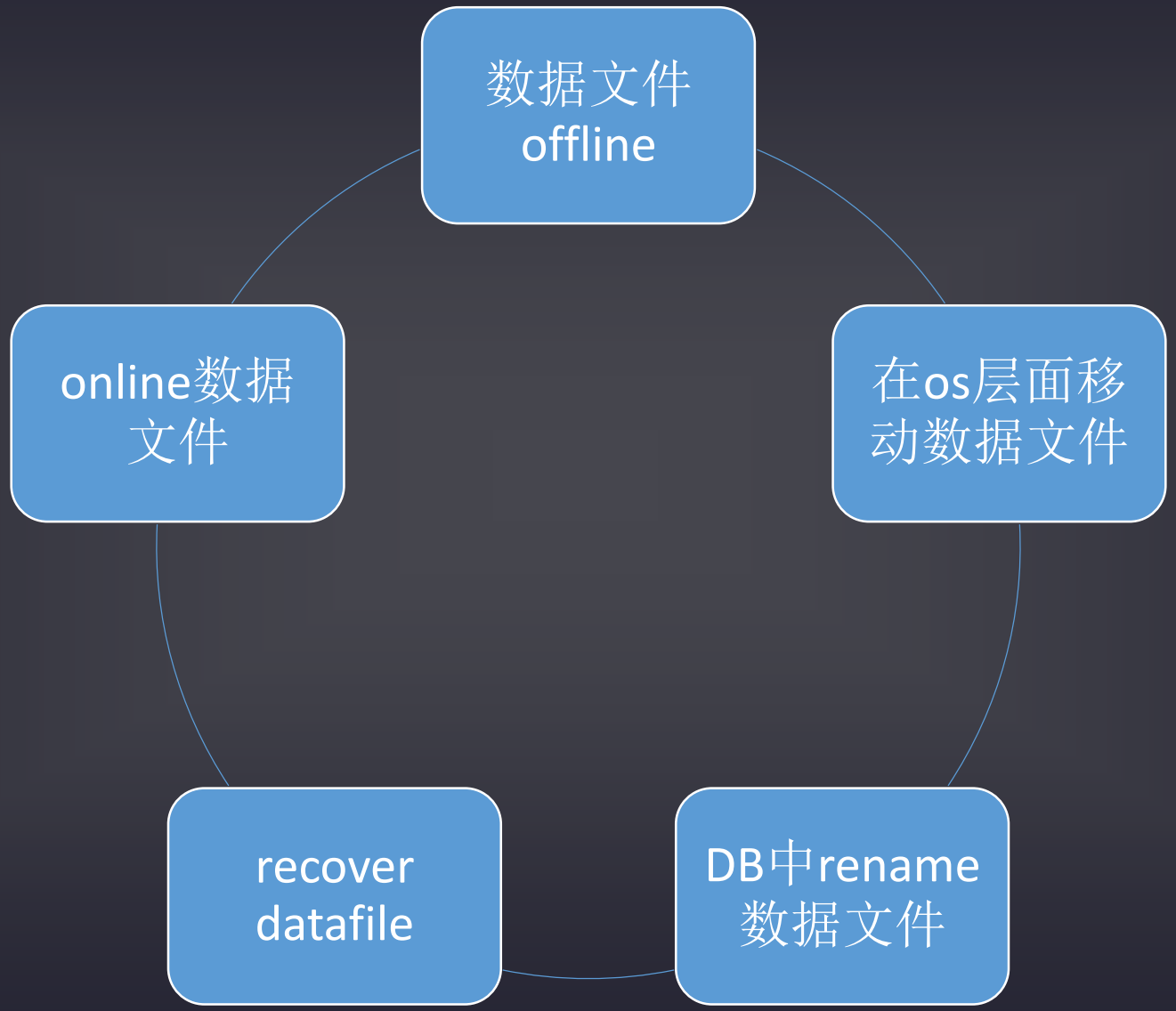
更有效的利用CEPH提供的API，对存储访问进行优化



最后利用openstack技术整合实现真正的数据库云化



Oracle 12C


数据文件
offline
在os层面移
动数据文件
DB中rename
数据文件
recover
datafile
online数据
文件

Oracle 12C

在Oracle 12c中直接Move 即可：

SQL> **alter database move datafile**
'/u01/app/oracle/oradata/dave/huaining.dbf' to
'/u01/app/oracle/oradata/dave/anqing.dbf';
Database altered.

这种移动不仅会修改控制文件中的信息，也会在OS级别物理的移动。

利用这个特性可以做如下事情：

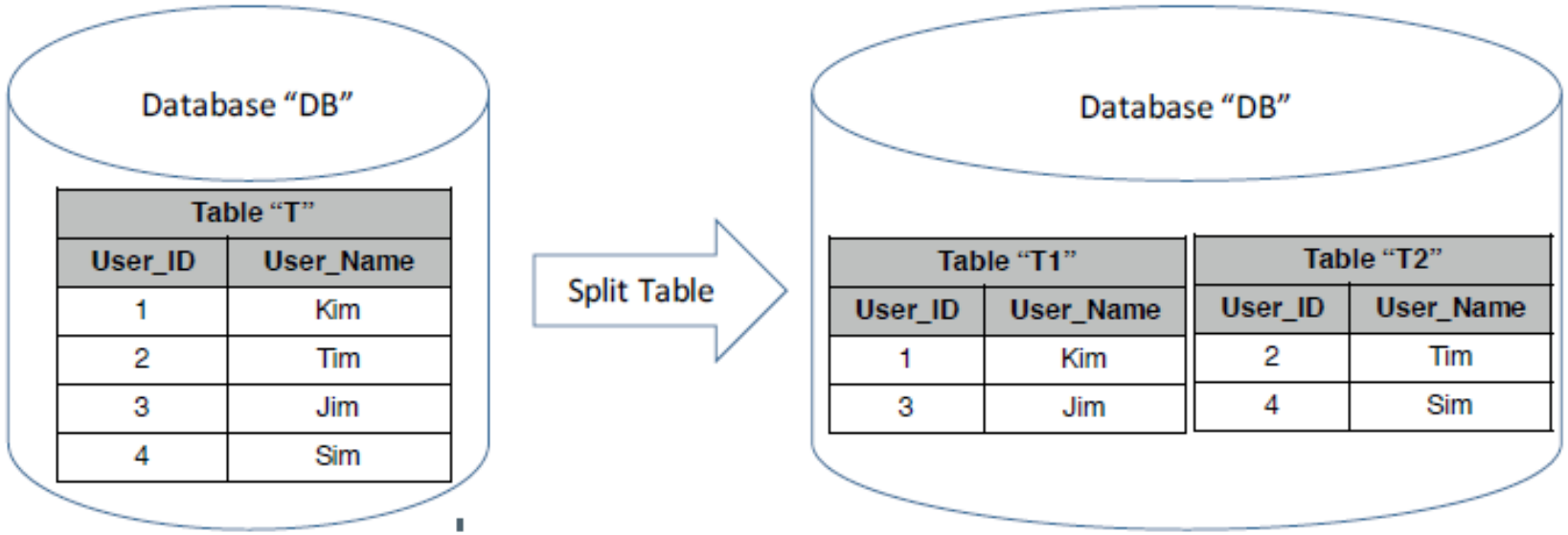
- 在线移动数据文件位置
- 在线重命名数据文件
- 在线移动数据文件从ASM到文件系统

Oracle 12C

sharding

Oracle 12C

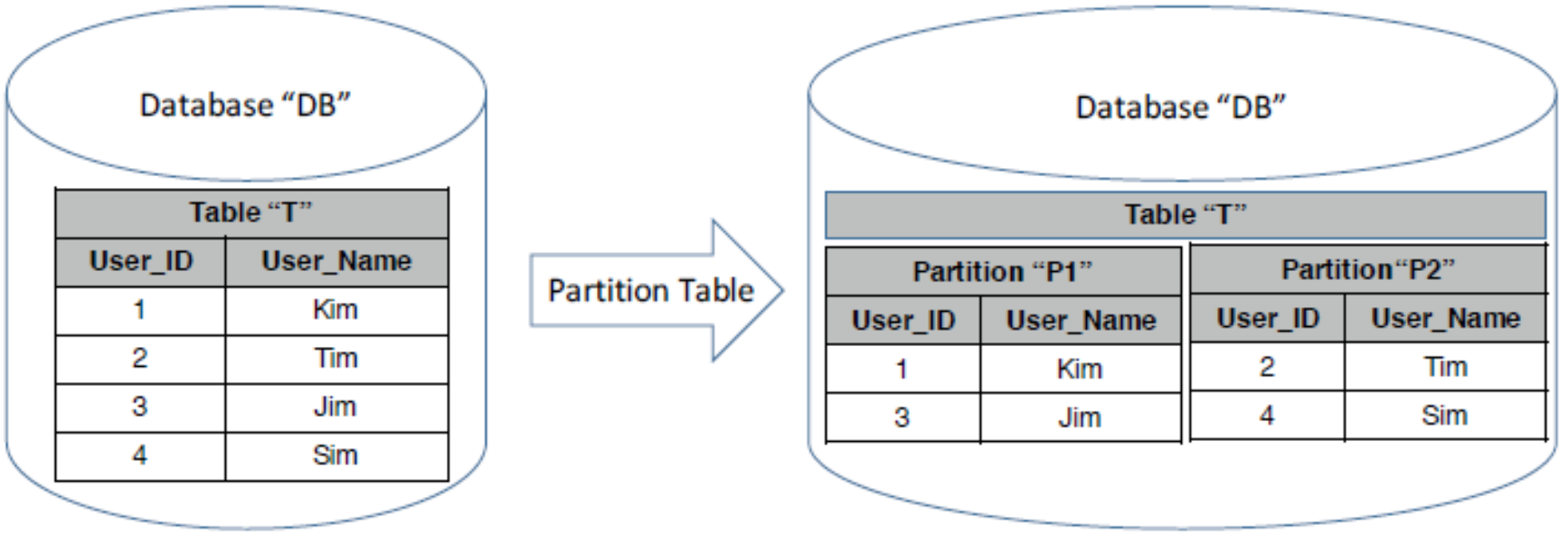
WHAT IS SHARDING EXACTLY




```
//Query single row
select User_Name from T1 where User_ID=1;
//Query all
create view T as select * from T1 union all select * from T2;
select count(*) from T;
```

Oracle 12C

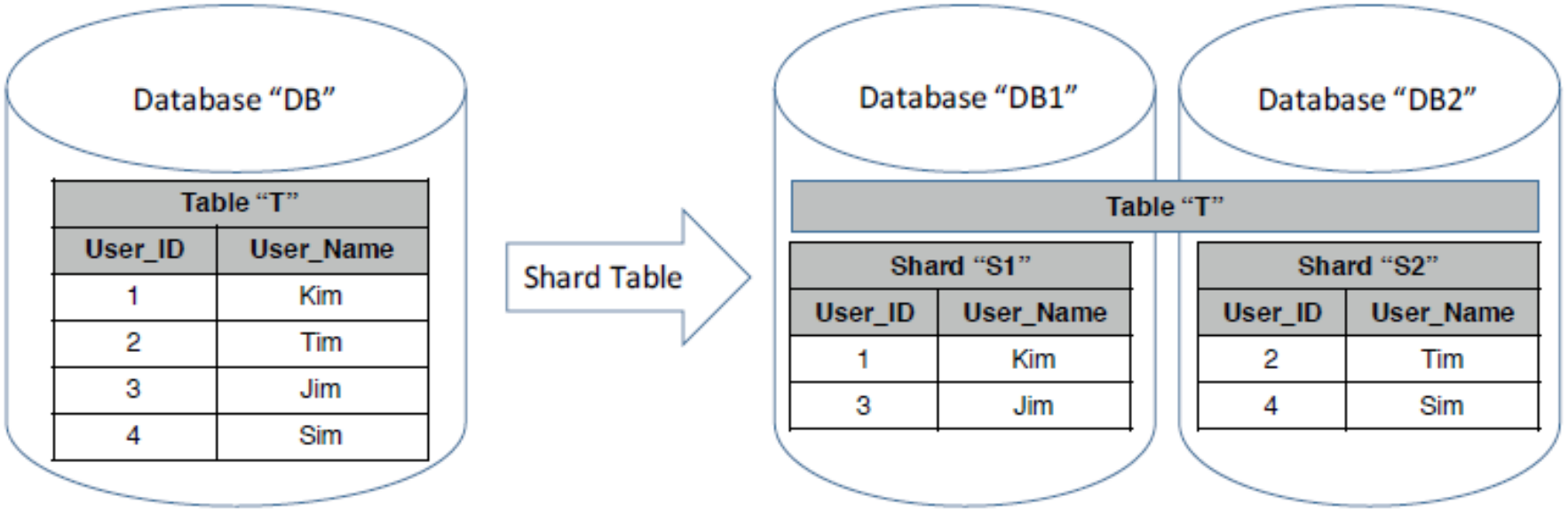
WHAT IS SHARDING EXACTLY




```
//Query single row
select User_Name from T where User_ID=1;
//Query all
select count(*) from T;
```

Oracle 12C

WHAT IS SHARDING EXACTLY




```
//Query single row
select User_Name from T where User_ID=1;
//Query all
select count(*) from T;
```

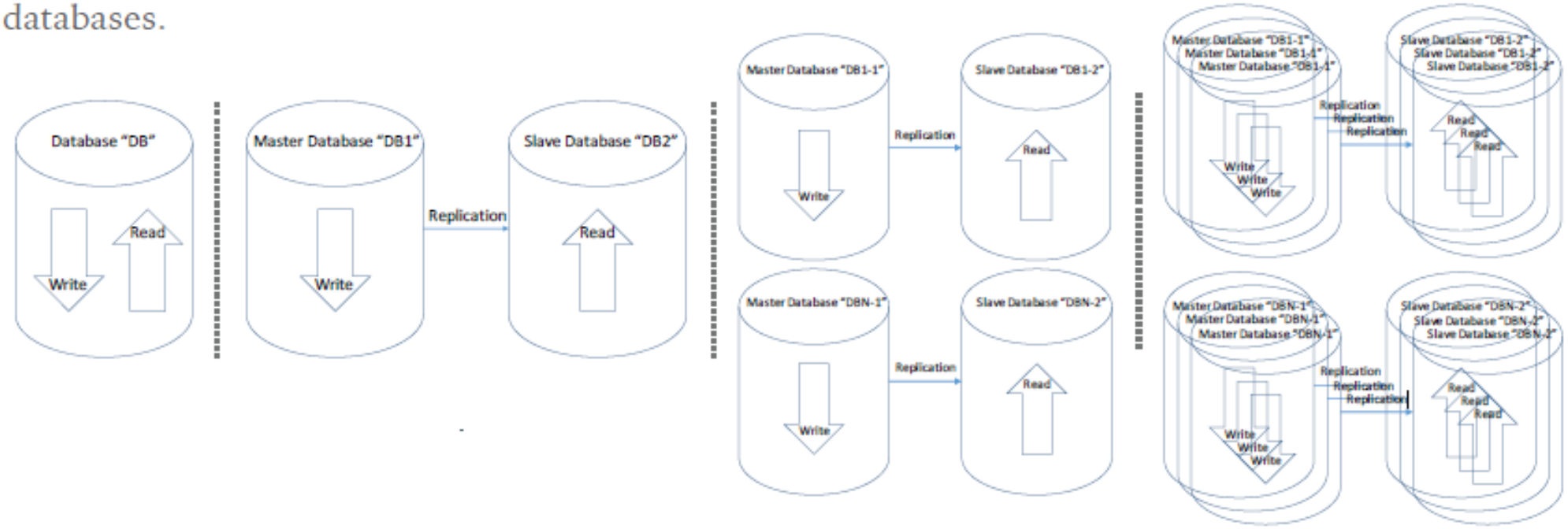
Oracle 12C

WHY NEED SHARDING? BENEFITS OF SHARDING

- Extreme Scalability
- Scale Out vs. Scale Up
- Fault Isolation
- No shared SAN
- Global Data Distribution
- Store particular data close to its consumers
- Pilot & Rolling Upgrades
- First test the changes on a small subset of data.
- Cost Down
- Because the size of a shard can be made arbitrarily small, deploying an SDB in a cloud consisting of low-end commodity servers with local storage becomes easy.

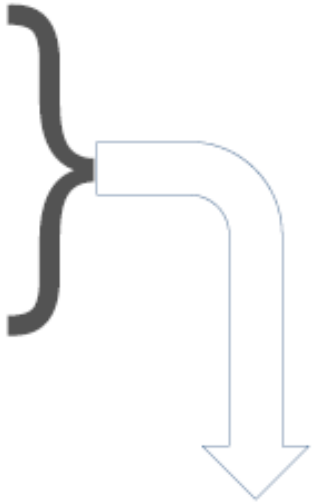
Oracle 12C

- All business on Single database
- All business on Master+Slave databases, write/read separated.
- Business separated, each business on separated Master+Slave databases.
- Area based/Hash based, table horizontal split, each part of business on separated Master+Slave databases.



Oracle 12C

- Is it transparent to application developer?
- Which database should access?
- How to sum query all databases?
- How to join query all databases?



Database access layer/Data access route layer

- Proxy-Free： MySQL Fabric, TDDL
- Proxy： MySQL Proxy, Atlas(Qihoo),Cobar

- Single point of failure
- Failover more complex
- Backups more complex
- Modification more complex

Oracle 12C

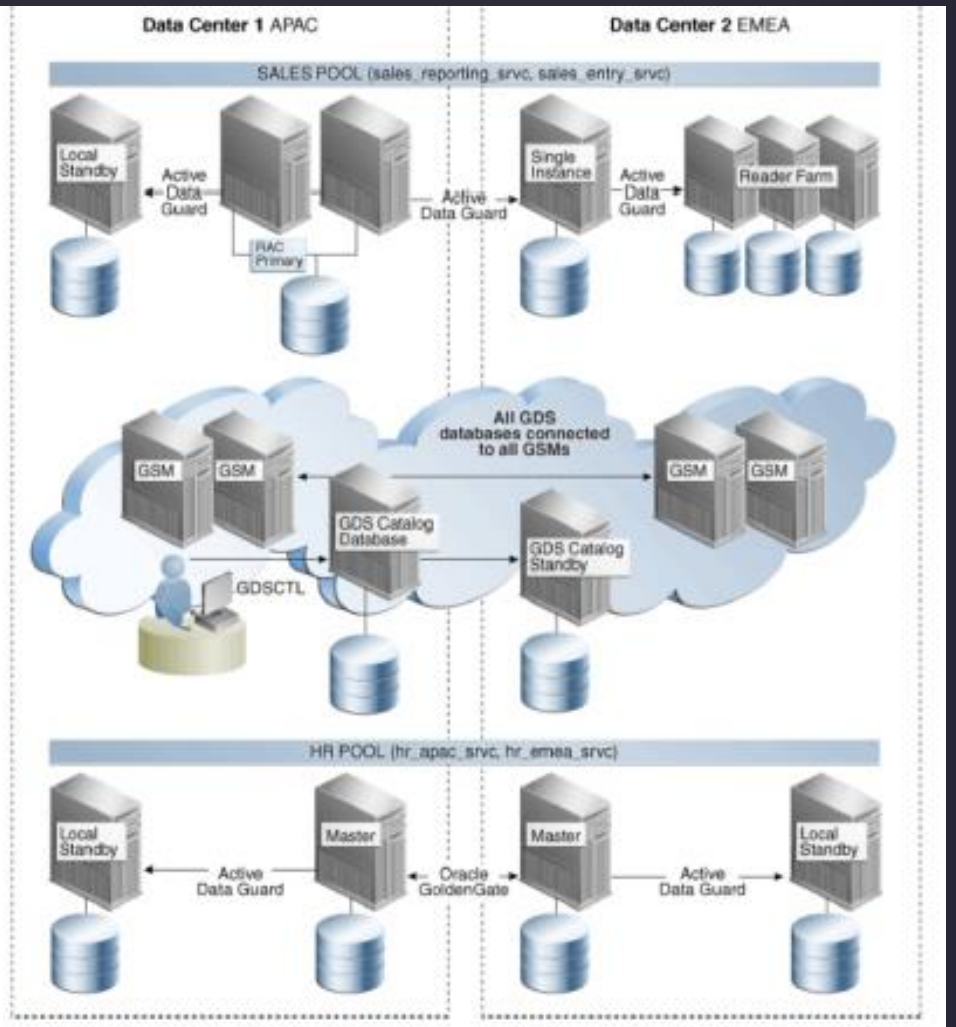
- SQL
  - MySQL with MySQL Fabric
  - Postgres-XC and Greenplum
  - Teradata
- No-SQL
  - Apache HBase
  - MongoDB
- New-SQL
  - Google Spanner

And NOW

ORACLE® *is coming!*

Oracle 12C

- What is GDS
  - Instance->Service->Global Service
- GDS Capabilities
  - Workload routing(region-based or lag-based)
  - Load balancing(Connect-time or Run-time)
  - Global service failover/role-based failover
- Released in Oracle database 12.1
  - GSM(Global Service Manager)
  - GDS Catalog
    - Must reside in Oracle database 12c
    - Recommended to co-hosted with RMAN catalog or OEM repository

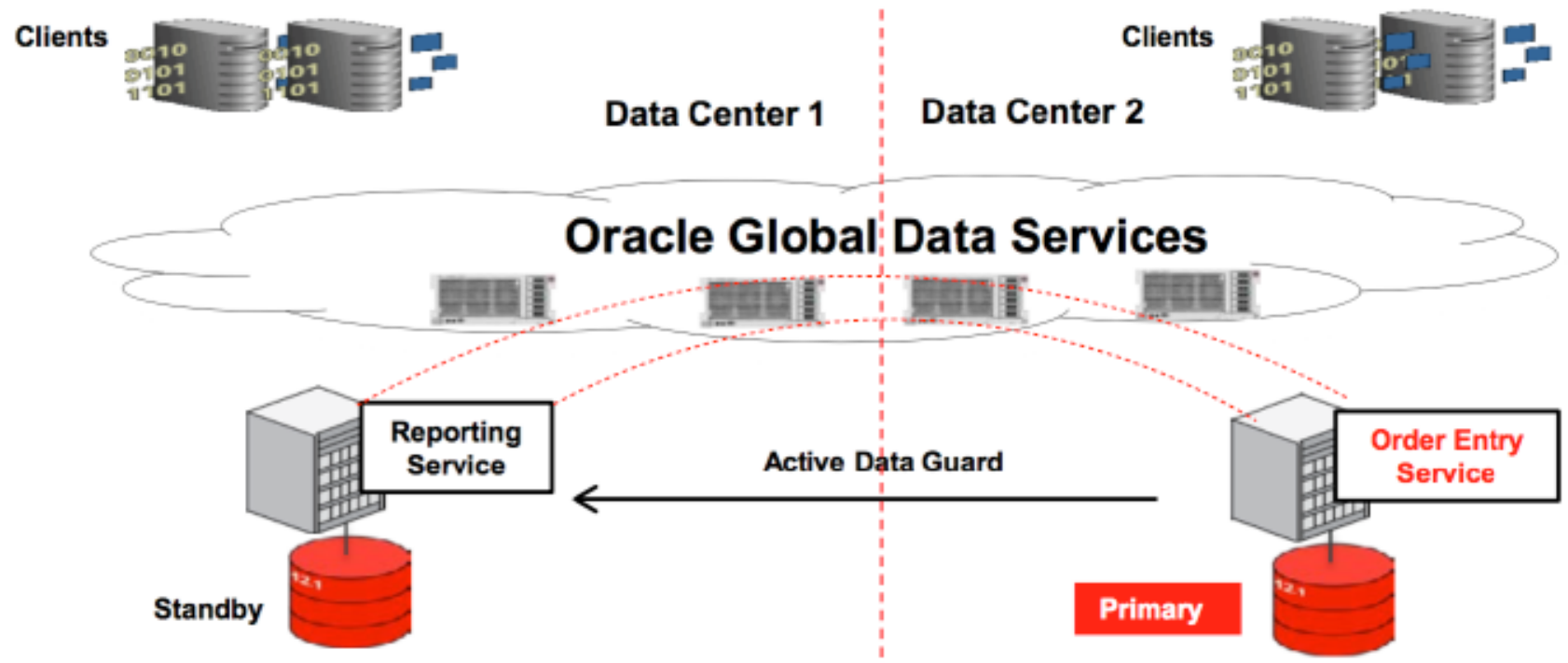


Oracle 12C

- Install GSM software on GSM servers
  - Minimum 1 GSM per region
  - Recommended 3 GSMs/region
- Setup GDS administrator accounts & privileges
- Configure GDS
  - Create GDS catalog
  - Add GSMs,Regions, Pools, Databases, Global Services
- Setup client connectivity
  - Clients connect to GSM instead of the database listener

Oracle 12C

- Order Service runs on Primary
- Reporting Service runs on Standby
- Upon Data Guard role change, GDS fails over services based on Role



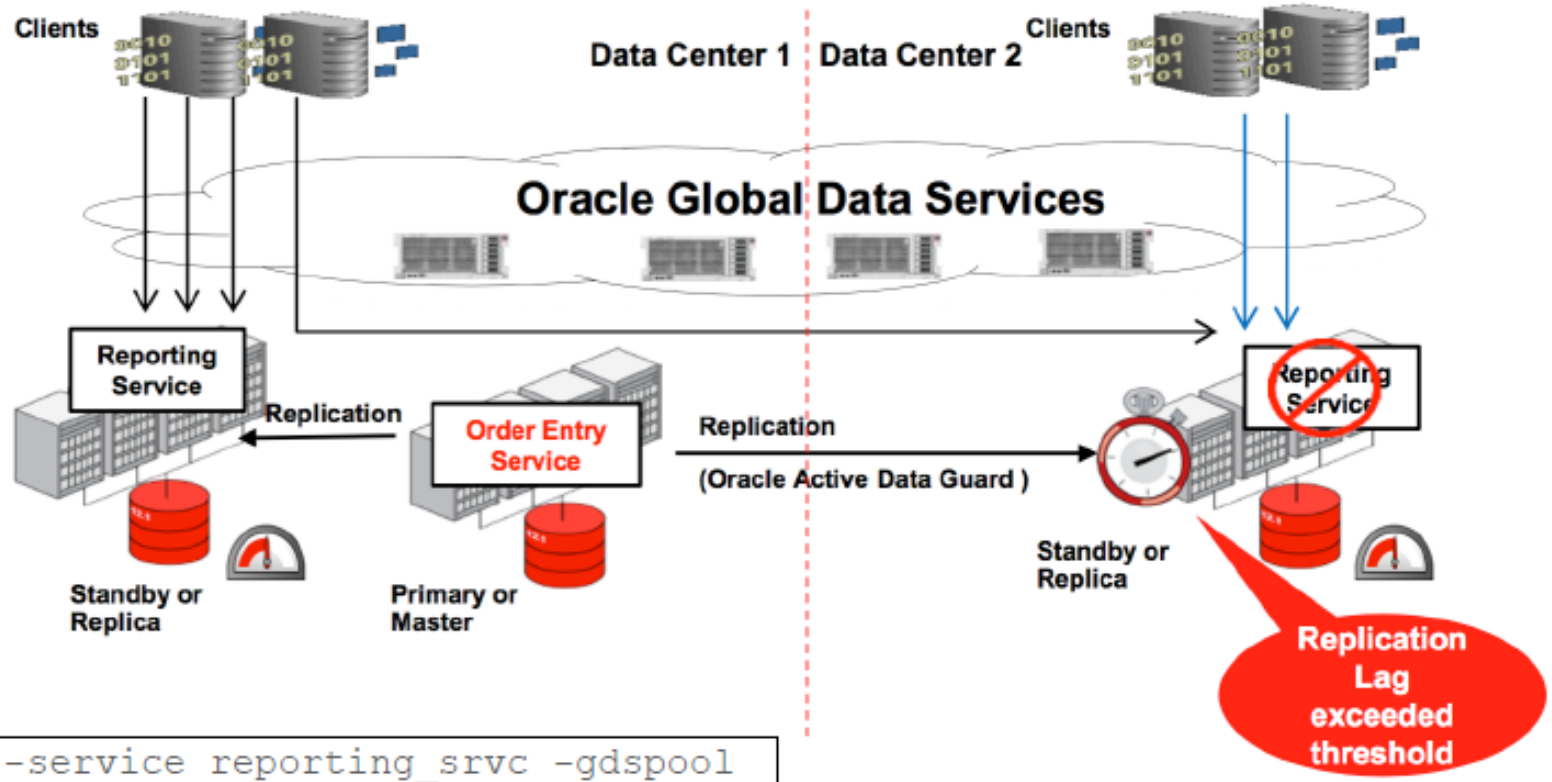

```
GDSCTL> add service -service order_srvc -gdspool sales -
preferred_all -role PRIMAY;

GDSCTL> add service -service reporting_srvc -gdspool sales
-preferred_all -role PHYSICAL_STANBY -failover_primay;
```

Oracle 12C

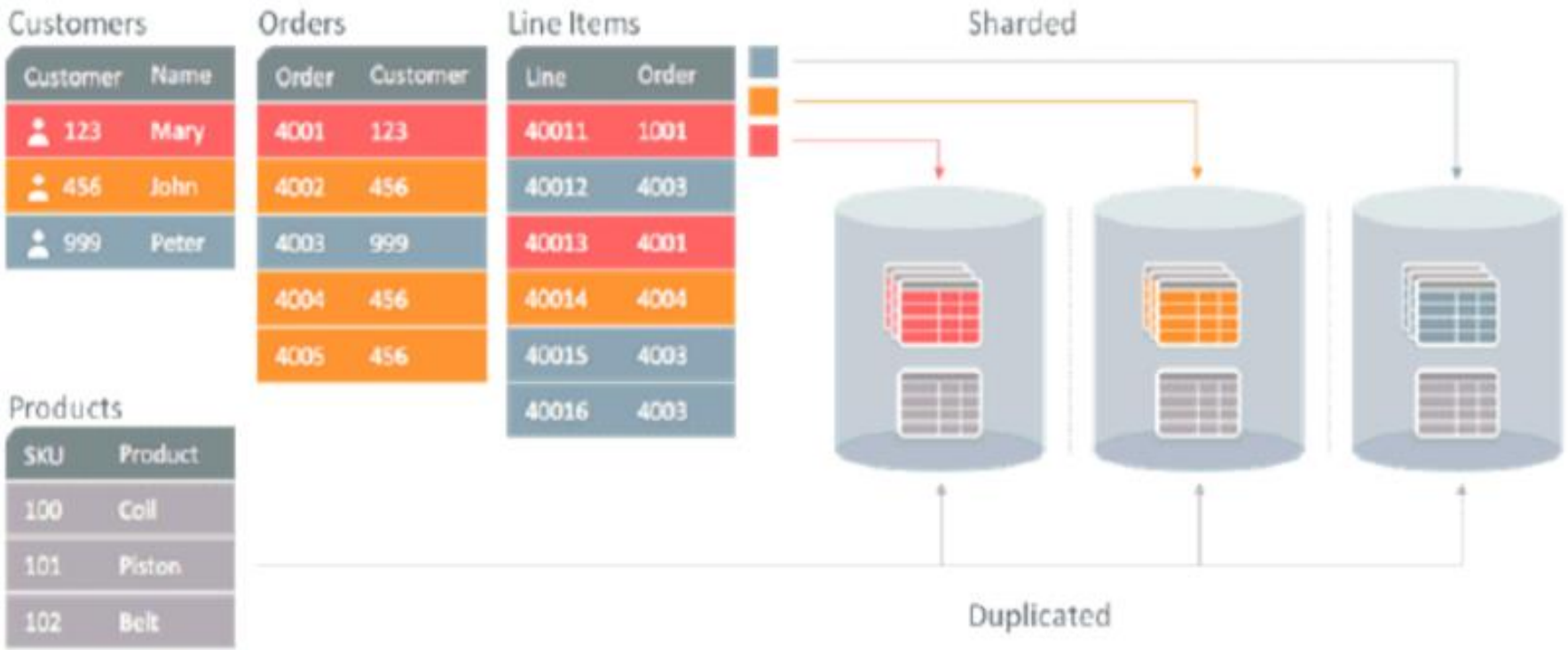
GDS USE CASES – REPLICATION LAG TOLERANCE IN ACTIVE DATA GUARD

- Specify replication lag limit for a service
- GDS ensures that service runs on Active Data Guard standby(s) with lag less than this limit
- Improved data quality




```
GDSCTL> add service -service reporting_srvc -gdspool
sales -preferred_all -role PHYSICAL_STANDBY -lag 180;
```

Oracle 12C



Oracle Sharding is implemented based on the Oracle Database partitioning feature.

Oracle Sharding is “Distribute Partitioning”.

Oracle 12C

“Unlike NoSQL data stores, Oracle Sharding provides the benefits of sharding without sacrificing the capabilities of an enterprise RDBMS, such as relational schema, SQL and other programmatic interfaces, complex data types, online schema changes, multi- core scalability, advanced security, compression, highavailability, ACID properties, consistent reads, and many more..

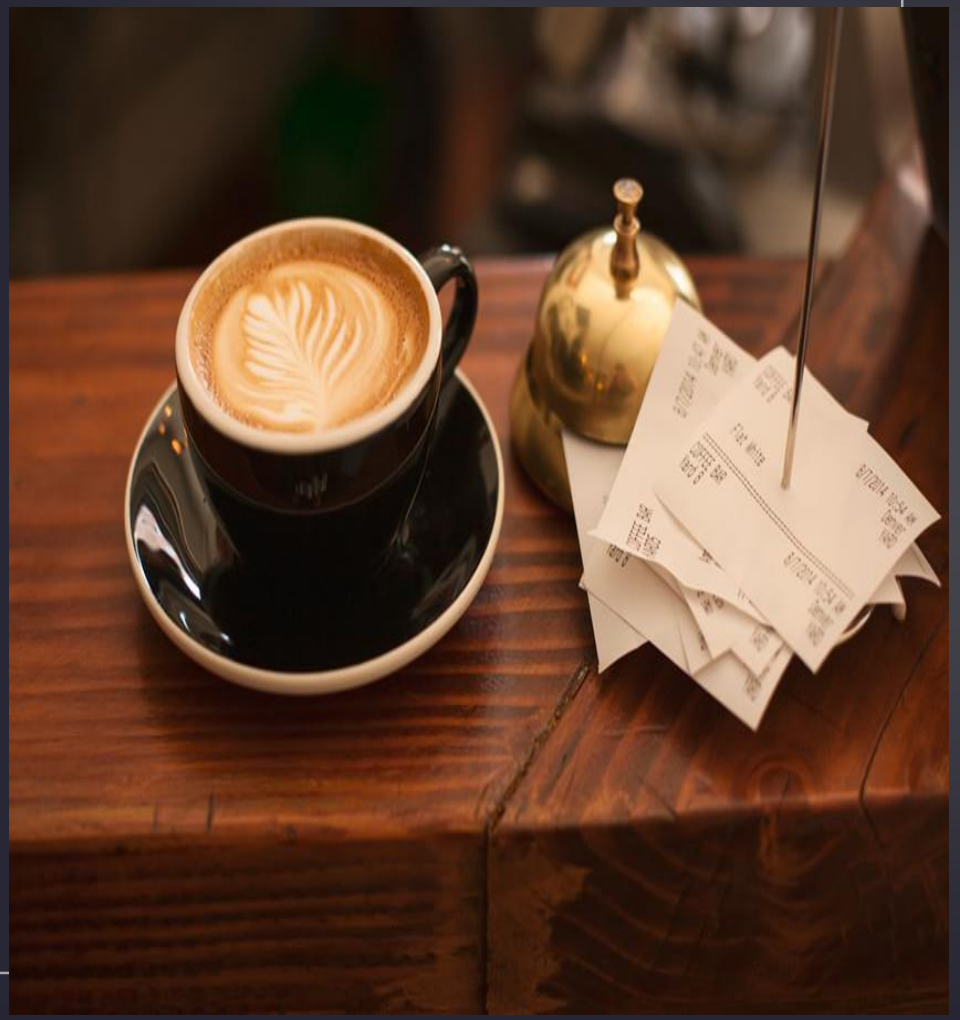
-Oracle® Database Administrator's Guide 12c Release 2 (12.2)

Oracle 12C

THINK TWICE BEFORE JUMPING INTO SHARDING

- **MySQL: Not exceed 20GB per table is optimised. (Concurrent processing capacity)**
- **How about we have a 2TB table? 100 shards!**
- **What does 100 shards mean?**
- **100 x86 servers**
- **1 Primary, 1 Slave -> 200 servers**
- **1 Primary, 2 Slaves -> 300 servers**
- **How many routers do you need? How many cables? How many spaces?**
- **Are you ready for maintaining so many servers?**
- **99% server stableness means: Server down/EVERY day.**

数据库云池

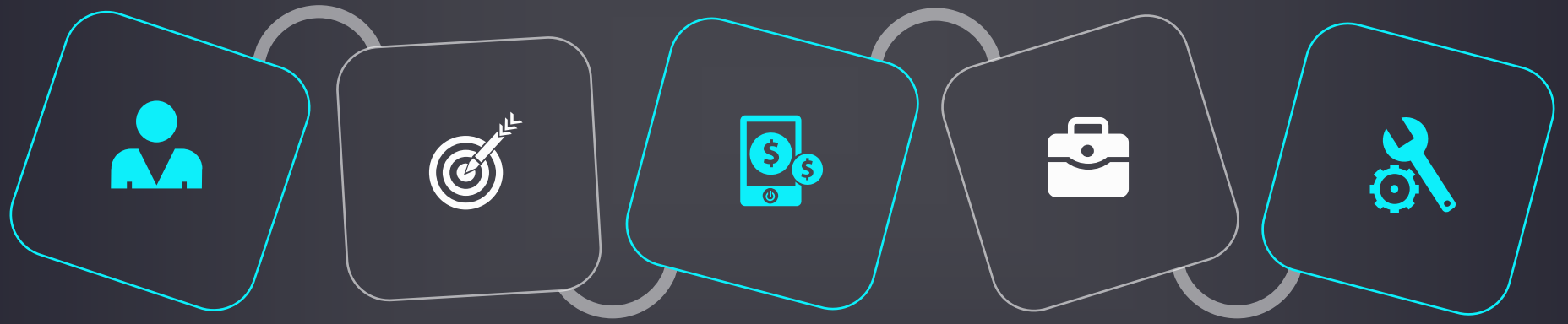


真正的数据库云池，运维人员从此过上幸福快乐的生活

中间件的替代者 Meteor

Web，Android，IOS代码统一

变态的reactive



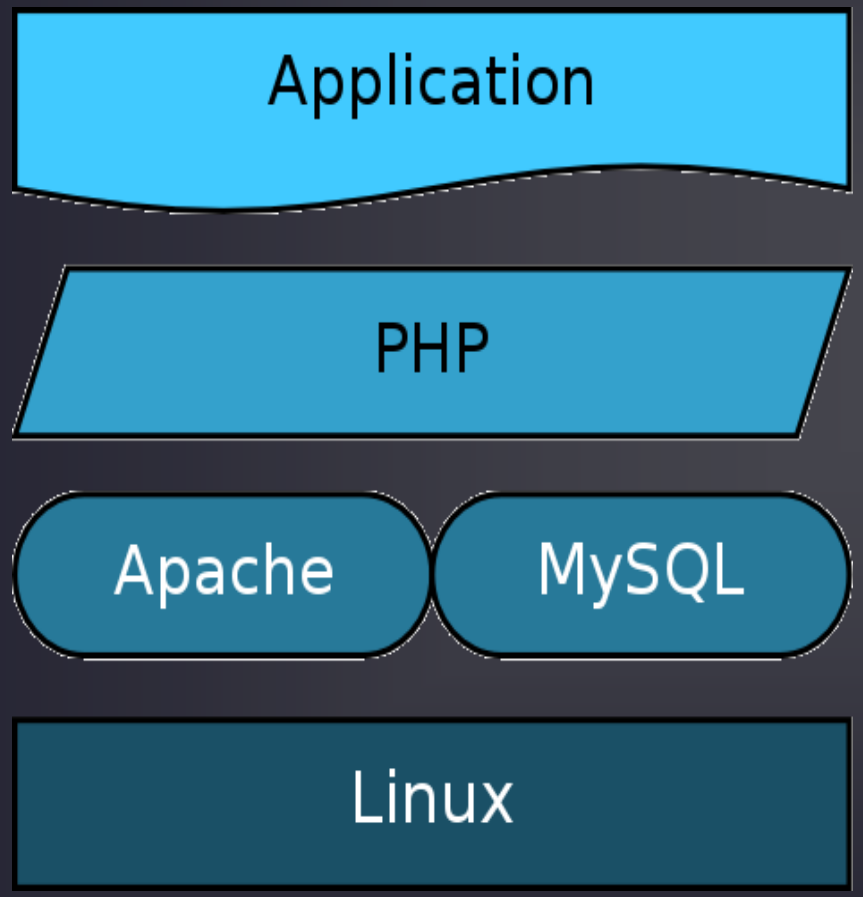
全栈统一

DDP及时响应无等待

本地操作处处响应

中间件的替代者 Meteor

互联网流行的技术栈

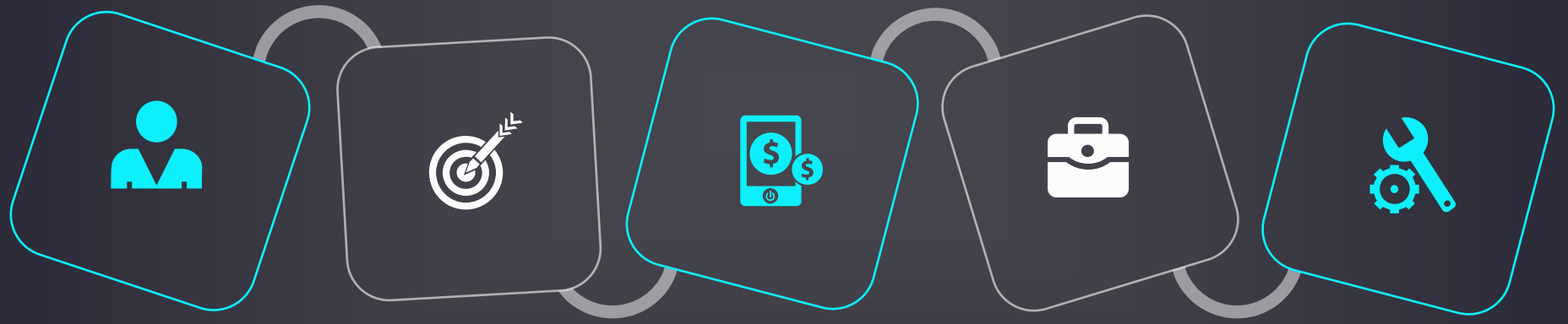


METEOR

中间件的替代者 Meteor

Web，Android，IOS代码统一

变态的reactive



全栈统一

DDP及时响应无等待

本地操作处处响应

THANK YOU