

# Über Communities mit rechenintensiven Aufgaben

---

Gabriel Stöckle (Uni Heidelberg)  
Ilya Agapov (DESY)

Virtuelle Forschungsumgebungen aufbauen – mit D-Grid

- CFD (Klima, Luft- und Raumfahrt, MHD, Energie,...)
- HEP (Datenanalyse, Detektor und Beschleunigerentwicklung)
- Biostatistik, Medical Imaging
- Astrophysik
- Media (Werbung, Film)

- Jülich, Oak Ridge, usw. (mehr als 1000 Knoten)
- Pros
  - Einzigste Möglichkeit für manche Probleme,  
z.B. komplexe CFD (Strömungssimulation)
- Contras
  - Zeit teuer
  - Begrenzte Verfügbarkeit
  - Begrenzte Infrastruktur

- Z.B. in Uni-Rechenzentren (weniger als 1000 Knoten)
- Pros
  - Gute Verfügbarkeit und Leistung wenn  
Kapazität  $\gg$  Nutzung
- Contras
  - Investition (Manpower, Infrastruktur)
  - Schlechte Verfügbarkeit Leistung wenn  
Kapazität  $\gg$  Nutzung
  - Nichtoptimale Nutzung (1-5% Auslastung)

- Rechenzeit ankaufen (z.B. Amazon EC)
- Pros
  - Gut für viele Aufgaben, z.B. WPA key hacks
- Contras
  - Nur standarte Infrastruktur verfügbar (MySQL, Apache usw.)
  - Bandbreite zu den Experimental-Instrumenten

- Resource Sharing und gemeinsame Infrastruktur
- Pros
  - Anpassbar
  - Kollaborative workflows möglich (LHC)
- Contras
  - Aufwand für Service-Betreiber

# Rechenintensiv ist oft kollaborativ

---

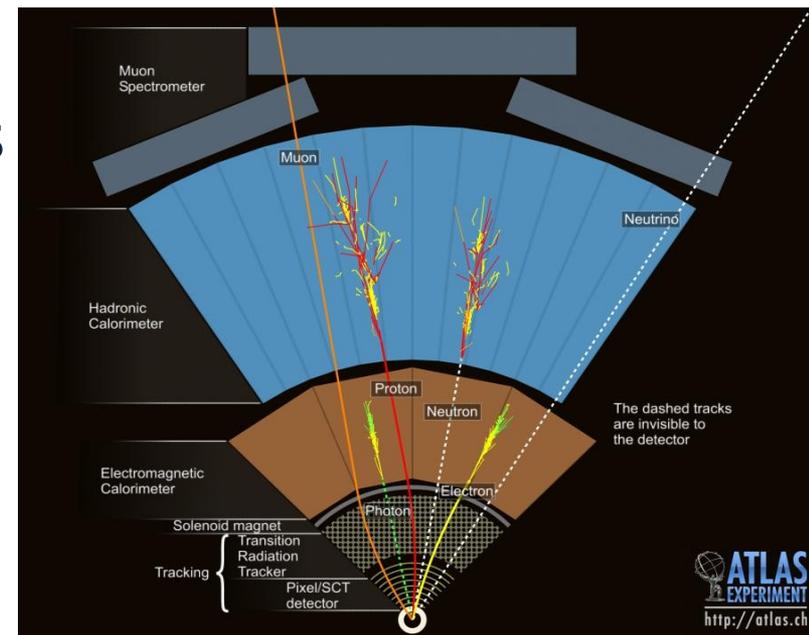
- Z.B. mehrere 1000 Leute am CERN
- Zugangsmechanismen, Metadaten, Verteilte Files, – teil der Umgebung
- Zugang zu Instrumenten (Teleskopen, Synchrotronen, usw.)
- Ermöglicht durch Grid-Middleware und Community-spezifische Lösungen

# Rechenintensiv ist oft datenintensiv

---

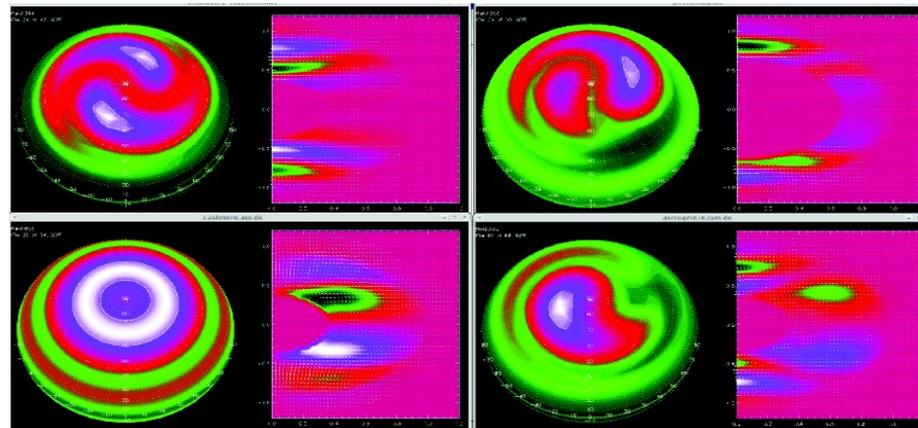
- Grosse Datensätze erzeugt
- Worldwide Zugang oft erwünscht (gridftp)
- Parallele lokale Filesysteme benötigt

- LHC experimente (ATLAS, CMS) – Partner aus aller Welt
- PB von Daten, Monte-Carlo und Reconstruction
- Grid – einzige Möglichkeit (um 2000)
- Multi-Tiered Architektur
- Andere HEP Communities (ILC)



- AstroGrid-D – Projekt, um die deutsche astronomische Community ins Grid zu bringen
- 14 beteiligte Institute
- Aufbauend auf der D-Grid Initiative und deren Infrastruktur und Standards
- Middleware: Globus Toolkit (GT4)
- Einige Zahlen:
  - Mehr als 100 user
  - Mehrere Compute nodes mit gesamt 1000 cores
  - Speicherkapazität von über 100 TB

- Parameter Studien/Atomic Jobs (Microlensing, Berechnung von Modellatmosphären, Magneto-Hydro-Dynamik): kleine, ähnliche Jobs werden gleichmäßig auf freie Ressourcen verteilt
- Steuerung von robotischen Teleskopen



- Rechenintensive Aufgaben in Klima, HEP, Astronomie, benötigen verteilte Umgebungen
- Grid-Lösungen sind dazu gut geeignet
- WissGrid/DGrid bieten dabei Unterstützung and (siehe weitere Präsentationen)