



## Grid Data Computing: "Entwurf und Realisierung eines erweiterten Verzeichisdienstes für das Grid Data Computing"

Diplomarbeit von  
Christian Dreyer

Universität Hamburg  
Fachbereich Informatik  
Arbeitsbereich VSIS

## Vorgehen

- Einführung / Grundlagen des Grid Computing  
Grid Data Computing, OGSA-DAI Entwurf und Architektur
- Vorstellung der Service Group Registry
- Entwurf und Realisierung des erweiterten Verzeichisdienstes
- Quellen
- *Anschließend: Fragen und Diskussion*

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

2

## Grid Computing

- **Definition**  
Eine Infrastruktur, die integrierte und kollaborative Nutzung von Computer, Netzwerken und Datenbanken, die von verschiedenen Personen und Organisationen betrieben werden, transparent ermöglicht.
- **Entstehung / Historie**
  - In den frühen 90er Jahren
  - Kopplung von Hochleistungsrechnern -> Grand-Challenge-Applikationen (rechenintensive Probleme)
  - Metacomputing oder Hypercomputing
  - Nachfrage nach leistungsfähigeren Ressourcen weiterhin Motivation
  - Begriff Grid -> Aus dem englischen Stromnetz (power grid)
  - Rechenleistung aus der Steckdose (MULTICS 50er Jahre)

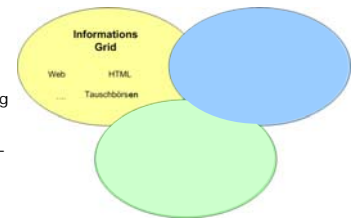
July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

3

## Arten von Grids

- **Informations Grids**
  - Entspricht dem WWW
  - Hyperlinkkonzept
  - Veränderung der Informationsverarbeitung
- **Ressourcen Grids**
  - Stellt Rechen-, Speicher- und Netzwerkleistung bereit
- **Service Grids**
  - Abbildung von Daten- und Rechendiensten auf eine abstrakte Ebene



July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

4

## Eigenschaften von Grid-Systemen

- **Lokale Autonomie**
  - Betrieb der Ressourcen geschieht lokal
  - Grid Software hat keinen Einfluss
- **Heterogenität der Ressourcen**
  - Rechnerarchitekturen, Softwaresysteme, Datenspeicher, Netzwerke, Spezialgeräte, menschliches Expertenwissen
- **Skalierbarkeit**
  - Von wenigen lokalen Knoten bis weltweit Millionen von Knoten
  - Mit Anforderungen wachsen
- **Dynamik und Adaptivität**
  - Flexible Reaktion auf Fehlerfälle

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

5

## Virtuelle Organisationen (VO)

- Gruppe von Personen / Unternehmen
- Gemeinsame Nutzung von Ressourcen
- Beliebige räumliche Trennung
- Gemeinsames oder ähnliches Ziel
- Z.B. Wissenschaft = Rechenleistung zur Datenanalyse
- Grid-Systeme häufig für spezielle VO angepasst → *keine „Standardsoftware“*

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

6

## Anwendungsszenario

- Einschränkung des Anwendungskontextes
- Schaffung von übersichtlichen Beispielen
- Erstellung einer VO
- Museumsinformationssystem
  - Mehrere Museen
  - Onlinekatalog der Kunstwerke und Ausstellungen
  - Interne Kataloge der Objekte
  - Anfragen durch Benutzer über Web-Oberfläche

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

7

## OGSA

- Open Grid Services Architecture (OGSA)
  - Erweiterbare Architektur für **Dienste** im Grid
  - Standardisiert durch das Global Grid Forum (**GGF**) -> s.u.
  - Definition von Methoden und Eigenschaften der Grid-Dienste
  - Entspricht einem **Komponentenmodell**
    - einfach Kopplung und Inanspruchnahme der Dienste
  - Generische Schnittstellen
    - Implementation ist austauschbar.

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

8

## Grid-Dienste

- Grid Service = Web Service
  - Erweitert um Routineoperationen
  - Definiert in Open Grid Services Infrastructure (OGSI)
  - Ebenfalls durch GGF standardisiert
- Open Grid Services Infrastructure (OGSI)
  - Beschreibt:
    - > Schnittstellendefinition (maschinenlesbar)
    - > Erweiterbarkeit von Diensten
    - > Lebenszeit von Diensten
    - > Referenzierung von Diensten
    - > Einheitliche Fehlerbehandlung
    - > Erzeugung von Diensten
    - > Uvm.

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

9

## Grid-Dienste (II)

- Schnittstellen
  - Definition über Web Service Description Language (WSDL)
  - WSDL 1.1 → keine Erweiterung von portTypes möglich
  - Einführung eines erweiterten portTypes (GWSDL)
  - Zur Aufnahme in WSDL 2.0 Standard eingereicht
- Lebenszeit-Management
  - Explizites und implizites Lebenszeit-Management
  - Implizit:
    - > 2 Zeitpunkte -> AFTER und BEFORE
    - > GMT, erweitert um den Wert INFINITY
  - Explizit: Spezielle Methode (destroy)

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

10

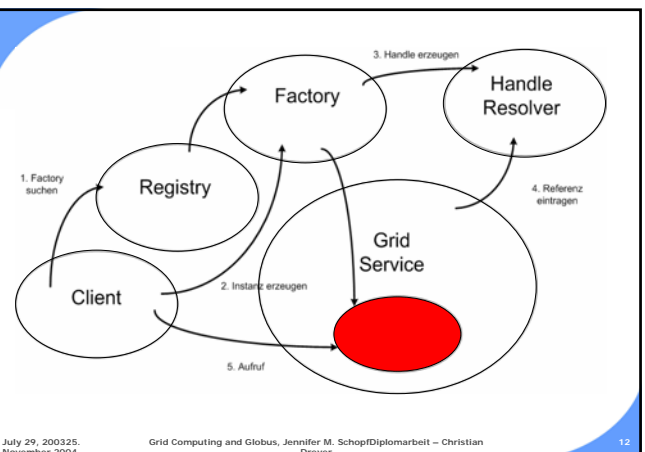
## Dienst-Referenzierung

- Zweischichtiges Referenzierungssystem
- Grid Service Handle (GSH) = abstrakter Verweis auf Dienst, unabhängig von konkreter Instanz und Implementation
- **GSH → HandleResolver → GSR**
- Grid Service Reference (GSR) = verweist auf Implementation, abhängig von Aufrufmechanismus
- Z. B. WSDL-Beschreibung bei SOAP-Dienst

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

11



July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

12

## Global Grid Forum (GGF)

- Mehrere Tausend Mitglieder
- Führt Standardisierungsbemühungen im Bereich des Grid Computing an
  - OGSA, OGSF ...
- Bereitstellung von „best practices“
- 3x jährliche globale Treffen
- Unterstützung der Grid-Benutzer

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

13

## Globus Toolkit

- Erstellt von der Globus Alliance
- Referenzimplementation der OGSI
- Java-basiert
- Implementation einiger Standard-PortTypes
- „Grid Service Hosting Container“
  - Web Service Hosting Container (Tomcat + Apache AXIS)
- Grundlage diverser Grid-Systeme
- Im Folgenden eingesetzt

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

14

## Grid Data Computing

- Datenbank-Management-Systeme nehmen zentrale Rolle ein:
  - Datenspeicherung
  - Datenorganisation
  - Datenzugriff
- Bisher nicht in Grid-Systeme integriert
- GGF → Neuer Forschungsbereich „Data“  
Verwendung von großen Datenmengen in Grid-Umgebungen
- DAIS-WG (Data Access and Integration Services)
  - Zentrale Arbeitsgruppe bezüglich der Fragestellung

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

15

## OGSA-DAI

- Open Grid Services Architecture – Data Access and Integration → OGSA-DAI
- Projekt durch **UK Database Task Force** gegründet
- Enge Zusammenarbeit mit GGF (DAIS-WG)
- Standardisierung der Schnittstellen
- Kooperation „Forschung und Wirtschaft“:
  - Oracle, IBM...
  - Epc, National e-Science Centre, ESNW, neresc.....

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

16

## OGSA-DAI (II)

- Erweiterung der OGSA-Spezifikation
- Datenquellen (z.B. relationale DB) sollen in OGSA-Rahmenwerk integriert werden
- Verteilte, heterogene DBMS wie eine einzige lokale Instanz
- Diverse Projekte und Einsatzgebiete...
- Software ist frei verfügbar / Open Source (nicht GPL, sondern eigene Lizenz)

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

17

## OGSA-DAI (III)

- AstroGrid
  - Infrastruktur von *virtuellen Sternwarten*
  - Netzwerk auf Basis von Web Services
  - Datenextrakte werden als OGSA-DAI-Dienste zur Verfügung gestellt.
- eDiaMoND
  - Grid-System für die Röntgenmammographie
  - Mammogramme = Large Binary Objects (LOBs)
  - In unterschiedlichen Datenbanken abgelegt
  - Über OGSA-DAI-Dienste zugänglich gemacht
    - Abstraktion von der einzelnen Datenquelle

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

18

## OGSA-DAI-Architektur

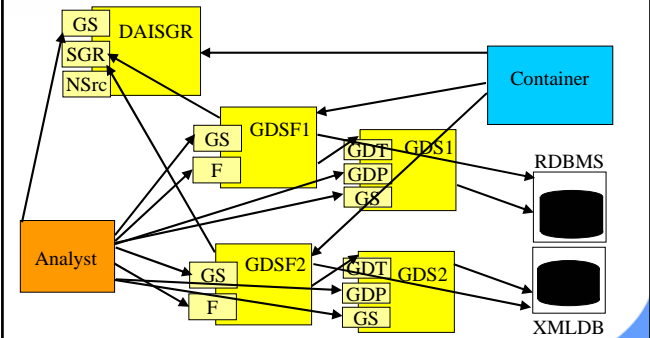
- Anzahl kooperierender Grid Dienste
- Agieren als Proxies für Datenbanksysteme
- Diese enthalten die eigentlichen Daten
- Unterschiedliche DBMS möglich:
  - Relationale DBMS (z.B. MySQL)
  - XML Datenbanken (z.B. Apache Xindice)
- Zugriff auf DB über spezielle Grid-Dienste

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. Schopf/Diplomarbeit – Christian Dreyer

19

## OGSA-DAI-Ablauf



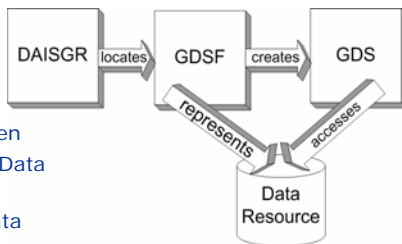
July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. Schopf/Diplomarbeit – Christian Dreyer

20

## Zusammenfassung

- Erzeugung von Registries und Factories
- Factories über Registries suchen
- Erzeugung der Data Services
- Anfragen an Data Services stellen
- Ergebnisse erhalten



July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. Schopf/Diplomarbeit – Christian Dreyer

21

## Service Grid Factory

- Erzeugung von Grid Data Services
- Aus dem objektorientierten Entwurf übernommen (siehe Gamma et.al. 1996)
- Persistente Dienste
- Konfiguration über Textdateien
- Implementation des OGSA-DAI portTypes **GridDataServiceFactoryPortType**
- Erweitert folgende OGS1 portTypes:
  - GridService portType
  - Factory portType



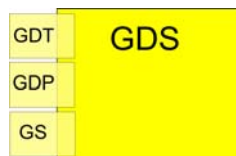
July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. Schopf/Diplomarbeit – Christian Dreyer

22

## Grid Data Services

- Basis aller OGSA-DAI Dienste
- Flüchtiger Grid-Dienst
- Ermöglichen Zugriff auf Datenquellen über ein dokumentorientiertes Verfahren
- Konfiguration über erzeugende Factory
- OGSA-DAI portType: **GDSportType**
- Erweiterungen: GridService, GridDataPerform, GridDataTransport



July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. Schopf/Diplomarbeit – Christian Dreyer

23

## GDS-Perform-Dokument

- Werden an GDS gesendet
- Einfache XML Dokumente (XML Schema Spec.)
- Enthalten Anfrage- und Änderungsoperationen
- **GridDataService::Perform** Operation als Parameter übergeben
- Enthalten Activities
  - Z.B. sqlQueryActivity
  - sqlUpdateActivity
  - xPathStatement...
- Auswertung durch spezielle Engine
- Ergebnis als GDS-Response-Dokument

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. Schopf/Diplomarbeit – Christian Dreyer

24

## Perform-Dokument – Beispiel

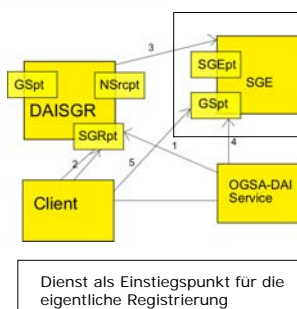
```
<sqlQueryStatement name="statement">
  <!-- value of first parameter -->
  <sqlParameter position="1" from="idvalue"/>
  <!-- value of second parameter -->
  <sqlParameter position="2">
    321
  </sqlParameter>
  <expression>
    select * from littleblackbook where
    id <= ? and id >= ?
  </expression>
  <webRowSetStream name=" statementResponse "/>
</sqlQueryStatement>
```

## OGSA-DAI Service Group Registry

- Verzeichnisdienst für OGSA-DAI Dienste
  - DAIServiceGroupRegistration
- PortType:
  - GridService
  - ServiceGroup
  - NotificationSource
- OGSI portTypes:
  - GridService
  - ServiceGroup
  - NotificationSource
- Beliebige OGSA-DAI-Dienste können sich registrieren
- Registry kann nach Diensten durchsucht werden

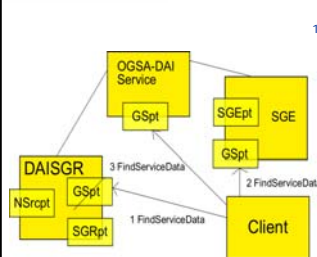


## Service Registration / Publication



- An- und Abmeldung an Registry über Methoden:
  - ServiceGroupRegistration::Add
  - ServiceGroupRegistration::Remove
- Unterschiedliche Anmelde-Möglichkeiten
  - Autonom (1)
  - Durch Benutzer oder anderen Dienst (2)
- Abmeldung:
  - Vom Dienst selbst initiiert (1)
  - Durch Benutzer (2)
  - Benutzer (5) oder Dienst (4)
  - GridService::Destroy des SGE
  - Durch die Registry selbst

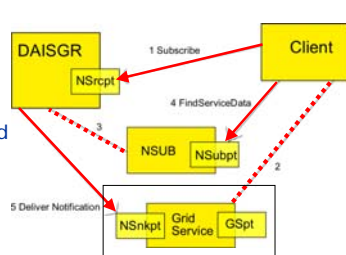
## Service Discovery



1. Befragung der DAISGR über die GridService::FindServiceData Operation
- Dienst kann direkt aufgerufen werden oder Registrierung des Dienstes verwalten

## Notification-Mechanismus

- Erlaubt Nachrichtenaustausch über spezielle Ereignisse
- Z. B. wenn ein neuer Dienst registriert wird



## Schwachstellen

- Zentrale Schwachstelle: Kommunikationsverhalten der Registry
- Eingeschränkte Schnittstellen und Funktionalitäten
  - Geringer Informationsgehalt über Datenquellen
  - Wenig ausgeprägte Anfrage- und Such-Möglichkeiten
  - Registry als Flaschenhals (Bottleneck)

## Erweiterungen

- Austausch zwischen Registries
- Extraktion von Metadaten aus den Datenquellen
- Erweiterung der Anfragemöglichkeit
- Abgleich zwischen Anfragen und Datenquellen-Metadaten

## Austausch zwischen Registries

- Grundlagen:
  - Es existiert mehr als eine Registry
  - Die Registries enthalten unterschiedliche Daten-Dienste (keine gleichen Kopien)
  - Z.B. verteilte VO mit verteilten Registries
- Einführung eines Austauschs:
  - Bei Anfrage größere Anzahl von Datenquellen als Ergebnis
  - Besser bewertete Datenquellen
  - Weniger fehleranfällig (ggf. redundante Registrierung von Datenquellen an unterschiedlichen Registries)

## Austausch zwischen Registries (II)

- Möglichkeiten:
  - Registries informieren sich gegenseitig über Änderungen
  - Anmeldung der Datenquellen an zahlreichen Registries
  - Weiterleitung der Anfrage an bekannte Registries

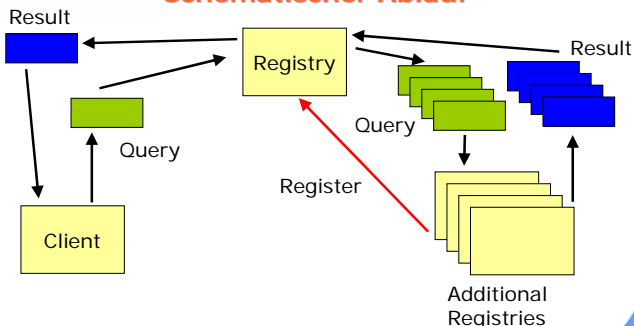
Erhöhtes Nachrichtenvolumen  
Welle von Aktualisierungsnachrichten  
Kritisch bei häufigen Änderungen

Anmeldevorgang verbraucht zu viele Ressourcen (aufwendig)  
Factory müsste alle Registries kennen, an denen sie registriert wurde

## Austausch zwischen Registries (III)

- Anfrage wird an bekannte Registries weitergeleitet
- Bildung einer Registry-Kette oder Registry-Hierarchie
- Ergebnisse (GSHs von Factories) werden zusammengeführt und als einheitliche Ergebnis geliefert (doppelte Vorkommen vermeiden).
- Weitergeleitete Anfrage = Basisanfrage
- Vorteile:
  - Hohe Modularität (Registries wie gewohnt benutzbar)
  - Kommunikation mit Klienten und Datenquellen bleibt bestehen
  - Änderungen auf Registry beschränkt

## Schematischer Ablauf



## Registrierung von Registries

- OGSA-DAI Registries erweitern OGSA portType **ServiceGroupRegistration**
- Alle Grid-Service-Arten können registriert und verwaltet werden (nicht nur OGSA-DAI Data Services)
- ServiceGroupRegistration::Add, -::Remove Operationen
- Auffinden durch ServiceGroupRegistration::FindServiceData (Suchparameter: portType = DAIServiceGroupRegistration)
  - Alle GSHs von OGSA-DAI-Verzeichnisdiensten

## Weiterleitung der Anfrage

1. Ermittlung der bekannten Registries (SGR::FindServiceData → DAISGR portType)
2. Liste von Registry-GSHs → bereinigen (Identität entfernen)
3. Erhaltene Anfrage an dieser Registries weiterleiten
  - Zähler (aktuelle Position in der Kette)
  - Maximalwert (maximale Anfragetiefe)
  - Liste abgearbeiteter Registries (Vermeidung von Mehrfachanfragen / Zyklen)

Abhängig von  
aktuellem  
Grid-System  
(Schätzwert)

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

37

## Weiterleitung der Anfrage (II)

- Besonderer Vorteil:  
Nicht erweiterte Registries können ebenfalls genutzt werden. Hier endet lediglich die Anfragekette.
- Kombination der Ergebnislisten (eigene Resultate und Ergebnisse weiterer Registries)
- Aufbereitung der **Ergebnisse** -> mehrfache Vorkommen löschen (oder Grad der Bekanntheit bestimmen ?)

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

38

## Realisierung

# WSDL

```
<operation name=" findChainedServiceData ">
  <input message="ogsdai: RequestChainingMessage "/>
  <output message="ogsi: FindServiceDataOutputMessage "/>
  ...
</operation>
<message name=" RequestChainingMessage ">
  <part name="parameters" element="ogsi: findServiceData "/>
  <part name="counter" type="xsd:integer"/>
  <part name="maxValue" type="xsd:integer"/>
  <part name=" completedDAISGR " type=" ogstdaitypes :GSHList"/>
</message>
<complexType name="GSHList">
  <sequence>
    <element name="DAISGRgsh"
      type="string"
      minOccurs="0"
      maxOccurs="unbounded"/>
  </sequence>
</complexType>
```

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

39

## Realisierung (II)

- Klassenerzeugung mit WSDL2Java
  - Behandlung der neuen Parameter
  - Ggf. Anfrage weiterleiten
  - Ergebnisse mit eigenen abgleichen
  - Resultat zurückgeben
- Relative simple Listen und Koordinationsoperationen

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

40

## Weiterleitung der Anfrage (III)

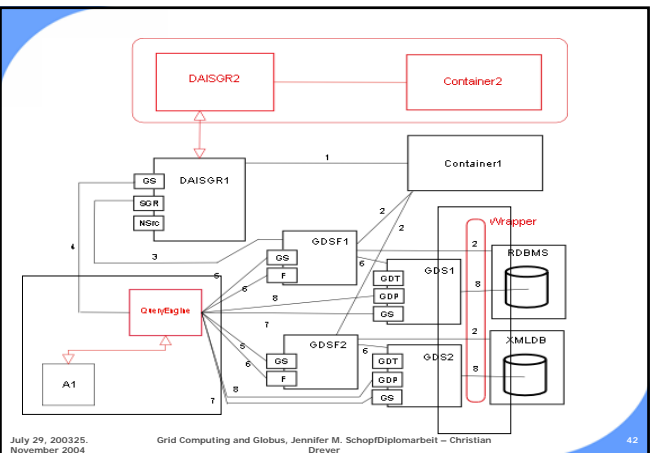
- Nicht weiter diskutiert:
  - Bewertung anderer Registries auf Basis der gelieferten Resultate
  - Caching der Ergebnisse anderer Registries

**Exkurs: Weiter  
Erweiterungen →**

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

41



July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

42

## Query Engine

- Einheitliche Schnittstelle für den Benutzer
- Nimmt Anfragen von diesem entgegen
  - Definiert einheitliche Anfrageformate
  - Angepasst an jeweiligen Anwendungskontext
- Übernimmt die Ablaufsteuerung:
  - Fragt die Registry in geeigneter Weise an
  - Stellt Anfragen an erhaltene Datenquellen
  - Bereitet die Ergebnisse auf
  - Stellt diese dem Benutzer dar
- Für Klienten einziger bekannter Dienst

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

43

## Database Wrapper

- Kapselung unterschiedlicher DBMS
- Erweitern die Schnittstelle auf der Backend-Seite → Vereinheitlichung
- Ermöglichen homogene Kommunikation
- Einsatz diverser DBMS (z.B. relationale Datenbanken [MySQL] oder XML-Datenbanken [Apache Xindice]) wird ermöglicht
- Fallunterscheidungen innerhalb des Quellcodes werden vermieden

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

44

## Extraktion von Meta-Daten

- Unterscheidung relationale DB und XML DB
  - Rel. DB: Tabellenstruktur (Tables)
  - XML-DB: Baumstruktur (Collections)
- Behandlung der unterschiedlichen Typen über die vorgestellten DB-Wrapper
- Benötigte Meta-Daten:
  - Relationale Datenbanken: Datenbankname, Tabellename, Spaltenname, Attribut + falls vorhandenen jegliche Art von Kommentar
  - XML Datenbanken: Collectionname, Dokumentname, Elementname, Elementtyp ... (Untertypen) ...

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

45

## Extraktion von Meta-Daten (II)

- Übergang zu „Meta-Kategorien“
  - Database → Datenbankname
  - Entity → Tabellename
  - Element → Collectionname
  - Type → Dokumentname
  - Attribut → Element
  - Attributtyp → Elementtyp
  - Unterelement → Unterelementtyp
- Hierarchische Anordnung
- XML Schema erstellen
- → Format für Meta-Daten

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

46

## Beispiel MD-Format

```
<database name=" KunstmuseumOslo " >
  <comment>
    Datenbank des Museums-Informations-Systems des
    Kunstmuseums OSLO.
  </comment>
  <entity name="Personen">
    <comment>Tabelle aller Personen</comment>
    <elements>
      <element name="Vorname">
        <comment>Vorname der Personen</comment>
        <type>VARCHAR2</type>
        <value>Michael</value>
      </element>
      ...
    </elements/>
  </entity>
  ...
</database>
```

Beispielwert (first row)

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

47

## Realisierung

- Über sog. **MetadataExtractor** (Wird in Konfigurationsdatei der Factory hinterlegt → Name der Java-Klasse)
  - Nach Einführung der Wrapper ist dieses überflüssig
- Neuer Extractor: GeneralMetadataExtractor
- Unabhängig von DBMS (Nutzung der DB Wrapper)
- Über Methode getDatabaseMetaData → Auslesen der Metadaten

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

48



## MD-XML-Schema

```
<xsd:element name="generalMetaData" type="gds: DatabaseType"/>
<xsd:complexType name="DatabaseType">
  <xsd:sequence>
    <xsd:element name="entity" type="gds:EntityType" minOccurs="1" maxOccurs="unbounded"/>
  </xsd:sequence>
  <xsd:attribute name="name" type="xsd:string" use="required"/>
</xsd:complexType>
<xsd:complexType name="EntityType">
  <xsd:sequence>
    <xsd:element name="element" type="gds:ElementType" minOccurs="1" maxOccurs="unbounded"/>
  </xsd:sequence>
  <xsd:attribute name="name" type="xsd:string" use="required"/>
</xsd:complexType>
<xsd:complexType name="ElementType">
  <xsd:sequence>
    <xsd:element name="comment" type="xsd:string" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="type" type="xsd:string" minOccurs="1" maxOccurs="unbounded"/>
    <xsd:element name="value" type="xsd:string" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="element" type="gds:ElementType" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
  <xsd:attribute name="name" type="xsd:string" use="required"/>
</xsd:complexType>
```

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

49

## Gewichtung von Meta-Daten

- Notwendig für die Erweiterung der Anfragen (Abgleich mit der gestellten Anfrage)
- Realisierung über austauschbare Schnittstelle
- Beispielimplementierung dieser Schnittstelle:
  - Bewertung über Tiefe im XML-Baum
  - Elemente am Wurzelknoten bedeuten bessere, aber auch grobere Übereinstimmung

XML Element	Ebene	Wert
database	1	16
entity	2	8
element	3	4
type	4	3
element	≥ 5	2
type	≥ 6	0

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

50

## Erweiterung der Anfragen

- Bisher: Suche nach speziellen portTypes z.B. gridDataServiceFactory portType
- Anschließend können Dienstdaten durchsucht werden
- Zu geringer Funktionsumfang
- Metadaten wurden bereits bereitgestellt
- → Erweiterung und Beibehalten der bisherigen Funktionalität (Kompatibilität)

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

51

## Einführung eines Query-Dokumentes

- In Anlehnung an Perform-Dokumente der GDS
- Vorteile
  - Hohe Modularität
  - Optimale Erweiterbarkeit
- Enthält keine *Activities*, sondern *Queries*
- Queries → Dokument → An Registry senden
- Ergebnisse bleiben unverändert (List of GSHs)
- In Registry folgt Abgleich mit Metadaten

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

52

## Typen von Queries

1. Entwicklungsphase (durchgeführt)
  - Keyword Query
    - > Suche nach Schlüsselwörtern (gleich gewichtet)
  - Prioritised Query
    - > Suche nach Schlüsselwörtern (Gewichtung angegeben)
  - Structured Query
    - > Suche nach angebenen strukturierten Informationen
    - > Z.B. Verfeinerung eines vorigen Ergebnisses
2. Entwicklungsphase (ausstehend)
  - SQL Query
  - XML Query

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

53

## Realisierung

- DAISGR muss neuen portType DAISGRQuery erweitern
  - Dieser ist für die Handhabung der Query-Dokumente verantwortlich
  - Operation → DAISGRQuery::query
- Umsetzung: Neues GWSL-Dokument, das den neuen portType beschreibt
- Erweiterung dieses neuen portTypes in Definitionsdatei des Registry-Dienstes (Schlüsselwort EXTENDS)

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian Dreyer

54

## Realisierung (II)

```
<gwsdl:portType name=" DAISGRQuery"
  extends="ogsi: GridService">
  <operation name="query">
    <input message="tns: QueryInputMessage "/>
    <output message="tns: QueryOutputMessage "/>
    ... <!-- Fault elements... -->
  </operation>
  <!-- Service data elements... -->
</gwsdl:portType>
```

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

55

## Realisierung (III)

- Queries realisiert als XML Schema und Java-Klassen
- Einführung einer QueryExecutionEngine (QEE) in Anlehnung an die Engine der PD
- Liest Queries aus Dokumenten und koordiniert die einzelnen Aufrufe

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

56

## Abgleich: Query – Metadaten

- Erneut Einführung einer speziellen, austauschbaren Schnittstelle
- Beispiel-Implementation über **MetaDataWordAppraiser**
- Fallunterscheidung nach Anfragetyp:
  - KeywordQuery: Gleiche Vorkommen werden bestimmt, Summe deren Bewertungen wird errechnet, bei Überschreiten eines Grenzwertes wird Data Service in Ergebnismenge aufgenommen (= *Prio. 1*)
  - PrioritisedQuery: Wie KeywordQuery + Modifikator, Mod. wird mit Bewertungsergebnis multipliziert, es können sehr effizient Prioritäten gesetzt werden

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

57

## Abgleich: Query – Metadaten (II)

- StructuredQuery: Strukturabgleich, übergebene XML-Struktur muss in gleicher oder ähnlicher Form in MD-Sammlung vorkommen
- Zukünftige Ausbaustufen:
  - Weitere Query-Arten
  - Reguläre Ausdrücke bei Abgleich integrieren
  - Weitere Verbesserung über Ontologie-System
  - Uvm.

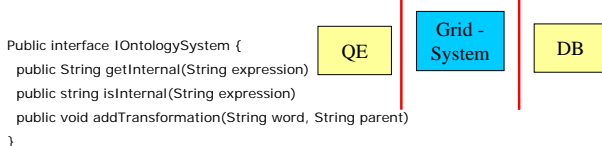
July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

58

## Fehlende Komponenten

- Ontologiesystem
  - Überführung der Begriffe (Metadaten und Anfragen) in eine interne Repräsentation
  - Erleichterung des Abgleichens zwischen Anfrage und MD
  - Durchführung an den äußeren Schnittstellen
  - Bisher: Einfache Schnittstellen und Begriffslisten



July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

59

## Zusammenfassung und Ausblick

- Grid Computing
  - Seit kurzer Zeit im Blickpunkt von Forschung
  - Bisher erreicht: Stabile und solide Basis für dienstorientierte Grid-Systeme
  - Steigende Anzahl kommerzieller Anbieter / Nutzer
- OGSA-DAI Entwurf
  - Entwicklung zu einem weit verbreitete Standard
  - Bekannte Mängel sollen bis Ende des Projektes (Sep. 2005) behoben werden
  - Aktuelle Version 4.0 stellt gut erweiterbare Grundlage dar

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

60

## Bewertung der erw. Registry

- Nahtlose Integration in bestehende Grid-Umgebungen möglich (ohne Anpassungen)
- Erzielte Vorteile hauptsächlich im Bereich der Schnittstellen (zw. den Komponenten)
  - Erweiterte Anfragemöglichkeiten
  - Zusätzliche Informationen über Datenquellen bekannt
  - Anfrageweiterleitung → quantitativ und qualitativ bessere Ergebnisse an Data-Services

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

61

## Literatur

- ALONSO, Gustavo: *Grid Computing*. In: Datenbank Spektrum 08/2004
- BERMAN, Fran; FOX, Geoffrey: *Grid Computing, Making the Global Infrastructure Reality*. John Wiley & Sons Ltd, 2003
- FOSTER, Ian; KISHIMOTO, Hiro; NICK, Jeffrey: *Open Grid Services Architecture (OGSA)*, 2003
- FOSTER, Ian; KESSELMAN: *The Grid 2: Blueprint for a new Computing Infrastructure*. 2nd Edition. Morgan Kaufmann, 2004

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

62

## Internetquellen

- <http://www.globus.org> – Globus Alliance
- <http://www.ggf.org> – Global Grid Forum
- <http://gridcafe.web.cern.ch> – Generelle Informationen über das Grid Computing
- <http://www.ogsadai.org.uk> – OGSA-DAI !!!

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

63

Ende

July 29, 200325.  
November 2004

Grid Computing and Globus, Jennifer M. SchopfDiplomarbeit – Christian  
Dreyer

64