



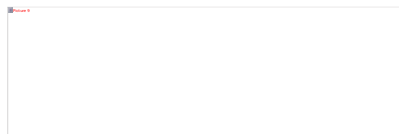
SAS 9.2 on RHEL Reference Architecture

Gary Geramanis, Partner Development Mgr., Red Hat
Doug Shakshober, Director Performance Engineering, Red Hat
Margaret Crevar, Mgr. Performance Lab, SAS

SAS on RHEL Reference Architecture (scale up)



NEC



SAS Mixed Analytics Workload: SAS created a multi-user benchmarking scenarios to simulate the workload of a typical Foundation SAS customer.

Testing: Engineers from Intel, NEC, SAS and Red Hat collaborated to run the SAS Mixed Analytics Workload on an 8 Socket, 8 Core Nehalem EX based NEC Server in the Intel lab in Chandler, AZ.

Key Test Objectives:

- Validate SAS 9.2 can scale up and perform on RHEL to exploit the capabilities of this new class of server
- Validate RHEL can excel at the extreme sequential I/O profile of a SAS workload.
- Document best practices for deploying and sizing SAS on RHEL
- Deliver key insights back to participating vendors

Results: Success against all objectives....

“These tests on the NEC platform have proven to us that Red Hat Enterprise Linux can scale vertically to exploit the potential of these servers while also delivering the intense I/O throughput that is characterized by SAS Analytics.”

*Craig Rubendall
SAS Director of Research and Development*

“The TCO and performance of SAS Analytics on Red Hat Enterprise Linux is a game changer that allows our customers to more effectively leverage the power of analytics within the confines of today’s shrinking IT budgets.”

*Anne Milley
SAS Sr. Director, Analytic Strategy*

Agenda

- SAS and Red Hat
- Red Hat Intro
- Reference Architecture Test
 - Workload
 - Test Configuration
 - Results/Observations
- RHEL Tuning Tips
- SAS Tuning Tips
- Q&A and Learn More

Red Hat and SAS

SAS' first Linux Release on RHEL 2002

SAS 9.2: Significant performance improvements for SAS Linux implementation

SAS internal GRIDs run RHEL

SAS-RHT Joint Engagement:

- Technology Alignment:
 - Roadmap exchanges
 - Lunch/Learn Series
 - SAS is included in RHEL Test Bed
 - SAS being used as benchmark for RHEL ABI Compatibility
 - SAS Cloud Substrate Evaluation (RHEL/KVM/MRG)
 - Performance Tuning
- Reference Architecture Development:
 - Analytics (scale up)
 - GRID
 - Virtualization (in Process)
 - EBI (in process)

Red Hat: *Vital Statistics*

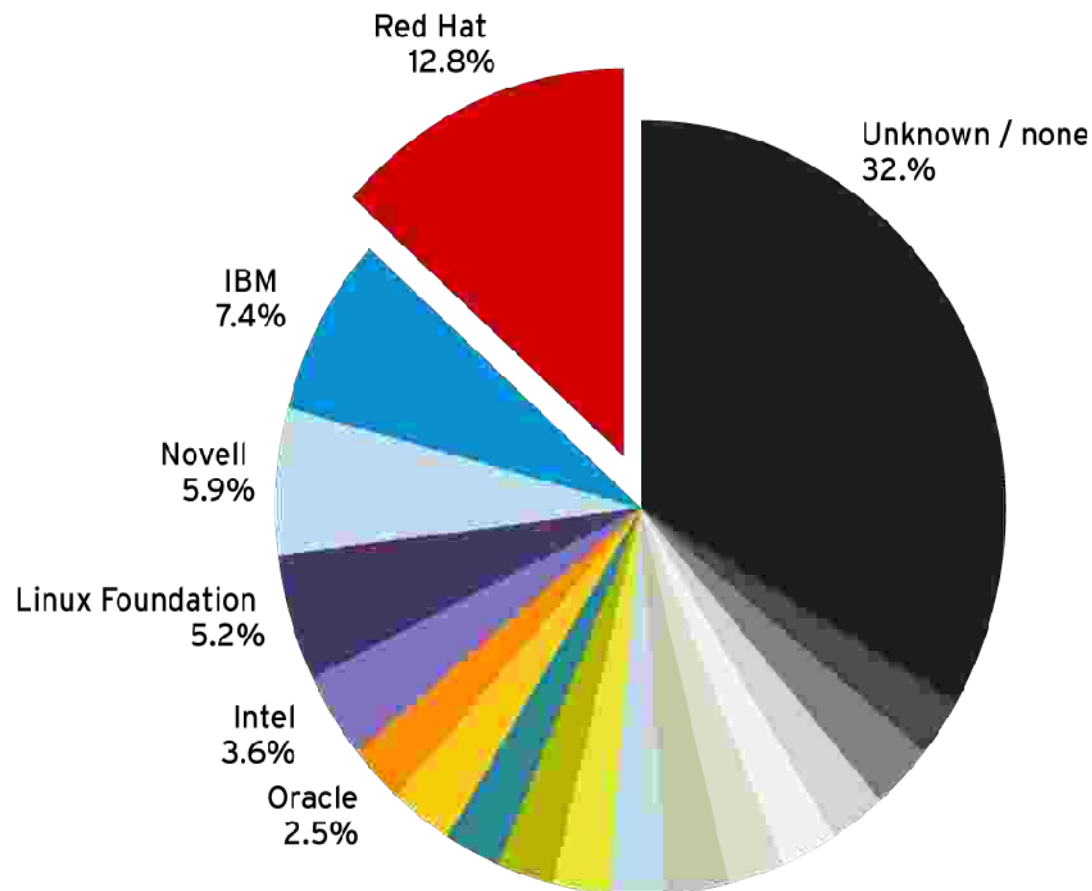
Headquarters in Raleigh, NC

- **Founded in 1993**
- **IPO, 1999 (NYSE: RHT)**
- **Added to S&P 500 2009**
- **HQ in Raleigh, NC: Over 3300 employees, in 65 worldwide offices in 27 countries**
- **Cash and investments: \$970 million (FQ4'10)**
- **FY10 revenues: \$748 million (up 15% from '09)**
- **Primary Markets: Financial, Govt. Healthcare, Education, Telco, High Tech, Computer Animation**

Red Hat contributions to open source

Red Hat is the leading contributor to Linux

RHEL is 65%* of the paid Linux Server Market
(more than 2x nearest competitor)



Source: <http://lwn.net/Articles/247582/>

* IDC Worldwide Linux Operating Environments 2008–2012
Forecast : Taking Linux to the Next Level

RED HAT PRODUCT PORTFOLIO

INFRASTRUCTURE

- Red Hat Enterprise Linux
- Red Hat Enterprise Linux Advanced Platform
- Red Hat Enterprise Linux Desktop
- Red Hat Enterprise MRG
- Red Hat Enterprise Virtualization
 - Red Hat Enterprise Virtualization Hypervisor
 - Red Hat Enterprise Virtualization Manager for Servers
 - Red Hat Enterprise Virtualization Manager for Desktops

MIDDLEWARE

- JBoss Enterprise Application Platform
 - Web Server
- JBoss Enterprise SOA Platform
- JBoss Enterprise Portal Platform
- JBoss Enterprise BRMS
- JBoss Enterprise Data Services
- JBoss Enterprise Communications Platform
- JBoss Enterprise Frameworks
- JBoss Developer Studio

MANAGEMENT

- Red Hat Satellite
- JBoss Operations Network
- Red Hat Directory Server
- Red Hat Certificate System

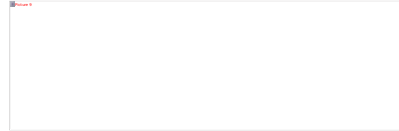
SERVICES

- Red Hat Consulting
- Red Hat Support Services
- Red Hat Training





NEC



Test Team Members:

NEC: Yuichi Kishida, Bobleigh Daniels, Daisuke Yamada, Jack Myers

SAS: Tom Keefer

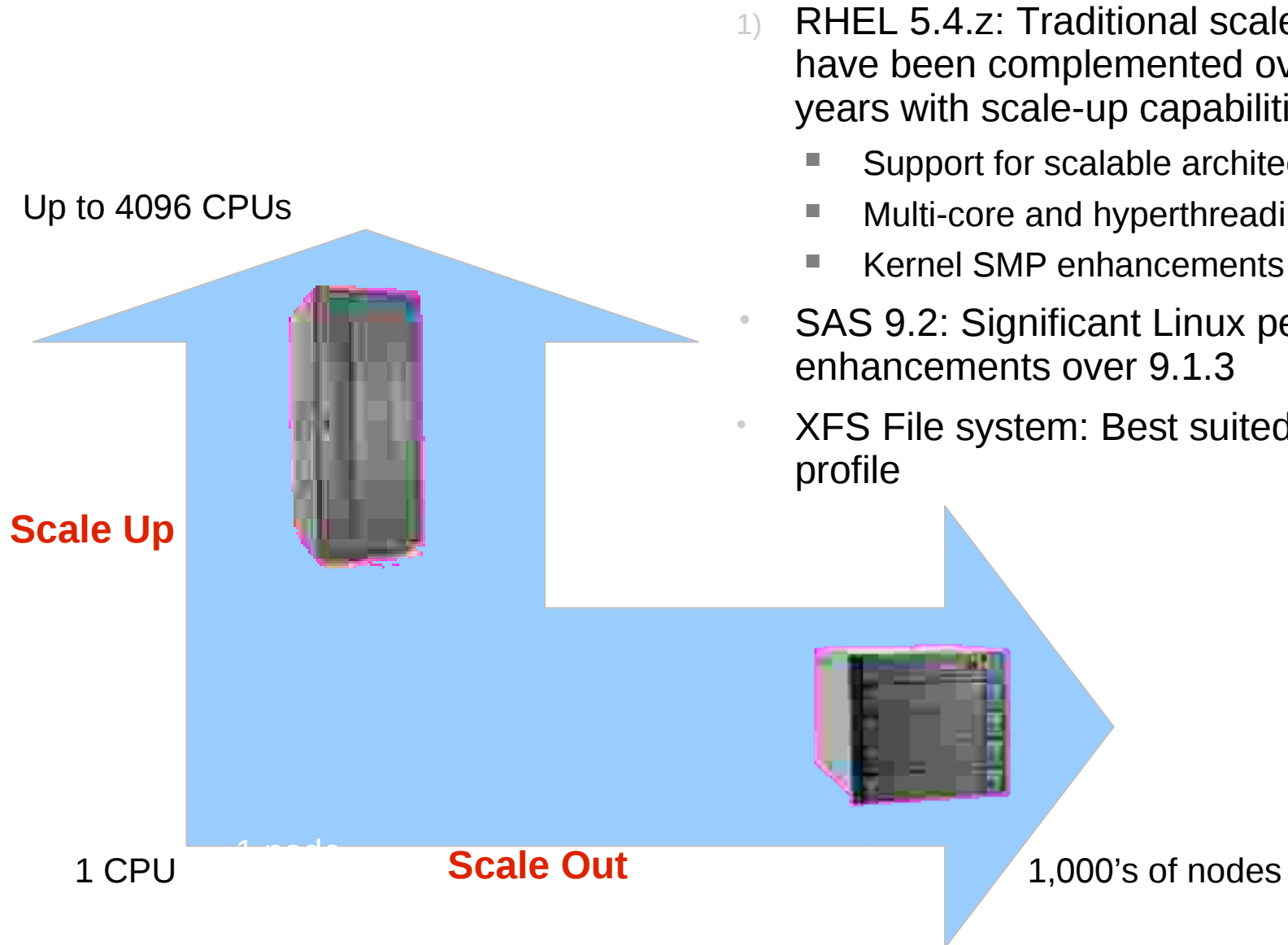
Red Hat: Douglas Shakshober, Barry J. Marson, Gary Geramanis

Intel: Sreekanth Keesara, David Baker, Mark Matusiefsky, Debra King,
Kulasinghe Priyalal, Tuan Bui, Carl Ralston, Agustin Gonzalez

Objectives

- Deliver compelling Business Analytics proof point on an NEC Express5800/A1080a 8 socket Xeon NHM-EX platform with SAS 9.2 on Red Hat Enterprise Linux
 - Demonstrate performance & near linear scalability up to 64 core
 - Demonstrate high sustained I/O throughput rates greater than 1GB/s
 - Document best practices for deploying and sizing SAS on RHEL
 - Deliver key learning's back to vendors

Key Elements to Success:



- 1) RHEL 5.4.z: Traditional scale-out capabilities have been complemented over the past two years with scale-up capabilities
 - Support for scalable architectures
 - Multi-core and hyperthreading
 - Kernel SMP enhancements
- SAS 9.2: Significant Linux performance enhancements over 9.1.3
- XFS File system: Best suited for analytics I/O profile

SAS Mixed Analytics Workload

SAS created a multi-user benchmarking scenarios to simulate the workload of a typical Foundation SAS customer. The goal of these scenarios is to evaluate the multi-user performance of SAS on various platforms. Various sized mixed analytic workloads were created to simulate many users utilizing CPU, RAM and I/O resources which SAS programs heavily utilize during typical program execution.

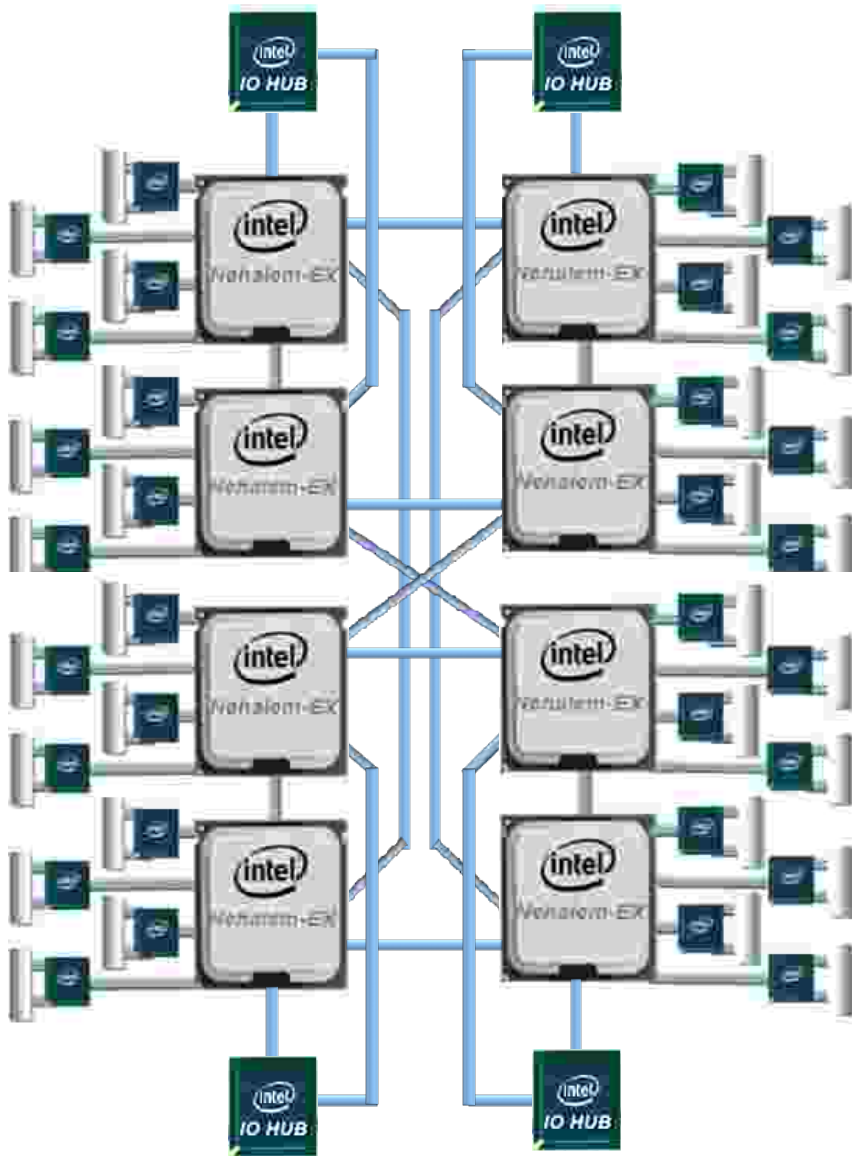
Workload:

- SAS datasets and text files; File sizes ranging from several kilobytes to 30 gigabytes in size
- Row counts up to 90 million
- Column counts up to 297
- All Jobs a Mix of CPU and I/O intensive jobs
- SAS Procedures: GLM, LOGISTIC, RISK, REG, MEANS, SORT, FREQ, SUMMARY, SQL
- Data volumes were designed to be larger than physical RAM in order to place realistic stress on the hardware and operating system file cache.

Test Execution 2 Scenarios (32 and 64 core): Each test scenario consists of a set of SAS jobs run in a multi-user fashion to simulate a typical SAS batch, Enterprise Guide user environment. Each scenario launches jobs simultaneously at a set interval to help simulate a multi-user environment where users come and go from the system. The test was designed to run in a period of 30 to 60 minutes.

	Scenario 1 32 Cores	Scenario 2 64 cores
Input Data	500GB	1 TB
Jobs*	216	432

NEC 8S – SAS 9.2 – Intel Nehalem-EX



8 CPU & I/O Topology

Scalable Architecture

- 8 cores/socket up to 64 Cores & 128 Threads
- Intel® Hyper-Threading Technology
- 24MB shared last level cache
- NEC NUMA BIOS Enhancements

High Performance Interconnect

- Intel QPI Technology x4 /CPU
- Up to 12.8 GB/s per link or 50.2GB/s per CPU
- 9X previous generation

Scalable Memory Interconnect

- Intel® Scalable Memory Interconnect w/Buffers
- 4, Buffered Memory Channels @ 1066MHz.
- 50GB/s /channel memory bandwidth
- 6X previous generation)

Advanced Reliability

- Machine Check Architecture (new to Xeon)
- Redundant Service Processors
- Red Hat OS Support*
- Advanced Hardware Transparency
(data reliability checks every cycle, link self-healing, alternative routing and clock failover)

64 Core Benchmark Configuration

Server configuration:

- NEC Express5800/A1080a 8CPU, 64 Core, 128 Thread
- 8 x Intel NEH-Ex 2.27 GHz processors
- 256GB (64 x 4GB DIMMs)
- 8 x 2port 4Gbps FC HBAs
- NUMA ON, HT ON, TURBO ON
- BIOS Revision : 01.20
- BMCFW version : 1.0.0.6



SAS 9.2 Foundation (64-bit) benchmark configuration:

- RHEL 5.4z, 2.6.18-164.11.1.el5 kernel
- SAS 9.2
- SAS Mixed Analytics workload w/64 Cores / 128 Threads “Scale factor”



Express5800



NEC

Storage configuration:

- NEC D-Series Storage Array
- 4 x D3 storage arrays, each D3 with four, 4 Gb/s, ports
- 2 x additional Disk Enclosure per D3 node
- 14 Luns (9 disks each)
- 1 TB SATA 7200 rpm disks
- 1 LVM with 4X9 disks in RAID 5 for input
- 1 LVM with 6X9 disks in RAID 5 for work
- 1 LVM with 4X9 disks in RAID 5 for output

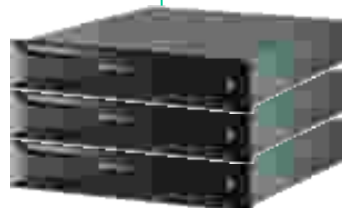
16 x 4Gbps FC ports

4 x 4Gbps FC ports

4 x 4Gbps FC ports

4 x 4Gbps FC ports

4 x 4Gbps FC ports



32 Core Benchmark Configuration

Server configuration:

- NEC Express5800/A1080a 4CPU, 32 Core, 64 Thread
- 4 x Intel NEH-Ex 2.27 GHz processors
- 128GB (32 x 4GB DIMMs)
- 4 x 2port 4Gbps FC HBAs
- NUMA ON, HT ON, TURBO ON
- BIOS Revision : 01.20
- BMCFW version : 1.0.0.6



SAS 9.2 Foundation (64-bit) benchmark configuration:

- RHEL 5.4z, 2.6.18-164.11.1.el5 kernel
- SAS 9.2
- SAS Mixed Analytics workload w/32 Cores / 64 Threads “Scale factor”



Express5800

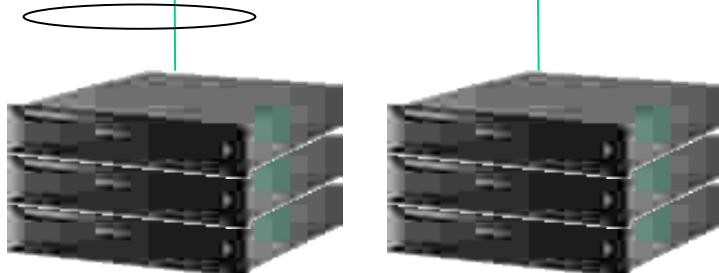


8 x 4Gbps FC ports

NEC

4 x 4Gbps FC ports

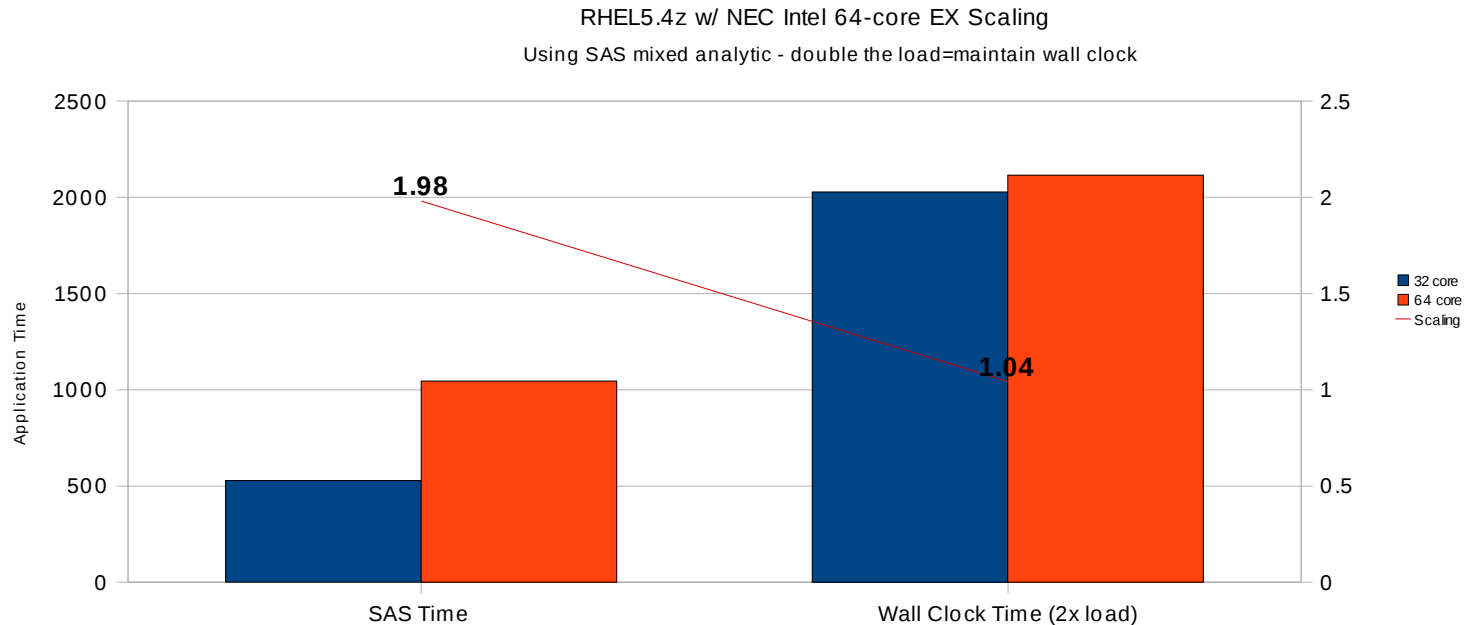
4 x 4Gbps FC ports



Storage configuration:

- NEC D-Series Storage Array
- 2 x D3 storage arrays, each D3 with four, 4 Gb/s, ports
- 2 x additional Disk Enclosure per D3 node
- 1 TB SATA 7200 rpm disks
- 1 LVM with 2X9 disks in RAID 5 for input
- 1 LVM with 4X9 disks in RAID 5 for work
- 1 LVM with 2X9 disks in RAID 5 for output

Results – SAS Time & Wall Clock accounting

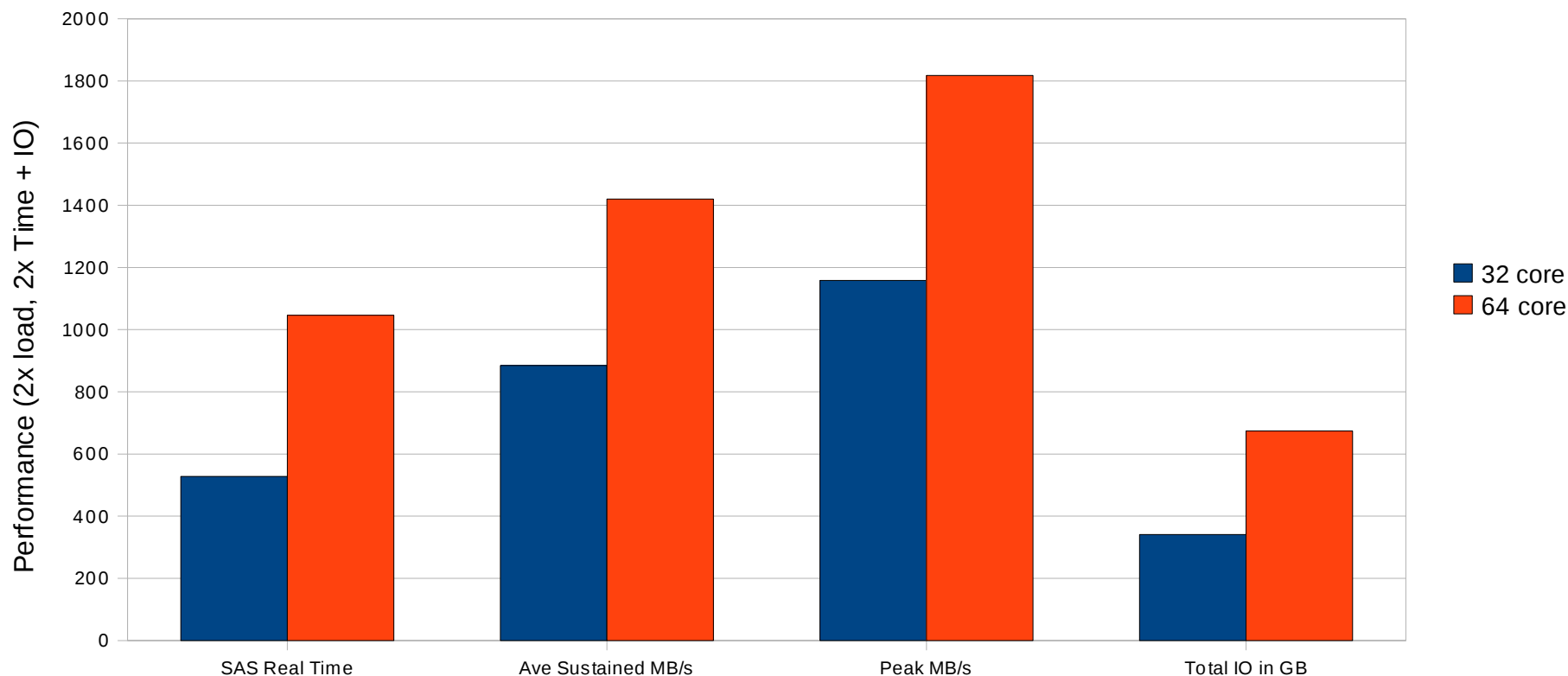


- More than 800 MB/s sustained i/o for 32 core, more than 1,800 MB/s sustained i/o for 64 cores
- 1.98X more i/o for 64C
- Twice as much load results in 2.04X CPU time¹ for 64 core load but because of double the amount of processors, elapsed time remains almost the same (1.04X). Almost perfect scalability!

Results – comparing I/O throughputs

RHEL5.4z w NEC Intel EX 64 core

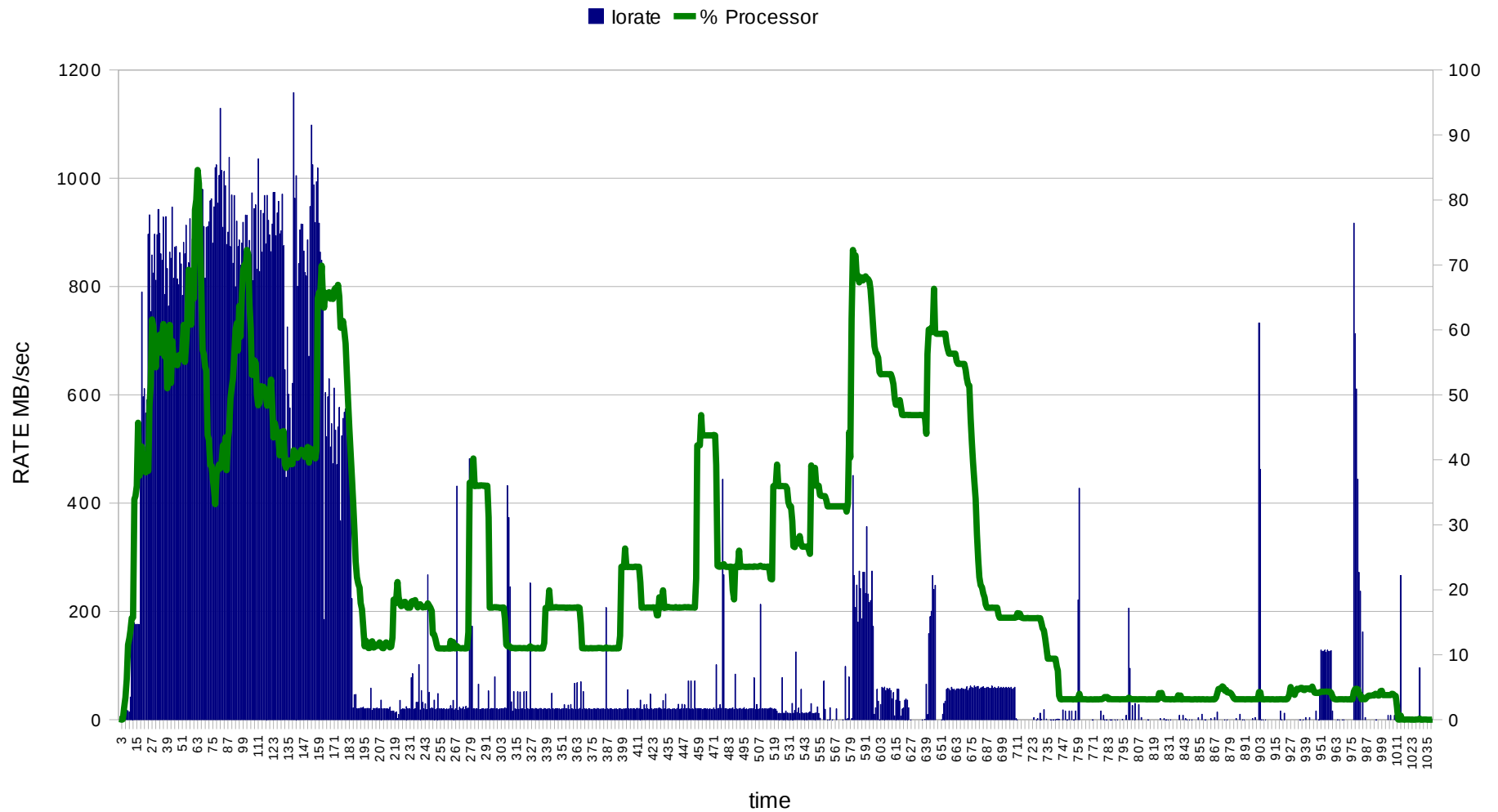
Using SAS mixed analytics IO workload



Twice more load, twice more resources, constant elapsed time = linear scalability!

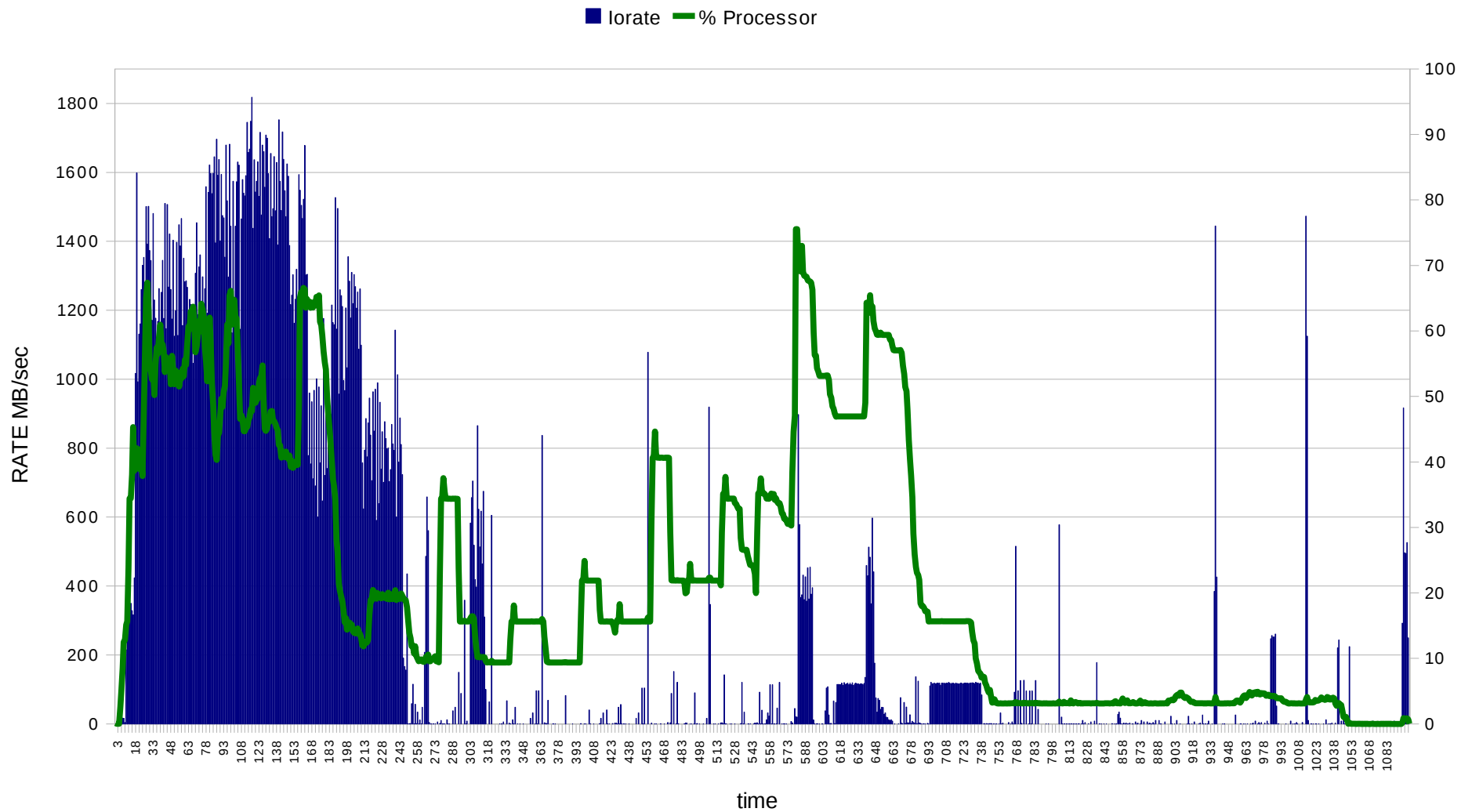
32 Core run

RHEL5.4z on NEC Intel EX 32-core, 128 GB mem, XFS
w/ SAS mixed analytic workload



64 core

RHEL 5.4z on NEC Intel EX 64-core, 256 GB, XFS
w/ SAS mixed analytic workload (cpu and IO data)



Red Hat Enterprise Linux: Observations

Validate performance features on Intel EX large SMP servers

- Support Intel's APIC Architecture for optimized Non-Uniform Memory Access (NUMA)
 - RHEL support on Multi-socket NEC Nehalem EX
 - Significant performance gain w/ NEC numa aware BIOS.
 - Linux process scheduler automatically optimizes SAS application processes.
- Near perfect scaling from 32 to 64 cores
- Scalability improved by both Turbo Boost and Hyperthreads

RHEL Observations & Tuning

- RHEL 5.4z used for testing and recommended for any Nehalem EX system.
- RHEL tuning best practices applied using Ktune*
- SELinux and unneeded services were disabled via chkconfig and sysctl parameters
- NUMA (non-uniform memory access) is RHEL 5 default on NEC and all Intel EX systems
- Check BIOS to ensure NUMA is enabled or disable memory interleave

*ktune will be discussed further under "General RHEL Tuning Best Practices"

Red Hat Enterprise Linux: Observations

I/O Observations & Tuning

- The I/O output from 8S server exceeded capacity of test storage configuration. Server utilization would have been significantly greater with more storage controllers.
- 1.78 GB/sec peak and 1.39 GB/sec sustained throughput for 64C runs
- 1.13 GB/sec peak and 885 MB/sec sustained throughput for 32C runs
- “Tunable” I/O stack essential to SAS performance
 - Tuned read-ahead on Logical Volume Manager (LVM) devices adjusted to 8192 bytes
 - Standard blockdev tool to adjust for large sequential access to LVM/filesystem
 - 30% improvement in performance using XFS over ext3 (XFS is better suited for large sequential IO in the SAS analytics workload)

Ktune: kernel tool to alter sysconfig parameters for High End Servers (aka NEC/EX)

- ktune to adjust Linux parameters
 - Install - `yum install ktune` (available > RHEL 5.3)
 - Run - `service ktune start`
- Alters
 - IO elevator for each device = from default to `elevator=deadline`
 - Power saving's governor defaults to `on_demand`, performance governor best for IO loads
 - `echo performance > /sys/devices/system/cpu/cpu0/cpufreq/scaling_governor`
 - Increase network buffers
 - Suggest lowering security fences Selinux – permissive, `auditd` service disable
- disable services that you don't need
 - `chkconfig –list`
 - `chkconfig auditd stop` (control after machine setting once rebooted)
- use EXT4, or XFS options in newer kernels

Tuning Filesystems

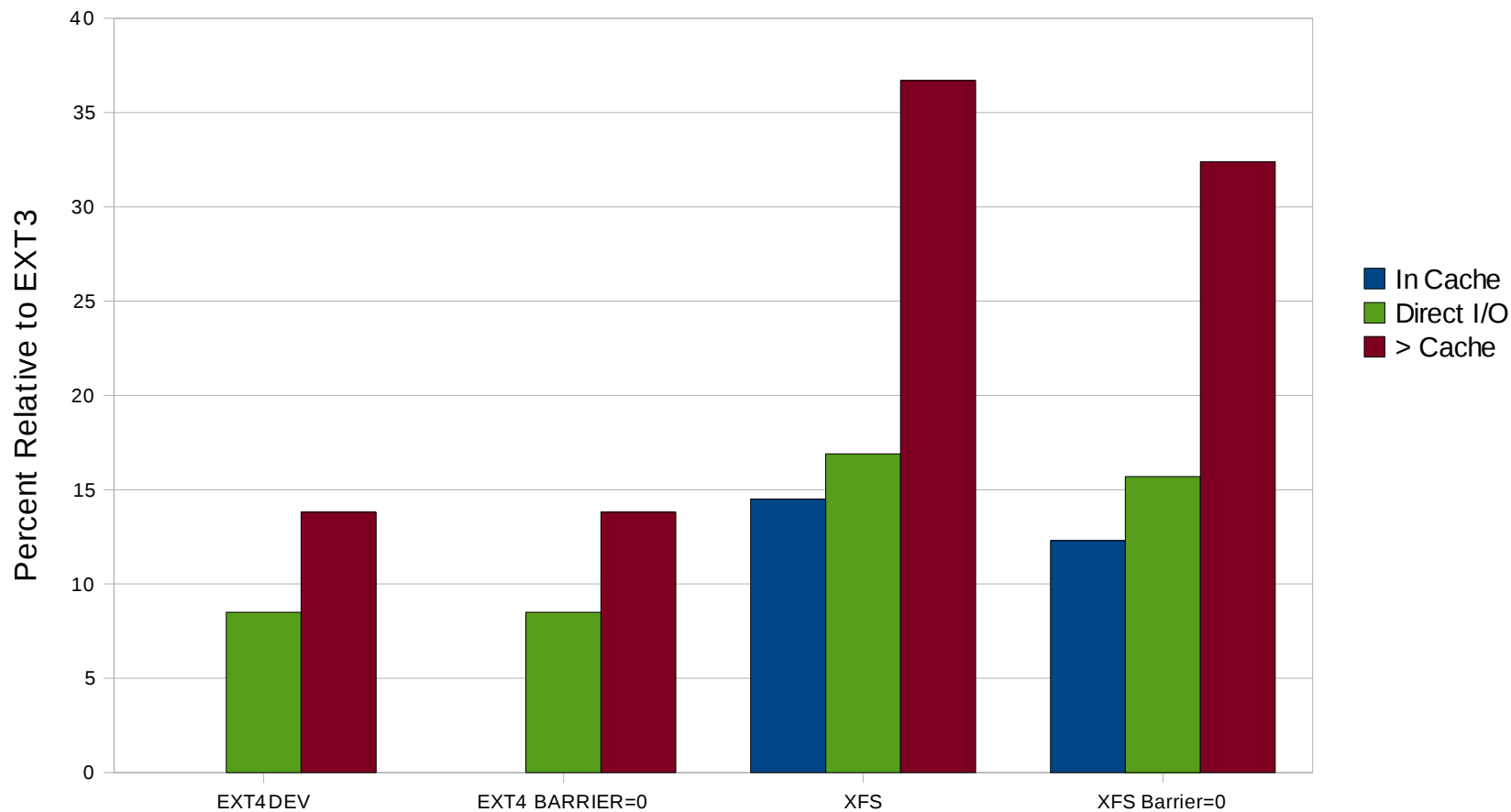
- Mount options
 - How precise you want date/time -noatime
 - Access controls acl or -noacl – NFS
- Tune2fs
 - Writeback options
 - Journal options -j, blocksize
- Lun optimizations
 - Blockdev w/ /dev and LVM
 - Readahead adjustment
 - [root@localhost ~]# blockdev --getra /dev/sda5
 - 256
 - [root@localhost ~]# blockdev --setra /dev/sda5 2048

RHEL Filesystems

- ext3: Journaled File System (Default for RHEL 5)
- ext4: Option as Tech Preview for RHEL 5
 - Scale beyond ext3's current 16TB limit
 - 10x improvement in fsck time.
- XFS: High Performance Journaled File System
 - Support for extremely large file systems of up to 100TB
- NFS: NFS4.0 – is default (NFS2 & 3 are also supported)
- GFS2: Global File System (shared disk file system) up to 16 nodes.

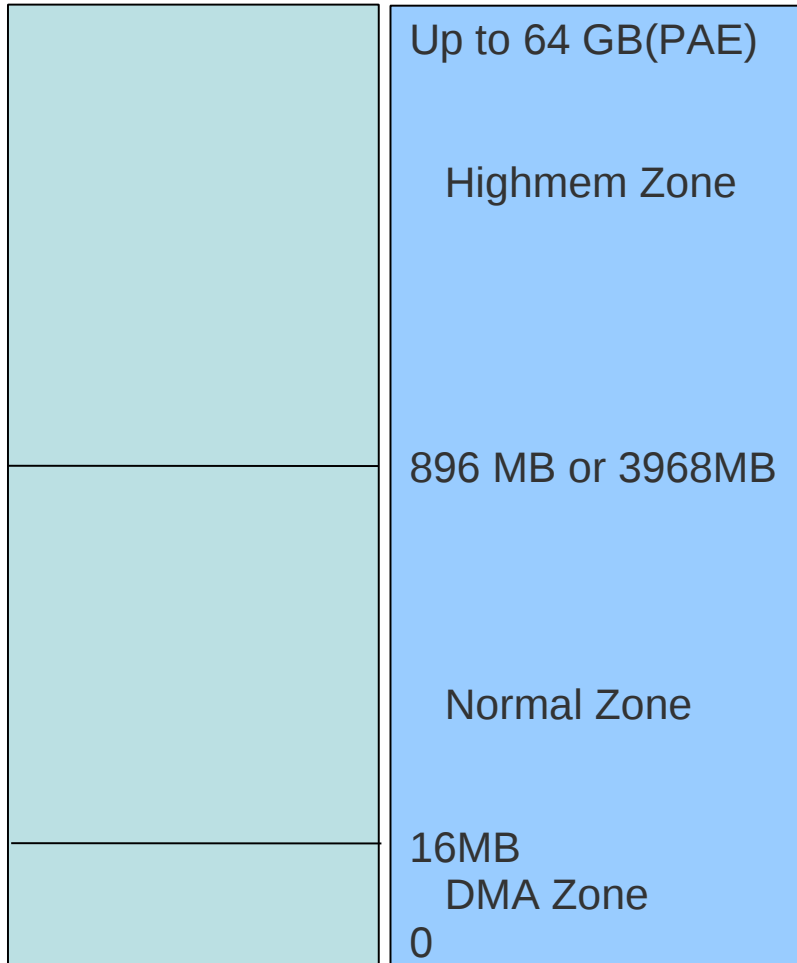
RHEL5.3 IOzone EXT3, EXT4, XFS eval

RHEL53 (120), IOzone Performance
Geo Mean 1k points, Intel 8cpu, 16GB, FC

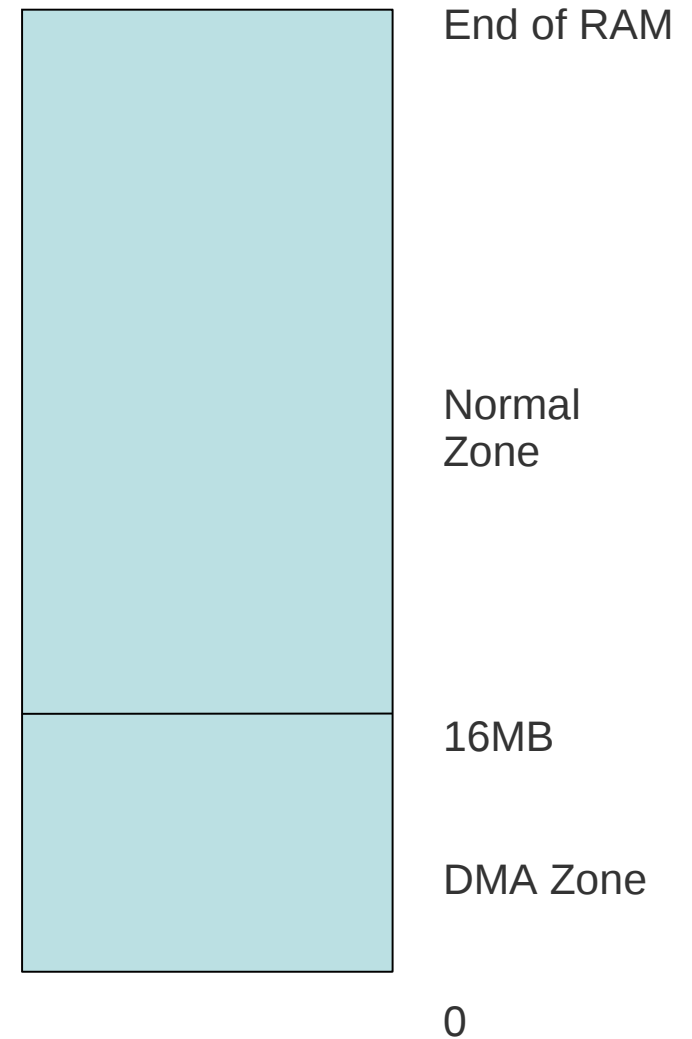


Memory Zones

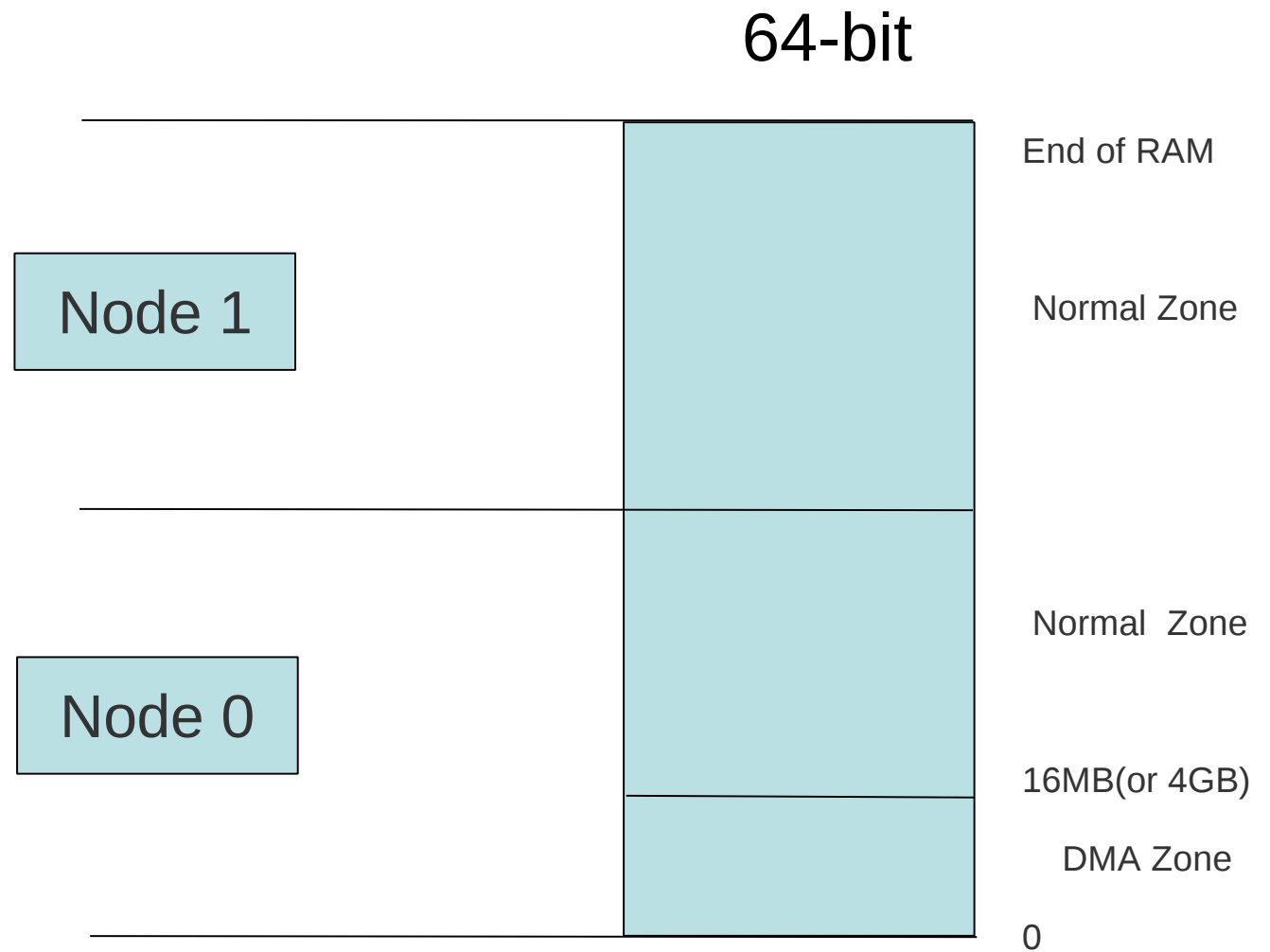
32-bit



64-bit



NUMA Nodes and Zones



General SAS Tuning Guidelines

Follow best practices for configuring IO

- Split permanent SAS data files, SAS WORK files, and SAS UTILLOC files into separate file systems
- Insure you have enough IO bandwidth to support the SAS application requirements
- Set the IO transfer unit size on storage array to match SAS BUFSIZE value
- Increase SAS BUFSIZE value if you are doing large volumes of IO

More information:

Look for detailed SAS Reference Architecture Whitepaper in May '2010 at...
<http://www.redhat.com/SAS>

Find all Red Hat Reference Architecture Papers at...
http://www.redhat.com/rhel/resource_center/reference_architecture.html

gary.geramanis@redhat.com
dshaks@redhat.com
Margaret.Crevar@sas.com

Or...

SAS@redhat.com

Questions?

