A Lightweight Platform for Managing Biomedical Simulation

Stefan Zasada
University College London
## AHE Motivation

- Problems for individual users installing/compiling/optimizing application
- Complexities of using computational infrastructures
  - Job workflows, from staging data to launching jobs
- Security can be complicated
  - Applying for certificates
  - Generating MyProxy
- A solution is needed to simplify usage for scientific end users → allow clinical scientists to run simulations at the click of a button.

<table>
<thead>
<tr>
<th>Usability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Usage</td>
<td>Security</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Workflow</th>
<th>RESTful interface</th>
</tr>
</thead>
</table>
What is AHE
- Virtualises Application on the computational resource
- Show components diagram
A number of constraints were placed on the design of AHE to ensure it met our requirements:

- To use AHE, the end user need only install the AHE client, and doesn't need any extra supporting tools
- It is assumed the client is NAT’d and behind a firewall, so the server cannot make any inbound connections to the client
- The client doesn’t maintain and state information locally on the applications it is used to launch
- The user shouldn’t have to update their AHE client when the middleware on the back end grid resources change.
What was AHE 2.0 used for & disadvantages

What’s new in AHE 3.0
- Java
  - Hibernate
  - JBPM
- Security handling
- Reliability
- RESTful interface

AHE 3.0: What’s New

- Re-Implemented in Java
  - Hibernate
  - JBPM
  - Can be deployed as an executable application using embedded Jetty Server
- Security Handling
  - Checks and Auto-Generate MyProxy Certificate
- RESTful Interface
- Can submit to QCG-Computing, Globus & Unicore back ends.
- Life cycle diagrams
- Talk about application Instance
App Instance Lifecycle
AHE grid interop was demonstrated at SC’08, with the AHE being used to submit jobs to both DEISA (Unicore) and TeraGrid (Globus). It can do this since AHE 2.0 can submit jobs to both Globus WS-GRAM and Unicore BES (it does this by transforming its internal state representation (a WS-ResourceProperties document) to JDD in the case of Globus or JSDL (in the case of Unicore BES) by applying an XSLT transform.
AHE can submit jobs to local resources as well as grid machines from a single interface, meaning it is useful for bridging the gap from campus level resources to TeraGrid etc.

It has both GUI and command line clients which interoperate with each other.

It provides a simple, consistent interface to a wide range of back end grid middlewares.
AHE GUI client provides a wizard to allow users to launch their applications – it steps through choosing a machine, staging input data, setting job arguments and launching the application.

The client uses a system of parser plugins to parse application input files (e.g. a NAMD job configuration file) to discover input and output data that needs to be staged. New plugins can be created for new applications, or data to be staged can be configured manually.
Notes from TG’11: Globus Online is only for file transfer - it doesn't cover all of the functionality that we tested in our paper. If this is the best GUI suggestion he can make, I take it that there isn't really a comparable GUI for Globus.

Globus Online website (this is not comparable to AHE or AHE+ACD or UNCORE.

Globus Online is a fast, reliable file transfer service that makes it easy for any user to move any data anywhere. Recommended by HPC centers and user communities of all kinds, Globus Online automates the time-consuming and error-prone activity of managing file transfers, so users can stay focused on what's most important: their research.
Biomedical Computing Requires Co-Allocation

- We can reserve multiple resources for specified time periods
- Co-allocation is useful for meta-computing jobs using MPIg, viz and for workflow applications
  - Ultimately, compute needs to be scheduled in with clinical workflow.
- We use HARC - Highly Available Robust Co-scheduler (developed at LSU) and
- QCG-Computing (developed at Poznan Supercomputing Centre)

The use of multiple machines makes cross site allocation necessary - users need to know that all of the resources will be available at the same time in order for the different MPIg components to be started.

It's also useful for applications that require real time visualisation.

AHE interfaces with HARC to provide advanced reservation capabilities.
Patient-specific medicine

- ‘Personalised medicine’ - use the patient’s genetic profile to better manage disease or a predisposition towards a disease
- Tailoring of medical treatments based on the characteristics of an individual patient

Why use patient-specific approaches?

- Treatments can be assessed for their effectiveness with respect to the patient before being administered, saving the potential expense of ineffective treatments

Patient-specific medical-simulation

- Use of genotypic and or phenotypic simulation to customise treatments for each particular patient, where computational simulation can be used to predict the outcome of courses of treatment and/or surgery
VPH-Share will provide the organisational fabric (the infrastructure), realised as a series of services, offered in an integrated framework, to expose and to manage data, information and tools, to enable the composition and operation of new VPH workflows and to facilitate collaborations between the members of the VPH community.
Several issues:

1. Location of data storage! What are the applicable laws and regulations governing the data (DPA, HIPPA) ? (Forgo)
2. Who to trust with the data? Adequacy of the security controls? Auditing in the cloud?
3. Who decides the level of assurance required?
- Get better image from VPH-Share slides D2.3

- AHE developed for VPH-Share cloud infrastructure, a gateway from cloud to HPC
- Provide access to Grid App as SaaS paradigm
- 4 flag ship workflow which may utilise HPC resources
AHE 3.0 and Taverna: Example

- AHE has been deployed with Taverna to execute batched HIV simulations.
- Molecular Dynamic Simulation using GROMACS
- The simulation involves two Applications
  - Grompp: Pre-processor application, reads the data and configuration files and converts it into a format that Mdrun can read.
  - Mdrun: molecular dynamic simulator.
- Uses AHE to manage submission to grid resources

Using AHE with Taverna for Virolab

- Get Job
- Execute and retrieve data
ContraCancrum: Towards patient specific lung cancer therapies

EGFR mutations for lung cancer

- Over expression of Epidermal Growth Factor Receptor (EGFR) is associated with cancer
- Target for inhibitory drugs
- Important mutations include deletions
- Again binding affinity calculations can be used to determine mutational effects

EGFR Tyrosine Kinase Domain

Applications used include: NAMD, CHARMM, AMBER...
Individualized Medicine Simulation Environment - IMENSE

Aims

• Central integrated repository of patient data for project clinicians & researchers
• Storage of and audit trail of computational results
• Interfaces for data collection, editing and display
• Ultimately provide a data environment for
  • Integration of multi-scale data
  • Decision support environment for clinicians

Critical factors for success and longevity

• Use standards & OS solutions
• Use pre-existing EU FP6/FP7 solutions & interaction with VPH NoE ToolKit
IMENSE Components

- **Data repository** – this is the key store for project data containing all patient data, and simulation data derived from the patient data.

- **Integrated web portal** – this provides the central interface from which users upload and access data sets, and analysis services. The interface provides users with the facility to search for patient data based on a number of criteria.

- **Web Services** – the web services platform implements required data processing functions.

- **Workflow environment** – the workflow environment provides a virtual experiment system, from which users can launch pre-defined workflows to automate moving data between the data environment and multiple data processing services.
• Goal: To make multiscale applications possible on European production e-Infrastructures.
• 5 application domains within the project,
  • + cooperation with external communities.
• EU FP7 project with partners throughout Europe.
• http://www.mapper-project.eu

Distributed Multiscale Computing workshop at this conference!
Points above summaries what has been said

Key point  → AHE provides a simple unified interface to a wide range of local and grid resources, giving the user a view on to a federated grid made up of machines from multiple providers.

AHE also provides a building block on which Gateways and other types of portal can be built.