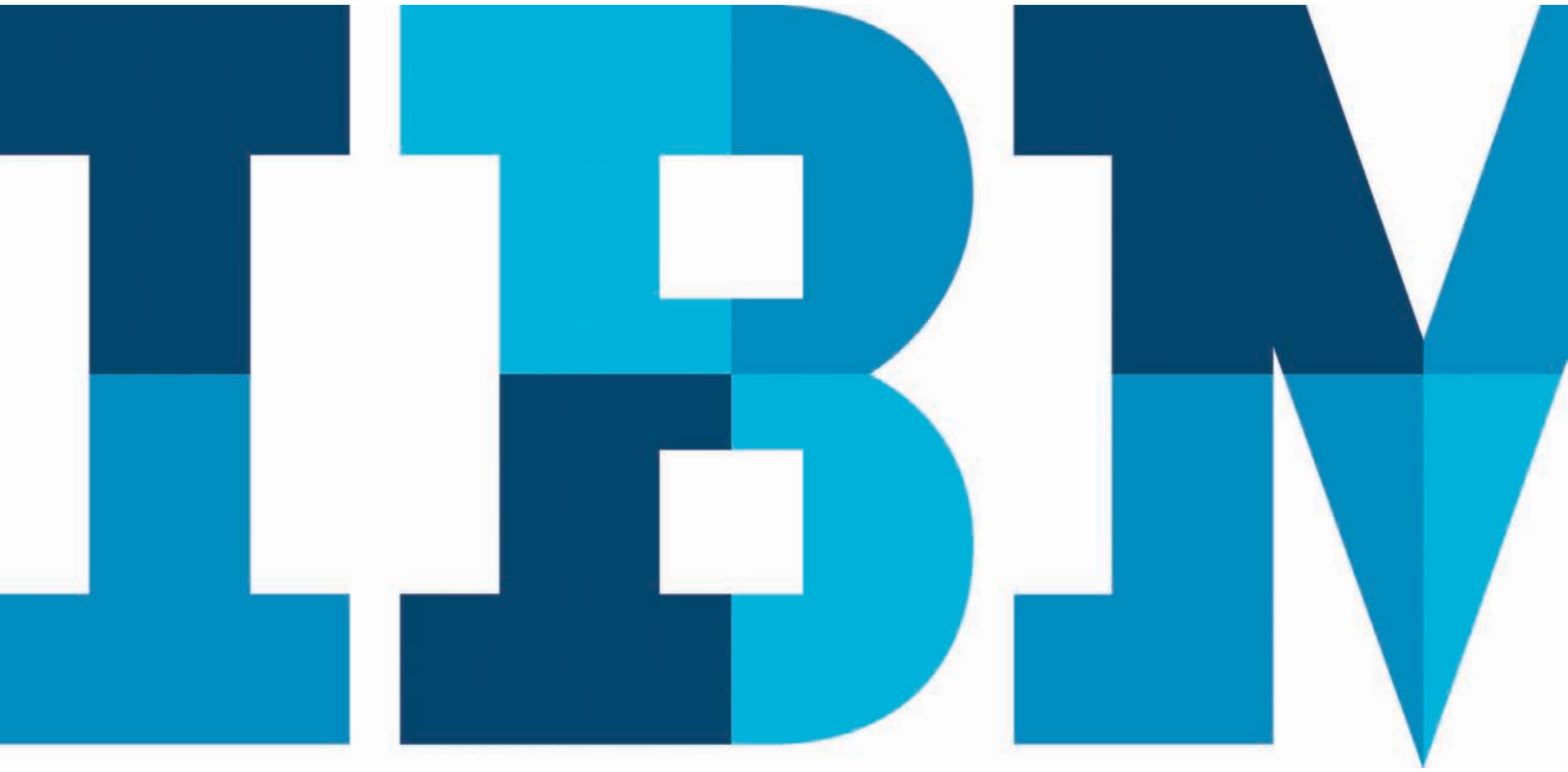
IBM Software

Thought Leadership White Paper

IBM InfoSphere Master Data Management

IBM InfoSphere MDM high-scale benchmark exceeds Oracle with twice the performance and reduced TCO



IBM

Contents

Executive summary

This paper shares recent performance benchmark results that demonstrate the superior performance and scalability of the IBM® InfoSphere® Master Data Management (InfoSphere MDM) solution. The performance benchmark was conducted by using IBM InfoSphere MDM 10.1, IBM WebSphere® Application Server Version 8 and IBM DB2® 10.1 for Linux, UNIX and Windows (LUW) with the DB2 pureScale® feature (DB2 pureScale). The InfoSphere MDM and WebSphere Application Server tier ran in an IBM PureFlex™ environment, and the DB2 pureScale database ran on a cluster of IBM System x® x3650 servers.

The paper demonstrates that the InfoSphere MDM solution running in this environment is a better-performing and more scalable MDM solution with a much lower total cost of ownership (TCO) than comparable offerings from Oracle,[1] making it the best MDM implementation in the marketplace today. The following list highlights some of the findings:

- More than 21,000 transactions per second (TPS) while executing a medium-complexity read workload
- Double the TPS using less than half the hardware when compared to the published Oracle result
- Good scalability, as demonstrated when the load on the system was increased up to three times (until the CPU capacity limit on the servers was reached)
- One-third to one-half of Oracle's reported response time for queries, even when throughput was more than double that of the reported Oracle result
- Similar response times for persistent transactions when compared to Oracle's results, even when throughput was 8 to 16 times higher

Business scenarios

Master data management (MDM) solutions help enterprise customers reduce costs, increase agility and achieve compliance. The IBM InfoSphere MDM solution offers a single, comprehensive and consistent view of the business and its customers, which enables decision makers to achieve better business results. InfoSphere MDM was designed with high performance and scalability requirements from its inception, and is thoroughly tested as part of a rigorous development cycle for each release.

The specifics of an individual MDM implementation will vary depending on the organization. As a proven and complete solution, InfoSphere MDM offers capabilities to meet a broad range of MDM needs. In numerous real-life scenarios, it serves as the foundation for the most significant business use cases: information consolidation within a single organization, information sharing across public sector agencies and collaborative authoring of production information between channel partners.

Delivering such capabilities requires a database architecture with exceptional levels of availability, scalability and application transparency. DB2 pureScale addresses all three of these requirements. With DB2 pureScale, you can scale out to meet business needs without changing your applications. DB2 pureScale is also designed to offer continuous availability through planned outages and more extreme multicomponent failures.

This paper showcases the performance of everyday transactions occurring in a real customer solution. The following transaction types were tested:

- **Medium Read:** Medium Read transactions are used to retrieve customer information. This transaction type retrieves customer names, addresses and contracts using a customer ID. This type of query is used in almost every transaction where a decision needs to be made whether to consolidate or to refresh the data.
- **Complex Read:** Complex Read transactions are similar to Medium Read transactions and are used when more details, such as contract components or contract role, are required.
- **Insert:** Insert transactions add new customer data to the MDM system.
- **Update:** Update transactions apply a heavier load than what is normally seen in a production system to generate more stress on the system under test. This transaction type updates the Person details, including name, address and customer account information on the contract, resulting in multiple changes per transaction.

Benchmark objectives

The main goal of this benchmark is to showcase the scalability and performance of the InfoSphere MDM solution by using the latest available (at the time of benchmark) IBM hardware and software, including PureFlex Systems, WebSphere Application Server and DB2.

Using workload specifications that are defined in an Oracle white paper, these tests demonstrate that the InfoSphere MDM solution exhibits superior performance and scalability and is the best of its breed. They also show that the InfoSphere MDM solution continues to meet very demanding performance requirements even when faced with significant business growth.

The benchmark had the following specific objectives:

- Build a scalable system with the highest possible throughput for the hardware
- Achieve the lowest possible sub-second response time
- Follow existing best practices and establish new best practices for high performance

Workload description

The workload for this benchmark was designed to mimic the day-to-day common services that run on an operational MDM system. To set up an operational MDM system, the database was prepared by loading a Parties table, which mimicked data consolidation from five different source systems in the same institute, such as a bank, insurance company or retail company. This consolidation was achieved by using Suspect Duplicate Processing functionality in InfoSphere MDM and resulted in 400 million conditioned parties as described above.

The data that was used to load the system is a representation of US census data in terms of ages, names and addresses.

The following table defines the transaction types that were used.

Transaction type	Description
Medium Read	Get Party details, including addresses, contracts and assets by using the Party ID from the operational system (133 attributes)
Complex Read	Get Party details (203 attributes)
Update	Update Party details (48 attributes)
Insert	Add new Party details (up to 62 attributes)

Transactions were individually executed using the InfoSphere MDM batch processor, a multithreaded, long-running Java application that submits large volumes of transactions into the InfoSphere MDM system for processing.

Benchmark environment

The software stack and the hardware stack that were used for this benchmark included the latest available products from IBM at the time of the benchmark, as shown in Figure 1.

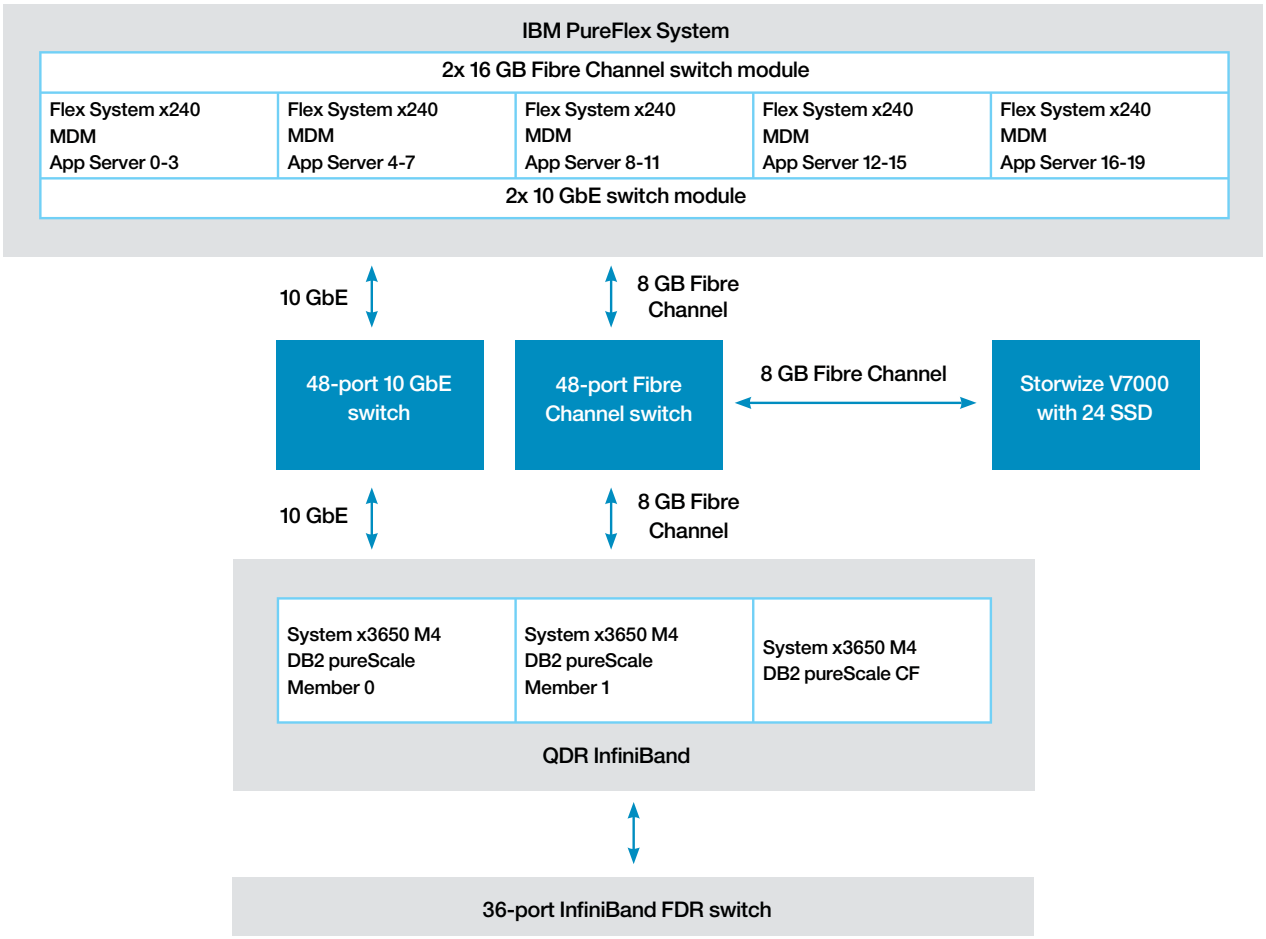
Software

- IBM InfoSphere MDM 10.1
- IBM DB2 10.1 for LUW with pureScale feature
- IBM WebSphere Application Server 8

Hardware

- Application servers
  - IBM PureFlex System (five IBM Flex System™ x240 for application servers)
- Database servers
  - IBM System x (three x3650 for database servers)
  - IBM Storwize® V7000 (twenty-four 300 GB solid-state drives)

All systems were equipped with 16-core Intel Xeon E5-2690 processors (2.9 GHz clock speed).



For more information about the products and acronyms, see the Glossary.

Figure 1. The benchmark environment.

Benchmark results

Performance data, including throughput, response time and CPU consumption were collected and analyzed.

As shown in Figure 2, InfoSphere MDM with WebSphere Application Server and DB2 pureScale delivered a throughput of 21,398 TPS, which is approximately twice the reported throughput for Oracle's MDM solution with Exalogic and Exadata when running the Medium Read workload. Similarly, IBM delivered 2.9 times higher throughput when running the Complex Read workload, 9 times higher throughput when running the Update workload and 16 times higher throughput when running the Insert workload.

As shown in Figure 3, InfoSphere MDM response times were three times faster (using 128 cores) than Oracle MDM response times (using 324 cores).

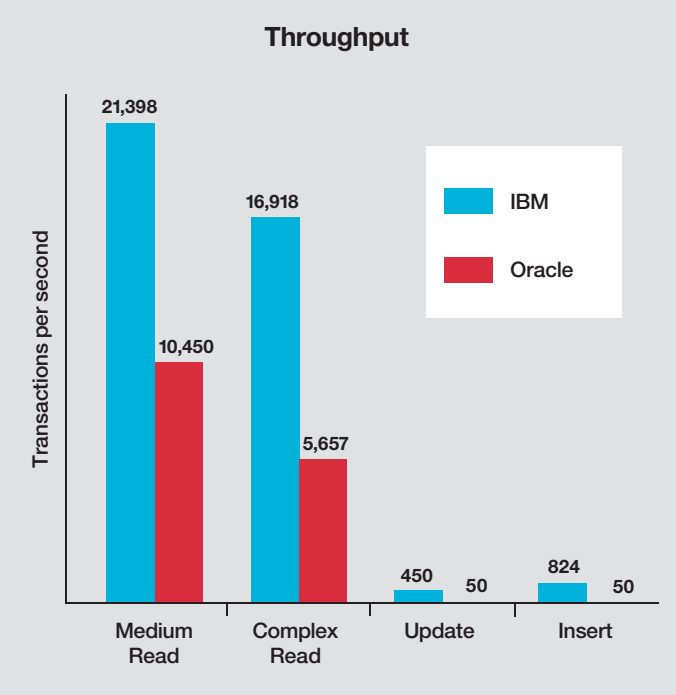


Figure 2. Performance of InfoSphere MDM compared to Oracle MDM while processing four transaction types of varying complexity.

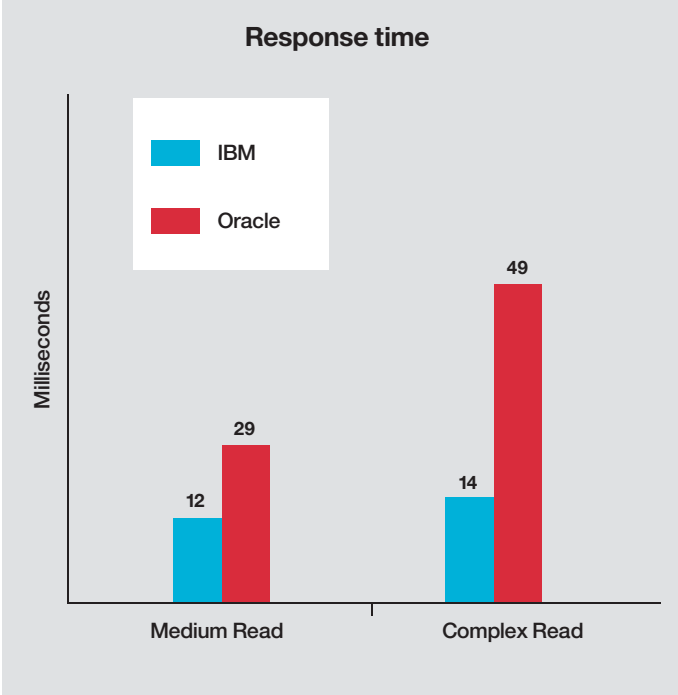


Figure 3. Response times of InfoSphere MDM compared to Oracle MDM while processing two different transaction types.

Figures 4 and 5 show the throughput curve when the number of concurrent threads in the InfoSphere MDM system is scaled up for Medium Read and Complex Read workloads, respectively.

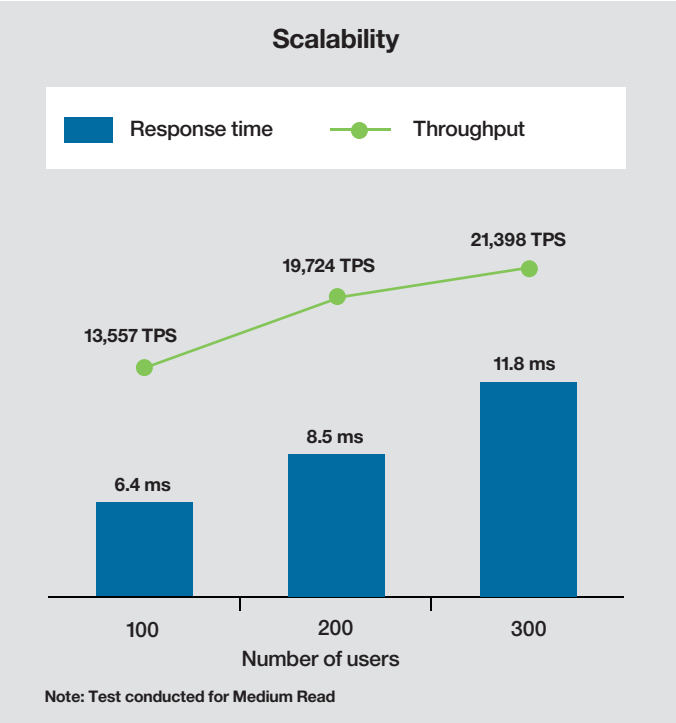


Figure 4. Scalability performance for Medium Read.

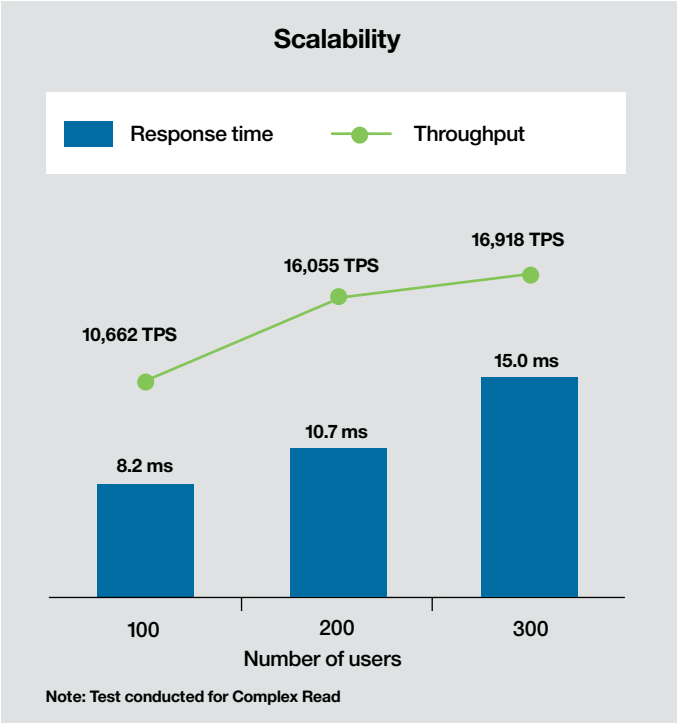


Figure 5. Scalability performance for Complex Read.

The IBM system is already well utilized at 100 users, and achieves higher throughput than what is reported in the Oracle case study with 300 users. The scalability from 100 users to 200 users already shows symptoms of being constrained by database CPU capacity. Scalability is further constrained when going from 200 to 300 users, because the system CPU capacity is close to the limit. For CPU utilization data from each test, see Appendix 2: Scalability test results for Medium and Complex Reads.

The results demonstrate that with sufficient resources, the InfoSphere MDM solution achieves near-linear scalability and a completely utilized system with very high throughput and low transaction response times.

Here is a quick summary of the results from both IBM and Oracle:

- **IBM solution:** InfoSphere MDM 10.1, WebSphere Application Server 8, DB2 pureScale 10.1 (IBM application tier: InfoSphere MDM and WebSphere Application Server on IBM PureFlex x240 ITE, total 5 ITEs, each ITE with 16c Intel Xeon E5-2690 2.9 GHz, 256 GB RAM; database tier DB2 pureScale on System x x3650 M4, total of three x3650 M4, each x3650 M4 with 16c Intel Xeon E5-2690 2.9 GHz, 384 GB) and Storwize V7000 (twenty-four 300 GB SSD drives). Scenario transactions per second: Medium Read 21,398, Complex Read 16,918, Update 450, Insert 824; October 21, 2012.
- **Oracle solution:** Oracle application tier using one-half rack Exalogic, database tier one-half rack Exadata. Scenario transactions per second: Medium Read 10,450, Complex Read 5,657, Update 50, Insert 50. Source: http://ebookbrowse.com/high-perform-computing-mdm-1396935-pdf-d243290221, June 2012.
  Note: The preceding Oracle source does not explicitly mention the Exadata model used. Calculations assume a one-half rack Exalogic x2-2 and a one-half rack Exadata x2-2. Results are valid as of October 21, 2012.

Summary

InfoSphere MDM provides a range of options for organizations that require different implementation styles. For years, InfoSphere MDM has been identified in Gartner reports as an industry leader on the basis of IBM's completeness of vision relative to the market, and IBM's ability to execute on that vision.[2] Given the strategic importance of leveraging the business value and unlocking the potential of master data in an enterprise, more and more companies are implementing their MDM solutions by using InfoSphere MDM products.

With an ever-growing number of lines of businesses connecting to MDM systems within the same company, requirements to serve more users and provide faster response times have increased quickly over the past few years. InfoSphere MDM products are designed with high performance and scalability considerations in mind, and are highly integrated with WebSphere Application Server and DB2.

An InfoSphere MDM solution with WebSphere Application Server running in a PureFlex environment using a DB2 pureScale cluster as the back-end database is an excellent choice for enterprises. This solution enables organizations to meet ever-growing business needs and to continue unlocking the business value of master data at very high speeds.

The performance benchmark demonstrates not only that the InfoSphere MDM solution delivers high performance with great scalability, but also that it achieves this at a lower cost than the Oracle MDM solution.

This high performance, together with the comprehensive functionality and styles that the InfoSphere MDM portfolio offers, make it clear that InfoSphere MDM is the best solution in the marketplace.

Appendices

Appendix 1: Performance test results

Table 1 demonstrates significantly higher performance of InfoSphere MDM on DB2 pureScale and WebSphere Application Server over Oracle MDM on Exalogic and Exadata.

	IBM solution			Oracle solution			
Scenario	**Definition**	**Throughput (TPS)**	**Response time (ms)**	**Definition**	**Throughput (TPS)**	**Response time (ms)**	**TPS gain**
Medium Read	Get Party details, including addresses, contracts and assets by using the Party ID from the operational system (133 attributes)	21,398	12	Get Customer details, including addresses, financial accounts and assets by using the Customer ID from the operational system (up to 25 attributes)	10,450	29	2x
Complex Read	Get Party details (203 attributes)	16,918	14	Get Customer details (up to 150 attributes)	5,657	49	2.9x
Update	Update Party details (48 attributes)	450	200	Update Customer details (25 to 40 attributes)	50	200	9x
Insert	Add new Party details (up to 62 attributes)	824	240	Add new Customer details (25 to 40 attributes)	50	200	16x

Table 1. Comparison of IBM solution and Oracle solution performance results.

Appendix 2: Scalability test results for Medium and Complex Reads

Table 2 shows the results of a scalability test for Medium and Complex Reads. In both cases, the database server was constrained by CPU.

Scenario	Number of users	Throughput (TPS)	Application server CPU	DB2 pureScale members CPU	CF CPU	Application server TPS/core
Medium Read	100	13,557	28.4%	50.0%	30.8%	596.70
	200	19,724	47.84%	80.8%	31.3%	515.36
	300	21,398	55.9%	87.3%	30.0%	478.49
Complex Read	100	10,662	27.9%	45.0%	31.5%	477.69
	200	16,055	49.7%	76.4%	30.6%	403.80
	300	16,918	63.2%	84.2%	30.3%	343.61

Table 2. Scalability test for Medium and Complex Reads.

Appendix 3: Glossary

CF: Cluster caching facility, used to facilitate global locking and buffer pool management in a DB2 pureScale environment.

DB2 pureScale: This IBM DB2 pureScale feature leverages proven technology from the DB2 for IBM z/OS® data-sharing architecture to bring the active-active shared-disk technology to open systems. The technology is primarily aimed at online transaction processing (OLTP) scale-out clusters.

PureFlex: IBM PureFlex System is built on elements of the IBM Flex System. IBM Flex System includes compute, storage, systems management and networking components that are pre-configured and pre-integrated to meet different IT infrastructure needs.

QDR: Quad data rate, referring to the throughput of the InfiniBand adapters.

SSD: Solid-state drive

For more information

To learn more about InfoSphere Master Data Management, please contact your IBM representative or IBM Business Partner, or visit the following resources:

- Implementing InfoSphere MDM for high performance (best practices): **ibm.com**/developerworks/data/bestpractices/infosphere/mdmhighperformance/index.html
- Achieving high availability and scalability with IBM InfoSphere MDM (best practices): https://www.ibm.com/developerworks/data/bestpractices/infosphere/mdmsha/index.html
- Monitoring and tuning InfoSphere Master Data Management, Part 1: Set goals and tune each layer of the infrastructure: **ibm.com**/developerworks/data/library/techarticle/dm-1010mdmservertuning1/index.html
- Monitoring and tuning InfoSphere Master Data Management, Part 2: Monitor the DB2 layer and learn about different monitoring tools: **ibm.com**/developerworks/data/library/techarticle/dm-1010mdmservertuning2/index.html
- InfoSphere MDM Best Practices Portal on IBM developerWorks®: **ibm.com**/developerworks/data/bestpractices/mdm/index.html
- DB2 pureScale on developerWorks: **ibm.com**/developerworks/data/products/db2/purescale
- Best practices for DB2 pureScale environments: http://ibm.co/1acxKvw
- IBM Redbooks®: "IBM PureFlex System and IBM Flex System Products and Technology": www.redbooks.ibm.com/abstracts/sg247984.html
- IBM Redbooks: "Implementing the IBM Storwize V7000 V6.3": www.redbooks.ibm.com/abstracts/sg247938.html
- "High Performance Computing and Master Data Management," an Oracle white paper: http://ebookbrowsee.net/high-perform-computing-mdm-1396935-pdf-d243290221
- Gartner. "Magic Quadrant for Master Data Management of Customer Data Solutions." October 14, 2013. www.gartner.com/reprints/ibm-v8-v7?id=1-1M60DM7&ct=131024&st=sb

Additionally, IBM Global Financing can help you acquire the software capabilities that your business needs in the most cost-effective and strategic way possible. We'll partner with credit-qualified clients to customize a financing solution to suit your business and development goals, enable effective cash management, and improve your total cost of ownership. Fund your critical IT investment and propel your business forward with IBM Global Financing. For more information, visit: **ibm.com**/financing

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Actual available storage capacity may be reported for both uncompressed and compressed data and will vary and may be less than stated.

[1] Oracle. "High Performance Computing and Master Data Management." June 2012. http://ebookbrowsee.net/high-perform-computing-mdm-1396935-pdf-d243290221

[2] Gartner. "Magic Quadrant for Master Data Management of Customer Data Solutions." October 14, 2013. www.gartner.com/reprints/ibm-v8-v7?id=1-1M60DM7&ct=131024&st=sb

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
November 2013

This document is current as of the initial date of publication and may be changed by IBM at any time.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary. It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

Using workload specifications that are defined in an Oracle white paper, these tests demonstrate that the IBM InfoSphere MDM solution exhibits superior performance and scalability and is the best of its breed.

Please Recycle

IMW14725-USEN-01