



**Oracle**数据库高性能高可用性

**ORACLE®**

# 提纲

## ➔ Oracle数据库最高可用架构

- Oracle数据库集群技术
- Oracle数据库分区技术
- Oracle GoldenGate技术
- 问答



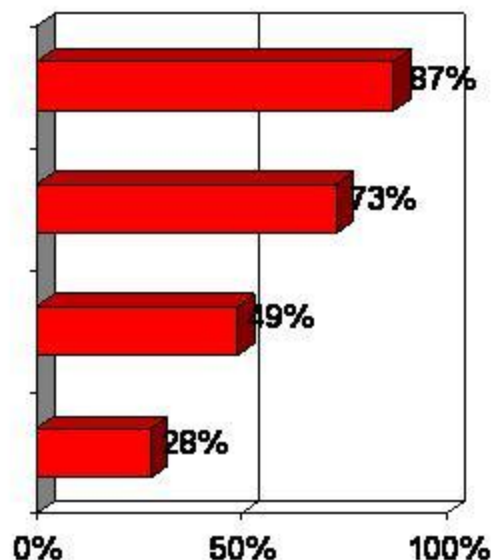
# 企业 IT 管理者面临的主要挑战

一致的端到端应用程序和服务性能保证

计划外的基础架构更改导致突发事件和停机

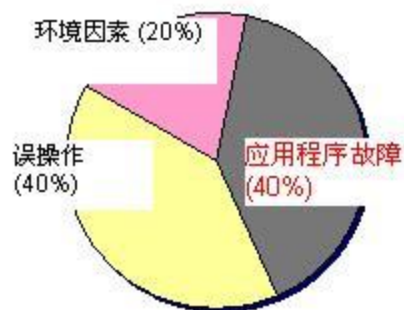
合并和新的应用程序项目对基础架构产生意外影响

网络对象配置错误



信息来源：  
Forrester Research,  
企业  
IT 基础架构管理者调查

计划外停机 (非 SOA)



计划外停机 (SOA)



计划停机



信息来源：  
Gartner 2006

# Oracle最大可用性体系结构

最佳HA 技术的完整集合

- 每个都是可以水平扩展的, 完全激活的, 以数据为中心的

最高可用性和最低的成本

**Real Application Clusters  
& Clusterware**

容错  
服务器水平扩展

联机升级

联机升级  
硬件和软件

**GoldenGate**

双活数据复制

**Data Guard**

完全激活  
故障切换到复制点

**Automatic Storage  
Management**

容错  
存储水平扩展

**Flashback**

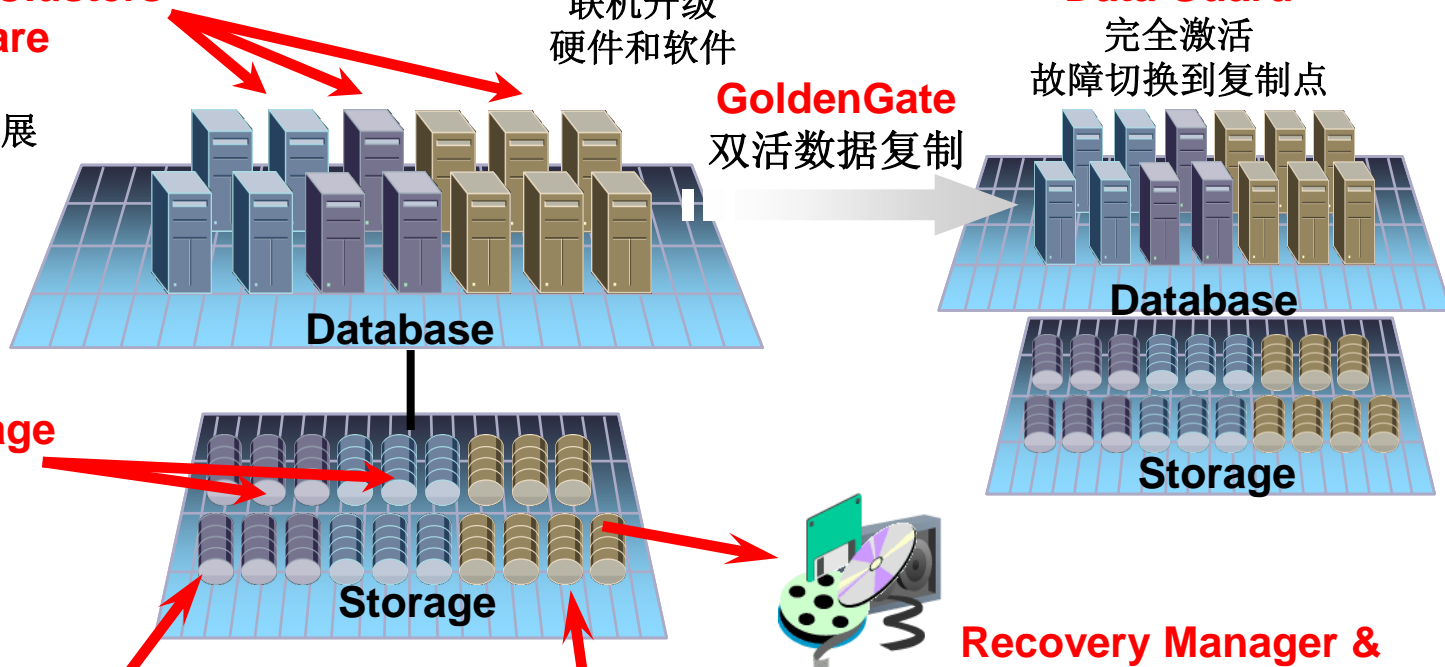
退回到指定时间点  
修正错误

联机重定义

联机重定义表

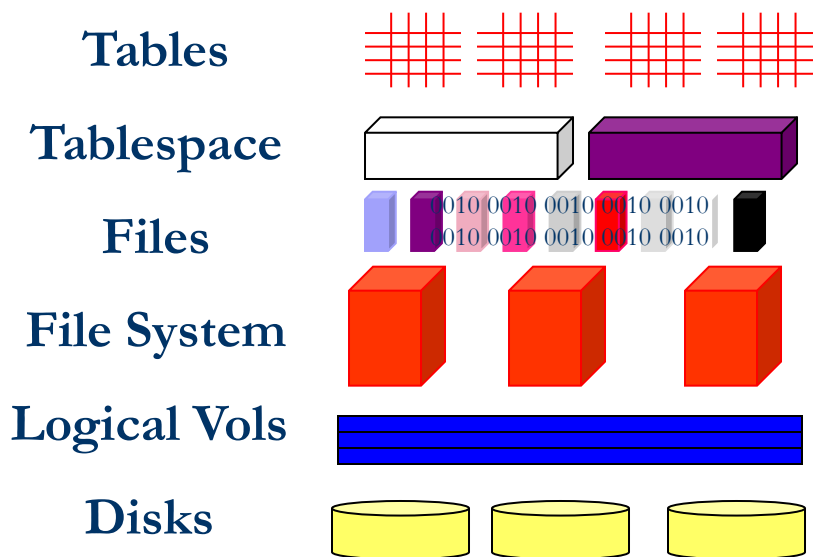
**Recovery Manager &  
Oracle Secure Backup**

低成本高性能  
数据保护 & 归档

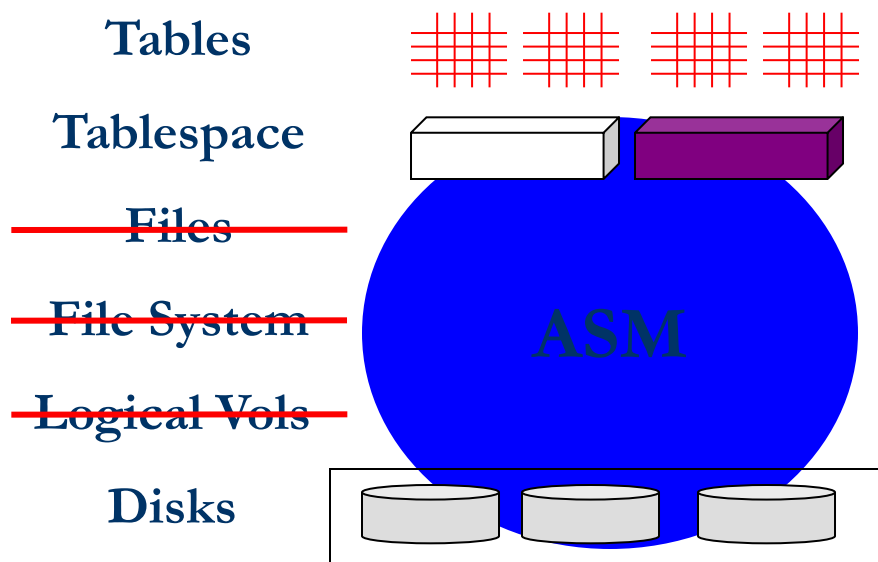


# 自动的存储管理 (ASM)

## 传统的方法



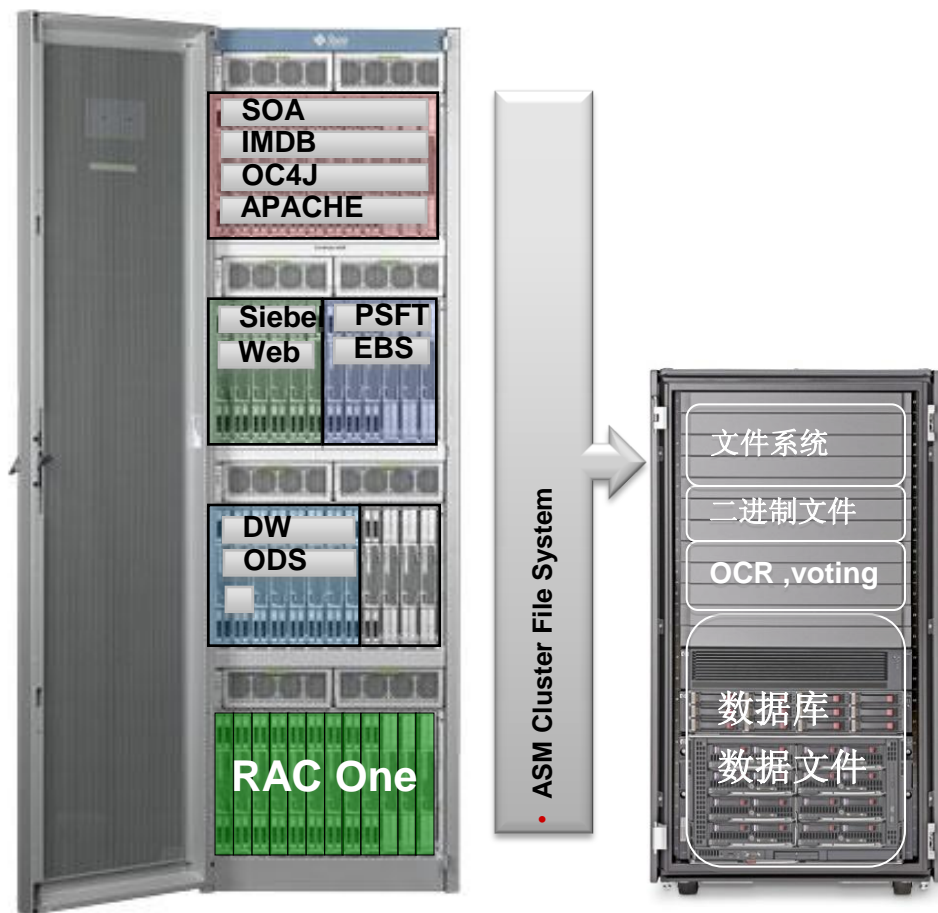
## Oracle ASM



降低管理成本的最佳方法就是降低复杂性

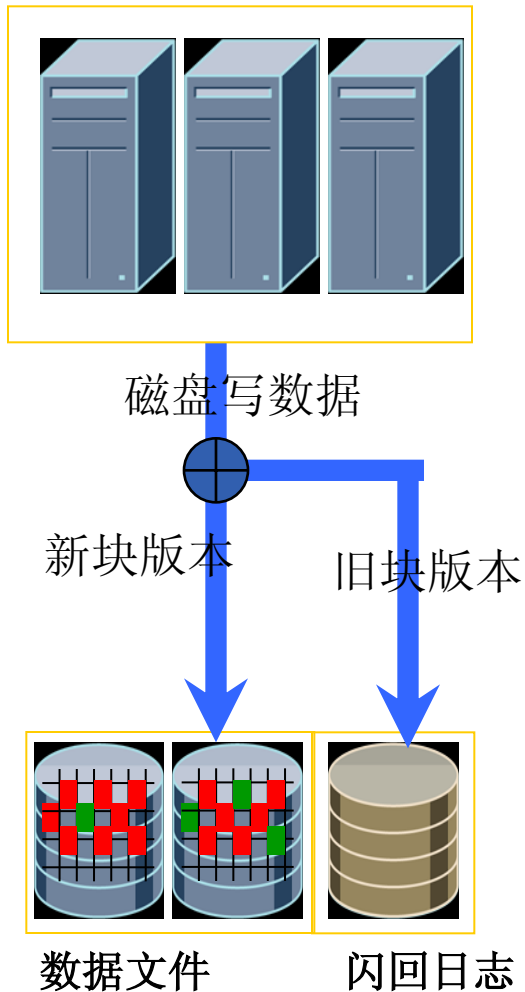
# Oracle 数据库 11g 第 2 版

## ASM 集群文件系统 (ACFS)



- 通用集群化或本地文件系统
- 数据存储在 ASM 中
- 继承所有 ASM 可管理性优势
  - 优化的磁盘布局
  - 联机添加/删除/重新平衡磁盘
  - 集成的镜像
- 只读快照
  - 文件系统多达 64 个时间点空间有效副本

# 闪回数据库



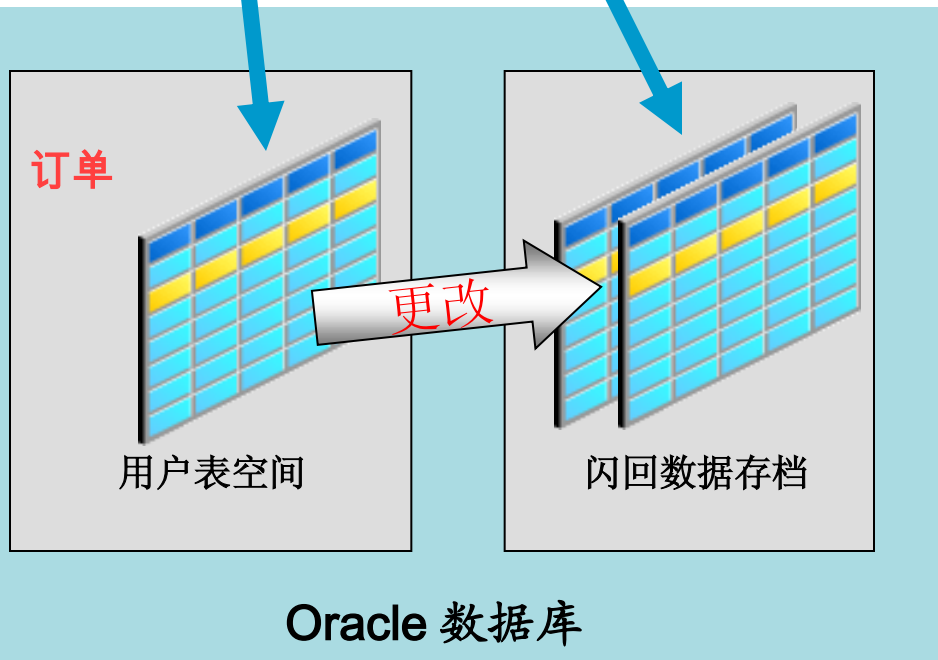
- 时间点恢复的新策略
- 闪回日志捕获更改的块的所有旧版本
  - 看作一种连续备份
  - 按日志重做将数据库恢复到某一时间
  - 仅恢复更改的块
- 快速 - 在几分钟而不是几小时内恢复
- 简单 - 单个命令即可进行恢复

将数据库闪回至 ‘2:05 PM’

数据库的“倒回”按钮

# Flashback Data Archive(Total Recall)

Select \* from orders  
AS OF  
'Midnight 31-Dec-2004'



- 长期保留 - 数年
- 自动存储对闪回数据存档中选定表做出的所有更改
  - 存档不能修改
  - 按保留策略清除老旧数据
- 使用闪回查询查看截至任何时间的表内容
- 使用
  - 更改跟踪/长期历史
  - ILM
  - 审计
  - 合规性



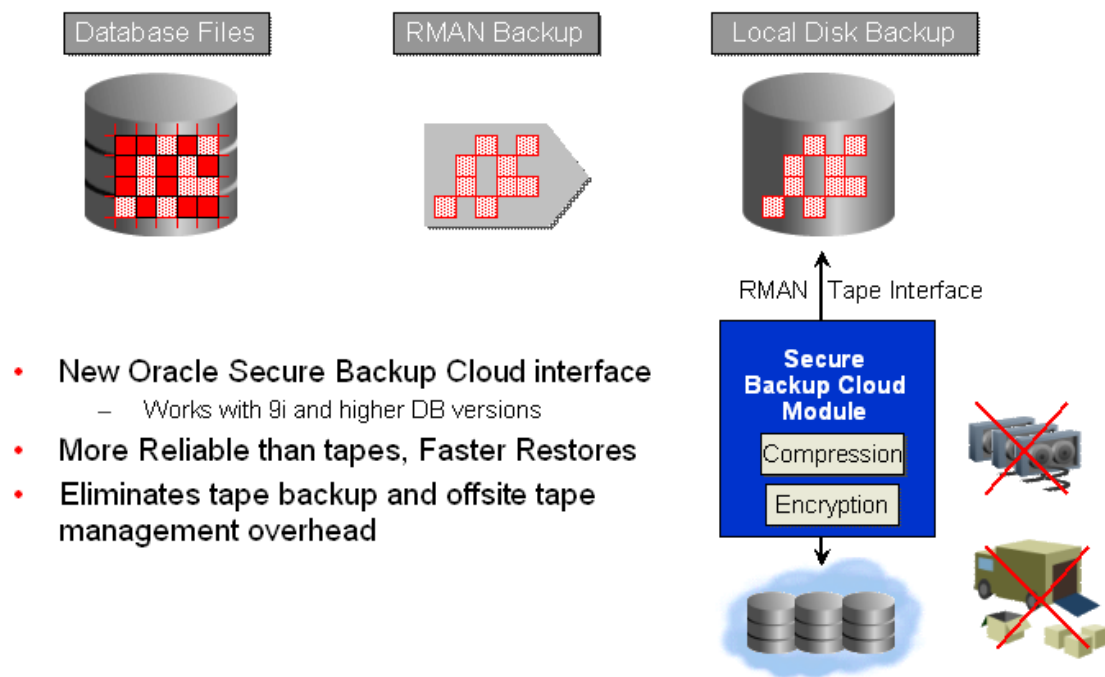
# 产品功能: Secure Backup

## 安全备份

### 功能:

- 为Oracle数据库提供了中央磁带备份管理平台，支持领先厂商的200多种不同的磁带设备
- 通过对备份到磁带的数据进行加密来防止备份磁带丢失或被盗时出现滥用敏感数据的情况
- 提供物理、虚拟和云备份目标
- 以最快的速度将Oracle数据库备份到磁带

### Oracle Database Backup in the Cloud



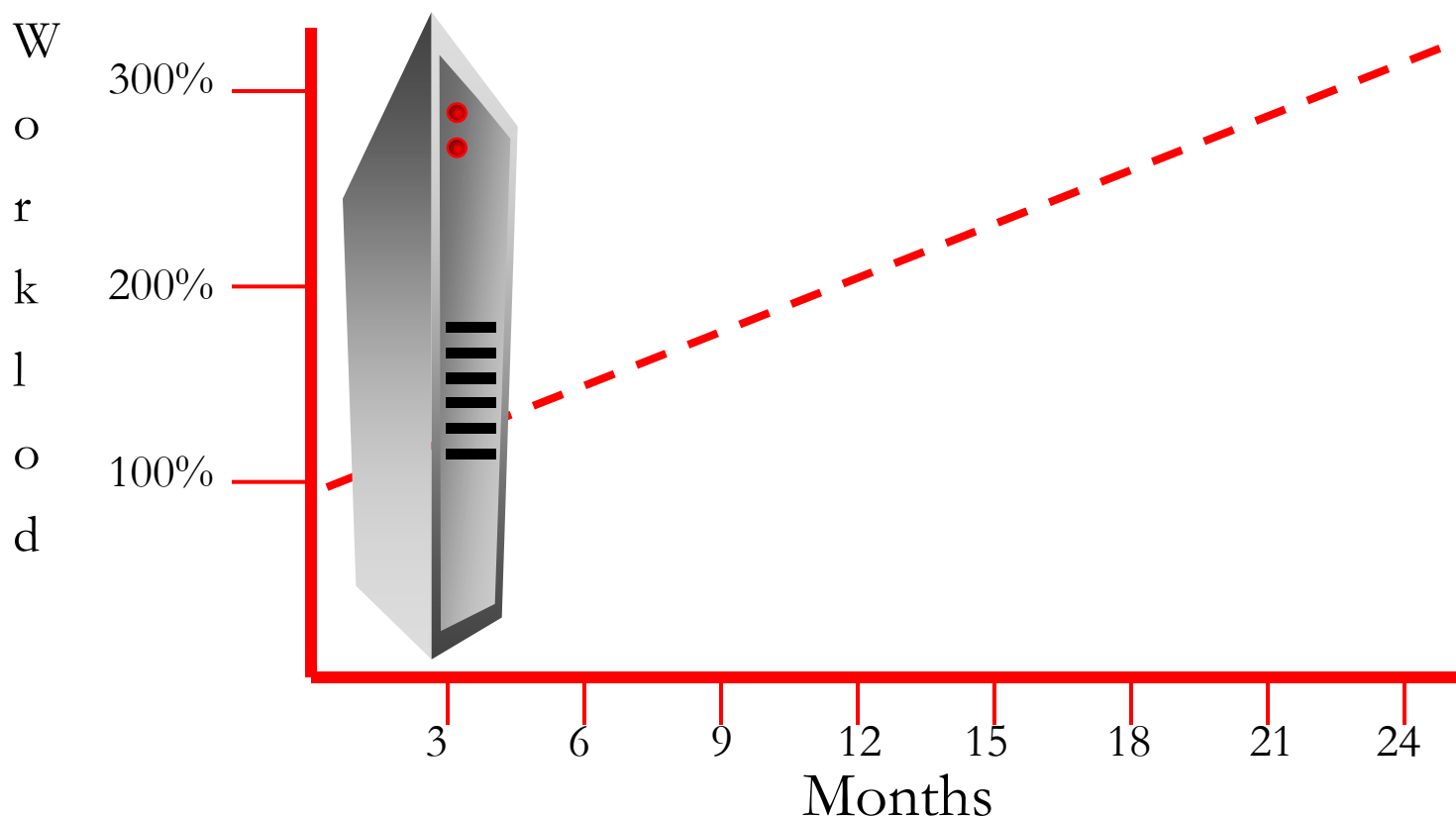
# 提纲

- Oracle数据库最高可用架构
- ➔ Oracle数据库集群技术
- Oracle数据库分区技术
- Oracle GoldenGate技术
- 问答



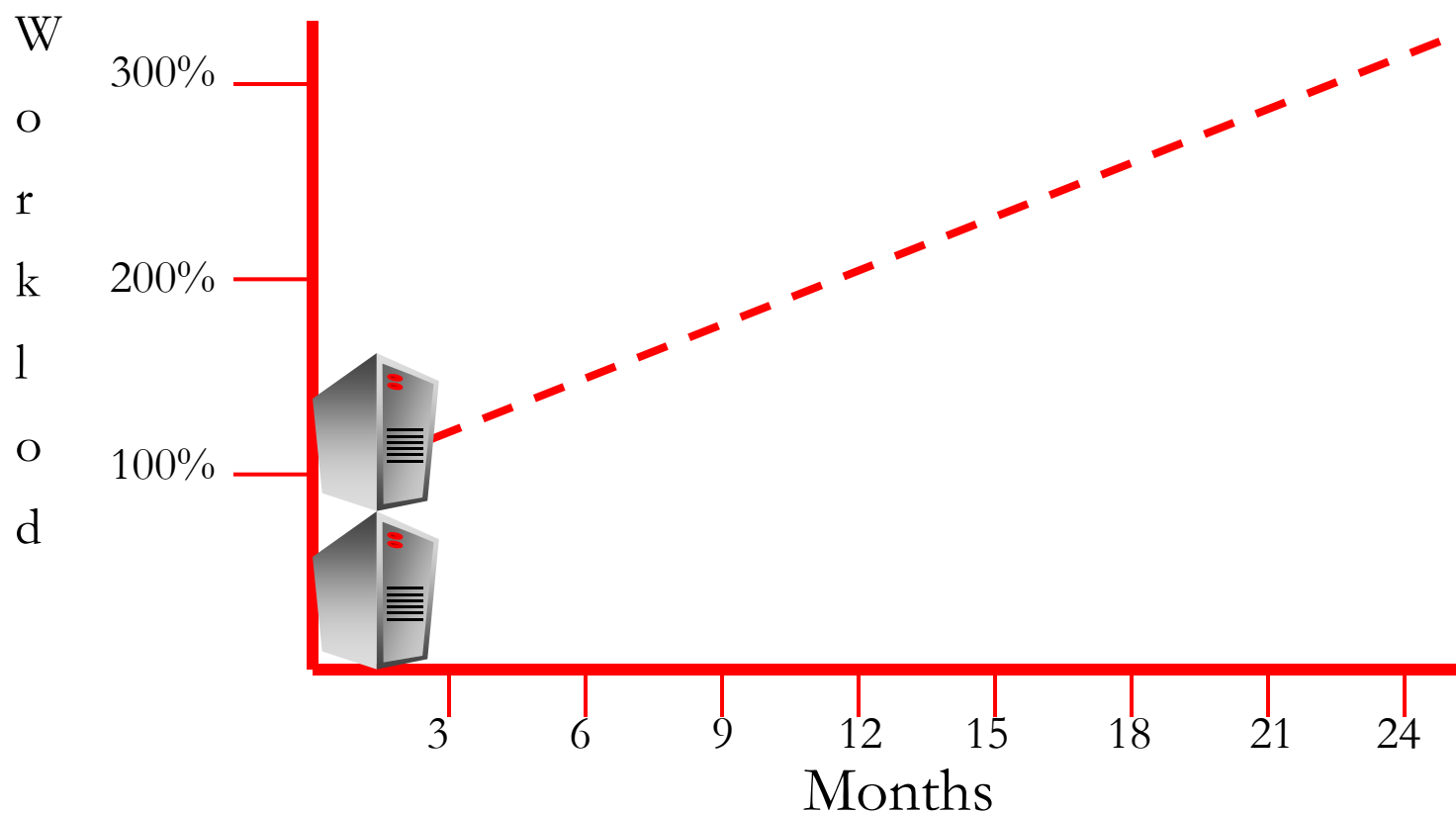
# Pay-as-you-grow 扩展性

为将来的工作量提供处理能力



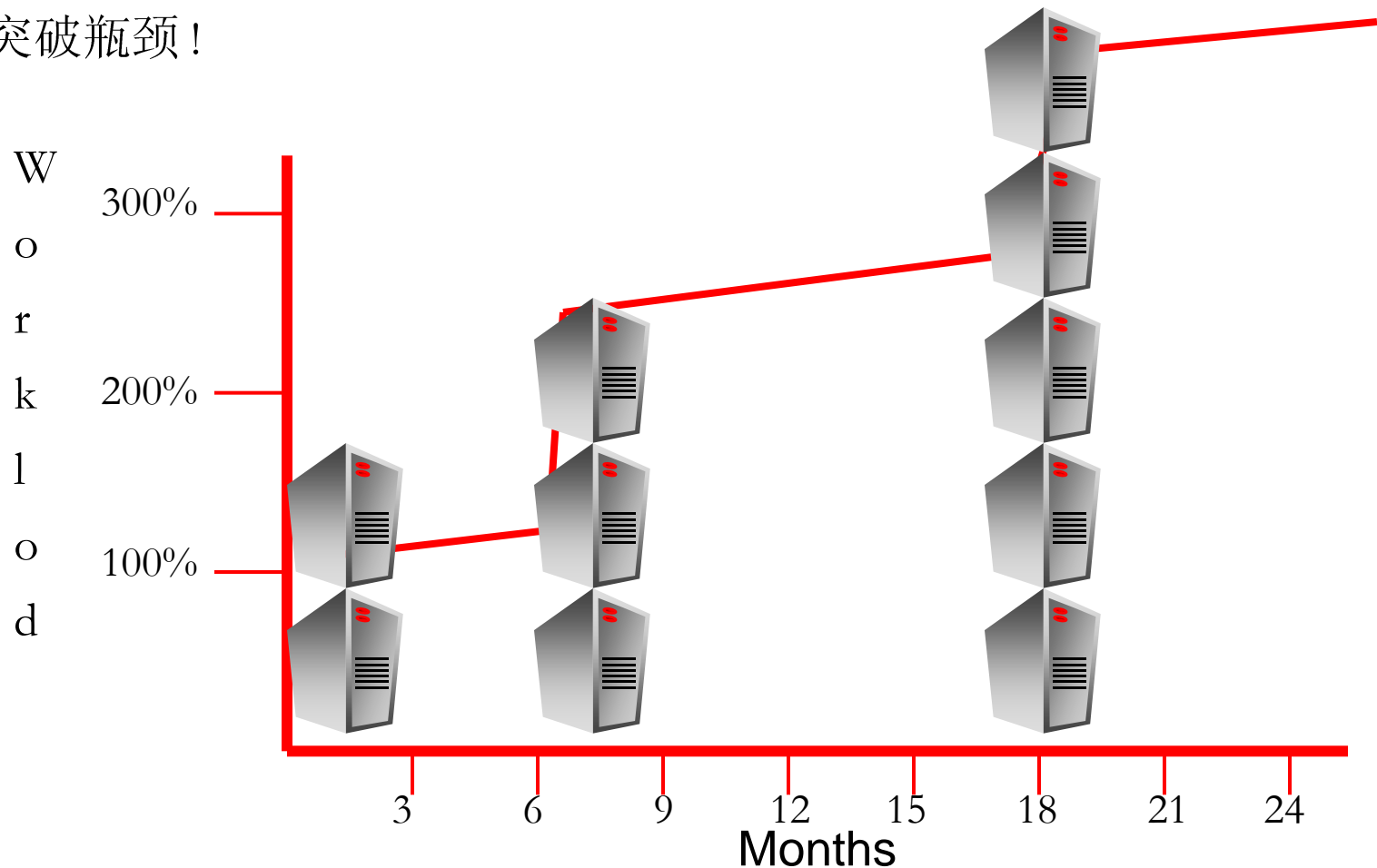
# Pay-as-you-grow 扩展性

如果使用 RAC，提供当前工作量的支持

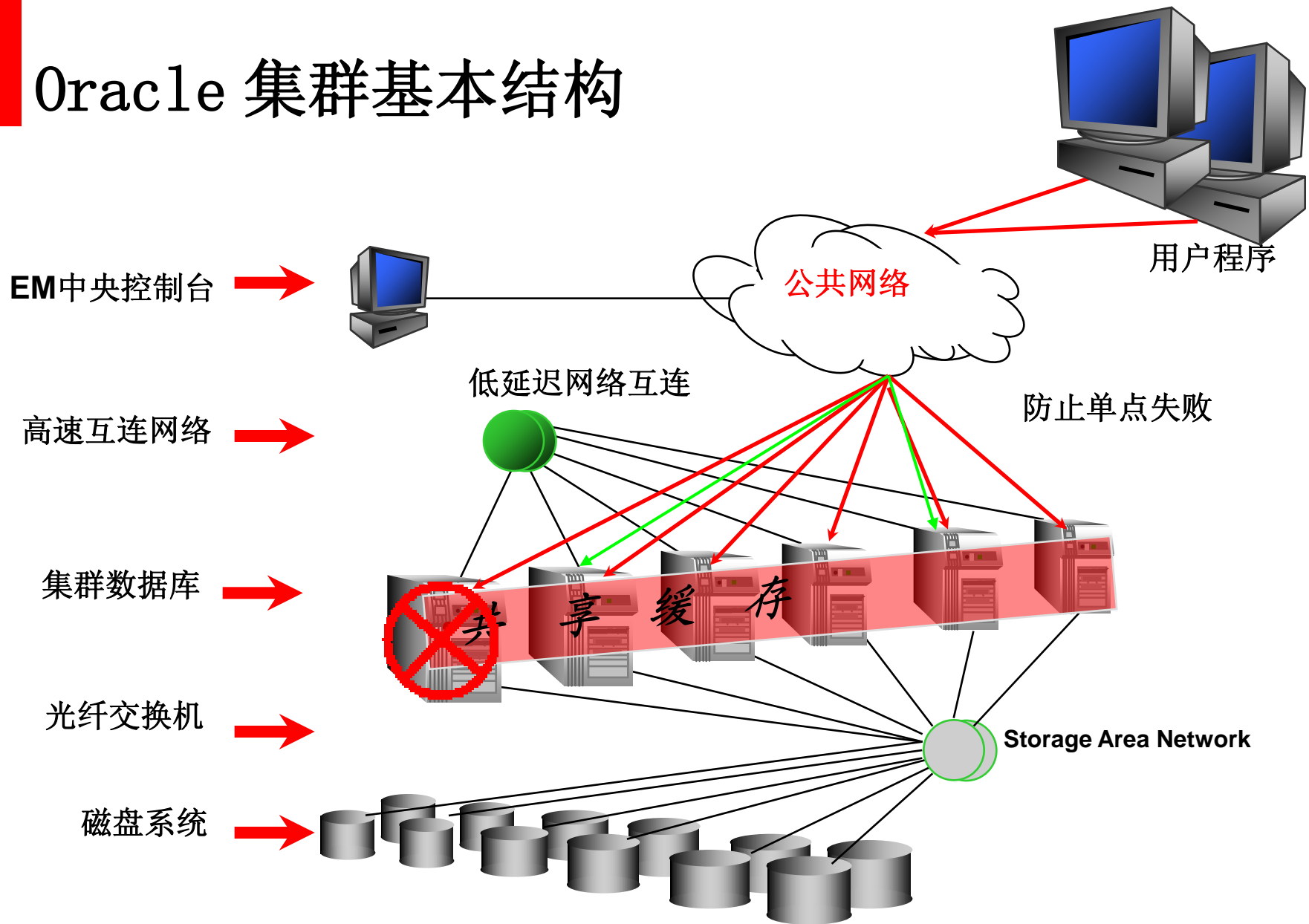


# Pay-as-you-grow 扩展性

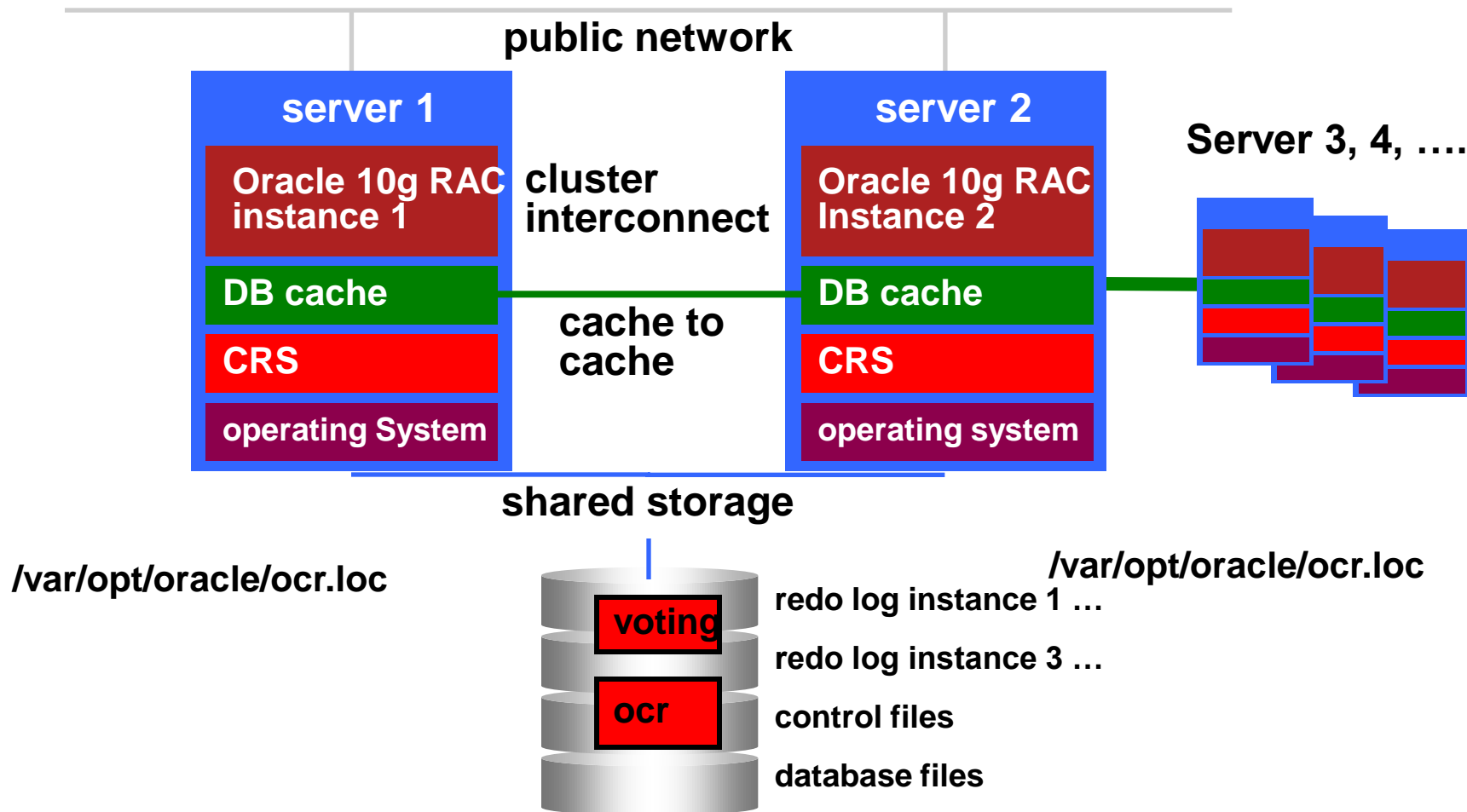
RAC 可以扩展适应未来工作量需求  
突破瓶颈！



# Oracle 集群基本结构



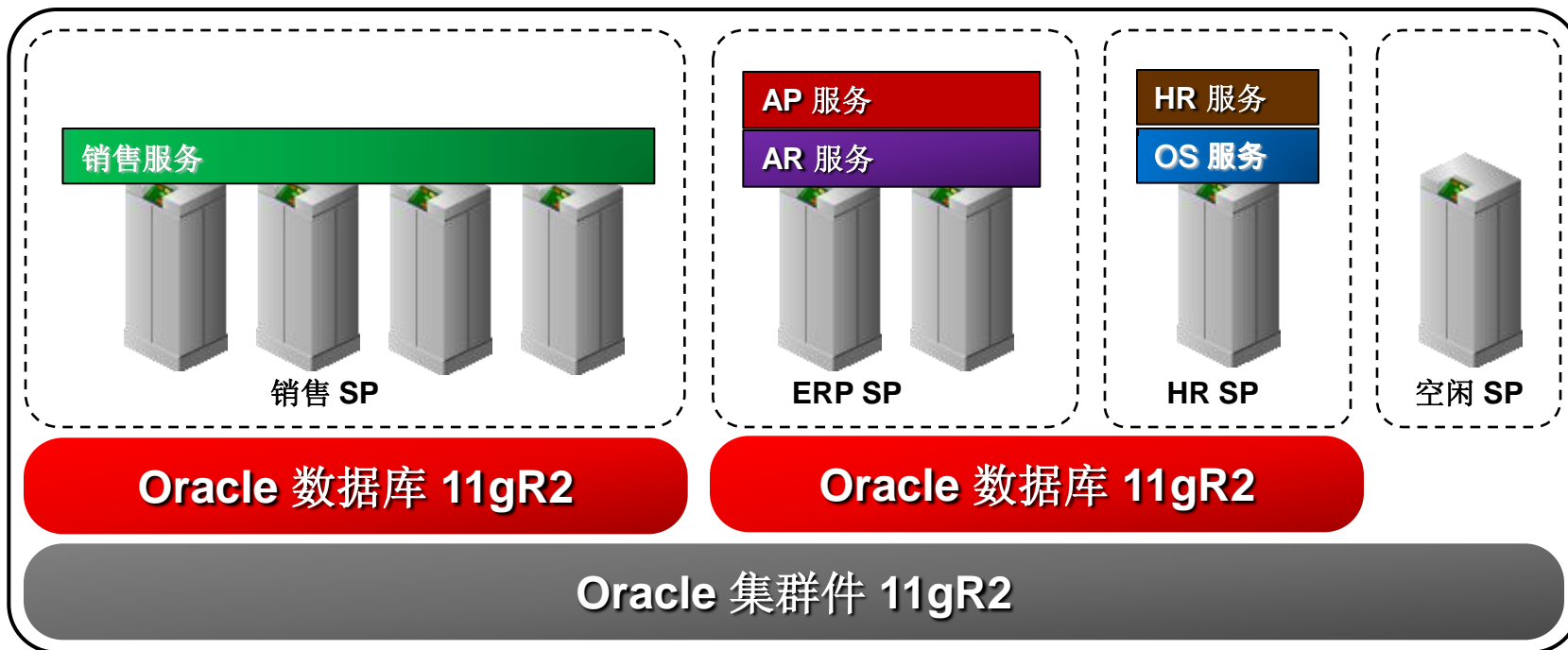
# Oracle 10g RAC架构



# Oracle 集群件 11g R2

## — 数据库服务器池

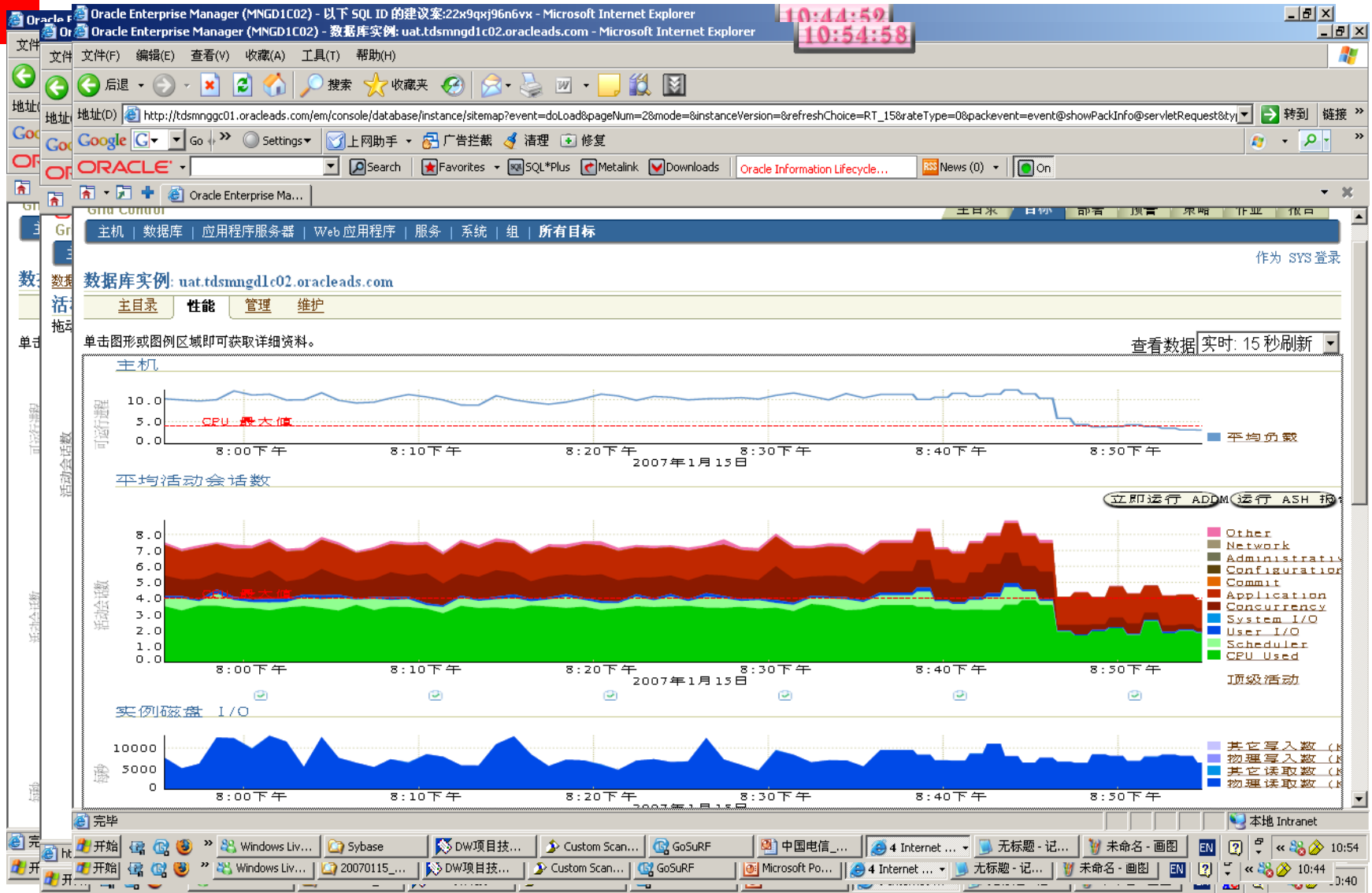
11.2新特性



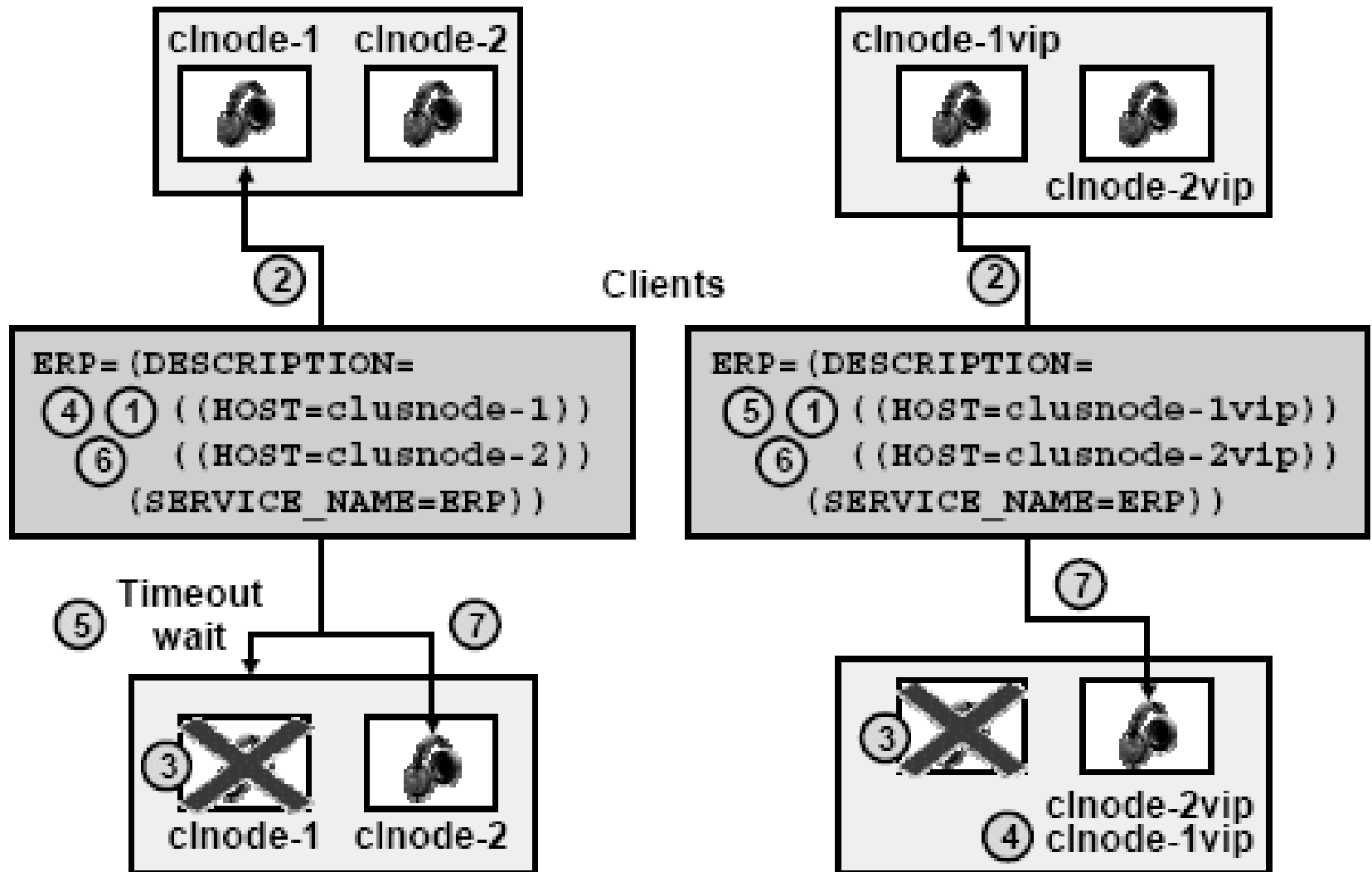
- 由策略动态管理服务的数据库资源
- 通过 Min、Max 和 Importance 属性控制可用性
- 轻松管理托管多个数据库的大型集群
- 如果池中服务器数低于最小值，则重新配置集群



# 性能监控和优化变得如此简单

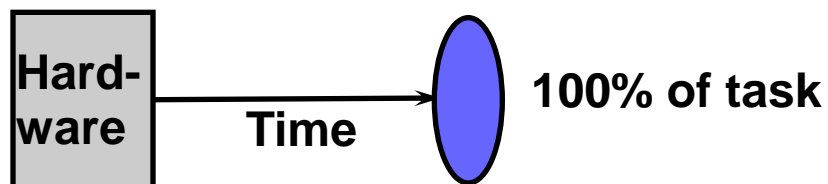


# Virtual IP Addresses and RAC

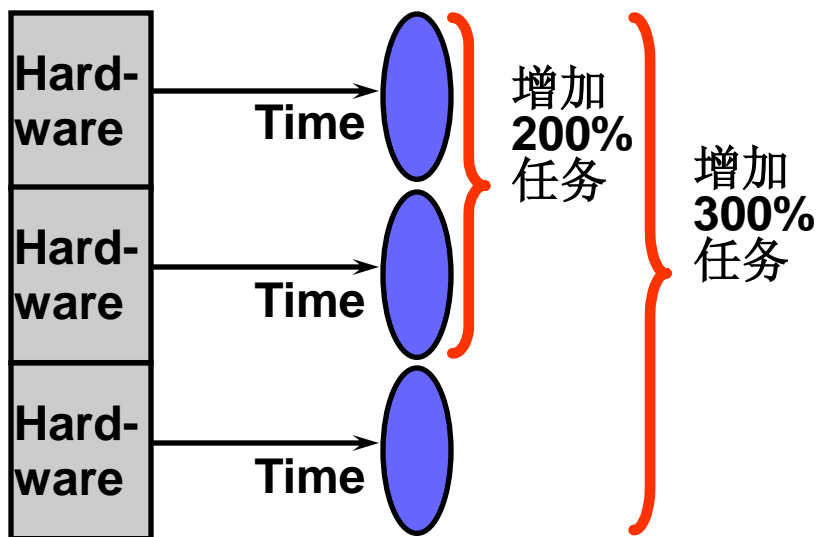


# 扩展性和性能提升

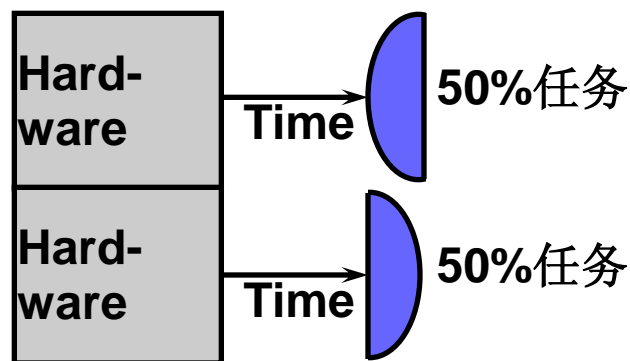
初始系统



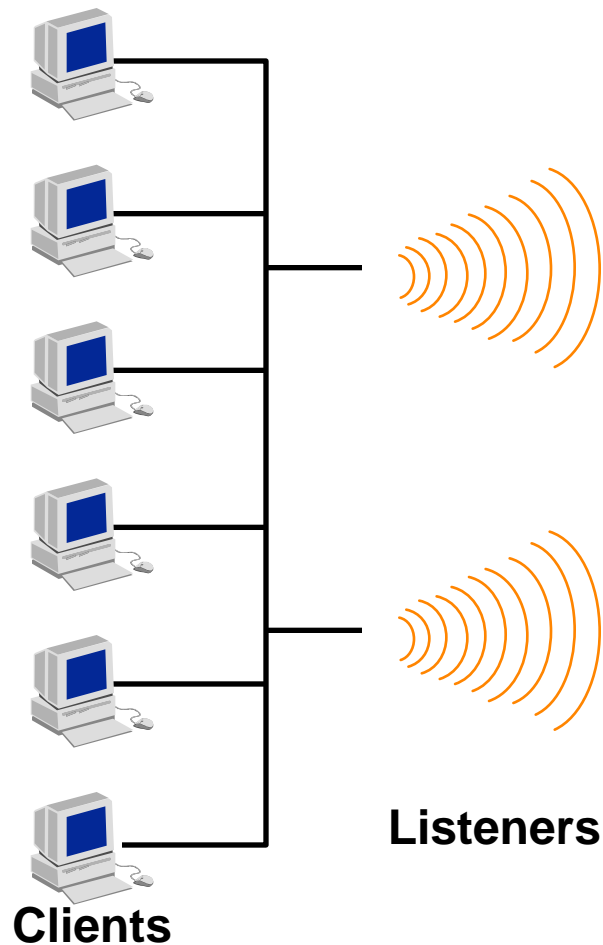
集群系统扩展性



集群系统性能提升



# Client Load Balancing



```
sales.us.acme.com=  
(DESCRIPTION=  
  (ADDRESS LIST=  
    (LOAD BALANCE=on)  
  (ADDRESS=  
    (PROTOCOL=tcp)  
    (HOST=sales1)  
    (PORT=1521))  
  (ADDRESS=  
    (PROTOCOL=tcp)  
    (HOST=sales2)  
    (PORT=1521)))  
(CONNECT DATA=  
  (SERVICE_NAME=  
    sales.us.acme.com) ) )
```

# TAF 配置举例

**RAC =**

**(DESCRIPTION =**

**(ADDRESS\_LIST =**

**(ADDRESS = (PROTOCOL = TCP)(HOST = 192.168.0.1)(PORT = 1521))**

**(ADDRESS = (PROTOCOL = TCP)(HOST = 192.168.0.2)(PORT = 1521))**

**(LOAD\_BALANCE = yes)**

**(FAILOVER = true)**

**)**

**(CONNECT\_DATA =**

**(SERVICE\_NAME = hh)**

**(failover\_mode =**

**(type = select)**

**(method = basic)**

**(retries = 20)**

**(delay = 15)**

**)**

**)**

**)**

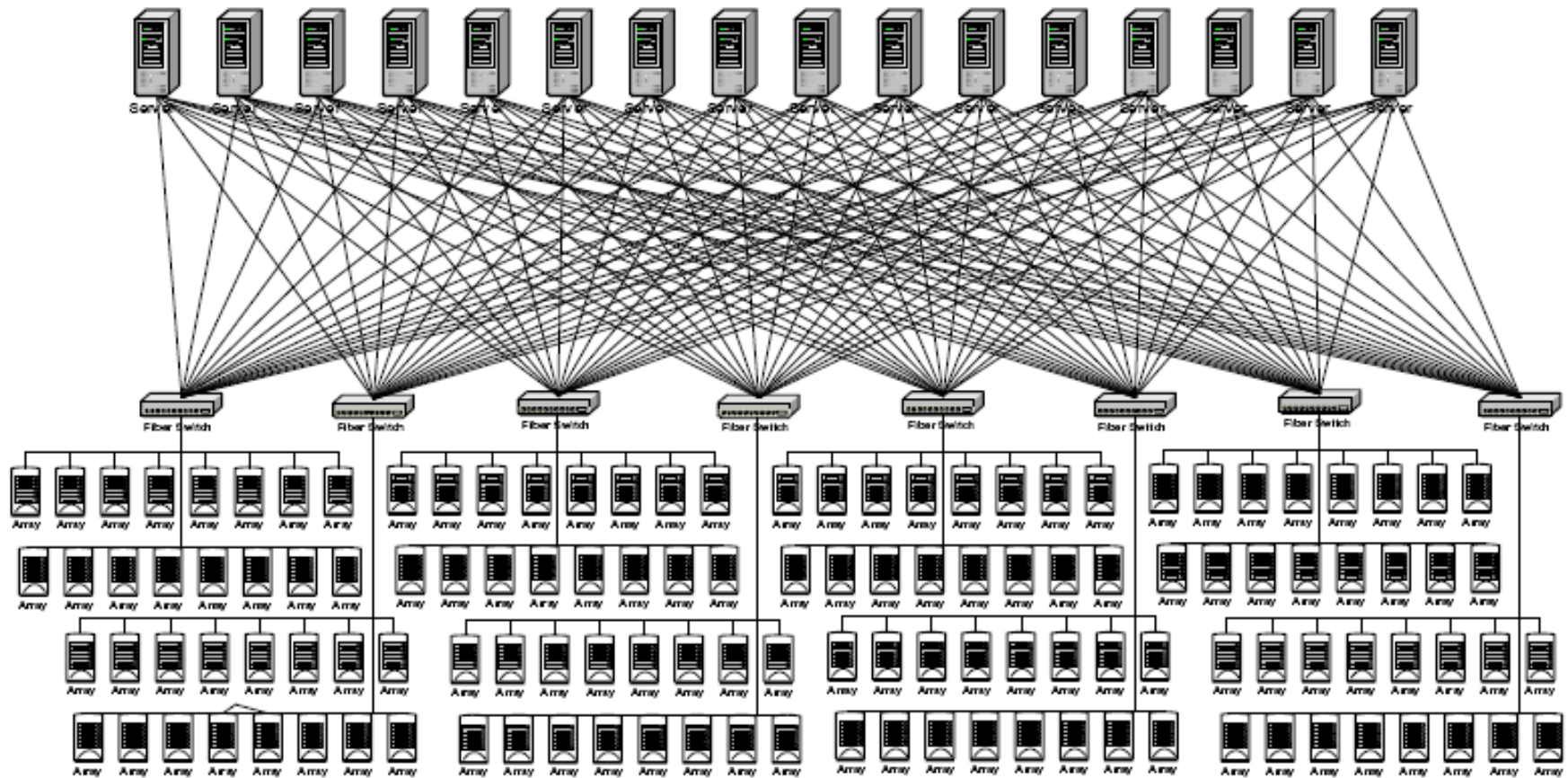
# Oracle Call Interface

TAF 支持的程序接口:

- OCI programs
- Java JDBC thick drivers (OCI drivers)
- ODBC connections
- SQL\*Plus
- Select statements

# Amazon 16节点RAC案例

16 - 4 Processor Servers with 8 Fiber Ports  
8 Fiber Switches  
128 - Fiber Arrays (1 port each)  
1920 - 72G 15K Disk Drives

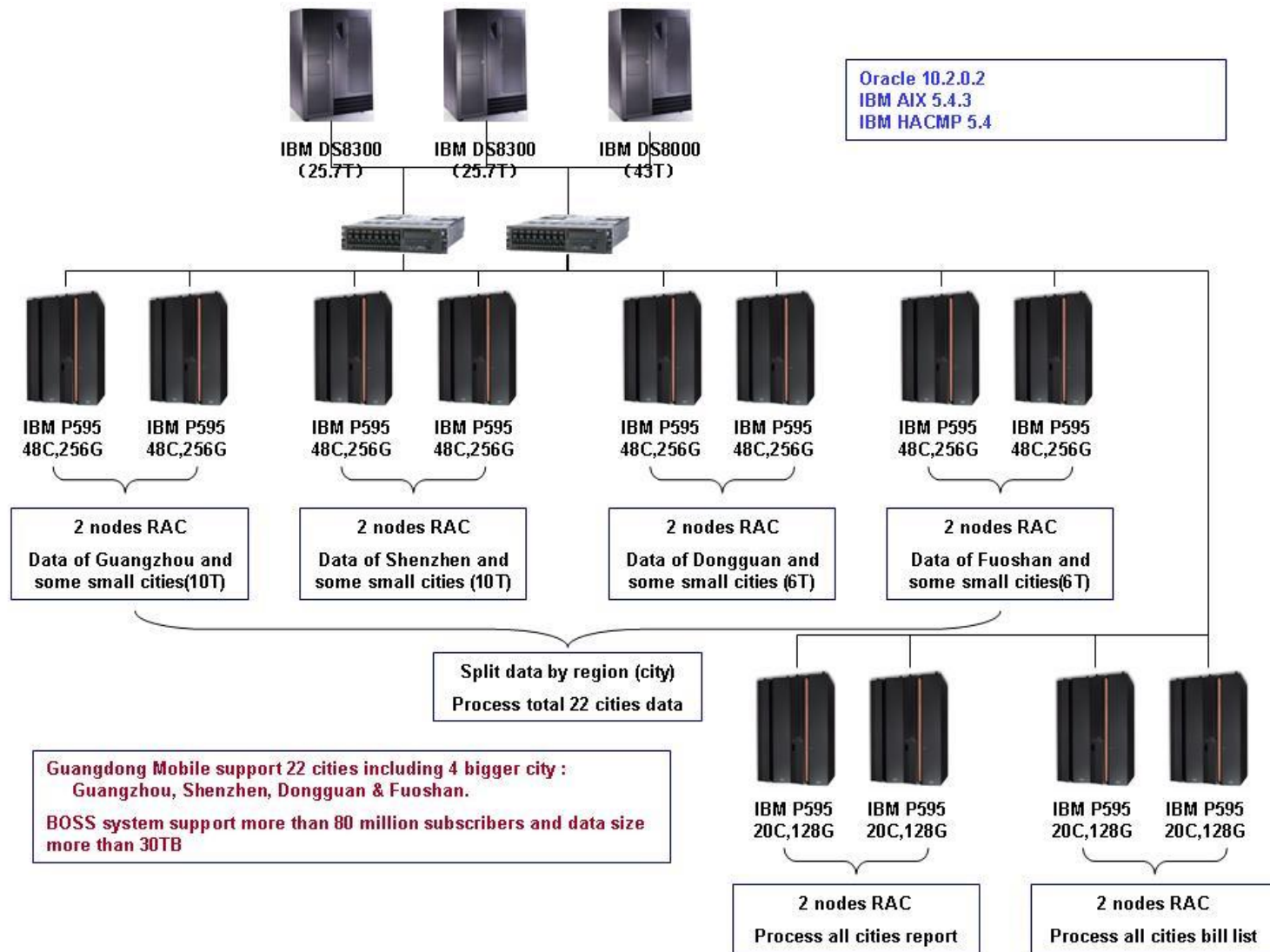


# 河南移动经分系统

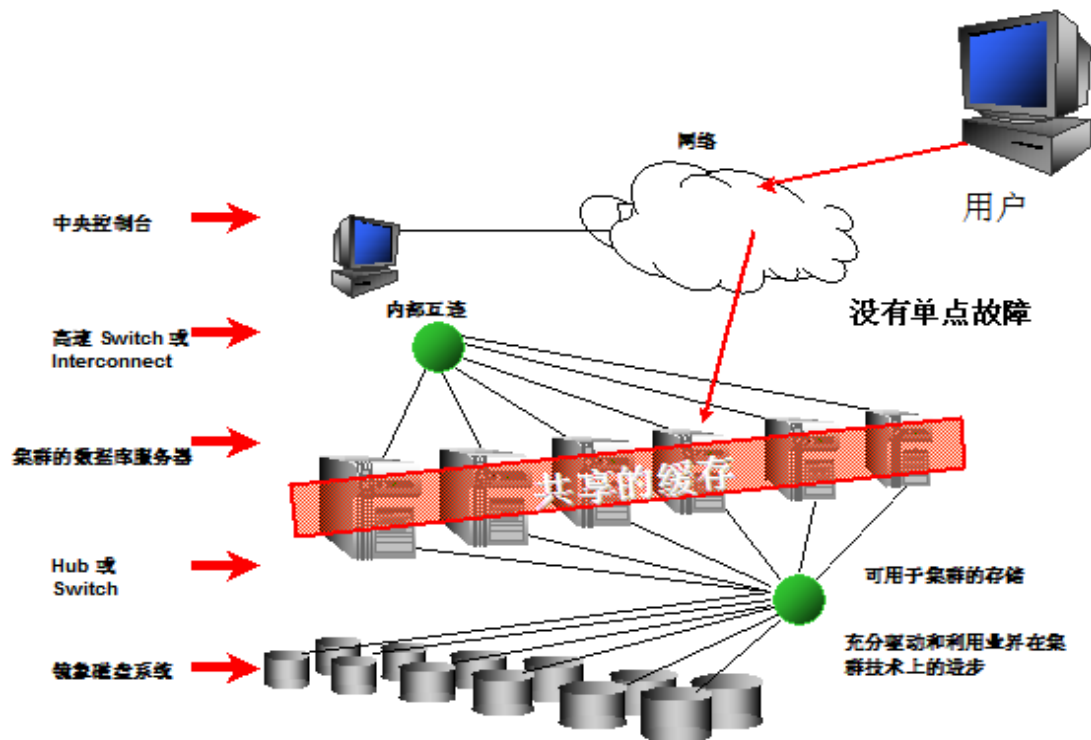
- 采用 Oracle 11gr1 v11.1.0.7, RAC, partitioning
- 经分系统A
  - 4台 HP Superdome
  - 每服务器 72core, 512G 内存
  - 数据量 200TB
  - Infiniband 内部互联
- 经分系统B
  - 3台 HP Superdome
  - 每服务器 72core, 512G 内存
  - 数据量 几十TB
  - Infiniband 内部互联



# 广东移动BOSS系统架构



# 真正应用集群技术



## 适用场景分析

- ❑ 对高并发、高性能并行计算要求的系统
- ❑ 要求稳定性且7\*24不间断连续运行的系统
- ❑ 对计算资源虚拟化需求，保障良好在线扩展性

## 性能提升

- ❑ 利用多个服务器集群互联，提升整个数据库的并行计算处理能力
- ❑ 实现负载均衡，充分发挥集群性能优势

## 高可用性

- ❑ 避免单机故障带来的停机
- ❑ 内建的容错功能。故障时可以自动切换到正常的节点
- ❑ 应用程序访问无需额外开发和修改

## 扩展性能力

- ❑ 可根据容量需求的增长为集群添加节点，从而能通过逐步扩充来节约成本
- ❑ 正常最多可实现100节点的集群数据库环境

# 提纲

- Oracle数据库最高可用架构
- Oracle数据库集群技术
- ➔ Oracle数据库分区技术
- Oracle GoldenGate技术
- 问答



# Oracle独有的表分区技术

- 表分区：范围、hash、列表、复合分区

添加分区

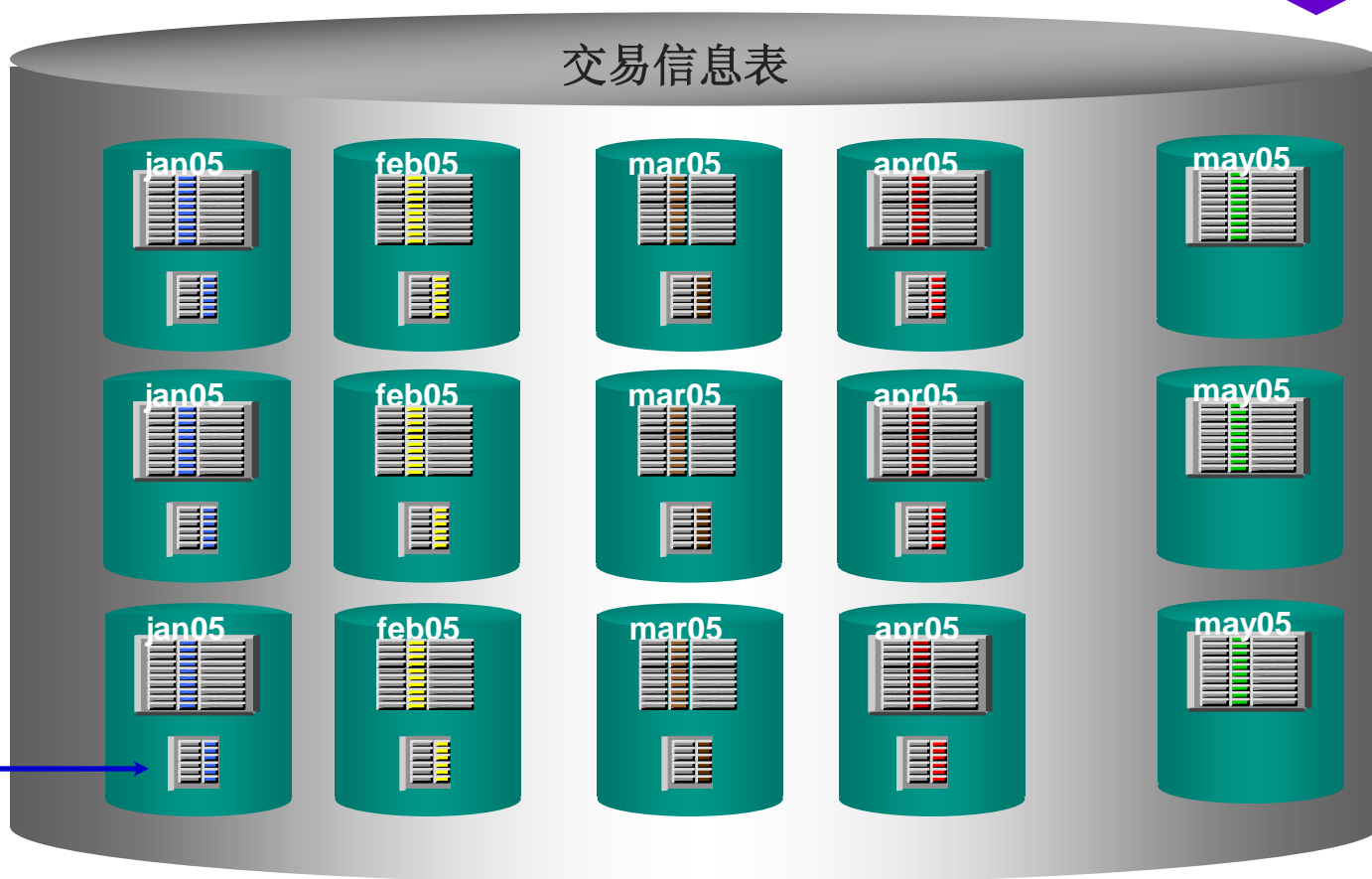
删除分区

北京

上海

黑龙江

本地索引



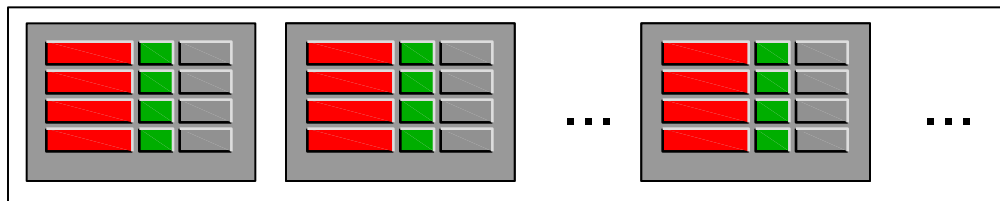
# Oracle 分区

## 10 年创新

数据库版本	核心功能
Oracle8 数据库	范围分区
Oracle8i 数据库	散列和组合分区
Oracle9i 数据库	列表分区
Oracle9i 数据库第 2 版	范围-列表组合分区
Oracle 数据库 10g	全局散列索引
Oracle 数据库 10g 第 2 版	每个表 1M 分区
Oracle 数据库 11g	引用分区 间隔分区 虚拟列分区 系统分区 完全组合分区

# 引用分区

**ORDERS 表**



2006 年 1 月 2006 年 2 月

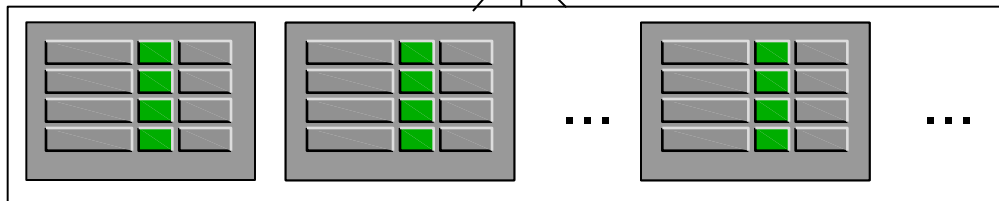
RANGE(**order\_date**)

主键 **order\_id**

按引用分区

通过 PK-FK 关系继承分区键

**LINEITEMS 表**



2006 年 1 月 2006 年 2 月

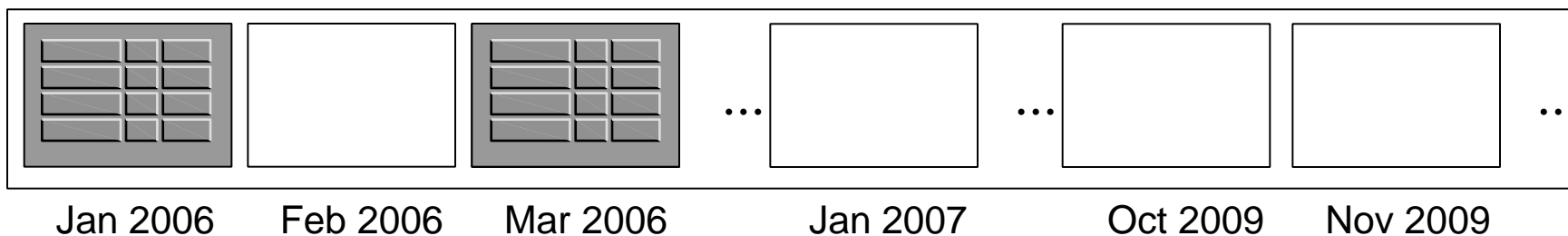
RANGE(**order\_date**)

外键 **order\_id**

# 间隔分区

```
CREATE TABLE sales (order_date DATE, ...)
PARTITION BY RANGE (order_date)
INTERVAL (NUMTOYMINTERVAL(1, 'month'))
(PARTITION p_first VALUES LESS THAN ('01-JAN-2006'));
```

Table SALES



↑ New segment is automatically allocated

```
INSERT INTO sales (order_date DATE, ...)
VALUES ('04-MAR-2006', ...);
```

# 虚拟列分区

Base table with all attributes ...

... is extended with the virtual (derived) column

... and the virtual column is used as partitioning key

```
CREATE TABLE accounts
(acc_no      number(10)      not null,
 acc_name    varchar2(50)    not null, ...
 acc_branch  number(2)       generated always as
                                (to_number(substr(to_char(acc_no),1,2)))
partition by list (acc_branch) ...
```

12500	Adams	12
12507	Blake	12
12666	King	12
12875	Smith	12

...

32320	Jones	32
32407	Clark	32
32758	Hura	32
32980	Phillips	32



# 系统分区

- 在数据表上允许应用来控制分区的使用
- 提供分区的优点，但分区策略以及数据的放置，由应用程序来控制
- 不使用分区关键字（与传统的表分区不同）
- 不提供传统意义上的分区裁剪功能

```
CREATE TABLE systab (c1 integer, c2 integer)
PARTITION BY SYSTEM
(
  PARTITION p1 TABLESPACE tbs_1,
  PARTITION p2 TABLESPACE tbs_2,
  PARTITION p3 TABLESPACE tbs_3,
  PARTITION p4 TABLESPACE tbs_4
);
```

## Using the system-partitioned table:

```
INSERT INTO systab PARTITION (p1) VALUES (4,5);
```

```
INSERT INTO systab PARTITION (1) VALUES (4,5);
```

```
ALTER TABLE systab MERGE PARTITIONS p1,p2
  INTO PARTITION p1;
```

# 完全组合分区

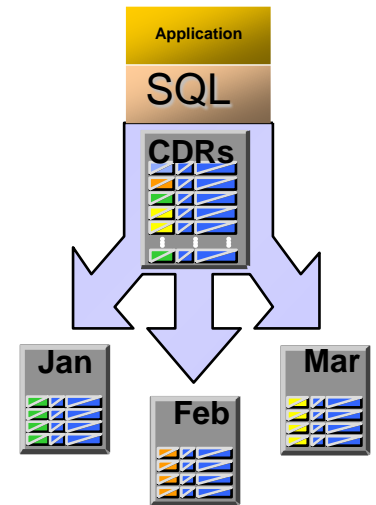
- 二维分区模式
- Oracle 数据库 11g 中的扩展

	范围	列表	散列
范围	11g	9i	8i
列表	11g	11g	11g

- 例如列表-范围：
  - 按国家/地区分区，然后再按星期分区
  - 按业务范围分区，然后再按星期分区

# 分区建议

- SQL Access Advisor 建议两种分区：
  - Hash
  - Interval
- Interval Partitioning
  - Range Partitioning的扩展
  - 自动建立range partitions
  - 插入新数据，自动分配Segment
  - 不需要手工创建新分区
  - 自动创建和维护Local index
- Hash partition 建议一般成对出现



# 提纲

- Oracle数据库最高可用架构
- Oracle数据库集群技术
- Oracle数据库分区技术
- ➔ Oracle GoldenGate技术
- 问答



恐怖事件

9. 11



NBC News

# 应急备份系统起源



系统割接业务中断



BOSS挂起业务中断



服务开通平台堵塞业务无法办理



极易引起客  
户投诉!!!

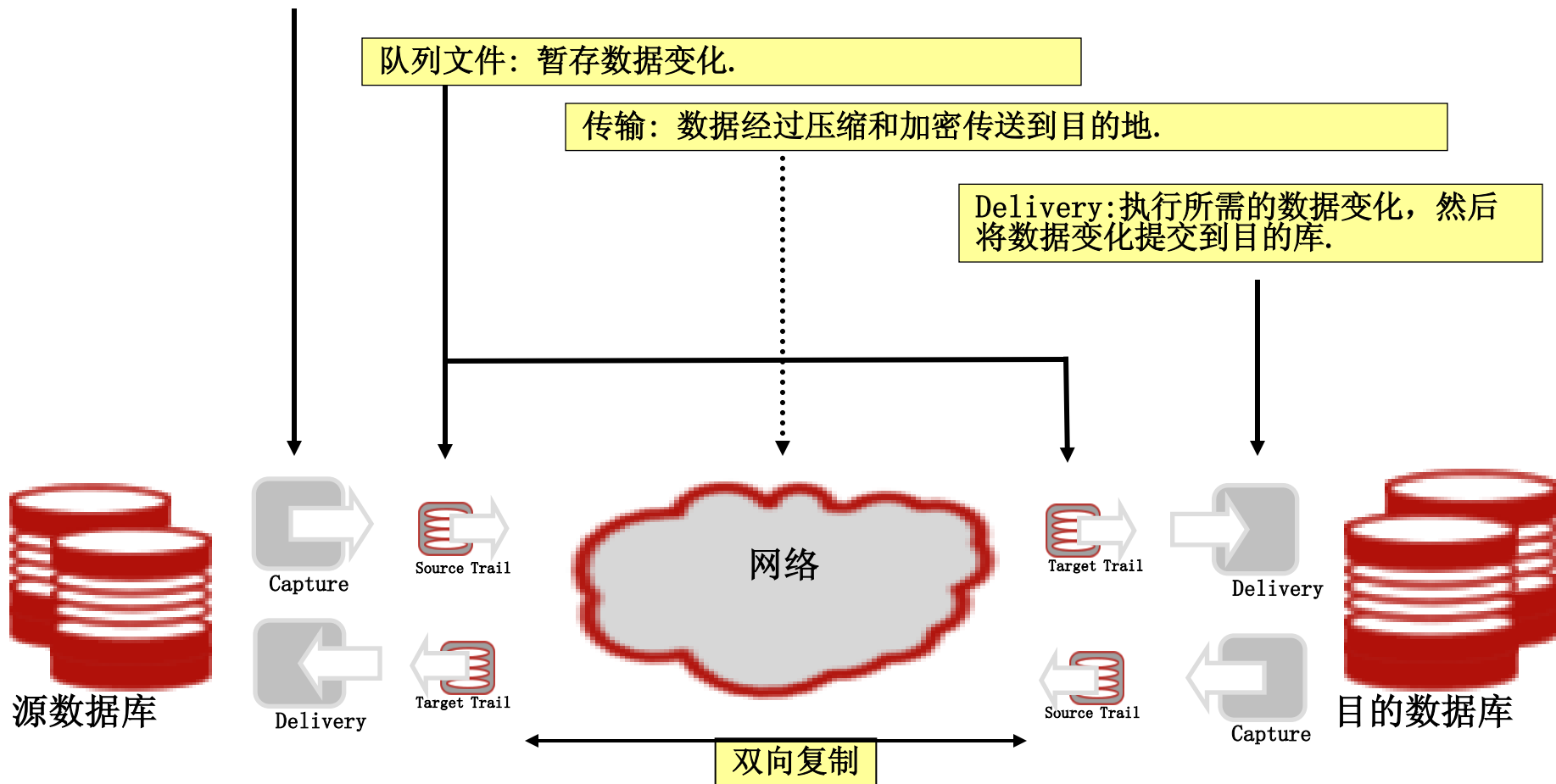
# GoldenGate工作原理

Capture: 实时读取交易日志捕捉数据变化并可实现过滤.

队列文件: 暂存数据变化.

传输: 数据经过压缩和加密传送到目的地.

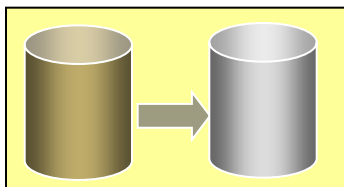
Delivery: 执行所需的数据变化, 然后将数据变化提交到目的库.



# 灵活的拓扑结构

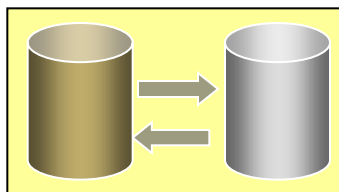
单向复制

备份、查询



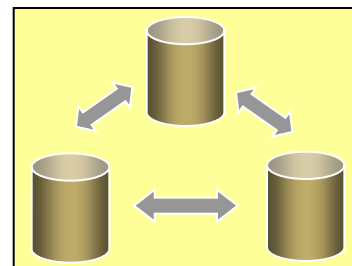
双向复制

灾备、双业务中心



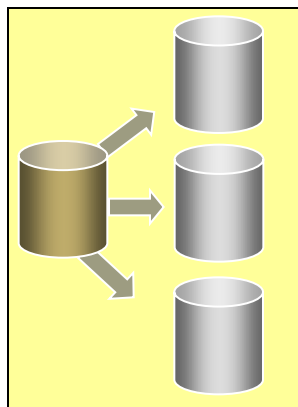
点对点

多业务中心



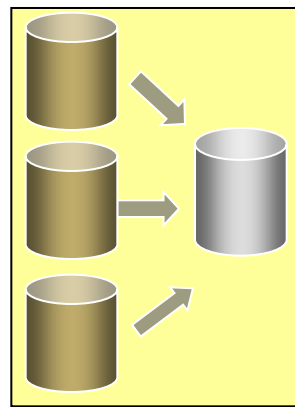
广播复制

数据分发



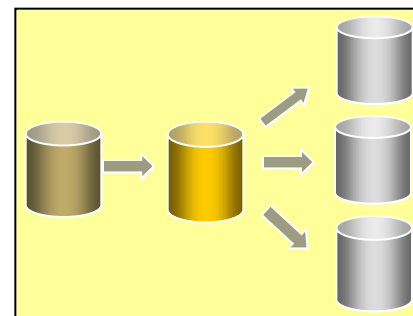
集中复制

数据仓库、N+1



多级复制

层次化企业数据

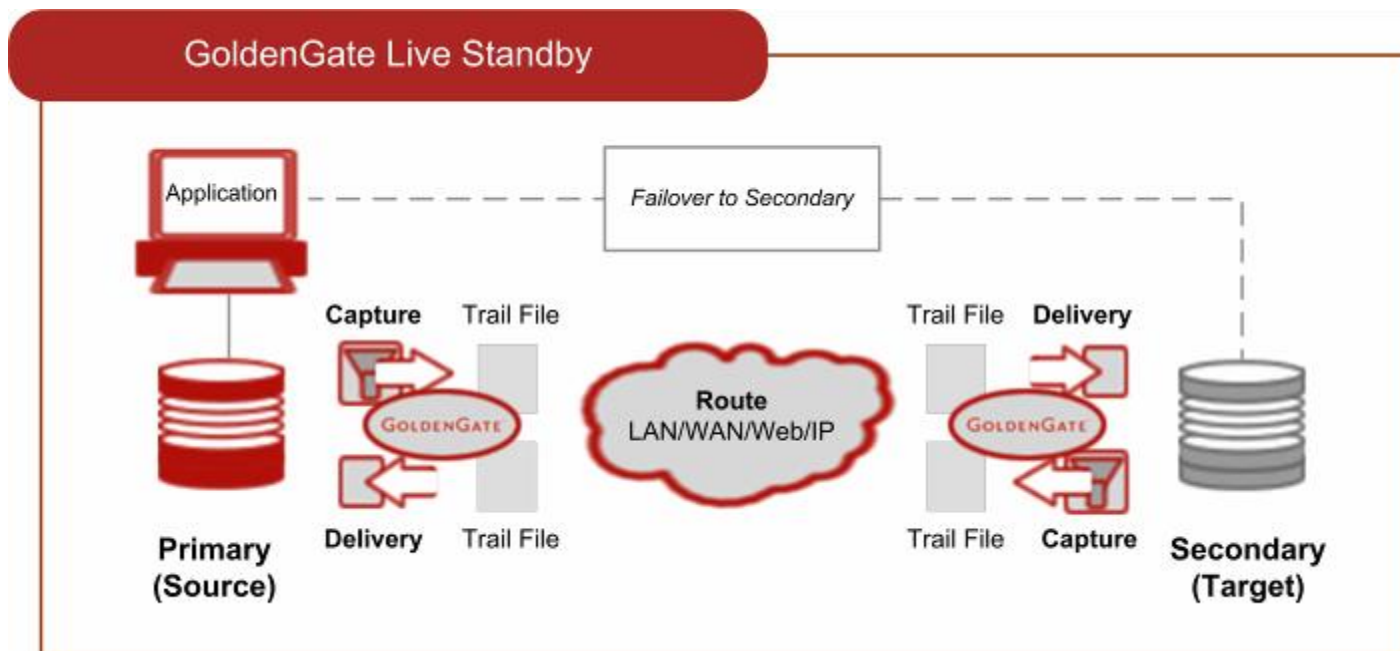




# GoldenGate软件支持的环境

Databases	O/S and Platforms
<p>Capture:</p> <ul style="list-style-type: none"><li>Oracle</li><li>DB2 UDB</li><li>Microsoft SQL Server</li><li>Sybase ASE</li><li>Teradata</li><li>Ingres</li><li>Enscribe</li><li>SQL/MP</li><li>SQL/MX</li></ul> <p>Delivery:</p> <ul style="list-style-type: none"><li>All listed above</li><li>MySQL, HP Neoview, Netezza</li><li>And ODBC compatible databases</li></ul>	<ul style="list-style-type: none"><li>Windows 2000, 2003, XP</li><li>Linux</li><li>Sun Solaris</li><li>HP NonStop</li><li>HP-UX</li><li>HP TRU64</li><li>IBM AIX</li><li>IBM z/OS</li></ul>

# 高可用性: 应急备份



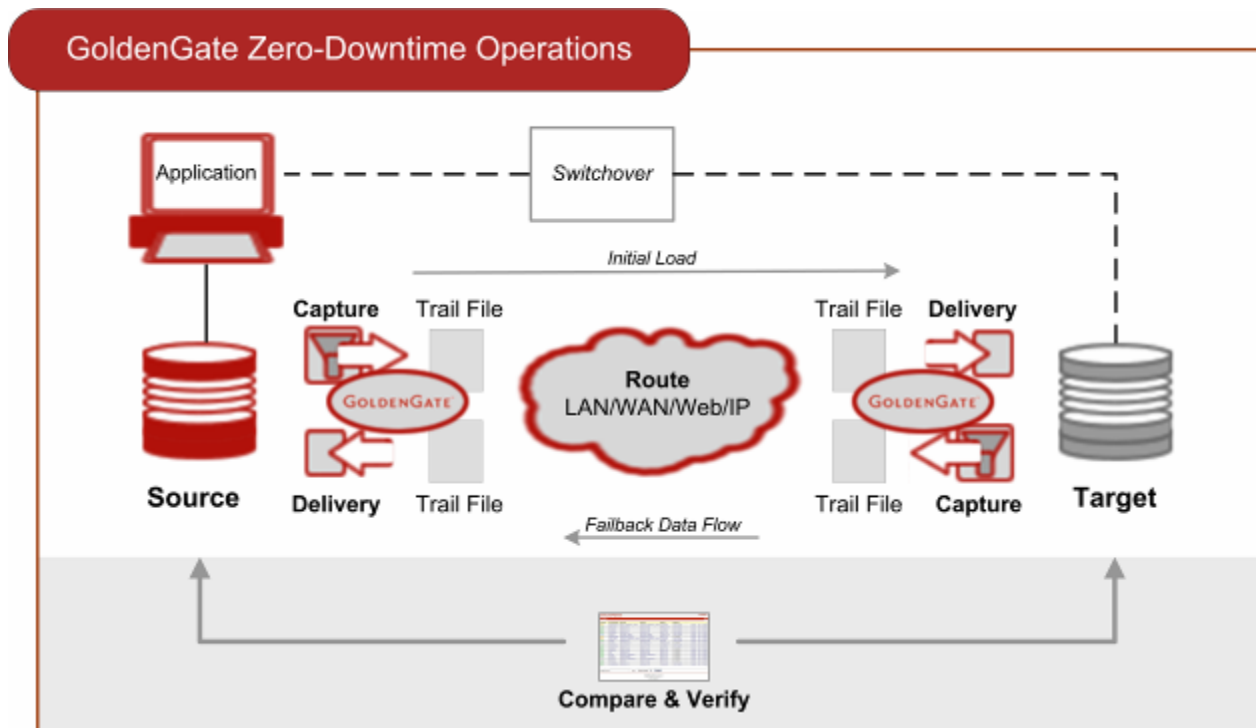
实现:

- 快速恢复和切换
- 最小化数据损失
- 重新同步主备两端数据

用途:

- 在非计划性停机时保持业务连续

# 高可用性: 减少计划内停机



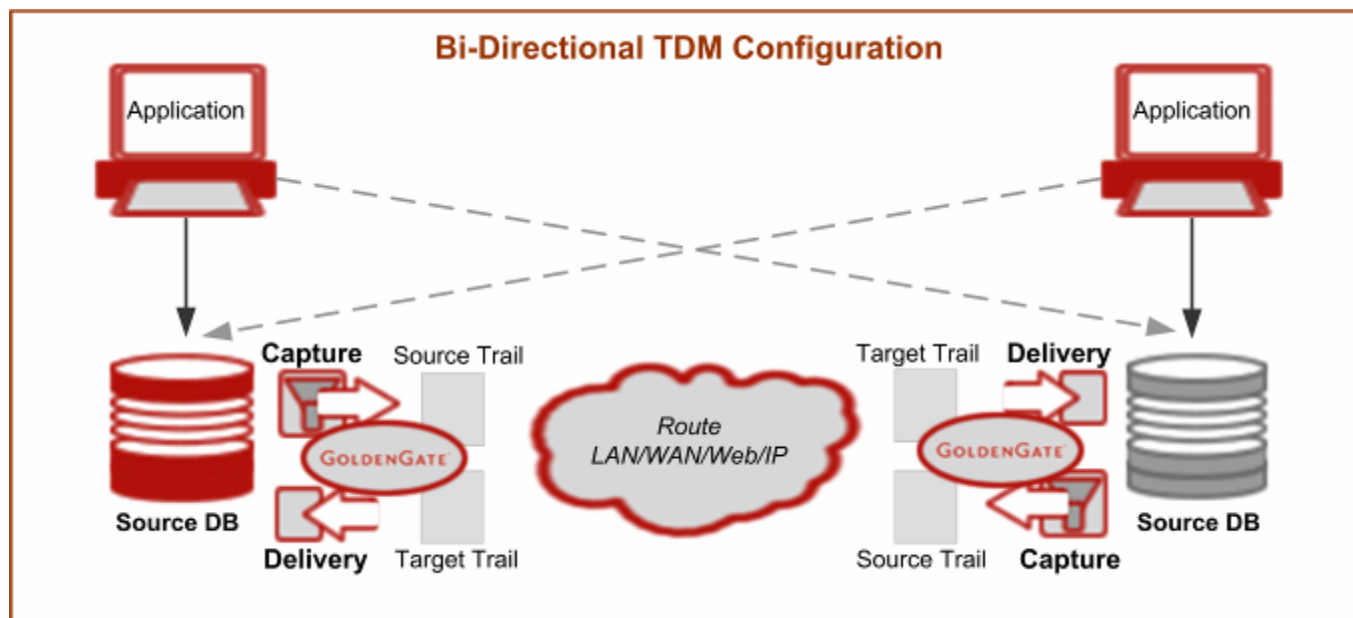
实现:

- 保障业务零或者近似零停机
- 滚动升级
- 降低业务中断带来的损失

用途:

- 保障系统/应用/数据库在升级、移植和维护期间业务的可用性

# 高可用性: 双业务中心 (Active-Active)

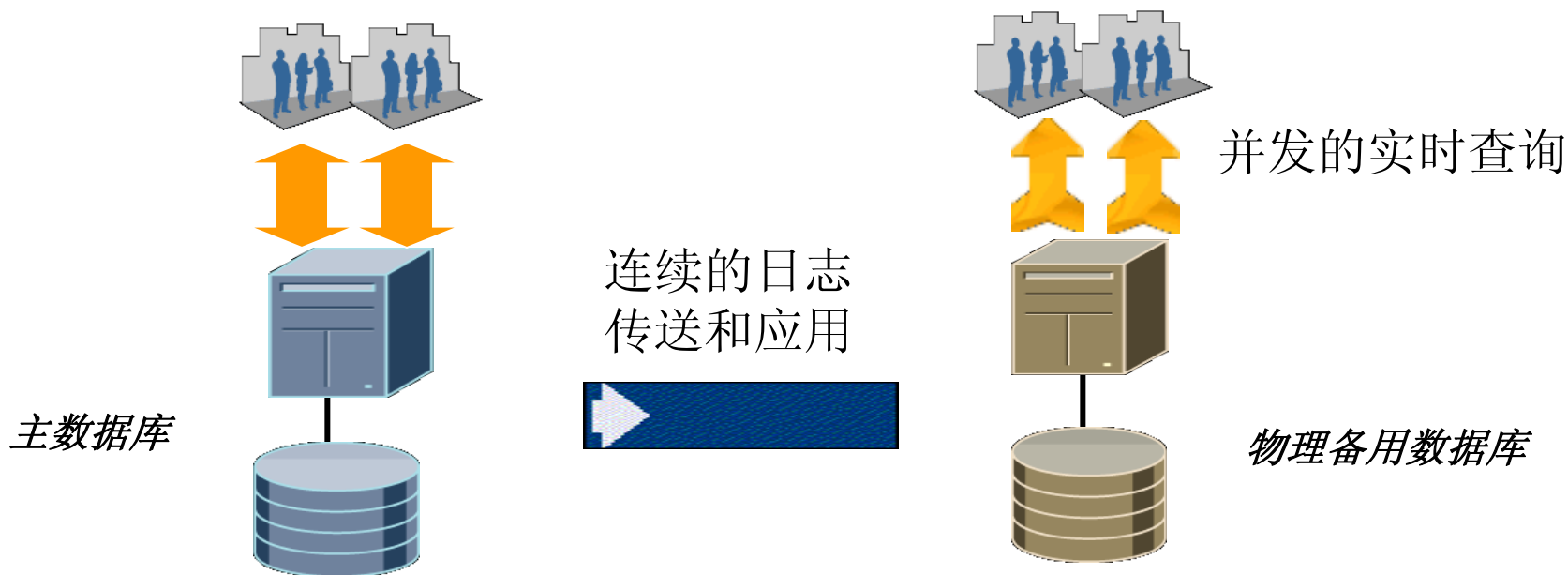


实现:

- 负载均衡, 提高系统整体性能
- 连续可用, 快速的容灾接管
- 冲突检测和处理

# 物理备用用于实时查询(11g)

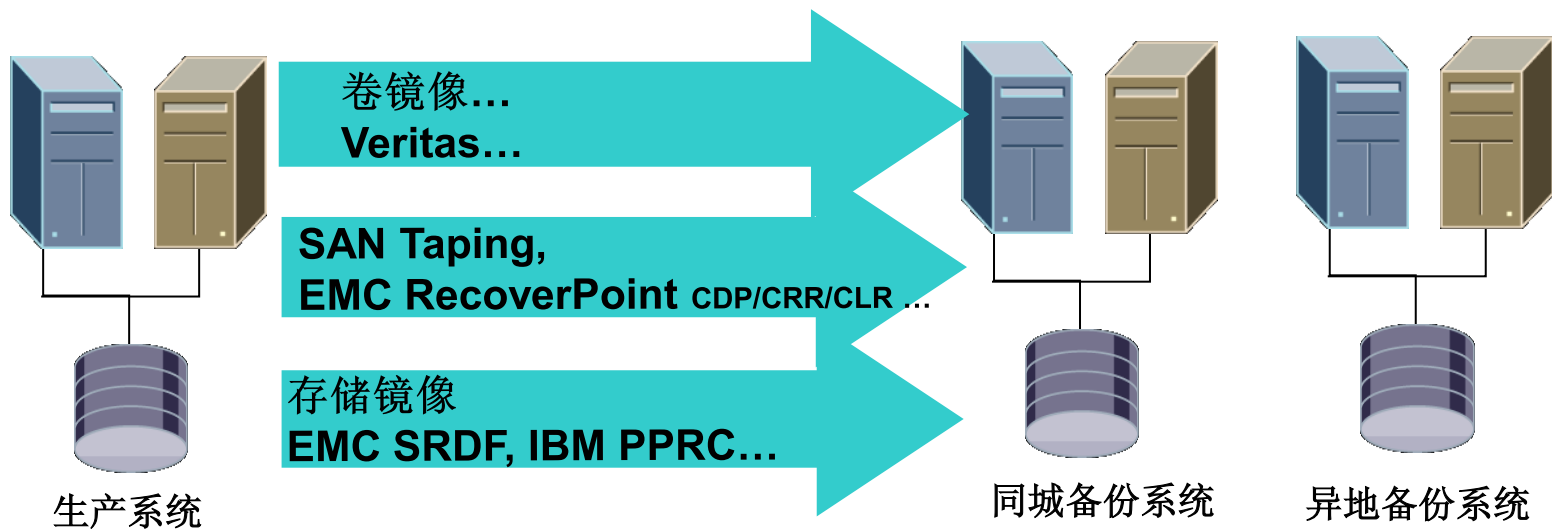
## Active DataGuard



- 在物理备用上的只读查询和日志应用可以同时进行
  - 在主 / 备用数据库上支持 RAC
  - 查询可以获得事物一致性结果
  - 可以处理所有数据类型

# 备份中心建设方式1 -

## 系统:系统 传统同构、准同构方式

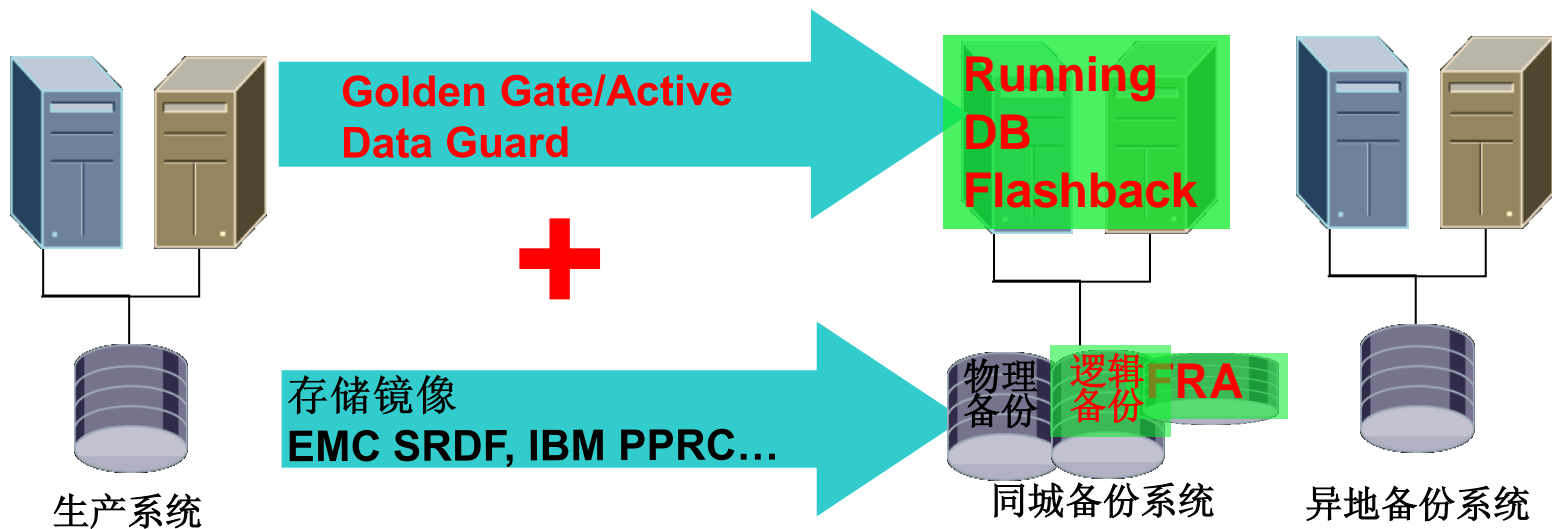


### 问题:

- 备份库非活动状态，难以确定可用性
- 错误(如坏块)被透明传递，会导致两套系统同时不能用
- 缺乏有效的方法去解决逻辑错误:
  - 人为错误
  - 联机数据修改
  - 软件升级
- 子系统众多，异构，设备型号难以统一，难以统一管理
- 独立的1:1备份，冗余度大，2xn
- 空闲的备份硬件

# 备份中心建设方式2 -

## 系统:系统 阵列同构 + OGG/Active DG方式



### Golden Gate提供:

- 一个备份库处于活动状态, 可以确定可用性
- 错误(如坏块)不会被透明传递, 事务级完整性

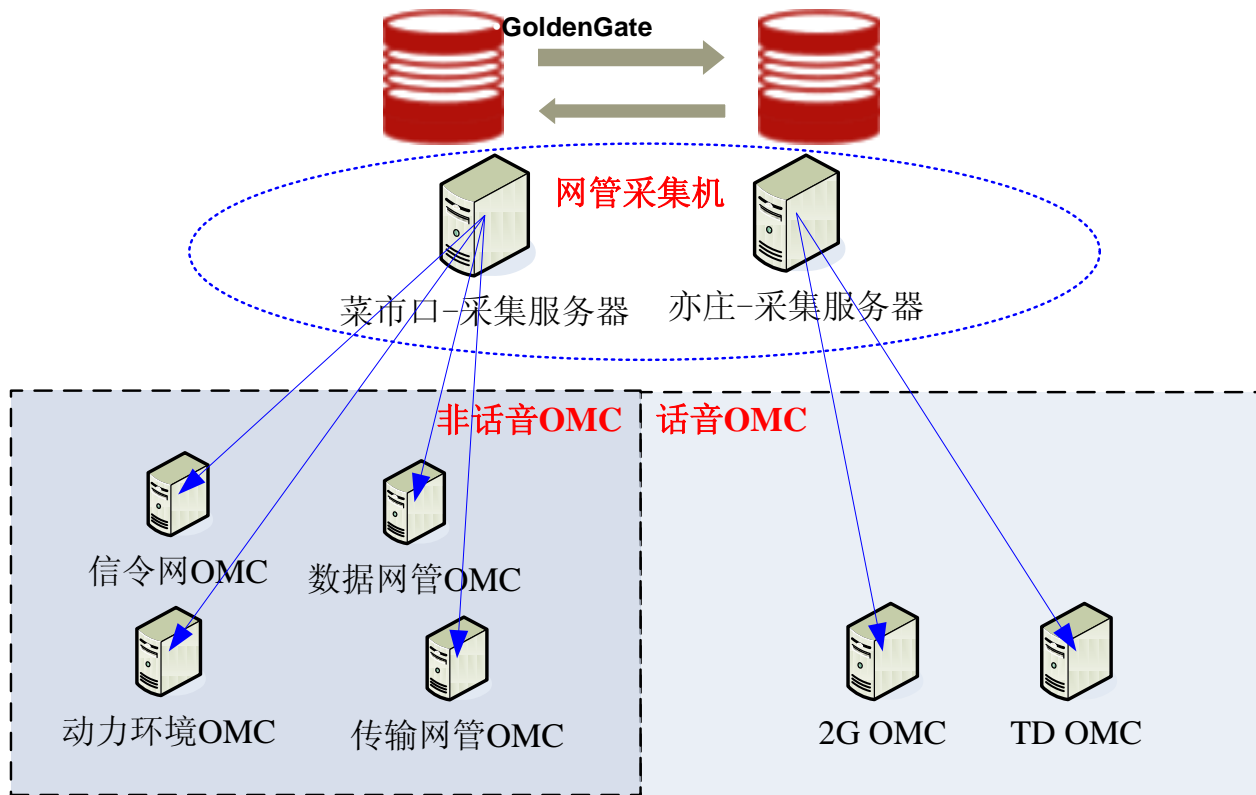
### 备份库Flashback解决逻辑错误:

- 人为错误
- 联机数据修改
- 软件升级
- 充分利用备份硬件 (报表、RMAN、测试...)
- 磁盘阵列备份的同步特性,  $RPO = 0$

### 系统改造要求:

- 备份系统分配一定的磁盘容量
- 增加OGG/Active DG选项

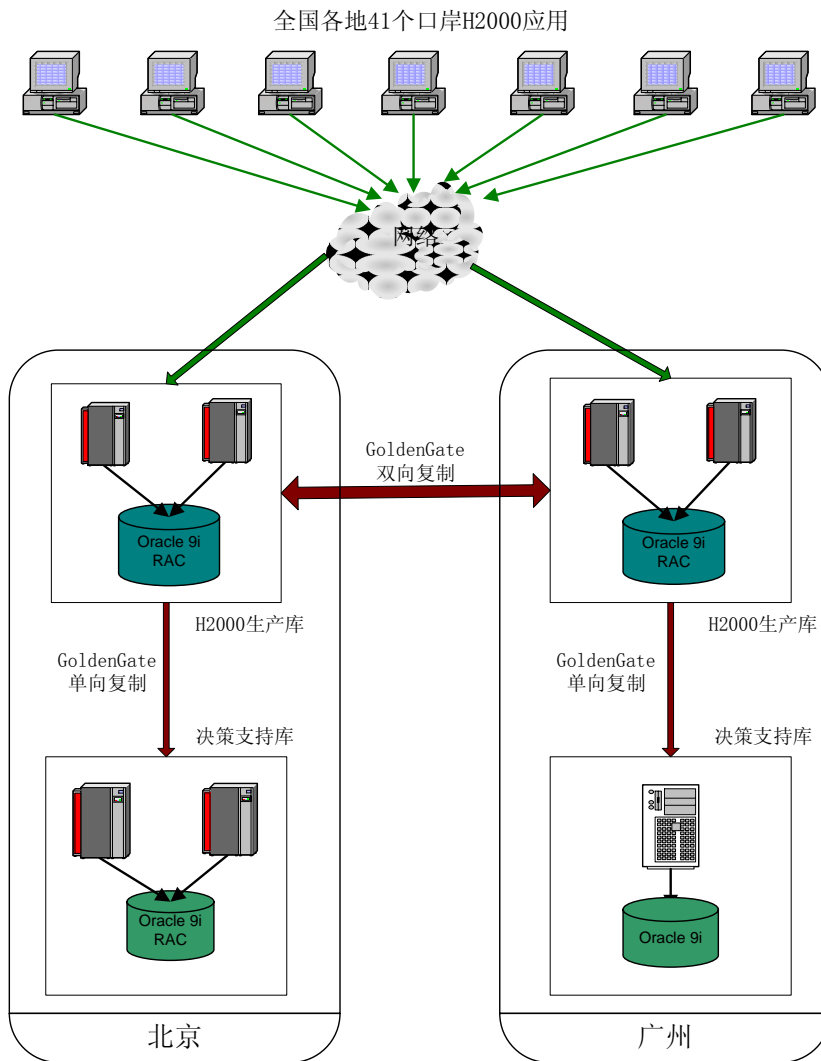
# 案例：北京移动



- 网管综合告警
- 同城异地
- 低带宽

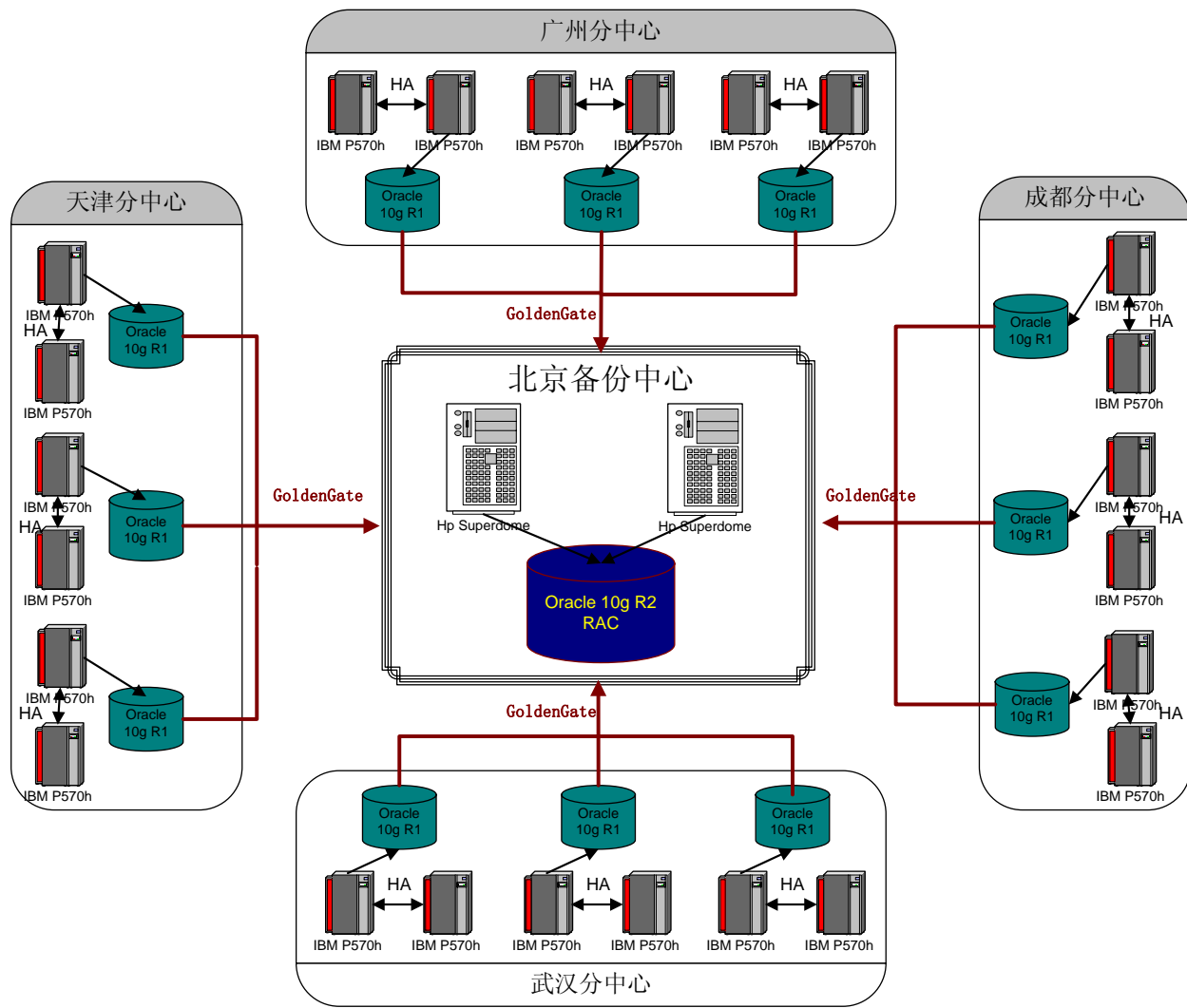


# 案例：中华人民共和国海关总署



- 远距离容灾
- 低带宽
- 低延时
- 可随时接管的双业务中心
- 快速业务切换

# 案例：国家体育总局体彩中心



- N+1模式容灾
- 各省市集中备份
- 远距离
- 低带宽

# Oracle最大可用性体系结构

最佳HA 技术的完整集合

- 每个都是可以水平扩展的, 完全激活的, 以数据为中心的

最高可用性和最低的成本

**Real Application Clusters  
& Clusterware**

容错  
服务器水平扩展

联机升级

联机升级  
硬件和软件

**GoldenGate**

双活数据复制

**Data Guard**

完全激活  
故障切换到复制点

**Automatic Storage  
Management**

容错  
存储水平扩展

**Flashback**

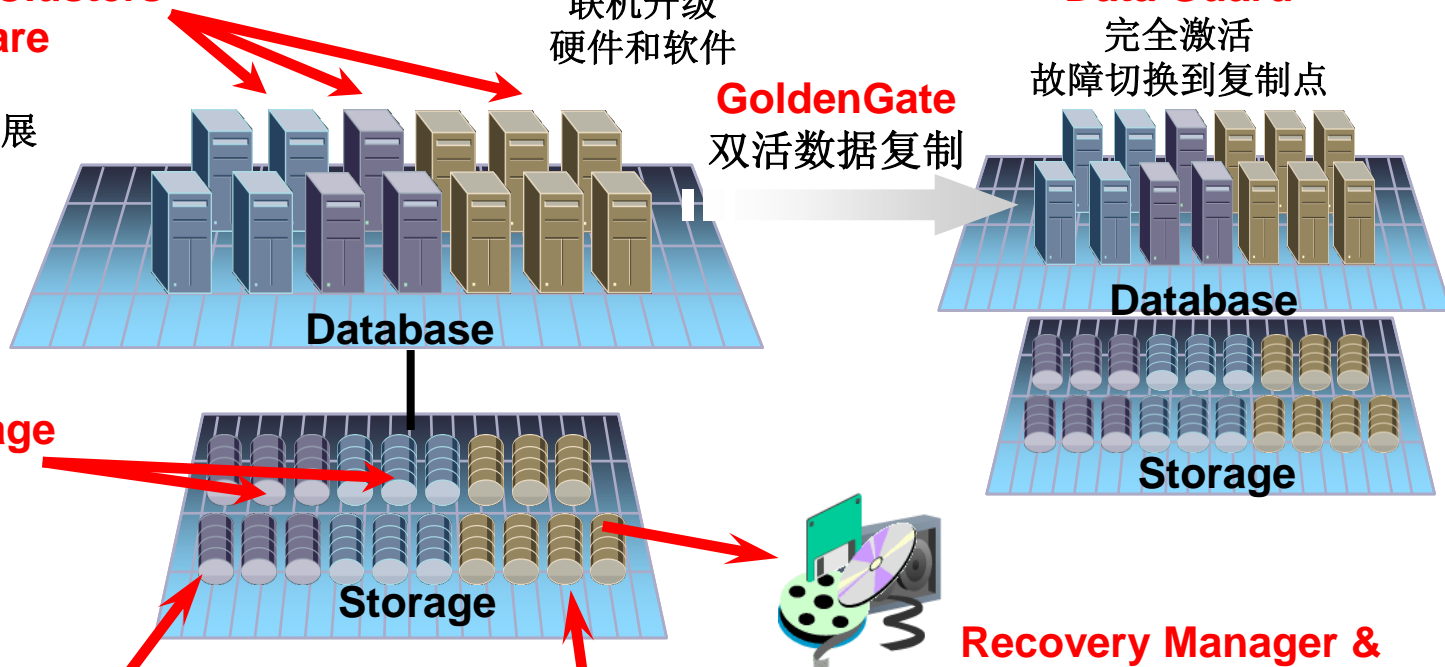
退回到指定时间点  
修正错误

联机重定义

联机重定义表

**Recovery Manager &  
Oracle Secure Backup**

低成本高性能  
数据保护 & 归档



ORACLE®

甲骨文

# 问答

ORACLE®

甲骨文