

***Web- und Gridservices
zur Überwindung von Heterogenität***

Bearbeiter: Lei Xia

16.07.2004

Gliederung

- Einleitung
- Formen von Heterogenität
- Grundlagen
- Web Services als Schnittstelle zu DBMS
- Grid Data Services
- Zusammenfassung und Ausblick

Einleitung

- Einsatz von inkompatibler Software und Hardware
- Kommunikation durch das Internet
- Integration der Software- und Hardwaresysteme
- Standardisierung erforderlich
- Web Services und Grid Services bieten eine mögliche Lösung

Formen von Heterogenität

- Technische Aspekte
- Programmiersprachen
- Schnittstellen
- Anfragesprachen
- Datenmodelle
- Semantik
- Schemas
- Strukturen

Grundlagen

- Web Services
 - Simple Object Access Protocol (SOAP)
 - Web Services Description Language (WSDL)
 - Universal Description, Discovery and Integration (UDDI)
- Grid
 - Virtuelle Organisationen
 - Open Grid Services Infrastructure (OGSI)
 - Open Grid Services Architecture (OGSA)
- Föderierte DBMS

Web Services

- Schnittstelle zwischen Client und Dienstanbieter
- basieren auf gängige Standards wie XML, HTTP, SOAP, WSDL, UDDI
- **plattform- und programmiersprachenunabhängig**
- **verbergen die konkrete Realisierung der Services**
- Database Web Services für den Zugriff auf Datenbanken

Simple Object Access Protocol (SOAP)

- XML-basiertes Protokoll für Informationsaustausch über das Internet
- unabhängig von Transportprotokoll
- One-way-message-Exchange
- Envelope: Identifizierung als SOAP- Dokument
- Header: Kontrollinformationen
- Body: applikationsspezifische Nutzinformationen
- Fault: Informationen über Fehler bei der Nachrichtenverarbeitung

Web Services Description Language (WSDL)

- beschreibt die Schnittstellen von Web Services
- basiert auf XML
- Beschreibung beantwortet drei Fragen: Was? Wie? Wo?
- Beschreibung unabhängig von den zur Realisierung verwendeten Technologien
 - Plattform
 - Programmiersprachen

Grid

- Nutzung von verteilt vorliegenden Ressourcen
- wichtige Merkmale eines Grid-Systems
 - Autonomie
 - Heterogenität der Ressourcen
 - Fehlende zentrale Kontrolle
- Grid Services sind Web Services
- Datagrid: Daten sind Ressourcen

Virtuelle Organisationen

- Zusammenarbeit einer Menge von Individuen aus verschiedenen Institutionen
- erfordert Middleware, die mehrere Datenquellen in eine einzelne Datenquelle integriert
 - *Verbergen der Formen von Heterogenität*

Open Grid Services Infrastructure (OGSI)

- Spezifikation der Grid Services
- standardisierte Routineoperationen:
 - Verfügbarkeitszeitraum (lifetime)
 - Benachrichtigung bei Zustandsveränderung (notification)
 - Instanziierung von Services (Factory)
 - Auffindung und Referenzierung von Serviceinstanzen (FindServiceData, GSH, GSR)
 - Fehlerbehandlung
 - Verwaltung der Gruppen von Serviceinstanzen
 - Zustandsinformationen (service data)
 - Vererbung von Schnittstellen

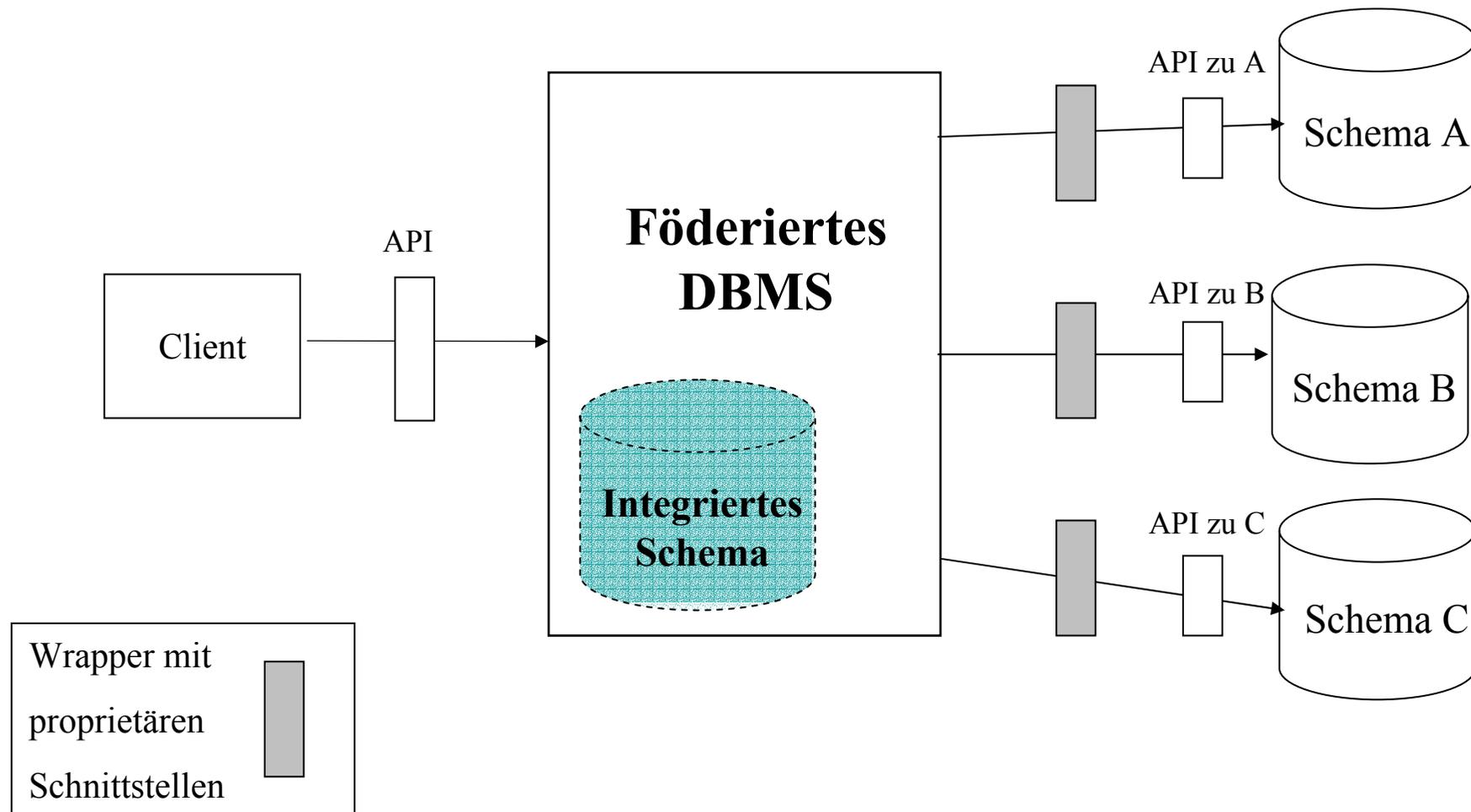
Open Grid Services Architecture (OGSA)

- basiert auf OGSF
- Definition von konkreten Schnittstellen für anwendungsorientierte Dienste
 - Execution Management Services
 - **Data Services**
 - Resource Management Services
 - Security Services
 - Self-Management Services
 - Information Services
 - Context Services

Föderierte DBMS

- Middleware zur Integration der Daten aus Datenquellen
- im Idealfall als ein einziges Datenbanksystem
- Verbergen von unterschiedlichen Plattformen, Datenmodellen, Strukturen, Anfragesprachen und Schnittstellen

Architektur eines föderierten DBMS



Web Services Object Runtime Framework (WORF)

- einfache Erzeugung von Web Services für den Zugriff auf DB2-Datenbanksystem
- DADX-Dokumente (Document Access Definition Extension) zur Spezifikation der Web Services.
- J2EE-Applikation
- unterstützte Applikationsserver:
 - WebSphere Application Server
 - Apache Jakarta Tomcat

Document Access Definition Extension (DADX)

- Beschreibung von Schnittstelle und Implementierung von Web Services für DB2 durch XML
- Beschreibung der Operationen durch SQL-Statements (inkl. Stored Procedures) und DAD-Dateien
 - Stored Procedures: Zusammenfassung mehrerer SQL-Statements
 - DAD: Abbildung zwischen XML-Daten und relationalen Daten
- Generierung der WSDL-Beschreibungen aus den DADX-Dokumenten

Zwei Arten von Operationen

- SQL-Operationen:
 - <query>: Anfrage auf den Daten in einer DB2-DB
 - <update>: Veränderung der Daten in der Datenbank (insert, delete, update)
 - <call>: Aufruf von Stored Procedures
- XML-Kollektionsoperationen:
 - <retrieveXML>: Generierung der XML-Dokumente durch Komposition der existierenden relationalen Daten
 - <storeXML> : Speichern der XML-Dokumente als relationale Daten in DB2

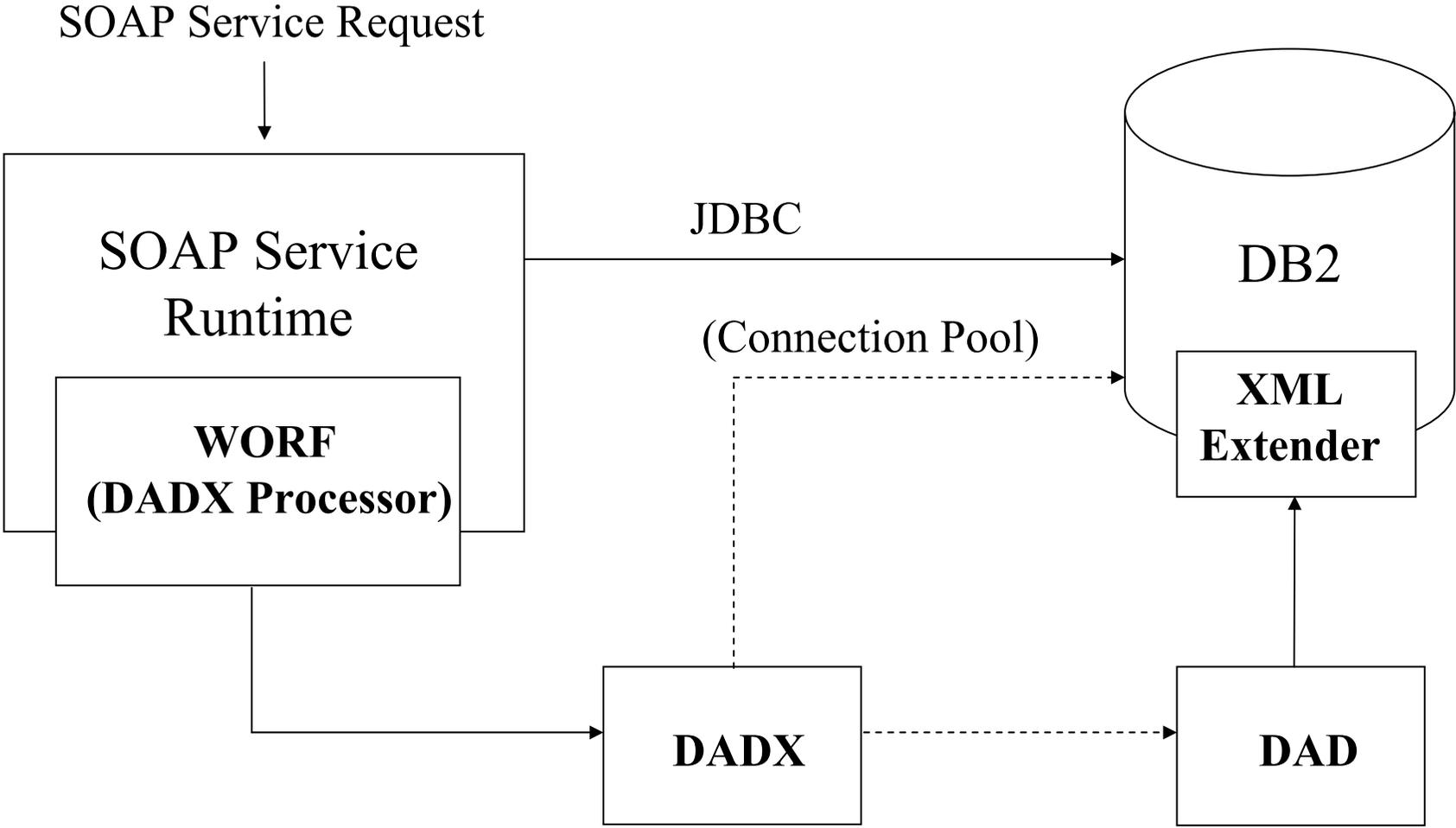
Beispiel von einer DADX-Datei

```
<?xml version="1.0" encoding="UTF-8"?>
  <DADX
    xmlns="http://schemas.ibm.com/db2/dxx/dadx"
    xmlns:wSDL="http://schemas.xmlsoap.org/wSDL/">
    operation ← <operation name="listDepartments">
      <query>
        Implementation { <SQL_query> select* from order_tab
                          where customer_name = :customer_name
                          </SQL_query>
        part ← <parameter name="customer_name" type="xsd:string"/>
      </query>
    </operation>
  </DADX>
```

DB2 XML Extender

- unterstützt die XML-Kollektionsoperationen
- ermöglicht Generierung und Speicherung von XML-Dokumenten in DB2-DB
- DAD zur Abbildung zwischen XML- und relationalen Daten
- Zwei Möglichkeiten zur Speicherung:
 - Speicherung des ganzen XML-Dokuments in DB2
 - Zerlegung der XML-Dokumente in relationale Daten (Shredding)

WORF und XML Extender



Vor- und Nachteile von WOLF

- Vorteile:
 - einfache Erzeugung von Database Web Services
 - standardisierte Spezifikation der Web Services mit XML
 - automatische Generierung von WSDL-Beschreibungen
 - Kombination von XML und relationalen Daten in einer DB
 - Verbergen der Implementierung
- Nachteile:
 - abhängig von SQL und JDBC
 - nur für DB2-DB

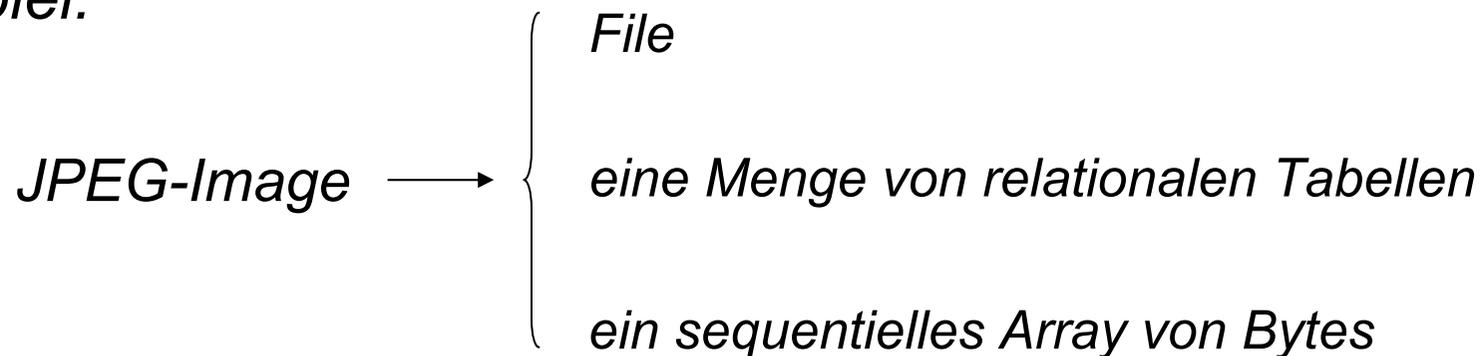
Grid Data Services

- Eine der von OGSA definierten Schnittstellen
- sind Web Services
- Verbindung zwischen Client und Datenanbietern
- Als Standardschnittstellen zu strukturierten Datenquellen
- Abstrahierung von den technischen Aspekten der Datenquellen
- Implementierung von den vier Basisschnittstellen
- WSDL zur Beschreibung der Schnittstellen

Datenvirtualisierung

- eine abstrahierte Sicht von Daten
- Ein Data Service implementiert eine Datenvirtualisierung
- mehrere Abstraktionsmöglichkeiten

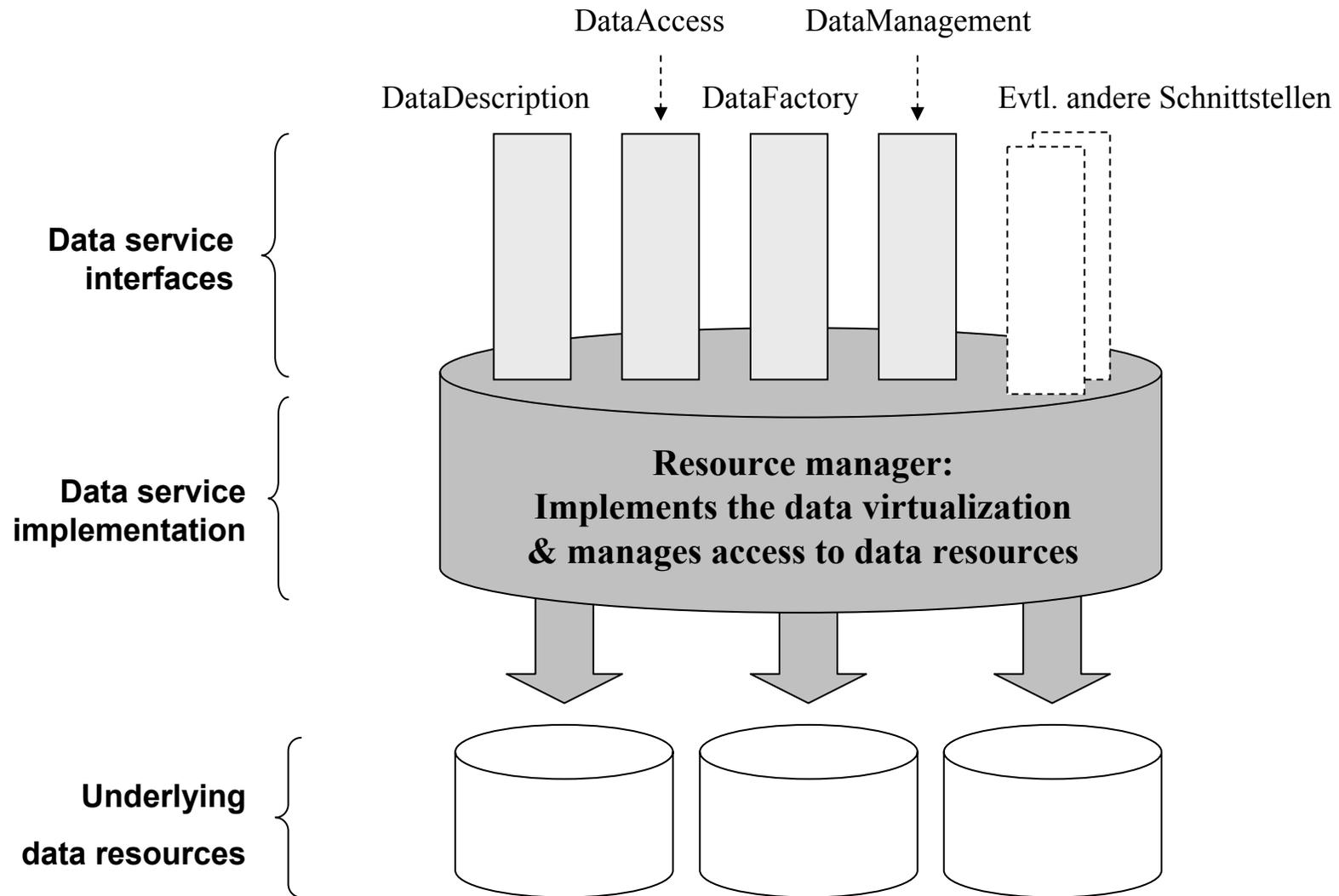
Beispiel:



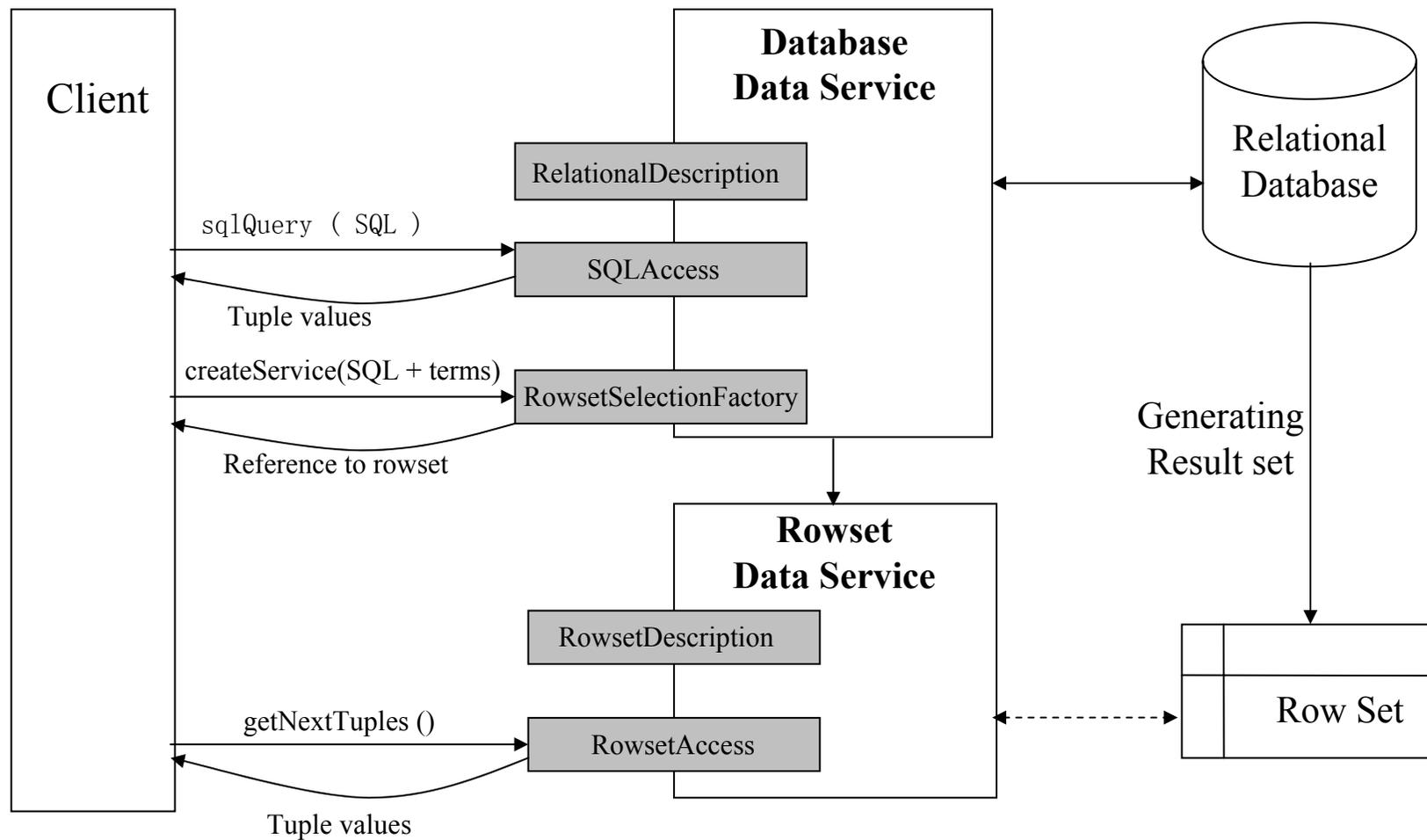
Basisschnittstellen

PortType	Beschreibung
DataDescription	Beschreibt den Inhalt und die Struktur eines Data Services.
DataAccess	Bietet Operationen für die Anfrage und Manipulation der Daten in einer Datenvirtualisierung.
DataManagement	Bietet Operationen für die Überwachung und Verwaltung der Datenvirtualisierung .
DataFactory	Bietet Operationen für die Erzeugung eines neuen Data Services mit einer Datenvirtualisierung.

Architektur von Data Services



Beispiel von Grid Data Service



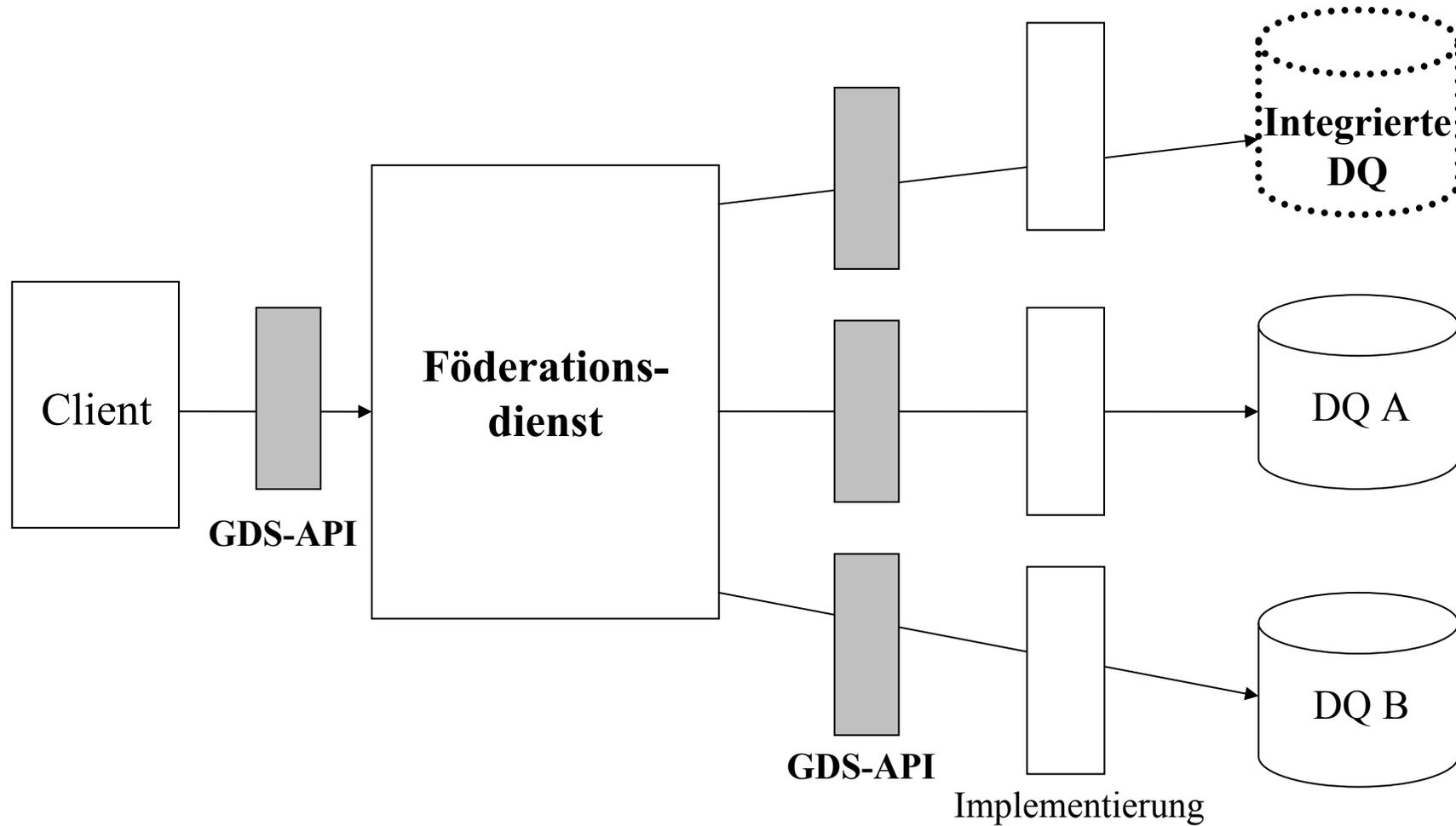
DataAccess-Schnittstelle

- Operationen für den Zugriff auf Datenvirtualisierungen
- Beschreibung der Operationen durch WSDL
- mögliche Verfeinerungen der Schnittstelle:
 - SQLAccess
 - XMLCollectionAccess
 - StreamAccess
 - BlockAccess
 - FileAccess

Grid Data Services als Grundlagen für föderierte DBMS

- hoher Aufwand für die Überwindung der Heterogenitätsformen
 - Unterschiedliche Quell-APIs
 - Nutzung proprietärer Wrapper-Schnittstelle
- Nutzung einer sprach- und systemunabhängige Standardschnittstelle (z.B. in Form von Grid Data Services)
- Überwindung der technischen und API-Heterogenität
- Unterstützung von Hierarchie der Föderationssysteme

Föderationsdienst mit Grid Data Service



Zusammenfassung und Ausblick

- Web Services sind plattform- und programmiersprachen-unabhängig
- Grid Data Services bieten standardisierte APIs
- Föderierte DBMS integrieren die heterogenen Datenquellen
- keine vollständige Lösung für alle Formen von Heterogenität
- Beschreibung der Zustandsinformationen von Web Services durch WSRF
- Entwicklung von WSDL 2.0