The High Performance Cluster for Lattice QCD Calculations: System Monitoring and Benchmarking

Lucas Fernandez Seivane
quevedin@mail.desy.de

Michal Kapalka
kapalka@icslab.agh.edu.pl

Supervisor: Andreas Gellrich
IT Division

September 2002

A cluster?!?

- Lattice QCD calculations are **complicated**
- We need a lot of **power**, so what do we do?

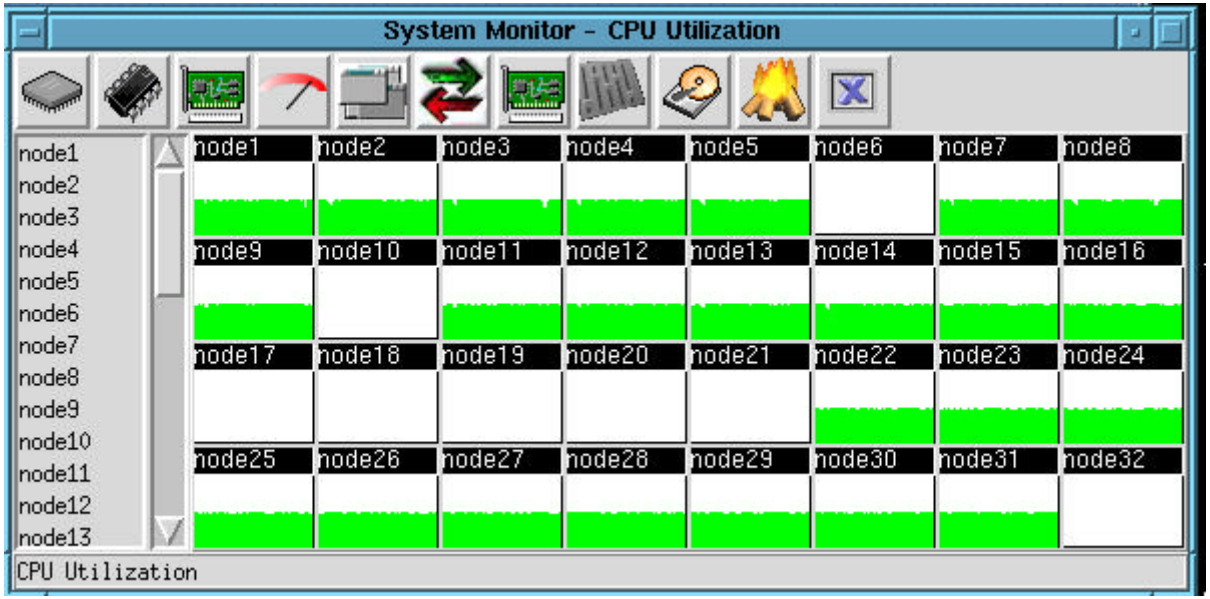
  Buy a **parallel computer**?

  Or maybe a **cluster** → good solution, but we have to **MANAGE** it!
- Manage = install software + test + **monitor** + **benchmark** + **analyse** + repair + etc.

Monitoring

- **Observing**, how things are going on
- **Measuring** CPU/network/whatever load
- Making **logs** (so nobody plays Quake on 500,000 $ cluster)
- And calling system administrator at 4 am when a node refuses working

Monitoring ‘lattice’ – before

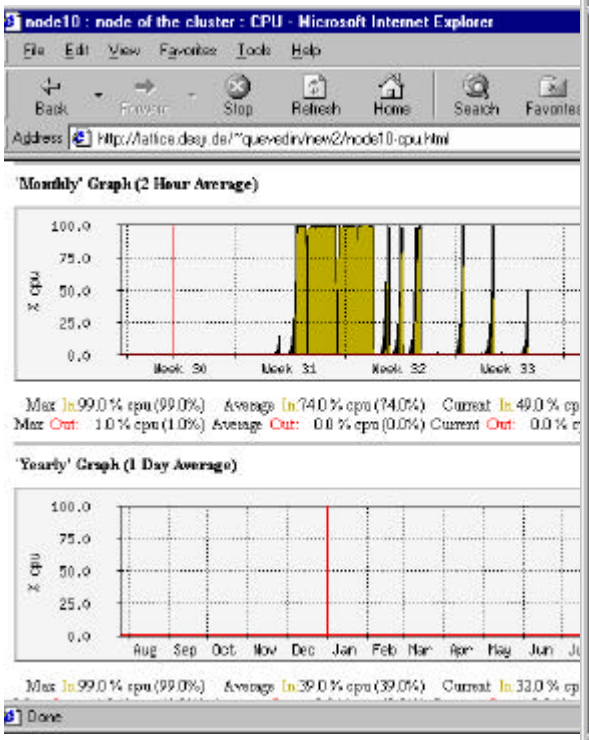


The lattice cluster = master + 32 nodes

Each node = dual Xeon CPU (16 x 1.7 GHz + 16 x 2.0 GHz), 1 GB RAM, 18 GB SCSI HDD

Network: Fast Ethernet + Myrinet

Monitoring – after



Lattice Cluster Monitor - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten

Adresse http://lattice.desy.de/cgi-bin/quevedin/cgi_clumon2.pl Wechseln zu Links

Lattice Cluster Monitor

Status: Mon Sep 2 17:55:00 2002 (Version: 2.0 of 31-08-2002)

Host	Date	Load	CPU	Mem [MB]	Swap [MB]	Fan	T [C]
Rack 1							
node1	Mon Sep 2 17:54:15 2002 45	1.00	0.99	952	1	5357	33.0
node2	Mon Sep 2 17:50:36 2002 264	1.04	0.99	377	0	5192	35.5
node3	Mon Sep 2 17:53:37 2002 83	1.00	0.99	365	0	5113	35.0
node4	Mon Sep 2 17:50:16 2002 284	1.01	0.99	364	0	5000	33.5
node5	Mon Sep 2 17:52:17 2002 163	1.07	1.00	363	0	5113	32.5
node6	Mon Sep 2 17:53:02 2002 118	0.00	0.00	608	0	5152	24.0
node7	Mon Sep 2 17:53:00 2002 120	1.08	0.99	366	0	5113	35.0
node8	Mon Sep 2 17:51:33 2002 207	1.02	0.99	360	0	5113	30.5
node9	Mon Sep 2 17:50:16 2002 284	1.00	0.99	361	0	5113	35.5
Rack 2							
node10	Mon Sep 2 17:53:00 2002 120	0.00	0.00	600	0	5113	28.0

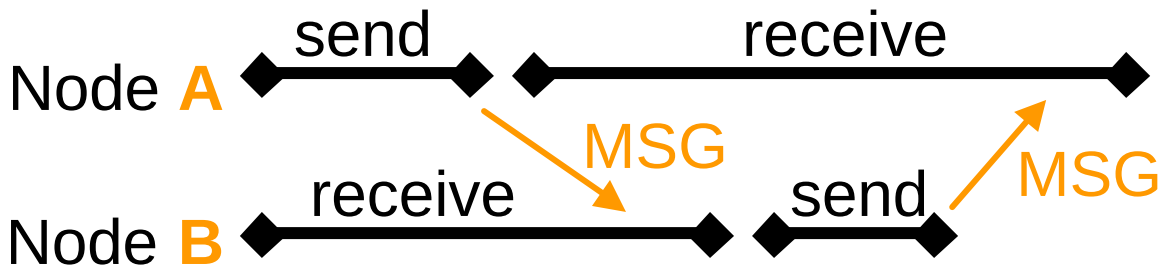
Internet

Benchmarking

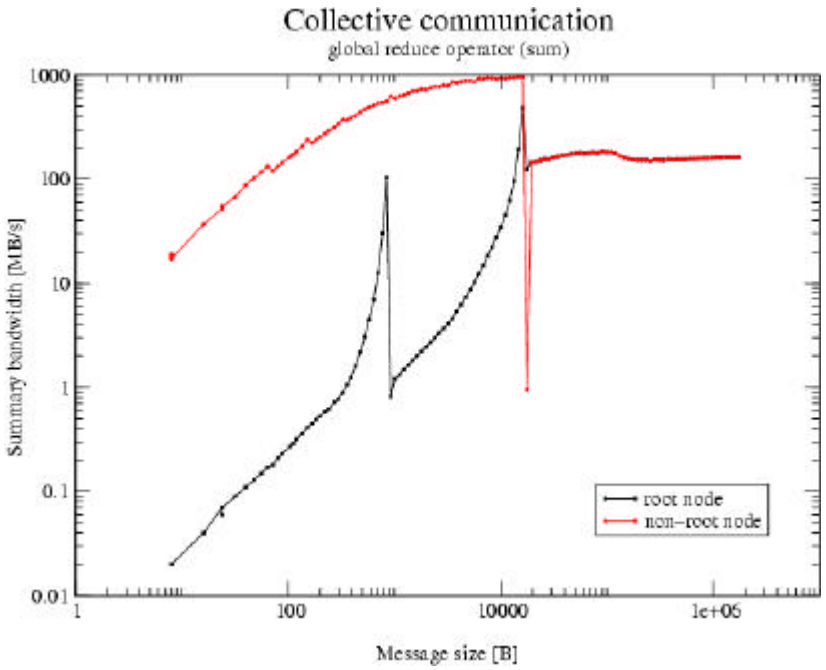
- **Benchmark** = compare or **test** (in order to improve things)
- On cluster we test:

  single nodes (CPU, memory,

  communication between them (here: MPI)
- **MPI** a standard → sending data = passing **messages** between processes



Benchmarking – results



Interpretation
is usually …

… NOT **trivial** ☹