

# GridKa: Das Grid Computing Zentrum Karlsruhe für Anwendungen aus der Hochenergiephysik

J. van Wezel, H. Marten, IWR

## Einleitung

Die deutschen Elementarteilchenphysiker veröffentlichten im Jahr 2001 ihre Anforderungen an ein Regionales Daten- und Rechenzentrum in einem Memorandum. Der Vorschlag, Rechenkapazitäten in einem Zentrum zu bündeln, um bevorstehende Anforderungen des *Large Hadron Colliders* (LHC) bei CERN zu meistern, wurde vom Forschungszentrum Karlsruhe schon im Dezember 2001 mit einem Plan und der Zusage zum Bau und Betrieb eines Groß-Rechenclusters beantwortet [1]. Aus den Anfängen als „*Regional Data and Computing Centre*“ Karlsruhe ging im Oktober 2002 das *Grid Computing Centre* Karlsruhe mit dem Namen GridKa hervor (Abb. 1). Das GridKa-Zentrum ist verbunden mit dem Institut für Wissenschaftliches Rechnen (IWR) des Forschungszentrums und ist in den Helmholtz-Forschungsbereich „Struktur der Materie“ eingebettet.

## Im GridKa arbeiten viele Forschungseinrichtungen zusammen

Der Aufbau eines der größten Computercluster in Deutschland ist ein langjähriges und komplexes Projekt. Nach der Bau- und Testphase wird der eigentliche Betrieb mindestens 15 Jahre dauern und erfordert ein ständiges Anpassen und Modernisieren der Hard- und Software und gegebenenfalls Erweitern der Komponenten für viele Jahre im Voraus. Die Erfahrung des IWR wird hierbei unterstützt von verschiedenen Experten. Im *Projekt Overview Board* begleiten und steuern Vertreter aus Physikausschüssen, BMBF und der Helmholtz-Gemeinschaft den Aufbau. Ein zweites Gremium mit Vertretern aus der Physik von allen beteiligten Universitäten und des Deutschen Elektronen-Synchrotrons DESY stärkt die technische Abstimmung zwischen Benutzerwünschen und Realisation.

Die Benutzer des GridKa-Zentrums sind in erster Linie Wissenschaftler der Deutschen Hochenergiephysik. Über Partnerschaften mit Institutionen wie den Helmholtz-Zentren DESY, GSI und Forschungszentrum Jülich oder CERN in Genf, RAL in Oxford, aber auch SLAC und FNAL in den USA, werden Erfahrungen ausgetauscht. So bekommen auch ausländische Benutzer die Möglichkeit, bei GridKa zu rechnen.

Die Rechenleistung des GridKa-Clusters wird vorerst für die Hochenergiephysik eingesetzt. Es wird aber ausdrücklich vorgesehen, künftig auch Kapazitäten für Disziplinen außerhalb der Teilchenphysik bereitzustellen, z.B. für die Astronomie, Astroteilchenphysik, Klimaforschung und Genetik.

## Messungen am Large Hadron Collider (LHC)

Der Large Hadron Collider ist ein Teilchenbeschleuniger, mit dem es möglich sein wird, die Bausteine der Materie weiter zu entschlüsseln und fundamentale Fragen aus der Physik zu beantworten. Der LHC wird am Europäischen Laboratorium für Teilchenphysik CERN in Genf in der Schweiz gebaut und benutzt den bestehenden Tunnel des früheren LEP-Ringes. Der unterirdische Beschleuniger hat einen Kreisumfang von 27 km und beschleunigt zwei Strahlen in entgegengesetzter Richtung. Die starken Magnetfelder, die die Teilchen auf ihrer Bahn halten, werden von supraleitenden Spulen bei Tempe-



Abb. 1: Seit der Einweihung schmückt sich das Grid Computing Centre Karlsruhe mit dem GridKa-Logo.

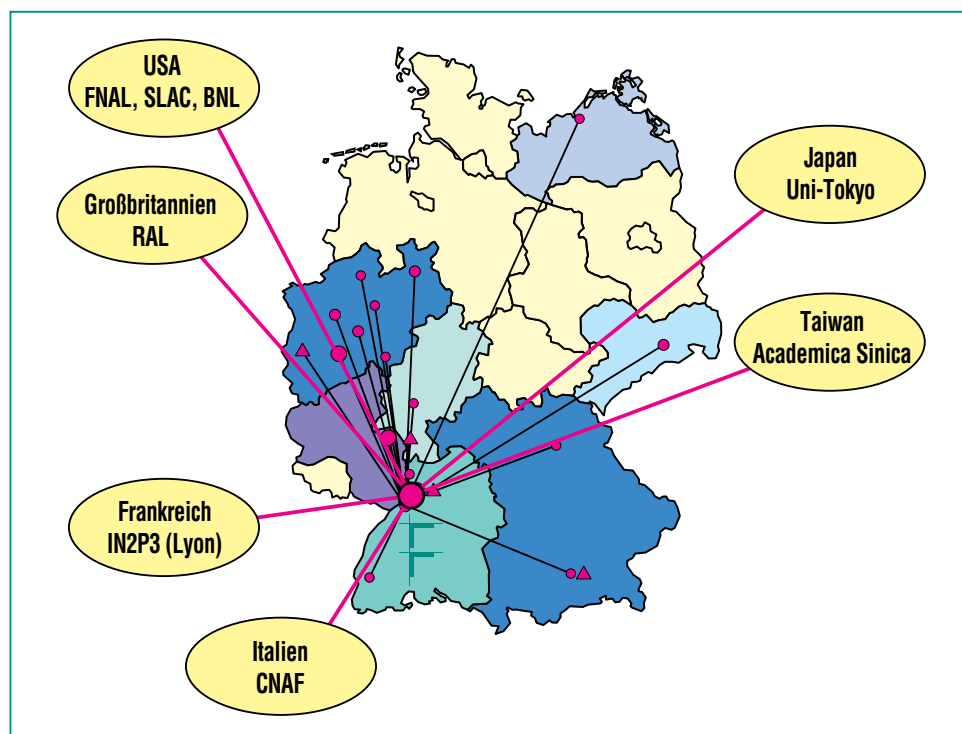
raturen um  $-270^{\circ}\text{C}$  erzeugt. Ab 2007 werden im LHC alle 25 ns Protonen mit einer Energie von 14 TeV aufeinander treffen. Daneben werden auch Strahlen von Bleikernen beschleunigt und kollidieren mit einer Aufprallenergie von 1150 TeV. An den Kollisionspunkten werden Detektoren aufgestellt, um die Reaktionen beim Aufeinandertreffen der Protonenstrahlen aufzuzeichnen. Die Experimente werden in bislang unerschlossenen Energiebereichen nach neuen Teilchen suchen. Das so genannte Higgs-Teilchen und der Ursprung der Masse sind grundlegende Fragen, die von den LHC-Experimenten erforscht werden.

Jeweils optimiert für spezifische Messungen kommen vier riesige Detektoren zum Einsatz. Neben der Konstruktion und dem Aufbau der Detektoren ATLAS, CMS, ALICE and LHCb werden auch Entwürfe für Software und Computeranlagen zur Speicherung und späteren Auswertung der Messdaten erarbeitet. Die Aufzeichnung und Auswertung der Daten stellt eine gewaltige Herausforderung dar: im LHC werden in jeder Sekunde rund 800 Millionen Teilchenkollisionen stattfinden. Das bedeutet, dass von Zehntausenden von Rechnern viele Petabyte Daten pro Jahr verarbeitet werden müssen. An dieser Verarbeitung ist das GridKa-Zentrum des Forschungszentrums Karlsruhe maßgeblich beteiligt und wird zusammen mit weltweit 7 bis 8 weiteren Rechenzentren den Kern des LHC Computing Grid bilden.

## Regionale Zentren

Ein einziges Rechenzentrum am CERN wäre logistisch und finanziell nicht in der Lage, die Aufgaben des LHC-Computing zu erfüllen. Wie Satelliten werden deshalb regionale Zentren jeweils einen Teil der Rechenleistung übernehmen. Der Report „Regional Centre Category and Service Definition“ [6] beschreibt Kategorien, basierend auf notwendigen Dienstleistungen für Rechenzentren, die am LHC-Computing teilnehmen. GridKa steht als Kategorie-1-Zentrum in einer Rechner- und Speicherkette, die bei CERN (Kategorie 0) anfängt und bei Kategorie 4, dem Arbeitsplatzrechner der Teilchenphysiker, endet. Zu der Gruppe von Kategorie-1-Zentren

gehören unter anderem IN2P3 in Lyon (F), RAL (GB) und FNAL in Batavia/Chicago (USA). Ein Kategorie-1-Zentrum (oder auf Englisch: Tier-1) wird neben Rechen vor allem Speicherkapazität anbieten und über eine direkte Hochgeschwindigkeitsnetzwerk-anbindung zum CERN die Messdaten entgegennehmen und speichern. Weiterhin wird von einem Kategorie-1-Zentrum eine sehr große Zuverlässigkeit und Betreuung rund um die Uhr erwartet. Die Kategorie-2-Ebene werden Zentren an Universitäten, Hochschulen und Forschungseinrichtungen bieten. Die Anforderungen an Benutzerunterstützung und Verfügbarkeit sind für Kategorie-2-Zentren geringer als für Zentren der Kategorie 1 (Abb. 2).



**Abb. 2:** Der Clusterrechner am GridKa wird betrieben für Universitäten und Forschungseinrichtungen in Deutschland. Innerhalb des LHC-Computing-Grid-Projektes arbeitet GridKa weltweit zusammen mit verschiedenen Kategorie-1-Zentren.

## Rechen- und Speicheranforderungen des LHC

Insgesamt vier große Detektoren sind im Bau zur Messdatenerfassung. Beispielsweise hat der ATLAS-Detektor etwa 200 Millionen Lesekanäle. Im Betrieb liefern diese 40 TB Messdaten pro Sekunde, die online über eingebaute Prozessoren und eine Rechneranlage auf 100 MB/s reduziert werden. Zusätzlich werden noch *Condition Data* gespeichert, die den Zustand des Detektors im Moment der Messung beschreiben. Man schätzt, dass der ATLAS-Detektor pro Jahr ca. 3 Petabyte Daten liefert, die auf Band gespeichert und an die Kategorie-1-Zentren verteilt werden müssen. Dies bedeutet, dass das GridKa-Cluster über lange Zeiträume Daten mit über 250 MegaByte pro Sekunde entgegen nehmen muss.

Die Rohdaten aus den Detektoren werden in erster Linie zur Rekonstruktion der Kollisionen verwendet, eine sehr rechenintensive Aufgabe. Ergebnis dieser Rekonstruktion sind die *Event Summary Daten* (ESD), die eine möglichst genaue Wiedergabe der Kollision enthalten. Die ESD müssen ständig online zur Verfügung stehen. Eine weitere Rechenaufgabe ist die Simulation der Messungen in einem Monte-Carlo-Verfahren. Die simulierten Messdaten werden später zur Auswertung der wirklichen Daten gebraucht.

Für alle Experimente liegen Planungen der Datenmengen und der Anforderungen an Rechenkapazität vor. Für die Datenspeicherung wird unterschieden zwi-

schen on-line auf Platte und offline auf Band gespeicherten Daten. Die meisten Daten werden nach der ersten Analyse nur noch selten gebraucht und müssen darum nicht auf teuren Platten gehalten werden. Die Aufbewahrung auf Band ist um ein Vielfaches billiger. Zu erwarten ist allerdings, dass sich in der Praxis die Anforderungen an Rechenzeit und Datenspeicher aufgrund neuer technologischer Entwicklungen in der Hardware und neuer Erkenntnisse in der Datenauswertung im Grid noch verschieben werden.

## Andere Experimente

Der LHC wird erst im Jahr 2007 in Betrieb gehen. Bis dahin arbeiten die vier LHC-Experimente zur Vorbereitung und zu Testzwecken mit simulierten Daten. Die deutsche Teilchenphysik ist allerdings nicht nur am LHC beteiligt, sondern auch an drei Experimenten in den USA engagiert: das sind BaBar am Stanford Linear Accelerator (SLAC) [3], sowie D0 und CDF am Fermi National Accelerator

Laboratory (FNAL) in Batavia/Chicago [4]. Ein weiteres Experiment außerhalb des LHC-Komplexes ist COMPASS am Super Protonen Synchrotron SPS des CERN [5], das als Vorbeschleuniger zum LHC weiter in Betrieb bleibt. Die meisten Experimente haben eine eigene Software entwickelt, die es erlaubt, weltweit verteilte Rechenzentren für Datenspeicherung und Analyse zu benutzen. Seit seiner Gründung hat GridKa einen signifikanten Anteil seiner stetig wachsenden Rechenkapazität für diese Nicht-LHC-Experimente zur Verfügung gestellt. Im Laufe der kommenden Jahre wird dieser Rechenzeitanteil von dem der LHC-Experimente übertroffen werden (Abb. 3).

## Grid und LCG

Über mehrere Jahre hatte das „Models of Networked Analysis at Regional Centres“ (MONARC) Projekt Basismodelle für das zukünftige LHC-Computerverfahren geliefert und ein Toolset zur Simulation von Daten produ-

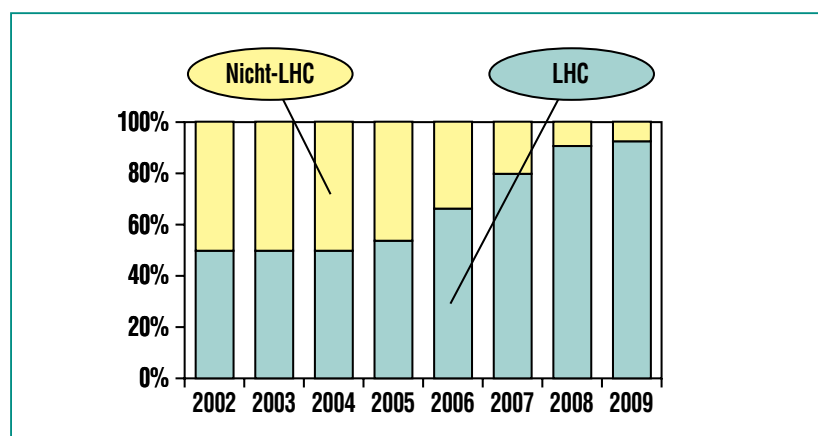


Abb. 3: Der Rechenzeitanteil der LHC-Experimente wird bei GridKa in den nächsten Jahren gegenüber demjenigen der Nicht-LHC-Experimente stetig anwachsen.

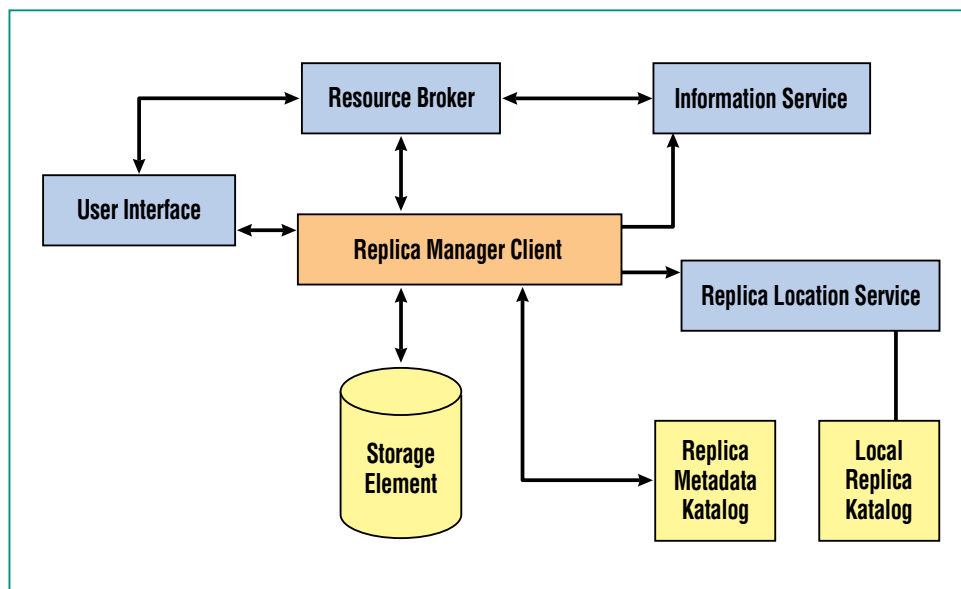
ziert, mit dem die Methoden für verteiltes Rechnen und Datenspeicherung ausprobiert werden konnten. Die MONARC-Studie hat zugleich Richtlinien für die technische Ausrüstung, Konfiguration und Dienstleistung von regionalen Computer-Zentren vorgeschlagen und damit die Basis für die Ausrüstung des GridKa-Zentrums gelegt.

Die Möglichkeit, Rechen- und Speicheraufgaben weltweit über Zentren zu verteilen und damit ein Computing-Grid zu bilden, ist die Aufgabe des LHC-Computing-Grid-(LCG)-Projektes. Innerhalb LCG werden die Entwicklungen der so genannten *Grid Middleware* zusammengeführt. Die *Middleware*-Software-Schicht bildet, oberhalb der Betriebssysteme der individuellen Rechner, mit Bibliotheken und Werkzeugen den Rahmen für die Anwendersoftware der LHC-Experimente. Die LCG-Software enthält Komponenten zur Benutzer-Authentifizierung über Virtuelle Organisationen, für eine Benutzerschnittstelle zum Abschicken und Verwalten von Programmen und zur Datenverwaltung und zum Datenaustausch über *wide area* Netzwerke (Abb. 4).

### LCG und GridKa

LCG hat im März diesen Jahres Version 2 der Middleware herausgegeben, die auf dem GridKa-Cluster installiert wurde. Der Aufbau liegt damit gut in der Planung.

Mit regelmäßigen *Data Challenges* oder Probeläufen wird nicht nur die Funktionsfähigkeit der Software getestet, sondern auch



**Abb. 4: Aufgabe der LCG „Replica Manager“ Middleware ist das Datenmanagement. Dateien und ihre Speicherorte werden in so genannte Katalogen gehalten.**

die Stabilität und Leistungsfähigkeit der einzelnen Kategorie-1-Zentren. Die Experimentgruppen setzen hierzu eigene Applikationen ein und benutzen die LCG-Software zum Starten der Programme und zur Verwaltung der zu verarbeitenden Daten.

Die LCG-Softwareentwickler werden künftig immer häufiger neue Versionen bereitstellen, weil jetzt alle Komponenten erstmals zusammen eingesetzt werden können. Ein wichtiger Teil der Arbeit am GridKa ist die Installation und Verifikation neuer Versionen oder Teilen von Versionen sowie die Rückkopplung mit den Entwicklern und Anwendern. Es entsteht ein Zyklus von Anwendung über Fehlerbehebung bis zur Herausgabe einer neuen Softwareversion. Die Cluster an den Zentren sind nicht alle gleich, eine fehlerfreie Funktion muss aber in vielen unterschiedlichen Einrichtungen

gewährleistet sein. Daher ist die Testphase für das LCG sehr wichtig.

Zugleich stellt das Einbeziehen der laufenden Produktion in die Weiterentwicklung des Clusters eine interessante Herausforderung an den Betrieb dar. Der Aufbau des Clusters ist hierfür in Blöcken organisiert, und wichtige Software- und Hardwarekomponenten sind redundant ausgelegt. Auf diese Weise können Teile des Clusters während der Installation neuer Software oder Hardware zeitweise außer Betrieb genommen werden, während die Produktion auf den übrigen Teilen mit geringerer Kapazität weiterläuft.

### GridKa Umfang und Erweiterung

Die Erweiterungen von GridKa verlaufen stufenweise und basieren auf Planungen der Teilchen-

physikexperimente und Fortschritten des LCG-Projekts. Nur weil jede Generation von Prozessoren und Speichermedien durch ständige Weiterentwicklung mehr Leistung bietet, sind die steigenden Anforderungen realisierbar. Die benötigte Anzahl der Prozessoren ist der zu erwartenden Vergrößerung der CPU-Leistung laut dem Mooreschen<sup>1)</sup> Gesetz laufend anzupassen. Die CPU-Anforderungen werden in den LHC-Modellen normiert auf die Leistung eines Intel Pentium-Prozessors von 1995 und mit kSI95 angegeben.

Der technologische Fortschritt in der verfügbaren Speicherdichte von Magnetbändern und Festplatten ist demjenigen der CPU-Leistung vergleichbar. Dennoch wird der größte Teil der Computerräume von GridKa im Endausbau mit Plattenspeichern und Magnetbandrobotern gefüllt sein (Abb. 5).

Im Laufe des Jahres 2004 wird das GridKa-Zentrum über das Deutsche Forschungsnetz (DFN) mit 10 Gigabit mit dem von DANTE<sup>2)</sup> betriebenen europäischen Netzwerk Géant für Wissenschaft und Bildung verbunden werden. Diese Verbindung verspricht eine Datentransfer-Kapazität von 1000 MegaByte/Sekunde. Für eine effiziente Nutzung von Netzen mit derart hohen Bandbreiten ist aber noch eine Menge Entwicklungs- und Optimierungsarbeit zu leisten. Das IWR kooperiert dazu eng mit dem DFN und CERN.

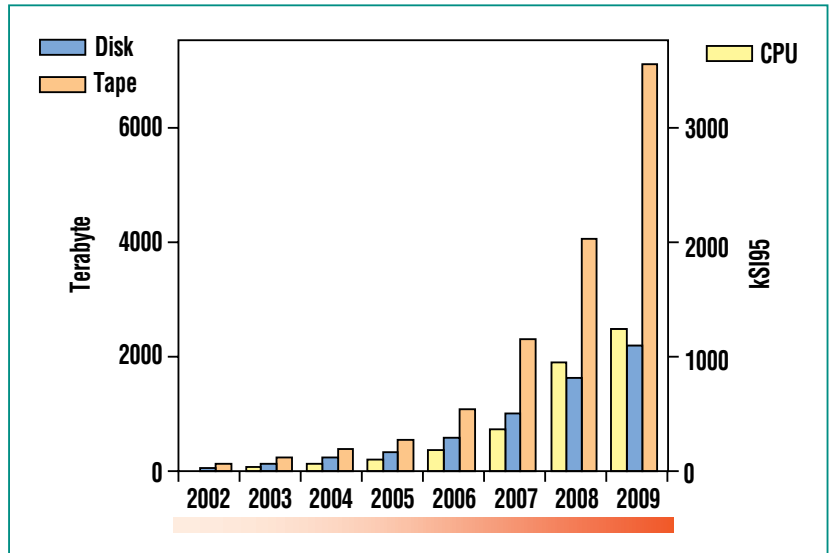


Abb. 5: Hardware-Erweiterung des GridKa-Clusters bis 2009.

## Zusammenfassung

Grid Computing verspricht die Bereitstellung verteilter Speicher- und Rechenleistung je nach Anforderung unterschiedlichster Nutzer. Die dazu benötigten Techniken werden im Rahmen des LCG-Projekts entwickelt. Davon wird letztendlich jeder Computer-Besitzer profitieren – so, wie das auch bei der Entwicklung des World Wide Web der Fall war. Die Anwendung von Grid-Werkzeugen für die Datenanalyse der Experimente am Large Hadron Collider bietet der Teilchenphysik die Möglichkeit, Rechenkapazität weltweit und daher effizienter zu nutzen. Das Cluster des Grid Computing Centre Karlsruhe stellt Rechen- und Speicherleistung für die deutsche Hochenergiephysik bereit und ist über LCG2-Grid-Software zugänglich.

## Literatur

- [1] *A Grid Computing Centre at Forschungszentrum Karlsruhe:*  
[http://lhccgrid.web.cern.ch/LHCgrid/PEB/Documents/FZK\\_RDCCG-answer-v7.doc](http://lhccgrid.web.cern.ch/LHCgrid/PEB/Documents/FZK_RDCCG-answer-v7.doc)
- [2] *MONARC –*  
<http://monarc.web.cern.ch/MONARC/docs/Phase2Report/Phase2Report.pdf>
- [3] *Web-Seite des BaBar experiment:*  
<http://www.slac.stanford.edu/BFROOT>
- [4] *Web-Seite der Experimente D0:*  
<http://www-d0.fnal.gov>  
<http://www-cdf.fnal.gov>
- [5] *Web-Seite des Compass Experiment:*  
<http://wwwcompass.cern.ch>
- [6] *Report of the LHC Computing Grid Project, Document identifier CERN-LCG-2002-16:*  
<http://lcg.web.cern.ch/LCG/Documents/>

<sup>1)</sup> Gordon Moore ist Mitgründer des Halbleiterproduzenten Intel und hat 1965 erstmals einen exponentiellen Anstieg der Transistorzahl pro integrierte Schaltung festgestellt.

<sup>2)</sup> Delivery of Advanced Network Technology to Europe, DANTE ist eine Non-profit-Organisation für Planung, Bau und Betrieb europaweiter Forschungsnetzwerke.