z/VM Workshop
Oracle on IBM Z Performance Tips/Update

David Simpson
IBM Z Systems Oracle Certified DBA / IT Specialist
simpson.dave@us.ibm.com

Agenda

Oracle on IBM Z Performance Tips

– Oracle Update

– Techcombank – LinuxONE Emperor II / DS8886 / LUNs with FS900 storage (managed SVC)

– US Government – Rockhopper II / FS9110

– Large Healthcare POC - Emperor II / FS9150 and FS900

– Flash Systems Demo (Poughkeepsie Benchmark Center)

Copyright and Trademark Information

- For IBM – can be found at http://www.ibm.com/legal/us/en/copytrade.shtml
- For Oracle – can be found at
  http://www.oracle.com/us/legal/index.html

- Oracle compliance:
  https://www.linkedin.com/feed/update/urn:li:activity:6534481473284694016

- Any performance results/observations in this presentation are purely for education and planning purposes. No Test results should be construed as indicative of any particular customer workload or benchmark result.

Popular Databases on LinuxONE

- **Relational Databases**

ORACLE Platinum Partner

MySQL

MariaDB

PostgreSQL

IBM Db2

- **Non Relational Databases**

redis

SCYLLA

mongoDB

cassandra

RethinkDB

CouchDB

IBM Cloudant®

APACHE GEODE

neo4j

©2018 IBM Corporation

Oracle/MongoDB Hands-on WildFire Class:

Topics Include:

IBM LinuxONE Oracle Solution Overview
IBM LinuxONE Overview
IBM z/VM Hypervisor Overview
Linux on IBM LinuxONE Systems Overview
Oracle Databases on LinuxONE Systems Overview
z/VM Labs Customization and Configuration
Linux Distribution and Installation requirements
Linux Installation lab
Storage provisioning and configuration on Linux
Logical Volume management (LVM2) Setup
Oracle Database Installation Lab
FCP/Multipathing
Oracle Database Migration
Tuning and Performance Aspects of Oracle database running on z/VM with Linux
Monitoring Tools and Performance
Data Collection for z/VM, Linux and Oracle database
z/VM Performance Toolkit, Linux and Oracle monitoring tools

MongoDB installation lab
Oracle and MongoDB on LinuxONE data integration Demo
Next Steps, Successful POCs, Tools and Resources Available

Let us know (email) if interested.

Oracle/MongoDB Demo:

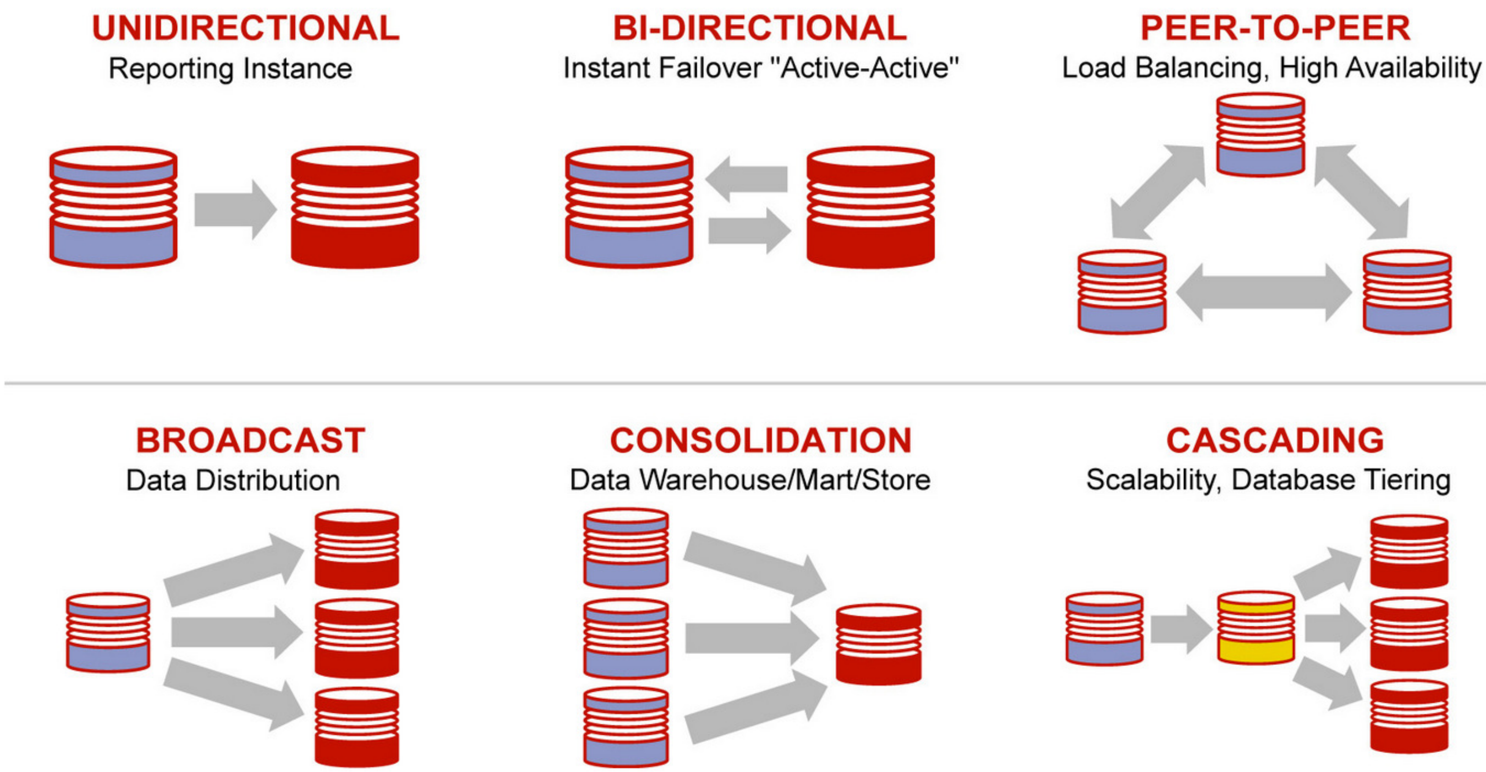
Enterprise OneView for MongoDb - Integrating Data from Oracle on LinuxONE, CICS from z/OS and node.js application onto MongoDB on LinuxONE

Enterprises are data driven and they do have multiple applications running on heterogeneous platforms like z/OS, LinuxONE systems and x86 platforms. These applications generate humongous data and still the enterprises struggle to see the relevant information to support their business operations. In this age the right data at the right time is the key ingredient for the success of their businesses.

In this live system demo, we will show how data from multiple sources of variable structured data can be integrated and shown in a single view, in a MongoDB running on a LinuxONE server. Please block 45 minutes to fully experience this demo.

Sam Amsavelu: Let us know (email) if interested.

GoldenGate – Oracle Data Integration Product



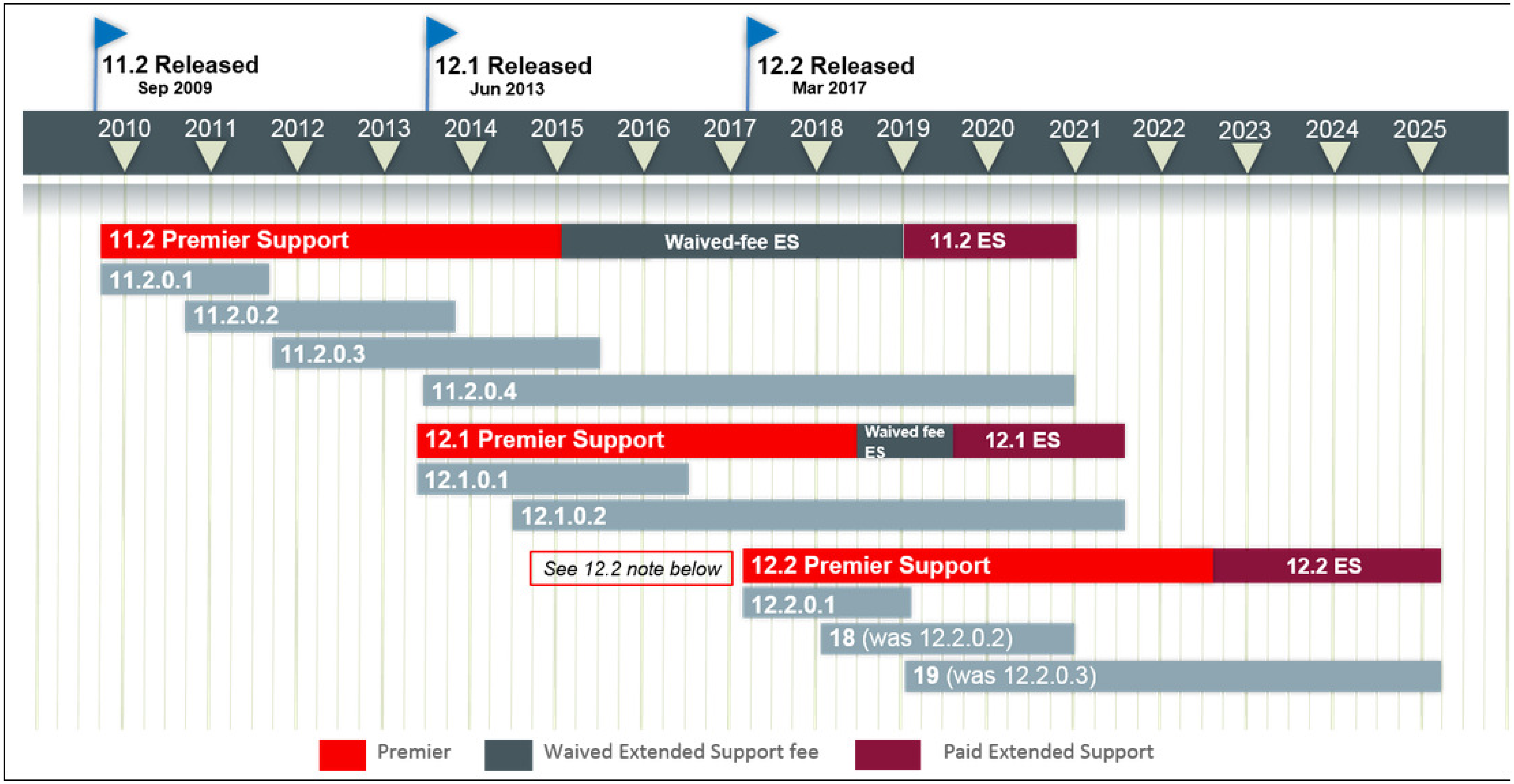
Source: Oracle® GoldenGate Administering Oracle GoldenGate for Windows and UNIX

Recent Oracle 19c Announcements

IBM

Linux x86-64	25-Apr-2019	23-Jul-2018	1-Mar-2017	22-Jul-2014	25-Jun-2013	27-Aug-2013
Oracle Solaris SPARC (64-bit)	26-Apr-2019	30-Jul-2018	1-Mar-2017	22-Jul-2014	25-Jun-2013	29-Aug-2013
Oracle Solaris x86-64 (64-bit)	Q2CY19	6-Aug-2018	1-Mar-2017	22-Jul-2014	25-Jun-2013	29-Aug-2013
Microsoft Windows x64 (64-bit)	Q2CY19	21-Aug-2018	16-Mar-2017	25-Sep-2014	9-Jul-2013	25-Oct-2013
HP-UX Itanium[7]	28-May-2019	20-Nov -2018	13-Apr-2017	14-Nov-2014	9-Jan-2014	10-Oct-2013
HP-UX PA-RISC (64-bit) *See footnote 6 below*	*Platform desupported* [6]	*Platform desupported* [6]	*Platform desupported* [6]	*Platform desupported* [6]	*Platform desupported* [6]	2-Jan-2014
IBM AIX on POWER Systems	28-May-2019	20-Nov -2018	13-Apr-2017	14-Nov-2014	9-Jan-2014	10-Oct-2013
IBM Linux on System z	06-June-2019	20-Nov -2018	6-Jun-2017	14-Nov-2014	9-Jan-2014	9-Jan-2014
Microsoft Windows (32-bit)	*Not planned*	*Not planned*	*Not planned*	*Not planned*	*Not planned*	25-Oct-2013
Platform	**19**	**18**	**12.2.0.1**	**12.1.0.2**[9]	**12.1.0.1**[2,10]	**11.2.0.4**[8]

Oracle's new Release roadmap (742060.1)



Release 12.2: New releases will be annual and the version will be the last two digits of the release year. The release originally planned as 12.2.0.2 will now be release 18, and the release originally planned as 12.2.0.3 will be release 19. Releases 18 and 19 will be treated as under the umbrella of 12.2 for Lifetime Support purposes. The current plan is for Oracle Database 19 to be the last release for 12.2.

Recent Oracle 19c Announcements

IBM

Release	Premier Support Ends	Extended Support Ends	Patching Ends	Notes and Exceptions*
19c	**31-Mar-2023**	**31-Mar-2026**	**31-Mar 2026**	(a.k.a. Long Term Support Release for 12.2) Extended Support fees will be required beginning Apr-2023.
18c	**TBD**	**N/A**	**TBD**	(a.k.a. Annual Release for 12.2) Patching will end 24 months after last on-premises platform releases 19c
12.2.0.1	**20-Nov-2020**	**N/A**	**20-Nov-2020**	Patching for 12.2.0.1 will end Nov 20, 2020. There is no Extended Support for 12.2.0.1.
12.1.0.2	**31-Jul-2018**	**31-Jul-2021**	**31-Jul-2021**	Extended Support fees waived through July 31, 2019. Beginning Aug 1, 2019 an ES service contract is required. For E-Business waiver, see *Extended Support Fee Waiver for Oracle Da Suite (Doc ID 2522948.1)*
12.1.0.1	**31-Aug-2016**	**N/A**	**31-Aug-2016**	Patching has ended for this release. There is no extended support for 12.1.0.1
11.2.0.4	**1-Jan-2015**	**31-Dec-2020**	**31-Dec-2020**	Premier Support ended 01-Jan-2015 but ES fees were waived through 31-Dec-2018. An EXS service contract is required starting 1 E-business waiver, see *Extended Support Fee Waiver for Oracle Database 12.1 and 11.2 for Oracle E-Business Suite (Doc ID 2522*

Oracle's Patch Set Support Update

Enhancement to existing Support for Linux on z Systems Servers

Patch Set Update – Linux on z

- **Policy Change on Patch Set Update (PSU)**
  - Beginning with the October 2009 Critical Patch Update release, Oracle will now deliver Patch Set Updates for all platforms on the release date including Linux on z.
- **What is a PSU and when is it provided?**
  - PSU is a bundle of patches Oracle recommends to apply. It consists of CPU, Generic patch bundle, RAC patch bundle and Data Guard patch bundles
  - Quarterly released
- **Benefit for Linux on z Customers**
  - Verified and tested before provided to the customer
  - Easy database maintenance
  - Recommended patches now also available for Linux on z
  - Reduces problem situation and downtime.
- **What About Critical Patch Updates (CPUs)?**
  - In the future single Critical Patch Updates are only available on request via service reqwuest (SR)

Raimund Reng, Oracle Support – September 2009

19c Oracle Autonomous Database: New Features

IBM

1. **Stability** –19c is the terminal release of Oracle 12c (just like 11.2.0.4 is terminal release of 11gR2 Version)
2. Automatic Indexing – Indexes can be created automatically
3. Real-time Stats + Stats Only Queries – Statistics can go stale between execution of DBMS_STATS statistics gathering jobs
4. Data-guard DML Redirect – Active Data Guard DML Redirection allows for incidental DML to be issued on an Active Data Guard standby database
5. Partial JSON Update support - You can now update a JSON document declaratively using the new SQL function json_mergepatch.
6. Schema-only Oracle accounts - Unused and rarely accessed database user accounts with administrative privileges can now become schema-(passwordless) only accounts.
7. DB REST API - perform DB operations using REST API (like creating a database, creating a user)
8. Partitioned Hybrid Tables - A hybrid partitioned table is a partitioned table in which some partitions reside in the database and some partitions reside outside the database
9. Social Sign-In Authentication- Social Sign-In preconfigured authentication scheme supports authentication with Google, Facebook, and other social network that supports OpenIDConnect or OAuth2 standards
10. DBA_REGISTRY_BACKPORTS - bugs fixed (data dictionary) by the patches applied to database

19c Oracle Autonomous Database - Indexes

- Fully automated process. Oracle will identify candidate indexes, verify effectiveness, perform online validations and t implement indexes where appropriate.
- DBA does not need to do anything.
- Oracle internally picks the candidate indexes and validates the index or indexes.

21.7.3 Enabling Automatic Indexing

Automatic indexing is disabled by default in an Oracle database. To enable automatic indexing, set the AUTO_IMPLEMENT_INDEXES initialization parameter to the Oracle database release number, for example, 19.1. You can disable automatic indexing by setting the AUTO_IMPLEMENT_INDEXES initialization parameter to NONE.

Source: https://docs-stage.oracle.com/en/database/oracle/oracle-database/19/admin/managing-indexes.html#GUID-6E31777C-3BE3-4510-90D5-C715644E00CB

Integrating Spectrum Scale Snapshots with Oracle Recovery Manager

Oracle Database Backup via snapshots

- *Oracle recently announced the official support for third-party snapshot technologies to create crash-consistent images*
  Source: My Oracle Support (MOS) note 604683.1
- *IBM Spectrum Scale snapshots can be used to reduce the time and administrative complexity required to create and restore Oracle databases, as well as reducing storage requirements of the backup target.*
- Oracle RMAN incorporates DB Block corruption checks
- http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102757 (TechDocs step by step guide)

Spectrum Scale snapshots

- IBM Spectrum Scale (GPFS) snapshots are a point in time copy of files and directories. If a block is changed or updated after creating a snapshot, there are two methods to preserve the old block contents.
- **Copy on Write (COW):**
- In this method, the block is pushed to the snapshot before writing new block contents.
- **Redirect on write:**
  In this method, a new block is allocated and written. The file metadata is changed such that the snapshot data points to the old block and file data points to the newly allocated block.

IBM Spectrum Scale 5.0: Certification with Oracle DB (GPFS)

IBM

Test Case:

- Seven snapshots representing everyday. Sunday, a level-0 (FULL) backup is taken. Monday to Saturday incremental backup and merge is done.
- The workload we ran against the database inserts one million rows; updates one million rows, update indexes.

Advantages of GPFS backup scheme:

- Daily backup window is reduced. Only changed blocks from previous day backed up. up.
- Changed blocks since previous incremental backup, are pushed into snapshot (**Block Push to snap**).
- For restore, since we always have a current level-0 backup copy, no need to merge, apply lots of archive logs during restore (faster RTO)
- Recovery time fast and customers can meet recovery time objective (RTO) goals.
- The table below shows amount of disk storage saved, **The original DB Size is 118 GB**

SnapShots	Backup Size	Incr Backup Size	Block Push to snap
0	118 GB	None	None
1 (snp_0717@16H56M)	125 GB	50 GB	6.8 GB
2 (snp_071717H19M @)	131 GB	63 GB	6.1 GB
3 (snp_0717@17H35M)	137 GB	54 GB	6.1 GB
4 (snp_0717@18H01M)	144 GB	60 GB	6.2 GB
5 (snp_0717@18H19M)	150 GB	63 GB	6.2 GB
6 (snp_0717@19H01M)	157 GB	62 GB	6.3 GB
7 (snp_0717@19H29M)	164 GB	60 GB	6.1 GB

Agenda

Oracle on IBM Z Performance Tips

– Oracle Update

– Techcombank – LinuxONE Emperor II / DS8886 / LUNs with FS900 storage (managed SVC)

– US Government – Rockhopper II / FS9110

– Large Healthcare POC - Emperor II / FS9150 and FS900

– Flash Systems Demo (Poughkeepsie Benchmark Center)

Techcommbank Customer Case

Customer Experiences: Techcombank LinuxONE – Emperor II / DS8886

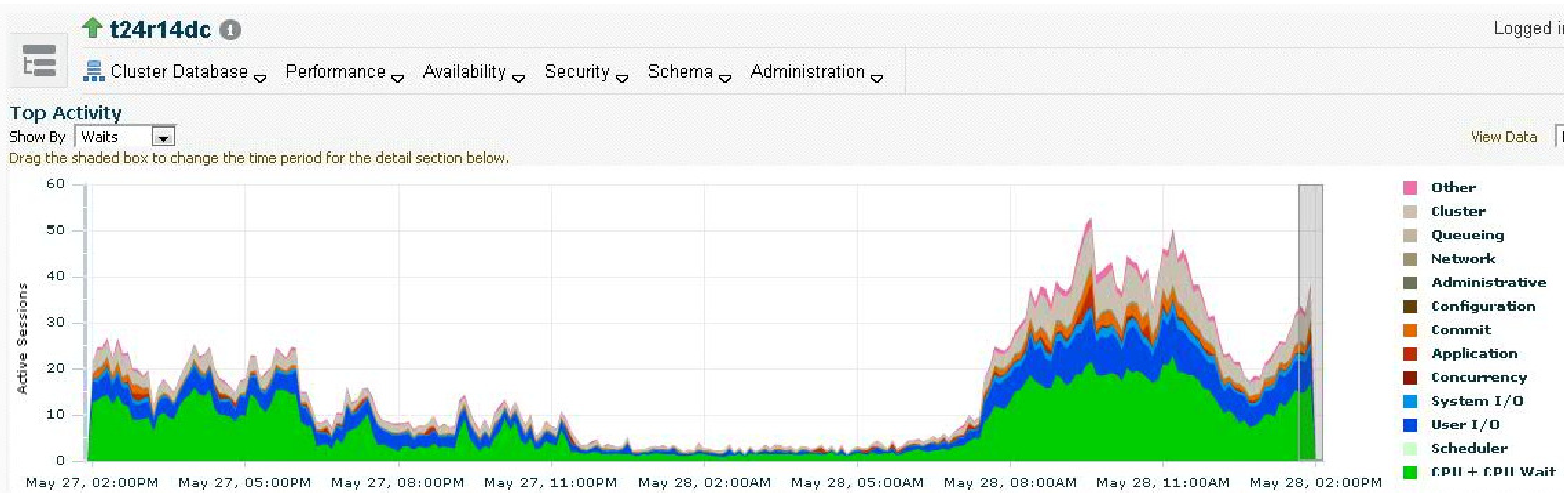
– Customer in Hanoi, Vietnam – key reference account for LinuxONE running Oracle

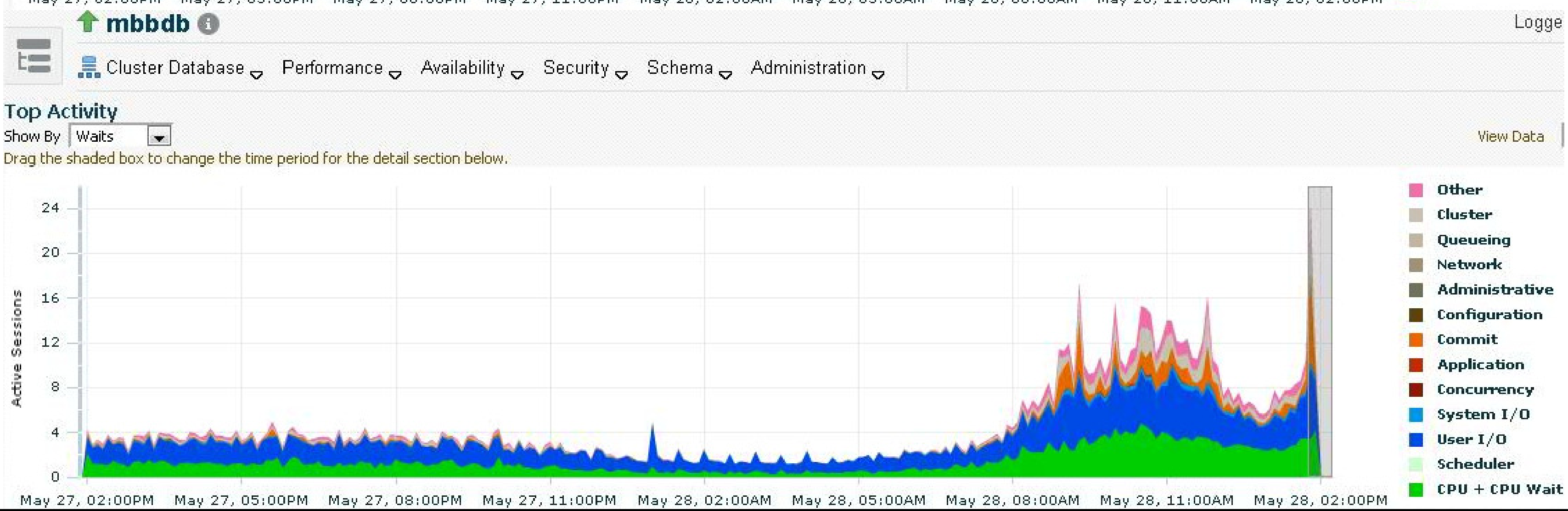
– Temenos Banking Application – benchmarked in Montpellier (MOP).

– System “pre-built” in Singapore then shipped to Hanoi (Code 20).

– Saved time prestaging getting firmware updates,
- z/VM installed with some pre IP/network information ahead of time.
- Some documentation created ahead of time for easier transition.

Customer Case TechComm Bank : “gc cr block lost” & Cluster Waits





Reallocate unused Large Page Memory

- Linux Guest had **413 GB** Linux virtual memory assigned
- Linux processes were running out of memory with Linux swap during gc cr block lost events

May 30 07:57:23 dc-core-db-01 kernel: **qethqoat: page allocation failure**: order:5, mode:0x10c0d0
May 30 07:57:23 dc-core-db-01 kernel: **lowmem_reserve**[]: 0 0 0
May 30 07:57:23 dc-core-db-01 kernel: 21 pages in swap cache
May 30 07:57:23 dc-core-db-01 kernel: Swap cache stats: add 540, delete 519, find 0/0
May 30 07:57:23 dc-core-db-01 kernel: Free swap = 33552268kB
May 30 07:57:23 dc-core-db-01 kernel: Total swap = 33554428kB

- Allocated **245,760 MB** of Linux large pages and increased Oracle SGA memory

SGA Target Size (M)	SGA Size Factor	Est DB Time (s)	Est Physical Reads
134,400	0.94	2,827,684	1,329,325,293
143,360	1.00	2,804,128	1,286,111,932
152,320	1.06	2,792,632	1,286,111,932
161,280	1.13	2,764,310	1,241,612,459
170,240	1.19	2,741,036	1,206,501,603
179,200	1.25	2,720,566	1,174,606,027
188,160	1.31	2,702,901	1,144,639,619
197,120	1.38	2,695,610	1,125,347,941
206,080	1.44	2,684,113	1,125,347,941
215,040	1.50	2,673,738	1,108,242,652
224,000	1.56	2,663,924	1,091,394,585

```
Finding 3: Undersized SGA
Impact is .87 active sessions, 9.17% of total activity.
-------------------------------------------------------
The SGA was inadequately sized, causing additional I/O or hard parses.
The value of parameter "sga_target" was "143360 M" during the analysis period.

   Recommendation 1: Database Configuration
   Estimated benefit is .52 active sessions, 5.48% of total activity.
   -------------------------------------------------------------------
   Action
      Increase the size of the SGA by setting the parameter "sga_target" to
      197120 M.
```

Additional Linux Kernel Parameters for Oracle RAC/Network

- Suggested kernel parameters in **/etc/sysctl.conf**.
- Prevent SYN packet loss, **net.ipv4.tcp_max_syn_backlog = 10000**. Must set **net.core.somaxconn** as well.
- Increase the CPU input packet queue length from the default of 1000 (netdev_max_backlog)
- **Strict Reverse Path filtering**. Strict mode is the default is to prevent IP spoofing from Distributed Denial-of-service attacks. Having strict mode enabled on private interconnect of an Oracle RAC database cluster may cause disruption of interconnect communication.
- Per Troubleshooting gc block lost and Poor Network Performance in Oracle **Doc ID 563566.1** increasing **wmem** parameters can reduce block loss.

Business Justification:

- These parameter changes are recommended in various best practice documents (see Appendix) will help avoid network related problems.

```
#net.ipv4.ip_local_port_range = 9000 65500 (duplicate)
# Per IBM Network Performance Guide
net.ipv4.tcp_max_syn_backlog = 10000
net.core.somaxconn = 1024
net.core.netdev_max_backlog = 25000
# per Red Hat Best Practice Guide
net.ipv4.conf.conf.enccw0/0/a800.rp_filter.rp_filter = 2
net.ipv4.conf.conf.enccw0/0/a900.rp_filter.rp_filter = 2
# increase wmem_default  & max per MOS Note:  563566.1
net.core.wmem_default = 4194304
net.core.wmem_max = 4194304
```

Kernel Change Summary:

net.ipv4.tcp_max_syn_backlog = 512 --> 10000
net.core.netdev_max_backlog = 1000 --> 25000
net.core.somaxconn = 128 --> 1024
net.core.wmem_max = 1048576 --> 4194304
net.core.wmem_default = 262144 -> 4194304
net.ipv4.conf.enccw0/0/a800.rp_filter = 1-> 2
net.ipv4.conf.enccw0/0/a900.rp_filter = 1 ->2

Network: Oracle Parameters for Oracle RAC:

1. **Description & Objective:**
   - Various Red Hat Network tuning parameters were suggested to be increased per Oracle, IBM and Linux Distro best practices.

Old:
net.ipv4.tcp_max_syn_backlog = 512
net.core.somaxconn = 128
New:
net.ipv4.tcp_max_syn_backlog = 10000
net.core.somaxconn = 1024

   - Observed evidence in the OS Watcher logs that various network sockets were over flowing.

From netstat on dc-01 - Jul 13th (Before)

1082 invalid SYN cookies received

11897 times the listen queue of a socket overflowed

11897 SYNs to LISTEN sockets dropped

2. **Observations and actual result:**
   - No packet loss observed after 5 days.
   - Observed some invalid SYN cookies received errors – which we want to continue to monitor.
   - No socket overflow or dropped packets in the netstat logs so far since system restart

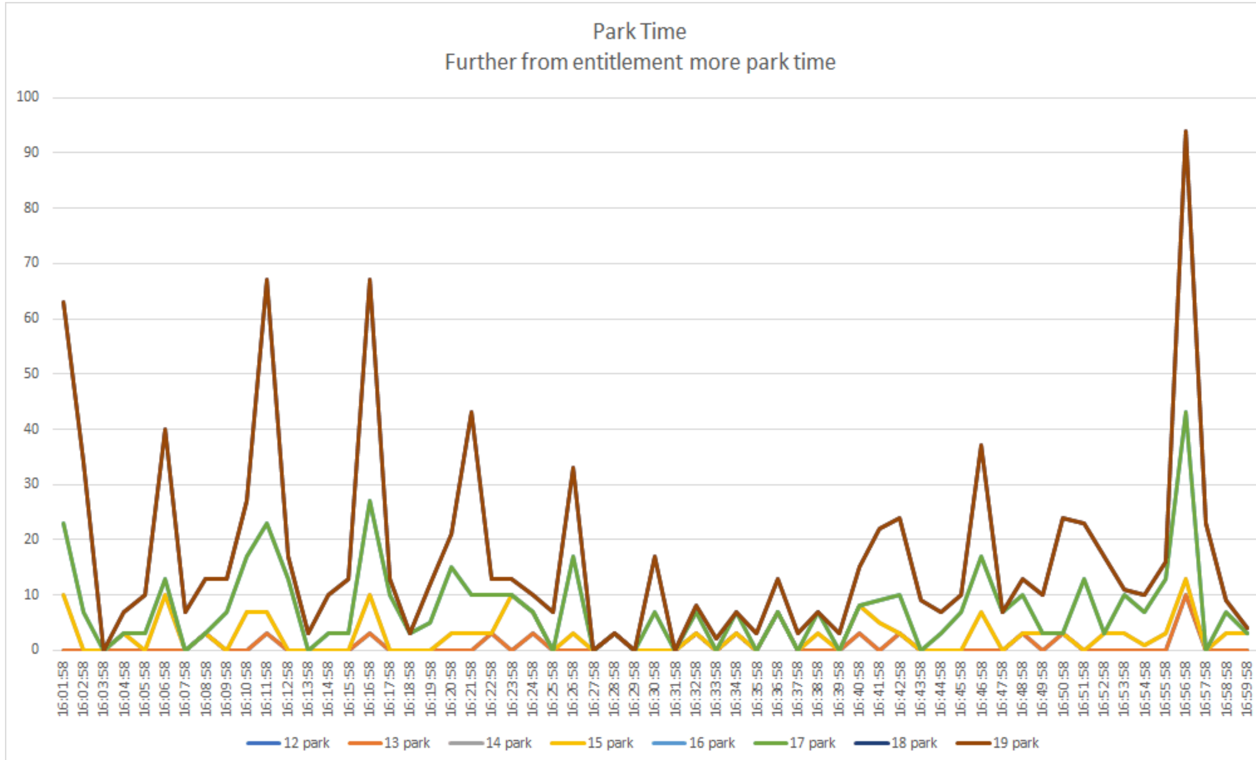
dc-core-db-01_netstat_18.07.26.1000.dat (After)
811 invalid SYN cookies received

3. **Analysis / Conclusion:** no packet loss, after memory and network parameter tuning.
If additional packet loss, per Reference: https://access.redhat.com/solutions/446963 and Oracle Support -> *Intermittent Slower Network and Connection Timeouts (Doc ID 1614134.1)* increase these network parameters even further.

LPAR Weights and Virtual CPU Allocations

- Each Linux defined with 20 virtual CPUs
- Only allowed 11 full time logical CPUs to run when system is busy.
- Best performance when configured with no more virtual CPUs than the sum of the logical vertical highs and vertical mediums, in this case **11-15 vCPUs**

Recommendation: Conduct a Capacity Planning Review based on real collected data and future growth plans for the system.



After: LPAR Weights and Virtual Status

IBM

1. **Description & Objective:**
   - Eliminate the system being dispatched on CPU Vertical lows leading to cpu contention
   - Configure Linux with no more virtual CPUs than the sum of the logical vertical highs and mediums, **11-15 vCPUs**.
   - Allow CPUs to do more meaningful database work

2. **Observations and actual result:**
   - CPU Wait went from 23% to 17% at a busier time period.
   - “Context switching” Diag 9C’s and Diag 44’s have gone from 11.8 / 20.8 K/s ➔ 2.6 / 2.9 K/s

07/16 (7:00-8:00).

DCCORE01 defined with 32 vCPUS with relative share of 100 mean CPU utilization 6.32 IFLs mean CPU wait 23%
issuing 11.8K DIAG x'9C's and 20.8K DIAG x'44's per second
Data discards and overflows were not detected.

Average Physical utilization on the CEC 9.76 IFLs
Mean system time 1.59 IFLs

DCCORE1 7/26 from 14:00-16:00

Average CEC PhysicalUtilization 1600

COREDB01 has 32 IFLs defined 16 are stopped with a relative share of 1600.

COREDB01's average CPU utilization 8 IFLS
average CPU Wait 17%

AVG diag x'44' = 2900/second
AVG diag x'9C'= 2600/second

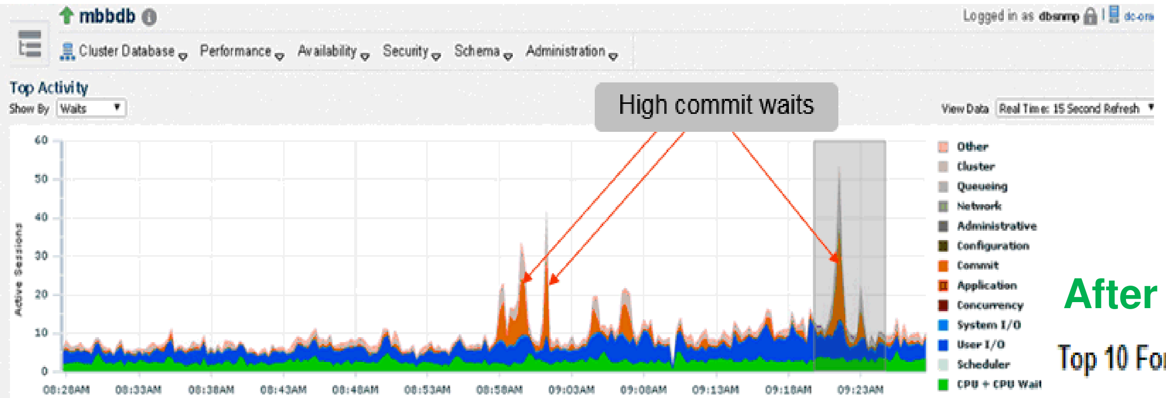
3. **Analysis & Conclusion:**

- LPAR weight recommendations helped more DB workload get completed.

I/O Performance: Mobile Banking Progress

IBM

Before moving redo log files to uncompressed disk:



Top 10 Foreground Events by Total Wait Time

Event	Waits	Total Wait Time (sec)	Wait Avg(ms)	% DB time	Wait Class
db file sequential read	6,212,435	5749	1	29.7	User I/O
DB CPU		5062.4		26.1	
log file sync	475,861	4020.8	8	20.7	Commit
reliable message	1,298,040	1473	1	7.6	Other
gc current block 2-way	751,650	363.8	0	1.9	Cluster
gc cr block 2-way	566,261	338.4	1	1.7	Cluster

After moving redo log files uncompressed disk:

Top 10 Foreground Events by Total Wait Time

Event	Waits	Total Wait Time (sec)	Wait Avg(ms)	% DB time	Wait Class
DB CPU		6468		33.1	
db file sequential read	2,926,400	5156.9	2	26.4	User I/O
reliable message	1,604,316	2418.1	2	12.4	Other
log file sync	511,468	998.2	2	5.1	Commit
gc current block 2-way	939,733	588.5	1	3.0	Cluster
gc cr block 2-way	676,792	573	1	2.9	Cluster
direct path read	27,151	430.8	16	2.2	User I/O
gc buffer busy acquire	108,743	213.2	2	1.1	Cluster
gc current grant busy	194,322	190.7	1	1.0	Cluster
gc current block busy	67,985	173.9	3	.9	Cluster

Redo Log Sync wait events went from 8ms to 2ms average

Case Study: Mobile Banking - SGA Target

- Other database running on **Linux Guest Mobile Banking**
  - **9,696 MB SGA** (PGA Target: 9,696)

- 156.94 GB Linux Guest

- Add 5GB to SGA? Candidate for Large pages?

SGA Target Size (M)	SGA Size Factor	Est DB Time (s)	Est Physical Reads	% Improvement
8,484	0.88	21,378,837	11,997,620,981	
9,696	1	17,646,582	7,299,154,944	
10,908	1.13	15,781,342	4,950,286,883	47.45
12,120	1.25	14,697,846	3,586,074,824	103.54
13,332	1.38	14,136,686	2,879,516,625	153.49
14,544	1.5	13,942,573	2,635,724,850	176.93
15,756	1.63	13,601,995	2,205,804,624	230.91
16,968	1.75	13,434,353	1,995,588,962	265.76

Memory Planning

IBM

- Plan for **ALL** Databases in sizings including Golden Gate
- Ensure databases come up with Large pages allocated (check Oracle alert log)

Oracle GoldenGate Memory Report:

```
---view report ET24DC
CACHEMGR virtual memory values (may have been adjusted)
CACHEPAGEOUTSIZE (default):                 8M
PROCESS VM AVAIL FROM OS (min):            128G
CACHESIZEMAX (strict force to disk):       96G
```

System	Parameter/DB	Original (MB)	Actual Memory
Large Pages Memory:			
DB1 Oracle SGA		143,360	197,120
DB 2 (Reporting) SGA		71,680	71,680
Linux Large Pages	vm.nr_hugepages	245,760	268,816
4K Memory:			
DB1 PGA usage		40,960	65,536
DB2 Reporting PGA		40,960	40,960
Oracle ASM (estimate)		2,048	2,560
User sessions Memory (DB1)	Connections*4.5MB	9,518	10,035
User sessions Memory (DB2)	Connections*4.5MB	450	900
GoldenGate			98,304
EM agent			16,998
Sub Totals		93,936	235,293
Total Memory		339,696	504,109
Memory + 10% overhead		373,665	
Linux Guest Memory	z/VM User Profile	438,952	472,783

Limit Network Traffic and High Availability

Recommend implementing Oracle service to run workload on one Linux guest node.
With failover to 2nd node for application affinity to help limit Cluster Waits.

Oracle Real Application Cluster across multiple LinuxONE machines for Higher Availability.

Wait Class	Waits	%Time -outs	Total Wait Time (s)	Avg wait (ms)	%DB time
Cluster	2,759,098	0	85,788	31.09	81.74
Concurrency	31,209	1	10,151	325.25	9.67
DB CPU			4,390		4.18
Other	919,399	90	1,394	1.52	1.33
User I/O	1,708,596	0	1,226	0.72	1.17

```
MERGE INTO F_TSA_STATUS USING DUAL ON (RECID = :RECID)
WHEN MATCHED THEN UPDATE SET XMLRECORD=XMLTYPE(:XMLRECORD, NULL, 1,
1)
WHEN NOT MATCHED THEN INSERT (XMLRECORD ,RECID)
VALUES(XMLTYPE(:XMLRECORD, NULL, 1, 1) ,:RECID)
  SELECT t.XMLRECORD.getClobVal() FROM F_OS_XML_CACHE t WHERE RECID
  =:RECID
      SELECT t.XMLRECORD.getClobVal() FROM F_OS_XML_CACHE t WHERE RECID
       SELECT t.XMLRECORD.getClobVal() FROM F OS TOKEN t WHERE RECID =:RECID
         SELECT t.XMLRECORD.getClobVal() FROM F LOCKING t WHERE RECID =:RECID
```

Agenda

Customer Experiences with LinuxONE and IBM Flash Systems:

– Techcombank – LinuxONE Emperor II / DS8886 / LUNs with FS900 storage (managed SVC)

– US Government – Rockhopper II / FS9110

– Large Healthcare POC - Emperor II / FS9150 and FS900

– Flash Systems Demo (Poughkeepsie Benchmark Center)

US Government - System Configuration

What systems did we configure including 16u Reserved Space?

– 3907 LR1 LinuxONE Rockhopper II: 16u reserved space (FC 617)

– Flash System 9110

– 2 Fiber Channel switches (NPIV), high availability

– Network switch (to simplify integration to client network)

– Small Intel Server (IBM xSeries) for Spectrum Control monitoring

Dynamic Partition Manager (DPM)

IBM

- DPM was really easy to use reconfigure LPARs, add cpu/memory – SIMPLE!
- Linux console access through a web browser was intuitive / easy to use.
- DPM does not currently support NVMe enclosure disks and ECKD storage devices

Filter Tasks Views

Select	Name	Status	Processors	Memory (GB)	Processor Utilization	Network Utilization	OS Na
	KVM1	Active	8	512.0	0 %	0 %	
	KVM2	Active	8	512.0	0 %	0 %	
	NSD1	Active	8	128.0	3 %	0 %	
	NSD2	Active	8	128.0	1 %	0 %	
	NSD3	Active	8	128.0	1 %	0 %	
	ORA1	Stopped	8	1,024.0			
	ORA2	Active	29	1,024.0	517 %	0 %	
	TEMS	Stopped	8	16.0			
	TEPS	Stopped	8	16.0			
	Util	Active	4	4.0	1 %	0 %	

Max Page Size: 500 Total: 10 Filtered: 10 Selected: 1

FS9150 (AF8) & FS9110 (AF7) Spec Sheet:

IBM

	AF8	AF7	V7000F
	8.2.0	8.2.0	8.1.0
4k read hit	2,400k	1,350k	1,400k
4k write hit	350k	270k	400k
4k random read	1,100k	560k	450k
4k random write	185k	95k	110k
4k 70/30 rw mix	500k	220k	220k
256k seq read (MB/s)	33,000	23,000	12,000
256k seq write (MB/s)	7,000	4,500	5,200

- FS9100 have NVMe flash drives with DRAID6 redundancy, compression & disk encryption
- Two node canisters, each with two 8-core processors & up to 768 GB memory for a total of 1.5 TB cache
- FS9100 includes SVC software capabilities FlashCopy / Global or Local Mirror capabilities
- Configured 2 four port -16 Gb Fibre Channel cards (8 Ports total) on one FS9110 array
- FS9150 can utilize up to 24 ports and support up to 2M IOPs
- Each FiconExpress 16s+ port can perform at up to 1.6 GB/s

Multiple Volumes Are Good

IBM Storage and SDI

The FlashSystem 9100 & Storwize V7000 are optimized for multiple volumes

*Around **30 volumes** are required to unlock the maximum performance*

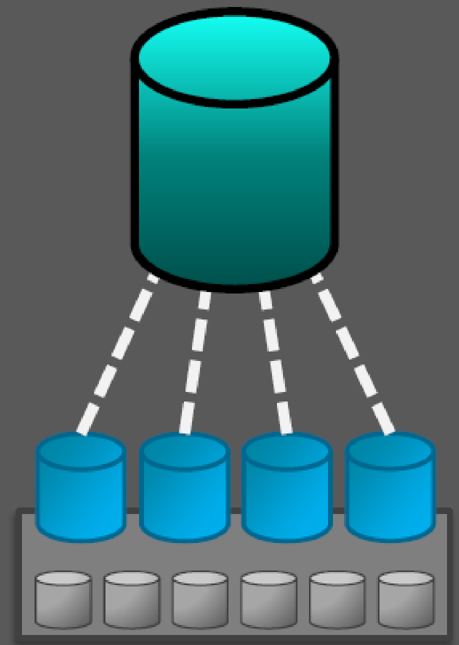
A workload can become unnecessarily limited when backed by a single volume

A single volume will be limited to up to 10% of the ultimate performance

If a single host or workload has a high performance requirement then consider creating multiple volumes and strip data across them at the host level (e.g. using Logical Volume Manager)

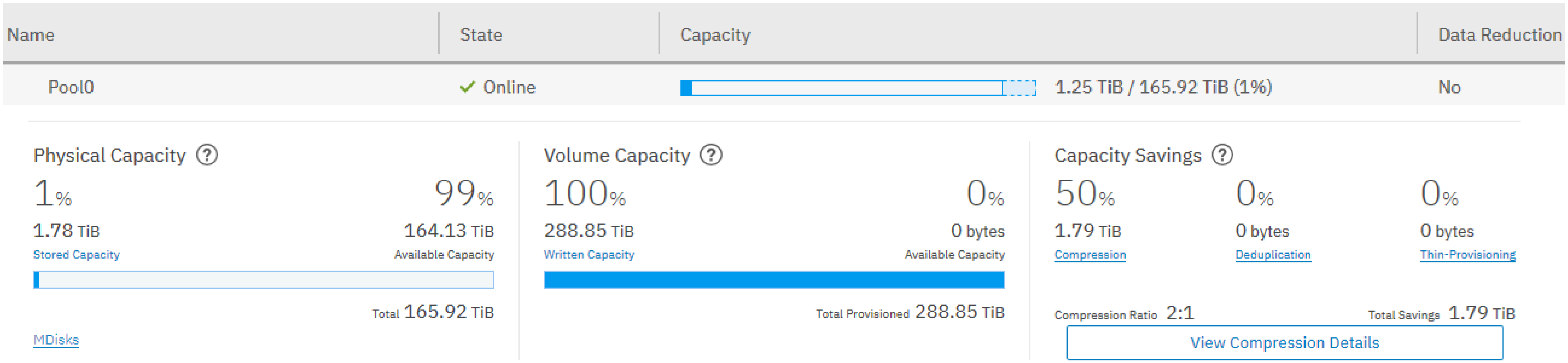
Adding volumes will initially scale performance linearly and allow the workload to be balanced across the ports and canisters

Verify the CPU core usage using the performance data



Note: For FS900 - 16 LUN volumes is optimal

FS9110 Storage Pool & Compression



- Storage pool created using all the storage (165.92 TB) as compressed storage with 288.85 TB allocatable for databases.
- All storage encrypted at the storage array level (encryption at rest).

Lessons Learned:

- Co-ordination of shipping
  - Multiple shipments to Washington Systems Center (WSC) and customer POC site with multiple components (LinuxONE, FS9110, SAN switches, network switches.
    - Weight and rack restrictions of components
    - Pack and unpack co-ordination
    - Shipping costs / Co-ordination of IBM service support representative (SSR)

- First in Enterprise (FIE) customer – simpler is better
  - Hardware Manager Console (HMC) restrictions with DS8882F
  - DS8882F line power is L6-20, LinuxONE LR1 uses L6-30 receptacles

- IBM Java and Sun Java
  - Differences with Apache Xerces Java libraries, resulted in CLASSPATH loading issues that required re-ordering of java libraries.

- Small 1u Intel / Power server for Spectrum Scale Monitoring
  - Helpful for non s390x support integration
  - Weblogic Apache Server plugin required Intel or Power server.

Agenda

Customer Experiences Oracle with Hyperconverged IBM Storage:

– Oracle Update

– Techcombank – LinuxONE Emperor II / DS8886 / LUNs with FS900 storage (managed SVC)

– US Government – Rockhopper II / FS9110

– Large Healthcare POC - Emperor II / FS9150 and FS900

– Flash Systems Demo (Poughkeepsie Benchmark Center)

Large Health Care Profile

IBM

Large Healthcare POC - Emperor II / FS9150 & FS900

– Reason for POC

- Running out of Data Center floor space
- Verify IT economic study findings
- Reduce energy costs
- Total Cost of Ownership (minor) of software

Platform	CPUs	Cores
Linux x86 64-bit	48	24

Snap Time	Load	%busy	%user	%sys	%idle	%iowait
16-Apr 01:00:37	4.14					
16-Apr 02:00:03	4.79	5.64	2.87	2.45	94.36	0.06
16-Apr 03:00:06	3.48	5.78	2.97	2.47	94.22	0.06
16-Apr 04:00:09	4.37	5.67	2.85	2.48	94.33	0.06
16-Apr 05:00:13	3.49	6.79	3.81	2.60	93.21	0.07
16-Apr 06:00:16	4.74	8.28	5.39	2.53	91.72	0.06
16-Apr 07:00:19	3.63	7.53	4.66	2.53	92.47	0.06
16-Apr 08:00:05	4.07	5.58	2.81	2.44	94.42	0.06
16-Apr 09:00:09	3.68	5.68	2.88	2.47	94.32	0.05
16-Apr 10:00:12	5.82	5.72	2.93	2.47	94.28	0.05
16-Apr 11:00:16	4.16	6.17	3.26	2.55	93.83	0.06
16-Apr 12:00:19	4.61	5.83	3.00	2.49	94.17	0.06
16-Apr 13:00:26	7.70	5.60	2.82	2.46	94.40	0.06
16-Apr 14:00:31	3.86	5.92	3.12	2.46	94.08	0.05
16-Apr 15:00:35	4.51	6.12	3.23	2.52	93.88	0.06
16-Apr 16:00:39	4.55	5.74	2.93	2.47	94.26	0.05
16-Apr 17:00:42	3.87	5.77	2.96	2.48	94.23	0.07
16-Apr 18:00:46	4.08	5.69	2.89	2.47	94.31	0.06
16-Apr 19:00:49	7.90	14.14	11.35	2.39	85.86	0.06
16-Apr 20:00:01	4.44	7.48	4.66	2.46	92.52	0.06
16-Apr 21:00:04	4.34	5.63	2.84	2.46	94.37	0.06
16-Apr 22:00:07	6.89	5.81	2.99	2.48	94.19	0.05
16-Apr 23:00:03	4.07	10.34	7.31	2.65	89.66	0.07
17-Apr 00:00:06	5.83	5.89	3.08	2.48	94.11	0.05

ExaSizer - Sizing Just Based on CPU load Observations:

IBM

Sizing Input:

Application Name	Hardware Description and Lookup Identification	# OEM Servers Enter # (Note: Fractional values must be less than 1.00)	Result	Peak Utilization Case-1 (Default Values) 0.0%	Peak Utilization Case-2 (Default Values) 0.0%	No.	Workl
PRPPR1	HS23 Xeon E5-2697v2 2.7GHz (2ch/24co)	1.00	1.00	5.7%	7.6%	6	Database
PRPPR2	HS23 Xeon E5-2697v2 2.7GHz (2ch/24co)	1.00	1.00	2.1%	3.3%	6	Database
IPDRM2PR1	HS23 Xeon E5-2697v2 2.7GHz (2ch/24co)	1.00	1.00	7.2%	8.8%	6	Database
IPDRB1PR2	HS23 Xeon E5-2697v2 2.7GHz (2ch/24co)	1.00	1.00	3.8%	4.9%	6	Database
P04DB05E1	x3650 M5 Xeon E5-2699v3 2.3GHz (2ch/36co)	1.00	1.00	6.1%	8.1%	6	Database
P04DB05E2	x3650 M5 Xeon E5-2699v3 2.3GHz (2ch/36co)	1.00	1.00	1.8%	1.8%	6	Database
P04DB07B1	x3650 M5 Xeon E5-2699v3 2.3GHz (2ch/36co)	1.00	1.00	16.3%	17.6%	6	Database
P04DB07B2	x3650 M5 Xeon E5-2699v3 2.3GHz (2ch/36co)	1.00	1.00	8.5%	9.5%	6	Database

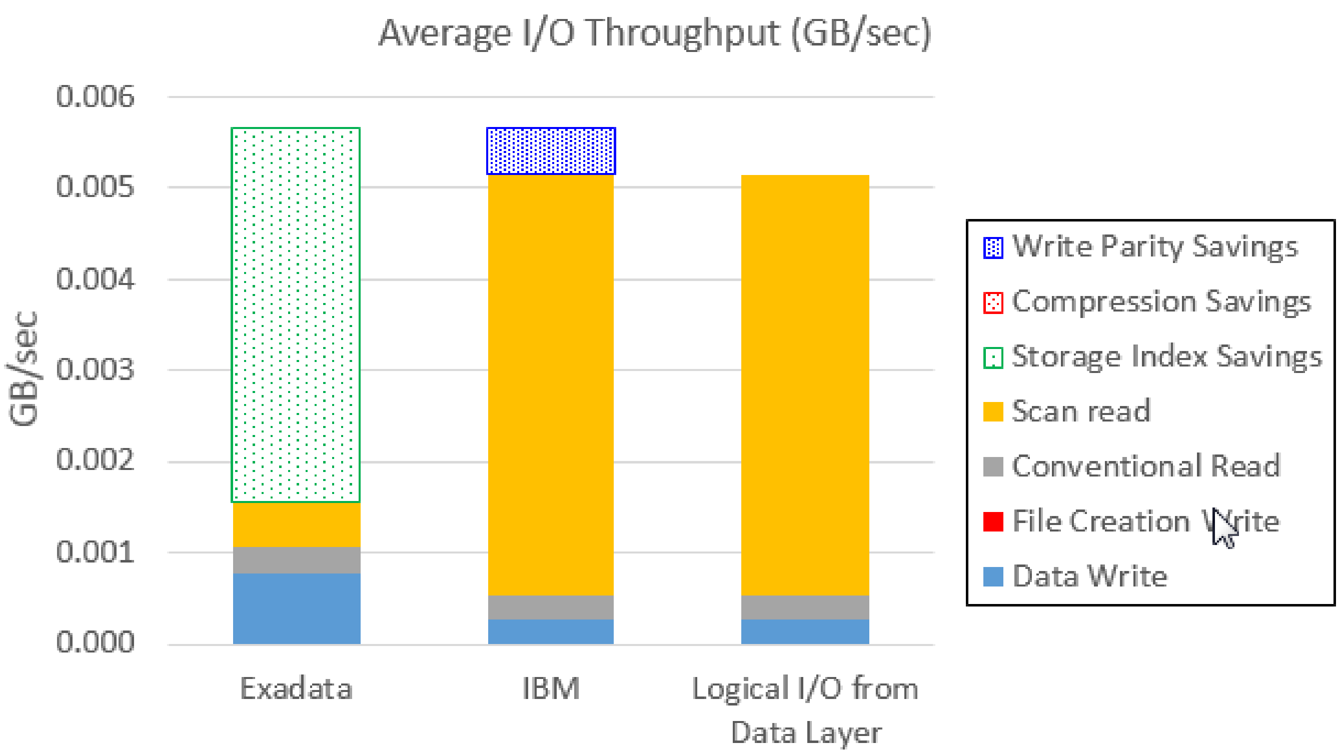
Sizing Results:

Required MIPS *(based on server inputs)*	9,647
Hypervisor Overhead % (zVM)	3%
Hypervisor Overhead MIPS	289
Peak Concurrency	100%
SDP	90%
MIPS required *(considering overhead and peak concurrency)*	9,936
MIPS required *(to maintain SDP)*	11,888

Target System	Cores (IFLs)	Estimated Utilization on Target System
LinuxONE Emperor II (z14)	**5**	84%

Sizing Notes:

- High benefit from storage indexes, partitioning the corresponding table and/or adjusting query could potentially significantly reduce IO requirements in an IBM environment.



Sizing Notes:

IBM

- Working with business partner, we sized the Oracle using 8K data blocks to 4 FS900 Systems

Est. Maximum Required Usable Storage Capacity (TB)	217.7	
Minimum required I/O throughput (GB/sec)	17.2	
read throughput	15.9	
write throughput	1.4	
Minimum required total IOPS	120,815	**High**
read IOPS	108,953	
write IOPS	11,862	
I/O density (IOPS/GB)	0.5	
Read:Write IOPS ratio	90:10	**Read dominant**
Read Sequential IOPS %	18%	
Write Sequential IOPS %	64%	
Average I/O transfer size - all (KB)	149.5	
read (KB)	152.7	**Mixed random/sequential**
write (KB)	120.6	**Mixed random/sequential**

zEnterprise Data Compression (zEDC) Testing

- One of top cpu consumers: nightly Oracle database backups.
- zEnterprise Data Compression (zEDC) Testing
  - Not Integrated with Oracle at this time
  - Can be used to compress / uncompress Oracle or Open source databases on disk

Without zEDC Card assist

```
[root@apsmd0051 ~]# time gzip samplerand.txt

real    0m38.037s
user    0m36.680s
sys     0m0.953s
[root@apsmd0051 ~]# time gunzip samplerand.txt.gz

real    0m5.709s
user    0m4.502s
sys     0m0.923s
```

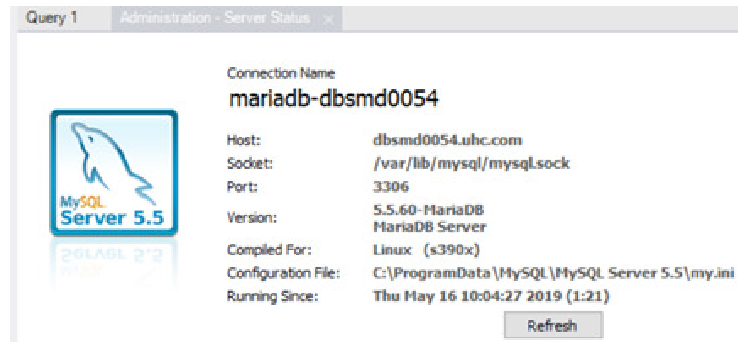
zEDC Card assist

```
[root@apsmd0051 ~]# time genwqe_gzip samplerand.txt

real    0m3.583s
user    0m0.115s
sys     0m1.094s
[root@apsmd0051 ~]# time genwqe_gunzip samplerand.txt.gz

real    0m5.446s
user    0m0.010s
sys     0m1.533s
```

	No zEDC	zEDC	%Improve
real time	38.037	3.583	961.59643
Total cpu	37.633	1.209	3012.7378

MySQL / MariaDB on LinuxONE vs MySQL on x86

IBM



MySQL 2.1x faster for insert,
MariaDB 2.53x faster on LinuxONE

mysql55_bench.x...	dbsmd0065 s390x MySQL 5.5.52		apvrt31478 x86 MySQL 5.5.52		dbsmd0065 s390x MariaDB 5.5.53	
Test	Wall clock (secs)	CPU (secs)	Wall clock (secs)	CPU (secs)	Wall clock (secs)	CPU (secs)
alter-table	12	0.01	15	0.09	11	0.01
ATIS	5	0.47	9	1.2	5	0.47
big-tables	5	0.81	6	1.4	5	0.8
connect	15	6.02	38	22.52	17	6.26
create	64	0.77	88	3.35	59	0.83
insert	711	43.58	1472	126.05	581	44.89
select	99	3.92	126	13.02	115	3.98
wisconsin	11	0.81	21	2.51	9	0.82

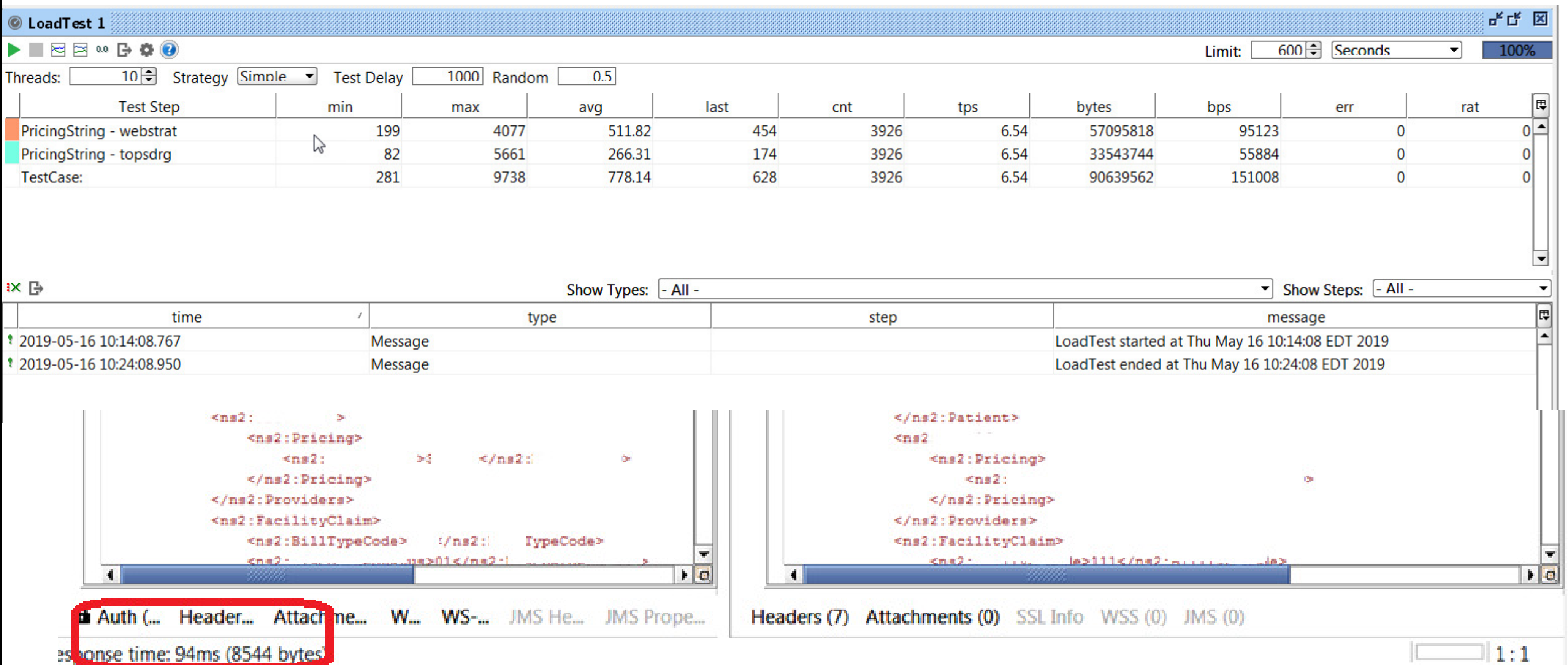
LinuxONE VMWARE LinuxONE/MariaDB

(MariaDB = MySQL + support from RedHat)

GGMAP - OpenShift on Intel vs LinuxONE Testing Results

IBM

Customer: Simulated transaction that mirrors Optum TOPS system calls about half-million times a day. Ran same transaction against our production RHEL6 platforms, it runs in about **325 ms** on average. Transaction below averages about **94 ms** on LinuxONE



Test Step	min	max	avg	last	cnt	tps	bytes	bps	err	rat
PricingString - webstrat	199	4077	511.82	454	3926	6.54	57095818	95123	0	0
PricingString - topsdrg	82	5661	266.31	174	3926	6.54	33543744	55884	0	0
TestCase:	281	9738	778.14	628	3926	6.54	90639562	151008	0	0

time	type	step	message
2019-05-16 10:14:08.767	Message		LoadTest started at Thu May 16 10:14:08 EDT 2019
2019-05-16 10:24:08.950	Message		LoadTest ended at Thu May 16 10:24:08 EDT 2019

I/O Performance VMWARE vs LinuxONE (IBM FlashSystem 900)

I/O Performance VMWARE:
10436 (8K) IOPs / 1543 MB/s

```
SQL> @io.sql
max_iops = 10436
latency  = 8
max_mbps = 1543

PL/SQL procedure successfully completed.
```

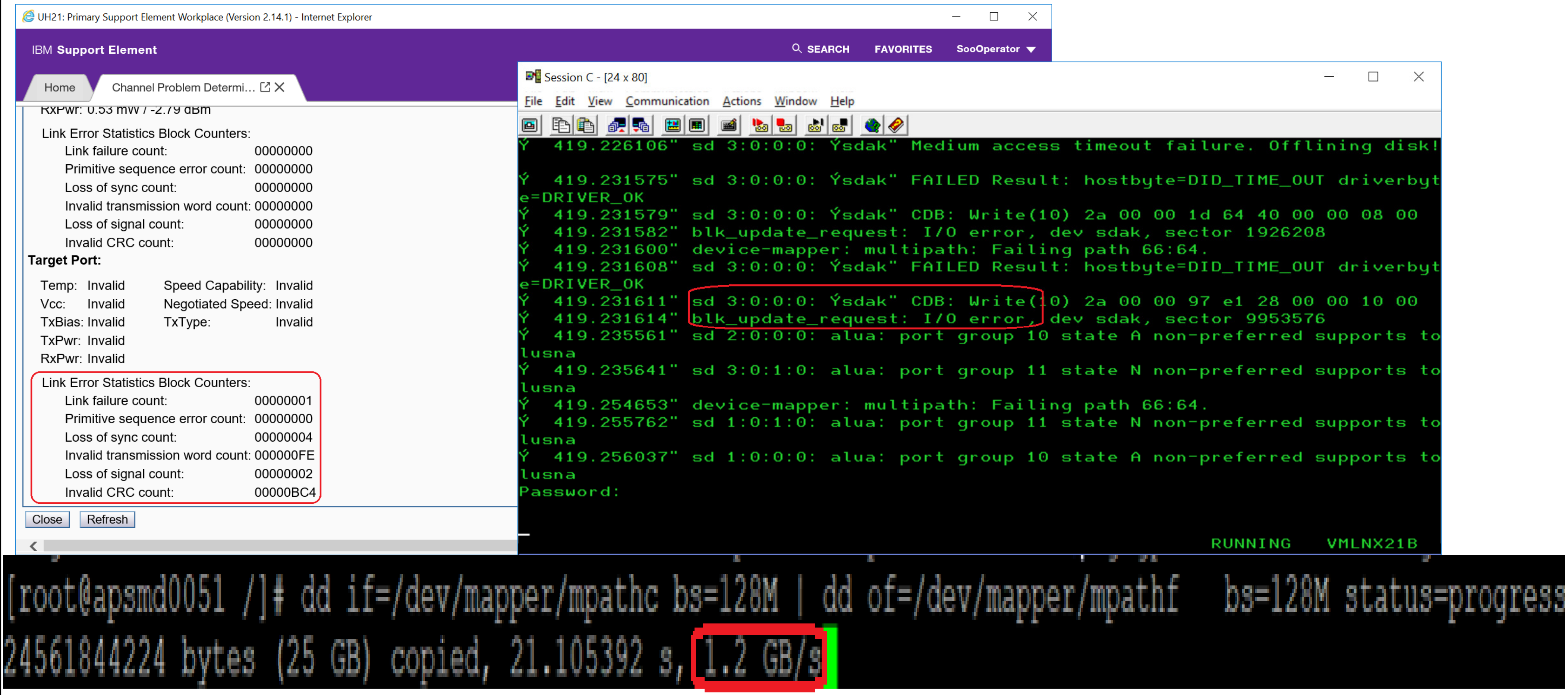
I/O Performance LinuxONE:
596145 IOPs (8K) / 26214 MB/s

```
SQL> @io.sql
max_iops = 596145
latency  = 0
max_mbps = 26214

PL/SQL procedure successfully completed.
```

LinuxONE – 50x more IOPs than x86

I/O Channel Path

– POC initial issues with I/O paths
– Cloning a Linux Image **20 MB/s** to 1.2 GB/s when bad I/O paths turned offline.



Agenda

Customer Experiences with LinuxONE and IBM Flash Systems:

– Oracle Update

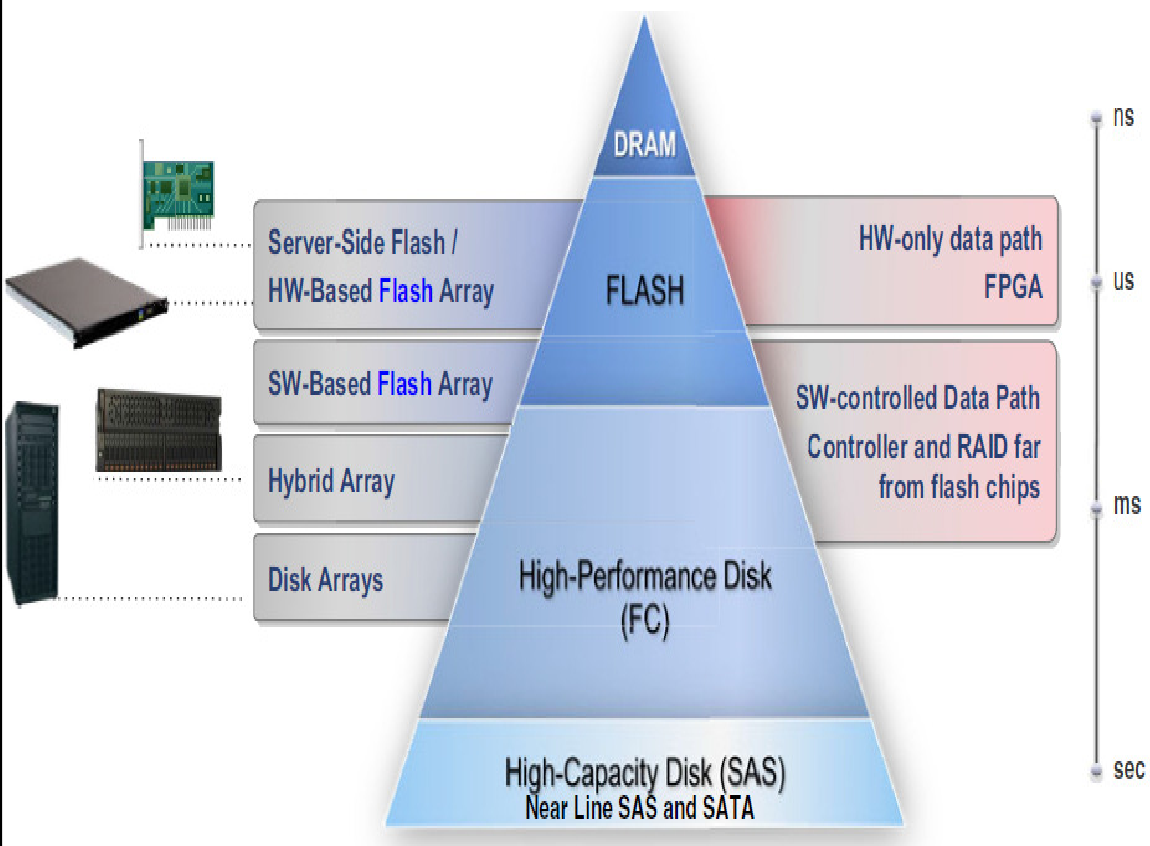
– Techcombank – LinuxONE Emperor II / DS8886 / LUNs with FS900 storage (managed SVC)

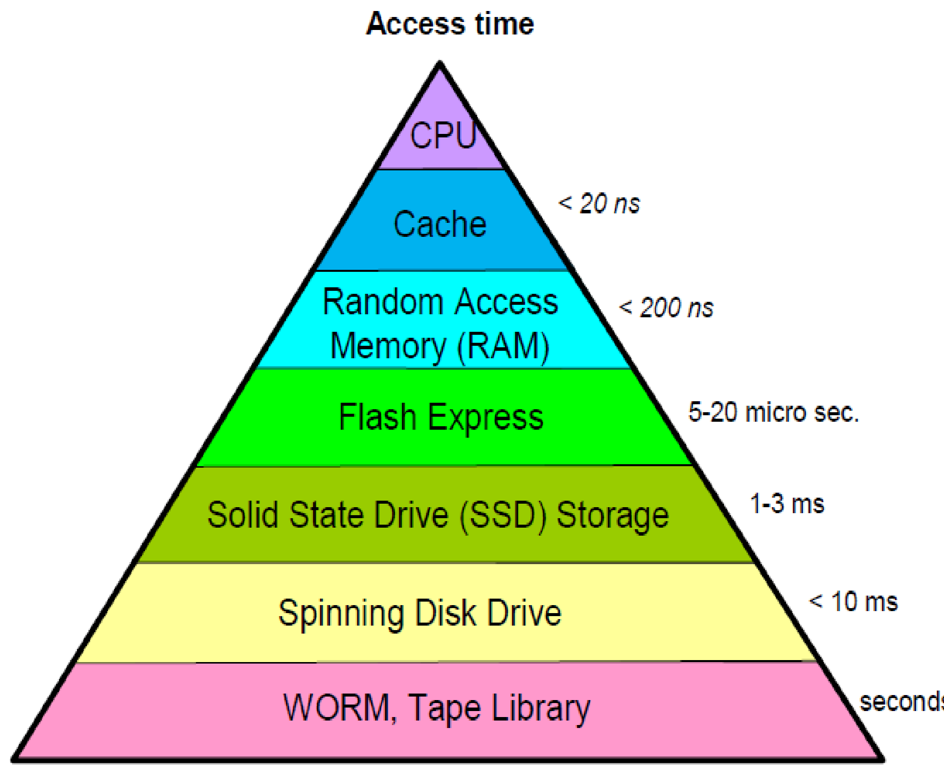
– US Government – Rockhopper II / FS9110

– Large Healthcare POC - Emperor II / FS9150 and FS900

– Flash Systems Demo (Poughkeepsie Benchmark Center)

Oracle Cache, Memory and I/O Access





Testing LinuxONE at an IBM Benchmark Center:

IBM

LinuxONE / Storage Testing Tools **:

– Oracle I/O Calibrate (random & sequential Oracle I/O) -
– SwingBench (synthetic Order Entry transactional workload)
– Vdvench (various I/O tests)
– Iometer (Open source agreement)
– Silly Little Oracle Benchmark (SLOB)
– HammerDB

** Oracle license restrictions in publishing Oracle benchmarks

CPU Test: Logical I/O

```
Load Profile                    Per Second
~~~~~~~~~~~~~~~               ---------------
               DB Time(s):             30.8
                DB CPU(s):             30.7
        Background CPU(s):              0.0
        Redo size (bytes):         15,265.8
    Logical read (blocks):     17,814,695.4
           Block changes:              45.3
   Physical read (blocks):              0.5
  Physical write (blocks):              0.1
        Read IO requests:               0.5
       Write IO requests:               0.0
            Read IO (MB):               0.0
           Write IO (MB):               0.0
```

Storage Test: Physical Read

```
Load Profile                    Per Second
~~~~~~~~~~~~~~~               ---------------
               DB Time(s):            122.5
                DB CPU(s):             19.1
        Background CPU(s):              0.0
        Redo size (bytes):          7,798.2
    Logical read (blocks):        613,151.0
           Block changes:              23.9
   Physical read (blocks):        596,755.3
  Physical write (blocks):              3.1
        Read IO requests:         596,749.5
       Write IO requests:               1.5
```

Oracle I/O Demo with an IBM FlashSystem

After switching to FlashSystem

Disk IO wait disappears and waiting is now on host CPU. This graph shows the effect of the low latency of FlashSystem and how it increases the host CPU utilization.