A Science Gateway for Molecular Simulations

Sandra Gesing, Peter Kacsuk, Miklos Kozlovszky, Georg Birkenheuer, André Brinkmann, Dirk Blunk, Sebastian Breuers, Gregor Fels, Richard Grunzke, Sonja Herres-Pawlis, Jens Krüger, Lars Packschies, Ralph Müller-Pfefferkorn, Patrick Schäfer, Thomas Steinke, Anna Szikszay Fabri, Klaus-Dieter Warzecha, Martin Wewior, and Oliver Kohlbacher

EGI User Forum, 11 April 2011

www.mosgrid.de

Funded by the BMBF grant: 01IG09006
Outline

• Motivation
• MoSGrid (Molecular Simulation Grid)
• MoSGrid science gateway
• Domain specific workflows
• Future work
Motivation

• Numerous applications for molecular simulations and docking, e.g.
  • Materials science
  • Structural biology
  • Drug design
• Sophisticated tools and algorithms support scientists
• High-performance computing facilities are available
Motivation

Drawbacks of using molecular simulations and docking
• Usability of tools is limited
• Complexity of methods
• Lack of graphical user interfaces
• Complexity of infrastructures
• Many end users lack computer science background

⇒ Need for self-explanatory and intuitive user interfaces
⇒ A science gateway for molecular simulations and docking
Web-based science gateways

• Single point of entry
• Possibility to customize views and tools
• Store user preferences
• No installation of software on the user’s side
• No firewall issues
MoSGrid

Molecular Simulation Grid (D-Grid project)

Goal

Providing users with Grid access to molecular simulation tools and docking tools via a web-based science gateway

www.mosgrid.de
MoSGrid in a Nutshell

- Portal: WS-PGRADE
- Workflow
- High-level middleware service level: gUSE
- Grid resources: UNICORE 6

Cloud File System XtreemFS

Structure

Recipe

Result

www.mosgrid.de
Credential Management

• User management based on Liferay features
  – Community management
  – Organization management
• X.509 user certificates
• SAML (Security Assertion Markup Language)
  – Minimize credential data transfers
  – Set of maximum hops for trust delegation
  – Usable for single sign-on infrastructures (e.g., Shibboleth)
Credential Management

Unicore Credential management

SAML Assertion

Browse assertion:

Keystore (.jks or .p12):
Einstellungen\Sandra\Eigene Dateien\certs\gesing2010.p12

Password:

SAML Assertion:

tere und Einstellungen\Sandra\Eigene Dateien\certs\Assertion

Generate trust delegation

Upload Message:
WS-PGRADE
gUSE Architecture

grid User Support Environment

User interface
WS-PGRADE

High-level middleware service layer
gUSE

Grid resources
middleware layer
UNICORE 6
gUSE Submitter

Interface GridService
• actionJobSubmit
• actionJobAbort
• actionJobOutput
• actionJobStatus
• actionJobResource

www.mosgrid.de
gUSE Submitter for UNICORE

**gUSE**

- JOB1
- JOB2
- JOB3
- JOB4

**UNICORE 6**

- actionJobSubmit
- UNICORE Atomic Services
  - 1 Security
  - 2 Registry
  - 3 Submit job
  - 5 Start job

**Workflow engine**

- JOBn

**UNICORE submitter (UCC lib)**

- 4 Upload data

**Resources**
Distributed Data Management

- XtreemFS is an object-based grid and cloud filesystem
- Ability to minimize data transfer
- Low latency, local availability through replication
- Grid Security Infrastructure (GSI) support

www.mosgrid.de
Distributed Data Management

- XtreemFS integration
  - Portlet
  - UNICORE
  - GSI support
- Data flow
  - WS-PGRADE
  - XtreemFS
  - Frontend nodes
  - Compute nodes
- UNICORE mediates data transfers
Domain Molecular Dynamics

- Study and simulation of molecular motion
- Provide a molecular dynamics service on multiple levels
- Direct upload of job descriptions
- Workflows and standard recipes for repeating tasks
- Analysis of relevant properties
Equilibration of Proteins

- Proteins from databases (e.g., the Protein Data Bank, PDB) do not necessarily represent a near-native conformation/configuration
- For all kind of production runs a minimization and an equilibration is an indispensable prerequisite
- Eases the work of experienced users
- Lowers the hurdle for novice users
UseCase: Gromacs_EQ

structure (pdb/gro)

EM.mdp (mdp)
grompp
topol.tpr

pdb2gmx structure (pdb)
editconf box (pdb)
genbox Solvated (pdb)
grompp adj. Top. (top/itp)
topol.tpr mdout.mdp

www.mosgrid.de
MD Portlet
Domain Quantum Chemistry

• Study and simulation of molecular electronic behavior relative to their chemical reactivity

• Survey - MoSGrid Community
  – First implementation for Gaussian
  – Then support for
    • Turbomole
    • GAMESS-US
    • Further relevant QC applications
Domain Quantum Chemistry

Gaussian Jobs

• Single input file
  – Defines molecular geometry and task

• Result
  – Not structured output
  – Platform dependent checkpoint file

• Integrated multi-step job option
  – Not usable for generalized workflows
Domain Quantum Chemistry

First prototype

• Workflow controlled by portlet
• Three phases
  – Pre-processing
  – Job execution
  – Post-processing
Domain Quantum Chemistry

Workflows

Assisted job creation
  – Guiding GUI
  – Most common options available

Pre-created job description
  – Upload of Gaussian job description file

Monitoring of jobs

Post-processing and presentation of results
Domain Quantum Chemistry

Preprocessing

• Portlet (GUI) supports common options

• Automatic generation and submit of job description
Domain Quantum Chemistry

Post-processing

• Parsing of result file

• Python scripts executed by portlet
  – Relevant information about molecular properties

• Data in CSV-Format saved and accessible
Future Work

• WS-PGRADE
  – Integration of the UNICORE IDB to offer drop-down boxes of available tools

• MD- and QC-Portlet
  – Adoption to gUSE workflow engine via the ASM (Application Specific Module) libraries

• Data Management
  – Extension of XtreemFS for SAML
Thank you!