DBAplus

数据库技术沙龙-武汉站

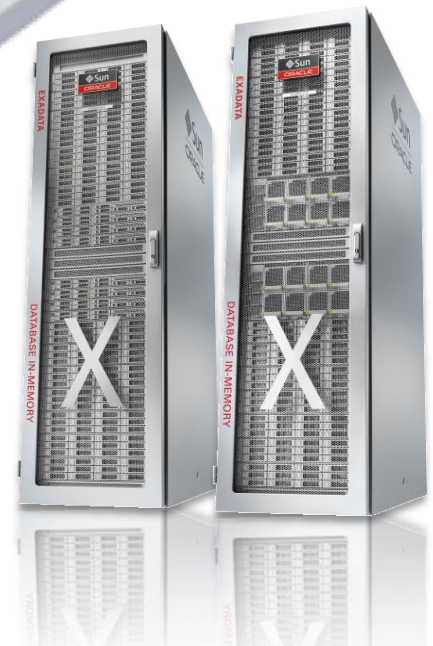
Exadata技术架构及保险行业实施

李明明

Exadata
基础技术架构介绍

Exadata released version

- Exadata V1
- 2008
- Exadata V2
- 2009
- Exadata X2
- 2010
- Exadata X3
- 2012
- Exadata X4
- 2013
- Exadata X5
- 2015
- Exadata X6
- 2016
- Exadata X7
- 2017



Exadata technology evolution

The best running platform of oracle database in all scenarios

- Exadata Cloud Service
- In-Memory Columnar in Flash
- Smart Fusion Block Transfer
- In-Memory Fault Tolerance
- Direct-to-wire Protocol
- JSON and XML offload
- Instant failure detection
- Network Resource Management
- Multitenant Aware Resource Mgmt
- Prioritized File Recovery
- 3D V-NAND Flash
- IO Priorities
- Data Mining Offload
- Offload Decrypt on Scans
- Software-in-Silicon

Smart Software

- Database Aware Flash Cache
- Storage Indexes
- Columnar Compression
- PCIe NVMe Flash
- Unified InfiniBand
- Smart Scan
- InfiniBand Scale-Out

Smart Hardware

- DB Processors in Storage
- Scale-Out Storage
- Scale-Out Servers

Exadata hardware architecture

Infiniband switch

Sun Datacenter InfiniBand Switch 36(x6-2)
A high-bandwidth low-latency 40 Gb/second InfiniBand network connects all the components inside an Exadata Database Machine.

Ethernet switch(Cisco)

External connectivity to the Exadata Database Machine is provided using standard 10 Gigabit Ethernet.

The scale-out architecture database platform

8 Database servers each with two 22-core x86 processors and 256 GB of memory (expandable up to 1.5 TB).
（X6-2）

Scale-out and intelligent storage servers

Two configurations：
High Capacity (HC)
four PCI Flash cards each with 3.2 TB (raw) Exadata Smart Flash Cache and twelve 8 TB 7,200 RPM disks
(X6-2 301 GB/S sql ,2million IOPS)
Extreme Flash (EF)
all-Flash configuration with eight PCI Flash drives, each with 3.2 TB (raw) storage capacity
(X6-2 350 GB/S sql ,2.4million IOPS)
At least two database servers and three storage servers, which can be elastically expanded by adding more database and/or storage severs as requirements grow

Expansion rack

1/4 .. 1/2, full(38U)
....
8 Rack
(X6-2)

Exadata software architecture

Exadata Database platform:
Oracle RAC

ASM
use libcell connect to cellsrv

Exadata storage node:
Storage Server Software

processes:
cellsrv
cellsrvstat
cellrssrm
cellrssmt
cellrsbkm
cellrsbmt
cellrsmmt
cellrsomt

physical disk Lun celldisk

griddisk asmdisk

CellCLI> list physicaldisk
12*8 TB Disk + 4*3.2 TB PCI flash = total 16
CellCLI> list lun
1:1 physicaldisk
CellCLI> list celldisk
1:1 lun
CellCLI> list griddisk
n:1 celldisk
CellCLI> create griddisk DATAC1_CD_00_jtcw01celadm07 celldisk=CD_00_jtcw01celadm07,size=xxxG
All cell griddisk(7 node) = all griddisk
= select count(*) from v$asm_disk;
select name ,TOTAL_MB/1024 GB from v$asm_disk;
(X6-2 1/2 rack)

Exadata software architecture

Exadata Smart Scan

The data search and retrieval processing can be offloaded to the Exadata Storage Servers. This feature is called Smart Scan. Using this Smart Scan, Oracle Database can optimize the performance of operations that perform table and index scans by performing the scans inside Exadata Storage Server, rather than transporting all the data to the database server.
Smart Scan capabilities includes :
1) Predicate Filtering
2) Column filtering
3) Join Processing
The definitive list of which functions are offloadable for your particular version is contained inV$SQLFN_METADATA.
SQL> select * from v$sqlfn_metadata where offloadable = 'YES';
A list of conditional operators that are supported by predicate filtering include =, !=, <, >, <=, >=, IS [NOT] NULL, LIKE, [NOT] BETWEEN, [NOT] IN, EXISTS, IS OF type, NOT, AND, OR.
Exadata Smart Scan FAQ (Doc. ID 1927934.1)

Exadata storage index

They are not indexes that are stored in the database like Oracle's traditional B-Tree or bitmapped indexes. They are not capable of identifying a set of records that has a certain value in a given column. Rather, they are a feature of the storage server software that is designed to eliminate disk I/O.
They work by storing minimum and maximum column values for disk storage units, which are 1 Megabyte (MB) by default and are called region indexes.
We can understand it in this way:
1， Dividing to storage on each ASM griddisk into chunks
2， Oracle Exadata storage cells can identify which areas of the disk storage will definitely not contain the values the query is interested in and avoid reading those areas.

Exadata Maximum availability architecture(MAA)

Active-active DB RAC clusters,
ASM镜像存储

冗余 的
数据库服务器
存储服务器
网络交换机
电源模块

最快RAC节点故障恢复
深度集成ASM镜像
最快备份 - RMAN offload to storage
最快Data Guard Redo Apply

主数据中心生产环境

同城灾备/异机房模块

异地灾备环境

Exadata Integrated software and hardware monitoring management

Oracle Exadata Database Machine Command-Line Interface (DBMCLI) utility

DBMCLI is the command-line administration tool for managing the database servers. DBMCLI runs on each server to enable you to manage an individual database server.
全面监控和恢复
硬盘，电池，InfiniBand端口，ILOM，CPU，内存，温度
CPU利用率, 内存利用率, 网络接口吞吐量，文件系统使用率
自动收集内核panic时console历史和告警
基于告警的阀值
告警配置类似存储服务器，并且通过E-mail和SNMP传送
部件故障时可以自动发送服务请求

Oracle Enterprise Manager

OEM

Exadata
保险行业实施分享

某保险公司重要业务系统

- 系统建设计目标与要求
- Exadata架构规划

某业务系统设计要求与目标

要求	具体描述	关注点
系统的扩展要求	建立企业弹性数据中心 满足未来2年的互联网快速业务发展需求、系统具有良好的横向扩展性	简单、动态、灵活的架构 资源的合理利用和分配 数据的集中整合
系统高性能要求	满足相关系统的高性能要求 满足业务高峰，尤其是“秒杀”活动的快速响应。	消除瓶颈，高效的业务数据访问 资源共享和负载均衡 更高的I/O吞吐量，更快的事务处理能力
稳定可靠性要求	具有可靠的软件体系支持。 内部软件核心软件安全可靠	冗余设计且多层次数据保护机制 内部软件的管理控制方面 资源使用的可控性
低成本、易管理	降低系统建设、系统维护及管理成本 统一管理，软硬件集中实时监控	提高资源利用率 系统集中监控，自动调优 简化管理，降低风险

业务设计目标

业务支撑目标对比去年：

业务单量总数X2
总出单时间/2
单位时间交易速度X4
在线人数X5

技术设计目标

1，中间件和数据库的并发扩容
2，应用计算能力提升、优化
3，DB计算能力和IO能力提升
4，实现更高可用
5，实现可扩展架构
6，网络扩容
7，软硬件一体化监控和运维
8，较易于改造和迁移实施
9，软硬件具有一定的性价比

Exadata 架构规划

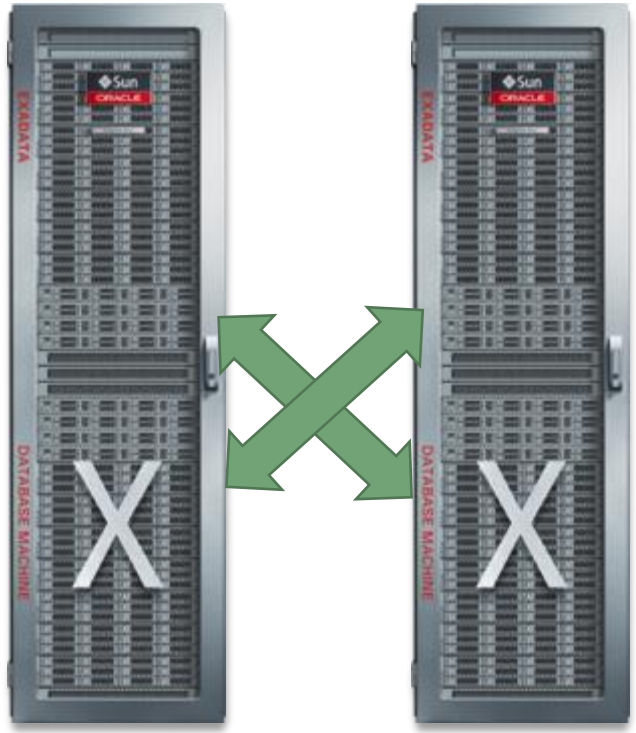
RAC DB1

应用A1

应用A2

应用An

RAC DG2



RAC DB2

应用B1

应用B2

应用Bn

RAC DG1

Q&A

感谢DBAplus和Oracle原厂支持

DBAplus

www.dbaplus.cn

THANK YOU !