### The Wilson HPC Cluster

Dalibor Djukanovic

Institut für Kernphysik Johannes Gutenberg-Universität, Mainz

#### Technical Seminar Desy Zeuthen 03.11.2009

< A >

- A - B - M

## Mainz Group

- Head of group: Prof. Hartmut Wittig
- Members:
  - 3 postdocs
  - 5 Ph. D. students
  - 1 diploma student
- IT department:

Head of group: Dr. Klaus Werner Krygier

- 2 system administrators
- 1 techincal staff member
- 2 electronics engineers
- 1 system admistrator dedicated to Cluster





What is Wilson and why do we need it?

2 Hardware

- 3 Management & Monitoring
- 4 Benchmarks



Hardware Management & Monitoring Benchmarks

### Outline



### What is Wilson and why do we need it?



< E

Hardware Management & Monitoring Benchmarks Summary Why do we need Wilson? Who is Wilson? Installation

### Physics



- 4 fundamental forces
- Strong force is strongest

<ロト < 同ト < 三ト

э

 Described by Quantum Chromodynamics (QCD)

Why do we need Wilson? Who is Wilson? Installation

### Quantum Chromodynamics (QCD)

• Nuclear Physics @ low energies

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} F^{a}_{\mu\nu} F^{a\,\mu\nu} + \sum_{f=u,d,s,\dots} \overline{\Psi}_{f} \Big\{ i \gamma^{\mu} \big( \partial_{\mu} - i g A^{a}_{\mu} T^{a} \big) - m_{f} \Big\} \Psi_{f}$$

- Looks simple, but it is hard to solve
- $\bullet$  Only way to solve this equation  $\Rightarrow$  numerical simulation
- Discretize Spacetime  $\Rightarrow$  Put the theory on a lattice
- Lattice QCD  $\Rightarrow$  mostly Monte Carlo simulations  $\Rightarrow$  this is why we need Wilson
- But still what is Wilson?

Hardware Management & Monitoring Benchmarks Summary Why do we need Wilson? Who is Wilson? Installation

### Wilson



• Kenneth G. Wilson: Nobel prize in physics 1982

- One of the founders of Lattice QCD
- We use Wilson fermions

#### Our own Wilson

- Dedicated platform for lattice QCD simulations
- Cluster delivered by Hewlett-Packard
- Installation October 2008  $\rightarrow$  Total cost: 1.098.000  $\in$

Hardware Management & Monitoring Benchmarks Summary Why do we need Wilson? Who is Wilson? Installation

### Pictures



#### We bascially started from scratch

- Preparation of the server room
- Installation of a proper cooling
- HP provided ready to go Installation

Hardware Management & Monitoring Benchmarks Summary Why do we need Wilson? Who is Wilson? Installation

### Pictures



▲□▶ ▲□▶ ▲ □▶

э

Hardware Management & Monitoring Benchmarks Summary Why do we need Wilson? Who is Wilson? Installation

### Pictures



# Air Condition First time a theorist needed a Riga crane! Dalibor Djukanovic The Wilson HPC Cluster

Hardware Management & Monitoring Benchmarks Summary Why do we need Wilson? Who is Wilson? Installation

### Pictures



<ロト <回ト < 回

Hardware Management & Monitoring Benchmarks Summary Why do we need Wilson? Who is Wilson? Installation

### Pictures



#### Sometimes it was close:



Dalibor Djukanovic The Wilson HPC Cluster

< □ > < 同 > < 回 >

Hardware Management & Monitoring Benchmarks Summary Why do we need Wilson? Who is Wilson? Installation

### Finally



Dalibor Djukanovic The Wilson HPC Cluster

æ

Infrastructure Compute Nodes Software

### Outline



### 2 Hardware

3 Management & Monitoring

### Benchmarks



< ∃ >

Infrastructure Compute Nodes Software

### Originally we had

- 250 DL 160 G5 Servers: Intel Xeon E5462 @ 2.8 GHz
- Problems with stability: Nodes crashed with kernel panic No obvious reason Not specific to node Uncorrelated to runtime
- Finally the complete System was replaced
   ⇒ some parts are olden than other:
   Power supplies
   IB HCA

Infrastructure Compute Nodes Software

### Now we have

#### Wilson & Lilly

- We actually have 2 Clusters:
  - Wilson (the big one): 280 dual socket quad core compute nodes Infiniband interconnect
  - Lilly (Little Willy): 22 dual socket quad core compute nodes Infiniband interconnect

#### This talk . . .

I will concentrate on Wilson, since Lilly is really just the little sister.

Infrastructure Compute Nodes Software

### Masternode

- Hewlett-Packard Proliant DL 385 G5 node:
  - 2 U rack mount server
  - 2 AMD 2356 QuadCore processors @ 2.3 GHz
  - Memory: 16 GB
  - 500 GB local SAS storage
  - 4 MSA 60 storagesystems (12 disks each):
    - $2\,\times\,500$  GB disks @ Raid 5 + 1 hot spare
    - $2\,\times\,1$  TB disks @ Raid 6  $+\,1$  hot spare
  - Red Hat Enterprise Linux 5.1

#### Used for

- NFS server
- Management & Monitoring
- Login node, job submission

< (□ )

Infrastructure Compute Nodes Software

### **Compute Nodes**

- Hewlett-Packard Proliant DL 165 G5 nodes:
  - 1 U rack mount server
  - 2 AMD 2356 QuadCore processors @ 2.3 GHz  $\Rightarrow$  2416 cores
  - Memory: 1 GB/Core  $\Rightarrow$  2.42 TB total
  - Infiniband HCA: Mellanox ConnectX MT25418  $4x\ DDR \rightarrow 20\ Gb/s$
  - 2 Hotswap 147 GB hard disk with Raid 1  $\Rightarrow$  44.4 TB local storage
  - Red Hat Enterprise Linux 5.1

Infrastructure Compute Nodes Software

## Cooling

#### Cooling Capacity per Rack

Each rack hosts 40 compute nodes  $\approx$  12 kW/rack Problematic using conventional air cooled racks

- 7 water cooled racks: HP MCS (Modular Cooling System) G1
- Maximum Cooling Capacity 35 kW
- Doors can be opened for maintenance
- Doors open automatically in case of cooling failure



Infrastructure Compute Nodes Software

- Infiniband:
  - The Switch:

Voltaire ISR 2012 288 Port IB Switch Voltaire ISR9024D-8032 24 Port IB Switch



- Ethernet:
  - 2 seperate Networks:
    - 1 GBit Network: login, storage connection ...
    - 1 100 MBit Network: serial console, virtual kvm, ...
  - Each rack has one switch for both networks
  - Bonded Uplink to master node

Infrastructure Compute Nodes Software

### Summary of the system



- 1 air cooled rack
- 36 TB gross storage
- 288 port IB switch
- 7 water cooled racks
- 280 compute nodes

Infrastructure Compute Nodes Software

### Experience with the hardware

Most problematic components:

- Hard disks in compute nodes (exchange 1 hd every 5 weeks)
- RAM: Usually a pile up of correctable ECC errors, but also uncorrectable ECC's
- System fans: Non redundant, i.e. if a fan is reported as missing system shuts down This happens frequently even though the fan is ok

#### In general

Hardware is robust and downtime due to hardware errors on the compute node side is negligible.

Infrastructure Compute Nodes Software

### Software

Compilers:

Openmpi, Mvapich, Mvapich2, GNU Compilers

- Almost all programs use Message Passing Interface (MPI)
- Batchsystem:

Wilson: "Nearest neighbour shouting"

Lilly: Torque, Maui

#### Big benefit of the machine

No restrictions to runtime and resources!

HP product: CMU Our Monitoring

### Outline



Hardware

- 3 Management & Monitoring
  - Benchmarks



< E

Cluster Management Utility (CMU) - Java program providing:

- Quick overview of Cluster Status
- Sensorinformation, e.g. cpu load, memory used, ...
- Several management tasks:
  - Power on/off compute nodes
  - Open serial console
  - Backup and clone nodes
  - Single and multiple window parallel shell
  - Simple alerting

HP product: CMU Our Monitoring

### CMU



・ロン ・部 と ・ ヨ と ・ ヨ と …

æ

HP product: CMU Our Monitoring

### **OS** Installation

Installation of the OS on compute nodes:

- CMU provides cloning mechanism
- Install one node by hand and backup hard disk  $\Rightarrow$  golden image
- $\bullet\,$  Distribute the golden image over the cluster  $\Rightarrow$  cloning
- Installation Steps:
  - Network boot the nodes to be cloned (1 master serves 1 slave)
  - Copy the golden image
  - Perform node specific configuration, e.g. IP-address
- Each cloned node serves as a master for another node
  ⇒ Cascading installation
- Masters and slave stay within a given network entity ⇒ reduction of network traffic

HP product: CMU Our Monitoring

### **OS** Installation

Installation of the OS on compute nodes:

- CMU provides cloning mechanism
- Install one node by hand and backup hard disk  $\Rightarrow$  golden image
- $\bullet\,$  Distribute the golden image over the cluster  $\Rightarrow$  cloning
- Installation Steps:
  - Network boot the nodes to be cloned (1 master serves 1 slave)
  - Copy the golden image
  - Perform node specific configuration, e.g. IP-address
- Each cloned node serves as a master for another node
  - $\Rightarrow \mathsf{Cascading} \text{ installation}$

#### Complete Cluster

Installation takes 1 hour with a 3 GB image.

HP product: CMU Our Monitoring

### Multiple Window parallel shell

# Useful for interactive multiple shell access, e.g. firmware update of $\mathsf{IB}\ \mathsf{HCA}$



Dalibor Djukanovic

The Wilson HPC Cluster

HP product: CMU Our Monitoring

### What do we use?

Intelligent Platform Management Interface (IPMI)

- Remote control of servers: Shutdown power off/on
  - boot order
- Collection of out of band information: CPU temperatures Fan speeds

Simple Network Management Protocol (SNMP)

# Challenge We have 280 nodes $\Rightarrow$ scalable solution needed

For monitoring we have

- Purchased hardware ⇒ usually Simple Network Management Protocol (SNMP), with vendor specific MIB
- Tailor made hardware  $\Rightarrow$  custom made readout in hardware and software

Purchased components giving sensor information via SNMP

- Water cooled racks
- UPS
- Switches
- Cooling system, i.e. chillers

Self-made

- Circuit breaker: Emergency power off, remote control via AVR micro controler
- Temperature and humidity sensors: attached to embedded systems, via serial to usb

HP product: CMU Our Monitoring

### Hardware for monitoring

Master node provides

- Centralized syslog server: syslog-ng
- SNMP-trap receiver

Additionally we have

- 2 virtual machines :
  - One machine runs conserver, archiving the console output of all machines
  - One machine constantly checks the air intake temperature of water cooled racks.
  - Initiate shutdown of cluster based on thresholds

#### What do we do with the information

We collect a lot of information and one quickly looses the overview!

HP product: CMU Our Monitoring

## Monitoring and Eventhandling

We want

- Framework to easily collect and archive data in form of rrd databases
- Graphical representation of data and other information via web interface
- Eventhandler for several user defined thresholds

We use

- Nagios
- Ganglia
- Cacti

HP product: CMU Our Monitoring

### Nagios

- Nagios is an extensible monitoring framework
- Nagios performs scheduled checks of all defined hosts and services, reporting the current state (Missing, Ok, Warning, Critical)
- Eventhandlers based on state change or actual state
- Nice escalation feature for notifications
- Another nice thing: Flapping detection

#### Example Lilly

Nagios checks rack temperatures of the Lilly Rack: Temperature at least 2 consecutive times  $> T_{crit}$  shutdown.

HP product: CMU Our Monitoring

### Ganglia

- Ganglia is a distributed monitoring system
- Daemon based datagatherer (gmond)
- Server collecting all information related to a cluster in rrd databases (gmetad)
- Nice feature:

Spoofing mechanism  $\Rightarrow$  inject data to the datastream of a certain node

We collect out of band information using the spoofing mechanism

HP product: CMU Our Monitoring

### Cacti

Webinterface for visualization of sensor information using rrd tools by Tobi Oettiker

- Easy to install and configure
- Creation of nice looking rrd graphs with a few mouse clicks
- Templates simplify graph creation for identical hardware
- Nice UI, e.g. zoomable graphs, timespan, ...
- Access control, using realms and per graph permissions
- Plugin architecture for extensions
- Consolidate interface for different packages, e.g. ganglia and nagios

HP product: CML Our Monitoring

### Cacti Sensor Information



Dalibor Djukanovic The Wilson HPC Cluster

HP product: CML Our Monitoring

### Cacti Nagios Interface



Dalibor Djukanovic The Wilson HPC Cluster

э

< 日 > < 同 > < 三 > < 三 >

HP product: CML Our Monitoring

### Cacti Ganglia



э

HP product: CMU Our Monitoring

### Cacti Syslog & SNMP

800			Castl- CAMM (CAst: Message Managment)			
			-C Q+ Compte			
and and and and	Desires Gauges					
Systems Description ( Distance)	-					Num la bar -
Adult Reports Dis Adventures						
		-				Tan Tan
-		domin't contain	- ket			and the second s
and p						-
10 M			8-1			
			Tank The	( Designer )		
			L'ANTING	CIT INTONE 1		
that at 1. Contract contract	Batters + M Suprittidant.					
1990 Januar Janu	C Arters Descripter	And .	and a	printly	Desiline -	manage
	C C C RETER	(arrest	-		2020-00-00 00-02111	SUTE capital that by it teaching
		diment.	-		2010/06/06 10:16:04	Dell' generales la Association
		19	-	. 499	200 10 10 10 10 10	-spittered average of the state of the second
		19	-	49	210.00.00.00.0	replaced protocol of 192.6.1 atomic 2
	E		-	1000	100010-0010-0110	Included the state (\$12504)
	- <b>0</b>		_		Same of Science of	salined south to the salestime
					THE OWNER WITH THE	April 75 Cens Laide
				. same	MOTORNE	terral FD Carlor tells rill regiment for Lasting
					and the state	The set of
			-		Dis C X R M R	Name PDI System medica cree
						Delle della dell
	C 4 0	104	4400	449	218-032-049-05	Annual and Processing Stationary at Society process Society and at 1912
	10 B B			and the second	2008 15:30 38:54 35	Terris PM to reduce cheek
	- <b>-</b>		100	144	20515121044	same and a find a find a first second second and a second se
	10 4 6	1N	100	( warning	DIGHTS TO BE MUST	ment projet fumine profess
	d 9 9	- 178	-	100	2010/10/06 08 04:00	Aurol Bernaria Farata
	5 G	10	141	i name	2006 15 20 20 54 54 20	Name Manual Contractor Film of Cases Participation Page.
	0.0	CN .	100	- sarry	2005-15-30 0054-25	Annual Add. may a 2016 and Sound Andreas and Angel
	0.4 6	19	Addr.	Contract of the local division of the local	2019/19/06 10:54:28	Name at writing previous hope with passe trading passes that
	24 C	19	100	manny	209 0 2 0 10 19	Served &1 years 2014 and source continues pay had these.
	C . C			-	2010/12/2010/02/201	samp RCT Regiment providence (F
	2 <b>2 9</b>	1.0	-	-	200412-02104104-031	Annel Registered RDNA yanget
	9 <b>4 9</b>	19	-	144	2010/02/02 02:02	Karran Regional RCNicy measure
	112.40	100	Auto-	144	285 15 10 10 14 28	Renal MC: Reprint provider beth 21

Dalibor Djukanovic The Wilson HPC Cluster

- 4 同 6 4 日 6 4 日 6

æ

HP product: CMU Our Monitoring

### In Case of Emergency

We have a three way shutdown procedure

- Software tries graceful shutdown
- Software initiates hard power off
- Circuit breaker coupled to thermostat:
  - $\Rightarrow$  physically disconnect power from rack

#### Last Line of Defense

Red Button switching off power





2 Hardware

- 3 Management & Monitoring
- 4 Benchmarks



< ∃ >

### With kind permission by Prof. Hartmut Wittig

#### Some preliminary benchmarks

- Benchmark based on typical lattice QCD operations: Application of Wilson Dirac Operator:  $\hat{Q}$ Norm squared of a vector:  $|| \cdot ||$ Linear combination: z = ax + y
- Weighted average gives synthetic QCD average performance
- Standard C Code (MPI)
- Optimization: SS3, prefetch
- Compare Hard scaling AMD 2356 with: BlueGene/L and BlueGene/P

### Hard scaling

- Increase the number of CPU's for a constant global problem size:
  - $\Rightarrow$  local problem size is decreasing with number of cpu's
- Two effects:

Problem might fit into cache

Surface/Volume ratio increases

- $\Rightarrow$  Communication becomes important
- We have results for 4 Lattices:  $192 \times 96^3$  $128 \times 64^3$ 
  - $96 \times 48^3$
  - $48 \times 32^3$

### Preliminary Hard Scaling: AMD vs BlueGene/L



### Preliminary Hard Scaling: AMD vs BlueGene/P





### Preliminary Hard Scaling: AMD vs BlueGene/L



### Preliminary Hard Scaling: AMD vs BlueGene/P





### AMD vs BlueGene

Clusterperformance in units of BlueGene/P racks:

 $\mathrm{Performance} = \frac{P_{\mathrm{AMDcore}} [\mathrm{MFlops}] \times 2240}{P_{\mathrm{BGPcore}} [\mathrm{MFlops}] \times 4096}$ 



Dalibor Djukanovic The Wilson HPC Cluster





2 Hardware

- 3 Management & Monitoring
- Benchmarks



< E



- Stable system with few downtimes
- 3.5 TFlops/User (Linpack)
- 2240 core Cluster running Lattice QCD code @ 3.7 TFlops sustained
  - $\Rightarrow 0.3 \in /MFlops$
  - $\Rightarrow$  Waste heat/cooling: 23 kW/TFlops
- Hard scaling shows up to 50 % drop (flatter curve for BlueGene)
   Need to compare with Nehalem system