Retour d'expérience et Performance Oracle sur POWER9 (démonstration et Benchmark clients)

Alain Cyr – cyralain@fr.ibm.com
Frederic Dubois – fred.dubois@fr.ibm.com
IBM Client Center Montpellier

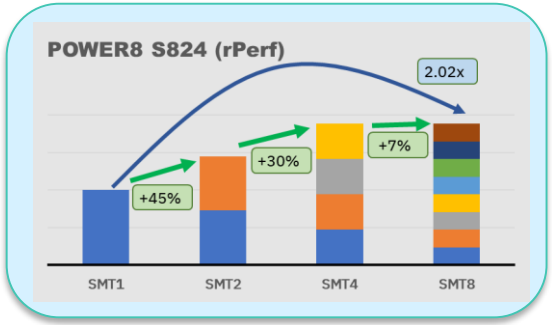
June 18th ,2019

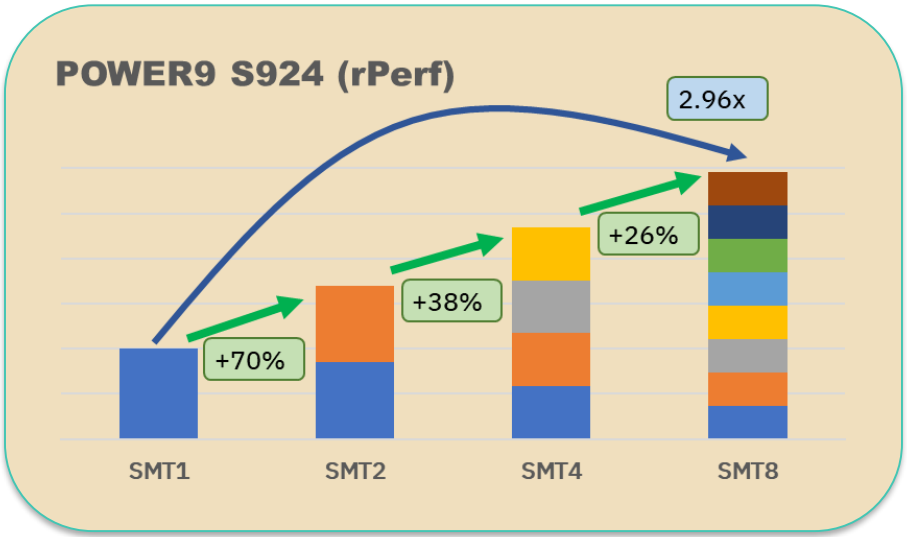
POWER9 Improved multi-threading options

Can dynamically shift between multi-threading modes to adapt to changing workload:

- **SMT1:** Largest unit of execution work
- **SMT2:** Smaller unit of work, but provides greater amount of execution work per cycle
- **SMT4:** Smaller unit of work, but provides greater amount of execution work per cycle
- **SMT8:** Smallest unit of work, but provides the maximum amount of execution work per cycle

POWER9 has nearly 50% better SMT scalability than POWER8





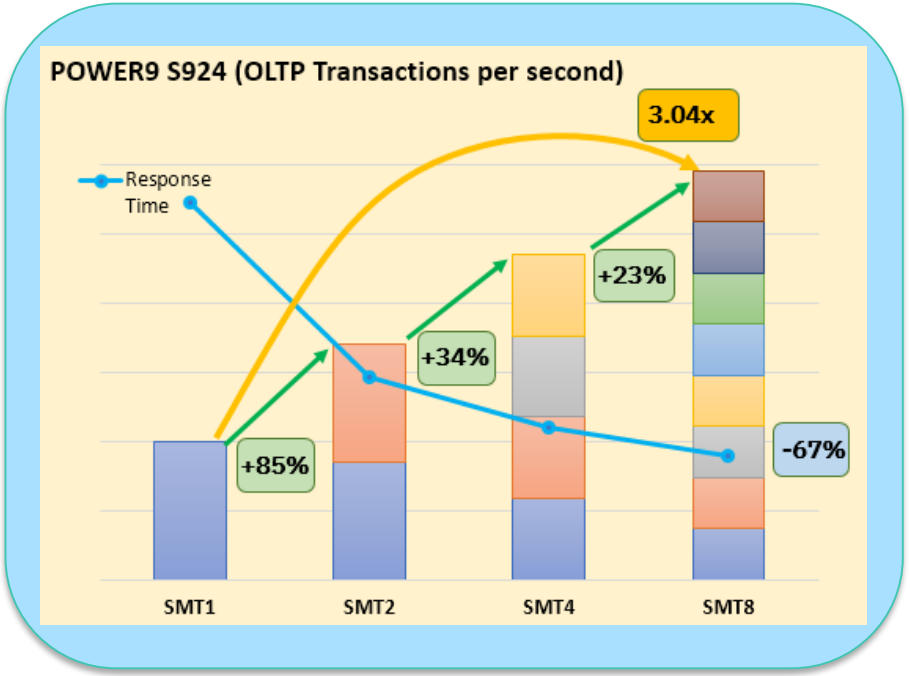
Based on published rPerf for S824 and S924 servers.

POWER9 SMT scalability proofpoint

Based on IBM internal OLTP workload testing, actual SMT scalability

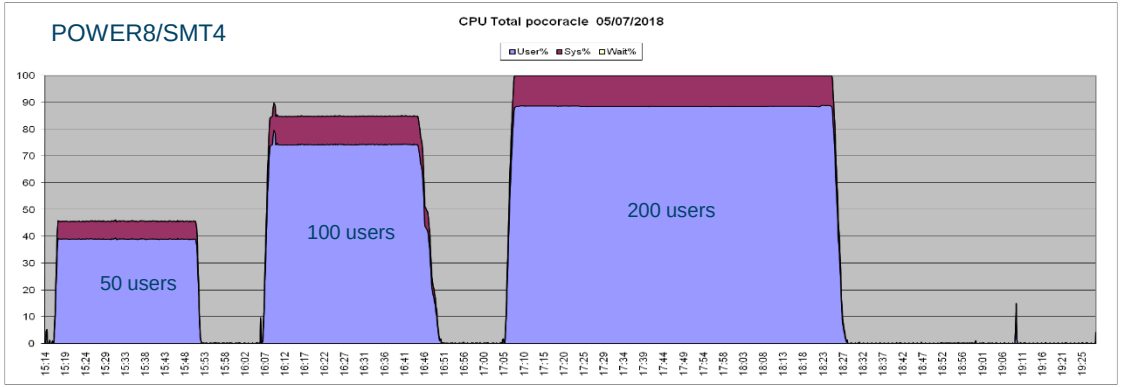
- **SMT1:** Largest unit of execution work
- **SMT2:** Smaller unit of work, but provides greater amount of execution work per cycle
- **SMT4:** Smaller unit of work, but provides greater amount of execution work per cycle
- **SMT8:** Smallest unit of work, but provides the maximum amount of execution work per cycle

POWER9 has nearly 50% better SMT scalability than POWER8



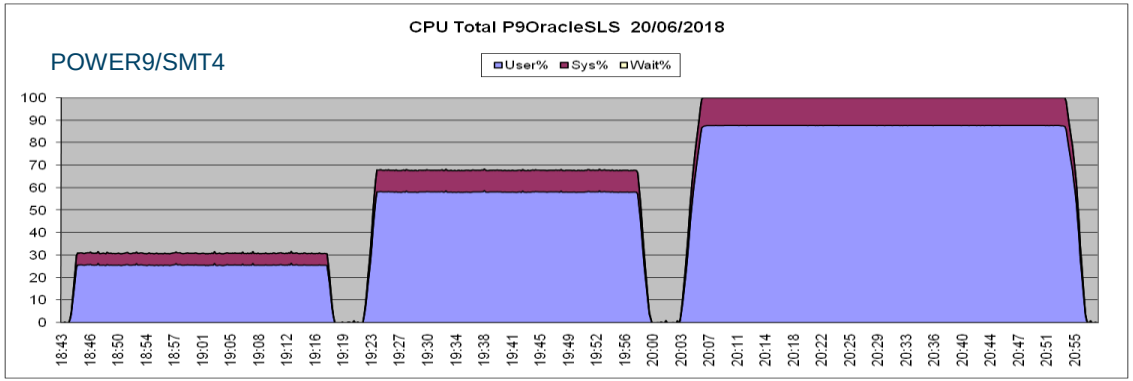
Based on IBM internal OLTP workload performance testing. Your results may vary!

TEST1 : OLTP – 50/100/200 Users – POWER8/SMT4 vs POWER9/SMT4



- 50 users: DBTIME = 54 mn
- 100 users: DBTIME = 180 mn
- 200 users: DBTIME = 507 mn

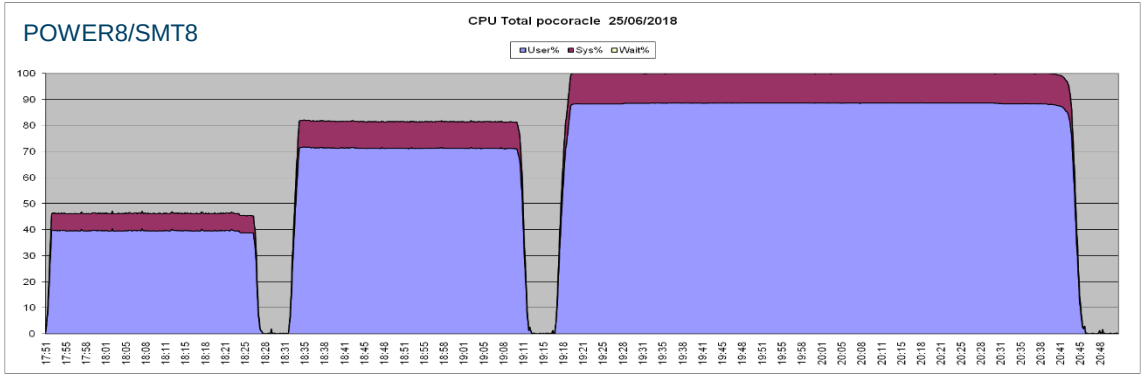
50 users : POWER9/SMT4 **1.50 x** time faster than POWER8/SMT4 : **-33% used CPU**
100 users : POWER9/SMT4 **1.60 x** time faster than POWER8/SMT4 : **-20% used CPU**
200 users : POWER9/SMT4 **1.75 x** time faster than POWER8/SMT4 : **RQ 80(P9) vs 150(P8)**



- 50 users: DBTIME = 35 mn
- 100 users: DBTIME = 108 mn
- 200 users: DBTIME = 291 mn

TEST2 : OLTP – 50/100/200 Users – POWER8/SMT8 vs POWER9/SMT8

POWER8/SMT8

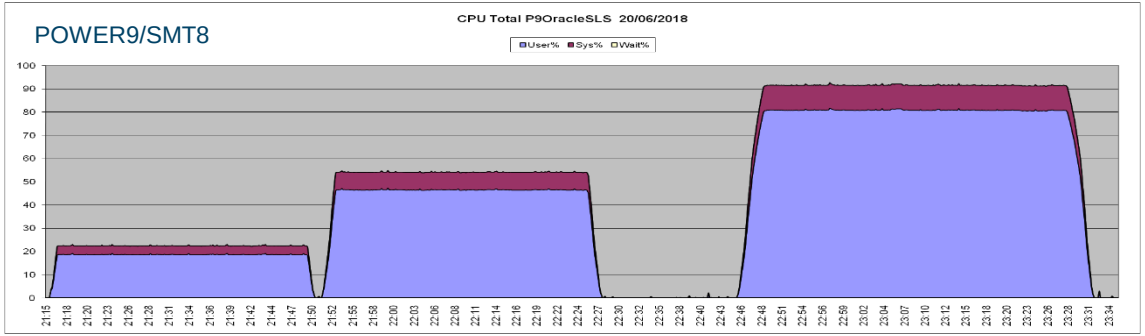


- 50 users: DBTIME = 53 mn
- 100 users: DBTIME = 170 mn
- 200 users: DBTIME = 425mn

The more POWER9 LPAR is loaded the faster it goes

50 users : POWER9/SMT8 **1.5 x** time faster than POWER8/SMT8 : **-48 % used CPU**
100 users : POWER9/SMT8 **1.7 x** time faster than POWER8/SMT8 : **-37% used CPU**
200 users : POWER9/SMT8 **2.2 x** time faster than POWER8/SMT8 : **-10% used CPU RQ 60(P9)/140(P8)**

POWER9/SMT8

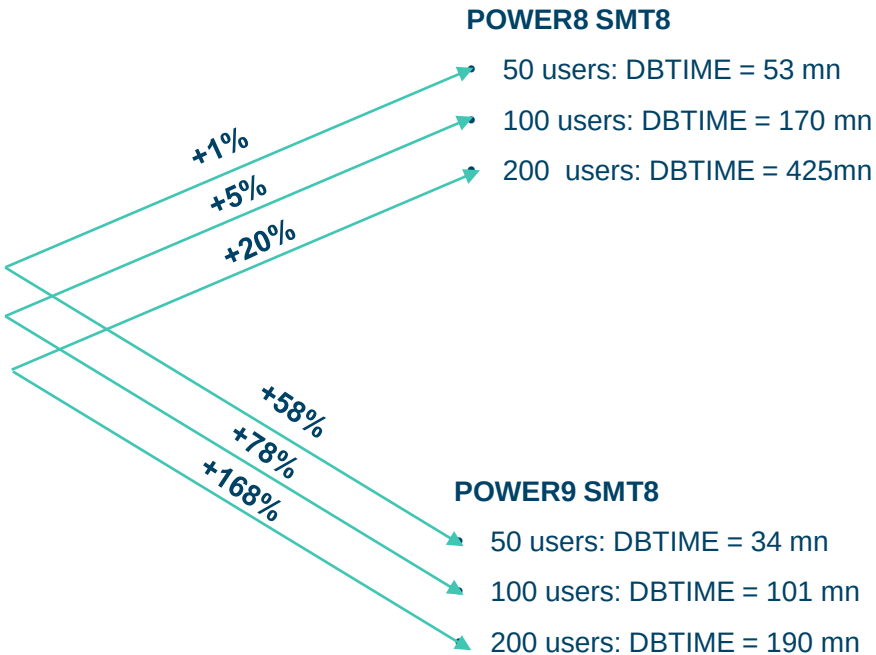


- 50 users: DBTIME = 34 mn
- 100 users: DBTIME = 101 mn
- 200 users: DBTIME = 190 mn

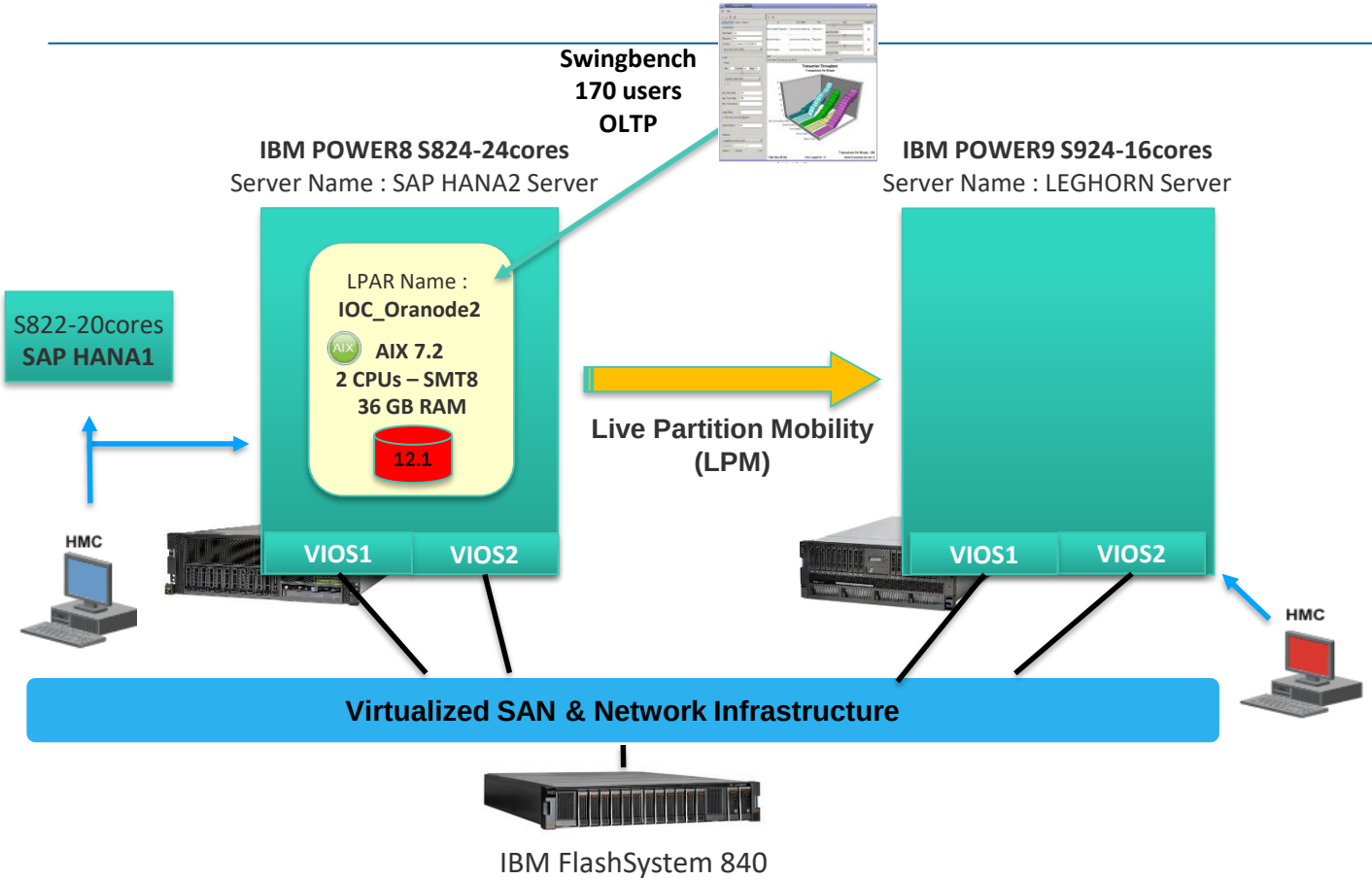
Test Summary

POWER8 SMT4

- 50 users: DBTIME = 54 mn
- 100 users: DBTIME = 180 mn
- 200 users: DBTIME = 507 mn



Same Oracle Licensing Cost



- **<u>Purpose of the Demo/Test</u>**
- Injection d'une charge OLTP orientée CPU sans goulot d'étranglement IO sur une infrastructure POWER8
- Capture des résultats de Performance sur le serveur POWER8
- Migration "Live" via LPM de la LPAR Oracle vers un serveur POWER9
- Capture des résultats de Performance sur le serveur POWER9
- Comparaison des résultats
- Aucune optimization entre les deux tests

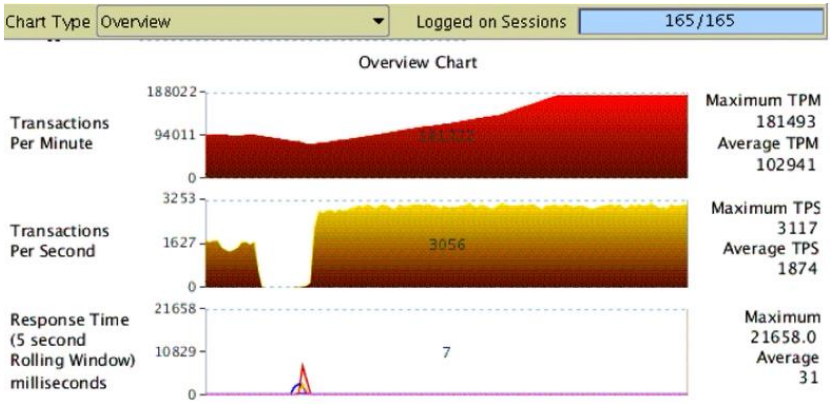
Cette démonstration présente une <u>illustration</u> des avantages de la technologie IBM POWER9.
Ce n'est pas un test de performance officiel

7

~ 70% CPU (vs 100% on POWER8)

~ 181k TPM (vs 121k on POWER8)

x 1.5 TPM
and
30% CPU Savings

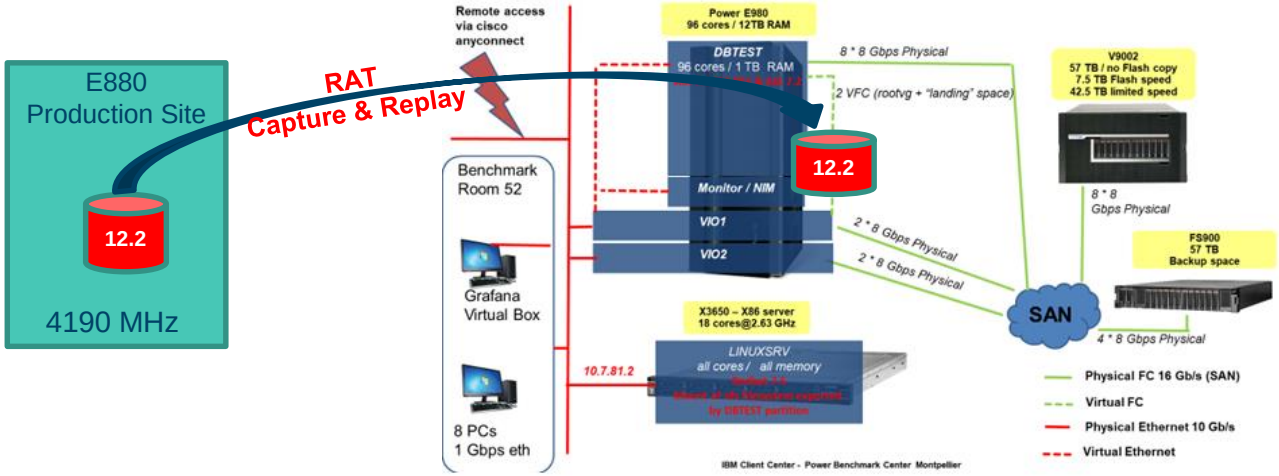



```
topas_nmon--m=Memory----------Host=oranode2-------Refresh=2 secs---23:03.53
 CPU-Utilisation-Small-View
                          0----------25----------50----------75----------100
CPU  User%  Sys% Wait% Idle%|          |          |          |           |
  0  56.5   8.5   0.0  35.0|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUssss                >
  1  48.5   8.0   0.0  43.5|UUUUUUUUUUUUUUUUUUUUUUUUUUssss                    >
  2  55.0   4.0   2.0  39.0|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUssW                 >
  3  50.0   8.0   0.0  42.0|UUUUUUUUUUUUUUUUUUUUUUUUUUUUssss                 >
  4  22.0   4.5   1.0  72.5|UUUUUUUUUUUss                                    >
  5  29.0   1.5   1.5  68.0|UUUUUUUUUUUUUU                                  >|
  6  27.0   2.5   0.5  70.0|UUUUUUUUUUUUUs                                   >
  7  25.0   4.5   0.0  70.5|UUUUUUUUUUUUss                                   >
  8  53.0  10.5   0.5  36.0|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUsssss               >
  9  52.5  10.5   4.0  33.0|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUssssWW              >
 10  56.0   4.5   0.5  39.0|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUss                 >
 11  55.0   8.5   1.5  35.0|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUssss               >
 12  27.9   7.0   0.0  65.2|UUUUUUUUUUUUUsss                                 >
 13  25.0   5.0   1.5  68.5|UUUUUUUUUUUUss                                  > |
 14  30.0   7.0   0.5  62.5|UUUUUUUUUUUUUUsss                                >
 15  33.0   4.5   1.0  61.5|UUUUUUUUUUUUUUUss                               >
Physical Averages          +-----------|-----------|-----------|-----------+
All  62.9   7.2   2.3  27.7|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUsssW          >|
                           +-----------|-----------|-----------|-----------+

 Dedicated-CPU-Logical-Partition
Partition:Number=16 "IOC_oranode2"
Flags:LPARed DRable SMT Dedicated Sharable Migratable Not-Donating.
Summary: Entitled= 2.00  Used     2.00 (100.0%) 12.5% of CPUs in System

CPU-Stats---------   Capacity----------------  ID-Memory--------------------
max Phys in sys  16  Cap. Processor Min  1.00  SPLPAR Group:Pool  32784:65535
Phys CPU in sys  16  Cap. Processor Max  4.00  Memory(MB) Min:Max  256:49152
Virtual Online    2  Cap. Increment      1.00  Memory(MB) Online   36864
Logical Online   16  Cap. Unallocated    0.00  Memory Region LMB   256MB min
Physical pool     0  Cap. Entitled       2.00  Time------------------Seconds
SMT threads/CPU   8   -MinReqVirtualCPU  1.00  Time Dispatch Wheel    0.0000
CPU---------Min-Max  Weight----------------  MaxDispatch Latency    0.0000
Virtual       1   4  Weight Variable       0  Time Pool Idle         0.0000
Logical       1  32  Weight Unallocated    0  Time Total Dispatch    1.9997
-------------------------------------------------------------------------------
Event= 6 post   unknown SerialNo Old=C00980 Current=C00980 When=23:02:36
-------------------------------------------------------------------------------
Shared_Pools   MaxPoolCapacity= 0.00     No Shared CPU Pool Authority
SharedCPU=0    EntPoolCapacity= 0.00     No Shared CPU Pool Authority
```

Customer Benchmark – E880 to E980 POWER Upgrade

A Brazilian Bank customer runs its banking application on Oracle with POWER8 systems. The growth rate leads to 30% per year increase of CPU consumption. In order to sustain the growth rate for the next 3 years, Customer wants to test the Power E980 to validate Capacity Planning and Validate Performance of its application on POWER9

The benchmark results show that POWER9 technology can save between 31% to 38% core consumption compared to Power8 technology.



Production System - Capture
P8_CORES = XX cores (P8) * (User% + Sys%)=YY cores.

PoC System - Replay
P9_CORES = XX cores (P9) * (User% + Sys%)=YY cores

P8 to P9 Benefits
CORE_SAVING%= 1 - P9_CORES/P8_CORES

	USR%	SYS%	SUM USR% + SYS%	Consumed cores
20c SMT8-P8	59.1	7.7	66.8	10.02
15c SMT8 -P9	69.9	7.3	77.2	15.44
Saving %				35.1 %

	USR%	SYS%	SUM USR% + SYS%	Consumed cores
27c SMT8 -P8	58.1	7.0	65.1	17.58
20c SMT8 – P9	47.4	6.9	54.3	10.86
Saving %				38.22 %

DWH on IBM POWER9, In-Memory Option Experience

Test Environment & Workload

- IBM Power Systems S924
- 24 Dedicated processor Cores
- AIX7.2 TL02
- 256GB Memory
- IBM FlashSystem 840

In-Memory additional cost
vs
reduced licensed cores ?
vs
Exadata DWH Secret Sauce

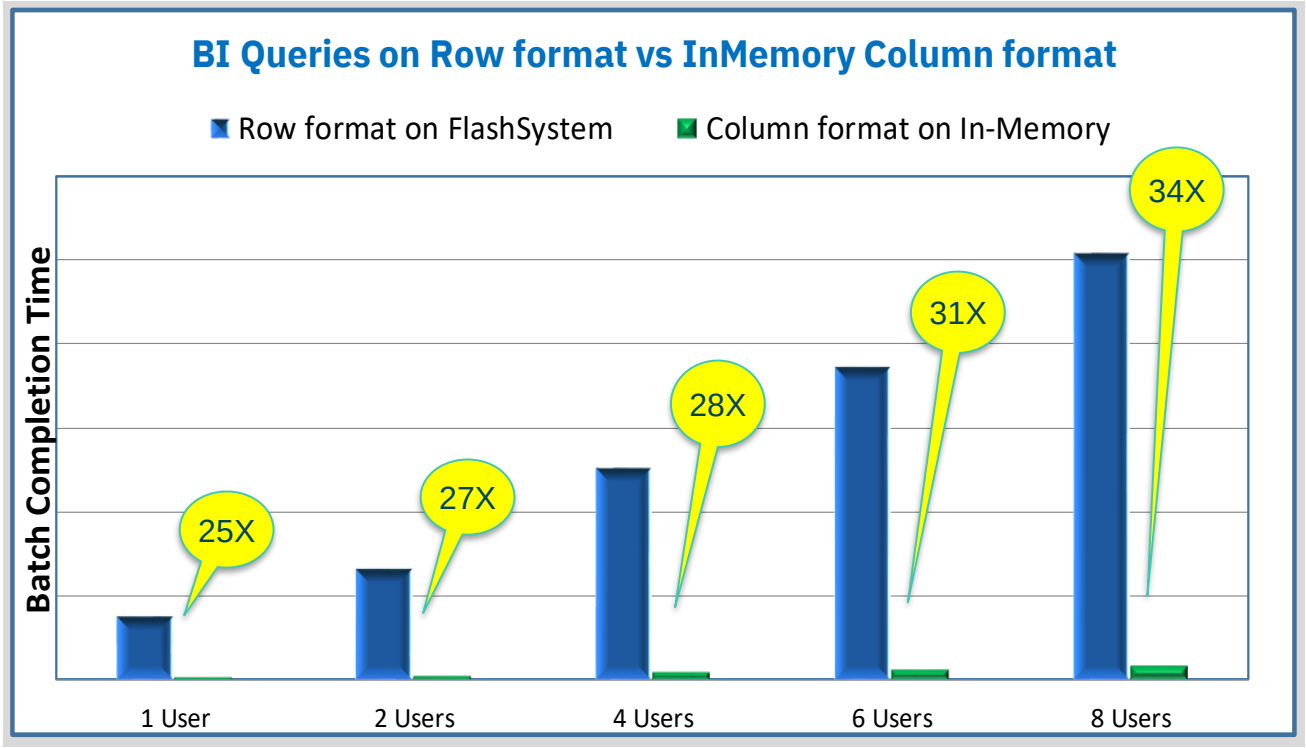
Database details

- Oracle DB 12c version 12.2.0.1 Single Instance
- 1TB BI Raw Data Base Case (STAR schema, Fact Table size 726GB)
- Used "MEMCOMPRESS FOR QUERY LOW" to compress 726GB table data to 200GB
- 250GB SGA (200GB allocated for In-Memory)
- 26 internally developed BI queries

BI Queries : In-Memory on POWER9 vs “Row” format

S924 delivers the Memory BW and CPU capacity needed to scale!

- BI workload run on 1TB data
- Number of users is how many concurrent batch sets are run
- Shorter bar is faster run time
- SMT8 drives high utilization with 8 users



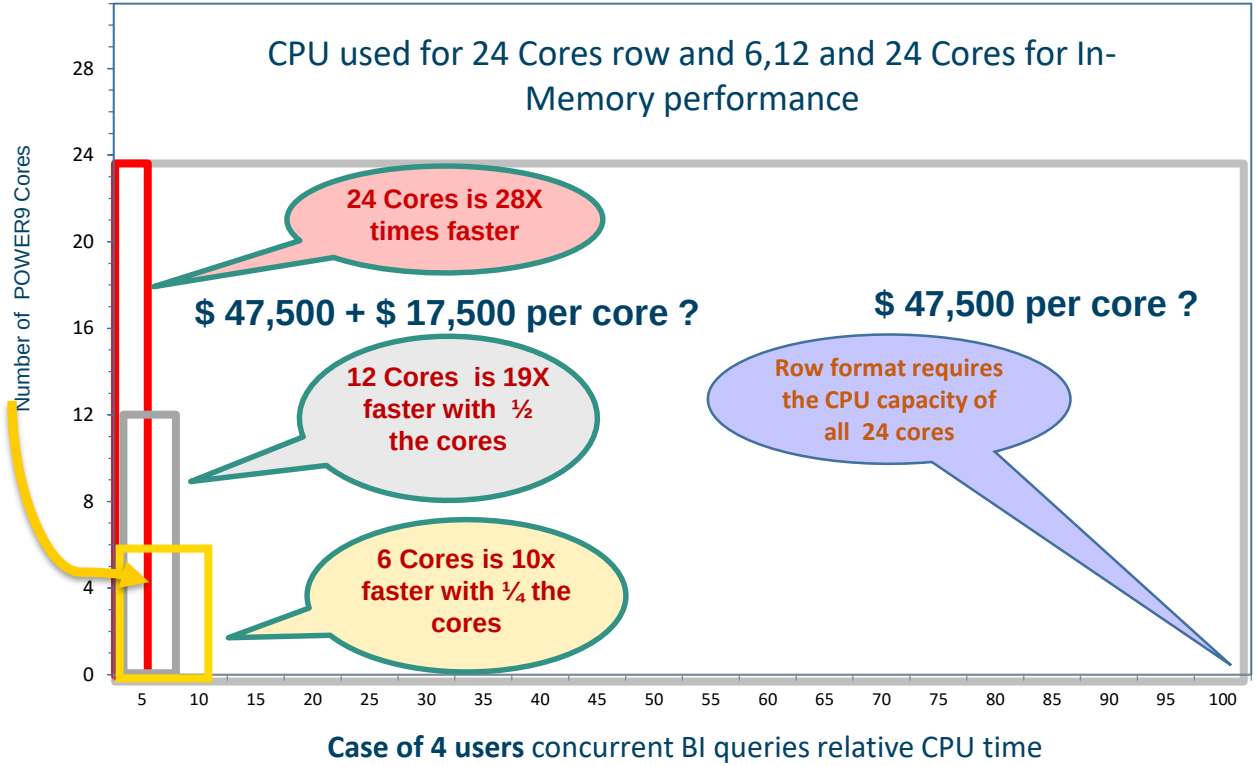
Benefits of IBM POWER9 for Oracle DB In-Memory

- Faster answers
- Less CPU power required
- Efficient license use

40X improvement to produce answers!

10x Faster Answer

4X less CPU



BI & OLTP workloads Co-deployment on POWER9 Infrastructure with Oracle DB In-Memory Option

Scale Out POWER9 are eligible to Oracle Standard Edition 2 Licensing

Oracle Standard Edition 2 Licensing – Priced by the number of Sockets

- Starting version 12.1.0.2, Oracle introduced Standard Edition 2 (only for up to 2 sockets machines)
- **All POWER9 Scale Out 1 & 2 sockets are eligible** (1 POWER8 chip = 1 licensed socket)
- Only POWER8 S814 was eligible due to DCM Technology (1 POWER8 chip = 2 licensed socket)

Oracle SE2 is the main socket based licensing solution

- Oracle SE (4 Sockets) and Oracle SE1 (2 Sockets) are withdrawn from marketing by Oracle
- **96% savings vs 1 socket 12 cores POWER9 Enterprise Edition ! (list price)**

Oracle RAC (Real Application Cluster) is supported till 18c

- When Oracle RAC feature is used with SE2, eligibility is determined by the total number of sockets in the cluster, not simply the number of sockets in an individual system

Higher DB consolidation with SE2 on POWER9 than Intel

- 2 sockets x86 servers up to 2*24 cores (2.1 GHz), S924 is 24 cores (3.4 to 3.9 GHz max)
- P9 is more than 2x performance per core and runs 4x more SMT threads vs x86 (HMT)
  - ➔ 2x more threads for more performance

Throughput and threads performance automagically tuned by SMT

- SE2 limits to 16 threads per Database – applied for any platform
- 16 threads per DB => 2 Cores SMT8 or 4 Cores SMT4, … 8 Cores SMT2, 16 cores SMT1

Oracle Standard Edition 2 – How it works

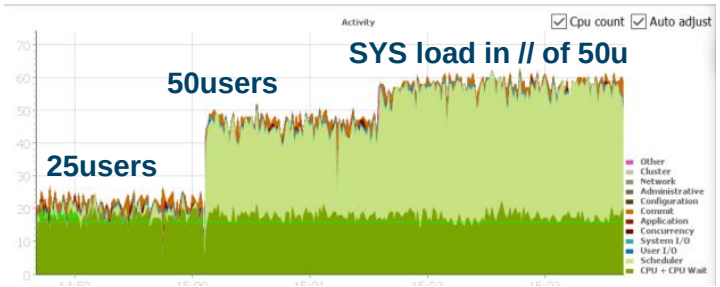
- Each SE2 database can use at maximum 16 Threads at any given time.
- When used with Oracle Real Application Clusters, each instance of the SE2 RAC Database can use a maximum of 8 threads at any given time.

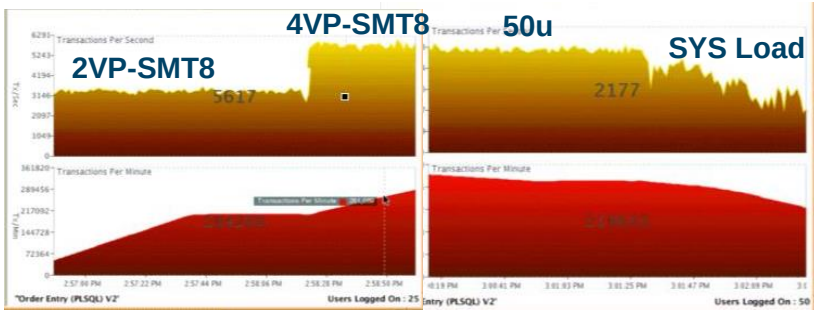
```
NAME                                 TYPE        VALUE
------------------------------------ ----------- -------
cpu_count                            integer     16
parallel_threads_per_cpu             integer     2
resource_manager_cpu_allocation      integer     16
SQL>
```

SE2 Database dans une LPAR avec 2 coeurs SMT8

```
NAME                                 TYPE        VALUE
------------------------------------ ----------- ---------
cpu_count                            integer     32
parallel_threads_per_cpu             integer     2
resource_manager_cpu_allocation      integer     16
SQL>
```

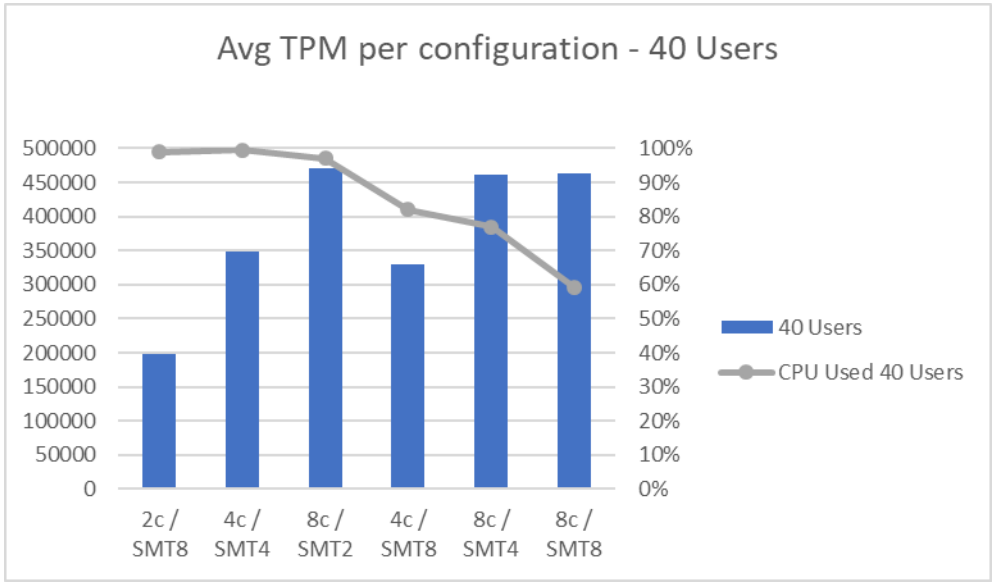
SE2 Database dans une LPAR avec 4 coeurs SMT8





- Only “user” session are impacted by Oracle SE2 scheduler
- Higher priority for user “SYS”

Oracle Standard Edition 2 – Performance Impacts and Sizing



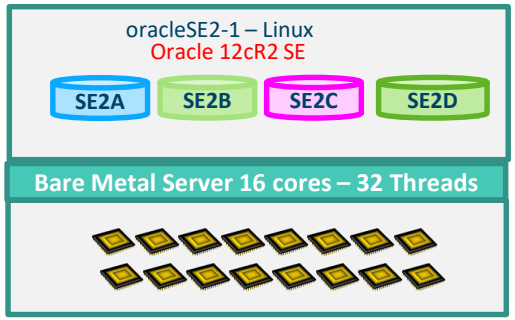
- 40 users load maximizes 16 threads SE2 Performance

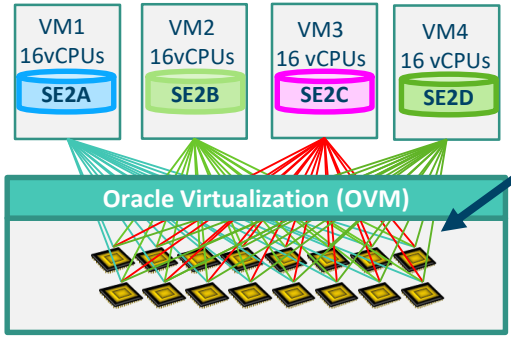
Maximize POWER9 Core performance to reach 16 SE2 Threads Limitation

Leverage POWER9 SMT Level to save CPU for additional SE2 databtases consolidation

AUTOMATED THROUGHPUT OPTIMIZATION
depending on Workload Profile (prod & non-prod), Consolidation, Performance …
based on the balance between cores and SMT allocation

Oracle SE2 – Consolidation Options on x86 Servers vs POWER



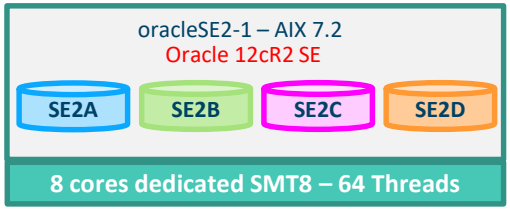


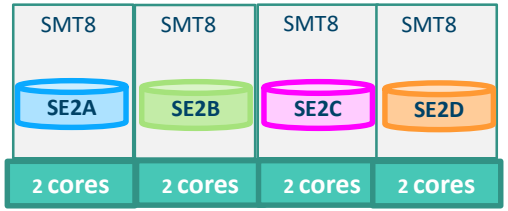
x86
Mapping CPU based on over provisioning

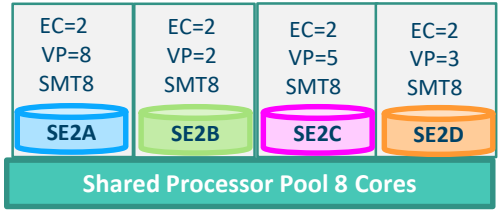
4 vCPUs to 1 physical Core ratio

CPU over commitment ratio too **High**
→ **Performance and stability issues can be raised**

Less DB isolation - - - **Resource control/Security** + + + **More Flexibility**







POWER9 : a Game Changer for Standard Edition 2

- Oracle Standard Edition 2 Licensing – Priced by the number of Sockets
- Oracle SE2 is the main socket based licensing solution – 65% of Oracle DB Customers run Standard Edition
  - 96% savings vs 1 socket 12 cores POWER9 Enterprise Edition ! (list price)
- Higher DB consolidation with SE2 on POWER9 than Intel
- Throughput and threads performance automagically tuned by SMT

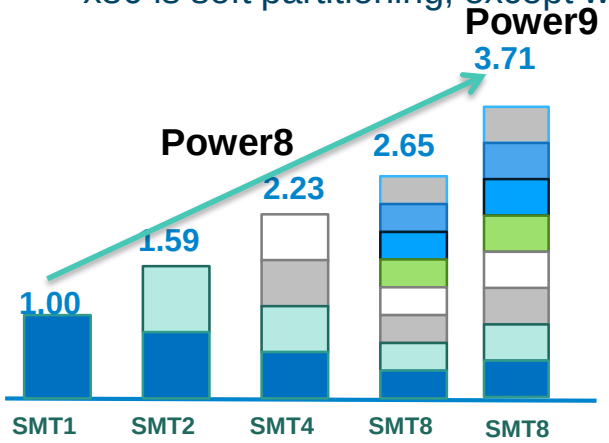
	IBM Power S914 (1ch / 6-core, 256 GB) 48 threads	Xeon Silver 4114 2.2 GHz (1ch / 10co – 192TB) 20 threads	IBM Power S922 (2ch / 16-core, 256 GB) 128 threads	Xeon Gold 6140 2.3 GHz (2ch / 36co – 384GB) 72 threads	IBM Power S922 (2ch / 20-core, 1TB) 160 threads	DL380 Gen10 8180M 28-Core 2.5GHz (2ch / 56co – 1TB) 112 threads
Server price HW Purchase + OS + Virtualization	30,725 $	25,495 $	53,658 $	71,995 $	96,816 $	81,235 $
3Year Maintenance Server + Linux OS + Virtualization	11,161 $	23,092 $	22,256 $	40,403 $	18,784 $	25,506 $
Oracle SE2 License (2 sockets)	$ 17,500	$ 17,500	$ 35,000	$ 35,000	$ 35,000	$ 35,000
3Yr Oracle SWMA (22% of TCA)	$ 11,550	$ 11,550	$ 23,100	$ 23,100	$ 23,100	$ 23,100
Nb of databases	**3**	**1,25**	**8**	**4.5**	**10**	**7**
Total TCO	**70,936 $**	**77,637 $**	**134,014 $**	**170,498 $**	**173,700 $**	**164,841 $**
Cost per database 3 years TCO	**$ 23,645**	**62,109 $**	**16,751 $**	**37,888 $**	**17,370 $**	**23,548 $**

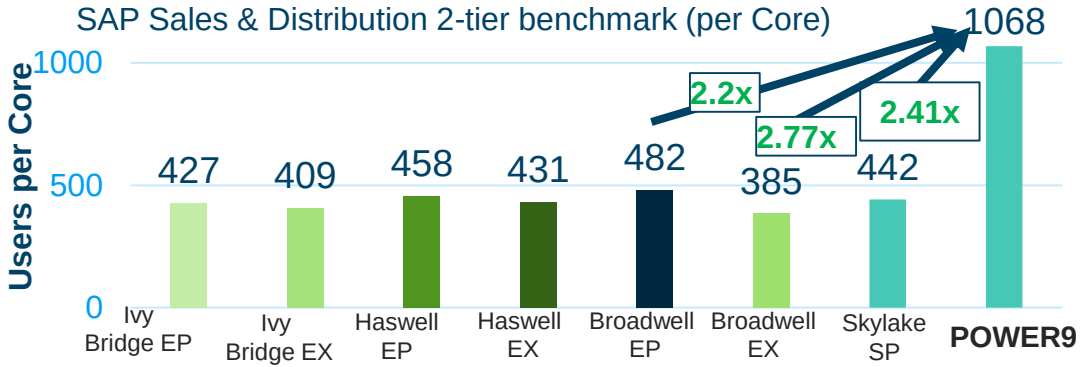
2.6x lessTCO per Database | **2.25x less TCO per Database** | **36% Better Price - Performance**

POWER Systems Performance versus Competitors

Oracle per core Licensing (aka EE)

- **40 to 50% per core performance improvement, POWER9 vs POWER8**
  performance benefit applies to 90% of the TCO (licensing cost)
  - an example of the P8 to P9 performance gain live demo https://bit.ly/2PLSkk1
- **Licensing core factor remains =1**
  - x86 =0.5 license but 2x perf per core and 2x vitualization efficiency
- **Virtualization is “hard partitioning” (pay what you use)**
  - x86 is soft partitioning, except weak OVM Oracle virtualization





Source: https://www.sap.com/dmc/exp/2018-benchmark-directory/#/sd

SAP TPC-C benchmark Comparison 8-sockets - April 2019

Intel Skylake and Cascade Lake – both 28 Cores per Chip

SAP TPC-C benchmark Comparison 8-socket server

Cascade Lake Server (Fujitsu)	**2712 SAPS per Core**
Skylake (Lenovo)	**2658 SAPS per Core**

IBM POWER9 has more than 6800 SAPS per Core

Only +2% more performance with a new CHIP generation
NO performance improvements due more Cores per Chip

Exadata to POWER Systems migration: Customer Case

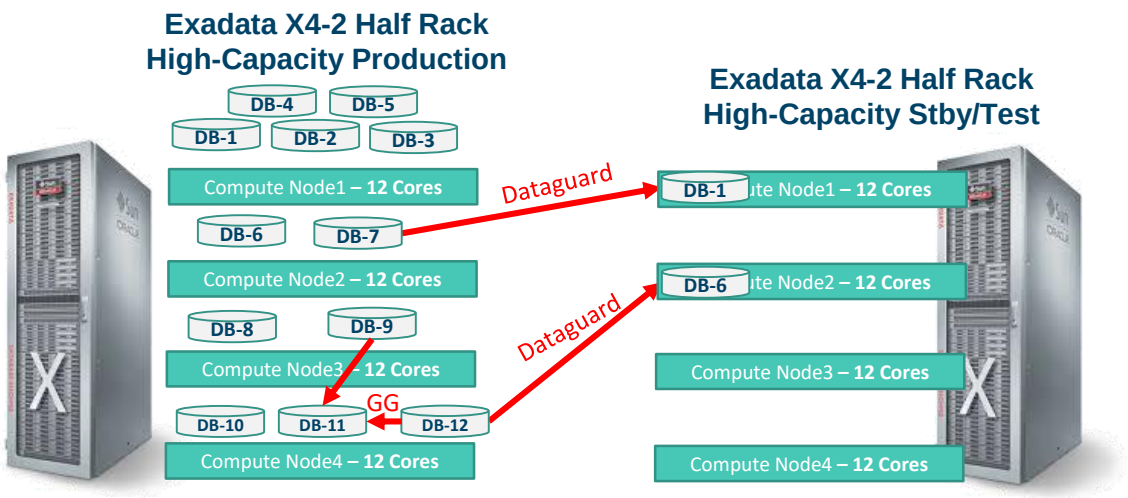
The **largest electrical energy** company in Norway owns, builds and operates critical energy infrastructure. Their most important task is to guarantee a safe and reliable supply of renewable energy to their customers – every hour, all year round. The network's uptime in 2016 was close to 99.99 percent, including both planned and non-planned incidents. Their solution is operated by a Nordic group offering technical consultancy services and IT solutions to customers in the areas of: Energy, Buildings, Industry, Telecom and infrastructure.

Pain Points
- Inflexible solution
- RAC Under Consideration for Scaling purpose
- Out of Storage Capacity
  - Storage Server Expansion is expensive
- End of Support
- Agressive Oracle Sales

Exadata X4

Key Client Benefits:
- Flexibility to scale and grow according to their needs, not limited within vendor packages
- IBM's clear message concerning platform choice and licensing, not fear, uncertainty and doubt (FUD) for license rules or migration difficulties
- Double performance per core and higher utilization reduces cost
- Processor pool for MAX utilization
- Better TCO

Customer - Initial Architecture

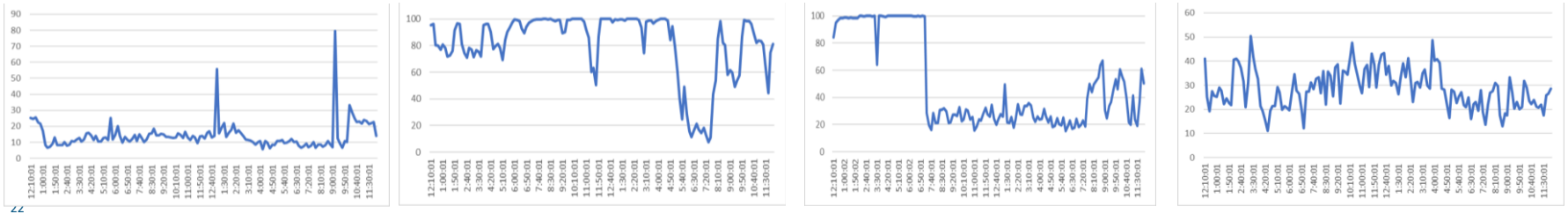


Exadata Pain Points

- Non-Flexible solution
- RAC under Consideration for Scaling purpose
- Out of Storage Capacity
  - Storage Server Expansion is expensive
- End of Support
- Agressive Oracle Sales

Production Exadata CPU Usage per Compute Node

High CPU Usage rate on critical Databases (100%)

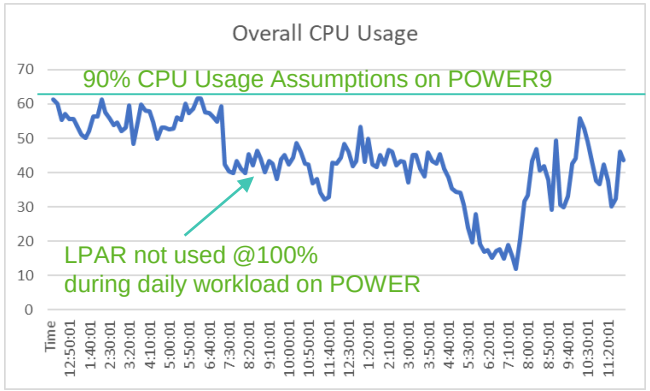


Exadata to POWER Systems Sizing

Exadata X4 ¼ Rack

48 Cores
24 Oracle Licenses

Consolidated View of 4 Compute Nodes
Exadata Usage Rate



Assumptions :

- 90% CPU Usage Assumptions on POWER9 at peaks
- Rest will run in normal situation at 70% in avg

20 Cores POWER9 – E950

20 Oracle Licenses

Exadata Specific Features and Workload Profile

- most Critical Databases have minimal benefit of Exadata secret sauce
- Other Databases do not leverage Exadata
- All Database Instances are very DB Cache friendly, Power Systems has better Memory performance

Exadata AWR Analysis

- On both critical Database, Oracle SQL statement tuning will definitely help in mitigating CPU consumption and thus reduce the number of required cores.

Oracle Databases Consolidation – Customer Case

Current Infrastructure

- Six-nodes RAC Cluster
- R730 2*Intel Xeon CPU E5-2667v4 @3.5Ghz (2ch, 8c)
- 48 Cores x86 – 24 Oracle EE Licenses
- Bare metal servers – no virtualization
- Same Architecture on both Prod and DR Sites

Requirements

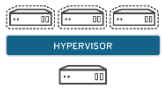
- Target Infra must provide same SLAs
- Best TCA/TCO Solutions

HW Refresh Options

- « bigger » and « newer » x86 .ie 2 or 4 * 24c Cascade Lake Server
- IBM POWER9 High-End E980/E950
- IBM POWER9 Scale-Out S92X

Sizing Estimate based on SAR Data

- Highest peak workload would require 17,9 Power9 cores activation on the proposed E950 configuration
- 99% of the peaks can be covered by 14,3 Power9 cores activation on the proposed E950 configuration
- 24 Core configuration can potentially cover up to > 60% capacity growth in the future



Virtualization

PowerVM vs VMWare

Performance

Standard Benchmarks Power vs x86
IDC QPI & CPM PrecisionIT
Sizing Methodology

License-cost Optimized

License as little as 1 core at a time; Isolate different SW stacks in different LPARs
Shared Processor Pool for licensing optimization and resource usage maximum

Servers

Optimized for Oracle Workloads
SAP HANA Potential Extension

References

x86 to POWER Cases
Power Market Share (Oracle)
Roadmap (19c, POWER10 …)