

Oracle TimesTen In-Memory Database on Oracle Exalogic Elastic Cloud

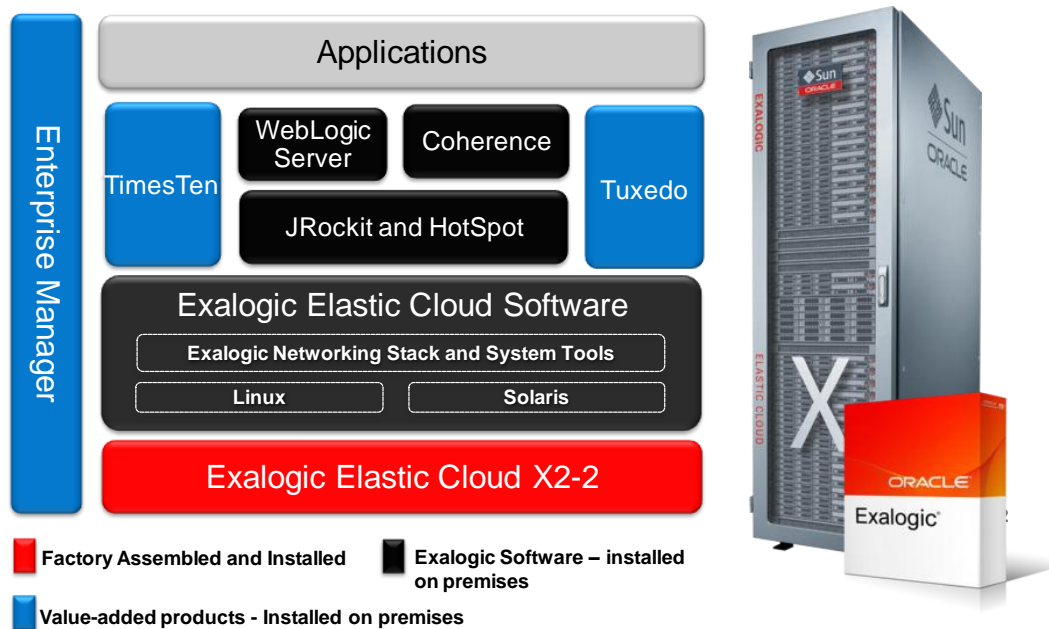
ORACLE WHITE PAPER | JUNE 2014



Executive Summary


With the growing speed of messages moving through business networks, the use of real-time processing to capture, analyze, and respond intelligently to key events has become the benchmark for corporate excellence. Not only are the execution and management of critical business processes important, customers expect highly tailored interactions and the utmost responsiveness from any company with which they do significant business.

Oracle TimesTen In-Memory Database (TimesTen) is a memory-optimized relational database. TimesTen provides real-time data management for performance critical systems deployed in the application tier to deliver blazing-fast response and real-time caching of application data from the Oracle Database to the TimesTen database.



Oracle Exalogic Elastic Cloud (Exalogic) is an integrated hardware and software system designed to provide a complete platform for a wide range of application types and widely varied workloads. Oracle Exalogic is intended for large-scale, performance-sensitive, mission-critical application deployments. It combines Oracle Fusion Middleware software and industry-standard Sun hardware to enable a high degree of isolation between concurrently deployed applications, which have varied security, reliability, and performance requirements.

Real-time OLTP applications can benefit greatly from the combined compute power of Exalogic and TimesTen.



Using the TimesTen TPTBM performance program running an 80-10-5-5 workload (80% read transactions, 10% update transactions, 5% insert transactions, and 5% delete transactions), throughput on each compute node reached 1.6 million Transactions per Second (TPS).

Using a workload simulating a Prepaid Mobile application or an Online Banking application, TimesTen achieved peak throughput of 8.7 million TPS running on a quarter-rack Exalogic configuration (8 compute nodes). More details of the tests are presented in subsequent sections of this paper.

Introduction

Oracle TimesTen In-Memory Database 11g has been certified for the Oracle Exalogic Elastic Cloud. This white paper describes configuration considerations, best practices and performance results of TimesTen running on Exalogic.

TimesTen is targeted to run in the application tier, close to applications, and optionally in-process with applications to deliver very low response time and very high throughput for performance-critical applications. TimesTen can be used as the database of record or as a relational cache database to the Oracle Database.

The enterprise architectures that derive the greatest benefit from real-time processing provide event, data, and transaction management in the application tier, empowering front-line systems with rapid response and deeper insight. It is not sufficient to merely collect and cache data next to applications, as is often the case with first-generation in-house efforts. Nor is it practical to locate the corporate database on the same tier with the applications.

Enterprise companies can extend their software infrastructures with TimesTen to create systems that are instantly responsive, highly scalable, and continuously available using the familiar SQL programming interfaces with minimal changes. These systems are used to increase customer loyalty, attract new customers, streamline operations, and avoid the costly alternative of proprietary software development.

The Exalogic certification was performed using the TimesTen 11.2.1.8.3 release with a quarter-rack Exalogic version 1.0 server. All TimesTen product components were installed and tested, including TimesTen Replication, Application-Tier Database Cache, integration with Oracle Clusterware, Oracle SQL Developer, and TimesTen Plug-in for Oracle Enterprise Manager. Our testing results showed that TimesTen is extremely fast, reliable, and resilient on Exalogic.

Hardware and Software Overview

The Exalogic test server was configured with a quarter rack consisting of 8 Compute Nodes.

Compute Node

Each of the 8 Compute Nodes has

- » Two sockets of 6-core Xeon X5670 processors at 2.93 GHz
- » 96GB RAM
- » 32GB Flash/Solid State Disk (internal storage local to each compute node)

The compute nodes are connected to a shared Storage Server.

Storage Server

Shared storage is provided by a dual-controller Sun 7320 storage server regardless of whether the configuration is a full, half or quarter rack. The storage server consists of two Intel x86_64 Xeon servers configured to run an active-standby Sun ZFS cluster. The storage server has 40 TB of disk space.

InfiniBand

The compute nodes are interconnected through a Sun Network QDR InfiniBand switch. The storage servers are also connected to the InfiniBand switch.

Standard Oracle Linux

For the purpose of running TimesTen, each compute node can be viewed as an individual Linux server. Kernel configuration parameters should be set following the guidelines in *the Oracle TimesTen In-Memory Database Installation Guide*.

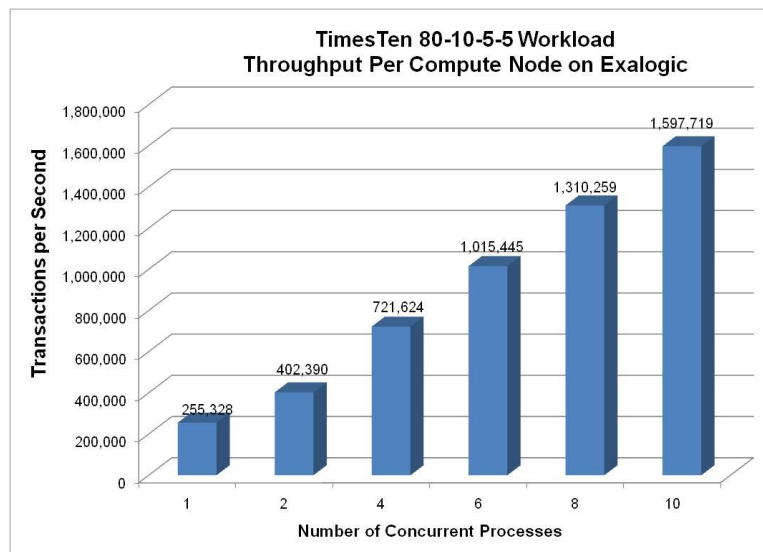
TimesTen Performance and Scalability on Exalogic

This section describes 3 different performance tests demonstrating TimesTen performance running on Exalogic.

Test #1: TPTBM 80-10-5-5 Workload

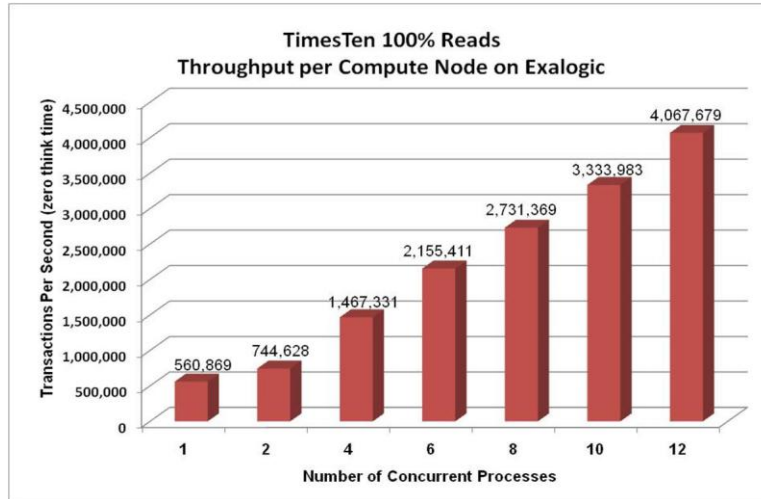
TPTBM is a sample program shipped with the TimesTen product to enable customers to conduct simple performance tests to assess base-line performance of the server where TimesTen is installed.

Using a workload of 80-10-5-5 with 80% reads (SELECT), 10% updates, 5% inserts, and 5% deletes, the peak throughput measured using 10 concurrent TPTBM processes achieved an unprecedented throughput of about 1.6 million transactions per second per compute node with very stable response time. Note that each compute node consists of 12 cores; we reserved 2 cores for the TimesTen data manager for transaction execution including transaction persistence to the disk on the shared storage server.



Test #2: TPTBM 100% Read Workload

The second test involved a workload of 100% read transactions. As shown in the chart below, TimesTen scales extremely well on Exalogic, achieving over 4 million reads per second with 12 concurrent processes executing SELECT queries as fast as possible (with zero think time).

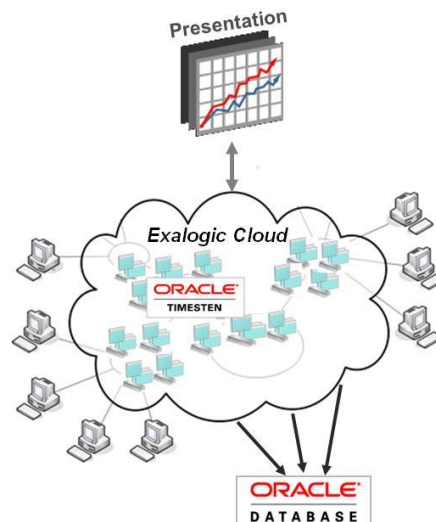


Test #3: Prepaid Mobile (PPM) and Online Banking (OLB) Application Workload

In this test, the PPM/OLB workload represents a typical real-time prepaid mobile application with 5 different types of transactions:

- » Authenticate user and account status
- » Check account balance
- » Update account profile
- » Refill, credit, or debit account
- » Search accounts with a low balance

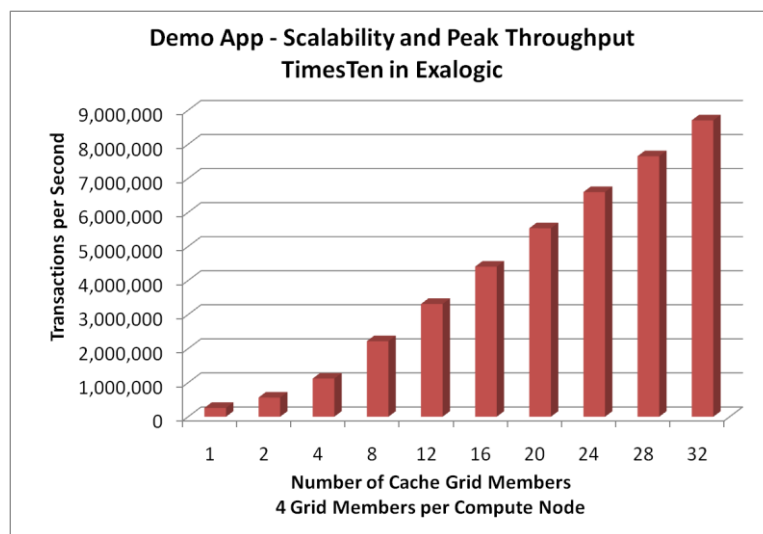
The application was implemented in Java using the JDBC interface to TimesTen. The workload transactions were executed and committed in TimesTen; the committed data were asynchronously propagated to the Oracle Database. Response time and throughput were captured in a TimesTen table for visual display by the presentation program, as shown in the figure below:



TimesTen In-Memory Cache Grid was configured on all 8 compute nodes in the Exalogic machine. Four cache grid members were configured for each compute node to take advantage of the compute power and available RAM.

Mobile prepaid subscriber applications and banking applications share similar transaction characteristics where real-time access is required for very high volume transactions.

As shown in the chart below, TimesTen scales very well as the number of grid members increases. With 32 grid members running the workload, the peak throughput reached an impressive 8.7 million transactions per second for all 8 compute nodes.



Installation

Exalogic File System Layout

From the operating system perspective, each compute node has 5 mounted file systems.

/ or the root volume is located on the SSD. Swap is also located on the SSD. In the first Exalogic release, SSD is the only storage internal to any compute node. All other storage is located on the shared storage server.

Each compute node has 4 NFS mounts to the storage server. These "NFS shares" are divided into 2 global shares and 2 local shares. Files placed on the local shares are visible only to the compute node that mounts them; files on the global shares are visible across all compute nodes. The file systems from the storage server appear as:

```
/u01/common/patches      # global share
/u01/common/images       # global share
/u01/<hostname>/dumps     # local share
/u01/<hostname>/general   # local share
```

Location of TimesTen Installation Binaries

TimesTen must be installed on every compute node where it is to be used. Sharing installations across compute nodes is not supported. Customers should not install TimesTen on a global share and use that instance on more than one compute node.

Oracle Clusterware (CRS) Considerations

The default NFS shares in the Exalogic machine are not mounted with noac option required for a CRS voting disk as specified in the *Oracle Clusterware Installation Guide*. It is recommended to create a new global share with the proper options, and then mount this file system to the compute nodes that will be clustered using CRS. Customers should ensure that enough IP addresses are provisioned, including any desired VIPs used by CRS.

TimesTen Database Files and Transaction Log Files

File Placement and Access

Each compute node has 32 GB of internal disk storage (SSD). In practice, only 14-16 GB of free space is available on these file systems since the Operating System installation and swap space use up the rest of the space. The free space on the SSD may not be adequate to hold TimesTen database checkpoint files or transaction log files. To set up appropriate data access control from each of compute nodes, the TimesTen database files and transaction log files should be placed on the local share storage.

NFS Shared File System

To place the TimesTen database and transaction log files on an NFS local shared file system, it is necessary to set the `-allowNetworkFiles` option in the `<install-dir>/info/ttendaemon.options` file before starting the TimesTen daemons.

Database Recovery

In the event of a node failure and if the database is not using TimesTen Replication for high availability, the database checkpoint files and the transaction log files of the failed compute node cannot be “recovered” using a TimesTen installation from a different compute node.

Customers who are not using TimesTen Replication should use `ttBackup/ttRestore` or `ttMigrate` utility to transfer database and transaction log files among compute nodes. Note that it is supported to take periodic backups on one compute node, store the files on a global share, and restore them to a distinct TimesTen installation on a second compute node.

Users can monitor the activity of the storage server, as described in the *Oracle Fusion Middleware Machine Owner's Guide*, by directing a browser to the Sun 7320 management interface at the following URL: `https://<storage-node>:215/#status/dashboard`

Performance Considerations

The `/u01/<hostname>/general` file system shows slightly higher transaction logging throughput than the `/u01/<hostname>/dumps` file system. Using connection attribute `LogFlushMethod=1`, (the default setting) provides higher write throughput than `LogFlushMethod=2`.

On Exalogic, the storage server is a shared resource across the compute nodes. Storage I/O activity on one compute node affects I/O activity on other compute nodes. It does not matter whether the I/O activity is to a local share or a global share. Local and global refer to the visibility of file systems; the I/O to a local share still goes through the shared storage server.

Given the shared nature of the storage server, it is recommended that customers use the `CkptRate` connection attribute (or `ttCkptConfig` built-in procedure) to throttle the rate of checkpoints. Un-throttled checkpoints can use up all the I/O bandwidth available on any one compute node, reducing the I/O bandwidth available on other compute nodes. A checkpoint rate of 10-25 MB/sec should not cause undue interference with logging nor with other compute nodes.

Since each of the compute nodes has 12 cores, it is recommended to set `LogBufParallelism <=12`. Setting it too low can impede throughput. Setting it to higher values can cause the log flusher to incur unnecessary overhead. An application that uses fewer than 16 active threads can set the value to the number of active threads or 12.



In summary, the following are examples of recommended settings:

Datastore=<file path for the database on local share>
LogFile=<directory on local share>
LogBufParallelism=12 # number of transaction log buffer strands
LogFlushMethod=1 # direct I/O used regardless of setting on NFS file systems
LogBufMB=1024 # larger value helps parallelism, async replication; max=1024
LogFileSize=1024 # >= LogBufMB (maximum value)
CkptRate=20 # MB/sec; adjust for your particular requirements
CkptFrequency= 600 # default frequency 10 minutes
CkptLogVolume=0 # governed by frequency and rate
PrivateCommands=1 # improves concurrency at expense of temp space
RecoveryThreads=12 # 12 core system
MemoryLock=<3 or 4> # setting for large memory systems

Benefit From InfiniBand

Presently, the TimesTen 11g release makes use of the InfiniBand via TCP/IP protocol (EoIB). Future releases of TimesTen may take advantage of other InfiniBand protocols.

TimesTen Resources

For additional product information, please reference the TimesTen Product Documentation set either

- » Online at <http://www.oracle.com/technetwork/database/database-technologies/timesten/documentation/index.html> or
- » Local copy of the documentation set that is optionally placed within the TimesTen installation directory <TimesTen Install Dir>/doc

The TimesTen technical forum is available on the Oracle Technology Network



https://community.oracle.com/community/developer/english/oracle_database/timesten_in-memory_database



Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

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

Hardware and Software, Engineered to Work Together

Copyright © 2014, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0614

 | Oracle is committed to developing practices and products that help protect the environment