Benchmark: Oracle Healthcare Transaction Base on Intel® Xeon

An Oracle White Paper
September 2006

v3

JOINT TESTING PROGRAM

- Oracle Software
- Intel Hardware

LOCATION

- Intel Solution Center
- Chantilly, Virginia, USA
- September, 2006

TEST VARIABLES

- Stored Data Volume
- Read/write API Transactions
- Messaging Throughput
- Hardware Configurations

EXECUTIVE OVERVIEW

Performance benchmarking of the Oracle Healthcare Transaction Base occurred at the Intel Solution Center in Chantilly, Virginia. These tests documented Health Level 7 (HL7) Version 3[1] messaging throughput and API transaction response times associated with increasing users, stored data volumes and changes in hardware configurations. Findings illustrate that good response times can be achieved on cost-effective hardware platforms while applying substantial system loads.

INTRODUCTION

The increasing volume of healthcare information, especially with the recent interest in large, regional healthcare information exchanges, has made cost-effective, high-performance, healthcare data platforms an imperative for the industry. Oracle and the Intel Digital Healthcare Group tested a scalable architecture that allows organizations to support a healthcare data exchange deployment. Hardware infrastructures, from Intel, along with healthcare software components, from Oracle, have the ability to scale from small community organizations to large regional delivery networks as well as supporting incremental additions in scale over time.

This performance benchmarking, conducted by Intel® Solution Services at its Chantilly center, document the performance characteristics of Oracle's Healthcare Transaction Base (HTB) under increasing loads. Benchmarking Configurations and Results are presented below.

KEY BENCHMARK OBJECTIVES

- Demonstrate HTB's performance and scalability characteristics for small and medium sized healthcare IT systems[2]
- Determine HTB's scalability characteristics with increasing stored data volumes
- Determine HTB's scalability characteristics with increasing system load (increasing transaction volume)

---

[1] HL7 Version 3 refers to HL7 standards that are based on the HL7 Version 3 Reference Information Model. See www.hl7.org.

[2] Larger database servers were placed out of scope for this exercize

- Determine response time characteristics
- Determine maximum system throughputs

Benchmarking configurations

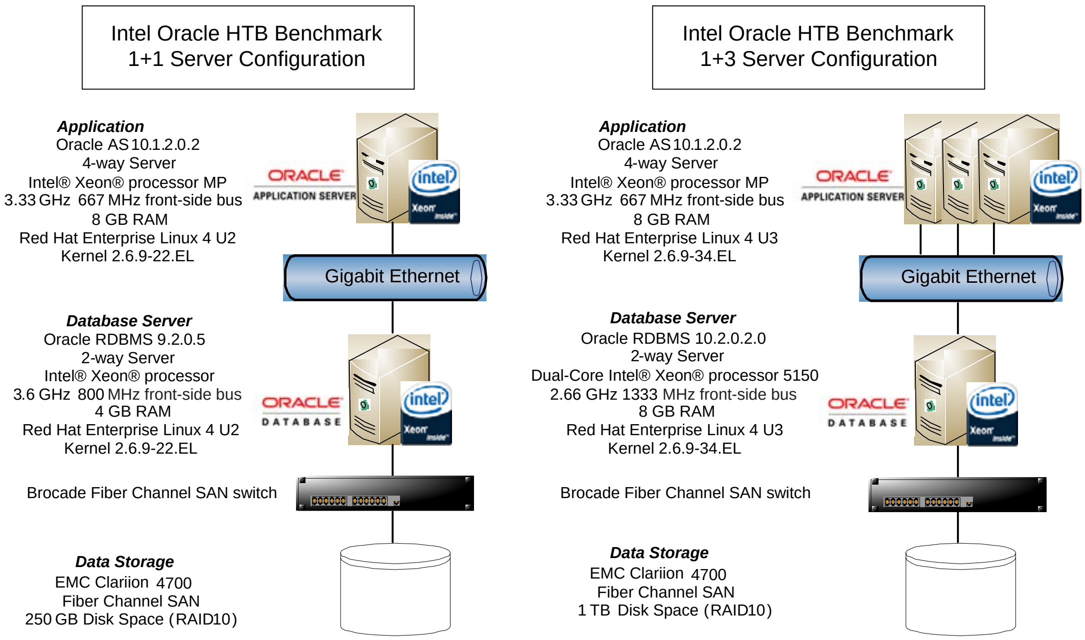

Table 1: Software Versions

Component	Version
Oracle Healthcare Transaction Base	V5.2
Oracle RDBMS	10.2.0.2.0 for the 1+ 3 server configuration, 9.2.0.5 for the 1+1 server configuration
Oracle Application Server	10.1.2.0.2
Oracle E-Business Suite	11.5.10.2
Mercury LoadRunner®	8.0

Table 2: 1 + 1 Server Configuration

1 + 1 Server Configuration

Oracle Database Server	1 x Intel Xeon 2-way 5100 3.6 GHz, 800 MHz front-side bus, 4GB RAM Red Hat Enterprise Linux 4 U2, Kernel 2.6.9-22.EL
Oracle Application Server	1 x Intel Xeon 5100 4-way 3.33 GHz, 667 MHz front-side bus, 8GB RAM Red Hat Enterprise Linux 4 U2, Kernel 2.6.9-22.EL
Storage	EMC Clarion 4700 250 GB SAN
Unique Patients	250,000
Years of Data	Dates for data ranged over five years
Encounters	625,000
Average Message Load	2,000 HL7 messages/hour

Table 3: 1 + 1 Server Configuration

1 + 3 Server Configuration

Oracle Database Server	1 x Intel® Xeon® Woodcrest-5150 2-way MP 2.56 GHz w/ 8GB RAM Red Hat Enterprise Linux 4 U3, Kernel 2.6.9-34.EL
Oracle Application Server	3 x Intel® Xeon® Paxville-7040 4-way MP 3.33 GHz w/ 8GB RAM Red Hat Enterprise Linux 4 U3, Kernel 2.6.9-34.EL
Storage	EMC Clarion 4700 1 TB SAN
Unique Patients	1,000,000
Years of Data	Dates for data ranged over five years
Encounters	2,500,000
Average Message Load	10,000 HL7 messages/hour

USER LOAD

- Users were added until transaction response times began to degrade. Results were reported at 75% of the peak number of users

APPLICATION MIXES

- Application mixes consisted of both read and write transactions

TRANSACTION THROUGHPUT

- Translates to Concurrent Users
- Response Time good up to 75% Load

BENCHMARKING METHODS

The number of users was increased on the system and key metrics were captured until the overall system throughput started to drop; i.e., the system started running out of resources to support additional load. These loads were applied by increasing the number of application users simulated with Mercury LoadRunner. Users executed transactions every thirty seconds. Unless otherwise specified, results were reported at 75% of this high point for read/write API transaction throughput. This 75% benchmark was chosen because most production systems will try to keep peak concurrent users within 50% to 75% of capacity in order to keep performance high under all loads.

HTB supports a large library of read and write Java API's for application development. A variety of API transaction mixes simulated various types of application load (numbers in parentheses refer to the approximate ratio of user activity in the system at any given instant):

- Application Mix 1 (TMX1): Results retrieval (90%), EMPI (10%)
- Application Mix 2 (TMX2): Clinical data entry (50%), EMPI (50%)
- Application Mix 3 (TMX3): Results retrieval (40%), clinical data entry (10%), order management (30%), patient registration (10%), EMPI (10%)

Each application mix was measured with a constant messaging load on the system (as noted in the configuration section). These HL7 V3 messages averaged 24KB in size and consisted of ADT, orders, labs and public health.

BENCHMARK RESULTS

Load Scalability

The load scalability tests demonstrate the effect of increasing transaction load on the response time.

1. 1+1 Server Configuration

TMX1	25% Load	50% Load	75% Load
Throughput (transactions/hour)	9,500	19,000	28,500
Concurrent Users[3]	80	160	240
Response Time (seconds)	0.24	0.24	0.29

[3] Assumption: a user executes one transaction every 30 seconds.

WRITE VERSUS READ TRANSACTIONS

- Write Transactions cause more load on system
- Read Response Times faster than Write Response Times
- Smaller configuration slowed more with write transactions than larger configuration

TMX2	25% Load	50% Load	75% Load
Throughput (transactions/hour)	5,000	10,000	15,000
Concurrent Users	40	80	120
Response Time (seconds)	0.43	0.43	0.59

MIXED TRANSACTION LOAD

- Illustrates good response time in both configurations

TMX3	25% Load	50% Load	75% Load
Throughput (transactions/hour)	11,000	22,000	33,000
Concurrent Users	90	180	270
Response Time (seconds)	0.20	0.20	0.26

2. 1+3 Server Configuration

TMX1	25% Load	50% Load	75% Load
Throughput (transactions/hour)	22,400	44,800	67,200
Concurrent Users	185	370	555
Response Time (seconds)	0.15	0.2	0.3

LARGER CONFIGURATION

- Improves write transaction response times
- Supports increased number of users
- Supports larger patient populations
- Supports larger message volumes

TMX2	25% Load	50% Load	75% Load
Throughput (transactions/hour)	15,000	30,000	45,000
Concurrent Users	130	260	390
Response Time (seconds)	0.25	0.3	0.4

TMX3	25% Load	50% Load	75% Load
Throughput (transactions/hour)	20,000	40,000	60,000
Concurrent Users	170	340	510
Response Time (seconds)	0.35	0.5	0.7

VOLUME SCALABILITY

This volume scalability test demonstrates the effect of increasing data volumes on overall throughput and response time. Application mix TMX3 was used for this test, and the throughput and response time were measured at 75% peak load on a dedicated, separate configuration[4] (distinct from the 1+1 or the 1+3 configurations).

EFFECT OF STORED DATA VOLUME

- No degradation in Response Time with stored data volume up to 1TB (maximum tested)

Stored Data Volume	25% Volume (250 GB)	50% Volume (500 GB)	100% Volume (1000 GB)
Response Time (seconds)	0.8	0.9	0.5[5]
Throughput (transactions/hour)	40,100	45,000	45,000

[4] Application Server: 3 x 4 way Intel Xeon processor MP, 3.33 GHz, 667 MHz front-side bus, 8 GB RAM; Database Server: 2 x 4 way Dual-Core Intel Xeon processor 7040, 3.0 GHz, 667 MHz front-side bus, 8 GB RAM

[5] The greater than anticipated drop in response time reflects additional tuning that was done on the system prior to recording the measurement. Without this additional tuning we would expect the response time to increase slightly from the 50% volume measurement.

PEAK SYSTEM PERFORMANCE

Maximum Transaction Throughput

Application Mix	1+1 Server Configuration (transactions/hour)	1+3 Server Configuration (transactions/hour)
TMX1	38,000	89,600
TMX2	20,000	60,000
TMX3	44,000	80,000

API TRANSACTION MIX SUMMARY

- Read API Transactions demonstrate higher system throughput
- Transactions correlate to concurrent users

Concurrent Users

These tests demonstrate the number of concurrent users that can be supported for each application mix on a given configuration. The number of users is measured at 75% transaction load on the system.

Application Mix	1+1 Server Configuration	1+3 Server Configuration
TMX1	240	555
TMX2[6]	120	390
TMX3	270	510

MESSAGING THROUGHPUT

- Includes storing messages in logs
- Includes transforming message data to relational format

Maximum Messaging Throughput

These tests demonstrate peak messaging throughput on a system without any other transaction load.

1+1 Server Configuration	10,000 messages/hour
1+3 Server Configuration	31,200 messages/hour

[6] The greater than expected increase in the number of concurrent users between the 1+1 server configuration run and 1+3 server configuration run for TMX2 is due to the fact that a poorly performing index was identified in the 1+3 server configuration run. Accordingly, additional database tuning was applied in mid-test.

DISCUSSION

HTB DESIGN CONSIDERATIONS

- HTB transforms messages to relational data

In the design of the HTB system, HL7 Version 3 messages are parsed and the data within the messages are written to a relational format that is optimized for transactional healthcare applications. Consequently, the native XML of the V3 messages is saved only in messaging logs destined for archive. Therefore, these messaging throughput results include the CPU cycles required for data transformation and relational storage of discrete data in addition to the simple processing of message text into database storage. This detail of design may be important when comparing these results to system designs that simply store the XML of the V3 messages.

DATA VOLUME MEASUREMENT

- Act Instance Count

This study introduces the concept of HL7 V3 "Act Instance Count" as a measure of HL7 V3 data volume in performance testing. In these results, the number of HL7 Encounter Acts and other HL7 Clinical Acts were reported. These two categories of HL7 Acts may be combined into a single measure of HL7 V3 data volume named "Act Instance Count." This Act Instance Count is easily measured in messages and in databases of different designs as long as the instance identifiers are preserved. "Act Instance Count," in addition to counts of unique patients and other V3 Entities, may allow better comparisons of benchmark performance metrics when comparing V3-based systems in the future.

SCALING TO POPULATION SIZE

- Higher messaging volumes
- Increasing application server numbers

This study illustrates the ability to scale to higher population numbers, higher messaging volumes and higher user volumes by adding additional CPU's and servers to the hardware configuration.

Of note is the dramatic increase in messaging throughput when moving from the 1+1 server to 1+3 server hardware configuration. This suggests that scaling to higher-volume messaging environments is best accomplished by scaling the hardware configurations.

INTERPRETING RESPONSE TIMES

- Average API Response Times
- End-user Response Times
- Response Time ranges

"API Response Times" should be interpreted as representing only a portion of the response time the end-user would experience in a graphical user interface (GUI) environment. If the GUI processing time added 0.5 seconds to an API Response Time of 0.5 seconds, the total response time the end-user would experience would be 1.0 seconds. That said, an average API Response Time of 0.5 seconds can be considered "good" in most GUI environments. The API Response Times in these test varied from 0.2 seconds to 1.0 second, depending on the volume of data returned.

These testing center results are best interpreted in light of messaging volumes and user loads reported from actual medical centers. However, a review of the literature does not produce exactly comparable data. The medical center messaging volumes appear to reflect data in HL7 V2 messages rather than HL7 V3 messages. Therefore, the results reported in this paper appear to be the first controlled tests on HL7 V3 message processing. Despite the qualifications noted in the comparisons below, these comparisons might be useful to organizations planning the move from HL7 V2 to HL7 V3-based systems:

STORED POPULATION NUMBERS IN TESTS

- Similar to reported target patient populations

STORED POPULATION VISITS IN TESTS

- Similar to ED Visits in reports
- Estimated at 1/3 visits for target patient populations

STORED CLINICAL DATA IN TESTS

- Estimated at ½ of total results reporting for target patient populations

AVERAGE MESSAGING LOAD IN TESTS

- Estimated at 40% than reported for reported patient populations

CONCURENT USERS

- Estimated higher than probable number of concurrent users in reported user populations when compared to potential user populations

Table 4: Comparison of HTB Test Data Volumes to Industry Reports

Five Years of Data--normalized from reports	**HTB Test Data**	**Metro Indian-apolis[1,2]**	**Utah Olympics Statistic[4]**	**Brigham & Women's Boston[5]**	**Saskatch-ewan, Canada[5]**
Unique Patients	1M	1.5M in Metro Area	Not Reported	Not Reported	1M
HL7 Encounter Acts	2.5M	2.5M Emergency (ED) Visits	4M ED plus Urgent Care Visits	2.6M Hospital Admission, Outpatient and ED Visits	Not Reported
HL7 Clinical Acts	27.5M	40M Results (normalized from informal report[2])	Not Reported	Not Reported	Not Reported
Messaging Load	10K per hour	6K per hour [2]	Not Reported	Not Reported	4K per hour (sim)
Potential # Users	NA	Not Reported	Not Reported	9,545	1,545
Concurrent # Users at 75% of peak	510 users (36,000 Trans-actions per hour)	Peak use 9700 "accesses" per hour for one hospital[2]	NA	NA	NA

[1] Finnell, JT, JM Overhage, PR Dexter, SM Perkins, KA Lane, CJ McDonald; "Community Clinical Data Exchange for Emergency Medicine Patients;" AMIA Symposium Proceedings 2003:pp235-8.

[2] McDonald, Clement J; "Regenstrief experience with electronic medical record systems-the good and the bad;" Stanford Center for Clinical Informatics, Stanford, CA, April 15, 2005.

[3] Gesteland, PH, RM Gardner, FC Tsui, JU Espino, RT Rolfs, BC James, WW Champman, AW Moore, MM Wagner; "Automated Syndromic Surveillance for the 2002 Winter Olympics;" JAMIA 2003;10:6:547-54.

4 Doolan, DF, DW Bates, BC James; "The Use of Computers for Clinical Care: A Case Series of Advanced U.S. Sites:" JAMIA 2003;10:94-107.

[5] McDaniel, JG; "Discrete-event Simulation of a Wide-area Health Care Network;" JAMIA 1995;2:220-237.

COMPARING TESTS TO PUBLISHED REPORTS

- Need for improved published statistics

The comparisons of the messaging volumes and user loads reported in this HTB study with messaging volumes and user loads in the referenced projects and medical centers indicates that these testing loads are comparable, at least within the same order of magnitude, to loads found in actual medical centers and smaller metropolitan areas. However, these actual production environments probably experience wider peaks and valleys in load over days, nights and weekends than those simulated in the laboratory environment. Better standardization of reporting in the literature would enhance the opportunity for realistic performance testing comparisons.

COMPARABILITY OF HL7 V2 AND V3 MESSAGE SIZE

- No method for comparing data volume density in messages

Another factor in comparison of HL7 V2 and HL7 V3 messages and systems is the question of data volume. The density of actual data in HL7 V2 "pipe and bar" messages is higher than that in HL7 V3 XML messages. Therefore, actual data volume cannot be compared based on message size alone when comparing HL7 V2 and V3 messages.

In the future, performance testing of systems with larger patient numbers and larger Act Instance Counts will be helpful in sizing hardware for larger metropolitan areas. Distinguishing the contribution of increasing numbers of application servers in combination with increasing size of database servers to performance benchmarking will also be helpful. Third, increasing the stored data volumes to represent longer periods of elapsed time and intensity of clinical activity will fine-tune these hardware requirements.

CONCLUSION

- HTB scales in stored data volume. When stored data volume was doubled, a fractional increase in response time was noticed.
- HTB scales in applied transaction/user load. Average response times increased within acceptable levels when the load on the system was scaled up. As the results demonstrate, systems should be sized based on the response times observed between 50% and 75% of peak throughput.
- In each of our tests, the database server CPU was the first resource to get completely consumed. This should be taken into account when application server/database server ratios are being determined. For this test configuration, the ratio of three application servers to one database server was optimum. However, for larger database servers, more application servers of this size would probably be needed to fully utilize the larger database server.
- Transaction throughput and messaging throughput scales with increased database processing power. We were not able to determine the effect of increasing application server processing power as we were constrained for resources on the database server.

Product Descriptions

Oracle Healthcare Transaction Base is a platform for the integration, development, and operation of healthcare applications across the healthcare continuum. It consists of a comprehensive, healthcare-specific data repository and standards-based information model coupled with a set of integrated services for

ORACLE HEALTHCARE TRANSACTION BASE

The following features are included in HTB:

- Clinical data repository
- Enterprise terminology normalization tools
- Extensive read/write Java API library based on the HL7 Reference Information Model

enterprise terminology, data normalization, security and auditing, and business process/workflow. The platform is designed specifically for the healthcare industry and supports meaningful data consolidation, rapid application development, and genuine interoperability across different systems. With Oracle HTB, organizations in the business of healthcare—from hospitals and insurers to government bodies and public health agencies—can more effectively integrate, manage, deliver and display information through the entire process of providing services. For more information, e-mail dan.russler@oracle.com or visit:
http://www.oracle.com/industries/healthcare/htb.html
http://www.oracle.com/industries/healthcare/healthcare-transaction-base-datasheet.pdf

http://www.oracle.com/industries/healthcare/index.html

INTEL DUAL-CORE SERVER ARCHITECTURE:

The following features are included in Dual-Core®:

- 64-bit Intel® Core Xeon microarchitecture
- 35% power reduction over previous Pentium processors
- Increased instructions per clock cycle
- Advanced Smart Cache
- Extended Memory 64 Technology
- Intel® I/O Acceleration

Intel Dual-Core: Intel Corporation's Dual-Core® sever platforms are based on the innovative 64-bit Intel® Xeon® processors. Combined with Fully Buffered DIMM (FBDIMM) technology, the Dual-Core® architecture improves performance 3 times over Intel's previous Pentium chips while cutting power use by 35%. Shared L2 cache allows each of the processors to share both code and data segments greatly increasing overall performance. The Fully Buffered DIMM architecture streamlines memory access for database buffer blocks providing a 3 X improvement over previous memory technology. The Intel® I/O Acceleration technology delivers up to twice the data movement while cutting CPU overhead by as much as 40 percent, supporting faster data transfer from database servers to requesting middle-tier applications. For more information, e-mail Stuart.A.Mathews@intel.com or visit:

http://www.intel.com/technology/computing/dual-core/

http://www.intel.com/design/intarch/dualcorexeon/benchmark_brief.pdf

http://www.intel.com/healthcare/

Intel® Solution Services is Intel Corporation's worldwide consulting organization, focused on driving business transformation and competitive advantage to companies through the practical application of new technologies, giving organizations the confidence to innovate. For more information visit:

http://www.intel.com/go/intelsolutionservices

Copyright 2006, Oracle. 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 is it 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, JD Edwards, PeopleSoft, and Siebel are registered trademarks of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

Oracle Healthcare Transaction Base Benchmark on the Intel Xeon Platform
September 2006
Authors: Anand Shroff, Dan Russler, MD

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com