

# Einführung in Cluster Computing

Robin Geyer, Stefan Höhlig, Jupp Müller

21. Januar 2009

Cluster bieten schon lange die Möglichkeit rechenintensive Probleme zu lösen. Durch immer preiswertere Hardware und die stete Entwicklung freier Software für Cluster, werden diese auch für Privatanwender oder kleine Unternehmen immer interessanter. Dieser Vortrag soll deshalb einen Einblick in diese Thematik geben: was ein Cluster eigentlich ist, Wahl der Distribution/Werkzeuge, Power Management, Verbindungsnetze, Software zur Verwaltung, Monitoring, Programmiermodelle, Probleme und Hürden etc. Dies und unsere Erfahrungen und Eindrücke von der Cluster Challenge 2008 sollen Starthilfe geben und Lust zum Experimentieren wecken.

## 1 Einführung

### Motivation HPC

Vielfältige ingenieurtechnische und wissenschaftliche Probleme lassen sich gut im Experiment untersuchen. Experimente werden mit steigender Komplexität aber immer teurer und müssen im Zweifelsfall mehrmals durchgeführt werden. Außerdem fallen dabei enorme Datenmengen an, die ausgewertet werden müssen. Versuche deren naturwissenschaftlicher Hintergrund hinreichend genau durch Modelle abgebildet werden kann, können auf Hochleistungsrechnern Simuliert werden. Kosten für aufwändige Experimente wie Crashtests, Windkanaltests oder Wachstumsexperimente entfallen dabei. Erst durch High Performance Computing-Systeme können die bei neuesten Versuchen entstehenden enormen Datenmengen überhaupt erst ausgewertet werden, wie zum Beispiel im LHC<sup>1</sup>.

Daneben bieten HPC-Systeme beispielsweise für Unternehmen hervorragende Möglichkeiten komplexe Aufgaben schnellstmöglich zu erledigen. Man denke an das Architekturbüro welches 3D Animationen zukünftiger Gebäude rendern muss oder an den Hersteller von Modellflugzeugen der seine Modelle ganz bequem am Computer testen kann.

Der Vortrag soll die Grundlagen zu folgende Fragen klären.

## 2 Cluster: Basics, Architektur und Vorüberlegungen

### Klärung der Begriffe

Was ist eigentlich ein Cluster? Wie ist ein Cluster aufgebaut? Was ist der Unterschied zwischen Capacity- und Capability Computing?

### Grundsysteme

Eingegangen wird auf verschiedene Rechnerarchitekturen aus denen man Cluster bauen kann. Habe ich überhaupt eine Wahl? Brauche ich Speicher, Festplatten und Grafikkarten? Wie viel Strom braucht ein Cluster? Wie kann ich Strom sparen? Wenn nicht selbst gebaut, woher kann ich solche Systeme beziehen? Was kostet ein Cluster in seinem Lebenszyklus?

### Verbindungstechnik(en)

Der Flaschenhals zwischen den Recheneinheiten ist meist das Verbindungsnetzwerk. Welche Technologien gibt es? Was ist Myrinet, Infiniband, Quadrics und 10GBE? Welche werden üblicherweise eingesetzt? Muss es immer ein Hochgeschwindigkeitsnetz sein? Nach welchen Kriterien kann man sich entscheiden?

---

<sup>1</sup>Teilchenbeschleuniger im Genfer Kernforschungszentrum

## **Anwendungsszenarien**

Gibt es den Alleskönner? Für wen eignet sich ein Cluster? Was kann ich tun, wenn eine eigene Maschine zu teuer ist? Wann ist ein Cluster für meine Zwecke ungeeignet? Was sind die Alternativen zu einem klassischen Cluster?

## **3 Systemsoftware**

Abhängig vom Grundsystem zeigen die Erfahrungen, dass es hinsichtlich der Systemsoftware einiges zu beachten gibt.

### **Betriebssystem / Systemsoftware**

Kann ich mein Cluster ausschließlich mit OpenSource Software betreiben oder muss ich kommerzielle Lösungen mit einbeziehen? Welche Linux-Distribution soll ich nehmen? Was geht, was nicht? Wie sieht das mit Windows HPC Server aus? Wie gut sind (GNU)-Compiler?

### **Kommunikationsbibliotheken / Programmiermodelle**

Wie will ich zwischen den Knoten kommunizieren? Muss ich das überhaupt? Was ist MPI, Cluster OpenMP und Co?

### **Verwaltung / Administration**

Was ist ein Batchsystem? Brauche ich so etwas überhaupt? Gibt es das rundum-sorglos-ich-administrier-mein-Cluster-in-10-Minuten Tool?

### **Monitoring und andere Helferlein**

Wie kann ich Systemauslastung oder Verfügbarkeit überwachen? Welche OpenSource Lösungen gibt es? Was macht das Leben mit dem Cluster einfacher?

## **4 Anwendungssoftware / Optimierung**

### **existierende Softwarelösungen verwenden**

Was gibt es für OpenSource Simulations-Software?

### **Software selbst entwickeln**

Wo kann ich paralleles Programmieren lernen? Wie kann ich mir es selber beibringen, wo fange ich an? Wie entwickle ich am besten? Wie kann ich debuggen?

### **Optimierung**

Ich habe bereits parallele Programme, wie finde ich heraus ob sie gut skalieren? Wie erkenne ich Fehler? Ist es möglich Programme zu optimieren?

## **5 Ausblick**

Was ist der aktuelle Stand der Forschung? Wie sieht die Zukunft der Cluster aus?