



## Diskless Cluster und Lustre – Erfahrungsbericht zum CHiC

Frank Mietke

Fakultätsrechen- und Informationszentrum (FRIZ) /  
Professur Rechnerarchitektur  
Technische Universität Chemnitz

Chemnitzer Linux-Tage 2007



# Inhaltsverzeichnis

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- 1 HPC in Chemnitz
- 2 CHiC – Projekt
- 3 Clusterarchitektur
- 4 Software-Umgebung
- 5 Und nun?



# Hochleistungsrechnen

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- Zusammenlegen von:
  - Prozessorleistung
  - Hauptspeicherkapazität
- High Performance Computing (HPC)
- High Throughput Computing (HTC)
- Anwendungen: Simulation und Datenverarbeitung
- Top500 Liste (öffentlich)



- Chemnitzer **L**inux **C**luster – CLiC (2000)

- Chemnitzer **H**ochleistungs **L**inux **C**luster – CHiC (2007)

221,6 GFlop/s



8,21 TFlop/s





- Mainstream Cluster
- Cutting-Edge Cluster
- Bleeding-Edge Cluster
- Vision

**HPC ist Testfeld für neue Technologien**



# Projektstart

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- Erste Gespräche Mitte 2003
- CHiC-Konsortium (24 Professuren)
- HBFVG Antrag April 2004
- Antragssumme 2,64 Mio Euro
- Genehmigung, Ausschreibung und Aufbau 2006



# Ziele des Projektes

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- Mindestens 1000 Prozessoren
- Hochgeschwindigkeitsnetzwerk
- Eigenbau/-design
- Open Source Software
- Verbesserung der Stabilität des Systems  
(Festplatte, Netzteil, Hauptspeicher)
- „Geringer“ Leistungsverbrauch ( < 200KW )



# Bauprojekt

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- Begrenzte Kühlleistung im Serverraum  
(→ wassergekühlte Schränke)
- Tragkraftproblematik
- Raumrekonstruktion ca. 1,7 Mio Euro

## Raumaufbau





# Clusteraufbau

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?



- 1 Monat Aufbauzeit
- 21,6 Tonnen Material (Schränke + Inhalt)
- 4200 Muttern und 4600 Schrauben notwendig
- 4900 Kabel mit 9800 Steckverbindern (8km Länge)
- 300 Manntage Aufwand



# Clusterarchitektur

CHiC-Bericht  
CLT-2007

Frank Mietke

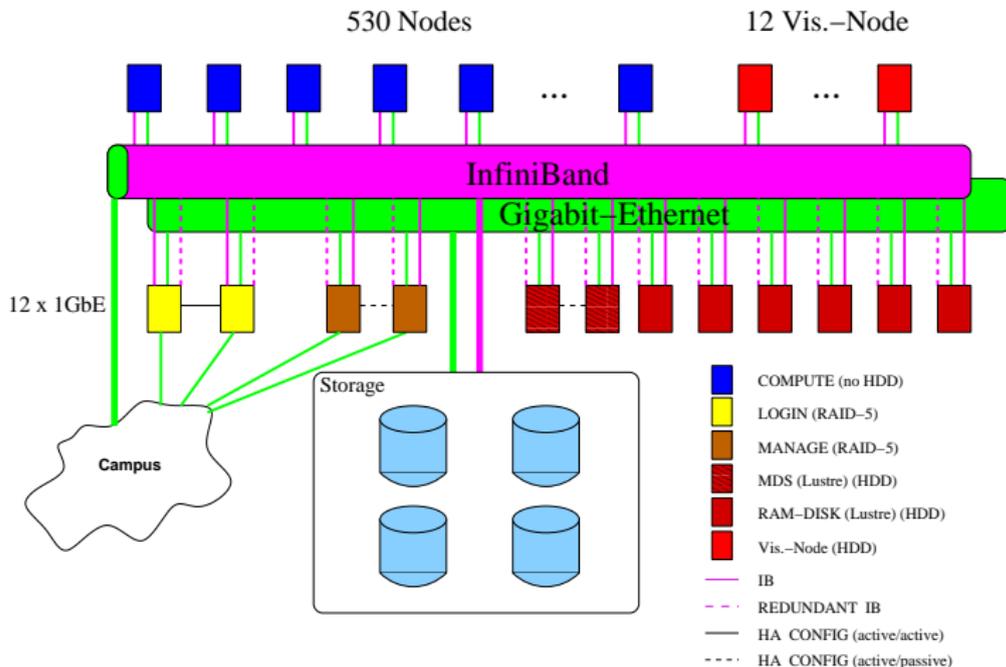
HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?





# Cluster Hardware

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- IBM Server Technologie
- AMD Opteron Dual Core Rev. F
- Voltaire InfiniBand 10Gbit/s
- 288-Port IB-Switche
- NVIDIA Quadro FX 4500 X2





# Storagearchitektur

CHiC-Bericht  
CLT-2007

Frank Mietke

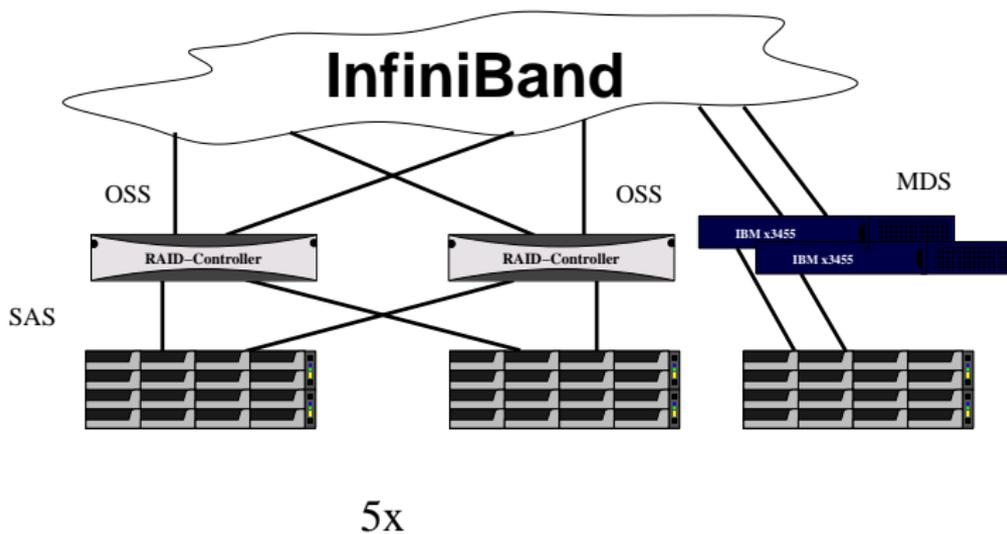
HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?





# Storage Hardware

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?



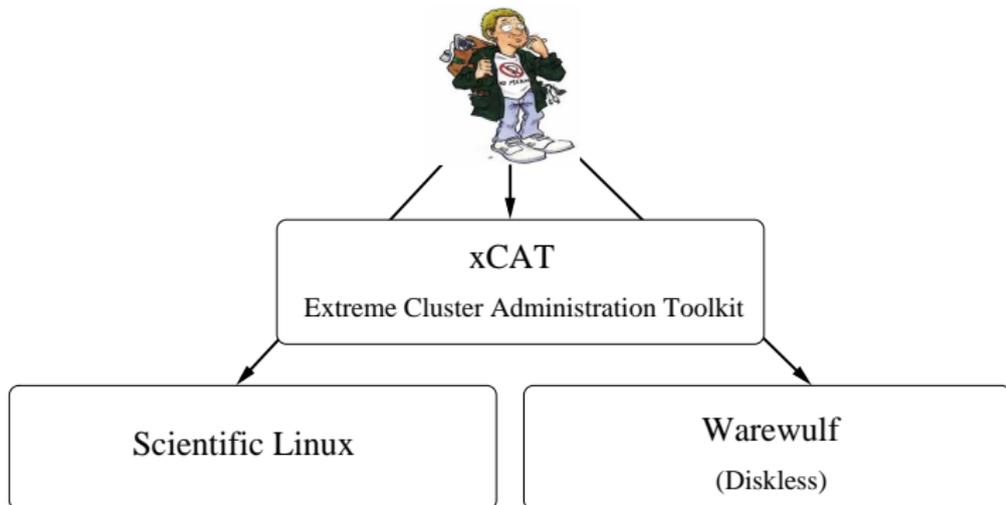
- 160x SATA (OSS)
- 16x SAS (MDS)
- 3,5GB/s (write)
- 1,7GB/s (read)
- RAID-5



- Scientific Linux 4.4
- Open Fabcris Enterprise Ed. 1.1
- Lustre 1.6 Beta 7
- Open MPI und MVAPICH
- GNU Compiler, EKOPath Compiler
- LoadLeveler Batchsystem



- xCAT ([www.xcat.org](http://www.xcat.org))
- Warewulf ([www.warewulf-cluster.org](http://www.warewulf-cluster.org))
- Scriptsammlungen/Konfigurationsdateien
- schlechte Dokumentation





# Schritte zur Inbetriebnahme (xCAT) – 1/3

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- 1** Definition des Clusters:
  - Knotennamen, Netzwerke usw.
  - 20+ Konfigurationsdateien
  - XX angepasste Scripte
- 2** Konfiguration der Switchtechnik:
  - Ethernet
  - InfiniBand (Switch + Gateway)
  - Terminal Server
  - Global Console Server
- 3** Sammeln der MAC-Adressen (Ser.Cons. oder Switch):
  - `$getmacs <noderange>`
  - `$makedhcp -allmac`



# Schritte zur Inbetriebnahme (xCAT) – 2/3

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- 1 Installations- und Imagetemplates:
  - SL-4.4
  - Warewulf
- 2 BIOS/FW Update 500+Knoten (IPMI)
- 3 Warewulf Knotenimage (Größe)
- 4 Serverskalierung (Dienste)
  - Offene TCP-Verbindungen
- 5 Synchronisierung der Knoten



# Schritte zur Inbetriebnahme (xCAT) – 3/3

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

- 1 Paralleles Dateisystem (Lustre)
  - Aufsetzen der RAIDs
  - Konfiguration des Metadatenserver (RAID-10 + Aktiv-Passiv HA)
  - Konfiguration der Object Storage Server (RAID-5 + Aktiv-Aktiv HA)
  
- 2 Software:
  - Compiler
  - MPis
  - Mathe-Bibliotheken
  - kommerzielle Nutzer-Software



## Fehler sind die Regel!

- Nagios (Client/Server)
- Zugriff über SSH/IPMI/Ser.Console/APC
- Auftretende Lastspitzen im Managementknoten
- Kommunikationsaufkommen beachten



# An die HPC-Front!

CHiC-Bericht  
CLT-2007

Frank Mietke

HPC in Chemnitz

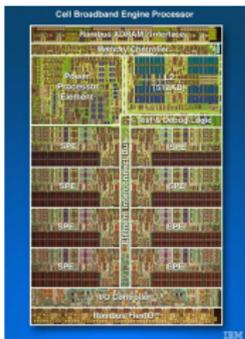
CHiC – Projekt

Clusterarchitektur

Software-  
Umgebung

Und nun?

Studien- und Diplomarbeiten sowie wiss. Mitarbeiter



<http://www.tu-chemnitz.de/informatik/RA/>



**Danke für Ihre Aufmerksamkeit**

**Clusterführung heute 18:30 Uhr**

**Mehr Informationen am IBM/Megware-Stand**