Reuters Market Data System

RMDS 6.3

Performance Test Results on IBM BladeCenter HS22

DISCLAIMER: The test results and recommendations contained in this report are made available for informational purposes only. By issuing this report, IBM does not guarantee similar performance results. All information contained herein is provided on an "AS-IS" BASIS WITHOUT WARRANTY OF ANY KIND. Obtaining repeatable, measurable performance results requires a controlled environment with specific hardware, software, network, and configuration in an isolated system. Adjusting any single element may yield different results. Additionally, test results at the component level may not be indicative of system level performance, or vice versa. IBM is pleased to provide this report for our customers. We appreciate that each organization has unique requirements, and therefore may find this information insufficient for its needs. Customers wishing to obtain custom analysis for their systems are encouraged to contact their local IBM representative.

Issue 2.1

Date of Issue: 12 01 2009

Contents

IBM, by publishing this document, does not guarantee that any information contained herein is and will remain accurate or that use of the information will ensure correct and faultless operation of the relevant service or equipment. IBM, its agents and employees, shall not be held liable to or through any user for any loss or damage whatsoever resulting from reliance on the information contained herein.

This document contains information proprietary to IBM and may not be reproduced, disclosed, or used in whole or part without the express written permission of IBM.

Any Software, including but not limited to, the code, screen, structure, sequence, and organization thereof, and Documentation are protected by national copyright laws and international treaty provisions. This manual is subject to U.S. and other national export regulations.

1 General Information

1.1 Objective

The objective of this document is to report the performance test results for RMDS 6.0, for a particular hardware and software platform. The test procedures are described in *Reuters RMDS 6.0 Performance Test Procedures and Results* document.

The goal of these tests is to measure throughput and latency through RMDS 6.0 infrastructure components, specifically the Point-to-Pont Server (P2PS) and Source Distributor. The tests are grouped into two categories:

- Update throughput using RSSL/RWF data (see 3.1)
- End-to-end RSSL/RWF latency using embedded timestamp (see 3.2)

1.1.1 Results Summary

- **End-to-end latency test:** 1.60 million updates per second with a mean latency on less than 1 millisecond
- **Source Distributor throughput:** 1.60 million updates per second
- **P2PS (no fan-out) throughput:** 1.60 million updates per second
- **P2PS (producer 50/50) throughput:** 4.5 million updates per second

1.2 Testing Methodology

For throughput testing, the ***sink_driven_src*** utility was used to generate update traffic, and the ***rmdstestclient*** utility was used to consume the updates. Level 1 data was used, with a Marketfeed (MF) update size of 140 bytes, and an equivalent Reuters Wire Format (RWF) update size of 74 bytes. Tests with no fan-out of updates used a 100,000 item watchlist. The infrastructure is tuned for maximum throughput, and the update rate was increased until the CPU limit was reached with no errors reported. Where needed, and as noted, multiple Source Distributors or multiple P2PSs were used to create the load necessary to measure the component under test.

The embedded timestamp approach was used to calculate end-to-end latency for Level 1 (Quotes and Trades) data. RMDS 6.0 end-to-end update latency is measured by using ***sink_driven_src*** as the publisher and ***rmdstestclient*** as the subscriber.

In the embedded timestamp approach, the publisher embeds timestamps into selected updates which the subscriber uses for latency calculations. In this scenario, the publisher and subscriber must be running on the same node for accurate timestamps.

1.3 Software Versions

1.3.1 RMDS

src_dist ver. mdh6.3.2
p2ps ver. p2ps6.3.4
rrcp as included in p2ps6.3.4

1.3.2 RMDS Test Tool

sink_driven_src (from MDH load above)
rmdstestclient (from P2PS load above)

1.3.3 Operating Systems

- Red Hat (RHEL 5.3 64 bit), Linux kernel 2.6.18-128.el5
- Chelsio Communications Linux cxgb3 version 1.3.0.23

1.4 Hardware

The performance tests were performed on a single IBM BladeCenter-H with Chelsio 10GbE adapter (TCP Offload Engine enabled) with the following components:

1.4.1 Compute nodes

5 IBM BladeCenter Server HS22 (7870) blades.
Each has:
  2 QC Intel Xeon X5570 processors (2.93 GHz, Turbo off);
  6 4GB 1333 MHz DDR III SDRAM;
  143GB HDD;
  2 integrated Broadcom 1GbE controllers;
  1 dual port Chelsio 10GbE Mezzanine Expansion Card (S320EM-BCH);

1.4.2 Blade chassis

1 BC-H (8852) which contains:
  1 Advanced Management Module;
  2 Power Modules;
  1 Cisco 3012 1GbE Switch Module (I/O Bay 1);
  2 Blade Network Technology (BNT) 6-port 10GbE Switch Modules (I/O Bays 7 & 9);
  5 HS22 (as mentioned above) in Blade Bays 1 to 5.

1.4.3 Network

Each blade is on 3 networks:
  1 1Gigabit Ethernet (GbE) network solely for management purposes and
  2 10Gigabit Ethernet (10GbE) networks for low latency and high throughput RMDS communications. TCP/IP Offload Engine (TOE) enabled.

Port 1 of the built-in dual-port1GbE is connected to Cisco 1GbE switch module.
Port 1 of the dual-port Chelsio 10GbE card is connected to BNT 10GbE switch module.
Port 2 of the dual-port Chelsio10GbE card is connected to another BNT 10GbE switch module.

2 Preparation for Performance Test

2.1 Network

All the performance tests were run where the machines were connected to a private network via 10 Gbps switches. All the network cards and switch ports were set to Auto Negotiate.

2.2 Hardware

All RMDS components were run on the same class of machine.

2.3 Operating System Configuration

Earlier tests have shown that the value chosen for ticks per second (tps) on the test application machine has a significant impact on latency measurement. Accordingly, a tps value of 1000 was used in these tests.

2.3.1 TCP and UDP Buffers

Any settings changed from the defaults are noted below:

Step	Procedure		
1	**OS**	**Enter the following lines in system file noted**	**System File**
	Linux	net.core.wmem_max = 8388608	***/etc/sysctl.conf***
		net.core.wmem_default = 8388608	
		net.core.rmem_max = 8388608	
		net.core.rmem_default = 8388608	
		net.ipv4.tcp_rmem = 4096 8388608 16777216	
		net.ipv4.tcp_wmem = 4096 8388608 16777216	
		net.ipv4.tcp_mem = 4096 8388608 16777216	

2.4 RMDS Configuration

The configuration template ***rmds.cnf.template*** was customized for the tests.

Config File	Description	Path
rmds.cnf.template	Configuration file	***./config***

2.5 Miscellaneous Notes

Any other significant deviations from the standard test procedures, or clarifications, are noted below (such as number/type of machines used, CPU binding policy, etc.):

Test	Deviation	Comments
All	CPU Binding	Linux *irqbalance* was disabled and all interrupts were handled by CPU 0. Linux *taskset* command was used to bind RMDS processes. The *rrcpd* daemons were bound to CPUs 1 to 4, the *src_dist* processes were bound to CPUs 5 to 7, and so were the p2ps processes (running on a separate blade).The *sink_driven_src* process was bound to CPU 1. *rmdstestclient* was bound to CPU 2.

All	TOE Enabled
	(TCP Offload Engine)
		toe.toe0_tom.max_host_sndbuf=131072
		toe.toe0_tom.rx_credit_thres=131072

3 Detailed Results

3.1 RSSL/RWF Update Throughput

- All the throughput numbers quoted here are for Level 1 data.
- The data file used in these tests has 1 update, with an update (data, not including header) size of 74 bytes in RWF.
- All of the tests with no fan-out used 100,000 item watchlist.
- In most of the throughput tests the individual processes were bound to particular CPU(s).
- ***sink_driven_src*** and ***rmdstestclient*** were used as the publisher and consumer of data.
- In some Source Distributor tests, two P2PSs were used to create sufficient load.

3.1.1 Standalone Source Distributor

Configuration Option	Transport	Max Throughput	Comments
Cache Disabled	RRCP	1.60 million updates per second	One P2PS and one src_Dist used

3.1.2 P2PS/LAN

Configuration Option	Mounts : Commonality	Transport	Max Throughput	Comments
Cache Disabled	No fan-out	RRCP	1.60 Million updates per second	One P2PS and one src_Dist used
Cache Disabled	100 mounts; Producer 50/50	RRCP	4.50 Million updates per second	One P2PS and one src_Dist used

3.2 End-to-End RSSL/RWF Latency

Latency is defined as the time for a data item to propagate through one or more RMDS components. “End to end” latency is defined as the delta between the time an update is posted by the publisher application to its API and the time the same update is received by the consuming application from its API, i.e. it includes both the latency contribution from the API and the core infrastructure components.

NOTES:

- Caching was disabled in both the Source Distributor and the P2PS during these tests.
- Optimized binaries of the RMDS infrastructure components were used.
- NTP was disabled on the tools node, as any drifts in time will affect the reported latency.
- Tests were run with 100,000 item watchlist and RWF data update size of 74 bytes [Data file (***sample.xml***) was used]. The update size is equivalent to a 140-byte IDN update.
- Latency tests were run at each update rate for at least 5 minutes, up to the maximum sustainable update rate for a given setup.
- Decode of data was turned on in these tests.

3.2.1 RRCP Backbone Results

Update Rate [74-byte RWF messages]	Mean Latency (microsec)	Std Deviation (microsec)	Maximum Latency (microsec)	Minimum Latency (microsec)	Number of Latency Points
50000	173.62	11.92	432	155	3000
100000	201.47	24.78	491	160	3000
150000	214.36	34.65	722	159	3000
200000	222.15	24.56	337	159	3000
250000	234.22	30.06	366	162	3000
300000	247.16	37.01	397	159	3000
350000	258.19	46.42	504	165	3000
400000	275.82	61.87	781	162	3000
450000	286.31	67.23	756	165	3000
500000	291.77	70.09	864	160	3000
550000	308.06	83.65	866	169	3000
600000	322.57	101.48	1499	181	3000
650000	338.2	120.28	1109	161	3000
700000	353.66	131.04	1427	182	3000
750000	367.98	139.01	1491	175	3000
800000	382.18	153.42	1501	171	3000
850000	400.1	176	1895	183	3000
900000	424.19	194.87	1681	171	3000
950000	445.89	230.74	2133	167	3000
1000000	460.46	240.05	2277	167	3000
1050000	468.45	247.35	2458	167	3000
1100000	467.28	269.74	2665	164	3000
1150000	487.95	302.68	2802	172	3000
1200000	508.35	313.11	2635	168	3000
1250000	536.19	330.76	2969	167	3000
1300000	583.68	368.43	3293	185	3000
1350000	592.43	369.8	3148	164	3000
1400000	634.16	443.33	3214	164	3000
1450000	713.44	506.38	3712	170	3000
1500000	739.22	554.39	3807	178	3000
1550000	802.66	627.63	7047	180	3040
1600000	913.56	1565.98	21927	170	3000