

ORACLE®

最大可用性架构 **Maximum Availability Architecture**

Ray Wang
资深技术咨询顾问



什么是**Maximum Availability Architecture**?



ORACLE

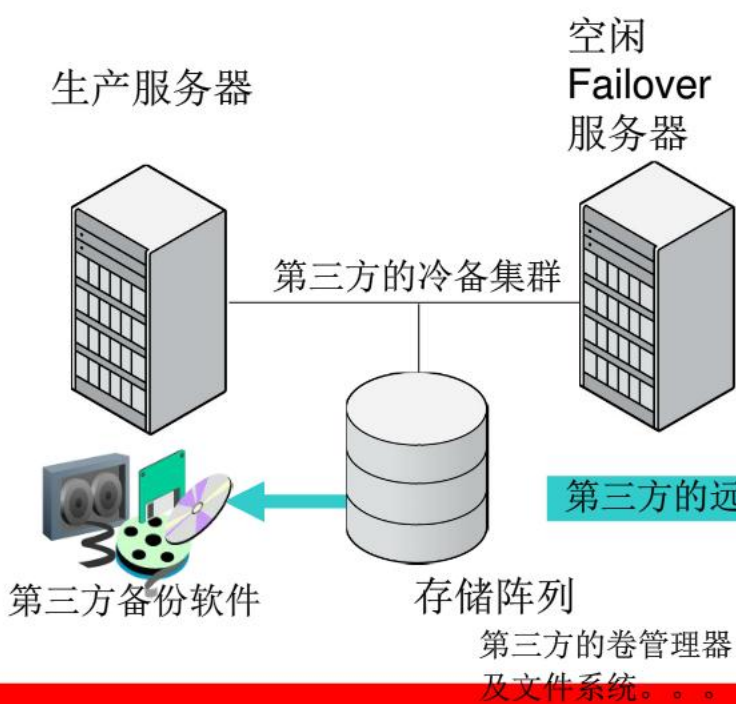
高可用业务问题

三个维度

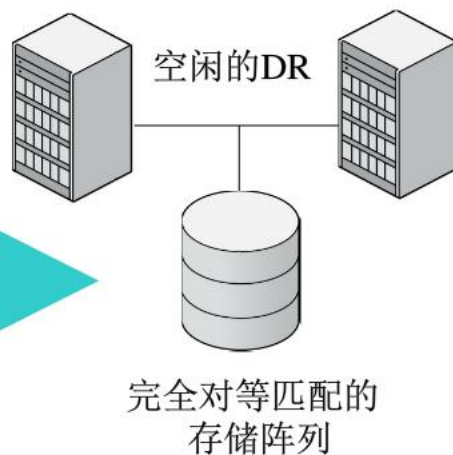
- 数据
 - 任何中断发生后你是否容忍数据丢失？（RPO）
- 时间
 - 任何中断发生后你的业务多长时间能恢复？（RTO）
- 成本
 - 是否所有系统组件都没有闲置？

所有这些都与数据中心运行降低风险相关...

传统的高可用方案



- 众多冗余的系统及存储
- 不可能知道容灾系统是否可用，直到你真的去尝试failover



当前的高可用要求

来自Oracle用户

需求	解决方案概要
中断保护	任何类型，任何地点
典型恢复时间	几秒到几分
测试频繁	无论何时
典型数据丢失	零丢失或者秒级
部署复杂	简单
需要集成方案	预集成
ROI	“告诉我投资回报!”



ORACLE

传统方案不能满足当前的客户需求

■ 高可用和容灾需求

- “我们需要跨地域的分布式多数据中心的部署，以最优的性能最快的速度响应全球的客户。”
- “我们需要升级数据库、操作系统还有硬件，但是我不想停应用，有办法吗？”
- “我们需要一个本地/远程的灾备节点，一旦灾难发生主系统停用了，立刻可以启用本地/远程的备用节点。”
- “我们需要构建数据仓库，要从很多不同的数据库抽取数据，而且希望数据越实时越好！”
- “我们的主系统在出报表的时候负荷太大了，能不能把这个任务剥离出来呢？”
- “现在数据量太大，备份数据时太耗资源，有没有好的解决方案”
- “当前业务程序变化较频繁，怎么能在不影响生产的情况下做一些基本的业务测试。”

MAA With Exadata

低成本、集成、全活动、高ROI

生产中心

RAC

- 可扩展性
- 服务器高可用

Flashback

- 人为错误修复

Online Redefinition,
Edition-based Redefinition,
Data Guard, GoldenGate

- 最小化维护、升级、迁移停机时间



ASM

- 卷管理

RMAN & Fast Recovery Area

- 磁盘备份



活动灾备中心

Active Data Guard

- 数据保护、容灾
- 查询卸载

GoldenGate

- Active-active 双活
- 异构



Oracle Secure Backup

- 备份到磁带或云

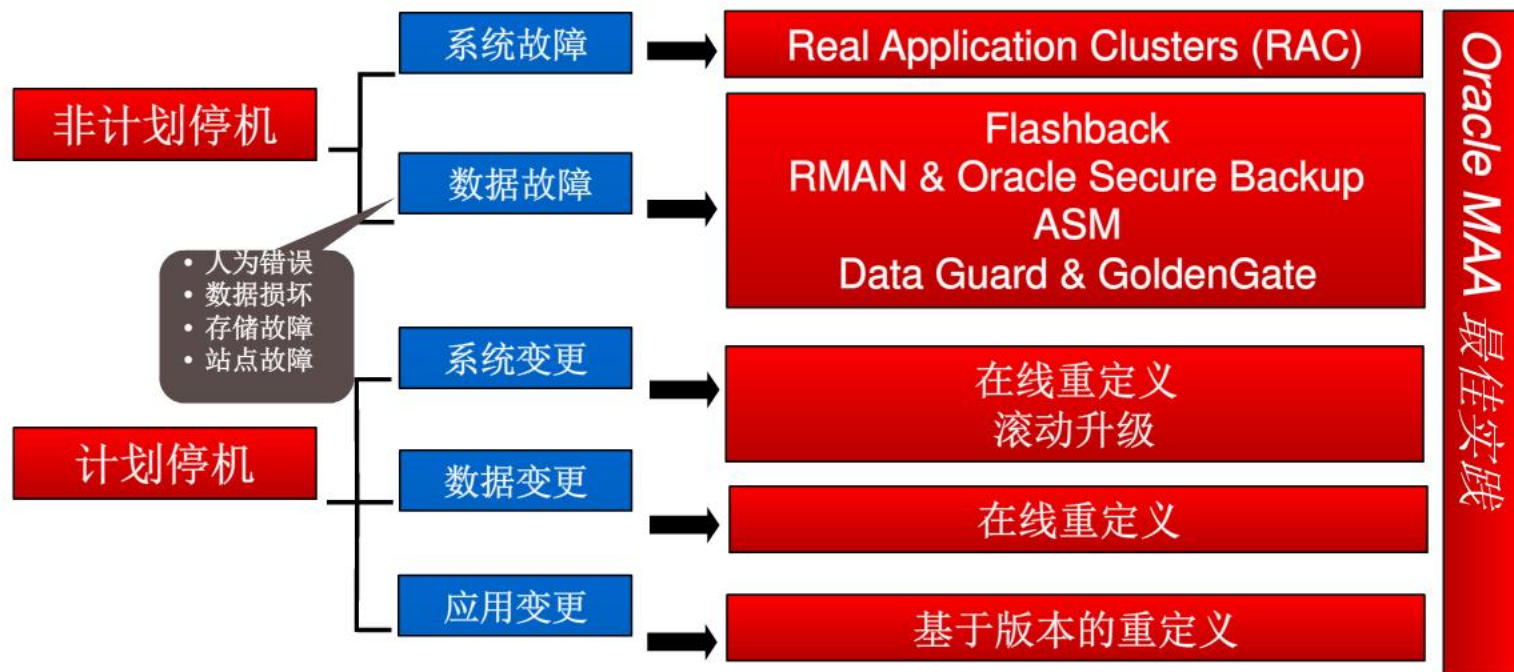
ORACLE

Oracle 数据库高可用性



ORACLE

Oracle数据库容灾高可用全面解决方案



ORACLE

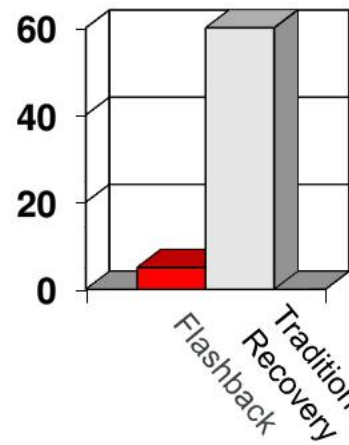
Flashback 闪回技术

快速、细粒度错误恢复

- Flashback彻底改变错误恢复手段：
 - 查看发生错误前某个点的正确数据
- 完成数据回退的时间和工作量取决于从错误发生起的数据库负载，和数据库大小无关！

$$\text{Correction Time} = \text{Error Time} + \cancel{f(\text{DB_SIZE})}$$

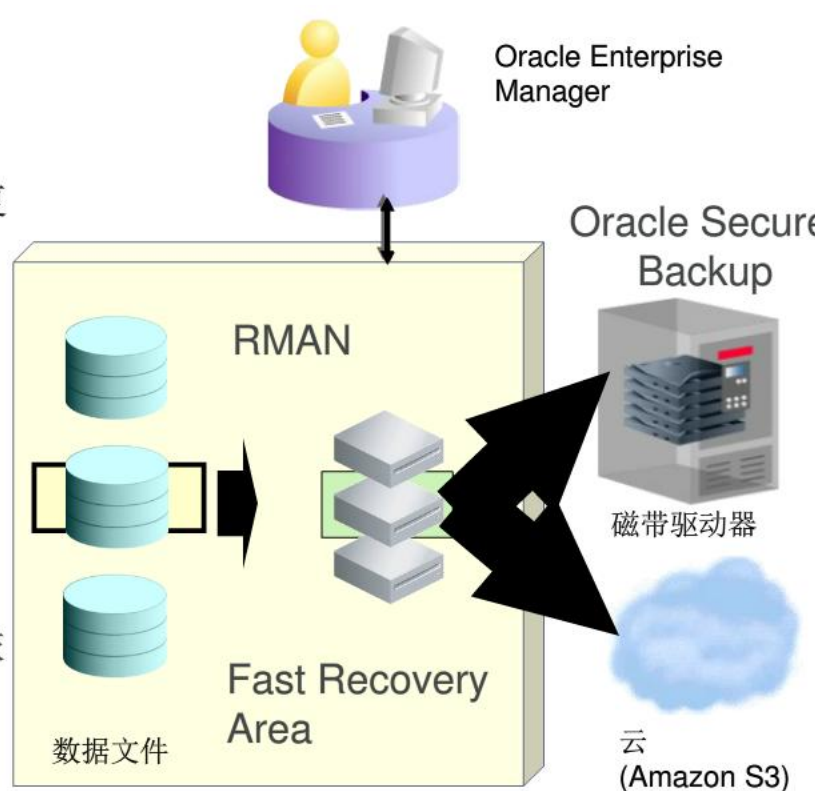
- 简单: **SQL> flashback database to <timestamp>;**
- 灵活: Flashback Query, Table, Transaction, Database, Drop



ORACLE

备份和恢复

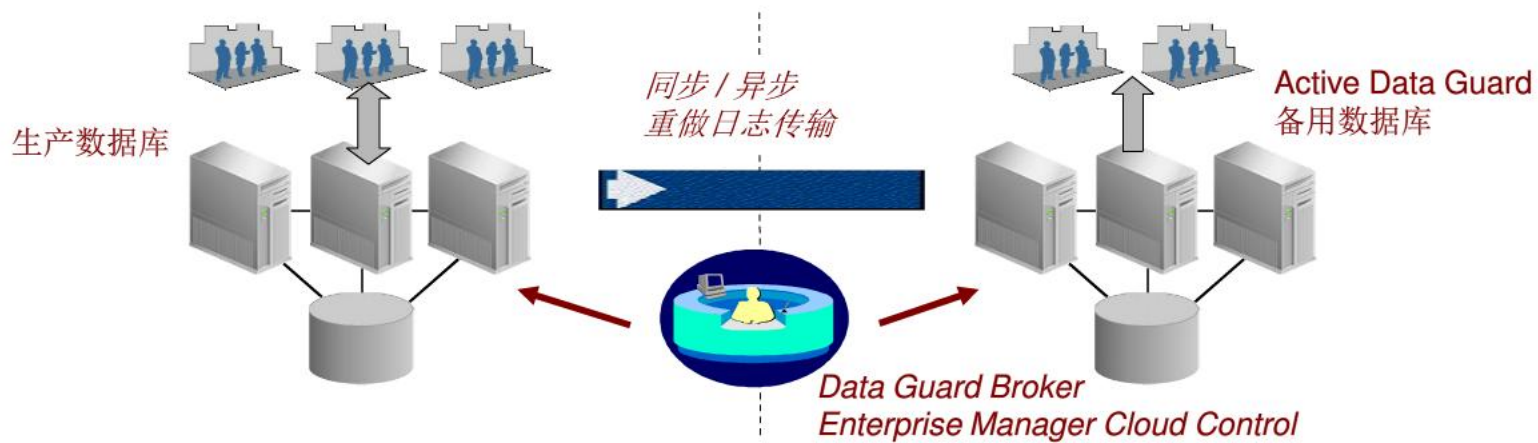
- Recovery Manager (RMAN): Oracle备份恢复的基础架构
- RMAN理解数据库文件格式和最佳恢复过程, 支持:
 - 块校验
 - 在线块级修复
 - 未使用块去重
 - 在线、多通道备份
 - 本地加密和压缩
- 集成的磁盘备份: Fast Recovery Area快速恢复区
- 集成的磁带和云备份: Oracle Secure Backup (OSB)



ORACLE

Oracle Data Guard

最佳数据容灾

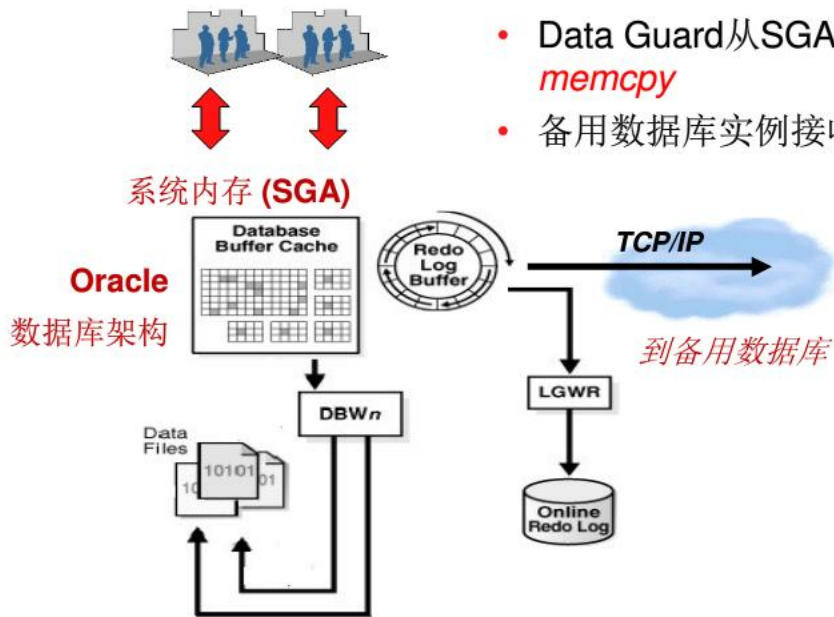


- Oracle数据库高可用和数据保护
- 单一配置支持30套备用数据库
- 备用数据库可以用于查询、报表、测试或者备份

ORACLE

数据库内置的复制

Data Guard: Why A Big Deal

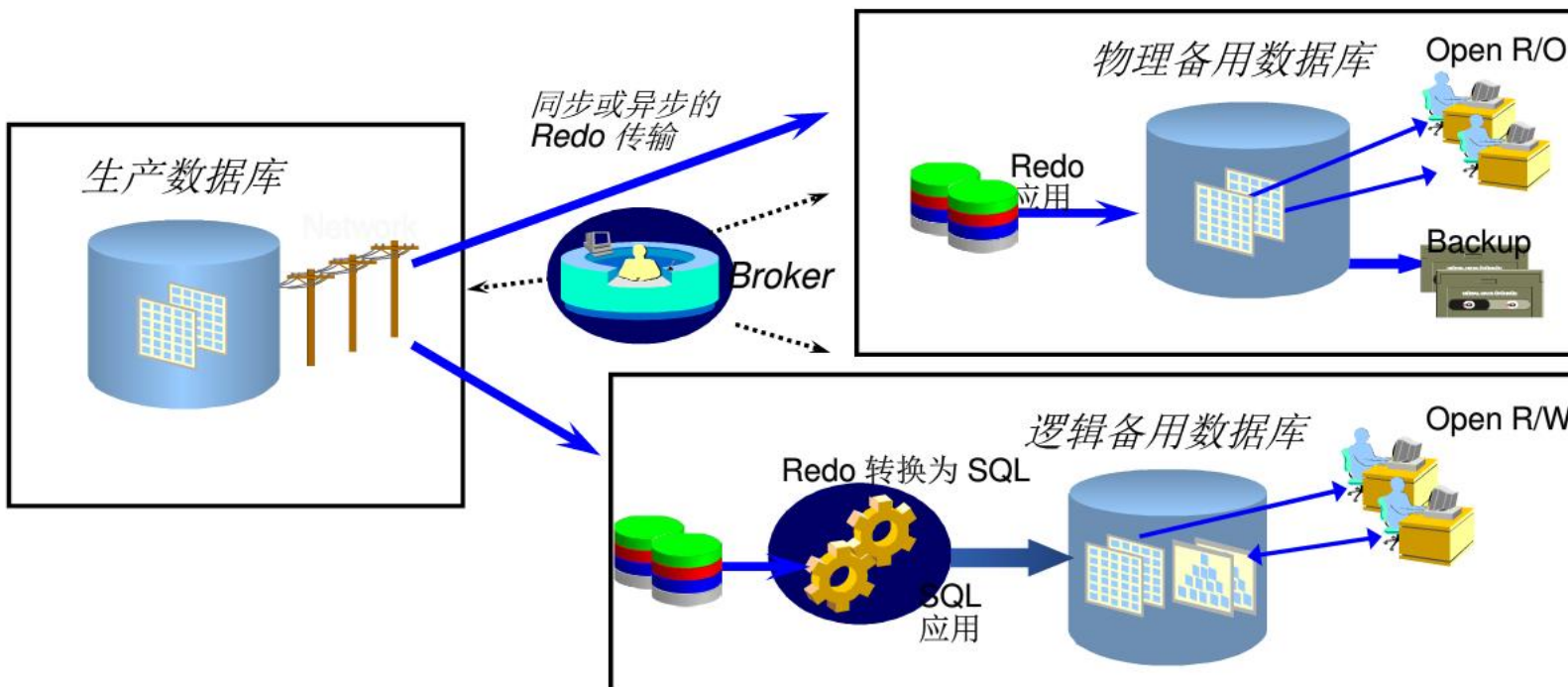


- Data Guard从SGA直接传输redo块: 像网络上的内存拷贝!
memcpy
- 备用数据库实例接收和应用Redo: 持续不断的数据校验

1. 更好的性能, 没有磁盘I/O
2. 对底层错误更好的隔离
3. 更好的网络利用率: 仅传输redo
4. 事物一致性: 总是
5. 坏块自动修复
6. 数据库内置应用failover

ORACLE

Oracle Data Guard 11g 架构



ORACLE

重做传输 vs. 保护类型

保护模式控制响应失败事件

模式	数据丢失风险	传输	是否没有来自备用数据库的确认
最大保护	零数据丢失 双重失败保护	同步	延迟主数据库，直到确认收到副本
最大可用	零数据丢失 单失败保护	同步	延迟主数据库，直到确认收到副本或者发生超时—然后继续处理
最大性能	有少量数据丢失的可能	异步	主数据库不会等待备用数据库的确认

LOG_ARCHIVE_DEST_n中的 NET_TIMEOUT 参数

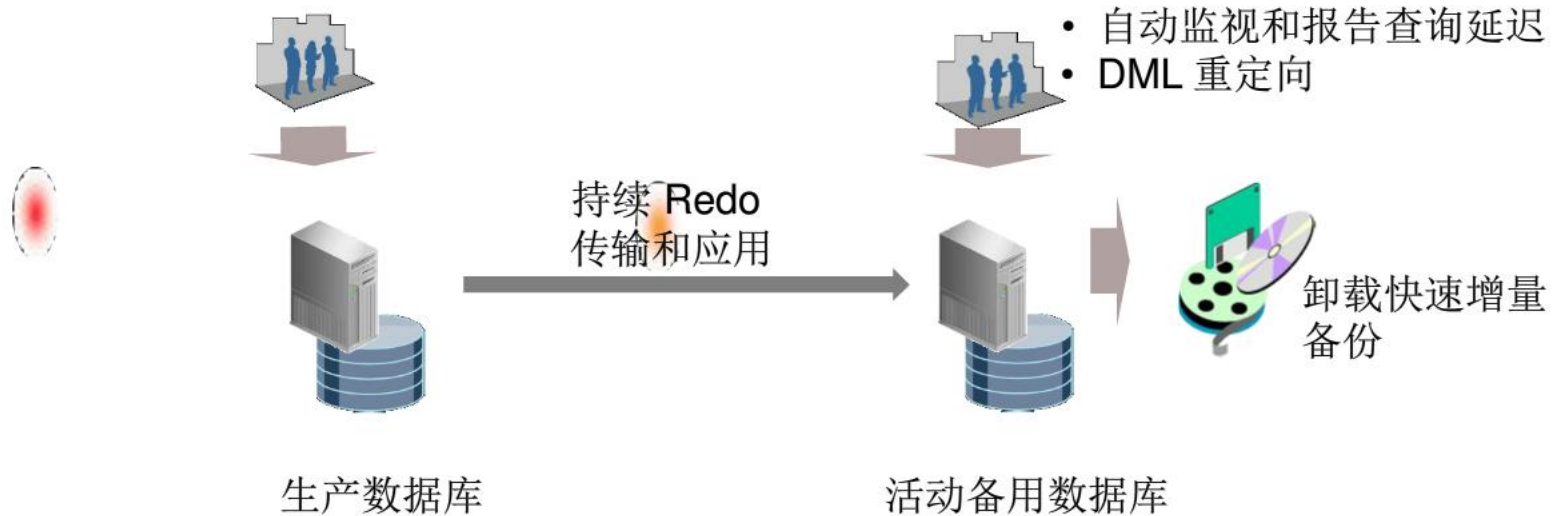
Data Guard 11g 缺省 = 30 seconds

Data Guard 10g 缺省 = 180 seconds

ORACLE

Active Data Guard

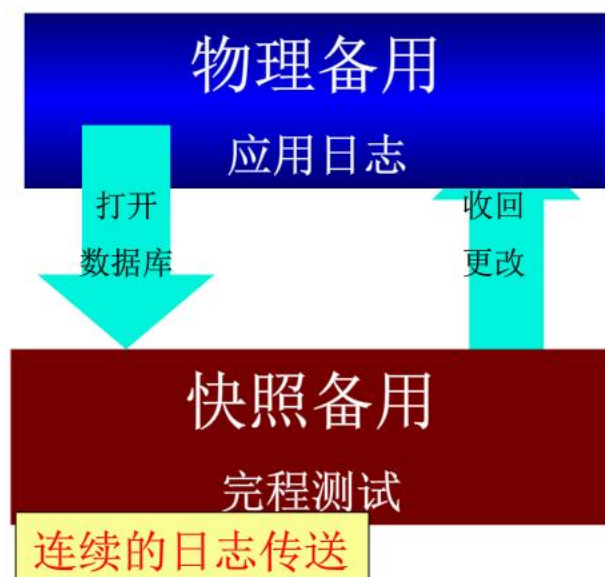
从灾备投资获得回报



ORACLE

ADG高可用方案

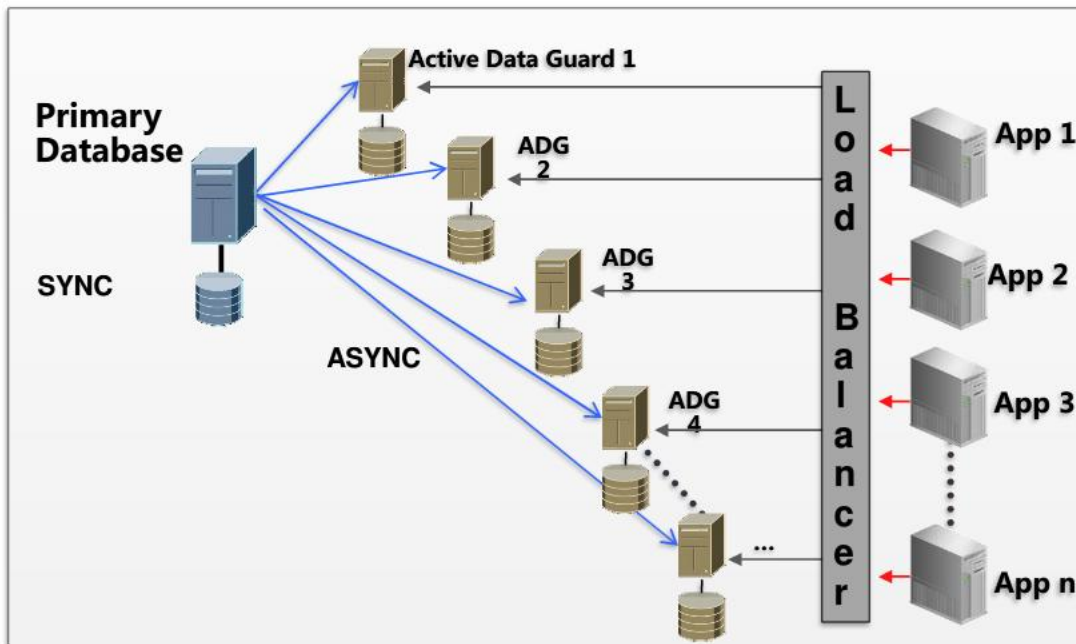
快照备用 – 利用备用数据库测试



- 转换物理备用到快照备用，并且打开数据库使测试应用程序可写
 - `ALTER DATABASE CONVERT TO SNAPSHOT STANDBY;`
- 丢弃测试改动并且通过应用日志保持与主数据库同步
 - `ALTER DATABASE CONVERT TO PHYSICAL STANDBY;`
- 保持零数据丢失
 - 但是不能实时查询或快速故障切换
- 避免资源的空闲等待时间与存储快照类似, 但是:
 - 同时提供了 DR 解决方案
 - 使用存储的单个副本

ORACLE

Active Data Guard: Reader Farm Scale-Out



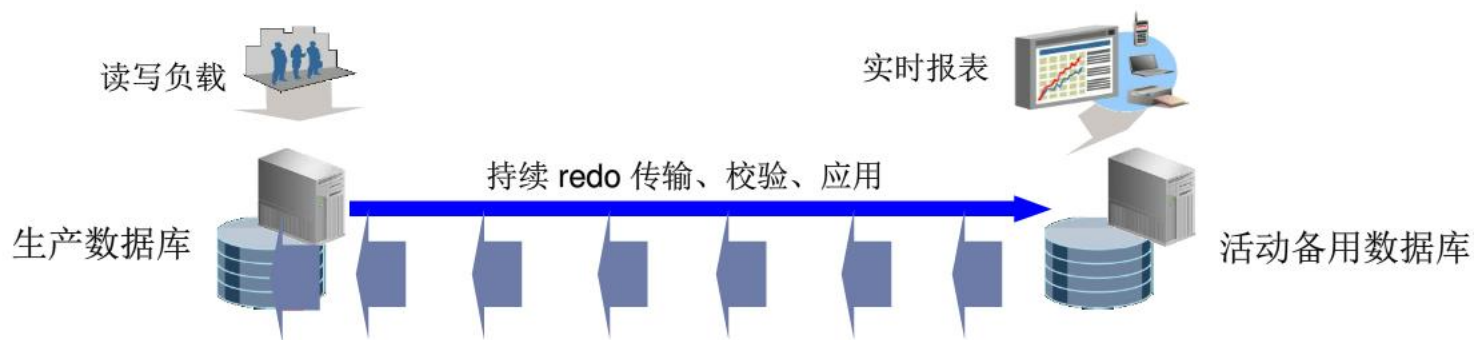
- 最多30个备用数据库
- 为“查询-下单”提供扩展

Active Data Guard: 自动坏块修复

在线坏块修复保证高可用

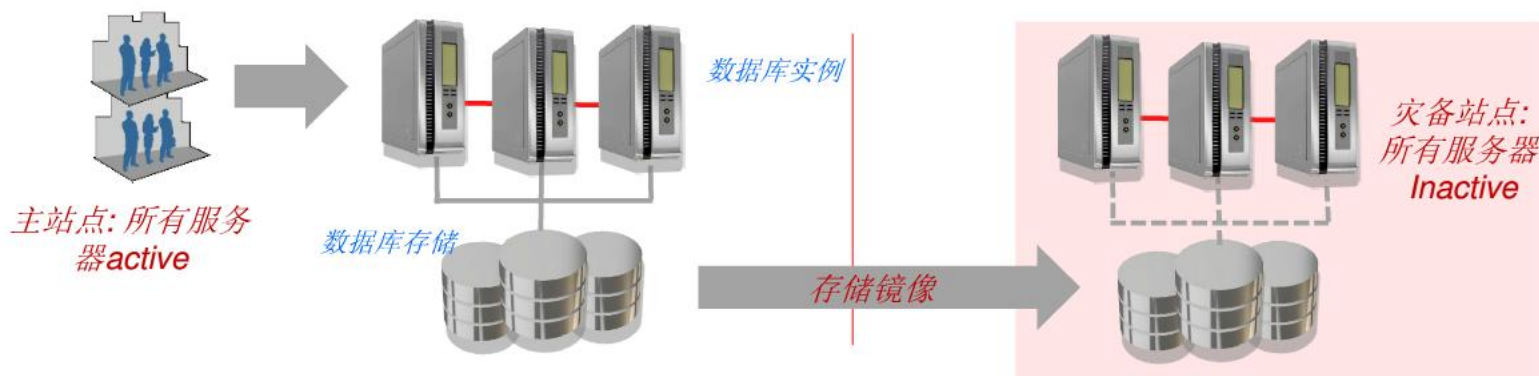
■ 自动数据块修复

- 当Oracle探测到生产数据库出现坏块，它会从活动的备用数据库拷贝一个好的数据块并自动修复，反之亦然
- 对用户和应用透明



ORACLE

通过存储镜像容灾



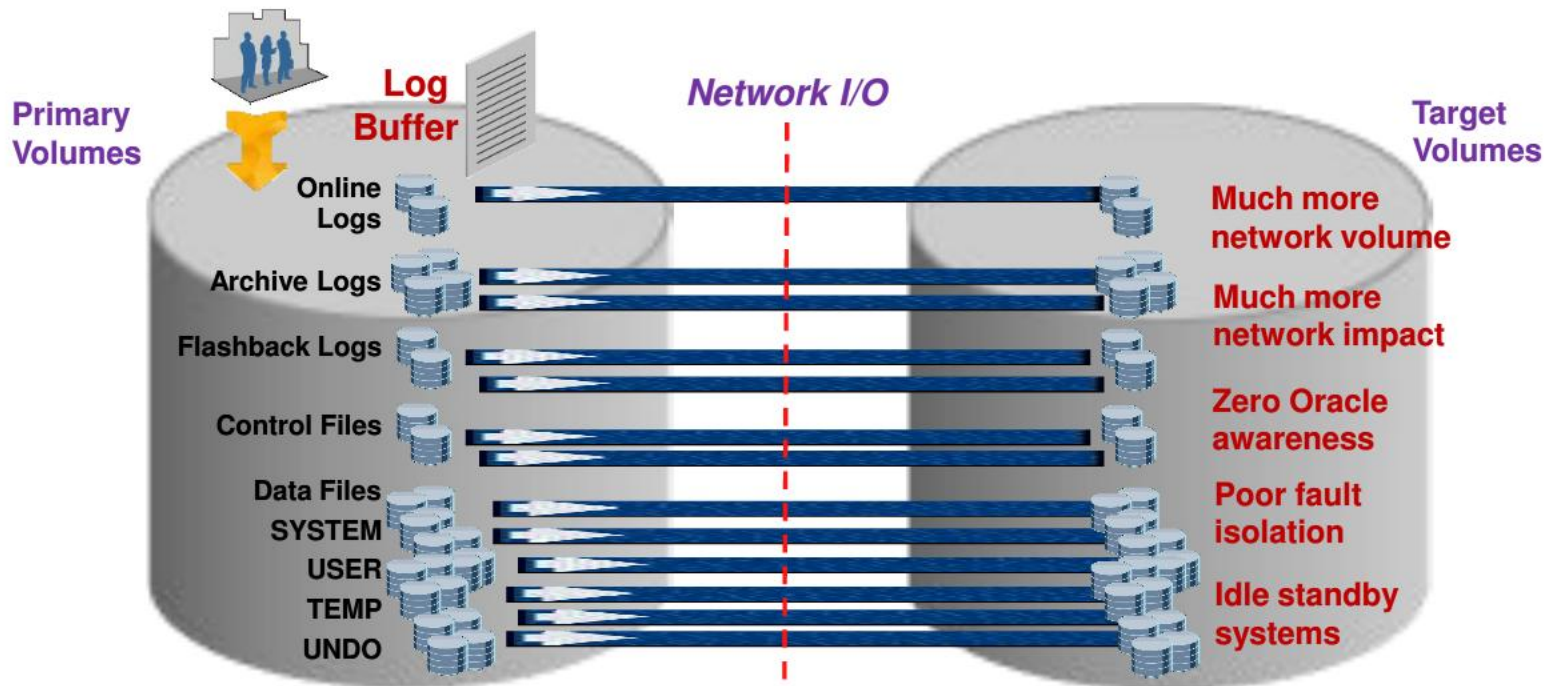
存储镜像: 冗余存储保护存储故障, 但是:

- 不能保护物理数据损坏
- 同步时灾备系统处于离线状态 – 没有实时数据校验
- 有限的距离, 锁定存储厂商, 手工failover, 没有滚动升级, 极高网络带宽使用

See: <http://www.oracle.com/technetwork/database/features/availability/dataguardremotemirroring-086151.htm>

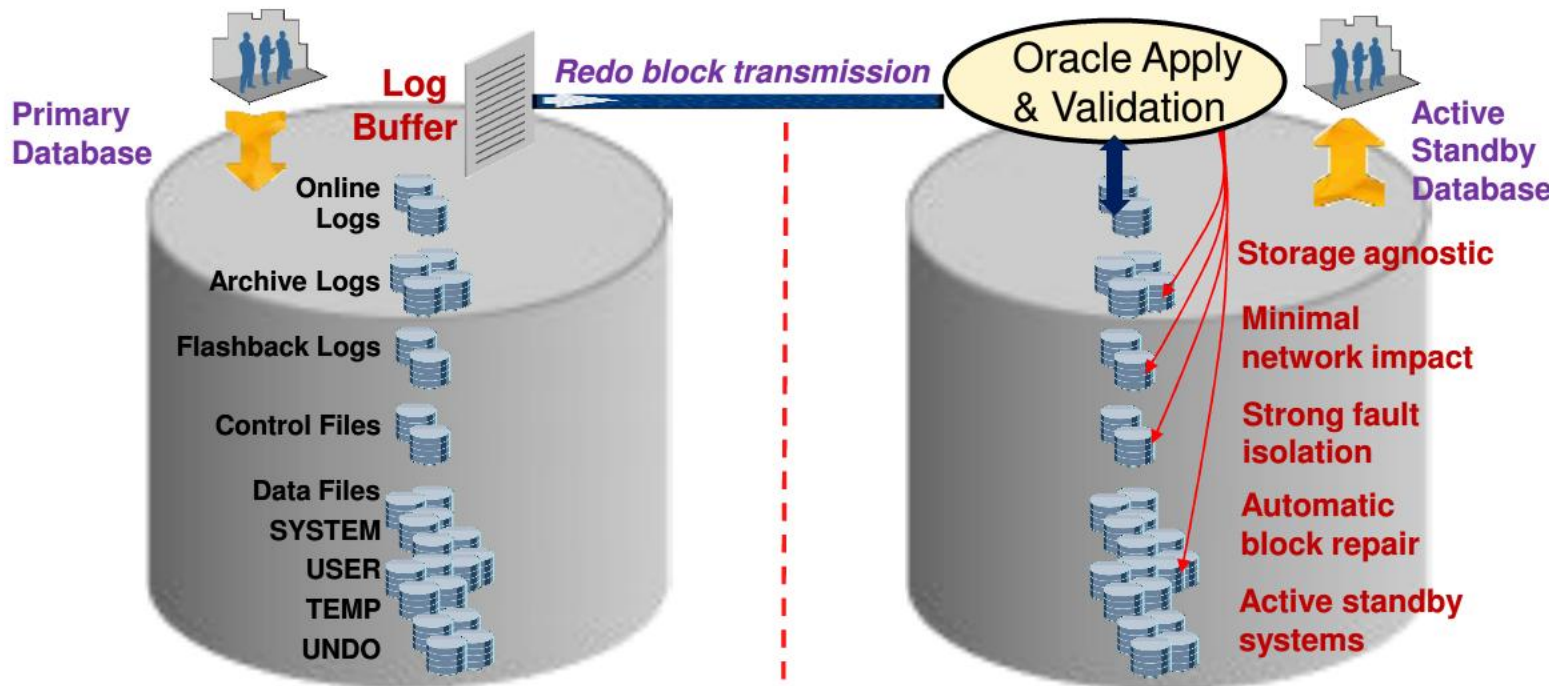
ORACLE

存储镜像缺点



ORACLE

Data Guard优势



ORACLE

Active DataGuard得到客户信赖的关键

实施维护简单	可以直接从主数据库生成备用系统上的备用数据库，不需要额外复杂的系统维护。	简单可靠
实时查询	备用数据库可以在日志应用的状态下以只读的方式打开,既支持同步模式也支持异步模式。	实时性好
支持DBEE特性	可以处理Oracle DB所有数据类型，支持Oracle DBEE的特性	兼容性好
数据保护	可以检查修复由于硬件存储错误所引起的数据文件讹误。故障切换可以马上进行	可信赖
增强的RMAN	减少大数据量数据库的备份资源压力	转移备份
快照模式	备用数据库的快照模式能够以读写的方式打开。	随时用于测试

ORACLE

Oracle GoldenGate



- 实时数据集成解决方案的行业领先者
- 和Oracle数据库的最佳集成融合
 - 异构基础架构持续的高可用性
 - 为报表、BI/EPM提供实时数据访问
 - 零停机迁移、升级Oracle数据库、硬件和应用

1200家以上客户，包括400家以上财富500强企业，涉及金融、通信、医疗、政府、零售、公用事业

- 全球最大的5家商业银行中的4家
- 全球最繁忙的3家ATM网络
- 全球最大的5家金融服务公司
- 全球最大的5家电信运营商
- 全球最大的5家食品和药品供应商中的3家

ORACLE

GoldenGate: 双活复制

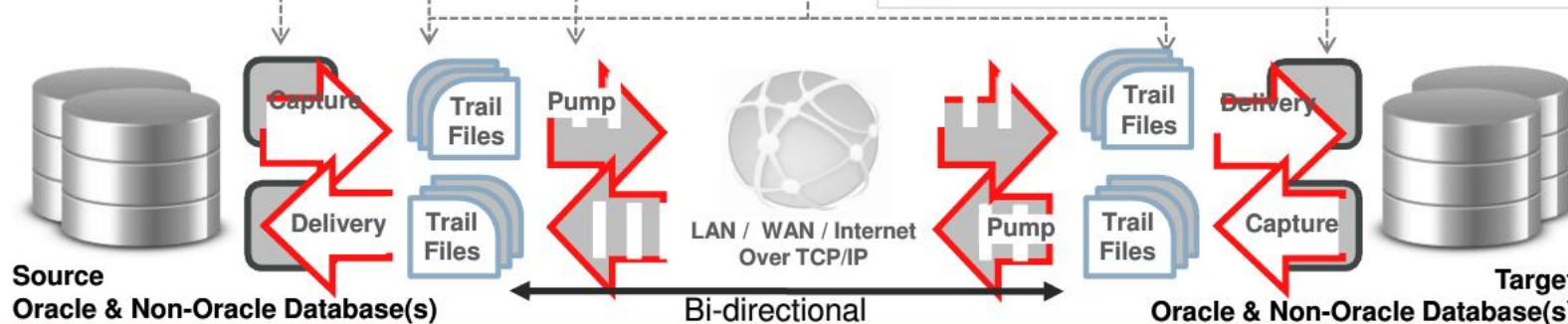
捕获: 通过读取重做日志捕获提交的交易 (可以筛选)

追踪: 存储和队列化数据

抽取: 分发数据

传输: 数据被压缩、加密并传输到目标服务器

投递: 按照交易一致性应用数据, 根据需要转换数据



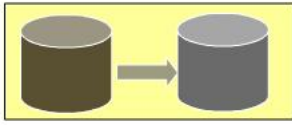
GoldenGate: 机构环境复制

数据库	操作系统和平台
Oracle GoldenGate Capture and Delivery: ▪ Oracle ▪ DB2 ▪ Microsoft SQL Server ▪ Sybase ASE ▪ Teradata ▪ Enscribe ▪ SQL/MP ▪ SQL/MX ▪ MySQL v 5.5 ▪ JMS message queues Oracle GoldenGate Delivery: ▪ All listed above, plus: ▪ TimesTen, Postgres ▪ Netezza, Greenplum, HP Neoview ▪ ETL products	Linux Oracle Solaris 11 Windows 2000, 2003, XP HP NonStop HP-UX IBM AIX v7.1 IBM z Series zLinux IBM i Series

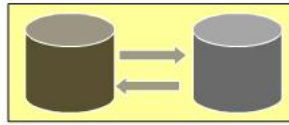
ORACLE

灵活的拓扑结构

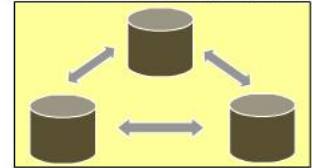
单向复制
报表



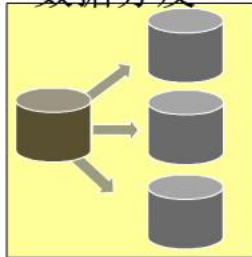
双向复制
立即接管



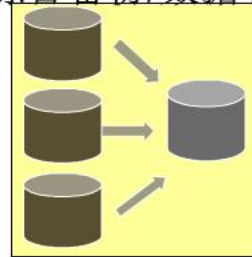
点对点
负载均衡, 高可用性及容灾



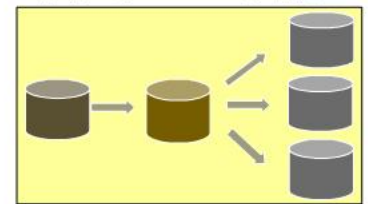
广播复制
数据分发



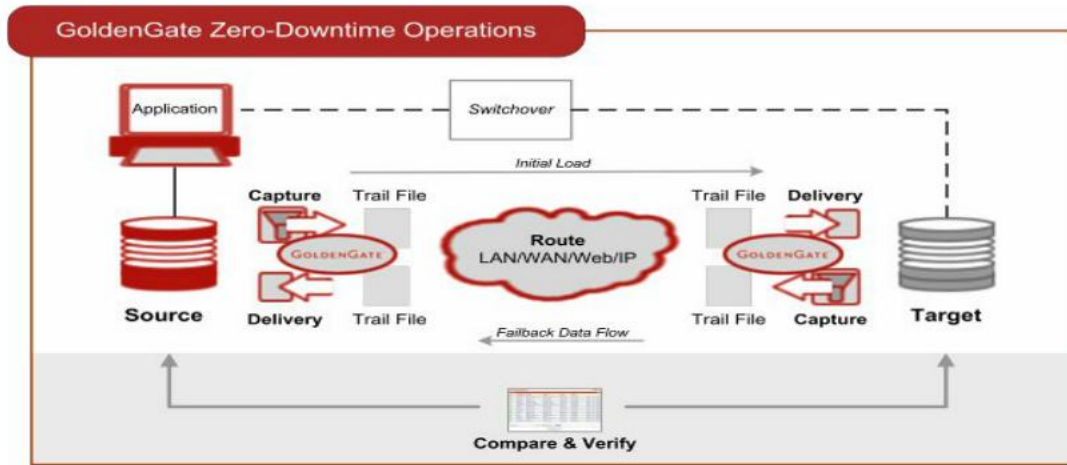
集中复制
综合备份/数据仓库



多级复制
层次化企业数据



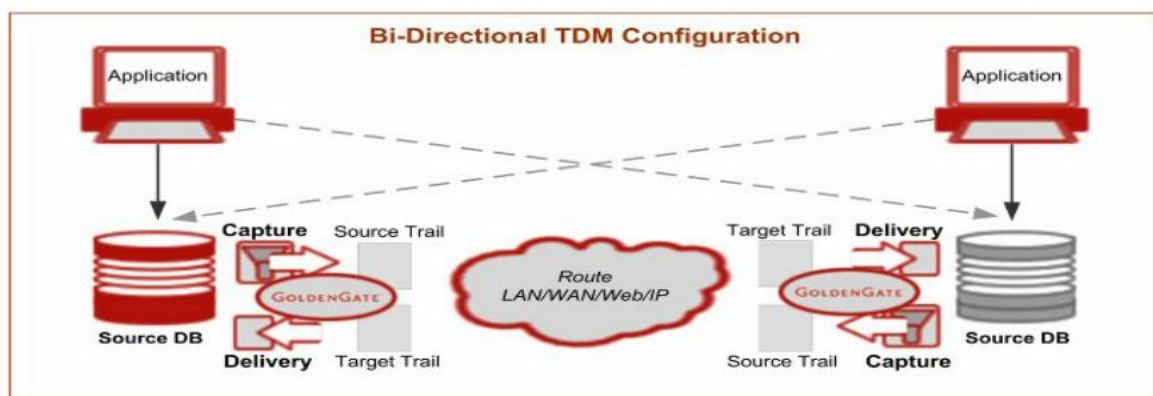
GoldenGate高可用性方案：零宕机操作



减少或者消除“计划停机时间”，包括平台，数据库或应用

- 迁移
- 升级
- 维护

GoldenGate高可用性方案：双活数据库



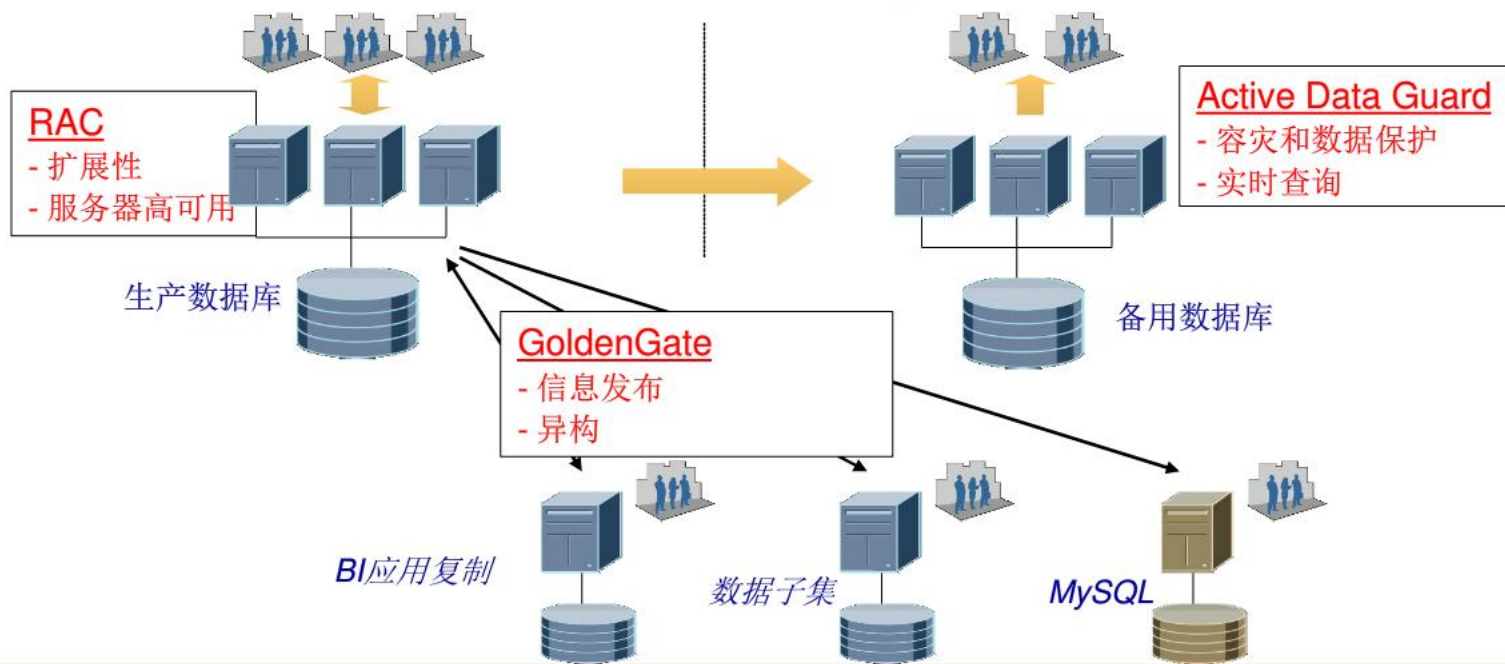
- 持续的可用性，单个节点中断不影响
- 事务的分布式处理，满足跨地域的部署要求，提高处理能力
- 性能的可伸缩性，可以增减节点
- 冲突检测 and 解决，必须的保证

GoldenGate的得到客户信赖的的关键

技术团队	GoldenGate作为Oracle的战略产品，从研发到售前售后已经全部融入了数据库的大团队	技术有保证 售后实施可靠
支持异构性	能够跨平台、跨品牌、跨版本支持各种主流数据库	扩展性好
DML语句性能	我们在联机事务处理OLTP系统上有着非常好的性能	高性能
数据一致性	提交再复制的机制，确保了源和目标的数据一致性	可信赖
带宽占用小	由于先进的抓取和压缩技术，占用带宽最小	适应性强
对主机占用少	得益于核心程序的高效，对主机CPU和内存的资源要求小	生产系统低占用

Active Dataguard + GoldenGate强强联合

提供整合的异构数据分发，高可用,灾备及双活复制



ORACLE

Active DataGuard与GoldenGate比较分析

Active Data Guard	GoldenGate
实现方式简单，全库数据保护	较多的数据层维护工作量
支持Oracle数据库所有特性	数据库支持有一定限制要求
实时性好，可以零数据丢失	实时性好
故障隔离，防止数据块损坏	
Oracle 11g及以上版本支持	主机无关，存储无关，数据库平台相关性弱
备用库在应用日志同时处于read only状态	支持双活，多数据中心，复制拓扑结构灵活
网络要求较低	网络要求很低
比较适合同城	适合异构环境，同城、异地

ORACLE

MAA In Exadata

ORACLE

MAA With Exadata

完整、集成的高可用

Real Application Clusters

ASM



RMAN / Fast
Recovery Area



Oracle Secure
Backup

Active Data Guard /
GoldenGate

WAN



- 全面的故障保护: 服务器、存储、网络、站点、数据块
- 人为错误保护: 数据库、表、行、交易
- 活动备用库: 通过打开备用库卸载查询负载、实现实时查询
- 在线索引和表重定义
- 在线打补丁和升级

ORACLE

Exadata HA 度量值



- 备份和恢复、闪回

- Exadata的备份表现出极限性能和可扩展性
 - 磁带备份测试每小时 7TB
 - 测试性能受限于测试环境只使用了14个磁带驱动器
- 按照10%变化率做增量备份到磁带： 10-50 TB/hr
- 全库备份到磁盘： 17+ TB/hr
- 按照10%变化率做增量备份到磁盘： 20-50 TB/hr
- 使用闪回，可以达到3 TB/hour 加载速度和1.7 TB/hour redo应用速度

- Data Guard

- 使用非压缩的全库容灾每小时加载2TB以上
- 压缩表减少redo和网络流量进一步增强Data Guard
- REDO应用： 批量处理:每秒615 MB、 OLTP每秒200MB

ORACLE

MAA 用户成功案例



ORACLE

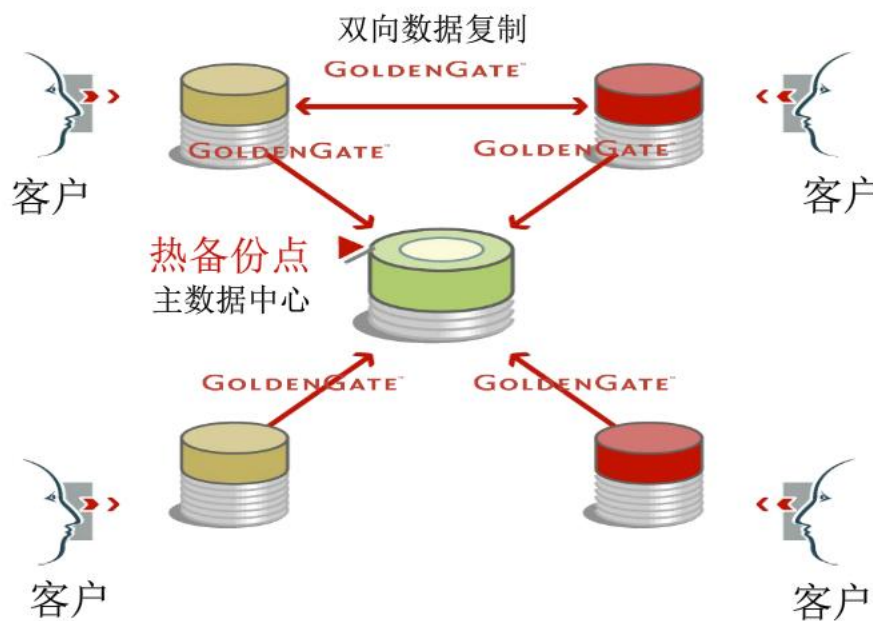
Bank of America美洲银行 层次化的HA实施



ORACLE

- **1000套Oracle数据库**
- **Oracle 数据库企业版, Enterprise Manager Grid Control, E-Business Suite, Oracle CRM, Hyperion, Fusion Middleware**
- **Oracle Database 11g Release 2**
 - **2012年7月1日核心数据库升级到最新版本**
 - **11gR2 升级不断进行, 10%的数据库已经升级**

- **Tiers 0 and 1**
 - 全面的**MAA**配置
- **Tiers 2 到 6**
 - 单实例**MAA** 配置 (非**RAC**)
- 所有层保持业务连续性
 - **99%**的生产数据库实施了**Data Guard**
 - **GoldenGate**被用于多目标的配置
 - 比如：一个内部的关键应用在洲际之间配置**3**向的复制



GoldenGate协助美国银行适应成长。

-Michele Schwappach,
Vice President and Senior Systems Manager

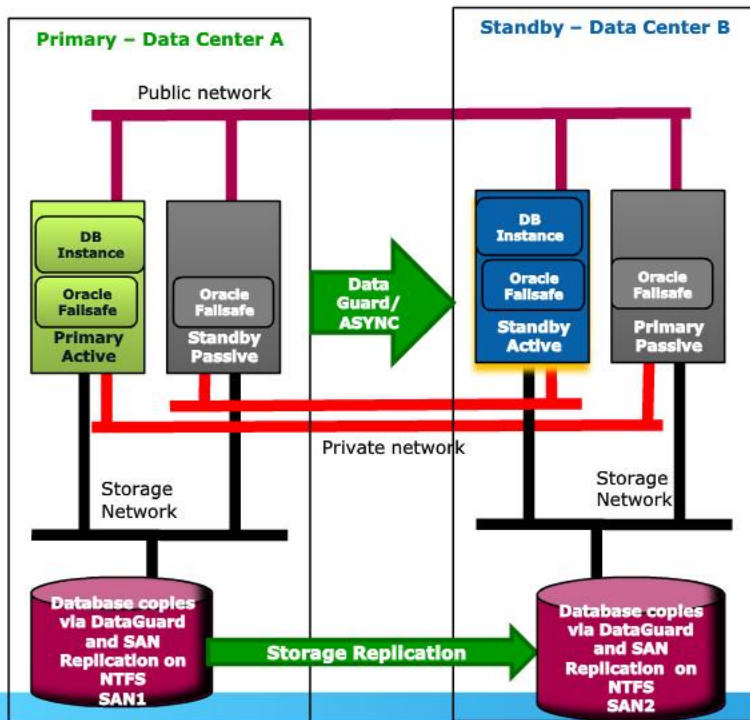
- 超过五千八百家分行及一万六千多部ATM
- 热备份数据中心从贯穿北美的四台主机(HP NSK)抽取数据
- 保障ATM 7X24小时可用

Intel HA 结合 DR

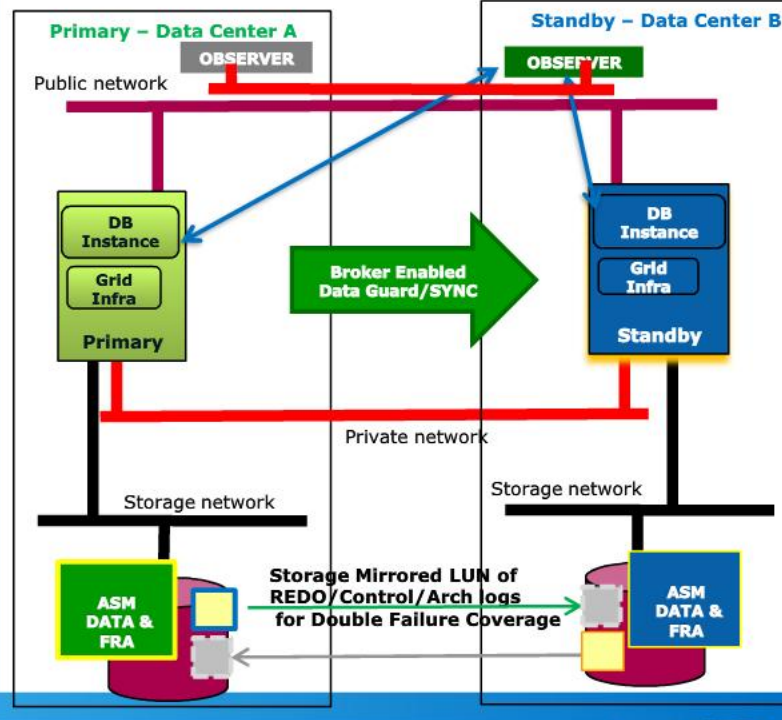


ORACLE

HA 1.0 DB 架构



HA 2.0 DB 架构



Comparison of Current HA 1.0 vs HA 2.0 DB HA/DR architecture illustrates simplicity of HA 2.0 stack and a thin database infrastructure footprint.



HA/DR Metrics with HA 2.0

Outages		SAN Replication /MSCS/ OFS/DG HA 1.0	OBSERVED DataGuard/Broker -FSFO HA 2.0
Unplanned	MTTR on Storage Failure	>1 hour	74 seconds
	MTTR on Site Failure	>1 hour	39 seconds
	MTTR on Node Failure	70 – 90 seconds	59 seconds
	MTTR on Instance Failure	70 – 90 seconds	50 seconds
Planned	DT needed for applying Oracle Patch Set Update	2+ hours	120 seconds

HA 2.0 allows lights-out role transition for DB/ASM/NODE/Storage/Network/Data Center failovers. Typical Failover time is around 60 seconds (includes 30 secs FSFO threshold) for unplanned outages. Observed failover time as low as 17 seconds (that includes 15 second FSFO threshold for that FMEA test) Graceful switchovers (for planned outage) take 120 seconds .

Paychex

在放弃存储镜像之后**MAA**带来的节省

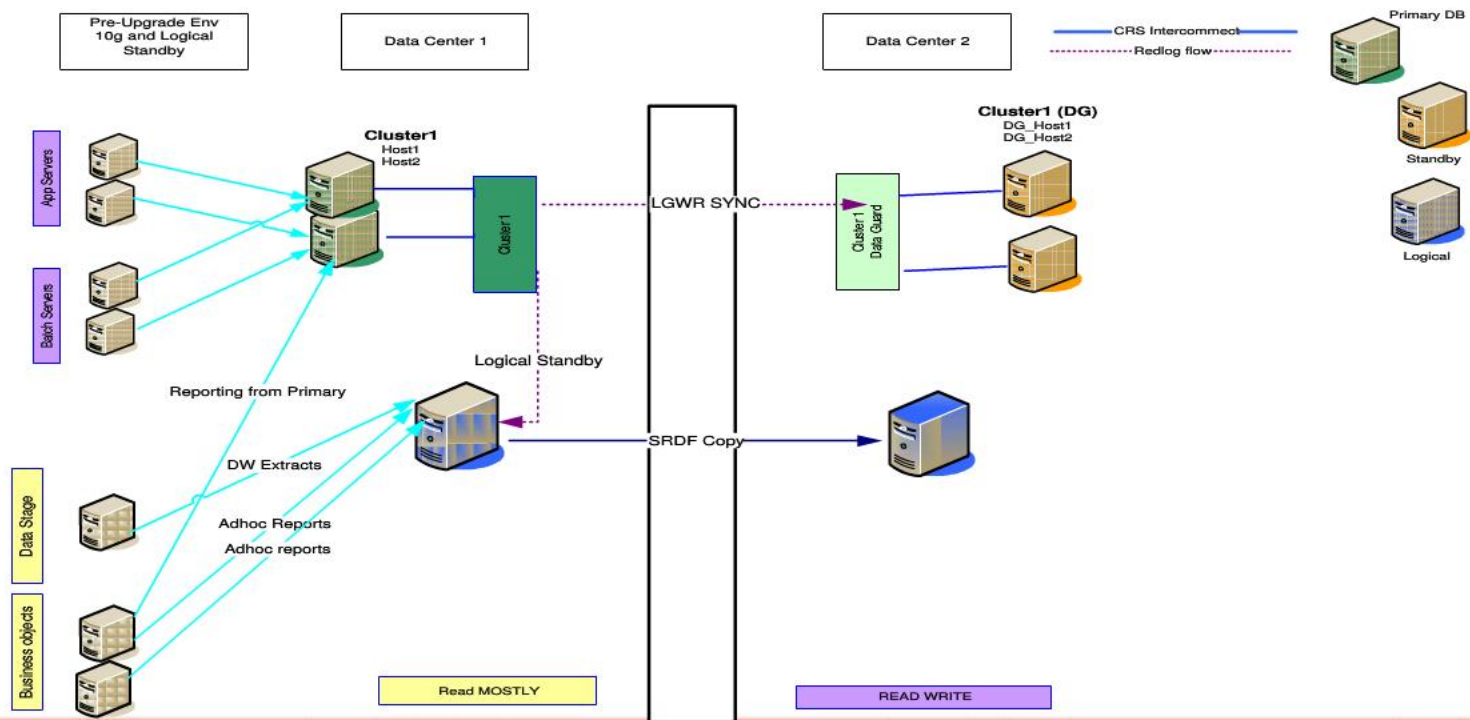


ORACLE

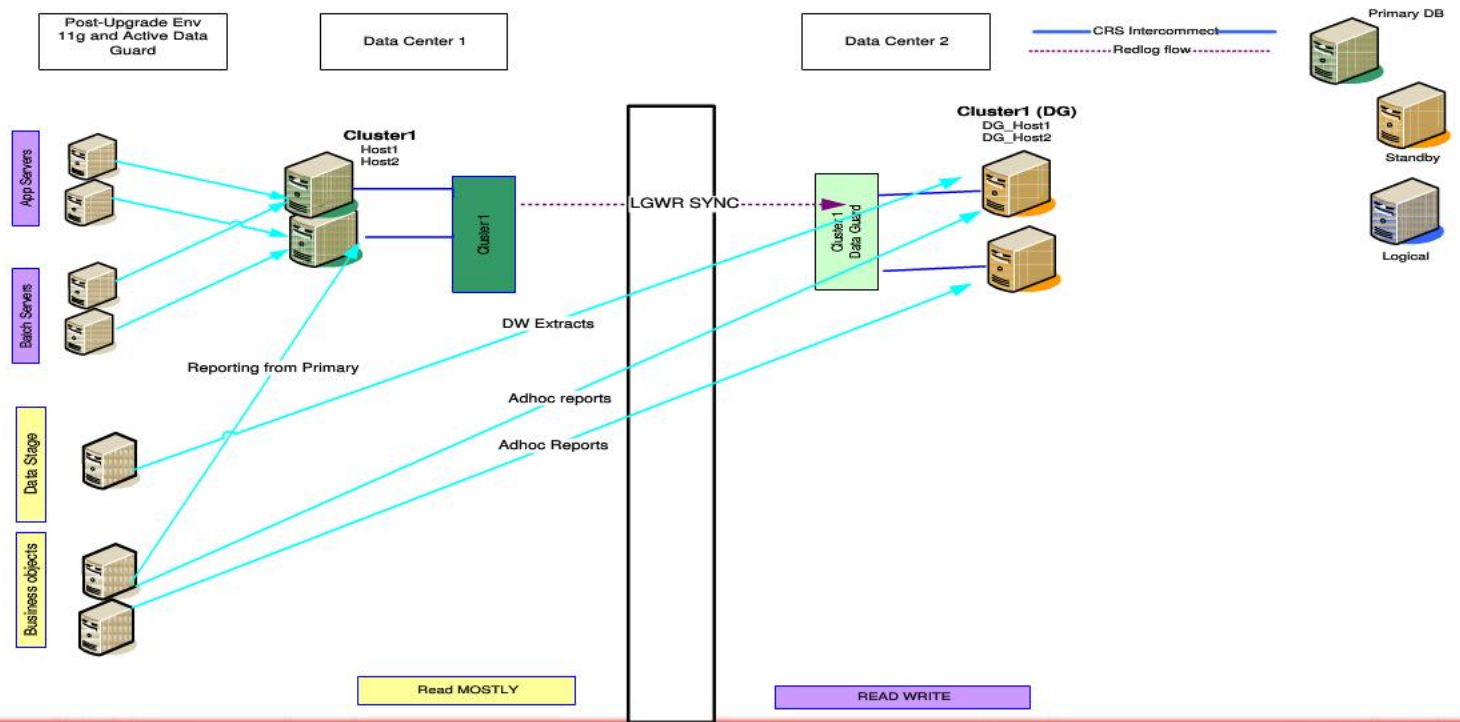


- Paychex Inc,是一家为564000家中小企业提供工资发放、人力资源、福利外包解决方案的行业领先供应商。
- 2011财年亮点 (截止May 31, 2011):
 - 20亿营收;
 - 8亿税前收入;
 - 5亿净利润
- 计算机世界 list of “Top 100 Best Places to Work in IT”.

升级11g前的架构



升级11g后的架构



结果

- 回收90TB存储!
- 更大的收获...回收光纤适配器端口用于平衡SAN负载
- 最大的收获...减少了SRDF 硬件复制的流量
- 逻辑备用需要大量维护...没有了! Logical



British Telecom英国电信 最小化计划停机时间



ORACLE

使用GoldenGate 无停机升级Siebel



业务挑战

1. 非常关键的Siebel CRM 呼叫中心应用
2. 每天15,000 呼叫中心代理, 25TB以上数据库
3. 在保证可回退方案的情况下无停机升级

解决方案

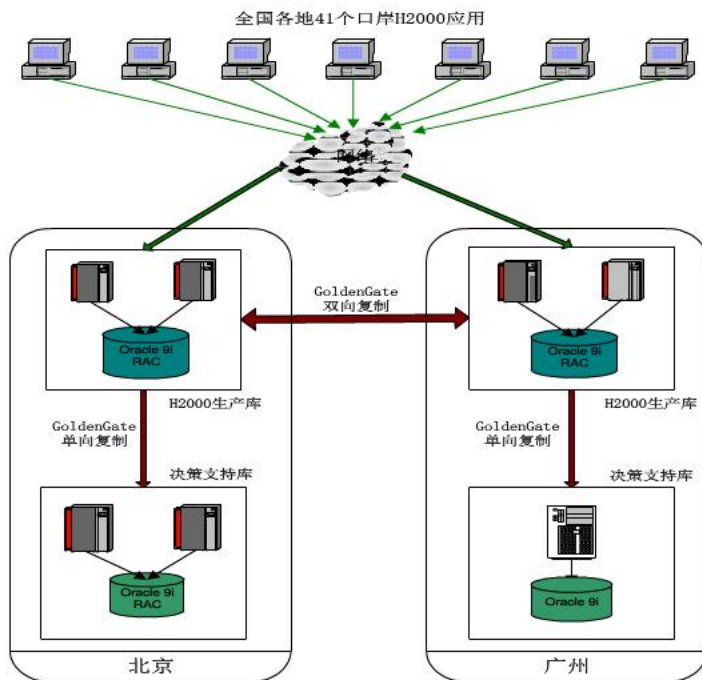
- GoldenGate 配置在Siebel 7.8 和 Siebel 8.0之间

投资回报

- 零数据库停机
- 减少用户停机
- 因为有之前的Siebel 环境可以回退, 所以降低了风险

ORACLE

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

Exadata实施MAA



ORACLE

Tier-1 Oracle Database HA at PayPal 贝宝

Primary Data Center

关键的支付处理数据库
每秒处理10万SQL

生产数据库
• RAC, ASM, FRA
• 10-30 TB

Data Guard
Redo 传输

异步

Active Data Guard Standby

- 卸载查询 – 余额、交易历史、风险分析、诈骗探测、行业特定分析等
- 自动块修复

GoldenGate 实时数据集成 *

ETL
Targets

WAN, 650+ miles

Data Guard 异步 Redo 传送

DR Data Center

Active Data Guard Standby

- 卸载查询和读

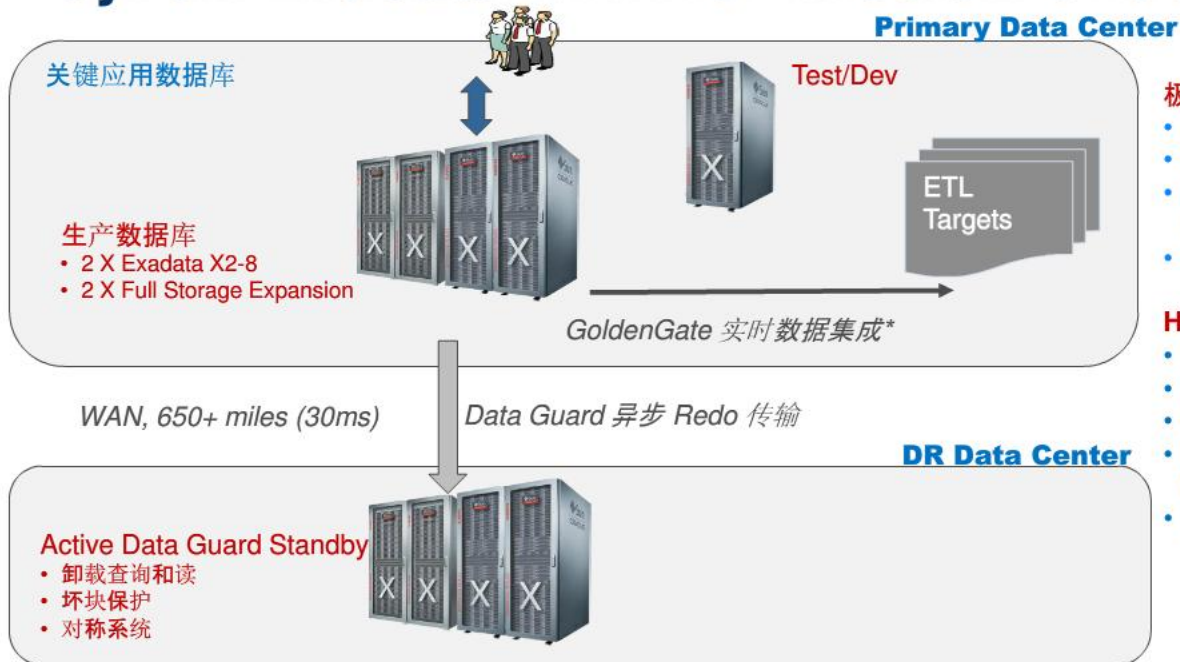
Data Guard
级联 Redo 传输

Physical Standby

- 支持 DR

* GoldenGate 也被用于最小化停机时间数据库升级

PayPal: Exadata MAA for Mission Critical OLTP



极限性能

- 每秒执行30万次以上处理
- 99.99%的关键交易实时分析
- 99.99%的处理平均响应时间40毫秒
- 比以前的系统性能提升10倍以上

HA and MAA

- 99.99% 高可用
- 所有磁盘组DG使用3倍冗余
- Active Data 自动块修复和DR
- 使用ASM, CRS, Data Guard和GoldenGate实现滚动升级
- 开启闪回

生产加灾备集群 = 8 Exadata Racks

3 套一致的架构 = 24 Exadata Racks + Test/Dev Resources supporting our Critical Applications.

目标

- 基于交易数据做实时查询
- 每天处理2000万笔资金交易
- 响应时间小于1秒
- 零数据丢失容灾保护

解决方案

- 2台1/4配置 Exadata V2 - Oracle Maximum Availability Architecture

"service has been live on 99.99% uptime. The service processes billions of Euros per week and achieves subsecond response times for online enquires." - Martin McGeough, Database Technical Architect

Pre-Exadata 4 ProdServers, 2 SANs



- **Electronic Payment System**
- 混合负载，大量OLTP和批处理

2011

Exadata V2 Quarter Rack



- 使用智能闪存和智能扫描，不修改应用，大幅提升性能

Active Data Guard
Synchronous Zero Data Loss

Exadata V2 Quarter Rack



Exadata
测试系统

- 零RPO, 15 分钟RTO
- 自动块修复
- Standby-first maintenance

快速可扩展性能

零非计划停机

更高的处理能力

容灾保护零数据丢失

"service has been live on 99.99% uptime. The service processes billions of Euros per week and achieves subsecond response times for online enquires." - Martin McGeough, Database Technical Architect


<540ms
response




20 million
transactions/day



ORACLE

业务挑战

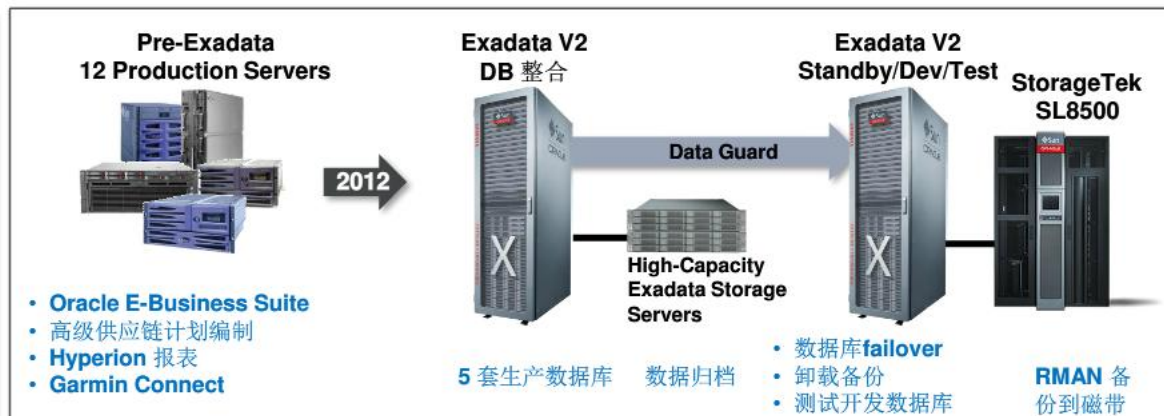
- 增加容错
- 标准化降低成本
- 简化支持模式
- 提升性能和扩展性

解决方案

- 2012: 整合12套服务器到2台

收益

"Oracle Exadata helped us meet our challenges by enabling us to consolidate to a single footprint."
- Ed Link, Vice-President IT, Garmin



更快的月终报表

99.95% 高可用

4倍用户增长

数据中心降低成本

Up to 11X

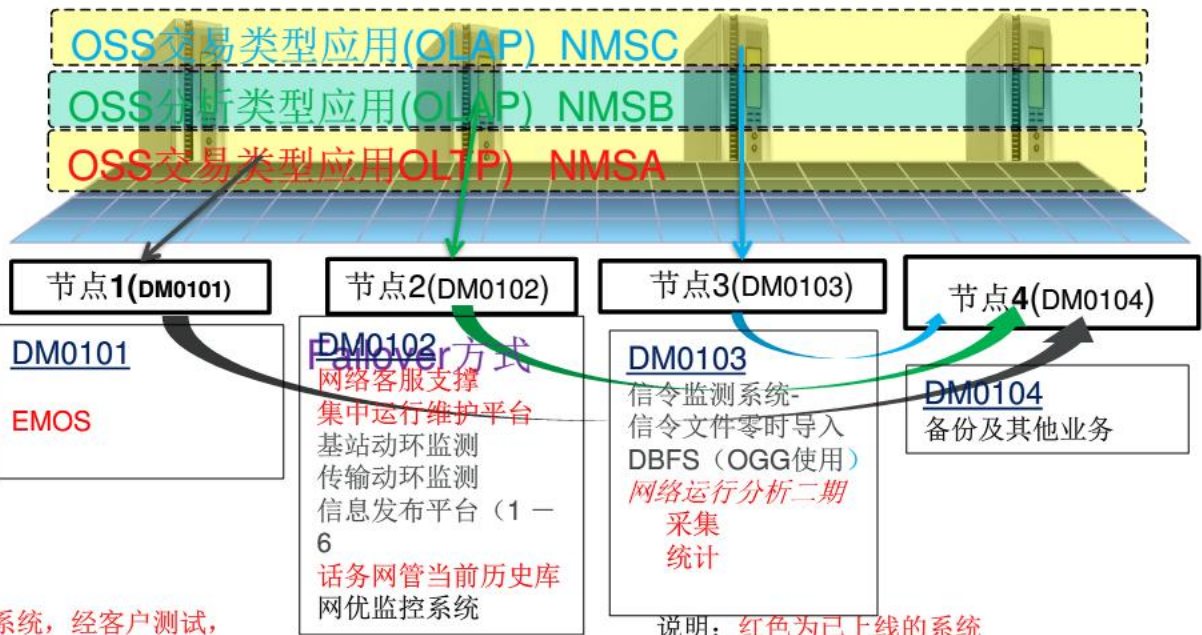


ORACLE

Data Guard部署建议



Half Rack
二分之一配



目前上线的系统, 经客户测试,
整体性能提升9倍。

Oracle on Oracle



ORACLE

Oracle Beehive

100,000+ 员工使用的协作套件技术



▪ 很多实时通讯:

ORACLE®



Email

每天270万邮件



Calendar

每天10万事件



文件共享

每天9.3万文件被阅读



聊天

每天220即时消息

ORACLE

Oracle Beehive: 充分验证的MAA

业务挑战

- 满足3年以上CPU和存储需要
- 提高响应时间
- 确保不间断运行

解决方案

- 2009: 迁移Beehive存储到 Exadata V1 存储
- 2011: 迁移到 Exadata X2-2

收益

"[Beehive] is our largest application in-house. It is Oracle's largest backend database."

- Campbell Webb, Vice-President IT, Oracle

Exadata V1 Storage Servers



- 96 V1 storage servers
- Post-Sun acquisition, CPU and disk oversubscribed

2011

Exadata X2-2 Production

Austin (Texas) Data Center



- 9 full-rack X2-2
- 2.3 Petabytes raw disk
- 48 TB flash
- 5,000+ peak TPS

Data Guard
1200+ miles
ASync

Exadata X2-2 Standby

Utah Data Center



- 9 full-rack X2-2
- ASM High Redundancy
- FRA, Flashback, RMAN, OSB
- StorageTek SL3000 Tape Library

满足容量增长
需求



更快的响应

5x – 60x



100% 高可用



ORACLE

更多信息

- **OTN HA Portal:**
<http://www.oracle.com/goto/availability>
- **Maximum Availability Architecture (MAA):**
<http://www.oracle.com/goto/maa>
- **MAA Blogs:**
<http://blogs.oracle.com/maa>
- **Exadata on OTN:**
<http://www.oracle.com/technetwork/database/exadata/index.html>
- **Oracle HA Customer Success Stories on OTN:**
<http://www.oracle.com/technetwork/database/features/ha-casestudies-098033.html>

Q & A

ORACLE