A Data Management System for UNICORE 6

UNICORE SUMMIT

Tobias Schlauch, German Aerospace Center

UNICORE Summit 2009, August 25th, 2009, Delft, The Netherlands

DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft

Outline

- Objectives
- Concept
- Implementation Overview
- Test Scenario AeroGrid
- Summary and Outlook

Objectives

Creation of a data management component for Grid environments

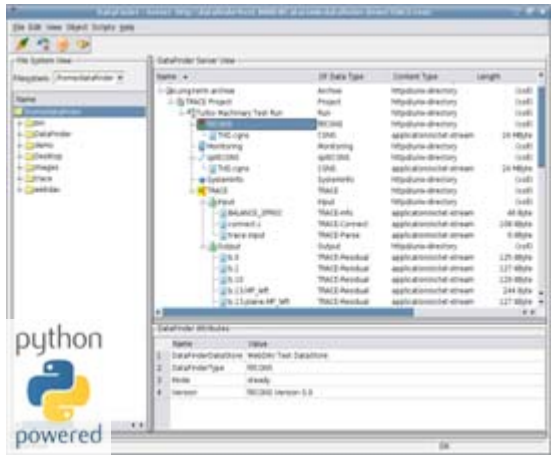
- Independence from specific Grid middleware system
- Transparent data management handling
- Focus on data structuring and metadata management

Constraints

- **Basis:** Data management software DataFinder
  - Compatibility with DataFinder
- Reference implementation on basis of UNICORE 6

DataFinder
Data Management System

- Based on ***open, stable*** standards (WebDAV)
- Specification of metadata and data structure through ***free-definable data models***
- Flexible inclusion of storage resources accessed by ***standardized data transfer interfaces***
- ***Platform-independent*** graphical user interfaces for data management and administrative tasks



Available as open source software (BSD license):

- http://sourceforge.net/projects/datafinder/

Concept
Basics

Abstraction of common data management concepts

- Introduction of:
  - Logical data object names
  - Logical storage resource names
  - Logical user names

Definition of functionalities on basis of these abstractions

- Data handling
- Privilege handling
- Data structuring and metadata handling through data models

Concept
Architecture - Principal Components

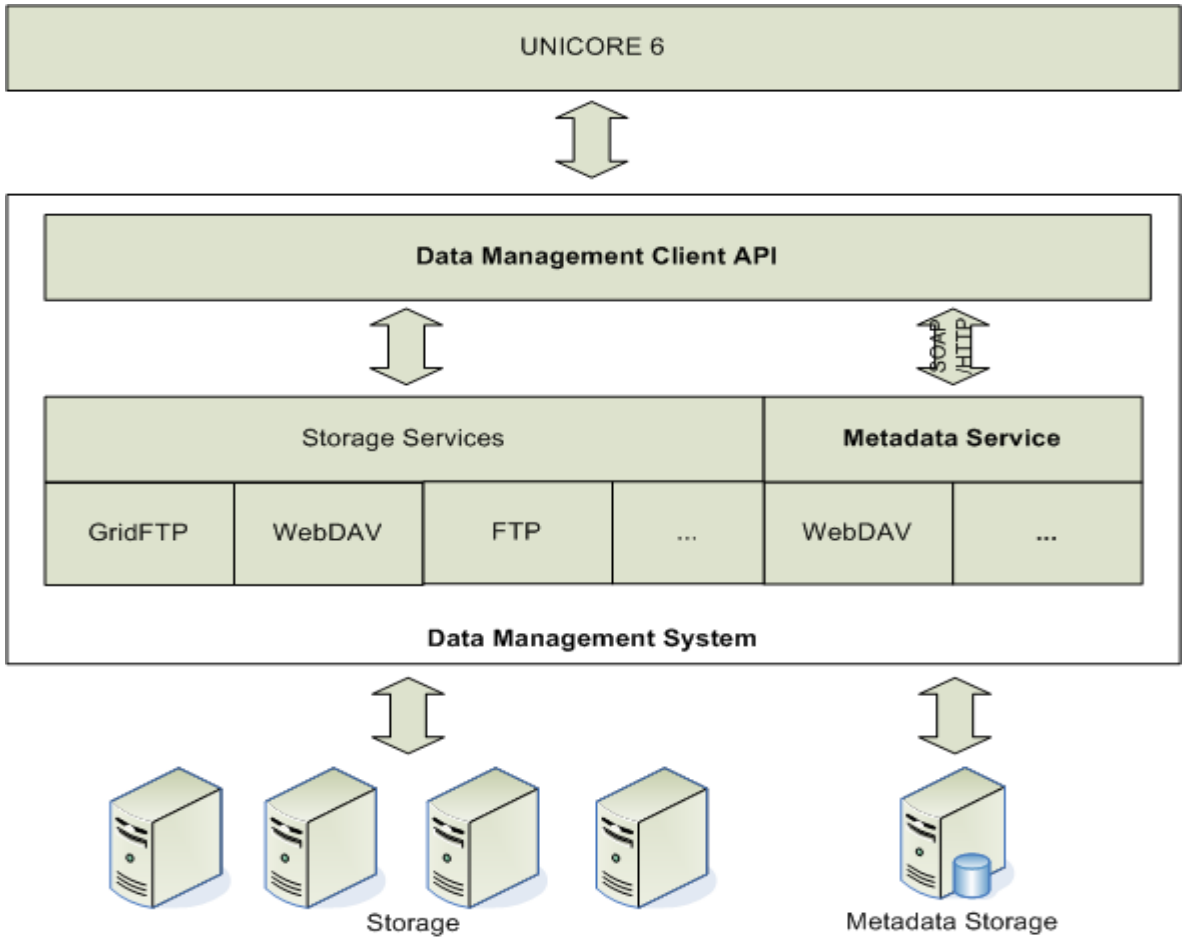
Metadata Service

- Management of logical data structure and metadata on basis of data models
- Privilege management
- Management of storage resources and location of data objects

Data Management Client API

- Structured storage of data using metadata service component
- Data transfer handling
  - Direct access of storage resources using common data transfer interfaces
- Basis of integration with Grid middleware

Concept
Architecture



Implementation Overview
Metadata Service – Basics (1)

Provision of functionalities through atomic Web Service

- Implementation based on Axis2 Web service framework
- Data structuring and metadata functionalities
- Administrative functionalities
- *Privilege handling is currently an open issue*

Development followed contract first principle

- Provision of an easy-to-use and extensible interface
- Specification of interface using WSDL 1.1
- Code generation

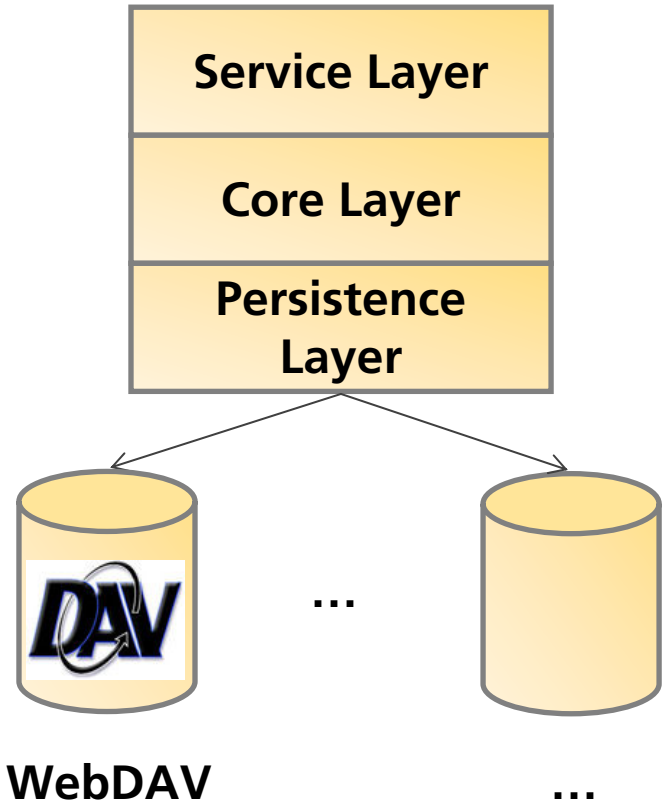
Implementation Overview
Metadata Service – Basics (2)

Usage of HTTP/HTTPs as transport protocol

- Point to point security through features of the transport protocol

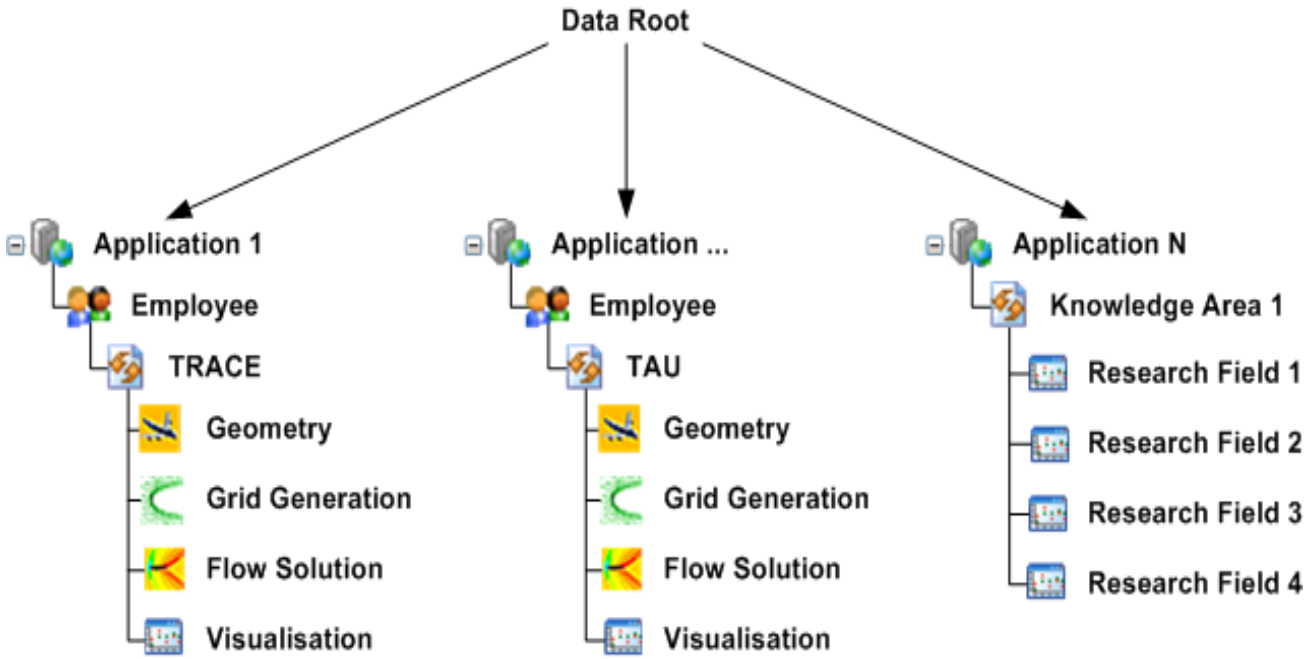
WebDAV server is used as storage backend

- WebDAV provides basic concepts simplifying implementation
- Compatibility with DataFinder
- Layered architecture ensures exchangeability of storage backend



Implementation Overview

Metadata Service – Logical Data Object Name Space



- Provision of ***integrated view*** on different data management applications

Implementation Overview
Metadata Service – Configuration

Every data management application owns its specific configuration

- Configuration consists of:
  - Data model
    - Logical data structure
    - Optional and required metadata
    - *Restrictions are controlled by the service*
  - Storage configuration
    - Data transfer interface
    - Parameter
    - *Used for location of specific data items*

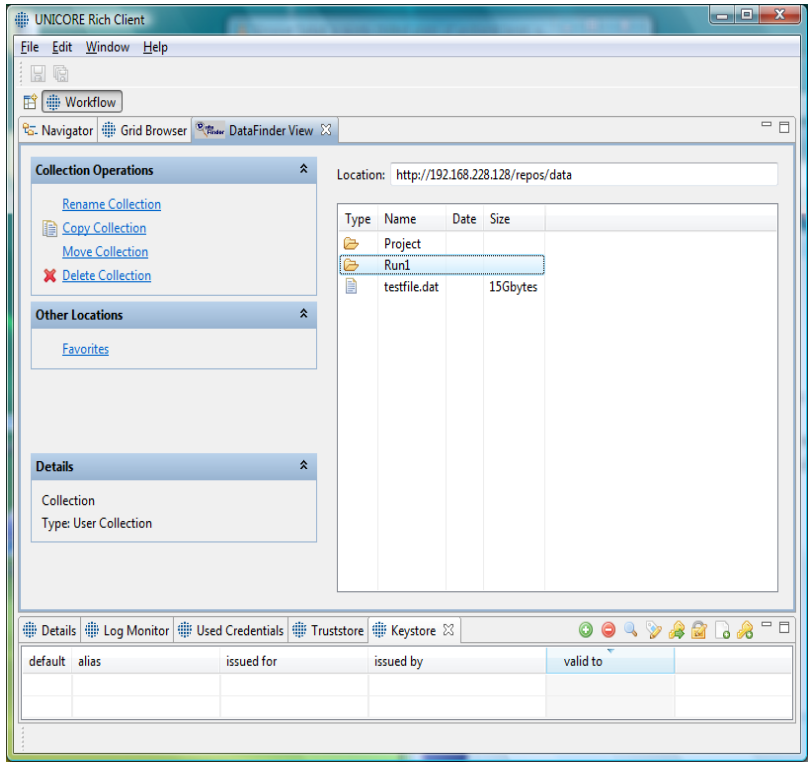
Implementation Overview
Data Management Client

Implementation as Eclipse plug-in

- *Eclipse Rich Client Platform* is widely adopted
- Easy integration with *UNICORE 6 Rich Client*
- Currently bundled with data transfer protocol WebDAV

Usable with different metadata service implementations

- Contract: Defined WSDL interface



Implementation Overview
Data Management Client - Features

General features

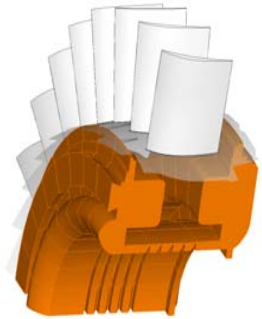
- Directory and file browsing
- Creation, modification and deletion of data
- Metadata manipulation
- Display of data models

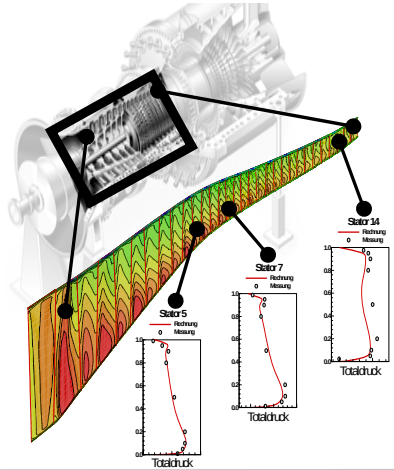
UNICORE specific features

- Selection of input files
- Selection of output files

Test Scenario AeroGrid
The AeroGrid Project

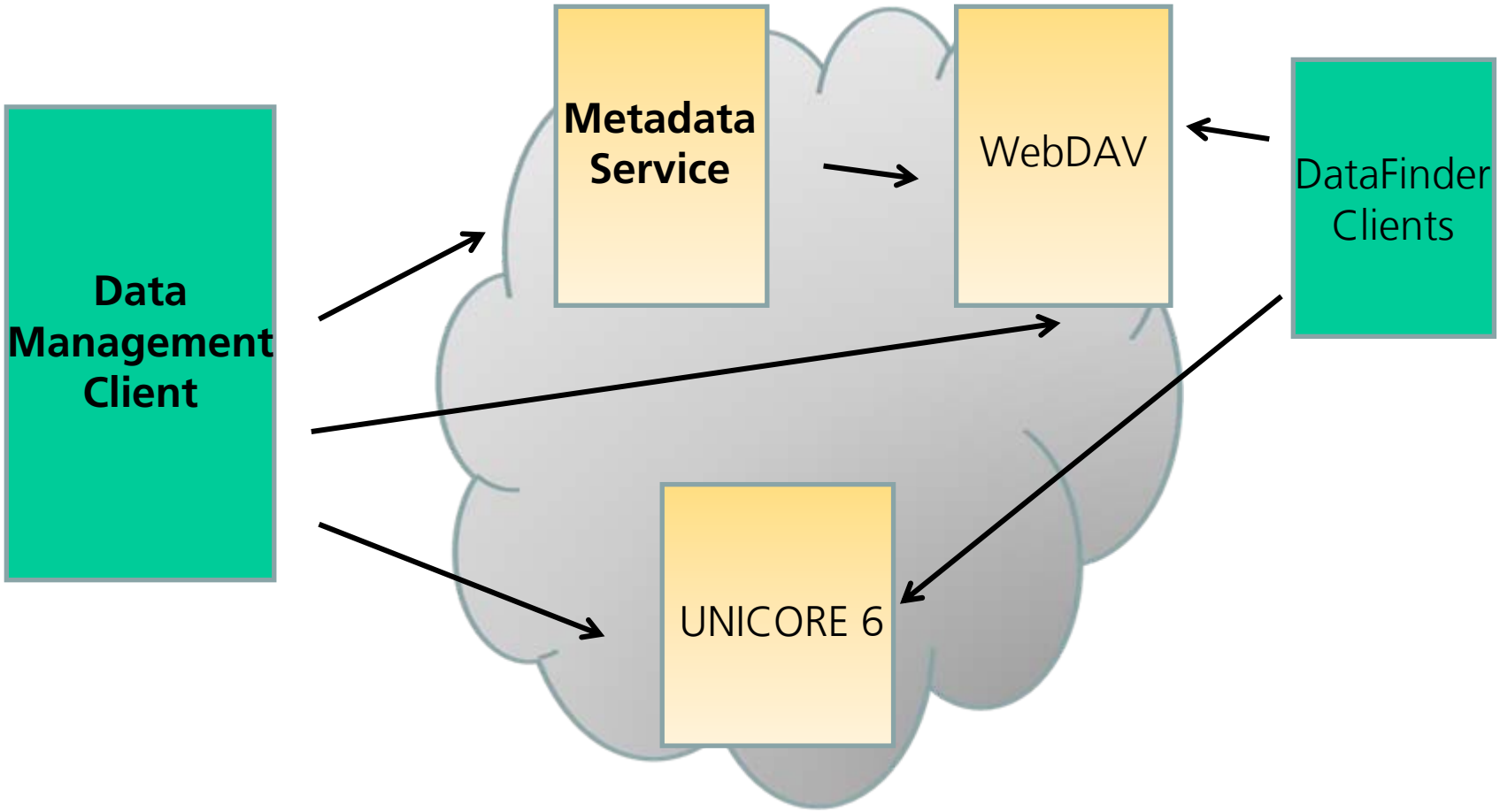
- ***Grid-based cooperation*** of industry, research centres and universities in the ***aerospace domain***
- Simulation of ***turbo machinery flows***
  - Usage of compute resource over AeroGrid
  - Cooperative design of turbo machinery components
  - Cooperative development of the CFD simulation code TRACE
- **Project Site:** http://www.aero-grid.de/
- DataFinder is used for ***simulation workflow execution*** and ***management of simulation results***





Test Scenario AeroGrid

Deployment Scenario



Test Scenario AeroGrid
Demonstrated Aspects (1)

Data management functionalities:

- Browsing of files and directories
- Deletion of files and directories
- Import and export of data files
- Display and modification of metadata

Administrative functionalities:

- Display of data models
- Display of storage configurations

Test Scenario AeroGrid
Demonstrated Aspects (2)

Compatibility with DataFinder

- Data models and storage configurations
- Data import and export

Submission of UNICORE jobs

- Usage of managed files as job input
- Export of job result files

Performance assessment

- *Still an open issue*
- Little overhead in comparison to DataFinder is expected

Summary and Outlook

Implementation of the data management component and its integration with UNICORE 6 is completed

- Metadata service
  - Data structuring and metadata management functionalities
  - Administrative functionalities
- Data management client
  - Data transfer handling using WebDAV
  - Integration with UNICORE 6

Different scenarios have been tested in scope of the AeroGrid project

- Basic functionalities and compatibility with DataFinder
- *Performance assessment is an open issue*

Summary and Outlook

Available as open source software (BSD license)

- Project Site: http://tor-2.scai.fraunhofer.de/gf/project/unicoredata

Planned extensions

- Privilege management for data objects and metadata based on ***Access Control Lists***
- Support of search queries based on metadata
- Extension of available data transfer interfaces

Project Team:

- Anastasia Eifer
- Tobias Schlauch

Deutsches Zentrum für Luft- und Raumfahrt e.V. in der Helmholtz-Gemeinschaft

- Thomas Soddemann

Fraunhofer SCAI

Supported by:

D-GRID

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

T-Systems

Thank for your attention!!!

Links

Project Site: http://tor-2.scai.fraunhofer.de/gf/project/unicoredata

DataFinder: http://sourceforge.net/projects/datafinder/

AeroGrid: http://www.aero-grid.de/

Contact

Email: Tobias.Schlauch@dlr.de