






**POWER9 Scale-Out & Scale-Up
Performance Review v18b**

- New rPerfs, Spectre/Meltdown, SMT
- Threads, Processor modes, Heat v GHz


Nigel Griffiths
IBM Power Systems: EMEA
Advanced Technology Support



This is not an IBM Announcement, not an
Official IBM Statement, not a Legal Document
It is Nigel's Personal Deck & Nigel's Opinions

nag@uk.ibm.com
@mr_nmon
<https://www.youtube.com/user/nigelgriffiths>



Please note

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM's sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Summary

Nigel is to look at the recently POWER9 rPerf numbers & compare with the updated POWER8 rPerf.

The new rPerf ranges for different over-clocking modes and threading levels - How do they work?

What to expect, if you upgrade POWER8 to POWER9?

Plus tuning the VP count to maximise efficiency and free up processors for other workloads.

POWER9 Performance Review

- or - What IBM forgot to tell clients about POWER9 !

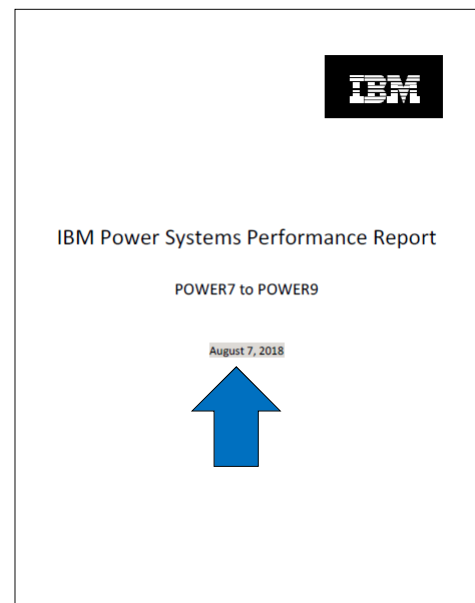
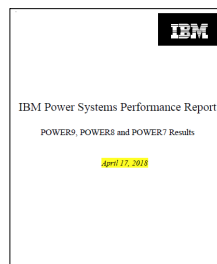
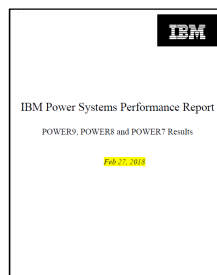
1. Detailed look at the "POWER9 Performance Report"
2. Comments on the Spectre/Meltdown numbers for POWER8
3. Explain the rPerf Ranges and the SMT1 to 8 numbers
4. Single threaded application are dead! **RIP**
5. Setting the POWER9 performance modes plus EnergyScale balancing heat and GHz
6. The "o" word
7. Getting your Server to over heat!
8. How is the POWER9 delivering better performance
9. What to do as you migrate POWER7 or POWER8 to POWER9
10. Monitoring the GHz, plus Temperature and Watts

Feb, April & August 2018

The Must Have Document

Google:

ibm power systems
performance report



https://www.ibm.com/systems/power/hardware/reports/system_perf.html

S914

New version of S914 numbers

If you switch from default **Dynamic Mode** to **Maximum Mode** an extra 9%
 [for reduced noise levels] [for high performance]

Section 2a – AIX Multiuser Performance (rPerf : POWER9) – Non-default Processor Power Mode Setting

All POWER8 and POWER9 results in this table reflect performance with firmware and Operating System updates to mitigate Common Vulnerabilities and Exposures issue numbers CVE-2017-5715, CVE-2017-5753 and CVE-2017-5754 known as Spectre and Meltdown.

Model	Processor / # Cores	Freq. GHz*	Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/ System	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8	Non-default EnergyScale Power Mode Setting
S914	p9/4	2.3 to 3.8	64/64	2/40/-		32.3	54.9	75.7	95.4	Max performance*
S914	p9/6	2.3 to 3.8	64/64	3/60/-		47.3	80.4	110.9	139.8	Max performance*
S914	p9/8	2.8 to 3.8	64/64	4/80/-		68.3	116.1	160.2	201.8	Max performance*

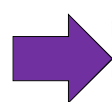
*S914 systems running in maximum performance mode may observe measurably higher sound levels under high utilization.



Section 2 – AIX Multiuser Performance (rPerf : POWER8 and up)

All POWER8 and POWER9 results in this table reflect performance with firmware and Operating System updates to mitigate Common Vulnerabilities and Exposures issue numbers CVE-2017-5715, CVE-2017-5753 and CVE-2017-5754 known as Spectre and Meltdown.

Model	Processor / # Cores	Freq. GHz*	Cache L1 (KB)	Cache L2/L3/L4 (MB)	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
S812	P8/4	3.00	32/64	2/32/128	31.3	45.3	58.9	63.0	
S822	P8/4	3.00	32/64	2/32/128	31.3	45.3	58.9	63.0	
S822	P8/6	3.80	32/64	3/48/128	56.4	81.9	106.4	113.8	
S822	P8/8	4.15	32/64	4/64/128	77.5	112.4	146.1	NA	
S822	P8/10	3.4	32/64	5/80/128	83.1	120.4	156.6	167.5	
S822	P8/8	3.00	32/64	4/64/128	60.9	88.4	114.8	122.9	
S822	P8/12	3.8	32/64	6/96/256	110.0	159.6	207.4	221.9	
S822	P8/16	4.15	32/64	8/128/256	151.1	219.2	284.9	NA	
S822	P8/20	3.4	32/64	10/160/256	161.9	234.8	305.2	326.6	
S922	p9/4	2.8 to 3.8	64/64	2/40/-	30.4	51.6	71.2	89.8	
S922	p9/8	3.4 to 3.9	64/64	4/80/-	68.4	116.3	160.5	202.3	
S922	p9/16	3.4 to 3.9	64/64	8/160/-	133.4	226.9	313.1	394.5	
S922	p9/10	2.9 to 3.8	64/64	5/100/-	74.0	125.7	173.5	218.6	
S922	p9/20	2.9 to 3.8	64/64	10/200/-	144.2	245.2	338.4	426.4	
S814	P8/4	3	32/64	2/32/128	31.3	45.3	58.9	63.0	
S814	P8/6	3	32/64	3/48/128	45.5	66.0	85.8	91.8	
S814	P8/8	3.7	32/64	4/64/128	67.3	97.5	126.7	135.6	
S914	p9/4	2.3 to 3.8	64/64	2/40/-	25.8	43.8	60.4	76.1	
S914	p9/6	2.3 to 3.8	64/64	3/60/-	37.7	64.1	88.5	111.5	
S914	p9/8	2.8 to 3.8	64/64	4/80/-	58.2	98.9	136.5	172.0	
S824	P8/6	3.8	32/64	3/48/128	56.4	81.9	106.4	113.8	
S824	P8/8	4.1	32/64	4/64/128	77.5	112.4	146.1	156.4	
S824	P8/12	3.8	32/64	6/96/256	110.0	159.6	207.4	221.9	
S824	P8/16	4.1	32/64	8/128/256	151.1	219.2	284.9	304.8	
S824	P8/24	3.5	32/64	12/192/256	197.0	285.6	371.3	397.3	
S924	p9/8	3.8 to 4.0	64/64	4/80/-	74.2	126.2	174.1	219.4	
S924	p9/16	3.8 to 4.0	64/64	8/160/-	144.7	246.0	339.5	427.8	
S924	p9/10	3.5 to 3.9	64/64	5/100/-	86.6	147.3	203.3	256.1	
S924	p9/20	3.5 to 3.9	64/64	10/200/-	169.0	287.2	396.4	499.5	
S924	p9/24	3.4 to 3.9	64/64	12/240/-	197.2	335.3	462.7	583.1	



POWER8 new rPerfs allowing for Spectre/Meltdown

If you do the maths it's a **5% to 6%** AVERAGE hit for the mixed rPerf workloads

Reaction: Wow! That is a LOW %

Superior engineering for LARGE "rPerf cocktail" of Workloads on LARGE systems pays off

Nigel's comments & not official IBM wording

IBM's official web page: <https://www.ibm.com/blogs/psirt/ibm-storage-meltdownspectre/>

Actual percentage is application dependant
 Warning: one average number can't represent every workload
 Your application could be better or worse.

All **POWER9** Scale-Out & Enterprise server firmware has Spectre/Meltdown protection

- POWER Users responsibility to check your operating systems version also has the fixes installed
- If you switch off firmware protection then the OS protection is also off

For client cases where top performance regardless of protection is demanded

1. Power off the server
2. Use ASMI to disable protection
Also means OS protection is off
3. Restart the server

Advanced System Management

Log out User ID: P9-S924-red FW910.00 (VL910_073)
admin

Expand all menus
Collapse all menus

- Power/Restart Control
- System Service Aids
- System Information
- System Configuration
 - System Name
 - Configure I/O Enclosures
 - Time Of Day
 - Firmware Update Policy
 - PCI Error Injection Policy
 - HSL Opticonnect Connections
 - I/O Adapter Enlarged Capacity
 - Hardware Management Consoles
 - Floating Point Unit Computation Test
 - Virtual Trusted Platform Module
 - Hypervisor Dispatch Wheel Time
 - PCIe Hardware Topology
 - Hardware Page Table Size
 - Console Type
 - Predictive Dynamic Memory Deallocation
 - High Frequency Trading
 - Speculative Execution Control**
 - Hardware Deconfiguration
 - Program Vital Product Data
 - Service Indicators
 - Power Management
 - Security
- Network Services
- Performance Setup
- On Demand Utilities
- Login Profile

Speculative Execution Control

Current Security Settings :
Speculative execution controls to mitigate user-to-kernel and user-to-user side-channel attacks

This feature is available only when the system is powered off.

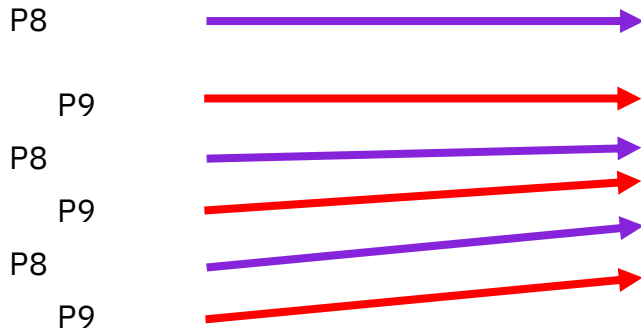
The Option to Switch only appears when the server is powered off

Now

Lets focus on Performance

This is doing my head in!!

Aaaaargh!



Section 2 – AIX Multiuser Performance (rPerf : POWER8 and up)

All POWER8 and POWER9 results in this table reflect performance with firmware and Operating System updates to mitigate Common Vulnerabilities and Exposures issue numbers CVE-2017-5715, CVE-2017-5753 and CVE-2017-5754 known as Spectre and Meltdown

Model	Processor # Cores	Freq. GHz*	Cache L1 (KB)	Cache L2/L3/L4 (MB)	LPAR Size/cores	rPerf S1	rPerf SMT2	rPerf SMT4	rPerf SMT8
S812	P8/4	3.00	32/64	2/32/128		31.3	45.3	58.9	63.0
S822	P8/4	3.00	32/64	2/32/128		31.3	45.3	58.9	63.0
S822	P8/6	3.80	32/64	3/48/128		56.4	81.9	106.4	113.8
S822	P8/8	4.15	32/64	4/64/128		77.5	112.4	146.1	NA
S822	P8/10	3.4	32/64	5/80/128		83.1	120.4	156.6	167.5
S822	P8/8	3.00	32/64	4/64/128		69.9	88.4	114.8	122.9
S822	P8/12	3.8	32/64	6/96/256		110.0	159.6	207.4	221.9
S822	P8/16	4.15	32/64	8/128/256		151.1	219.2	284.9	NA
S822	P8/20	3.4	32/64	10/160/256		161.9	234.8	305.2	326.6
S922	p9/4	2.8 to 3.8	64/64	2/40/-		30.4	51.6	71.2	89.8
S922	p9/8	3.4 to 3.9	64/64	4/80/-		68.4	116.3	160.5	202.3
S922	p9/16	3.4 to 3.9	64/64	8/160/-		133.4	226.9	313.1	394.5
S922	p9/10	2.9 to 3.8	64/64	5/100/-		74.0	125.7	173.5	218.6
S922	p9/20	2.9 to 3.8	64/64	10/200/-		144.2	245.2	338.4	426.4
S814	P8/4	3	32/64	2/32/128		31.3	45.3	58.9	63.0
S814	P8/6	3	32/64	3/48/128		45.5	66.0	85.8	91.8
S814	P8/8	3.7	32/64	4/64/128		67.3	97.5	126.7	135.6
S914	p9/4	2.3 to 3.8	64/64	2/40/-		25.8	43.8	60.4	76.1
S914	p9/6	2.3 to 3.8	64/64	3/60/-		37.7	64.1	88.5	111.5
S914	p9/8	2.8 to 3.8	64/64	4/800/-		58.2	98.9	136.5	172.0
S824	P8/6	3.8	32/64	3/48/128		56.4	81.9	106.4	113.8
S824	P8/8	4.1	32/64	4/64/128		77.5	112.4	146.1	156.4
S824	P8/12	3.8	32/64	6/96/256		110.0	159.6	207.4	221.9
S824	P8/16	4.1	32/64	8/128/256		151.1	219.2	284.9	304.8
S824	P8/24	3.5	32/64	12/192/256		197.0	285.6	371.3	397.3
S924	p9/8	3.8 to 4.0	64/64	4/80/-		74.2	126.2	174.1	219.4
S924	p9/16	3.8 to 4.0	64/64	8/160/-		144.7	246.0	339.5	427.8
S924	p9/10	3.5 to 3.9	64/64	5/100/-		86.6	147.3	203.3	256.1
S924	p9/20	3.5 to 3.9	64/64	10/200/-		169.0	287.2	396.4	499.5
S924	p9/24	3.4 to 3.9	64/64	12/240/-		197.2	335.3	462.7	583.1

Every one converted this to a spreadsheet to analyse the numbers

P8 single GHz

P9 GHz Range

Eh!

What settings decide the top or bottom GHz and so the rPerf?

S824	P8/6	3.8	32
S824	P8/8	4.1	32
S824	P8/12	3.8	32
S824	P8/16	4.1	32
S824	P8/24	3.5	32
S924	p9/8	3.8 to 4.0	64
S924	p9/16	3.8 to 4.0	64
S924	p9/10	3.5 to 3.9	64
S924	p9/20	3.5 to 3.9	64
S924	p9/24	3.4 to 3.9	64

Model	Processor / # Cores	Freq. GHz*	Cache L1 (KB)	Cache L2/L3/L4 (MB)	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
S924	p9/8	3.8 to 4.0	64/64	4/80/-		74.2	126.2	174.1	219.4
S924	p9/16	3.8 to 4.0	64/64	8/160/-		144.7	246.0	339.5	427.8
S924	p9/10	3.5 to 3.9	64/64	5/100/-		86.6	147.3	203.3	256.1
S924	p9/20	3.5 to 3.9	64/64	10/200/-		169.0	287.2	396.4	499.5
S924	p9/24	3.4 to 3.9	64/64	12/240/-		197.2	335.3	462.7	583.1

*POWER9 frequency is expressed from Typical to Max GHz

This End of Table comment is not actually explained any where!
What decides the GHz that your server is running at?

“ST” means **Single Threaded**

Model	Processor / # Cores	Freq. GHz*	Cache L1 (KB)	Cache L2/L3/L4 (MB)	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
-------	---------------------	------------	---------------	---------------------	------------------	----------	------------	------------	------------

Now we get rPerf's for different threading levels (SMT=1, 2, 4 or 8)

- This is new and frankly confusing

IBM had rPerfs for different SMT setting for many years for older HW but did not generally share them

The low thread count = low rPerf numbers are fairly normal.

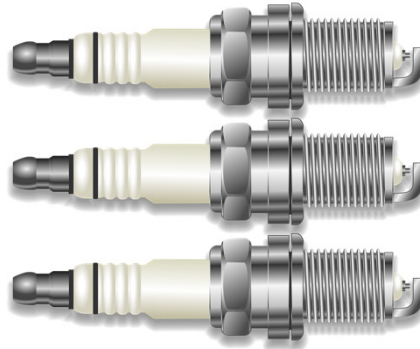
Hopefully, avoid some tricky situations.

Analogy

Take out all but one spark-plug on your:

- BMW Mini &
- BMW 7 series

Then compare the car?



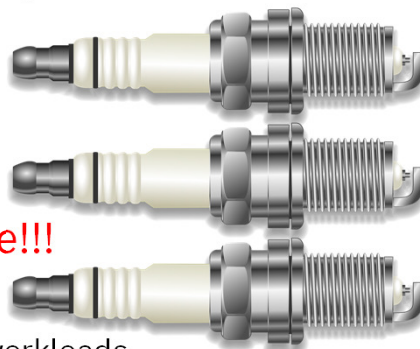
Analogy

Take out all but one spark-plug on your:

- BMW Mini &
- BMW 7 series

Then compare them?

Yes, both cars are terrible!!!



The same with single-threaded workloads

1. These are also terrible workloads
2. We have known this for 25 years

Mr Pessimistic 

rPerf prediction single threaded

POWER8 S824 16 core 4.1 GHz = 151 rPerf (SMT8=304.8)

POWER9 S924 16 core 4.0 GHz = 144 rPerf (SMT8=427.8)

→ 5% down

-5%

EK!!!

Can we demo that in a "micro benchmark"?

1. Set application running on POWER8
2. Then Live Partition Mobility to POWER9 (still in POWER8 mode)
3. Does the application speed up or slow down!!!

LPM POWER8 to POWER9

```

Updates/s=135 @@@
SerialNum=F9D494
LPM-Num =14
LPM-Name=brass4-1a82b37e-00000027
  
```

POWER8

Serial Number & colour
 Change as LPM happens
 135 Worm movements/s
 Predicted 125.5 per second on POWER9

Results on beta HW

May differ in the GA releases

LPM POWER8 to POWER9

```

Updates/s=135 @@@
SerialNum=F9D494
LPM-Num =14
LPM-Name=brass4-1a82b37e-00000027
  
```

POWER8

Serial Number & colour
 Change as LPM happens
 135 Worm movements/s
 Predicted 125.5 per second on POWER9

Jump from 135 to 192 = up **42%**

Results on beta HW

May differ in the GA releases

Worms is a **VERY** unusual app!

```

Updates/s=192
SerialNum=C00190
LPM-Num =14
LPM-Name=brass4-1a82b37e-00000027
  
```

POWER9

LPM POWER8 to POWER9

Don't Panic!

POWER9 is a BIG performance jump

For Multi-threaded applications as promised

Even at slightly reduced GHz for these initial models

Mr Optimistic 😊

		rPerf SMT8
S824	P8/6	113.8
S824	P8/8	156.4
S824	P8/12	221.9
S824	P8/16	304.8
S824	P8/24	397.3
S924	p9/8	219.4
S924	p9/16	427.8
S924	p9/10	256.1
S924	p9/20	499.5
S924	p9/24	583.1

$(583.1-397.3)/397.3*100$
+46.7656% Scientific

+47% **Techie**
+50% **Marketing**

Don't Panic!

+47%

EnergyScale
balancing:
GHz and Heat

Next we explain the
Performance mode
and the GHz

Energy & CPU GHz Balance



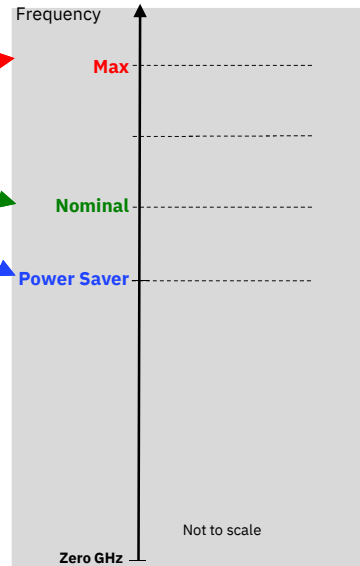
Important Frequencies

Max

Nominal – Fixed normal GHz

Power Saver – Fixed reduce GHz
– Reduces electrical power use = saves money

Zero GHz – The server is powered off!



Only for Technical Customers with P6, P7 or P8 using Dynamic (Favour performance) Mode

POWER7/8 Energy & CPU GHz Balance

POWER8

Important Frequencies

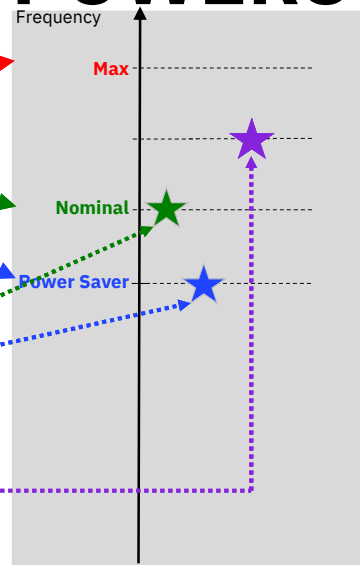
Max

Nominal – Fixed normal GHz

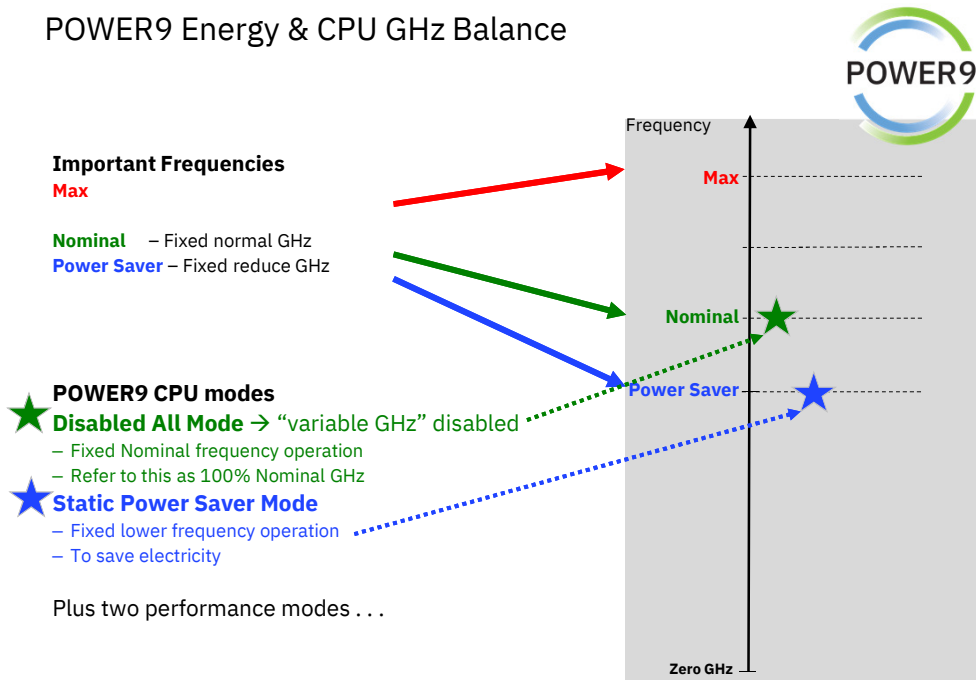
Power Saver – Fixed reduce GHz
– Reduces electrical power use = saves money

With POWER7 and POWER8

- ★ **Disabled All Mode** → “variable GHz” disabled
– Most servers ran at Nominal GHz
- ★ **Static Power Saver Mode**
– Rarely used – it reduces electricity use = reduced costs
- ★ Some used **Dynamic Favour Performance Mode**
– Which allows variable GHz



POWER9 Energy & CPU GHz Balance



POWER9 Energy & CPU GHz Balance

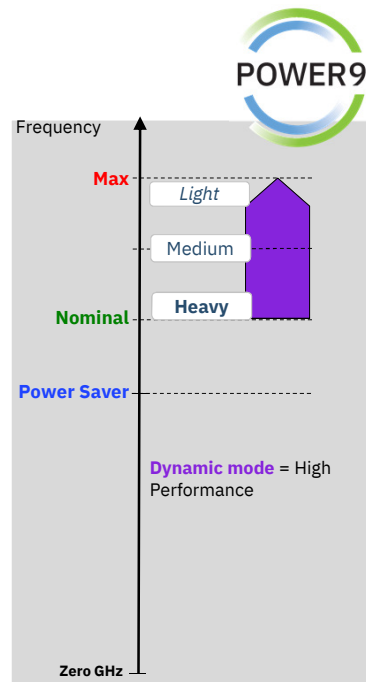


Dynamic Performance Mode

- Near **Max** GHz
 - Workload
 - Light core use or many cores idle
- Middle of the range GHz
 - Workload
 - Medium workload [for example 40% - 60% CPU Utilisation]
 - Some cores idle
- Near **Nominal** GHz
 - Workload
 - Worst case heavy workload or all cores very busy



Not so often at this end
Often at this end



POWER9 Energy & CPU GHz Balance



Maximum Performance Mode

Maintain top frequencies under heavy workloads

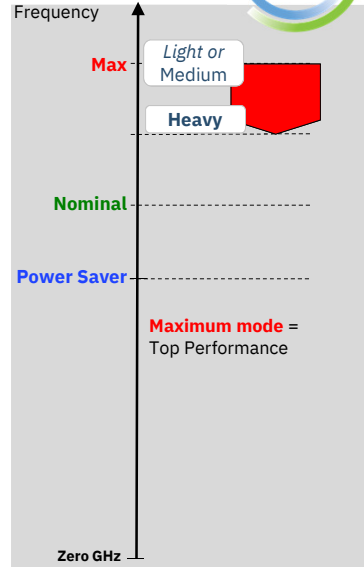
- Near **Max** GHz
Workloads **Light or Medium**
Light or Medium core use or few cores idle
- Middle of the range GHz
Workload **Heavy**
Only for the busiest workloads
- Below **Nominal** – Never
- Stays at Max GHz when idle, for sudden peak fast response
- Excellent for performance in “lights out” computers room
- May increase fan speed (noise) & electricity used



Often at this end

Not so often at this end

Note: If room temperature > 27 C then the GHz is reduced, as necessary to Nominal



POWER9 Energy & CPU GHz Balance

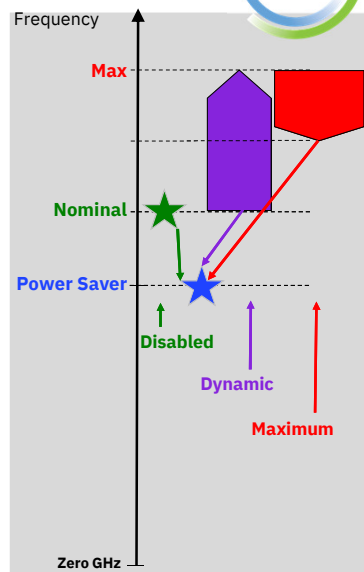
Totally independent selectable feature from the four modes

Idle Power Save = on / off

Different setting on the HMC ASMI Menu*
When sure CPUs are “idle” - it lower CPU GHz to save electricity

- ★ **Static Power Saver Mode**
 - Already at lower GHz
- ★ **Disabled All Mode** → “variable frequency” disabled
 - Lower GHz when System Idle (after many seconds)
- ◆ **Dynamic Performance Mode**
 - Lower GHz when Socket Idle (for fraction of a second)
- ◆ **Maximum Performance Mode**
 - Lower GHz when System Idle (after many seconds)

Don't fiddle with the other Idle Power Save settings unless told to by IBM Support

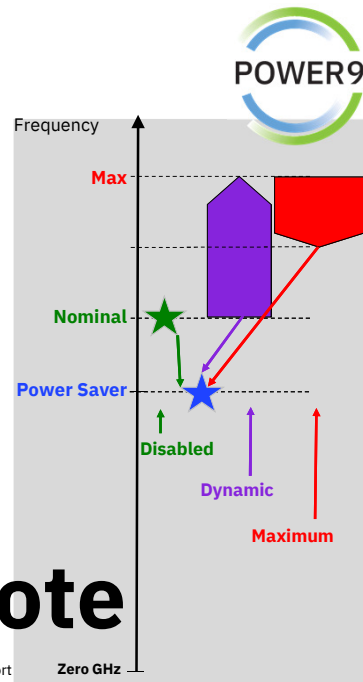


POWER9 Energy & CPU GHz Balance

Idle Power Save = on / off

Different setting on the HMC ASMI Menu*
When sure CPUs are "idle" - it lower CPU GHz to save electricity

- ★ **Static Power Saver Mode**
 - Already at lower GHz
- ★ **Disabled All Mode** → "variable frequency" disabled
 - Lower GHz when System Idle (after many seconds)
- 🏠 **Dynamic Performance Mode**
 - Lower GHz when Socket Idle (for fraction of a second)
- 🔴 **Maximum Performance Mode**
 - Lower GHz when System Idle (after many seconds)



Don't fiddle with the other "Idle Power Save settings" unless told to by IBM Support

POWER9 Energy & CPU GHz Balance

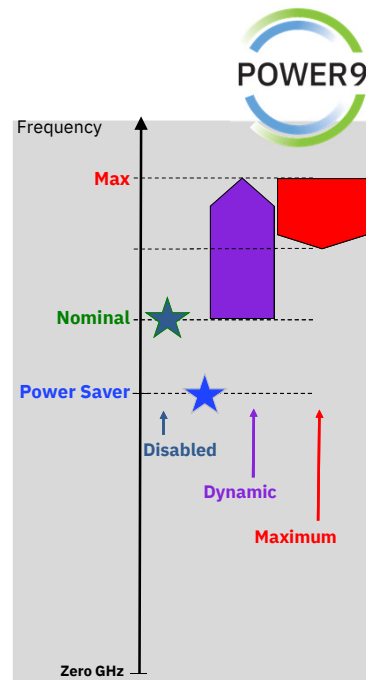
What is the default mode for each POWER9 Scale-Out Model?

- 🔴 S922 → **Maximum Performance mode** (also L922)
- S924 → **Maximum Performance mode**
- E950 → **Maximum Performance mode**
- E980 → **Maximum Performance mode**
- 🏠 S914 → **Dynamic Performance**

Why different?

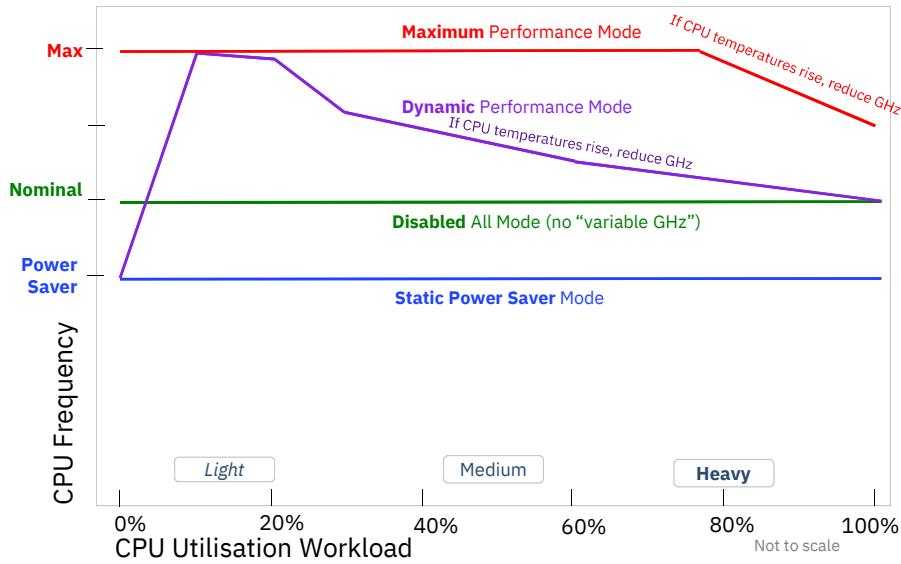
Some S914 are used in the office work space where the extra fan noise may be unacceptable

If in a cool computer room, you can switch to **Maximum Performance** to gain the extra performance





POWER9 EnergyScale CPU Frequency depends on Utilisation Workload



Not to scale
Graph lines are illustrative only i.e. don't go estimating GHz values from percentages

How to set the Performance Modes?

Setting the Performance mode on the HMC

Hardware Management Console

hmc15 Resources > All Systems > P9-S924-brass > General Settings

General Settings

View or change the general and advanced settings for the managed system.

General Properties

View or change the general properties for the managed system. [Learn More](#)

System Name: P9-S924-brass

IP Address: 10.255.128.2

Reference Code:

Machine Type: 9009-42A

Serial Number: 7800190

System Location:

FW910.00 (UL910_038)

Maximum Partitions: 320

Service Partition: Unass

Group Tags:

Description:

Power off when the last logical partition is shut

Migration

System Actions

- Operations
- Power Off
- Power Management
- Schedule Operations
- Launch Advanced System Management**
- Rebuild System
- Change System Password
- Attention LED
- Connections
- Templates
- Updates
- Legacy
- Partitions
- Properties
- General Settings
- Processor, Memory, I/O
- Power VM
- Virtual I/O Services

Advanced System Management

Copyright © 2002, 2011 IBM Corporation. All rights reserved.

P9-S924-brass FW910.00 (UL910_038) Update Access Key Exp Date (YYYY-MM-DD): 2018-03-01

User ID: admin

Password: *****

Language: English

Log in

Session expired.

Welcome

Machine type-model: 9009-42A
Serial number: 7800190
Date: 2018-1-17
Time: 12:32:57 UTC
Service Processor: Primary (Location: U78D2.001.WZS008U-P1-C1)

The User Status table which displays user ids and the status of user ids are only visible to users logged in with at least admin privileges.

Default user: admin
Password: admin

Setting the Performance mode (alternative)

Hardware Management Console

hmc15 Resources > All Systems

All Systems

View and monitor the state, health, and capacity information of all the systems that are connected to the HMC.

Select All

Actions

- View System Partitions
- View System Properties
- View Performance Dashboard...
- Performance Data Collection
- Turn Attention LED Off...
- Schedule Operations...
- Updates
- Serviceable Events Manager...
- Reference Code Log...
- Reset or Remove System Connection...
- View All Actions

Operations

- Power Off
- Power Management
- Schedule Operations
- Launch Advanced System Management (ASM)
- Rebuild System
- Change System Password

Attention LED

- Turn Attention LED Off
- Identify Attention LED
- Test Attention LED

Connections

- Service Processor Status
- Reset or Remove System Connection
- Disconnect another Management Console

Templates

- Deploy System from Template
- Create Partition from Template
- Capture Configuration as Template with Physical I/O
- Capture Configuration as Template without Physical I/O

Updates

- View System Information
- Change Licensed Internal Code for the Current Release to a New Release
- Check System Readiness
- SR-IOV Firmware Update

Legacy

- Partition Availability Priority
- View Workload Management Groups
- Manage System Profiles
- Manage Partition Data
- Restore
- Initialize
- Backup
- Delete
- Utilization Data
- Change Sampling Rate

Advanced System Management

Copyright © 2002, 2011 IBM Corporation. All rights reserved.

P9-S924-brass FW910.00 (UL910_038) Update Access Key Exp Date (YYYY-MM-DD): 2018-03-01

User ID: admin

Password: *****

Language: English

Log in

Session expired.

Welcome

Machine type-model: 9009-42A
Serial number: 7800190
Date: 2018-1-17
Time: 12:32:57 UTC
Service Processor: Primary (Location: U78D2.001.WZS008U-P1-C1)

The User Status table which displays user ids and the status of user ids are only visible to users logged in with at least admin privileges.

Default user: admin
Password: admin

Setting the Performance mode

Can be change online
with immediate effect 😊

Disable all modes
The processor clock frequency will be set to its fixed, nominal value.

Enable Static Power Saver mode
Enabling this feature reduces power consumption by lowering the processor clock frequency and voltage to fixed values. This reduces the power consumption of the system while delivering predictable performance.

Enable Dynamic Performance mode
Enabling this feature causes the processor frequency to vary based on workload and active core count. As the workload/active core count decreases, the processor uses less power, which enables the frequency to be increased above nominal. During periods of very low utilization, the processor frequency will be reduced in order to save energy. This mode provides consistent performance across all environmental operating conditions.

Enable Maximum Performance mode
Enabling this feature causes the processor frequency to vary based on workload and active core count. As the workload/active core count decreases, the processor uses less power, which enables the frequency to be increased above nominal. In this mode, the allowed socket power is increased to the maximum value, which results in top performance along with increased fan noise and higher power consumption. In more stressful environmental conditions, performance may vary.

Setting the Performance mode

Disable all modes
The processor clock frequency will be set to its fixed, nominal value.

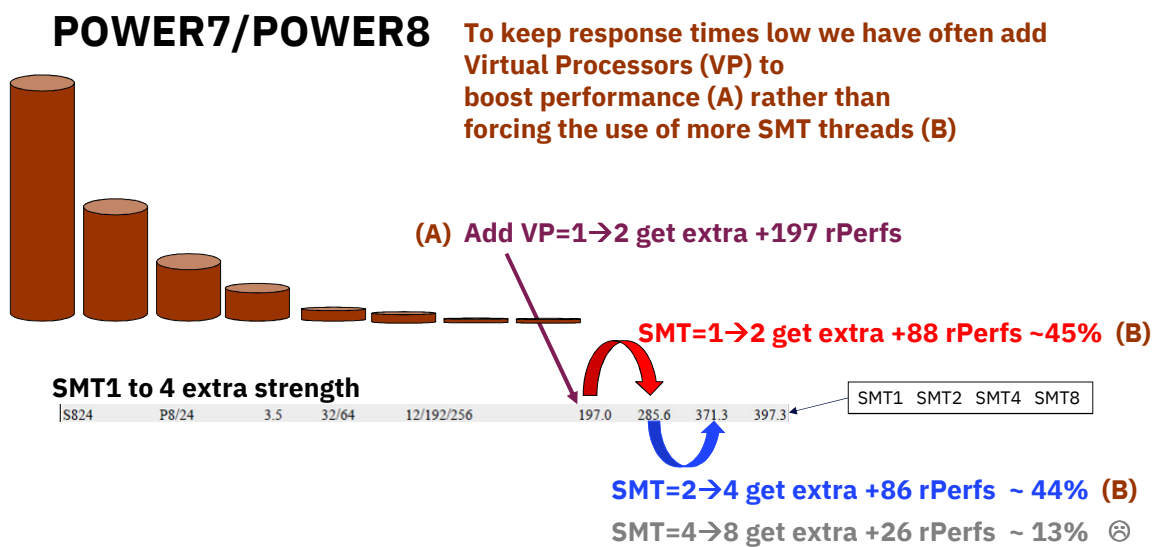
Enable Static Power Saver mode
Enabling this feature reduces power consumption by lowering the processor clock frequency and voltage to fixed values. This reduces the power consumption of the system while delivering predictable performance.

Enable Dynamic Performance mode
Enabling this feature causes the processor frequency to vary based on workload and active core count. As the workload/active core count decreases, the processor uses less power, which enables the frequency to be increased above nominal. During periods of very low utilization, the processor frequency will be reduced in order to save energy. This mode provides consistent performance across all environmental operating conditions.

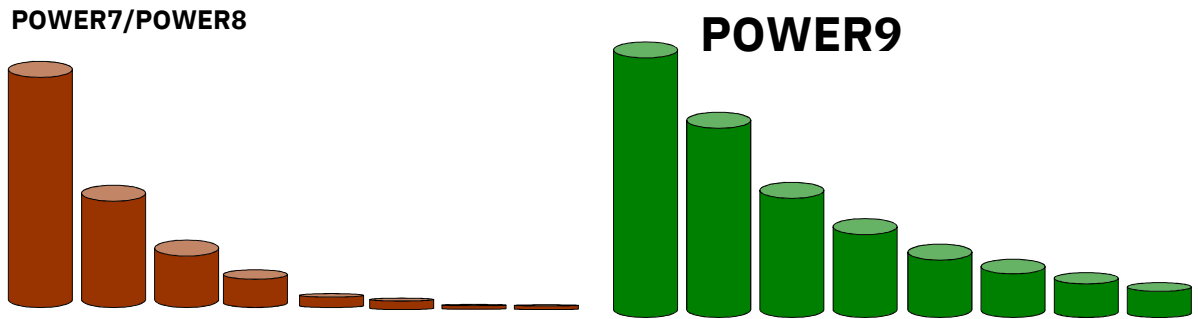
Enable Maximum Performance mode
Enabling this feature causes the processor frequency to vary based on workload and active core count. As the workload/active core count decreases, the processor uses less power, which enables the frequency to be increased above nominal. In this mode, the allowed socket power is increased to the maximum value, which results in top performance along with increased fan noise and higher power consumption. In more stressful environmental conditions, performance may vary.

Lets talk about CPU thread strength

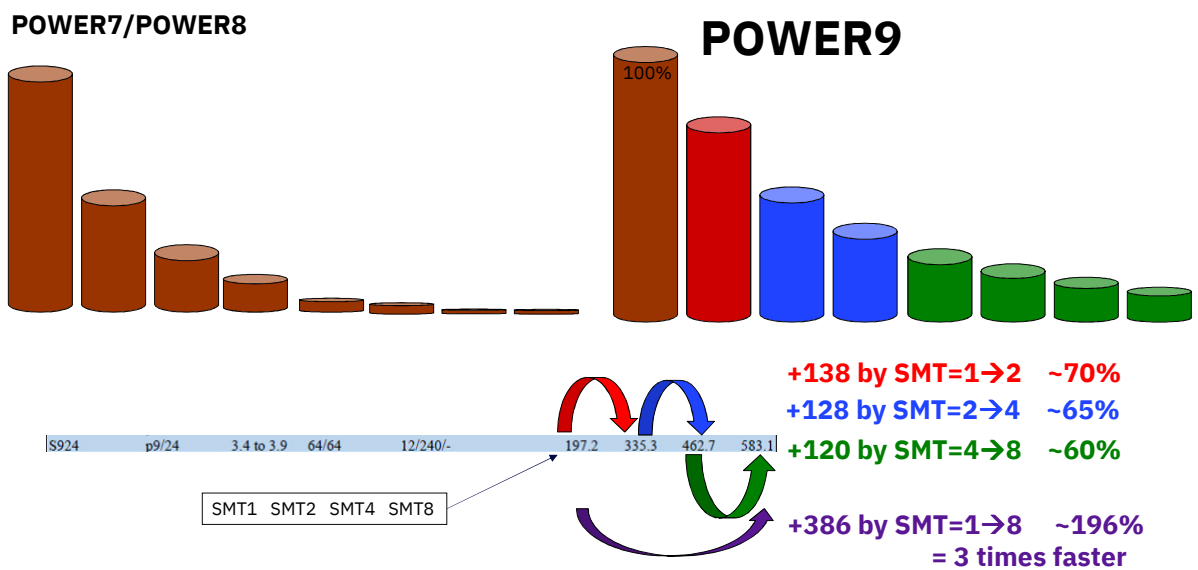
POWER9 thread strength is a primary benefit for higher performance



POWER9 thread strength is a primary benefit for higher performance



POWER9 thread strength is a primary benefit for higher performance



WOW!!
How did IBM
developers do
that?

- 1 Shrunk the chip
- 2 They double the transistor count
- 3 Boosted the bandwidth between CPUs, to memory, to adapters
- 4 Beefed the internal components of the CPU used by SMT
→ see next slide

POWER9 Fused core STRENGTH



POWER8 SMT8 Core Resources

<p>Issue of VSU and AGEN</p> <ul style="list-style-type: none"> • 2x load AGEN / simple-ALU • 2x load/store AGEN • 2x scalar-64b / vector-128b • 2x FXU 	
<p>Vector Scalar Unit (VSU) Pipes</p> <ul style="list-style-type: none"> • 2x FP (64b/128b) + Complex (128b) • 2x ALU (128b) • 2x Permute (128b) • 1x Decimal FP • 1x Cryptography 	
<p>Fixed Point (FXU) Pipes</p> <ul style="list-style-type: none"> • 2x ALU (64b) • 2x FX-MUL + Fixed Divide (64b) 	
<p>Load Store Unit (LSU) Slices</p> <ul style="list-style-type: none"> • 64kB, 8-way Data Cache • Up to 4 DW load or 2 store • 1x Store complete 	

POWER9 SMT8 Core Resources

<p>Issue of VSU and AGEN</p> <ul style="list-style-type: none"> • 8x scalar-64b / 4x vector-128b • 8x load/store AGEN 	
<p>Vector Scalar Unit (VSU) Pipes</p> <ul style="list-style-type: none"> • 8x FP + FX-MUL + Complex (64b slice) • 8x ALU + Simple (64b slice) • 4x Permute (128b) • 4x Quad Fixed (128b) • 4x Fixed Divide (64b) • 2x Quad FP / Decimal FP • 2x Cryptography 	
<p>Load Store Unit (LSU) Slices</p> <ul style="list-style-type: none"> • 64kB, 8-way Data Cache • Up to 8 DW load or store • 2x Store complete 	

x4

x4

x2

x2

x2

What does this mean?

Moving P7 or P8 → P9

REDUCE VP to

RAISE the use of those
powerful threads

“Thread Harvesting”

What does this mean?

Moving P7 or P8 → P9

Recalculate

Entitlement as P9 has
higher rPerfs

“POWER9 Harvesting”

In Practice

When sizing an upgraded
or using Live Partition Mobility
between POWER8 & POWER9

Same principles for
POWER6 or POWER7 → POWER9

Section 2 – AIX Multiuser Performance (rPerf : POWER8 and up)

All POWER8 and POWER9 results in this table reflect performance with firmware and Operating System updates to mitigate Common Vulnerabilities and Exposures issue numbers CVE-2017-5715, CVE-2017-5753 and CVE-2017-5754 known as Spectre and Meltdown.

Model	Processor / # Cores	Freq. GHz*	Cache L1 (KB)	Cache L2/L3/L4 (MB)	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
S812	P8/4	3.00	32/64	2/32/128		31.3	45.3	58.9	63.0
S822	P8/4	3.00	32/64	2/32/128		31.3	45.3	58.9	63.0
S822	P8/6	3.80	32/64	3/48/128		56.4	81.9	106.4	113.8
S822	P8/8	4.15	32/64	4/64/128		77.5	112.4	146.1	156.4
S822	P8/10	3.4	32/64	5/80/128		83.1	120.4	156.6	167.5
S822	P8/8	3.00	32/64	4/64/128		60.9	88.4	114.8	122.9
S822	P8/12	3.8	32/64	6/96/256		110.0	159.6	207.4	221.9
S822	P8/16	4.15	32/64	8/128/256		151.1	219.2	284.9	304.8
S822	P8/20	3.4	32/64	10/160/256		161.9	234.8	305.2	326.6
S922	p9/4	2.8 to 3.8	64/64	2/40/-		30.4	51.6	71.2	89.8
S922	p9/8	3.4 to 3.9	64/64	4/80/-		68.4	116.3	160.5	202.3
S922	p9/16	3.4 to 3.9	64/64	8/160/-		133.4	226.9	313.1	394.5
S922	p9/10	2.9 to 3.8	64/64	5/100/-		74.0	125.7	173.5	218.6
S922	p9/20	2.9 to 3.8	64/64	10/200/-		144.2	245.2	338.4	426.4
S814	P8/4	3	32/64	2/32/128		31.3	45.3	58.9	63.0
S814	P8/6	3	32/64	3/48/128		45.5	66.0	85.8	91.8
S814	P8/8	3.7	32/64	4/64/128		67.3	97.5	126.7	135.6
S914	p9/4	2.3 to 3.8	64/64	2/40/-		25.8	43.8	60.4	76.1
S914	p9/6	2.3 to 3.8	64/64	3/60/-		37.7	64.1	88.5	111.5
S914	p9/8	2.8 to 3.8	64/64	4/80/-		58.2	98.9	136.5	172.0
S824	P8/6	3.8	32/64	3/48/128		56.4	81.9	106.4	113.8
S824	P8/8	4.1	32/64	4/64/128		77.5	112.4	146.1	156.4
S824	P8/12	3.8	32/64	6/96/256		110.0	159.6	207.4	221.9
S824	P8/16	4.1	32/64	8/128/256		151.1	219.2	284.9	304.8
S824	P8/24	3.5	32/64	12/192/256		197.0	285.6	371.3	397.3
S924	p9/8	3.8 to 4.0	64/64	4/80/-		74.2	126.2	174.1	219.4
S924	p9/16	3.8 to 4.0	64/64	8/160/-		144.7	246.0	339.5	427.8
S924	p9/10	3.5 to 3.9	64/64	5/100/-		86.6	147.3	203.3	256.1
S924	p9/20	3.5 to 3.9	64/64	10/200/-		169.0	287.2	396.4	499.5
S924	p9/24	3.4 to 3.9	64/64	12/240/-		197.2	335.3	462.7	583.1

Model	Processor / # Cores	Freq. GHz*
-------	---------------------	------------

rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
----------	------------	------------	------------

S824	P8/6	3.8
S824	P8/8	4.1
S824	P8/12	3.8
S824	P8/16	4.1
S824	P8/24	3.5
S924	p9/8	3.8 to 4.0
S924	p9/16	3.8 to 4.0
S924	p9/10	3.5 to 3.9
S924	p9/20	3.5 to 3.9
S924	p9/24	3.4 to 3.9

56.4	81.9	106.4	113.8
77.5	112.4	146.1	156.4
110.0	159.6	207.4	221.9
151.1	219.2	284.9	304.8
197.0	285.6	371.3	397.3
74.2	126.2	174.1	219.4
144.7	246.0	339.5	427.8
86.6	147.3	203.3	256.1
169.0	287.2	396.4	499.5
197.2	335.3	462.7	583.1

S924

Example LPAR:

POWER8 24 cores but mostly SMT=2 use, 80% Util

Model	Processor / # Cores	Freq. GHz*	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
S824	P8/6	3.8	56.4	81.9	106.4	113.8
S824	P8/8	4.1	77.5	112.4	146.1	156.4
S824	P8/12	3.8	110.0	159.6	207.4	221.9
S824	P8/16	4.1	151.1	219.2	284.9	304.8
S824	P8/24	3.5	197.0	285.6	371.3	397.3
S924	p9/8	3.8 to 4.0	74.2	126.2	174.1	219.4
S924	p9/16	3.8 to 4.0	144.7	246.0	339.5	427.8
S924	p9/10	3.5 to 3.9	86.6	147.3	203.3	256.1
S924	p9/20	3.5 to 3.9	169.0	287.2	396.4	499.5
S924	p9/24	3.4 to 3.9	197.2	335.3	462.7	583.1

S924

Example LPAR:

POWER8 24 cores but mostly SMT=2 use, 75% Util

Model	Processor / # Cores	Freq. GHz*	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
S824	P8/6	3.8	56.4	81.9	106.4	113.8
S824	P8/8	4.1	77.5	112.4	146.1	156.4
S824	P8/12	3.8	110.0	159.6	207.4	221.9
S824	P8/16	4.1	151.1	219.2	284.9	304.8
S824	P8/24	3.5	197.0	285.6	371.3	397.3
S924	p9/8	3.8 to 4.0	74.2	126.2	174.1	219.4
S924	p9/16	3.8 to 4.0	144.7	246.0	339.5	427.8
S924	p9/10	3.5 to 3.9	86.6	147.3	203.3	256.1
S924	p9/20	3.5 to 3.9	169.0	287.2	396.4	499.5
S924	p9/24	3.4 to 3.9	197.2	335.3	462.7	583.1

Solution:

POWER9 20 cores but mostly SMT=2 use, 75% Util

S924

Example LPAR:

POWER8 24 cores but mostly SMT=2 use, 75% Util

Model	Processor / # Cores	Freq. GHz*	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
S824	P8/6	3.8	56.4	81.9	106.4	113.8
S824	P8/8	4.1	77.5	112.4	146.1	156.4
S824	P8/12	3.8	110.0	159.6	207.4	221.9
S824	P8/16	4.1	151.1	219.2	284.9	304.8
S824	P8/24	3.5	197.0	285.6	371.3	397.3
S924	p9/8	3.8 to 4.0	74.2	126.2	174.1	219.4
S924	p9/16	3.8 to 4.0	144.7	246.0	339.5	427.8
S924	p9/10	3.5 to 3.9	80.6	147.5	203.3	256.1
S924	p9/20	3.5 to 3.9	169.0	287.2	396.4	499.5
S924	p9/24	3.4 to 3.9	197.2	335.3	462.7	583.1

Solution:

- 1 POWER9 20 cores but mostly SMT=2 use, 75% Util
- 2 POWER9 10 cores but mostly SMT=8 use, 85% Util

11 cores
= 281 rPerf

E950

Model	Processor / # Cores	Freq. GHz	Inst/Data Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/System	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
E850C	p8/16	4.22	32/64	8/128/256	16	156.3	226.6	294.7	315.3
E850C	p8/24	4.22	32/64	12/192/384	24	230.6	334.3	434.6	465.1
E850C	p8/32	4.22	32/64	16/256/512	32	304.8	442.0	574.6	614.8
E850C	p8/20	3.95	32/64	10/160/256	20	183.6	266.2	346.1	370.3
E850C	p8/30	3.95	32/64	15/240/384	30	270.8	392.6	510.5	546.2
E850C	p8/40	3.95	32/64	20/320/512	40	358.0	519.1	674.8	722.0
E850C	p8/24	3.65	32/64	12/192/256	24	205.8	298.4	388.0	415.1
E850C	p8/36	3.65	32/64	18/288/384	36	303.6	440.2	572.3	612.3
E850C	p8/48	3.65	32/64	24/384/512	48	401.4	582.0	756.5	809.5
E950	P9/16	3.6 - 3.8	64/64	8/160/256	16	151.0	256.7	354.2	446.3
E950	P9/20	3.4 - 3.8	64/64	10/200/256	20	179.4	304.9	420.8	530.2
E950	P9/22	3.2 - 3.8	64/64	11/220/256	22	185.9	316.1	436.2	549.6
E950	P9/24	3.15 - 3.8	64/64	12/240/256	24	198.9	338.1	466.5	587.8
E950	P9/32	3.6 - 3.8	64/64	16/320/512	32	294.4	500.6	650.8	870.4
E950	P9/40	3.4 - 3.8	64/64	20/400/512	40	349.8	594.7	820.7	1,034.1
E950	P9/44	3.2 - 3.8	64/64	22/440/512	44	362.6	616.4	850.7	1,071.9
E950	P9/48	3.15 - 3.8	64/64	24/480/512	48	387.8	659.3	909.9	1,146.4

POWER8 Cores=48 SMT=2 ~580 rPerf

POWER9 Cores=40 SMT=2

POWER9 Cores=24 SMT=8 Squeeze VP and E

E950

Model	Processor / # Cores	Freq. GHz	Inst/Data Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/ System	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
E850C	p8/16	4.22	32/64	8/128/256	16	156.3	226.6	294.7	315.3
E850C	p8/24	4.22	32/64	12/192/384	24	230.6	334.3	434.6	465.1
E850C	p8/32	4.22	32/64	16/256/512	32	304.8	442.0	574.6	614.8
E850C	p8/20	3.95	32/64	10/160/256	20	183.6	266.2	346.1	370.3
E850C	p8/30	3.95	32/64	15/240/384	30	270.8	392.6	510.5	546.2
E850C	p8/40	3.95	32/64	20/320/512	40	358.0	519.1	674.8	722.0
E850C	p8/24	3.65	32/64	12/192/256	24	205.8	298.4	388.0	415.1
E850C	p8/36	3.65	32/64	18/288/384	36	303.6	440.2	572.3	612.3
E850C	p8/48	3.65	32/64	24/384/512	48	401.4	582.0	756.5	809.5
E950	P9/16	3.6 - 3.8	64/64	8/160/256	16	151.0	256.7	354.2	446.3
E950	P9/20	3.4 - 3.8	64/64	10/200/256	20	179.4	304.9	420.8	530.2
E950	P9/22	3.2 - 3.8	64/64	11/220/256	22	185.9	316.1	436.2	549.6
E950	P9/24	3.15 - 3.8	64/64	12/240/256	24	198.9	338.1	466.5	587.8
E950	P9/32	3.6 - 3.8	64/64	16/320/512	32	294.4	500.6	690.8	870.4
E950	P9/40	3.4 - 3.8	64/64	20/400/512	40	349.8	594.7	820.7	1,034.1
E950	P9/44	3.2 - 3.8	64/64	22/440/512	44	362.6	616.4	850.7	1,071.9
E950	P9/48	3.15 - 3.8	64/64	24/480/512	48	387.8	659.3	909.9	1,146.4

POWER8 Cores= 48 SMT=4 ~750 rPerf

POWER9 Cores= ~36 SMT=4

POWER9 Cores= ~28 SMT=8 Squeeze VP and E

51

E950

Model	Processor / # Cores	Freq. GHz	Inst/Data Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/ System	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
E850C	p8/16	4.22	32/64	8/128/256	16	156.3	226.6	294.7	315.3
E850C	p8/24	4.22	32/64	12/192/384	24	230.6	334.3	434.6	465.1
E850C	p8/32	4.22	32/64	16/256/512	32	304.8	442.0	574.6	614.8
E850C	p8/20	3.95	32/64	10/160/256	20	183.6	266.2	346.1	370.3
E850C	p8/30	3.95	32/64	15/240/384	30	270.8	392.6	510.5	546.2
E850C	p8/40	3.95	32/64	20/320/512	40	358.0	519.1	674.8	722.0
E850C	p8/24	3.65	32/64	12/192/256	24	205.8	298.4	388.0	415.1
E850C	p8/36	3.65	32/64	18/288/384	36	303.6	440.2	572.3	612.3
E850C	p8/48	3.65	32/64	24/384/512	48	401.4	582.0	756.5	809.5
E950	P9/16	3.6 - 3.8	64/64	8/160/256	16	151.0	256.7	354.2	446.3
E950	P9/20	3.4 - 3.8	64/64	10/200/256	20	179.4	304.9	420.8	530.2
E950	P9/22	3.2 - 3.8	64/64	11/220/256	22	185.9	316.1	436.2	549.6
E950	P9/24	3.15 - 3.8	64/64	12/240/256	24	198.9	338.1	466.5	587.8
E950	P9/32	3.6 - 3.8	64/64	16/320/512	32	294.4	500.6	690.8	870.4
E950	P9/40	3.4 - 3.8	64/64	20/400/512	40	349.8	594.7	820.7	1,034.1
E950	P9/44	3.2 - 3.8	64/64	22/440/512	44	362.6	616.4	850.7	1,071.9
E950	P9/48	3.15 - 3.8	64/64	24/480/512	48	387.8	659.3	909.9	1,146.4

POWER8 Cores= 32 SMT=4 ~570 rPerf

POWER9 Cores=~28 SMT=4 - SMT4 to SMT4 release 4 cores

POWER9 Cores=~23 SMT=8 - SMT4 to SMT8 release 9 cores

52

E950

Model	Processor / # Cores	Freq. GHz	Inst/Data Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/ System	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
E850C	p8/16	4.22	32/64	8/128/256	16	156.3	226.6	294.7	315.3
E850C	p8/24	4.22	32/64	12/192/384	24	230.6	334.3	434.6	465.1
E850C	p8/32	4.22	32/64	16/256/512	32	304.8	442.0	574.6	614.8
E850C	p8/20	3.95	32/64	10/160/256	20	183.6	266.2	346.1	370.3
E850C	p8/30	3.95	32/64	15/240/384	30	270.8	392.6	510.5	546.2
E850C	p8/40	3.95	32/64	20/320/512	40	358.0	519.1	674.8	722.0
E850C	p8/24	3.65	32/64	12/192/256	24	205.8	298.4	388.0	415.1
E850C	p8/36	3.65	32/64	18/288/384	36	303.6	440.2	572.3	612.3
E850C	p8/48	3.65	32/64	24/384/512	48	401.4	582.0	756.5	809.5
E950	P9/16	3.6 - 3.8	64/64	8/160/256	16	151.0	256.7	354.2	446.3
E950	P9/20	3.4 - 3.8	64/64	10/200/256	20	179.4	304.9	420.8	530.2
E950	P9/22	3.2 - 3.8	64/64	11/220/256	22	185.9	316.1	436.2	549.6
E950	P9/24	3.15 - 3.8	64/64	12/240/256	24	198.9	338.1	466.5	587.8
E950	P9/32	3.6 - 3.8	64/64	16/320/512	32	294.4	500.6	690.8	870.4
E950	P9/40	3.4 - 3.8	64/64	20/400/512	40	349.8	594.7	820.7	1,034.1
E950	P9/44	3.2 - 3.8	64/64	22/440/512	44	362.6	616.4	850.7	1,071.9
E950	P9/48	3.15 - 3.8	64/64	24/480/512	48	387.8	659.3	909.9	1,146.4

POWER8 Cores=30 SMT=8 ~540 rPerf

POWER9 Cores=22 SMT=8 - even with SMT8 we release 8 cores

53

E980

Note: LPARs sizes are NOT whole Server

E880C	p8/32	4.35	32/64	16/256/512	32	334.5	485.0	630.6	674.8
E880C	p8/64	4.35	32/64	32/512/1024	32	669.0	970.1	1,261.1	1,349.4
E880C	p8/96	4.35	32/64	48/768/1536	32	1,003.5	1,455.2	1,891.7	2,024.2
E880C	p8/128	4.35	32/64	64/1024/2048	32	1,338.1	1,940.2	2,522.3	2,698.8
E880C	p8/40	4.19	32/64	20/320/512	40	399.8	579.7	753.6	806.4
E880C	p8/80	4.19	32/64	40/640/1024	40	799.6	1,159.3	1,507.1	1,612.6
E880C	p8/120	4.19	32/64	60/960/1536	40	1,199.4	1,739.1	2,260.8	2,419.1
E880C	p8/160	4.19	32/64	80/1280/2048	40	1,599.1	2,318.8	3,014.4	3,225.4
E880C	p8/48	4.00	32/64	24/384/512	48	456.0	661.3	859.7	919.8

Four 48 Core LPARs one on each CEC/node/drawer on 1 to 4 node E980

8 core
10 core
11 core
12 core

Model	Processor / # Cores	Freq. GHz	Inst/Data Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/ System	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
E880C	p8/96	4.00	32/64	48/768/1024	48	912.0	1,322.6	1,719.2	1,839.6
E880C	p8/144	4.00	32/64	72/1152/1536	48	1,368.2	1,983.8	2,578.9	2,759.4
E880C	p8/192	4.00	32/64	96/1536/2048	48	1,824.2	2,645.0	3,438.5	3,679.3
E980	p9/32	3.9 - 4.0	64/64	16/320/512	32	304.8	442.0	574.6	614.8
E980	p9/64	3.9 - 4.0	64/64	32/640/1024	32	609.6	884.0	1,149.2	1,229.6
E980	p9/96	3.9 - 4.0	64/64	48/960/1536	32	914.4	1,326.0	1,723.8	1,844.4
E980	p9/128	3.9 - 4.0	64/64	64/1280/2048	32	1,219.2	1,768.0	2,298.4	2,466.2
E980	p9/40	3.7 - 3.9	64/64	20/400/512	40	399.8	579.7	753.6	806.4
E980	p9/80	3.7 - 3.9	64/64	40/800/1024	40	799.6	1,159.3	1,507.1	1,612.6
E980	p9/120	3.7 - 3.9	64/64	60/1200/1536	40	1,199.4	1,739.1	2,260.8	2,419.1
E980	p9/160	3.7 - 3.9	64/64	80/1600/2048	40	1,599.1	2,318.8	3,014.4	3,225.4
E980	p9/44	3.58 - 3.9	64/64	22/440/512	44	417.6	616.4	809.9	850.7
E980	p9/88	3.58 - 3.9	64/64	44/880/1024	44	835.2	1,232.8	1,619.8	1,711.4
E980	p9/132	3.58 - 3.9	64/64	66/1320/1536	44	1,199.0	1,739.0	2,260.8	2,419.0
E980	p9/176	3.58 - 3.9	64/64	88/1760/2048	44	1,598.7	2,317.8	3,013.6	3,225.2
E980	p9/48	3.55 - 3.9	64/64	24/480/512	48	429.7	730.5	1,008.1	1,270.2
E980	p9/96	3.55 - 3.9	64/64	48/960/1024	48	859.4	1,461.0	2,016.2	2,540.4
E980	p9/144	3.55 - 3.9	64/64	72/1440/1536	48	1,289.1	2,191.5	3,024.2	3,810.6
E980	p9/192	3.55 - 3.9	64/64	96/1920/2048	48	1,718.8	2,922.0	4,032.3	5,080.7

1 x 1270 = 1270
2 x 1270 = 2540
3 x 1270 = 3810
4 x 1270 = 5080

E980

E880C	p8/32	4.35	32/64	16/256/512	32	334.5	485.0	630.6	674.8
E880C	p8/64	4.35	32/64	32/512/1024	32	669.0	970.1	1,261.1	1,349.4
E880C	p8/96	4.35	32/64	48/768/1536	32	1,003.5	1,455.2	1,891.7	2,024.2
E880C	p8/128	4.35	32/64	64/1024/2048	32	1,338.1	1,940.2	2,522.3	2,698.8
E880C	p8/40	4.19	32/64	20/320/512	40	399.8	579.7	753.6	806.4
E880C	p8/80	4.19	32/64	40/640/1024	40	799.6	1,159.3	1,507.1	1,612.6
E880C	p8/120	4.19	32/64	60/960/1536	40	1,199.4	1,739.1	2,260.8	2,419.1
E880C	p8/160	4.19	32/64	80/1280/2048	40	1,599.1	2,318.8	3,014.4	3,225.4
E880C	p8/48	4.00	32/64	24/384/512	48	456.0	661.3	859.7	919.8

4 node E880C
beaten by
3 node E980
= 25% less cores

8 core

10 core

11 core

12 core

Model	Processor / # Cores	Freq. GHz	Inst/Data Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/ System	LPAR Size# cores	rPerf		rPerf	
						ST	SMT2	SMT4	SMT8
E880C	p8/96	4.00	32/64	48/768/1024	48	912.0	1,322.6	1,719.2	1,839.6
E880C	p8/144	4.00	32/64	72/1152/1536	48	1,368.2	1,983.8	2,578.9	2,759.4
E880C	p8/192	4.00	32/64	96/1536/2048	48	1,824.2	2,645.0	3,438.6	3,679.3
E980	p9/32	3.9 - 4.0	64/64	16/320/512	32	307.8	523.3	722.2	910.0
E980	p9/64	3.9 - 4.0	64/64	32/640/1024	32	615.7	1,046.7	1,444.4	1,820.0
E980	p9/96	3.9 - 4.0	64/64	48/960/1536	32	923.5	1,570.0	2,166.6	2,729.9
E980	p9/128	3.9 - 4.0	64/64	64/1280/2048	32	1,231.4	2,093.4	2,888.8	3,639.9
E980	p9/40	3.7 - 3.9	64/64	20/400/512	40	371.5	631.5	871.5	1,098.1
E980	p9/80	3.7 - 3.9	64/64	40/800/1024	40	743.0	1,263.1	1,743.0	2,196.2
E980	p9/120	3.7 - 3.9	64/64	60/1200/1536	40	1,114.5	1,894.6	2,614.5	3,294.3
E980	p9/160	3.7 - 3.9	64/64	80/1600/2048	40	1,486.0	2,526.1	3,486.0	4,392.4
E980	p9/44	3.58 - 3.9	64/64	22/440/512	44	399.7	679.5	937.7	1,181.4
E980	p9/88	3.58 - 3.9	64/64	44/880/1024	44	799.4	1,358.9	1,875.3	2,362.9
E980	p9/132	3.58 - 3.9	64/64	66/1320/1536	44	1,199.0	2,038.4	2,813.0	3,544.3
E980	p9/176	3.58 - 3.9	64/64	88/1760/2048	44	1,598.7	2,717.8	3,750.6	4,725.8
E980	p9/48	3.55 - 3.9	64/64	24/480/512	48	429.7	730.5	1,008.1	1,270.2
E980	p9/96	3.55 - 3.9	64/64	48/960/1024	48	859.4	1,461.0	2,016.2	2,540.4
E980	p9/144	3.55 - 3.9	64/64	72/1440/1536	48	1,289.1	2,191.5	3,024.2	3,810.6
E980	p9/192	3.55 - 3.9	64/64	96/1920/2048	48	1,718.8	2,922.0	4,032.3	5,080.7

E980

E880C	p8/32	4.35	32/64	16/256/512	32	334.5	485.0	630.6	674.8
E880C	p8/64	4.35	32/64	32/512/1024	32	669.0	970.1	1,261.1	1,349.4
E880C	p8/96	4.35	32/64	48/768/1536	32	1,003.5	1,455.2	1,891.7	2,024.2
E880C	p8/128	4.35	32/64	64/1024/2048	32	1,338.1	1,940.2	2,522.3	2,698.8
E880C	p8/40	4.19	32/64	20/320/512	40	399.8	579.7	753.6	806.4
E880C	p8/80	4.19	32/64	40/640/1024	40	799.6	1,159.3	1,507.1	1,612.6
E880C	p8/120	4.19	32/64	60/960/1536	40	1,199.4	1,739.1	2,260.8	2,419.1
E880C	p8/160	4.19	32/64	80/1280/2048	40	1,599.1	2,318.8	3,014.4	3,225.4
E880C	p8/48	4.00	32/64	24/384/512	48	456.0	661.3	859.7	919.8

1 CEC/node/drawer
48 core VM E880C
beaten by
32 core VM E980
= 25% less cores

8 core

10 core

11 core

12 core

Model	Processor / # Cores	Freq. GHz	Inst/Data Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/ System	LPAR Size# cores	rPerf		rPerf	
						ST	SMT2	SMT4	SMT8
E880C	p8/96	4.00	32/64	48/768/1024	48	912.0	1,322.6	1,719.2	1,839.6
E880C	p8/144	4.00	32/64	72/1152/1536	48	1,368.2	1,983.8	2,578.9	2,759.4
E880C	p8/192	4.00	32/64	96/1536/2048	48	1,824.2	2,645.0	3,438.6	3,679.3
E980	p9/32	3.9 - 4.0	64/64	16/320/512	32	307.8	523.3	722.2	910.0
E980	p9/64	3.9 - 4.0	64/64	32/640/1024	32	615.7	1,046.7	1,444.4	1,820.0
E980	p9/96	3.9 - 4.0	64/64	48/960/1536	32	923.5	1,570.0	2,166.6	2,729.9
E980	p9/128	3.9 - 4.0	64/64	64/1280/2048	32	1,231.4	2,093.4	2,888.8	3,639.9
E980	p9/40	3.7 - 3.9	64/64	20/400/512	40	371.5	631.5	871.5	1,098.1
E980	p9/80	3.7 - 3.9	64/64	40/800/1024	40	743.0	1,263.1	1,743.0	2,196.2
E980	p9/120	3.7 - 3.9	64/64	60/1200/1536	40	1,114.5	1,894.6	2,614.5	3,294.3
E980	p9/160	3.7 - 3.9	64/64	80/1600/2048	40	1,486.0	2,526.1	3,486.0	4,392.4
E980	p9/44	3.58 - 3.9	64/64	22/440/512	44	399.7	679.5	937.7	1,181.4
E980	p9/88	3.58 - 3.9	64/64	44/880/1024	44	799.4	1,358.9	1,875.3	2,362.9
E980	p9/132	3.58 - 3.9	64/64	66/1320/1536	44	1,199.0	2,038.4	2,813.0	3,544.3
E980	p9/176	3.58 - 3.9	64/64	88/1760/2048	44	1,598.7	2,717.8	3,750.6	4,725.8
E980	p9/48	3.55 - 3.9	64/64	24/480/512	48	429.7	730.5	1,008.1	1,270.2
E980	p9/96	3.55 - 3.9	64/64	48/960/1024	48	859.4	1,461.0	2,016.2	2,540.4
E980	p9/144	3.55 - 3.9	64/64	72/1440/1536	48	1,289.1	2,191.5	3,024.2	3,810.6
E980	p9/192	3.55 - 3.9	64/64	96/1920/2048	48	1,718.8	2,922.0	4,032.3	5,080.7

E980

E880C
 192 core in SMT=2
 or
 144 core in SMT=4
 beaten by
 all these on E980

96 core SMT=8
 50% less cores

Model	Processor / # Cores	Freq. GHz	Inst/Data Cache L1 (KB) Per core	Cache L2/L3/L4 (MB)/ System	LPAR Size# cores	rPerf ST	rPerf SMT2	rPerf SMT4	rPerf SMT8
E880C	p8/32	4.35	32/64	16/256/512	32	334.5	485.0	630.6	674.8
E880C	p8/64	4.35	32/64	32/512/1024	32	669.0	970.1	1,261.1	1,349.4
E880C	p8/96	4.35	32/64	48/768/1536	32	1,003.5	1,455.2	1,891.7	2,024.2
E880C	p8/128	4.35	32/64	64/1024/2048	32	1,338.1	1,940.2	2,522.3	2,698.8
E880C	p8/40	4.19	32/64	20/320/512	40	399.8	579.7	753.6	806.4
E880C	p8/80	4.19	32/64	40/640/1024	40	799.6	1,159.3	1,507.1	1,612.6
E880C	p8/120	4.19	32/64	60/960/1536	40	1,199.4	1,739.1	2,260.8	2,419.1
E880C	p8/160	4.19	32/64	80/1280/2048	40	1,599.1	2,318.8	3,014.4	3,225.4
E880C	p8/48	4.00	32/64	24/384/512	48	456.0	661.3	859.7	919.8
E980	p8/96	4.00	32/64	48/768/1024	48	912.0	1,322.6	1,719.2	1,839.6
E980	p8/144	4.00	32/64	72/1152/1536	48	1,368.2	1,983.8	2,578.9	2,759.4
E980	p8/192	4.00	32/64	96/1536/2048	48	1,824.2	2,645.0	3,438.6	3,679.3
8 core	1 node E980 p9/32	3.9 - 4.0	64/64	16/320/512	32	307.8	523.3	722.2	910.0
	2 node E980 p9/64	3.9 - 4.0	64/64	32/640/1024	32	615.7	1,046.7	1,444.4	1,820.0
	3 node E980 p9/96	3.9 - 4.0	64/64	48/960/1536	32	923.5	1,570.0	2,166.6	2,729.9
	4 node E980 p9/128	3.9 - 4.0	64/64	64/1280/2048	32	1,231.4	2,093.4	2,888.8	3,639.9
10 core	1 node E980 p9/40	3.7 - 3.9	64/64	20/400/512	40	371.5	631.5	871.5	1,098.1
	2 node E980 p9/80	3.7 - 3.9	64/64	40/800/1024	40	743.0	1,263.1	1,743.0	2,196.2
	3 node E980 p9/120	3.7 - 3.9	64/64	60/1200/1536	40	1,114.5	1,894.6	2,614.5	3,294.3
	4 node E980 p9/160	3.7 - 3.9	64/64	80/1600/2048	40	1,486.0	2,526.1	3,486.0	4,392.4
11 core	1 node E980 p9/44	3.58 - 3.9	64/64	22/440/512	44	399.7	679.5	937.7	1,181.4
	2 node E980 p9/88	3.58 - 3.9	64/64	44/880/1024	44	799.4	1,358.9	1,875.3	2,362.9
	3 node E980 p9/132	3.58 - 3.9	64/64	66/1320/1536	44	1,199.0	2,038.4	2,813.0	3,544.3
	4 node E980 p9/176	3.58 - 3.9	64/64	88/1760/2048	44	1,598.7	2,717.8	3,750.6	4,725.8
50% less cores	1 node E980 p9/48	3.55 - 3.9	64/64	24/480/512	48	429.7	730.5	1,008.1	1,270.2
	2 node E980 p9/96	3.55 - 3.9	64/64	48/960/1024	48	859.4	1,461.0	2,016.2	2,540.4
	3 node E980 p9/144	3.55 - 3.9	64/64	72/1440/1536	48	1,289.1	2,191.5	3,024.2	3,810.6
	4 node E980 p9/192	3.55 - 3.9	64/64	96/1920/2048	48	1,718.8	2,922.0	4,032.3	5,080.7

So how does this relate to the System Performance Report: GHz range and rPerfs ?

Warning: Secret chart next!



POWER9 Scale Out Frequencies & Defaults

Default Mode

Model	Default Mode	Feature Code	Number of Cores	Static Nominal Frequency Disable All mode	Dynamic Performance mode GHz Range	Maximum Performance mode Typical GHz Range
S924 H924	Maximum Performance	EP1G	12 cores	2.75 GHz	2.75 to 3.9 GHz (max)	3.4 to 3.9 GHz (max)
		EP1F	10 cores	2.9 GHz	2.9 to 3.9 GHz (max)	3.5 to 3.9 GHz (max)
		EP1E	8 cores	3.3 GHz	3.3 to 4.0 GHz (max)	3.8 to 4.0 GHz (max)
S914	Dynamic Performance	EP12	8 cores	2.8 GHz	2.8 to 3.8 GHz (max)	3.15 to 3.8 GHz (max)
		EP11	6 cores	2.3 GHz	2.3 to 3.8 GHz (max)	2.8 to 3.8 GHz (max)
		EP10	4 cores	2.3 GHz	2.3 to 3.8 GHz (max)	2.8 to 3.8 GHz (max)
S922 H922	Maximum Performance	EP19	10 cores	2.5 GHz	2.5 to 3.8 GHz (max)	2.9 to 3.8 GHz (max)
		EP18	8 cores	3.0 GHz	3.0 to 3.9 GHz (max)	3.4 to 3.9 GHz (max)
		EP16	4 cores	2.3 GHz	2.3 to 3.8 GHz (max)	2.8 to 3.8 GHz (max)
L922	Maximum Performance	ELPX	12 cores	2.3 GHz	2.3 to 3.8 GHz (max)	2.7 to 3.8 GHz (max)
		EPPW	10 cores	2.5 GHz	2.5 to 3.8 GHz (max)	2.9 to 3.8 GHz (max)
		ELPV	8 cores	3.0 GHz	3.0 to 3.9 GHz (max)	3.4 to 3.9 GHz (max)

Notes:

1. Frequencies outlined in Red reflect the default mode (i.e. frequency range) for that particular system
2. In order to reach maximum frequency, some cores may need to be turned off

This is from the POWER9 EnergyScale - Configuration & Management web page
<https://www.ibm.com/developerworks/community/wikis/home?lang=en-gb#!/wiki/Power%20Systems/page/POWER9%20EnergyScale%20-%20Configuration%20&%20Management>



POWER9 Scale Out Frequencies & Defaults

“unnamed”
Frequency

Model	Default Mode	Feature Code	Number of Cores	Static Nominal Frequency Disable All mode	Dynamic Performance mode GHz Range	Maximum Performance mode Typical GHz Range
S924 H924	Maximum Performance	EP1G	12 cores	2.75 GHz	2.75 to 3.9 GHz (max)	3.4 to 3.9 GHz (max)
		EP1F	10 cores	2.9 GHz	2.9 to 3.9 GHz (max)	3.5 to 3.9 GHz (max)
		EP1E	8 cores	3.3 GHz	3.3 to 4.0 GHz (max)	3.8 to 4.0 GHz (max)
S914	Dynamic Performance	EP12	8 cores	2.8 GHz	2.8 to 3.8 GHz (max)	3.15 to 3.8 GHz (max)
		EP11	6 cores	2.3 GHz	2.3 to 3.8 GHz (max)	2.8 to 3.8 GHz (max)
		EP10	4 cores	2.3 GHz	2.3 to 3.8 GHz (max)	2.8 to 3.8 GHz (max)
S922 H922	Maximum Performance	EP19	10 cores	2.5 GHz	2.5 to 3.8 GHz (max)	2.9 to 3.8 GHz (max)
		EP18	8 cores	3.0 GHz	3.0 to 3.9 GHz (max)	3.4 to 3.9 GHz (max)
		EP16	4 cores	2.3 GHz	2.3 to 3.8 GHz (max)	2.8 to 3.8 GHz (max)
L922	Maximum Performance	ELPX	12 cores	2.3 GHz	2.3 to 3.8 GHz (max)	2.7 to 3.8 GHz (max)
		EPPW	10 cores	2.5 GHz	2.5 to 3.8 GHz (max)	2.9 to 3.8 GHz (max)
		ELPV	8 cores	3.0 GHz	3.0 to 3.9 GHz (max)	3.4 to 3.9 GHz (max)

↑
Six Models

↑
Nominal Frequency

↑
Max Frequency



POWER9 Enterprise Frequencies & Defaults

Model	Default Mode	Feature Code	Number of Cores	Static Nominal Frequency Disable All mode	Dynamic Performance mode GHz Range	Maximum Performance mode Typical GHz Range
E950	Maximum Performance	EPWT	12cores	2.8 GHz	2.8 to 3.8 GHz (max)	3.15 to 3.8 GHz (max)
		EPWY	11 cores	2.85 GHz	2.85 to 3.8 GHz (max)	3.2 to 3.8 GHz (max)
		EPWS	10 cores	3.0 GHz	3.0 to 3.8 GHz (max)	3.4 to 3.8 GHz (max)
		EPWR	8 cores	3.3 GHz	3.3 to 3.8 GHz (max)	3.6 to 3.8 GHz (max)
E980	Maximum Performance	EFB3	12cores	2.9 GHz	2.9 to 3.9 GHz (max)	3.55 to 3.9 GHz (max)
		EFB4	11 cores	3.0 GHz	3.0 to 3.9 GHz (max)	3.58 to 3.9 GHz (max)
		EFB2	10 cores	3.15 GHz	3.15 to 3.9 GHz (max)	3.7 to 3.9 GHz (max)
		EFB1	8 cores	3.4 GHz	3.4 to 4.0 GHz (max)	3.9 to 4.0 GHz (max)

“unnamed” Frequency

Nominal Frequency

Max Frequency

On Twitter?



```
# lparstat -E 1 3
System configuration: type=Shared mode=Uncapped smt=8 lcpu=16 mem=16384MB ent=1.00
Physical Processor Utilisation:
-----Actual-----
user  sys  wait  idle  freq  vmm  sys  wait  idle
1.126 0.016 0.000 0.000 3.7GHz[113%] 1.272 0.018 0.000 0.000
1.127 0.016 0.000 0.000 3.7GHz[113%] 1.274 0.018 0.000 0.000
1.124 0.016 0.000 0.000 3.7GHz[113%] 1.271 0.018 0.000 0.000
#
# lsattr -El proc0
frequency 3900000000 Processor Speed False
smt_enabled true Processor SMT enabled False
smt_threads 8 Processor SMT threads False
state enable Processor state False
type PowerPC_POWER9 Processor type False
#
```

3.7 GHz = 13% Over-clocking (oops!)

3.3 billion Hz = 3.3 GHz Nominal

P9 GHz **part 1:** #POWER9 servers in practice run at (max) ~3.8-4 GHz, other server chips eat our dust! I see: normal GHz + overlocking, I am told to not use the "o" word, oops! #EnergyScale guys say run full speed but will lower GHz, if getting hot like your air-conditioning fails!

P9 GHz **part 2:** #POWER9 servers **How to get too hot!** If you don't have: max CPU count + max memory size + max disks + max high-speed adapter AND max server workload + computer room is warm then your server may never get too hot and still be at that (max) GHz. I know as I tried!!!

P9 GHz **part 3:** One quirk on AIX: commands like lparstat -E 1 9 report the varying current GHz but others report the non-overlocking (oops!) GHz value called Nominal So don't worry is you buy 3.9 GHz but nmon or lsattr -El proc0 reports a lower Nominal GHz between 2.3 to 3.3 GHz

On the
S922/S924/E950/E980:

What can we find out
what is happening?

Applies to the other models too.

AIX - POWER9 Nominal Frequency

Ksh script:

```
machine=$(lsattr -El sys0 -a modelname -F value)
cpus=$(lsdev -Cc processor | grep Available | wc -l | sed 's/ //g')
procstr=$(lsdev -Cc processor | grep Available | head -1 | cut -d' ' -f1)
proctype=$(lsattr -El $procstr | awk '/^type/ {print $2}')
Hz=$(lsattr -El ${procstr%% *} -a frequency -F value)
echo $machine cpu=$cpus type=$proctype Hz=$Hz
Output: IBM,9009-42A cpu=8 type=PowerPC_POWER9 Hz=3234000000
3.2 GHz
```

These code lines are from the publicly available **rperf** korn shell script

Or

```
# lsattr -El proc0      ← Note: not all LPARs have a proc0!!
frequency 3234000000    Processor Speed      False
smt_enabled true       Processor SMT enabled False
smt_threads 8          Processor SMT threads False
state      enable       Processor state      False
type       PowerPC_POWER9 Processor type        False
```


AIX - POWER9 Current Frequency Monitoring

Usage: `lparstat -E [Interval [Count]]`

- Reports the current CPU frequency averaged for the LPAR
- **`lparstat -E 1 999`**

Usage: `mpstat -E [Interval [Count]]`

- Reports the current CPU frequency per Virtual Processor
- **`mpstat -E 1 999`**

Without the *Interval* parameter they report a single statistic since the last LPAR boot

There appears to be no tooling to capture to logs
Nominal GHz, Current GHz or percentage (nsp)
nor the Performance mode (Dynamic mode or Maximum mode)

AIX - POWER9 Current Frequency Monitoring Examples

```
# lparstat -E 1 333
System configuration: type=Shared mode=Uncapped smt=8 lcpu=16 mem=16384MB ent=1.00
Power=Dynamic-Performance
Physical Processor Utilisation:
-----Actual-----
user  sys  wait  idle  freq  user  sys  wait  idle
-----
0.323 0.003 0.000 0.674 3.9GHz [119%] 0.386 0.004 0.000 0.611
0.323 0.003 0.000 0.674 3.9GHz [119%] 0.385 0.003 0.000 0.611
0.323 0.003 0.000 0.674 3.9GHz [119%] 0.385 0.003 0.000 0.611
...
Current GHz [Overclocking]
```

**Not good for long term data
capturing or graphing**

```
# mpstat -E 1 444
System configuration: lcpu=16 ent=1.0 mode=Uncapped
vcpu      pbusy      physc      freq      scaled physc
-----
0          0.3264 [ 33%] 0.9981 [100%] 3.9GHz [119%] 1.1906 [119%]
8          0.0000 [ 0%] 0.0003 [ 0%] 3.9GHz [119%] 0.0003 [ 0%]
ALL        0.3264 [ 33%] 0.9983 [100%] 3.9GHz [119%] 1.1909 [119%]
...
Current GHz [Overclocking]
```

AIX - POWER9 Current Frequency for logging

1) nmon does not log current CPU MHz/GHz stats **only Nominal MHz at the start**
– but nor does anything else!

2) Best I could find was: lparstat -X -o lparstat.xml -E 1 999

```
then grep "<nsp>" lparstat.xml
Output: <nsp>109</nsp>
        <nsp>109</nsp>
        <nsp>112</nsp>
```

nsp = Nominal Speed Percentage – multiply by Nominal GHz → current GHz

*Very Embarrassing
This is being fixed*

3) Nigel's new data collector "njmon" that outputs ~650 stats in JSON format

```
# ./njmon -s2 -c3 | grep -I _mhz
  "nominal_mhz": 3234.0,
  "current_mhz": 3529.6
  "nominal_mhz": 3234.0,
  "current_mhz": 3529.2
  "nominal_mhz": 3234.0,
  "current_mhz": 3528.8
```

Gets the GHz from libperfstat library

Use Python to load JSON into
Python dictionary then

inject in to online graphing tools:

InfluxDB + Grafana or Splunk

<http://nmon.sourceforge.net/pmwiki.php?n=Site.Njmon>

Linux - POWER9 Frequency Monitoring in Native non-Virtualised AC922/LC922

EnergyScale status can be obtained from dmesg:

```
# dmesg|grep cpufreq
[ 2.003516] powernv-cpufreq: cpufreq pstate min 91 nominal 55 max 0
[ 2.003597] powernv-cpufreq: Workload Optimized Frequency is enabled in the platform
```

i.e. not PowerVM

Nominal frequency range

```
# cat /sys/devices/system/cpu/cpu0/cpufreq/scaling_available_frequencies
3283000 ...
```

Full Frequency range

```
# cat /sys/devices/system/cpu/cpu0/cpufreq/scaling_boost_frequencies
3800000 ..
```

Current running frequency of any core:

```
# cat /sys/devices/system/cpu/cpu0/cpufreq/cpuinfo_cur_freq
2316000
```

Test max frequency achieved in the system:

```
# ppc64_cpu --frequency
min: 3.776 GHz (cpu 143)
max: 3.777 GHz (cpu 73)
avg: 3.777 GHz
```

Use **cpupower** tool to query and set frequency

available frequency steps from cpupower will list only the nominal range, but user can select full frequency range to set and it will take effect.

Linux - POWER9 Frequency Monitoring for PowerVM

EnergyScale status can be obtained from dmesg:

```
# $ dmesg|grep MHz
[ 0.000000] time_init: decrementer frequency = 512.000000 MHz
[ 0.000000] time_init: processor frequency = 3234.000000 MHz ← Nominal MHz
```

Or

```
grep clock /proc/cpuinfo
clock      : 3234.000000MHz
clock      : 3234.000000MHz
clock      : 3234.000000MHz
...
```

nmon for Linux does the best it can from available data

IBM i - POWER9 Frequency Monitoring

IBM iDoctor for IBM i

IBM iDoctor for IBM i displays the CPU rate for the IBM i partition over time on the Collection Overview graph. The CPU rate for the partition is the ratio of scaled to unscaled processor utilized time, expressed as a percentage. The processor utilized time is the accumulation of non-idle virtual processor SPURR and PURR over each time interval.

WRKSYSACT

The Work with System Activity (WRKSYSACT) command displays the Average CPU rate since last refresh for the partition in output shown on the display station. The Average CPU rate for the partition is the ratio of scaled to unscaled processor utilized time, expressed as a percentage. The processor utilized time is accumulation of non-idle virtual processor SPURR and PURR for the interval since the last refresh.

IBM i Collection Services

Database file QAPMJOBMI contains time series data by task, primary thread, and secondary thread. Scaled and unscaled CPU times, both charged and used, are available to calculate average CPU rate for processing activity of tasks and threads.

Database file QAPMSYSTEM contains time series system-wide (i.e. partition) accumulations of performance data. Scaled and unscaled CPU times are accumulated for various categories of processor usage. The ratio of scaled to unscaled time is the average CPU rate for the category of time accumulation. The processor utilized time is accumulation of non-idle virtual processor SPURR and PURR for the time interval.

Note: As of IBM i 7.3, the QAPMCONF database file key "NF" contains the processor nominal frequency in MHz. The processor nominal frequency can be used to convert average CPU rate to average processor frequency.

In practice

S924 16 CPU (8 core + 8 core) Room temp=22C
 + 256 GB RAM + 8 disks + 4 simple adapters
 - 3.3GHz nominal, 3.5GHz unnamed, 4GHz max

Wow!!! That is a low GHz = under-clocking

- I check and Idle-Power-Saver=On !!
- 2 VIOS + 1 LPAR very idle

Start a workload

- Quick rise to 3.9 GHz (nearly max GHz)
- LPAR has VP=2 out of 16 CPU **19% overclock**
- Utilisation% low: Hinting not using SMT=8

Switch off Idle-Power-Saving

- Mode changed (Note: name should be Max.)

Stop workload

- CPU stay at Max GHz=Max mode (no saving)

No matter the workload - server runs at Max GHz
 = does not clock down because it is not hot

```
# lparstat -E 1 3
System configuration: type=Shared mode=Uncapped smt=8 lcpu=16
mem=16384MB ent=1.00 Power=Static

Physical Processor Utilisation:
-----Actual-----
user  sys  wait  idle   freq  user  sys  wait  idle
-----Normalised-----
0.324 0.004 0.000 0.673 2.3GHz [ 69%] 0.222 0.002 0.000 0.776
0.323 0.003 0.000 0.673 2.3GHz [ 69%] 0.222 0.002 0.000 0.776
0.324 0.003 0.000 0.674 2.3GHz [ 69%] 0.222 0.002 0.000 0.776

Start CPU hungry programs
user  sys  wait  idle   freq  user  sys  wait  idle
-----Normalised-----
0.324 0.003 0.000 0.673 3.7GHz [111%] 0.368 0.003 0.000 0.628
0.323 0.003 0.000 0.673 3.9GHz [119%] 0.385 0.003 0.000 0.611
0.323 0.003 0.000 0.674 3.9GHz [119%] 0.385 0.003 0.000 0.611

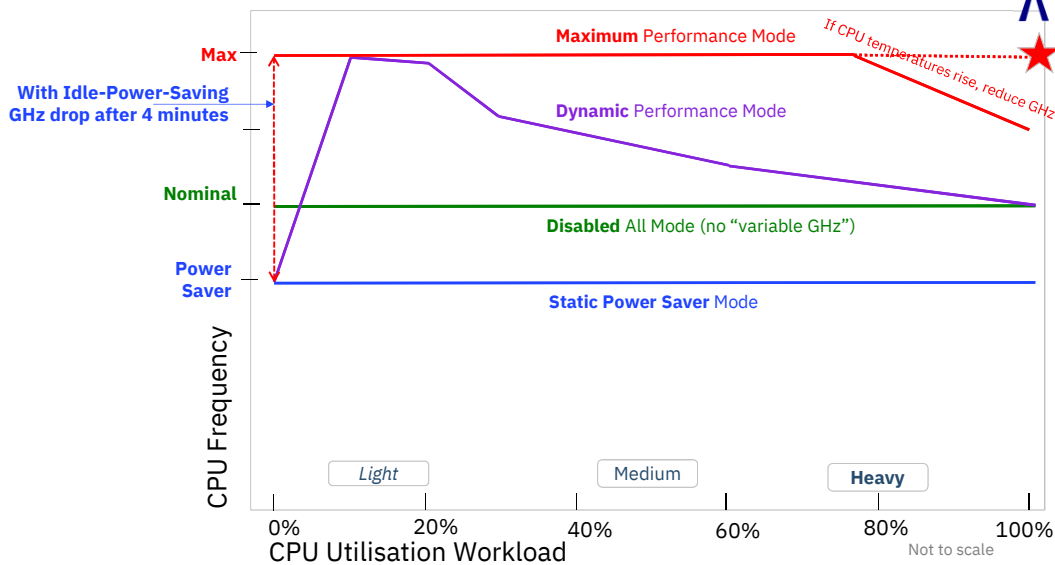
Switch off Idle-Power-Saving
System configuration: type=Shared mode=Uncapped smt=8 lcpu=16
mem=16384MB ent=1.00 Power=Dynamic-Performance

Physical Processor Utilisation:
-----Actual-----
user  sys  wait  idle   freq  user  sys  wait  idle
-----Normalised-----
0.322 0.003 0.000 0.675 3.9GHz [119%] 0.384 0.003 0.000 0.612

Switch off CPU hungry programs
-----Actual-----
user  sys  wait  idle   freq  user  sys  wait  idle
-----Normalised-----
0.000 0.001 0.000 0.999 3.9GHz [119%] 0.000 0.001 0.000 0.999
0.000 0.000 0.000 1.000 3.9GHz [119%] 0.000 0.001 0.000 0.999
0.000 0.000 0.000 1.000 3.9GHz [119%] 0.000 0.000 0.000 0.999
#
```

POWER9 EnergyScale

CPU Frequency depends on Utilisation Workload



Why no GHz reduction when 100% busy?

1. POWER9 with 8 cores per chip (max is 12) → less heat
2. Computer room at 22C → efficient fan cooling
3. Some CPUs might be at 99%
but
4. Memory size & load is low → low heat
5. Disks 8 out of 18 + not busy → low heat = using FC SAN disk
6. No high speed adapters (40Gbs+) → low heat

FSP decides it is safe to run at Max GHz (not need to reduce the GHz)

Conclusions:

1. **Meltdown/Spectre hit is small for rPerf workloads due to IBM Power System superior engineering ☺**
2. **POWER9 continues EnergyScale variable CPU frequency methods to maximise performance**
3. **POWER9 big jump in rPerf (CPW) and SMT performance**
4. **If not a full config, cool room & less than 100% busy then you may never detect a reduced GHz**
5. **Get the best from POWER9 by using SMT=8, if necessary: reducing VP and E to force SMT use**
6. **Monitor your computer room temperature & GHz !**
 - AIXpert blog: HMC REST API or Raspberry pi
 - njmon for current MHz
7. **We all need to focus on removing those Single Threaded application curse**

POWER9 is rated: ASHRAE A2

ASHRAE =

- The American Society of Heating, Refrigerating and Air-Conditioning Engineers



A2 =

- 10C - 35C (~59 F to 95 F)
- 20-80% RH (relative humidity)
- 3050m max (above sea level)

Most data centres are A1 or A2

- A1 is 15 C to 32 C (~59 F to 90 F)

Notices and disclaimers

© 2018 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.

U.S. Government Users Restricted Rights – use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed “as is” without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply.”

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those

customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Notices and disclaimers continued

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, ibm.com and [names of other referenced IBM products and services used in the presentation] are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.