



Das Globus Toolkit

Praktische Ansätze für GRID
Computing Konzepte.


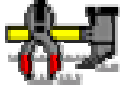






Inhaltsverzeichnis

1. Einführung
 - 1.1. Symbollegende
 - 1.2. Einordnung
2. Motivation und historische Entwicklung
3. Gliederung des Projektes
4. Globus Toolkit
 - 4.1 Grundsätze
 - 4.2 Security
 - 4.3 Resource Management
 - 4.4 Information Management
 - 4.5 Data Transfer
5. Zusammenfassung und Ausblick



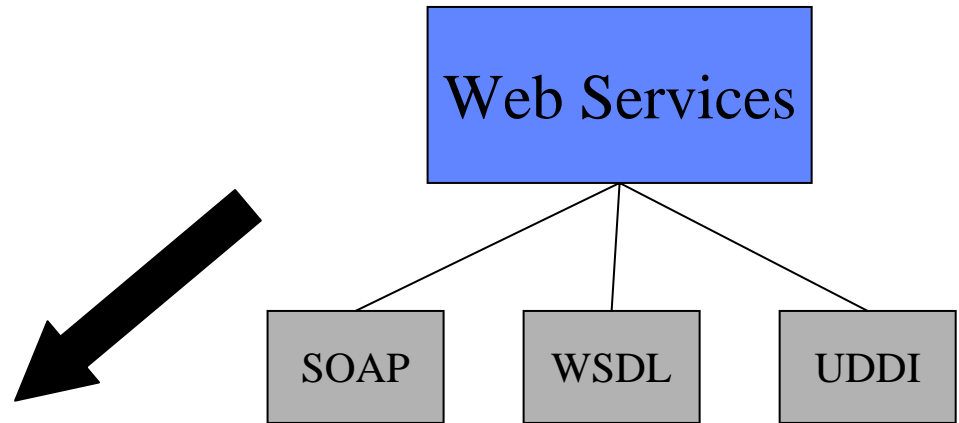
Symbollegende

- Globus Projekt 
- Globus Toolkit 
- Security 
- Resource Management 
- Information Management 
- Data Transfer 



Einordnung in den Seminarkontext

- Bisher:
 - Technische und betriebswirtschaftliche Ausführungen zum Thema Web Services.
 - Inter- Applikationskommunikation
 - Information - Orientierung



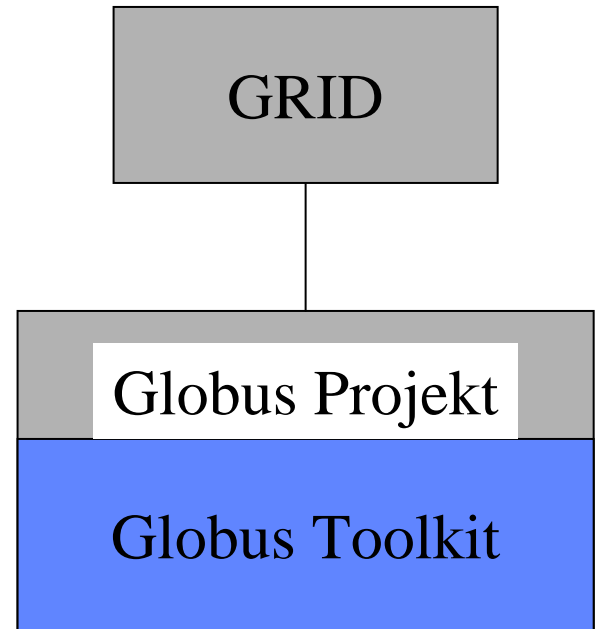
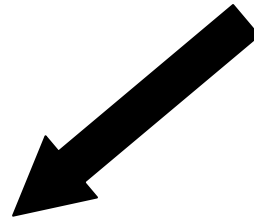
Verbindung beliebiger Dienstleistungen zu neuen Anwendungen



Einordnung in den Seminarkontext

- Heute:
 - Infrastruktur zur integrierten, kollaborativen Nutzung von (Computing) Ressourcen von verschiedenen Organisationen

**Konstruktion einer GRID
Software Infrastruktur**





Infrastruktur

- Alle institutionellen und materiellen Einrichtungen für Daseinsfürsorge und ökonomische Entwicklung [Duden]
- Alle technischen Einrichtungen zum Austausch von Informationen
- Infrastruktur ist eine Grundlage für andere Vorgänge, die, einmal etabliert, unsichtbar im Hintergrund agiert [Star, Ruhleder]

Globus Projekt / Toolkit



- Globus Projekt:
 - Einen Ansatz der Entwicklung einer GRID–Software – Infrastruktur.
 - Integrierte Sammlung von grundsätzlichen GRID Computing Dienstleistungen



- Globus Toolkit :
 - Bezeichnung für diese „Bag of Services“
 - Referenzsoftware
 - Effiziente und einfache Steuerung des GRID

Motivation 1/2

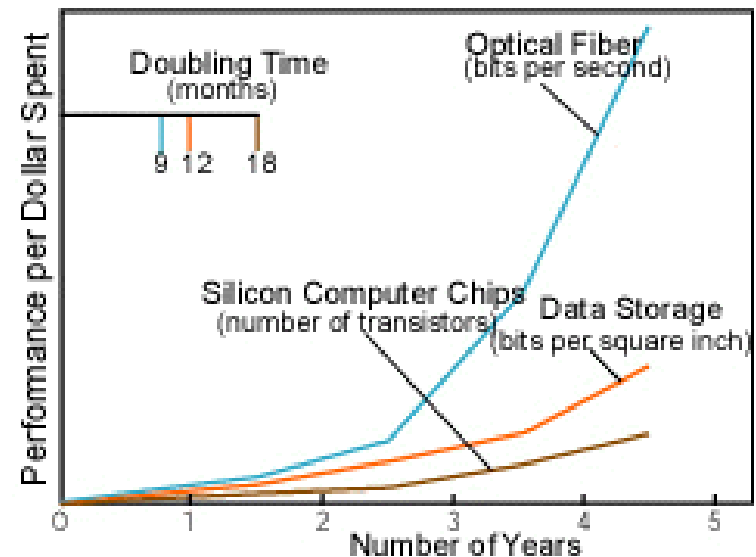
- Das Internet bietet eine universelle Netzwerkaritektur
- Unternehmenskultur hat sich hin zum Teamwork und gemeinsamer Problemlösung orientiert
- Netzwerkgeschwindigkeit wächst schneller als Prozessorgeschwindigkeit



Deshalb stärkere Konzentration auf Vernetzung als auf reine Prozessorgeschwindigkeit erforderlich

Motivation 2/2

- Netzwerkperformance verdoppelt sich alle 9 Monate
- Computerperformance verdoppelt sich alle 18 Monate



Quelle: Moore's Law vs. storage improvements vs. optical improvements.
Graphik: Scientific American (Jan-2001) by Cleo Vilett, source Vined Khoslan, Kleiner, Caufield and Perkins.

Historische Entwicklung

Bis zum Beginn des Globus Projektes gab es mehrere verschiedene Ansätze zur Umsetzung eines Grids:

- MPICH-G2: Nachrichtenaustausch
- CoG Kits, GridPort: Portal construction, based on N-tier architectures
- GDMP, Data Grid Tools, SRB: replica management, collection management
- Condor-G: Workflow Management
- Legion: Objektmodellierung für Grid Programmierung
- Cactus: Framework zum Lösen von numerischen Problemen

Historische Entwicklung

- Nachteil: Alle Programmierumgebungen mussten Protokolle und Dienste vom scratch implementieren
- Die Programmierumgebungen sollten jedoch source code gemeinsam nutzen können, welche
 - die Kernfunktionalität implementiert
 - robust, konsistent und
 - open source ist.

Dies führte zum Globus Projekt und zum „Sanduhr- prinzip“

Historische Entwicklung

- Das Projekt wurde 1996 von der Argonne National Laboratory, Universität von Südkalifornien und der Universität von Chicago gegründet
- Wichtige Nutzer des aus dem Projekt Globus entstandem Toolkit sind das European DataGrid project, NASA Information Power Grid project und das NSF's National Technology Grid
- Beispiel Testbed von 1998: GUSTO (Globus Ubiquitous Supercomputing Testbed):
 - 330 Computer, 3600 Prozessoren
 - Peak Leistung von 2 TeraFlops

Globus Projekt - Gliederung

1. Intention / Erfolg
2. Team / Sponsoren
3. Aufbau
 - 3.1 Research
 - 3.2 Testbeds
 - 3.3 Applications
 - 3.4 Software Tools

Intention / Erfolg

- Durch das WWW wurde das Konzept des Zugangs und der Verfügbarkeit von Informationen revolutioniert
- Globus möchte eine ähnliche Revolution im Bereich des „Computing“ erreichen.
- Applikationen könnten dann z.B. Supercomputer, Massenspeichereinrichtungen, Sattelitenbilder – kommunikation nutzen



Globus ist unter den 100
signifikantesten technischen
Innovationen 2002



Mitarbeiter von Globus GII
Next Generation Award For
Advanced Computing
Infrastructure

Team / Sponsoren

- Team:

- [Argonne National Laboratory](#)
- [The University of Chicago](#)
- [University of Southern California Information Sciences Institute](#)
- [National Center for Supercomputing Applications at the University of Illinois Urbana-Champaign](#)

- Unterstützung:

- [DARPA](#)
- [Teile der NASA](#)
- [U.S. Department of Energy](#)
- [National Science Foundation](#)

Aufbau - Research

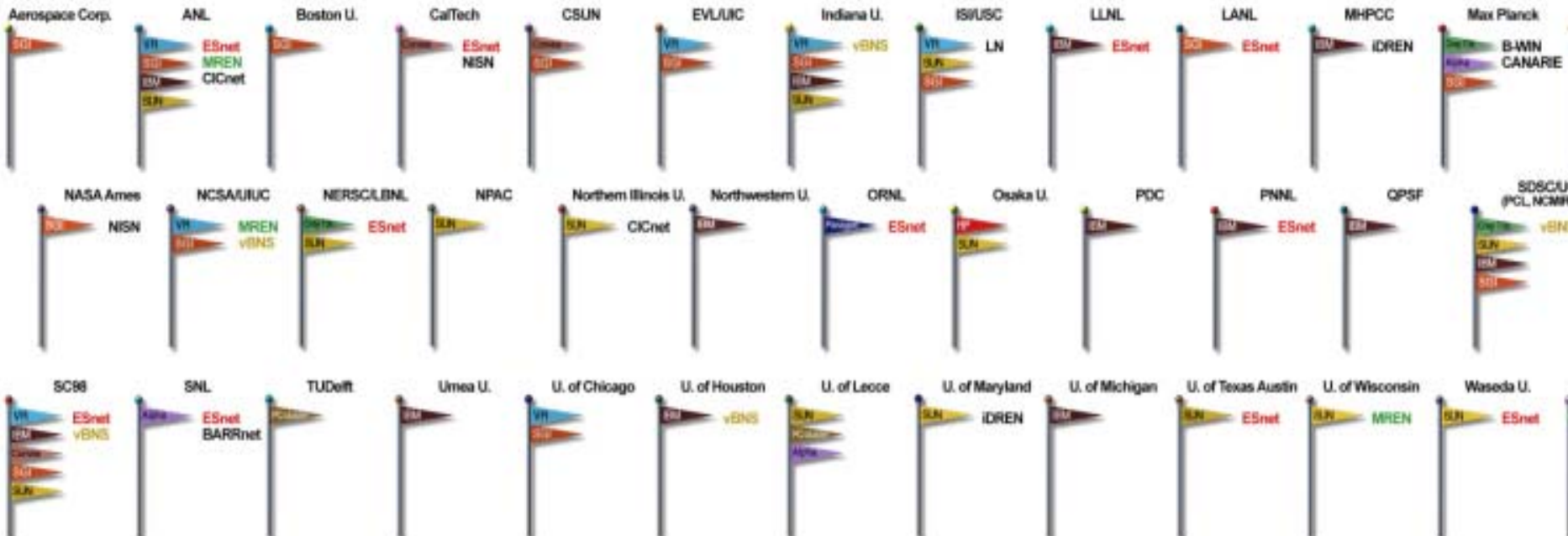
- Forschungsvorhaben mit Konzentration auf GRID Software Infrastruktur Aspekte und Problematiken in der Entwicklung von Anwendungen, die GRID Services benutzen
- Bereiche:
 1. Resource Management
 2. Application Development Enviroments
 3. Data Management and Access
 4. Information Services
 5. Security

Aufbau - Testbeds

- Unterstützung für Planung und Bau von ausgedehnten hochperformanten GRID Architekturen
- Zu Forschungszwecken und für produktiven Einsatz für z.B. ingenieurtechnischen Aufgaben
- Beispiele:
 - National Technology GRID (NPACI, NCSA)
 - GUSTO

GUSTO Computational Grid Testbed

as of November 1998

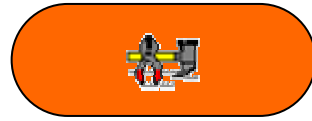


Aufbau - Applications

- Entwicklung von umfassenden gridbasierten Anwendungen für und in Kooperation mit Ingenieuren und Wissenschaftlern
- Beispiele:
 - Distributed Supercomputing (Cactus Computational Computing)
 - Smart Instruments
 - Teleimmersion (NICE)
 - Andere wie SFExpress, Neph

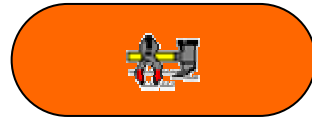
Aufbau – Software Tools

- Unterstützung des Aufbaus von Computing GRIDs
- Vereinfachung der Entwicklung von Anwendungen für die Nutzung der GRID Infrastruktur
- Globus Toolkit:
 - Open Source
 - Aktuell Version 2.2
 - In Planung 3.0 (Implementation OGSA Spezifikation)

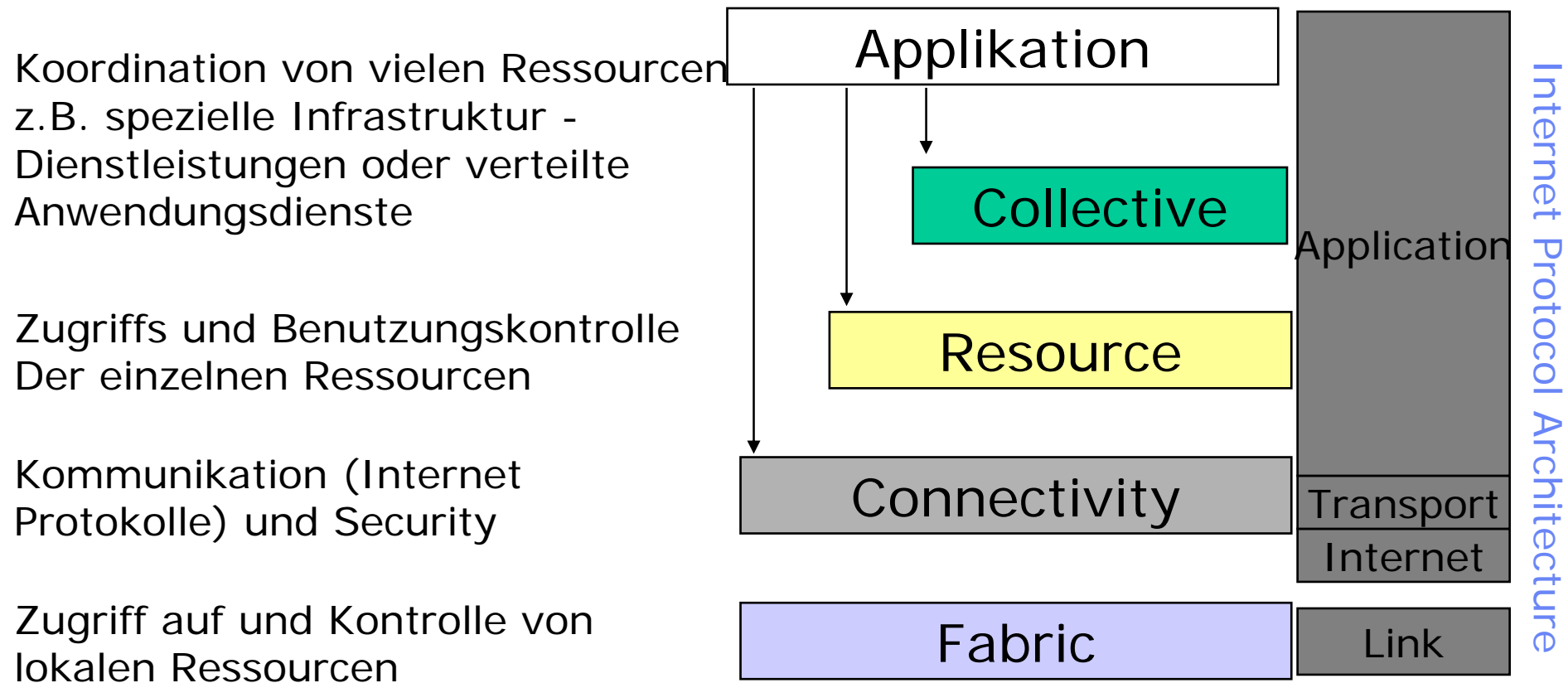


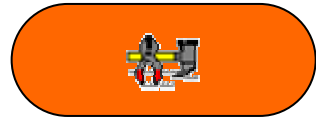
Globus Toolkit – Grundsätze 1

- Computation GRIDs sollen eine Vielzahl von Applikationen und Programmiermodellen unterstützen
- Portierung von traditionellen Anwendungen auf GRID Strukturen muss in einzelnen, klar abgegrenzten Schritten möglich sein
- Schichtmodell des Globus Toolkits vergleichbar mit Internet Protocol Suite (Key Protokolle – Connectivity Layer und Resource Layer)



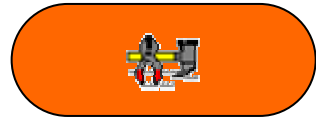
Schichten der Globus Architektur





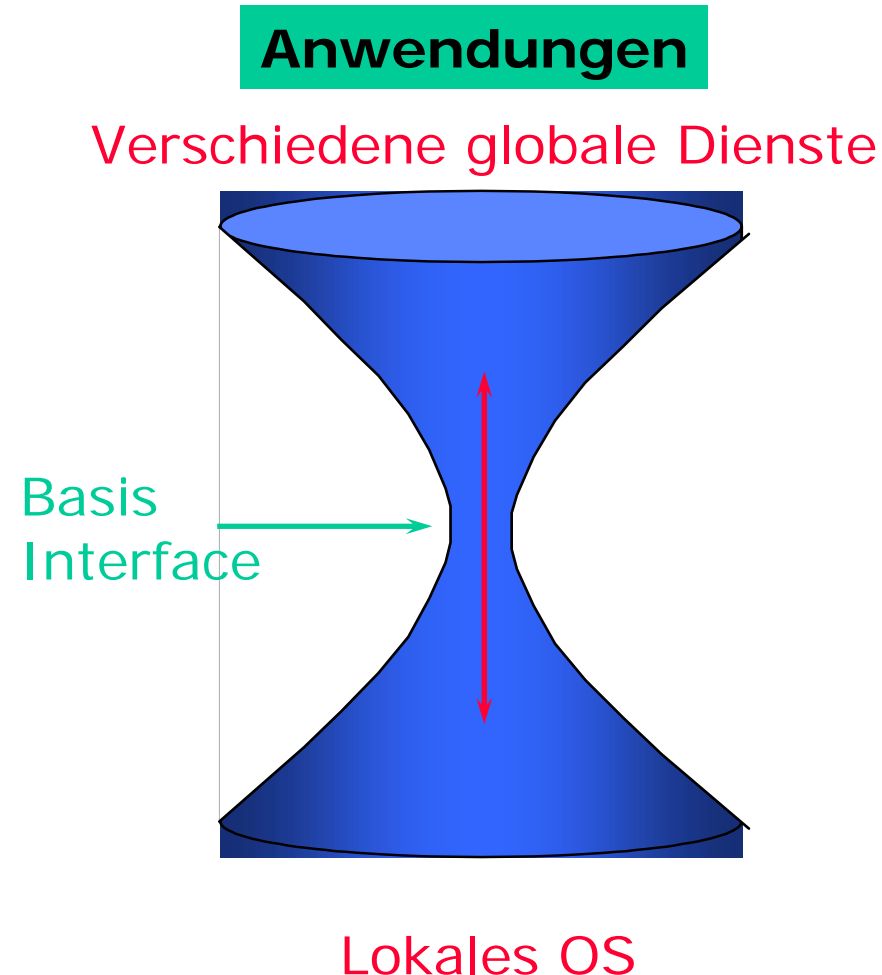
Globus Toolkit – Grundsätze 2

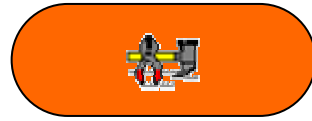
- Unterscheidung zwischen lokalen und globalen Services
- Standardisierten Zugriff auf lokal unterschiedliche Ressourcen durch Einsatz des Sanduhrprinzips
- Gewisse Transparenz und Kontrolle der untergeordneten Services wird durch transparente Schnittstellen ermöglicht
- „Information – Rich – Environment“ ist nötig, um die sich ständig ändernden Grid Umgebung, ohne feste Konfiguration, zu erfassen, sich ihr anzupassen und in ihr zu entwickeln.
- Dynamische Informationsaustausch zwischen Anwendung und Ausführungsumgebung



Sanduhrprinzip

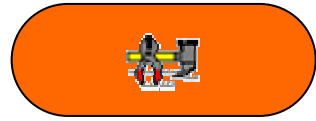
- Zentral: Einfaches Interface das einheitlichen Zugriff auf unterschiedlich implementierte lokale Dienste bietet
- Überlagernde komplexere globale Service können dann auf diesem Interface aufsetzen





Aufbau des Toolkits-Protokolle

- The Globus Toolkit™ zentriert sich auf vier Kern Protokollen:
 - Connectivity layer:
 - *Security*: Grid Security Infrastructure (GSI)
 - Resource layer:
 - *Resource Management*: Grid Resource Allocation Management (GRAM)
 - *Information Services*: Grid Resource Information Protocol (GRIP)
 - *Data Transfer*: Grid File Transfer Protocol (GridFTP)



Aufbau des Toolkits – API

Drei verschieden Arten von APIs:

- Portability and convenience API/SDKs
- API/SDKs , die die vier Hauptprotokolle implementieren
- Collective layer API/SDKs

Bsp.:

Zu 1) globus_common - kümmert sich um Modul De-/Aktivierung, die Threads, Konditionen und Treiber für die verschiedenen Ereignisse

Zu 2) globus_io - unterstützt gängige Kommunikationsprotokolle, beinhaltet auch den Sicherheitsaspekt

Zu 3) API für Broker- Systeme, die auf die Ressourcen zugreifen, oder Index- Server, die die Verwaltung der Ressourcen übernehmen und Informationen über die Ressourcen Allokation beinhalten



Security

- Sicherheitsproblematik
- Sicherheitskonzept des Globus Toolkits
- GSI
- Ausblick



Problematik der Sicherheit

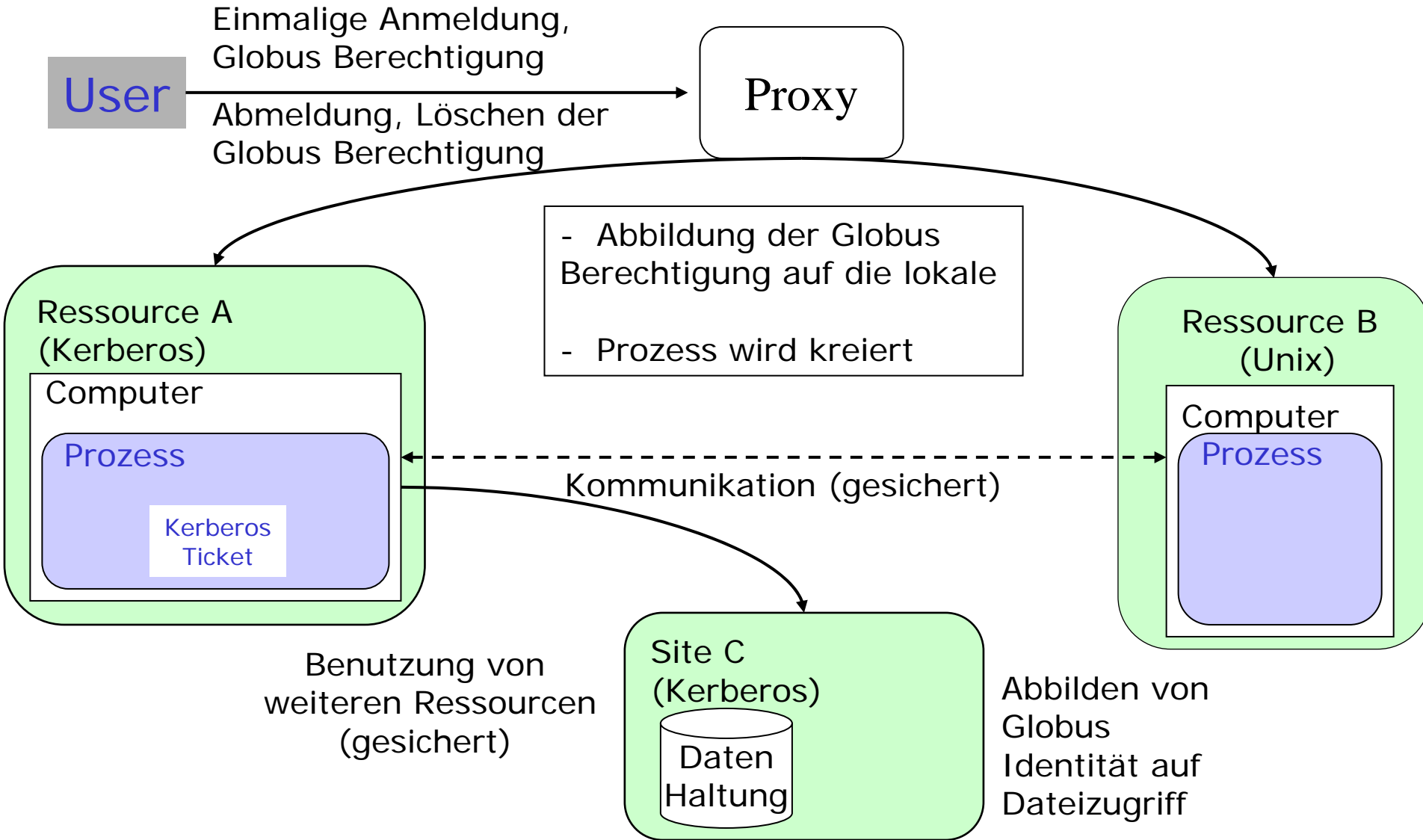
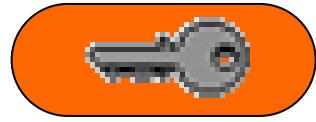
- Kommunikation: „Jeder mit Jedem“
- Sensible Daten bei beiden Kommunikationspartnern
- N-Way Sicherheitskonzept
- Viele GRID Teilnehmer haben unterschiedliche lokale Sicherheitskonzepte (Keberos, Unix, Windows)
- Lokale Sicherheitssysteme können Eigendynamik besitzen und komplex aufgebaut sein
- Sicherheitsmechanismen müssen standardisiert, „sicher“ und gut integrierbar sein

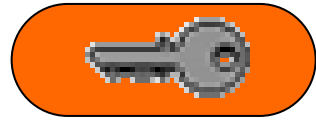


Sicherheitskonzept Globus Toolkit

- Wichtige Grundlage: Security Policy (GSI - Protokoll)
 - Bedeutung von Sicherheit für das System
 - Welche Komponenten sollen gesichert werden
 - Wogegen soll gesichert werden
- Entwicklung einer einheitlichen Schnittstelle (entsprechend der GT SW Infrastruktur)
- Schnittstelle muss logische Einheit bilden, die in ihrer internen Struktur dynamische Prozessbildung und die entsprechende Prozesskommunikation enthält
- Einheitliche Anmeldung und Bewältigung des Konflikts der lokalen Heterogenität durch Abbilden der „Globus Identität“ auf das jeweilige lokale System
- Sicherheitsalgorithmen und –prozesse werden nach dem GSS (Generic Security Services) Standard entwickelt

GSI





Security - Ausblick

- Unterstützung des Delegationsprinzips in der internen Struktur von Ressourcen
- Benutzerkontenverwaltung
- Policy Sprachen
- Gruppenrichtlinien (CAS Community Authorization Service)



Resource Management

Das Resource Management muss einen sicheren kontrollierten Zugriff auf entfernte heterogene Computerressourcen gewährleisten und diese auch managen, dabei

- Identifizierung und Authorisation
- Ressourcen entdecken und charakterisieren
- Ressourcen reservieren und zuteilen
- Berechnung beobachten und kontrollieren

Dies wird durch das GRAM Protokoll, Resource Brokern und Zuteilern, für Sicherheit das GSI und für Entdeckung das MDS erreicht.

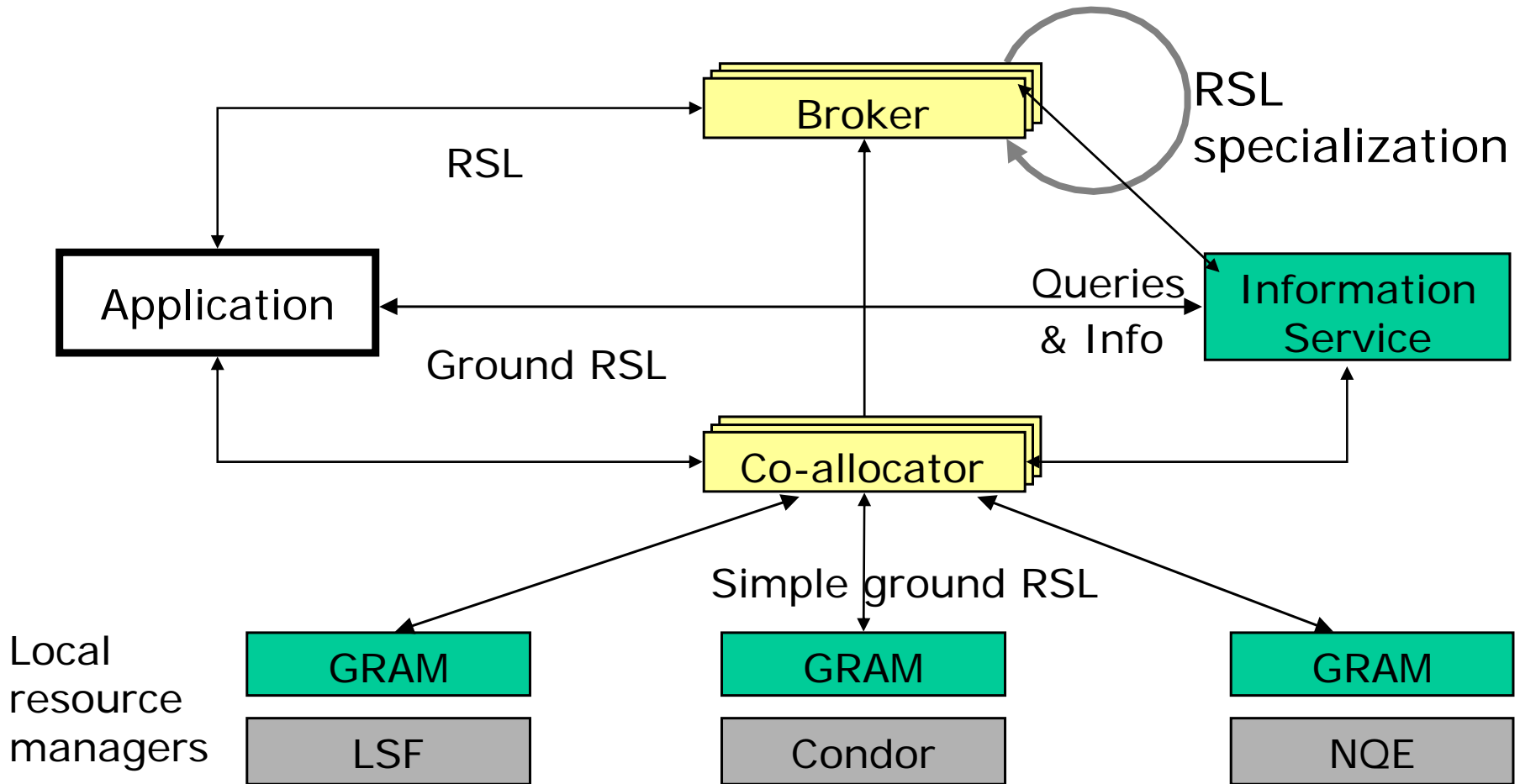


Resource Management

- Das Grid Resource Allocation Management (GRAM) Protokoll und die Client API erlaubt es Programme auf entfernten Rechnern trotz lokaler Heterogenität zu starten.
- Zur Kommunikation über die Anforderungen an Ressourcen wird die Resource Specification Language (RSL) benutzt.
- Die Schichtenarchitektur ermöglicht eine Definition unter Beachtung von GRAM von Applikations-spezifischen Resource Broker und Co-Zuteiler



Resource Management





Information Management

System Informationen sind für Operationen und Applikationen eine kritische Resource,

- welche Ressourcen stehen zur Verfügung
- wie ist der Status des Grids
- wie kann man die Nutzung optimieren

Dabei sind folgende Informationen nützlich:

- Betriebssystem, IP Adresse, Bandbreite, Latenz etc.



Information Management

Das Toolkit benutzt dabei zwei verschiedenen Informations Server:

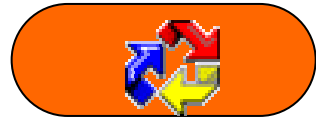
- Resource Description Services
 - liefert Informationen über bestimmte Ressourcen (GRIS)
- Aggregate Description Services
 - liefert ein Verzeichnis über Informationen, die durch GRIS gesammelt wurden (GIIS)



Information Management

Die Verwaltung wird durch das **Metacomputing Directory Service (MDS)** übernommen:

- nutzt Lightweight Directory Access Protocol (LDAP) zur Abfrage
- Informationszugriff auf ein verteiltes Verzeichnis
 - Verzeichnis durch eine Gruppe von LDAP Servern gebildet
 - Wobei jeder Server für spezielle Funktionen optimiert ist
- Verzeichnis kann upgedated werden durch:
 - Informationsprovidern und Tools
 - Applicationen (i.e., users)
 - Backend Tools, die Informationen auf Anfrage bilden
- Informationen sind dynamisch für Tools und Applikationen erreichbar

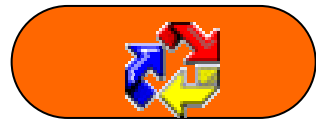


Information Management

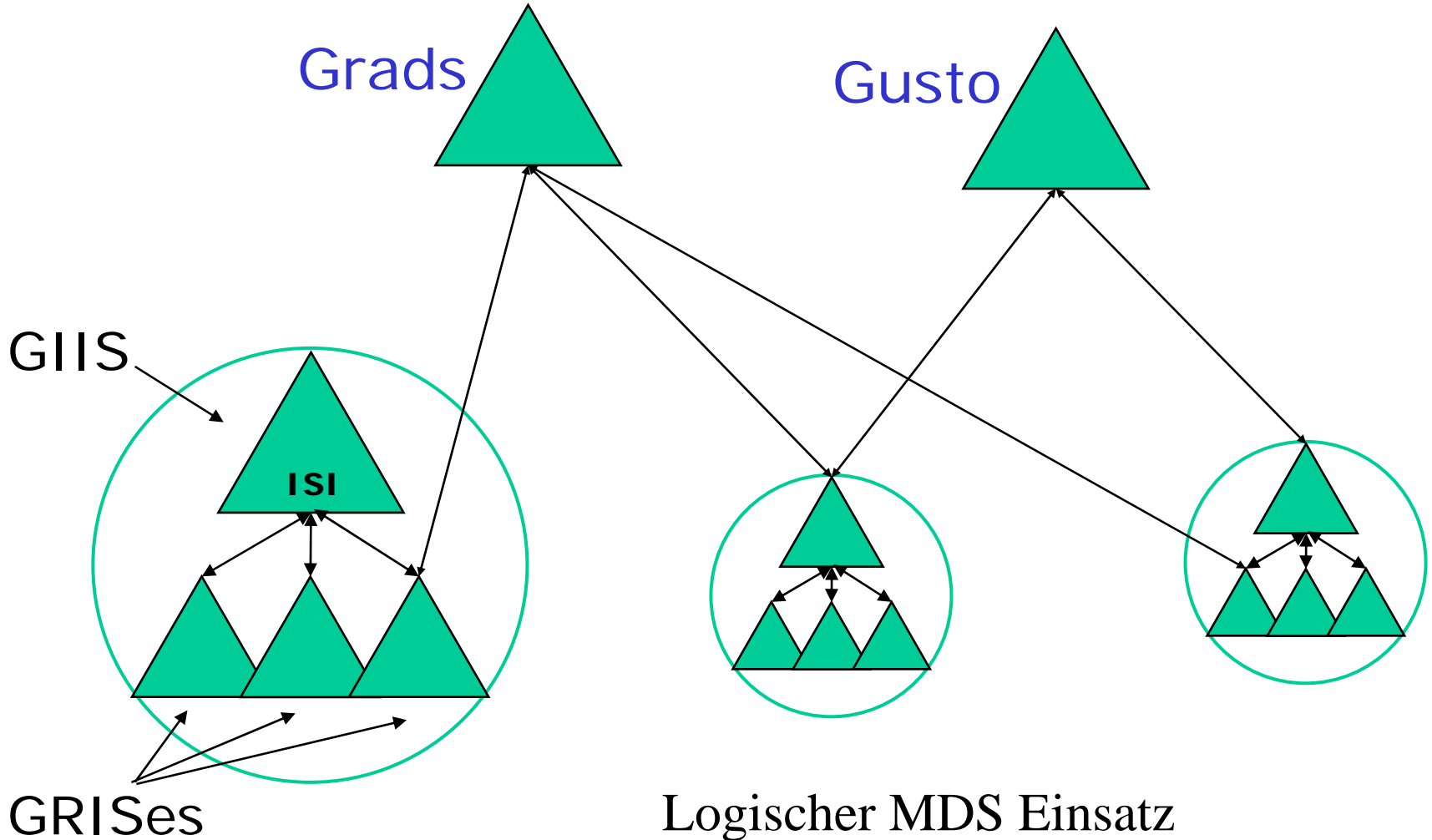
Wie schon erwähnt gibt es zwei Arten von MDS Servern:

- Grid Resource Information Service (GRIS)
 - liefert Informationen über spezifische Resource
 - konfigurierbar zur Unterstützung multipler Informationsprovidern
- Grid Index Information Service (GIIS)
 - liefert Informationen, die von mehrfachen GRIS Servern gesammelt wurde
 - ermöglicht effiziente Anfragen über Informationen, die über mehrere GRIS Server verteilt sind

Beide Dienste benutzen das LDAP zur Anfrage



Information Management





Data Transfer

Charakteristische Merkmale eines Grids in Bezug auf Daten:

- große Anzahl an Daten, Netzwerkressourcen
- Datenhaltung in unterschiedlichen administrativen Gebieten (Regeln und Gesetze beachten)

Notwendige Eigenschaften eines Daten Transfers:

- Hohe Geschwindigkeit, Zuverlässigkeit
- automatische Entdeckung der „besten“ Kopie
- Transparenz
- Sicherheit vor unbefugten Datenzugriff
- Dezentralität



Data Transfer

Das Toolkit implementiert zwei Data Grid Komponenten:

- Data Transfer and Access
 - einheitliches Protokoll nötig, das sicher, effizient, flexibel und erweiterbar ist
- Replica Management Architecture
 - mehrere Kopien von Dateien auf verschiedenen Servern



Data Transfer

Problem für Data Transfer und Access: bisherige Protokolle sind

- inkompatibel
- zum Teil proprietär, wodurch man für jedes Protokoll Clients installieren müsste
- Nicht alle nötigen Features unterstützen

Lösung: GridFTP, wobei

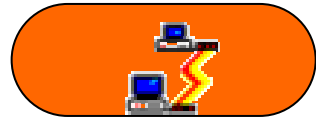
- GridFTP > FTP
- und nicht nur auf File Transfer begrenzt



Data Transfer

Warum FTP?

- viele nötigen Features unterstützt
- durch Patches erweiterbar, wodurch z.B. eine Verknüpfung mit GSI möglich ist
- standardisiert
- Verständnis für Funktionsweise sehr gross und guter Support



Data Transfer

Das Replica Management verwaltet die Verknüpfung zwischen logischen Namen für Dateien und Verzeichnisse und den (mehreren) physischen Standorte(n)

Dies ist für Applikationen wichtig, die ein Grid über viele verteilte nationale und regionale bilden, wobei zentral alle Dateien für die Applikation verwaltet werden und diese Daten für die Bearbeitung geteilt und Kopien an die einzelnen Subinstanzen verteilt werden.



Data Transfer

Dafür wird ein „Replica Catalog“ angelegt mit einer zuverlässigen Kopienbildung:

- als Schicht auf anderen Grid Diensten: GSI, Transport etc.
- benutzt LDAP als Katalog-Format und Protokoll aus Gründen der Konsistenz

Vorteil:

- diese Dienste sind für eine grosse Vielfalt an Situationen nutzbar



Zusammenfassung

- Inkrementelle Entwicklung / Portierung basierend auf dem Schichtensystem
- Einfache einheitliche Schnittstellen
- Benutzung von bekannten Techniken
- Erfolgreicher Einsatz in der Praxis
- Globus Toolkit bildet Referenz also Quasistandard
- Noch wenige Applikationen, die die Potentiale nutzen
- Kein breiter Einsatz des Toolkits zur Zeit möglich



Ausblick

- Computer werden nicht mehr gekauft sondern Ressourcen werden gemietet, geleast
- Computerbegriff ändert sich zur dynamischen, kollaborativen Ansammlung von Speicher, Prozessoren, etc.
- Netzwerkzentrierung



Quellen

- <http://www.globus.org>
- <http://www.bode.cs.tum.edu/~gerndt/home/Teaching/WS2001/GridComputing/Proseminar.htm>
- <http://www.heise.de/ix/artikel/2002/09/121/>
- Foster, Kesselmann „The Grid Blueprint for a new Computing Infrastructure“, Morgen Kaufmann, 1998