ORACLE

Why Exadata is the Best Platform for Database In-Memory

May 2020 | Version 3.2

PURPOSE STATEMENT

This document provides an overview of features and enhancements when using Database In-Memory with Exadata. It is intended solely to help you assess the business benefits of using Database In-Memory with Exadata and to plan your I.T. projects.

DISCLAIMER

This document in any form, software or printed matter, contains proprietary information that is the exclusive property of Oracle. Your access to and use of this confidential material is subject to the terms and conditions of your Oracle software license and service agreement, which has been executed and with which you agree to comply. This document and information contained herein may not be disclosed, copied, reproduced or distributed to anyone outside Oracle without prior written consent of Oracle. This document is not part of your license agreement nor can it be incorporated into any contractual agreement with Oracle or its subsidiaries or affiliates.

This document is for informational purposes only and is intended solely to assist you in planning for the implementation and upgrade of the product features described. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described in this document remains at the sole discretion of Oracle.

Due to the nature of the product architecture, it may not be possible to safely include all features described in this document without risking significant destabilization of the code.

TABLE OF CONTENTS

INTRODUCTION

Oracle Database In-Memory (Database In-Memory) transparently accelerates analytic queries by orders of magnitude, enabling real-time business decisions. Database In-Memory uses a "dual-format" architecture that enables data to be maintained in both row format and a pure in-memory columnar format. This columnar format allows data to be scanned much faster than row formatted data. Database In-Memory is able to futher speed up scan performance by taking advantage of SIMD (Single Instruction, Multiple Data) vector processing and In-Memory Storage Indexes . With Database In-Memory it is possible to scan billions of rows per processor core per second purely in-memory, and this now makes it feasible for businesses to run real-time analytics on their critical business data without impacting the performance of their existing systems.

With the benefits of Database In-Memory, does it matter what platform you run your database on? Yes, the Oracle Exadata Database Machine (Exadata) has been the preferred platform for running Oracle Database since its release in 2008, and it provides distinct advantages for running Database In-Memory as well. The following are key advantages that Exadata uses with Database In-Memory:

- Exadata efficiently scales Database In-Memory
- Exadata provides a very fast interconnect with special protocols to speed up Database In-Memory scale-out
- Exadata provides high storage bandwidth to quickly populate the Database In-Memory column store
- In-Memory columnar formats in flash cache
- Exceed DRAM limits and transparently scale across Memory, Flash and Disk
- Exadata is Oracle's Database In-Memory development platform
- Exadata is a database consolidation platform and Database In-Memory further enables consolidation opportunities
- In-Memory fault tolerance
- In-Memory Aggregation optimization can be offloaded to Exadata storage cells
- Database In-Memory support for Active Data guard only on Exadata
- Automatic In-Memory only on Exadata
- In-Memory external tables only on Exadata

In this paper we will examine each of these points and explain in detail why Exadata is the best platform for running Database In-Memory.

EXADATA EFFICIENTLY SCALES DATABASE IN-MEMORY

Exadata uses a scale-out architecture for both database servers and storage servers. The Exadata configuration carefully balances CPU, I/O and network throughput to avoid bottlenecks. As an Exadata system grows, database CPUs, storage, and networking are added in a balanced fashion ensuring scalability without bottlenecks. This scale-out architecture can accommodate any size workload and allows seamless expansion from small to extremely large configurations while avoiding performance bottlenecks and single points of failure.

In a Real Application Clusters (RAC) environment, objects with the INMEMORY attribute specified can be distributed across the cluster by rowid range, by partition or by subpartition. Exadata is architected to accommodate the increased parallelism and interconnect messaging when the In-Memory column store (IM column store) is distributed across multiple RAC nodes.

EXADATA PROVIDES A VERY FAST INTERCONNECT WITH SPECIAL PROTOCOLS TO SPEED UP DATABASE IN-MEMORY SCALE OUT

Exadata uses an InfiniBand interconnect between the database servers and storage servers. Each InfiniBand link provides 40 Gb/second of bandwidth – many times higher than traditional storage or server networks. Further, Oracle's interconnect protocol uses direct data placement (DMA – direct memory access) to ensure very low CPU overhead by directly moving data from the wire to database buffers with no extra data copies. The InfiniBand network has the flexibility of a LAN network, with the efficiency of a SAN. By using an InfiniBand network, Exadata ensures that the network will not bottleneck performance. The same InfiniBand network also provides a high performance cluster interconnect for RAC nodes. When scaling out Database In-Memory on Exadata this high-speed transfer and large bandwidth for messaging between IM column stores keeps the IM column stores transactionally consistent and in sync with each other. This enhances scale out for distributed objects as well as objects that have been duplicated.

EXADATA PROVIDES HIGH STORAGE BANDWIDTH TO QUICKLY POPULATE THE DATABASE IN-MEMORY COLUMN STORE

When data is initially populated into the IM column store it is read directly from disk in its row format, converted to a columnar format and then compressed. The faster you can read the data, the faster you can complete the population process. Exadata storage offers outstanding IO performance ensuring the data population process is not I/O bound.

The population process is conducted by a set of background worker processes. These worker processes can operate in parallel to populate the IM column store as fast as data can be read off of disk and CPUs can process that data. This is where the high I/O performance and CPU resources of Exadata come into play to make the population of the IM column store as fast as possible. The number of background worker processes can also be controlled to take further advantage of Exadata's scalability.

Database In-Memory will also repopulate IMCUs when the number of stale entries in an IMCU reaches a staleness threshold. Again, with Exadata's high I/O performance this can occur in the background with no noticeable effect on application performance.

IN-MEMORY COLUMNAR FORMATS IN FLASH CACHE

Exadata uses a unique set of software algorithms to implement database intelligence in storage, NVMe flash storage, and RDMA over Converged Ethernet. A full rack Exadata X8-2 Database Machine with HC Storage Capacity offers 360TB of flash, which is nearly 30X the capacity of DRAM, and can achieve up to 350 GB/second of analytic scan bandwidth from SQL[1].

It is now possible to store data in the In-Memory columnar format in the flash cache in an Exadata environment. This enables all of the In-Memory optimizations (accessing only the compressed columns required, SIMD vector processing, storage indexes, etc.) to be used on a much larger amount of data.

When the INMEMORY_SIZE parameter is set to a non-zero value objects accessed using a Smart Scan will be brought into Exadata flash cache and will be automatically converted into the In-Memory columnar format. The data will be rewritten in the background into Database In-Memory columnar format, and this will result in all subsequent accesses to the data benefitting from all of the In-Memory optimizations when that data is retrieved from the flash cache.

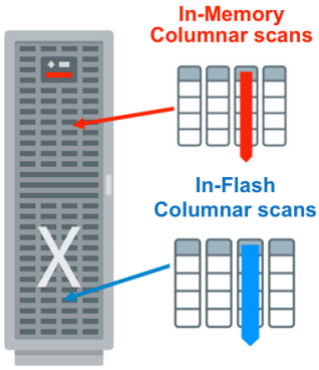

Figure 3. All of the benefit of In-Memory columnar now available on Exadata Flash

EXCEED DRAM LIMITS AND TRANSPARENTLY SCALE ACROSS MEMORY, FLASH AND DISK

With Exadata, your application can make use of all storage tiers (memory, flash, & disk) without having to be aware of where the data resides or suffer suboptimal performance when not all of the data resides in-memory in the IM column store. On Exadata data can reside in the IM column store, in the database buffer cache, in flash storage in columnar or row format, or on disk storage and your application never needs to be aware of data location because Oracle Database can seamlessly access that data.

When data resides on Exadata storage servers, Exadata's Smart Flash Cache feature or In-Memory Column Cache can dramatically accelerate Oracle Database processing by speeding I/O operations. Exadata Smart Flash Cache provides intelligent caching of database objects to avoid physical disk I/O. Exadata storage also provides an advanced compression technology, Hybrid Columnar Compression (HCC), that typically provides 10x level of data compression and boosts the effective data transfer by an order of magnitude.

1 Exadata Database Machine X8-2 Data Sheet

This means that all data access, and not just data that has been populated into the IM column store in DRAM, will be as efficient as possible.

Exadata also includes Smart Scan, a unique technology that offloads data-intensive SQL operations into the Oracle Exadata storage servers. This is similar to and complements Database In-Memory processing by pushing SQL processing into the Exadata storage servers when data is not in the IM column store. Data filtering and processing occurs immediately and in parallel across all storage servers as data is read from disk or flash. Exadata Smart Scan reduces database server CPU consumption and greatly reduces the amount of data moved between storage and database servers. This enables scaling and efficient SQL processing across all storage tiers whether data resides in the IM column store, on flash storage or on disk storage.

EXADATA IS ORACLE'S DATABASE IN-MEMORY DEVELOPMENT PLATFORM

Exadata is the development platform for Database In-Memory. Thus, Database In-Memory issues are discovered and fixed on Exadata first. Exadata is also the primary platform for Oracle Database testing, HA best practices validation, integration and support. The same reasons it is the best platform for Oracle Database apply to Database In-Memory.

EXADATA IS A DATABASE CONSOLIDATION PLATFORM AND DATABASE IN-MEMORY FURTHER ENABLES CONSOLIDATION OPPORTUNITIES

Database consolidation is one of the major strategies that organizations use to achieve greater efficiencies in their operations. Increasing the utilization of hardware resources while reducing administrative costs are primary goals of consolidation projects. Exadata is optimized for Data Warehouse and OLTP database workloads, and its balanced database server and storage grid infrastructure make it an ideal platform for database consolidation. Exadata is a modern architecture featuring scale-out industry-standard database servers, scale-out intelligent storage servers and a high-bandwidth low-latency InfiniBand network that connects all servers and storage. In many ways Database In-Memory "completes" Exadata by applying in-memory performance techniques that are similar to those that are used by Exadata on flahs and disk. Exadata allows customers to simultaneously optimize performance and cost for analytic workloads by using Database In-Memory columnar formats in-memory and flash in conjunction with existing Exadata features to increase consolidation capacity for all data. The result is a solution that gives the speed of DRAM, the IOPs of flash, and the cost effectiveness of disk.

IN-MEMORY FAULT TOLERANCE

Given the shared nothing architecture of the IM column store in a RAC environment, some applications may require a fault-tolerant option. On Exadata it is possible to mirror the data populated into the IM column store by specifying the DUPLICATE subclause. This means that each In-Memory Compression Unit (IMCU) populated into the IM column store will have a mirrored copy placed on one of the other nodes in the RAC cluster. Mirroring the IMCUs provides in-memory fault tolerance as it ensures data is still accessible via the IM column store even if a node goes down. It also improves performance, as queries can access both the primary and the backup copy of the IMCU at any time.

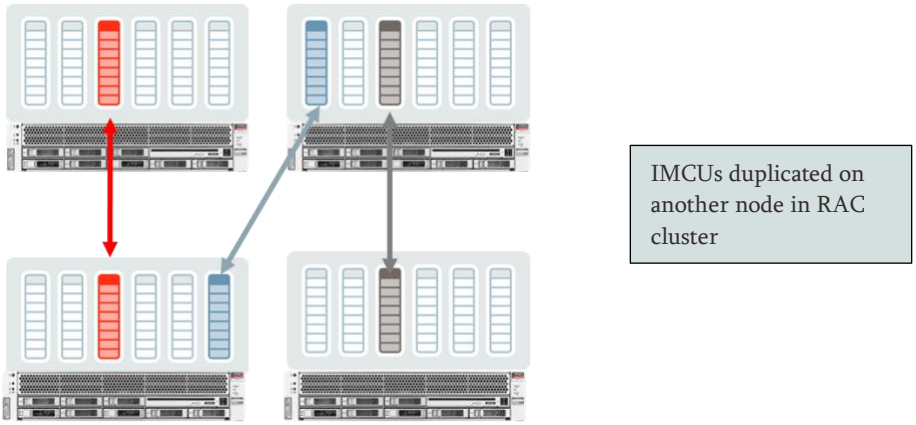

Figure 1. Objects in the IM column store on an Exadata Database Machine can be mirrored to improve fault tolerance

Should a RAC node go down and remain down for some time, the only impact will be the re-mirroring of the primary IMCUs located on that node. Only if a second node were to go down and remain down for some time would the data have to be redistributed.

If additional fault tolerance is desired, it is possible to populate an object into the IM column store on each node in the cluster by specifying the DUPLICATE ALL sub-clause. This will provide the highest level of redundancy and provide linear scalability, as queries will be able to execute completely within a single node.

The DUPLICATE ALL option may also be useful to co-locate joins between large distributed fact tables and smaller dimension tables. By specifying the DUPLICATE ALL option on the smaller dimension tables a full copy of these tables will be populated into the IM column store on each node. In the example in Figure 2, when a query joins a partition of the sales table to one or more of the dimension tables all of the data required for the join will be in the local node, avoiding having to fetch data across nodes to complete the join.

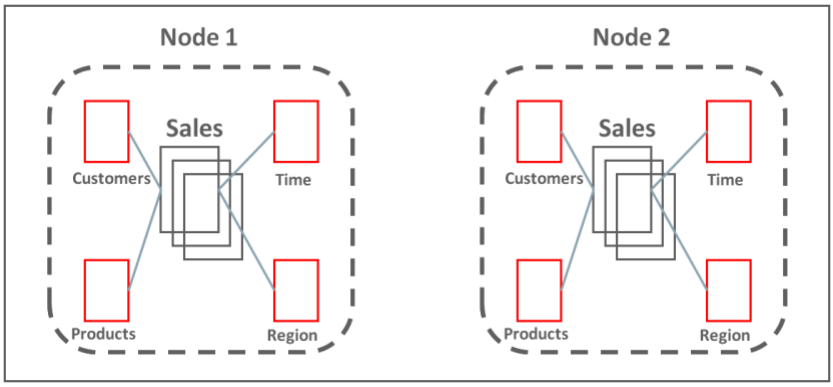

Figure 2. Distributed fact table with duplicated dimension tables

IN-MEMORY AGGREGATION OPTIMIZATION CAN BE OFFLOADED TO EXADATA STORAGE CELLS

With the introduction of Database In-Memory comes the new In-Memory Aggregation optimization, or vector group by feature. In-Memory Aggregation (IMA) provides new SQL execution operations that accelerate the performance of a wide range of analytic queries against star and similar schemas. These include the KEY VECTOR USE and VECTOR GROUP BY operations which enable the use of a vector transformation plan that minimizes the amount of data that must flow through the execution plan. This minimizes the amount of CPU used as compared to alternative plans.

The result of this is that IMA can transform joins to KEY VECTOR filters on the fact table and aggregate data in a single pass while lowering CPU use. This is extremely fast when the entire table resides in the IM column store, but what if the entire table doesn't fit into the IM column store? On Exadata when tables are accessed, and they have not been populated in the IM column store, IMA is enhanced by the ability to offload the KEY VECTOR USE operation to Exadata storage servers. This might occur when the table is partitioned and only the most recent partitions are loaded into the IM column store and the other partitions are on disk. The offload capability distributes key vector processing across Exadata storage servers and minimizes the volume of data that must be returned to the database nodes.

DATABASE IN-MEMORY SUPPORT FOR ACTIVE DATA GUARD ONLY ON EXADATA

Oracle Active Data Guard is the most comprehensive solution available to eliminate single points of failure for mission critical Oracle databases. It prevents data loss and downtime in the simplest and most economical manner by maintaining a synchronized physical replica of a production database at a remote location. If the production database is unavailable for any reason, client connections can quickly, and in some configurations transparently, failover to the synchronized replica to restore service. It also eliminates the high cost of idle redundancy by allowing reporting applications, ad-hoc queries, and data extracts to be offloaded to read-only copies of the production database.

Active Data Guard is unique in using a highly parallelized process to apply changes to a standby database for best performance while enforcing the same read consistency model as the primary database. Active Data Guard has been tightly integrated with Database In-Memory on Exadata, providing users the ability to enable the IM column store on the primary, standby or both environments.

With synchronized physical replication and read-consistency, in-memory processing on Active Data Guard is a viable solution for running read-only workloads instead of running those on the primary. It makes it possible to run real-time analytics on the standby database with no impact on the production database, making productive use of the standby database resources, and at the same time increasing the total columnar capacity of the system.

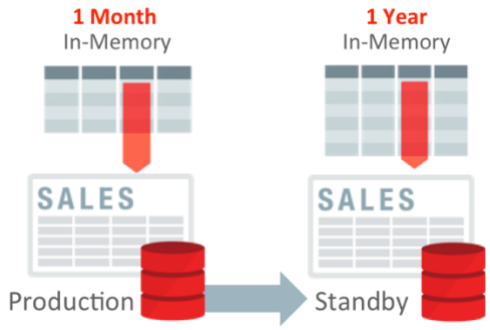

Figure 4. Example of how the IM column store on the standby database can have very different content to the primary

By considering the standby environment as a separate database, Database In-Memory makes it possible to populate the same or a different set of tables or table partitions in-memory on the primary and on the standby database. Just as Active Data Guard maintains a synchronized physical replica of the production database, it also maintains the contents of the IM column store ensuring transactionally consistent results as of the query SCN.

AUTOMATIC IN-MEMORY ONLY ON EXADATA

Automatic In-Memory (AIM) is available on Exadata to automatically manage the contents of the IM column store. If the sum of the space of the segments that have been enabled for in-memory exceeds the available memory in the IM column store then AIM will kick in and manage the IM column store space. AIM automatically manages objects populated into the IM column store using access tracking and column statistics. With AIM enabled segments can be automatically evicted from the IM column store to make room for the population of more active segments.

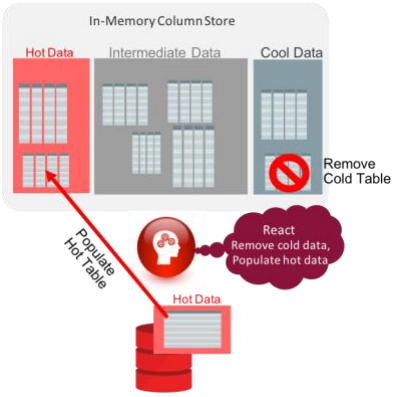

Figure 5. AIM management of IM column store segments

IN-MEMORY EXTERNAL TABLES ONLY ON EXADATA

Database In-Memory supports populating external tables directly into the IM column store on Exadata. The data does not have to be materialized in the row store first and this can be very useful for short-term data that must be scanned repeatedly in a short time span, data aggregated by NoSQL tools that must be joined to relational data, and data that must be queried by both Oracle Database and NoSQL tools.

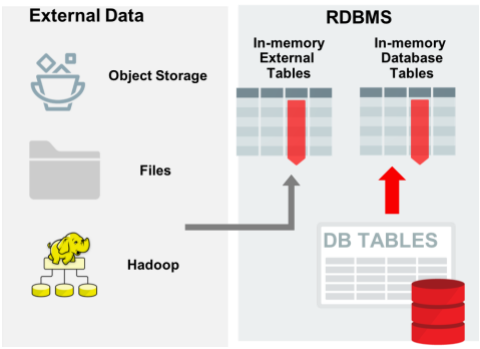

Figure 6. Database In-Memory external table support

CONCLUSION

Exadata is the best platform for running Oracle Database and Database In-Memory. Database In-Memory takes full advantage of Exadata's unique hardware features, enabling better performance than any other hardware platform. These features include a very fast interconnect enabling IM fault tolerance and scale-out, high storage bandwidth and IOPs enabling fast IM column store population, seamless access to all storage tiers and the running of mixed workload environments. Exadata is also an excellent consolidation platform with support for Oracle Trusted Partitions to limit the number of Oracle software licenses to just those that are needed and elastic configurations so that you only configure the hardware you need. All of this along with Oracle's commitment to ensuring that all hardware and software components are pre-configured, pre-tuned and pre-tested to work seamlessly together for the best possible performance and reliability in the industry make Exadata the best platform for running Oracle Database and Database In-Memory.

CONNECT WITH US

Call +1.800.ORACLE1 or visit oracle.com.
Outside North America, find your local office at oracle.com/contact.

blogs.oracle.com facebook.com/oracle twitter.com/oracle

Why Exadata is the Best Platform for Database In-Memory
May, 2020
Author: Andy Rivenes