Scientific Workflows
With UNICORE 6

Bastian Demuth (b.demuth@fz-juelich.de)
Jülich Supercomputing Centre (JSC)
What is UNICORE?

- **UNiform Interface to COmputing Resources**
  - Grid middleware
  - Access to heterogeneous resources
  - Focus on high performance computing (HPC)
  - Seamless, secure, and intuitive
- Initial development started in 1997
- Sourceforge project since Summer 2004
More than a decade of European research and infrastructure projects

And many others, e.g.
Infrastructures using UNICORE

- D-Grid (German national Grid)
- DEISA (Distributed European Infrastructure for Supercomputing Applications)
- National German supercomputing center NIC
- Gauss center for supercomputing (3 German HPC centers)
- PRACE (European PetaFlop infrastructure, starting-up)
UNICORE in D-Grid

- All D-Grid resources accessible via UNICORE, GLite and Globus
- Core D-Grid sites committing existing resources to D-Grid
  - Approx. 700 CPUs
  - Approx. 1 PByte of storage
- Additional Sites received extra money for buying compute clusters and data storage
  - Approx. 2000 CPUs
  - Approx. 2 PByte of storage
UNICORE in DEISA

- Consortium of leading national HPC centers in Europe
- Operate a persistent, production quality, distributed, heterogeneous HPC environment
- UNICORE as Grid middleware
  - On top of DEISA’s core services:
    - Dedicated network
    - Shared file system
    - Common production environment at all sites
  - Used e.g. for workflow applications

IDRIS – CNRS (Paris, France), FZJ (Jülich, Germany), RZG (Garching, Germany), CINECA (Bologna, Italy), EPCC (Edinburgh, UK), CSC (Helsinki, Finland), SARA (Amsterdam, NL), HLRS (Stuttgart, Germany), BSC (Barcelona, Spain), LRZ (Munich, Germany), ECMWF (Reading, UK)
UNICORE 6 Design Principles

- Open source (BSD license, SourceForge project)
- Extensible
- Interoperable (e.g. JSDL, WS-RF 1.2, OGSA-BES)
- Strong security: X.509, proxy and VO support
- Workflow and application support tightly integrated
- Variety of clients: graphical, command-line, portal, etc.
- Quick and simple installation and configuration
- Support for many operating and batch systems (written in JAVA and Perl)
Ease of installation

tar.gz based installer is also available
... and even runs on Windows
UNICORE 6 Architecture

- **Portal clients**
  - command-line client
  - Eclipse-based client

- **Gateway**

- **Service Registry**
  - UNICORE Atomic Services
  - OGSA-*
  - XNJS + TSI
  - UNICORE hosting env.

- **UNICORE Atomic Services**
  - Local RMS (e.g. Torque, LL, LSF, etc.)

- **OGSA-***
  - Local RMS (e.g. Torque, LL, LSF, etc.)

Clients and applications
- authentication
- job execution and data storage
- parallel scientific jobs of multiple end-users on target systems

Local RMS (e.g. Torque, LL, LSF, etc.)
UNICORE 6 Workflows – Design Goals

- Ease of installation, configuration, and usage
- Performance and scalability
- Flexibility and extensibility
  - Service oriented architecture
  - Domain specific workflow languages
  - Pluggable resource brokering strategies
- Fault tolerance
- Security through trust delegation
- Reproducibility of results
- Support for existing applications without modification
UNICORE 6 Workflows – Architecture

Gateway

Portal clients
command-line client
Eclipse-based client

Workflow Engine
Tracing Service

Service Orchestrator
Resource Information Service

Service Registry

Clients and applications
authentication
workflow execution
brokering and job management
job execution and data storage
parallel scientific jobs of multiple end-users on target systems

Clients and applications

Service Orchestrator

Gateway

Workflow Engine
Tracing Service

Service Registry

Clients and applications
authentication
workflow execution
brokering and job management
job execution and data storage
parallel scientific jobs of multiple end-users on target systems

Gateway

Portal clients
command-line client
Eclipse-based client

Workflow Engine
Tracing Service

Service Orchestrator
Resource Information Service

Service Registry

Clients and applications
authentication
workflow execution
brokering and job management
job execution and data storage
parallel scientific jobs of multiple end-users on target systems

Gateway

Portal clients
command-line client
Eclipse-based client

Workflow Engine
Tracing Service

Service Orchestrator
Resource Information Service

Service Registry

Clients and applications
authentication
workflow execution
brokering and job management
job execution and data storage
parallel scientific jobs of multiple end-users on target systems

Gateway

Portal clients
command-line client
Eclipse-based client

Workflow Engine
Tracing Service

Service Orchestrator
Resource Information Service

Service Registry

Clients and applications
authentication
workflow execution
brokering and job management
job execution and data storage
parallel scientific jobs of multiple end-users on target systems
UNICORE 6 Workflows – Auxiliary Services

- Location Manager
  - Mapping: logical file names => physical locations
  - Supports wildcards
- Tracer (Provenance service)
  - Stores all messages with timestamps
  - Different use cases (state of execution, Grid usage & congestion, benchmarks, reproducibility of results)
- Resource Information service
  - Provides data for brokering
- Virtual Organisation service
  - Based on SAML
  - Supports arbitrary attributes
- Licence Management service
UNICORE 6 Workflows – Rich Client

- UNICORE Rich Client
  - Based on Eclipse => extensible
  - Graphical job preparation and workflow editing
  - Submission, monitoring, and tracing of workflows
  - Visualisation of job outcomes (application specific)
UNICORE 6 Workflows – Commandline Client

- UNICORE Commandline Client
  - Based on Apache Commons CLI and Groovy
  - Textual job preparation and workflow editing
  - Submission of workflows
  - Monitoring of workflows
  - Fetching job outcomes
UNICORE 6 Workflows – Portal Client

- Chemomentum Web portal
- Based on JSP, JavaScript and AJAX
- Parametrisation of pre-defined workflows
- Simple user interface
- Execution monitoring and inspection of outcomes
Data Management System

- Handle large volumes of data
- Annotate data with *extensible* metadata (workflow, applications, ...)
- Automatic extraction of additional metadata from files (e.g. chemical structures used, image thumbnails)
- Transparent access to external data sources (e.g. chemical databases)
- Comfortable client API and GUI client

http://www.chemomentum.org
UNICORE 6 Workflows – Outlook

- Human interaction with running workflows
  - Suspend/resume
  - Predefined breakpoints
  - Variable modification
- Automated conversion of workflows to portal pages
- Tests and optimisation of brokering strategies
- Additional workflow structures
UNICORE

join the developer community, software, source code, documentation, tutorials, mailing lists, community links, and more:

http://www.unicore.eu