G-Fluxo: A Workflow Grid portal ready for Computational Chemistry

CESGA

Aurelio Rodríguez, PhD

Applications Technician

aurelio@cesga.es

CENTRO DE SUPERCOMPUTACIÓN DE GALICIA

ESTABLISHED IN 1993 IN SANTIAGO DE COMPOSTELA

CESGA

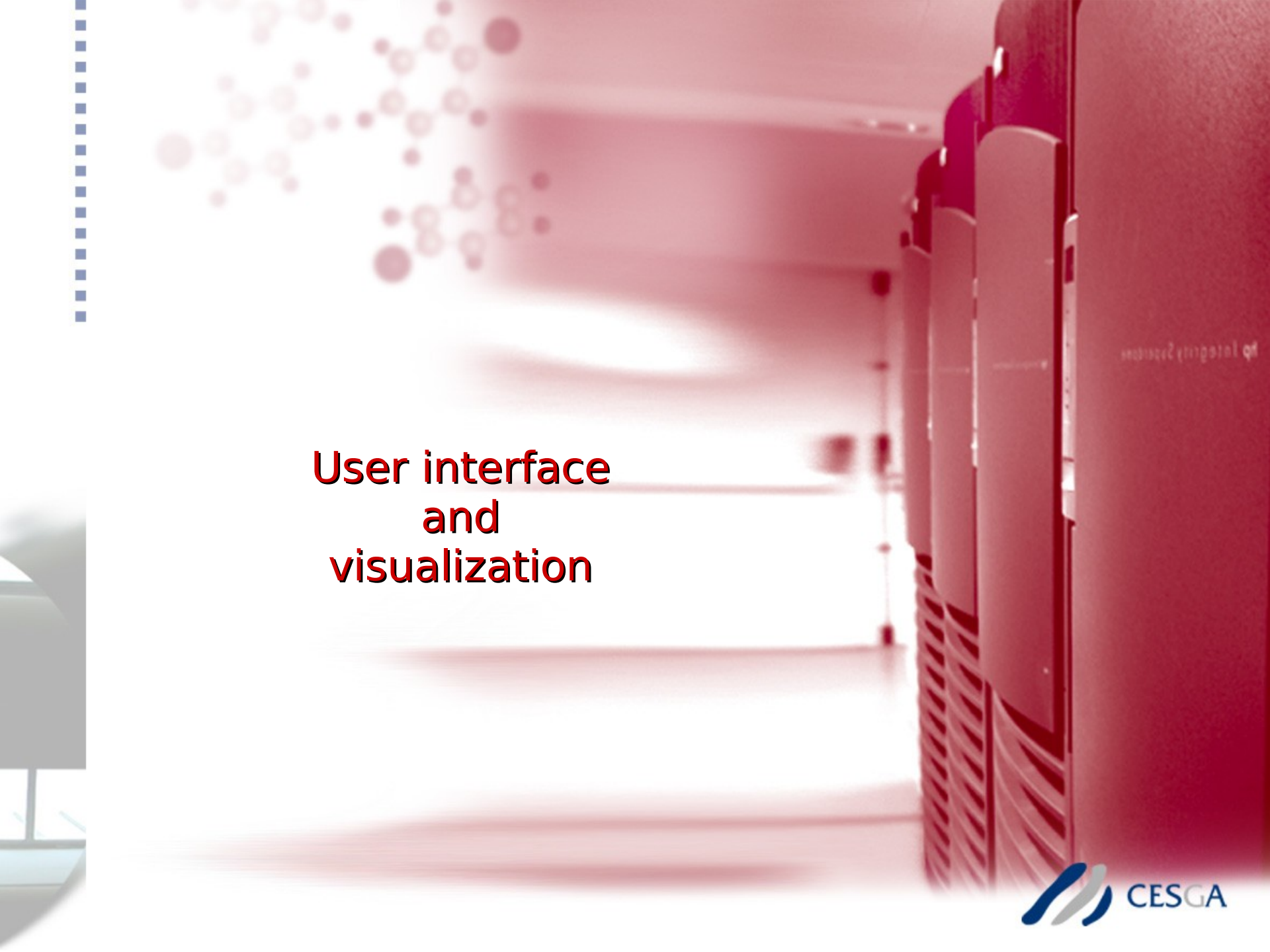
CESGA

Index

- User interface and visualization
  - Inherent problems on Distributed Infrastructures
  - Portals
- G-Fluxo proposal
  - P-Grade
  - Local Infrastructures (not Grid related). DRMAA
  - DAG Workflow (condor supported)
- Meta Workflow (XPDL)

CESGA

User interface and visualization



Inherent problems on Distributed Infraestructures

- Authentication
- Distributed Data
  - Data centralization is desirable
- Job management
  - Job submission to High Throughput Computing resources
  - Job monitoring (Visualization steering)

CESGA

Portals

E-scientists' concerns:

- How to concentrate on my own research if the tool I would like to use is in continuous change?
- How can I learn and understand the usage of the Grid?
- How can I develop Grid applications?
- How can I execute Grid applications?
- How to tackle performance issues?
- How to use several Grids at the same time?
- How to migrate my application from one Grid to another?
- How can I collaborate with fellow researchers?

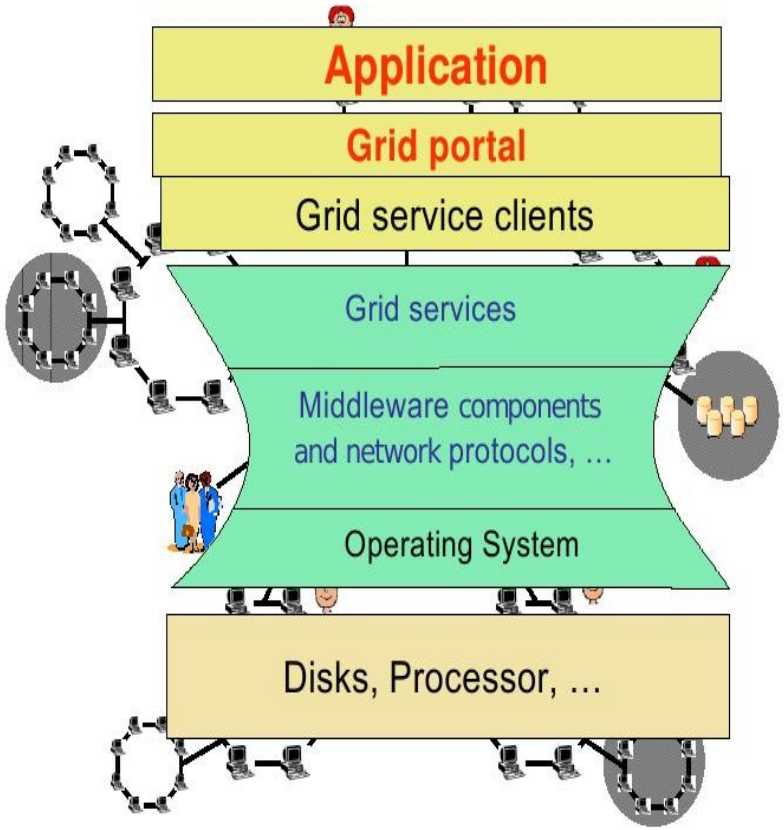
A good Grid Portal must answer these concerns!!!

CESGA

Portals

Typical functionalities of a Grid portal:

- User authentication
  - Web security (e.g. name&passwd) grid security (proxies)
- Data management
  - Web protocols / grid protocols (e.g. HTTP)/(e.g. GridFTP)
- Job management
- Local resource/distributed resources
- Visualization
  - Application input/output/execution progress



CESGA

Support for Emerging Technologies

Welcome to Open Cirrus™ the HP/Intel/Yahoo! Open Cloud Computing Research Testbed

Open Cirrus is an open cloud-computing research testbed designed to support research into the design, provisioning, and management of services at a global, multi-datacenter scale.

Open Cirrus is a collaboration of several organizations:

HP Labs

Intel Research

Yahoo! Research

University of Illinois at Urbana Champaign

Karlsruhe Institute of Technology, Germany

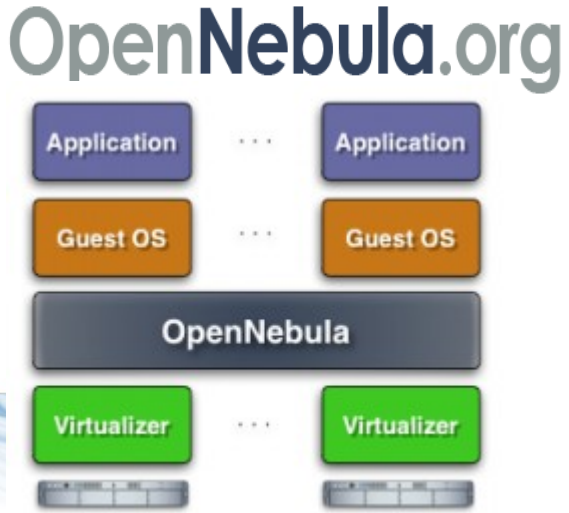
Infocomm Development Authority, Singapore

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers.

Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios.

Sign Up For Amazon EC2



Ian Foster, Yong Zhao, Ioan Raicu, Shiyong Lu, "Cloud Computing and Grid Computing 360-Degree Compared", arXiv:0901.0131v

CESGA

G-Fluxo proposal

P-Grade

- General purpose, workflow-oriented computational Grid portal. Supports the development and execution of workflow-based Grid applications – a tool for Grid orchestration
- Based on GridSphere-2
  - Easy to expand with new portlets (e.g. application-specific portlets)
  - Easy to tailor to end-user needs

Solves Grid interoperability problem at the workflow level

http://www.lpds.sztaki.hu/pgportal

CESGA

P-Grade

Main features:

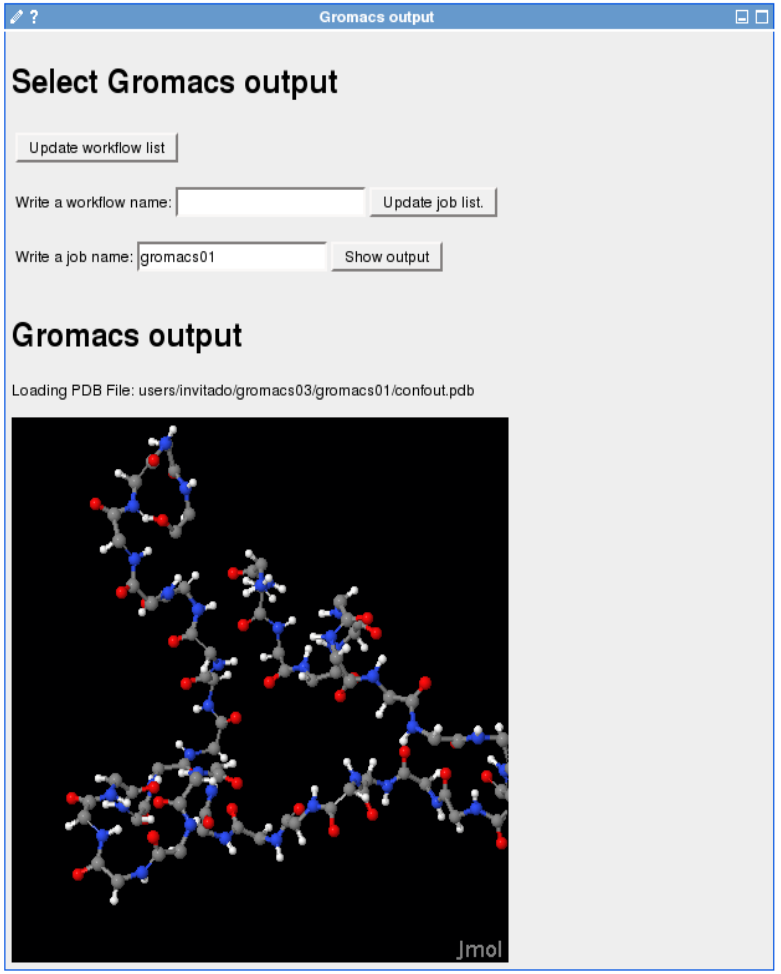
- Interoperability with Globus Toolkit 2, Globus Toolkit 4, LCG and gLite grid middlewares
- Secured grid access mechanisms using X.509 certificates and proxy credentials
- Built-in graphical editor to design and define grid workflows and grid parameter studies
- Integrated workflow manager that orchestrates the fault tolerant execution of grid applications
- On-line workflow, job monitoring and visualization facilities
- Multi-grid access mechanism to utilize multiple grid implementations simultaneously
- MPI execution in Globus and gLite grid environments
- Graphical MDS and BDII grid information system browser
- Support to include local and remote storage files into grid applications
- User level storage quota management to protect against server overloading
- Workflow import-export-archive service

CESGA

P-Grade

Expanding its functionality:

- A graphical interface for specific applications:
  - A portlet to show the simulation output Gromacs based on Jmol. The user can see the job results in an interactive and 3D way.



Web3D?

CESGA

Local Computing Infrastructures

- What happens if a specific job requires HPC resources (a high demand of memory, low latency/high bandwidth network) or have a private license only available for an specific machine?
- A E-scientist have access to different computing resources:
  - A Personal Computer
  - A Computer Cluster (Lab or institution)
  - Grid resources
  - Supercomputing center

eGee
Enabling Grids
for E-sciencE

CompChem
Virtual
Organization

CESGA

Local Computing Infrastructures

A modification in P-Grade have been done so it is possible to configure "grids" formed by just one local computational resource.

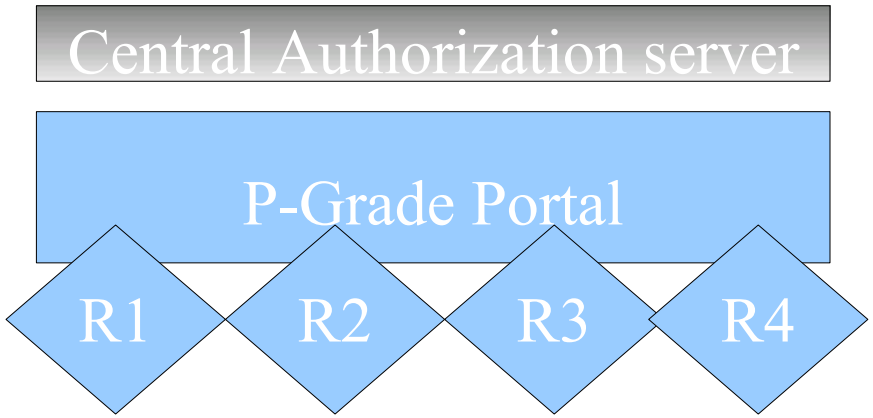
- The local resource must be:
  - Accessible by ssh
  - A DRMAA C supported scheduler must run on it

CESGA

Local Computing Infrastructures: A centralized schema

It is possible to support a centralized authorization schema based on LDAP (Lightweight Directory Access Protocol)

- The local resource must:
  - Use a accounting policy based on LDAP
  - Export a local file system through NFS (The P-Grade portal must be able to mount it)
  - A DRMAA C supported scheduler must run on it



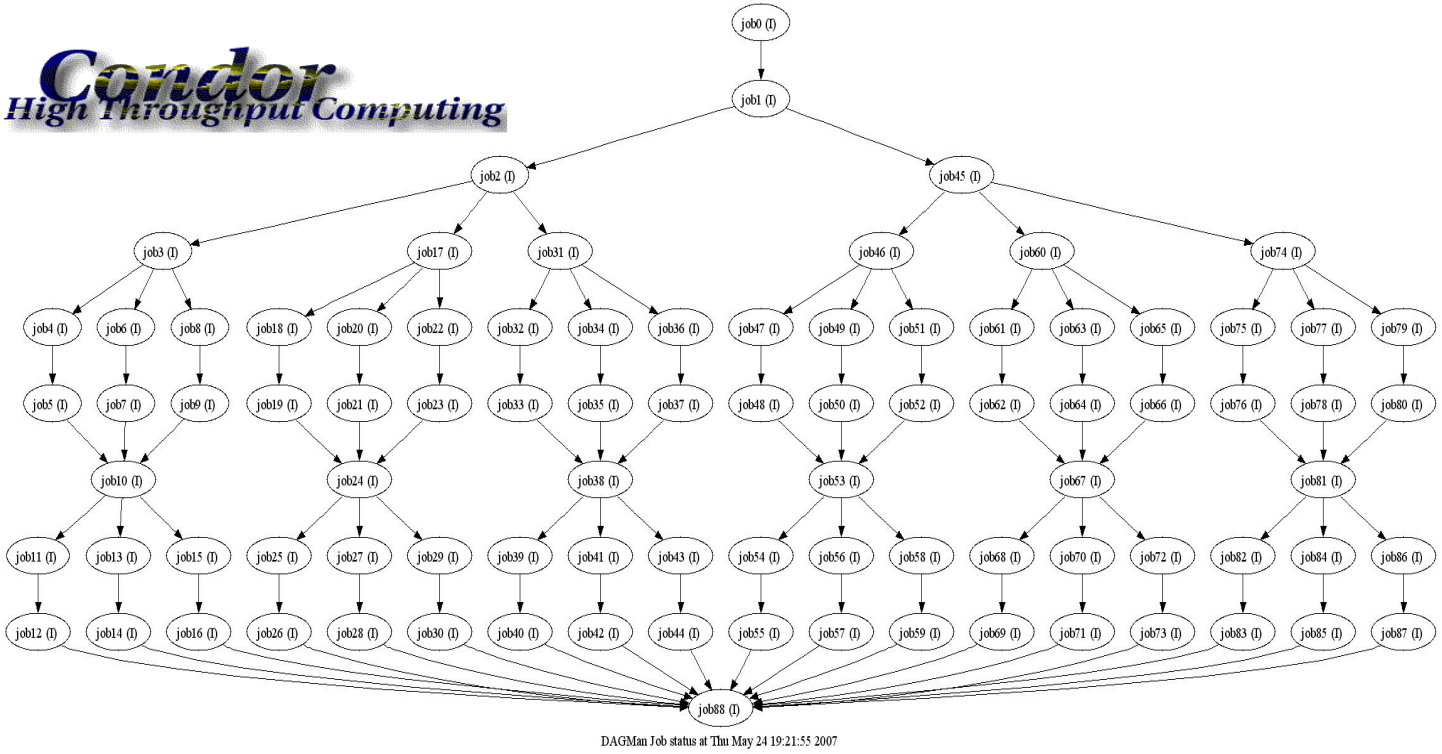
OpenLDAP™
http://www.OpenLDAP.org

CESGA

DAG Workflow (condor suported)

Workflow support:

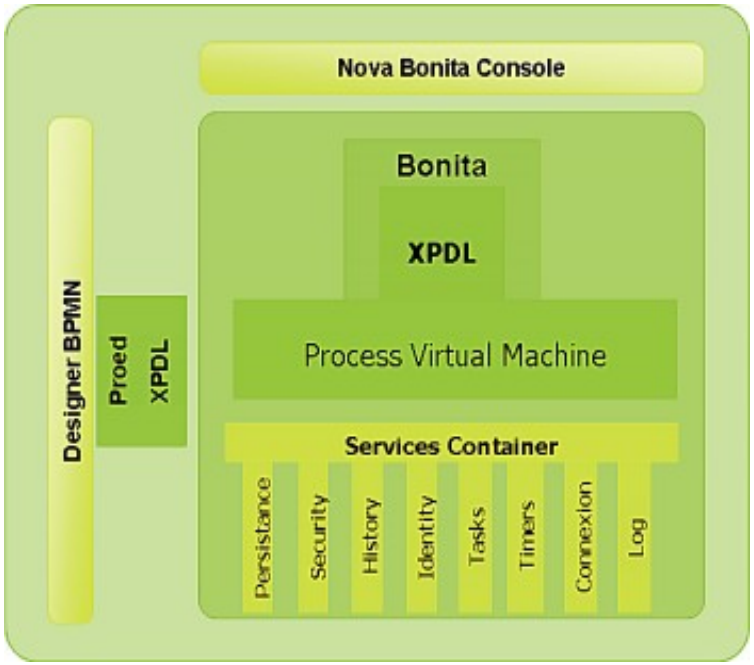
- P-Grade support DAG Workflows through the use of condor for job management.
- In this way quite complex workflows are supported



CESGA

Meta Workflow (XPDL)

- In the future, P-Grade DAG Workflow submission will be linked with the Nova Bonita Console:
  - A workflow open source solution
  - Support for the standard XPDL process definition language



Meta Workflow (XPDL)

- Why we are interested in XPDL:
  - Format standardized by the Workflow Management Coalition (WfMC) to **interchange Process definitions** between different workflow products, ie between different modeling tools and management suites.
  - BPEL and XPDL are entirely different yet complimentary standards. **BPEL is an "execution language"** designed to provide a definition of web services orchestration. It defines only the executable aspects of a process, when that process is dealing exclusively with web services and XML data. BPEL does not define the graphical diagram, human oriented processes, subprocess, and many other aspects of a modern business process

CESGA

Meta Workflow (XPDL)

- Bonita Features:
  - **Powerful BPM API** covering deployment, definition, runtime and history BPM data
  - Support for **XPDL 1.0 activities** : Join, Split, Activity
  - Support of main **XPDL 1.0 elements**: Datafield, DataType, Participant, Transition, RedefinableHeader,Transition Restriction, Package...
  - **Advanced entities/resources**: Hooks, mappers, performer assignments and activities multiinstantiators (via XPDL extended attributes)
  - **Subprocesses support**
  - Activities **multi-instantiation** support
  - **Iterations/cycles support**
  - **Task Management module** handling init, ready, executing, finished, dead, suspend and resume states

CESGA

Meta Workflow (XPDL)

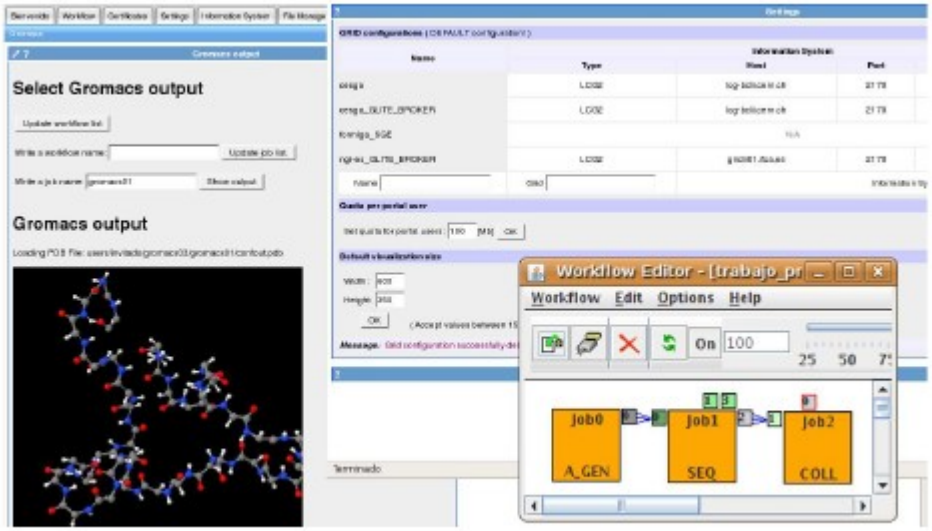
- **BPM Designer (ProEd):**
  - Eclipse and Desktop versions
  - "Easy BPM project" creation wizard available in Eclipse version
  - Graphical support for advanced Nova Bonita entities: hooks, mappers, performers and instantiators
  - "Smart" conditions editor: graphical definition of complex conditions based on multiple operators and variables types
  - Automatic generation of start and end BPM steps
- **BPM Console**
  - Web 2.0 console supporting both desktop and traditional portal layout modes
  - Monitoring vs Worklist applications (portlets)
  - Internal user repository handling access rights to applications
  - Automatic generation of forms vs customized forms
  - Console customization capabilities: on the fly page creation, add/remove applications and widgets, look and feel...
  - Applications (portlets) and widgets support

CESGA

Meta Workflow (XPDL)

Bonita | The Open Source Workflow project

- Scientific Simulation (XPDL definition)
  - Roles
  - Iterative
  - Conditions
  - Decision Taken Steps
    - P-Grade Job (Complex task through a DAG Workflow)



CESGA

Meta Workflow (XPDL)

- Still we need application communication:
  - Common data format: for example Q5COST
  - Wrappers
  - Conversion routines

Open Babel

CESGA

Summarizing...

- G-Fluxo try to give answers to the computational needs a e-scientist could have:
  - Orchestration of the use of many different Computational infrastructures
  - Specific support for applications
  - Scientific Simulations Modelling:
    - Implementation of different methodologies
    - Reproduction

CESGA

Acknowledgements

- P-Grade people (MTA-SZTAKI)
- Eduardo Gutierrez, Javier López Cacheiro (CESGA)
- Xunta de Galicia for the funding through the INCITE project 07SIN001CT
- COST D37 action

CESGA

Thanks for your attention!! Questions??

CESGA