



RED HAT<sup>®</sup>  
ENTERPRISE LINUX<sup>™</sup>

# RED HAT ENTERPRISE LINUX 7 TECHNICAL OVERVIEW

Scott McCarty  
Senior Solutions Architect, Red Hat  
01/12/2015

# Performance Tuning Overview

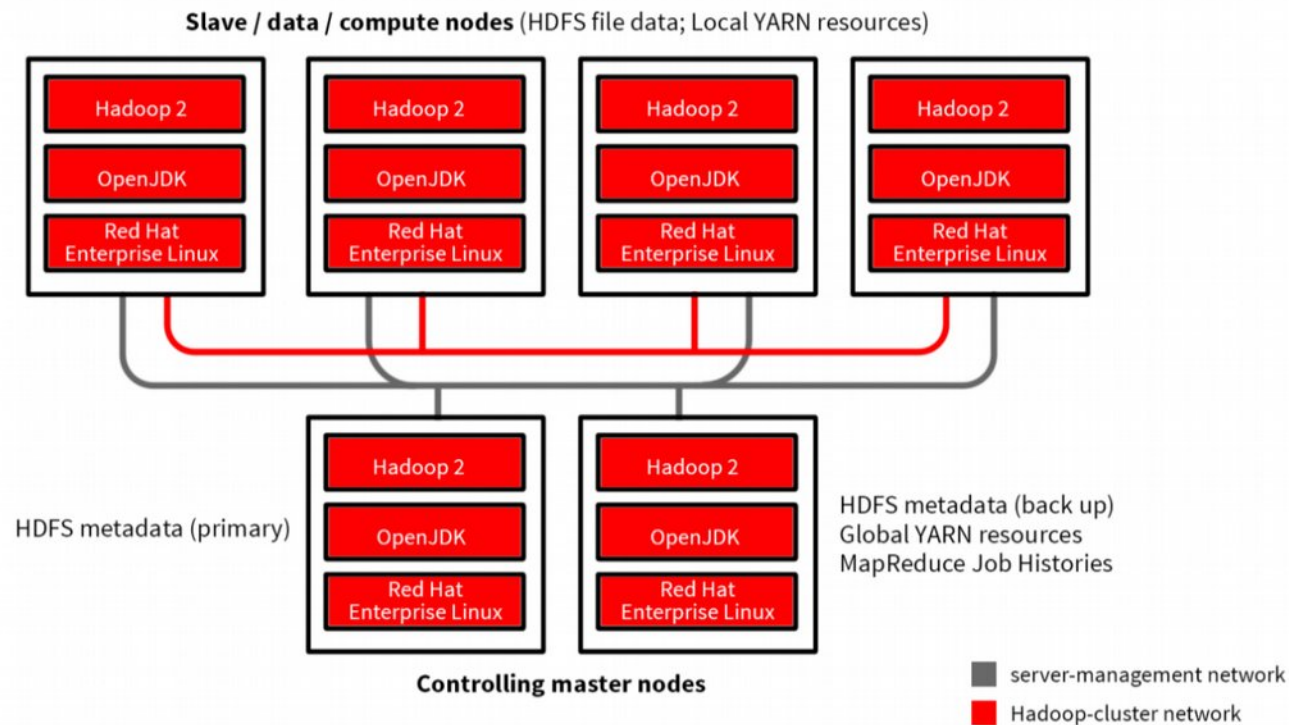
- Little's Law:  $L = A \times W$  (Queue Length = Average Arrival Rate x Wait Time)
  - The length of the waiting line for a resource depends on the average rate at which new requests arrive multiplied by the average amount of time a request spends in the system.
- Wait Time:  $W = S + Q$  (Wait Time = Service Time + Queue Time)
  - The sum of the amount of time the request waits for a resource to become available combined with the amount of time it takes to service the request.
- Queue Stability & Forced Flow Law
  - For a queue to maintain the same length, the rate at which new requests arrive must be the same as the rate at which requests are completed
  - System Throughput = Resource Throughput x Resource Visit Count

# Performance Tuning Overview (continued)

- Variables
  - Latency
  - Throughput
  - Power Usage
  - Jitter
- Standard Process
  - Baseline -> Make Change -> New Baseline
- Final Thoughts
  - Typically compared to the old system
  - Does not necessarily always mean better\*

# Standard Distributed Architecture

- Master Nodes: HDFS NameNode service, YARN resource manager and MapReduce Job History manager
- Data Nodes: HDFS DataNode service and the YARN distributed NodeManager

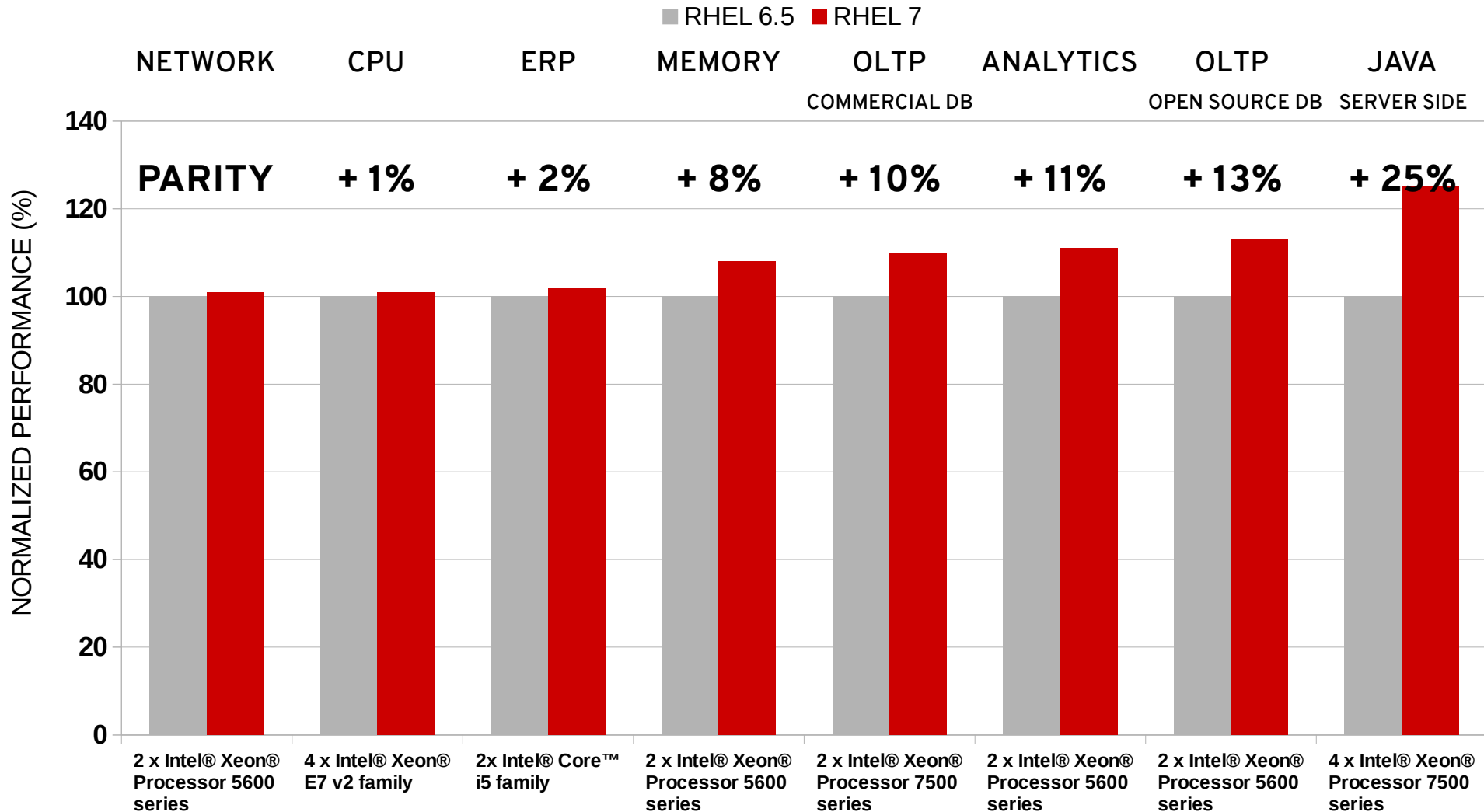


# SOLID PERFORMANCE ACROSS WORKLOADS

## RHEL 7 VS RHEL 6.5



PERFORMANCE GAINS ACROSS WIDE RANGE OF WORKLOADS  
AND MULTIPLE GENERATIONS OF HARDWARE



# CHOICE OF FILE SYSTEMS



- Scale file systems to 500TB with new default filesystem XFS
- Scale to 50TB with ext4
- Btrfs also available<sup>1</sup>
- Parallel NFS v4 provides improved performance and throughput

Type	Supported Limit	Root	Boot	Comments
Single-node				
XFS	500TB	Yes	Yes	System default
ext4	50TB	Yes	Yes	Driver allow access to older versions (ext2, ext3).
btrfs <sup>2</sup>	50TB	Yes	Yes	
Network/Multi-node				
GFS2	2-16 nodes	Yes	No	Shared-storage file system

<sup>1</sup> Available as a Technology Preview

# Tuned Recommendations

- Physical
  - NameNode and DataNode:
    - `tuned-adm profile enterprise-storage`
- Virtualized
  - NameNode and DataNode:
    - `tuned-adm profile virtual-guest`
    - `tuned-adm profile virtual-host` (for underlying host)

# Behind the Scenes

- Scheduler Changes
  - Deadline executes I/O Operations (IOPs) through the concept of "batches" which are sets of operation
  - ELEVATOR="deadline"  
# These are the devices, that should be tuned with the  
  
ELEVATORELEVATOR\_TUNE\_DEVS="/sys/block/{sd,cciss,dm-,vd}\*/queue/scheduler"
- Sysctl.conf Changes: Increases latency, but also throughput
  - kernel.sched\_min\_granularity\_ns = 10000000
  - kernel.sched\_wakeup\_granularity\_ns = 15000000
  - # The generator of dirty data starts writeback at this  
# percentage (system default is 20%)  
vm.dirty\_ratio = 40



# Links

- Red Hat & HortonWorks Data Platform Reference Architecture
  - <http://www.redhat.com/en/resources/exploring-the-next-generation-of-big-data-solutions-with-hadoop-2-on-red-hat-enterprise-linux-6>
- Douglas Shakshober (Shak) and Larry Woodman Performance Tuning:
  - Part 1: <https://www.youtube.com/watch?v=fATEiBJ3pKw>
  - Part 2: <https://www.youtube.com/watch?v=km-vLELmWLS>
- Deadline Scheduler
  - [http://en.wikipedia.org/wiki/Deadline\\_scheduler](http://en.wikipedia.org/wiki/Deadline_scheduler)
- Adjusting CFS parameters
  - <http://www-01.ibm.com/support/knowledgecenter/linuxonibm/liaai.saptuning/saptuningadjust.htm>



RED HAT<sup>®</sup>  
ENTERPRISE LINUX<sup>™</sup>

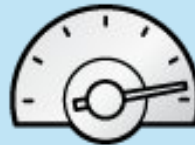
# Thank You

Scott McCarty  
Senior Solutions Architect, Red Hat  
01/12/2015

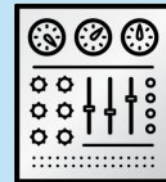
# PERFORMANCE ENHANCEMENTS WITH RED HAT ENTERPRISE LINUX 7



**BUILT-IN PERFORMANCE  
PROFILES SIMPLIFY  
CONFIGURATION**



**MONITORING WITH  
PERFORMANCE CO-PILOT  
AND THERMOSTAT**



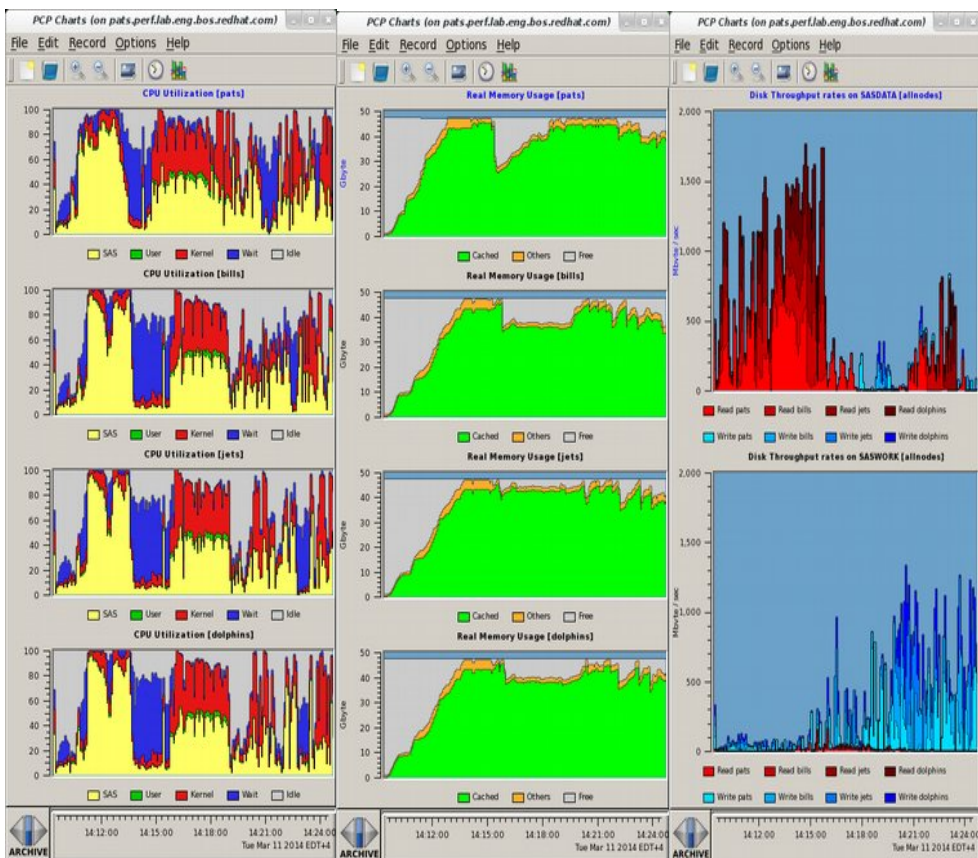
**FINE-TUNE PERFORMANCE  
WITH ENHANCED TOOLING  
VIA TUNA AND TUNED**

# OPTIMAL PERFORMANCE VIA PROFILES

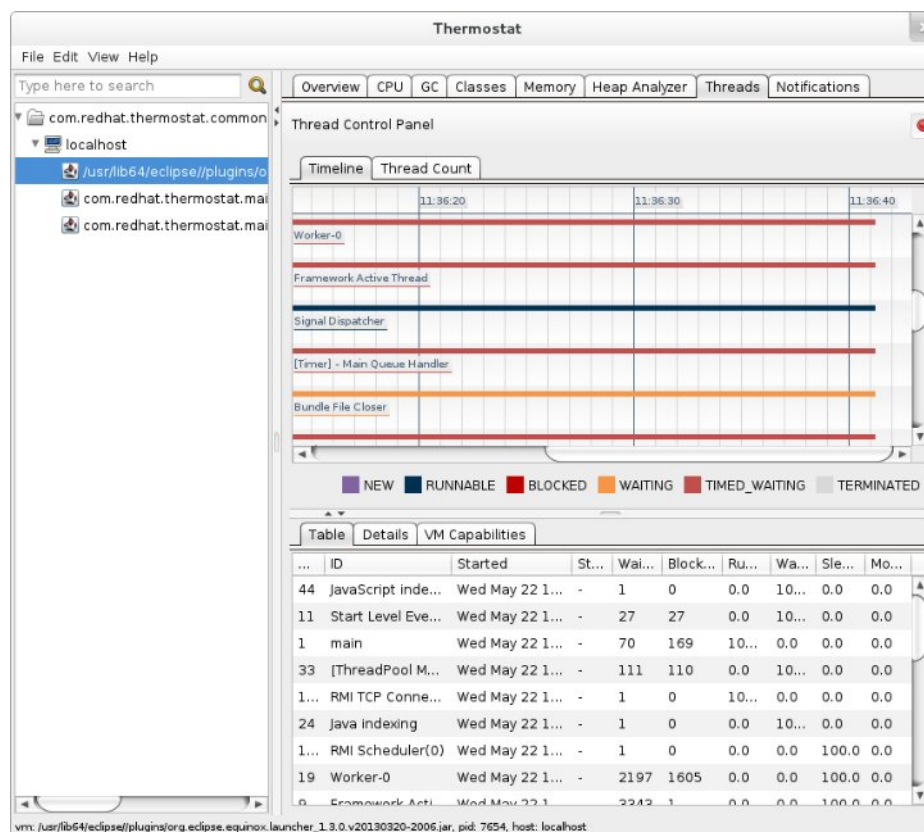


Optimal performance management via enhanced performance tuning at install, simplified instrumentation and tuning features, and performance monitoring tooling

## PERFORMANCE CO-PILOT (PCP)



## THERMOSTAT (FOR JVMs)



# PROFILING AND MONITORING WITH *TUNA*



- Tool for fine grained control
- Display applications / processes
- Displays CPU enumeration
- Socket (useful for NUMA tuning)
- Dynamic control of tuning
  - Process affinity
  - Parent & threads
  - Scheduling policy
  - Device IRQ priorities, etc

The screenshot shows the Tuna application window titled "Tuna (on dhcp47-40)". It features a table for CPU usage across two sockets and a larger table for process details.

Socket 0			Socket 1		
Filter	CPU	Usage	Filter	CPU	Usage
<input checked="" type="checkbox"/>	1	2	<input checked="" type="checkbox"/>	0	12
<input checked="" type="checkbox"/>	3	0	<input checked="" type="checkbox"/>	2	5
<input checked="" type="checkbox"/>	5	0	<input checked="" type="checkbox"/>	4	1
<input checked="" type="checkbox"/>	7	0	<input checked="" type="checkbox"/>	6	3
<input checked="" type="checkbox"/>	9	0	<input checked="" type="checkbox"/>	8	8
<input checked="" type="checkbox"/>	11	0	<input checked="" type="checkbox"/>	10	0
<input checked="" type="checkbox"/>	13	0	<input checked="" type="checkbox"/>	12	0
<input checked="" type="checkbox"/>	15	0	<input checked="" type="checkbox"/>	14	3

PID	Policy	Priority	Affinity	VolCtxSwitch	NonVolCtxSwitch	Command Line
10600	OTHER	0	0-15	437	1	/usr/libexec/notification-area-applet --oaf-activ
10504	OTHER	0	0-15	544	1	/usr/libexec/notification-area-applet --oaf-activ
11065	OTHER	0	0-15	3796	1	/usr/libexec/notification-daemon
11066	OTHER	0	0-15	781	1	/usr/libexec/notification-daemon
7487	OTHER	0	0-15	2113	2	/usr/libexec/polkit-1/polkitd
8669	OTHER	0	0-15	120	1	/usr/libexec/polkit-gnome-authentication-agent
3286	OTHER	0	0-15	3502	1	/usr/libexec/postfix/master
7641	OTHER	0	0-15	35	2	/usr/libexec/pulse/gconf-helper
26428	OTHER	0	0-15	<b>5514763</b>	132	/usr/libexec/qemu-kvm -S -M rhel6.3.0 -enable
26510	OTHER	0	0-15	<b>6018050</b>	146	/usr/libexec/qemu-kvm -S -M rhel6.3.0 -enable
26599	OTHER	0	0-15	<b>366744</b>	88	/usr/libexec/qemu-kvm -S -M rhel6.3.0 -enable
26352	OTHER	0	0-15	<b>400968</b>	126	/usr/libexec/qemu-kvm -S -M rhel6.3.0 -enable
7497	OTHER	0	0-15	23	2	/usr/libexec/rtkit-daemon
10573	OTHER	0	0-15	238	1	/usr/libexec/trashapplet --oaf-activate-iid=OAF
10473	OTHER	0	0-15	231	1	/usr/libexec/trashapplet --oaf-activate-iid=OAF
7596	OTHER	0	0-15	1626	7	/usr/libexec/udisks-daemon
10470	OTHER	0	0-15	1124	1	/usr/libexec/wnck-applet --oaf-activate-iid=OAF
10567	OTHER	0	0-15	1155	1	/usr/libexec/wnck-applet --oaf-activate-iid=OAF
3310	OTHER	0	0-15	154	0	/usr/sbin/abrt
2856	OTHER	0	0-15	2	0	/usr/sbin/acpid
3371	OTHER	0	0-15	21	0	/usr/sbin/atd
3500	OTHER	0	0-15	<b>10875</b>	1	/usr/sbin/certmonger -S -p /var/run/certmonger
6561	OTHER	0	0-15	1814	2	/usr/sbin/console-kit-daemon --no-daemon
3512	OTHER	0	0-15	<b>10814</b>	0	/usr/sbin/dnsmasq --strict-order --bind-interfac
2654	OTHER	0	0-15	49367	2672	/usr/sbin/ehtcmd --pidfile /var/run/ehtcmd.pid
7387	OTHER	0	0-15	485	3	/usr/sbin/gdm-binary -nodaemon
2108	OTHER	0	0-15	1	0	/usr/sbin/gdm-binary -nodaemon