Distributed visualization services on the grid for science & engineering

Lakshmi Sastry

(lakshmi.sastry@stfc.ac.uk)

Scientific Applications Group

e-Science Centre

Science & Technology Facilities Council

UK

Presentation overview

- Introduction

  – STFC, eSC, Scientific Applications Group

- Integrated services to support UK science & Engineering
  - National Grid service
  - Atlas Data Store and Digital Curation Centre

- Visualization on the grid

  – the need

  – approaches to visualization in distributed environments

  – the requirements

- Challenges & Future

  – user expectation

  – hardware and software issues

  – resource allocation

  – XtreemOS

Introduction

STFC – Science & Technology Facilities Council – runs large scale science Facilities to support UK/international scientific research & development

- Harwell & Daresbury Science & Innovation Campuses, employing about 2000 staff; 4500 including all smaller science based enterprises on both sites
- ISIS, Diamond, CLF, HPC, SSTD, PPD (Tier1, EGEE)
- e-Science Centre formed in 2000 to support HPC needs for very large scale collaborative science
- Scientific Applications Group is the R&D wing also responsible for implementing technological solutions
  - HPC simulations
  - Data integration from experiments, observations
  - Metadata and ontology framework for data sharing
  - High resolution visualization of extremely large datasets
  - Image reconstruction
  - Data analysis
  - Trusted collaborations between VO
  - OS level embedding of services

Science & Technology Facilities Council
e-Science

E-Science Services

- National Grid Service provides secure access to distributed compute and data storage resources 24/7
- Capable of storing PBs of data per day (e.g. Tier 1)
- Data Management, curation
  - cataloguing infrastructure using open standards, metadata schema, ontology infrastructure
- Internationally recognised expertise and contribution to Web2.0 technologies and open science forum
- State of the art R&D of trusted security in distributed environments
- Software development and services for simulations, visualization, image analysis, data analysis and collaboration
- Embedding services into operating systems to support seamless real-time use
- Beginning to work with Arts & Humanities communities

National Grid Service

Integrated access to computation and data

- Some core provision
- Some special or restricted services
- Increasingly dominated by partner owned resources
- Access for UK academic researchers
- Funded by HEFCE + RCUK
- Service started in September 2004
- ~£7m core funding over 6 years.
- Currently ~1000 registered compute users

NGS

- Vision of NGS

    To provide coherent electronic access for UK researchers to all computational and data based resources and facilities required to carry out their research, independent of location.

    Providing integrated interfaces and infrastructure for research communities to access and share distributed resources

- NGS has four cores sites
- Leeds, Manchester, Oxford, RAL
- Also has partner resources
    - Cardiff University; HPCx; Lancaster University; Queens University of Belfast ; University of Glasgow Computing (2) ; University of Westminster
- And affiliate resources
    - Edinburgh; Imperial College London; Keele University; University of Bristol ; University Of Oxford Particle Physics Department ; University of Reading ; University Of Southampton ; RHUL.
- Partners provide a range or resources
    - Compute, data, hosting, community portals …

# The NGS Core, Partners & Affiliates, Summer 2008

From data and compute to analysis on the Grid

- Data management and Computation are well understood and well used
- The increased amount of resulting data cannot be handled by traditional data flow visualization systems
- Need to support data exploration, moving away from visualization
    - E.g: single isosurface does not convey much, even though compute intensive
    - E.g: 2 seconds animation of OCCAM 1/12 deg data takes 2.5 days of computation
    - E.g computational steering with user in the loop to control simulation is not possible
    - E.g: scientists want to analyse their data, not just visualise

- Solution:
    - Visualization on the grid as an application in itself to support scalability, real-time, use of advanced techniques and access to advanced tools, processing data where it is generated or stored

Visualization on the grid
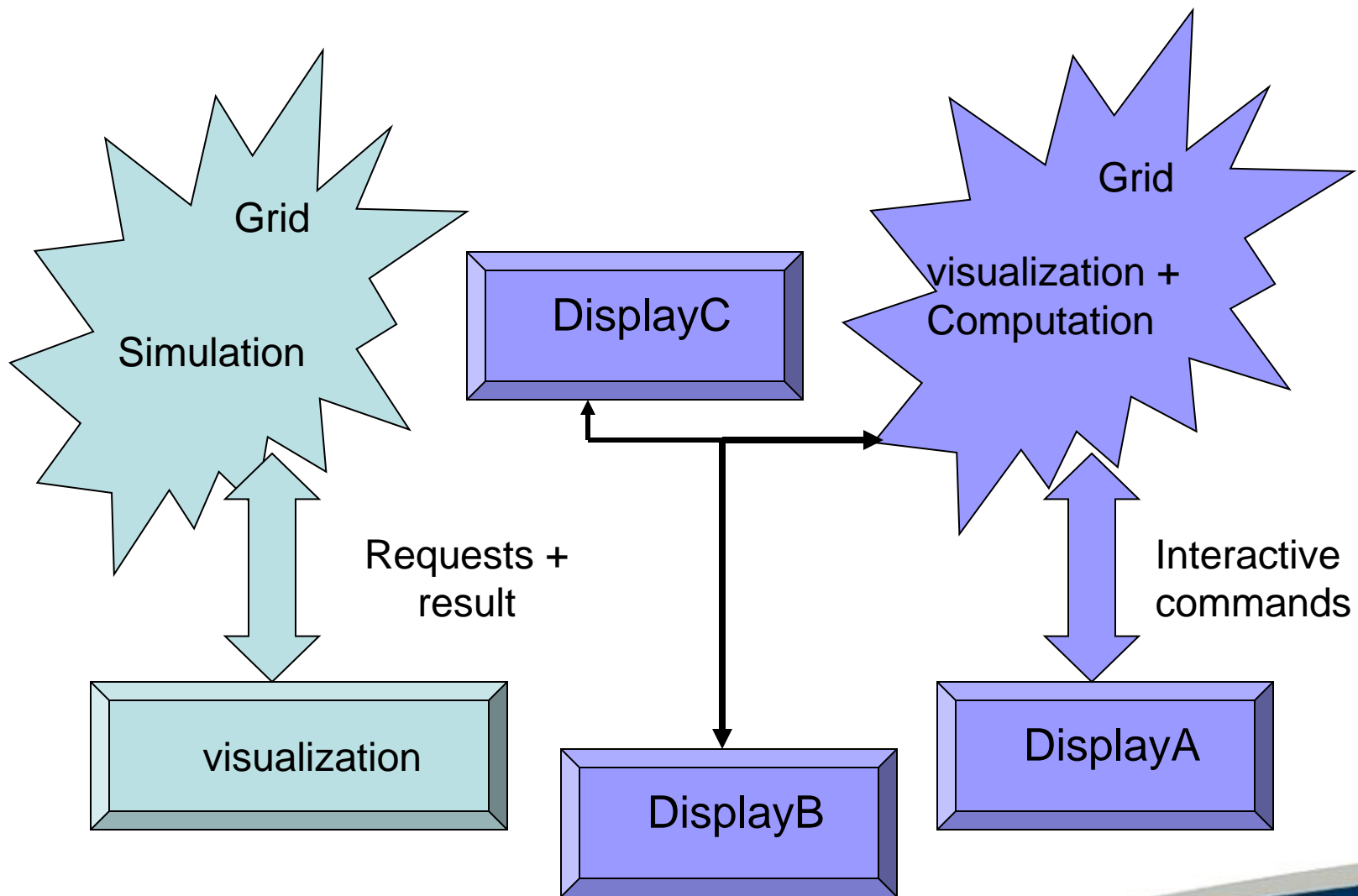
Leading projects to supports scalable visualization

– Dedicated HP visualization systems such as SGI Onyx with Visualizer

– IRIS Explorer, AVS, ParaView, PV3

–CoreGrid's GVK for immersive reality applications

–RAVE

– eViz

–Reality Grid

– Visualization portals

– STFC visualization services

What we needed to support

– Scalability

– Real-time analysis

– Collaborative steering of simulations with real-time intervention and control

– Integration with simulations and observational data from other sources

– Interface with databases to support persistence

– SOA based architecture to handle firewalls and standardised interface

– Interoperability

– Reuse

– A range of applications and users

All the above within the NGS environment

Integrated Visualization and Data Analysis Service at STFC

- – Dedicated hardware configured to support multiple applications and users
- – Redesigned compute intensive visualization techniques to exploit distributed, parallel, high performance resources
- – Web services based interface to the software
- – End to end user support
- – Customisation of user data and applications
- – Interface and display to grid services from a range of desktop tools such as Matlab, IDL.

STFC visualization Cluster

- Dedicated 17 dual processor node Linux based graphics cluster

  - Chromium based remote visualization with ability to deliver real time high resolution images and geometry

  - Parallel rendering using some or all of the nodes

  - Variety of software toolkits

    - VTK, ParaView, IRIS Explorer and in-house software for visualization

    - Chromium, VirtualGL and VRJuggler for tiling, graphics data marshalling and device configuration and use

    - NAG data mining components

  - Can support tiled, stereo, mono and desktop display devices

  - Linked to high performance compute and data clusters on site via 1Gb link and 4Gb infiniband interconnect between nodes

- Interfaces through a variety of desktop tools

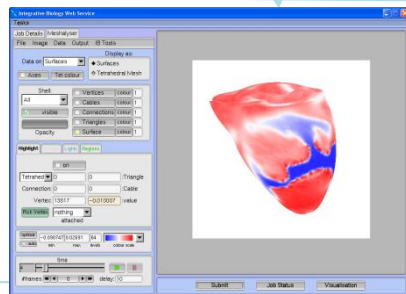# Schematic Overview of STFC Advanced visualization on the Grid
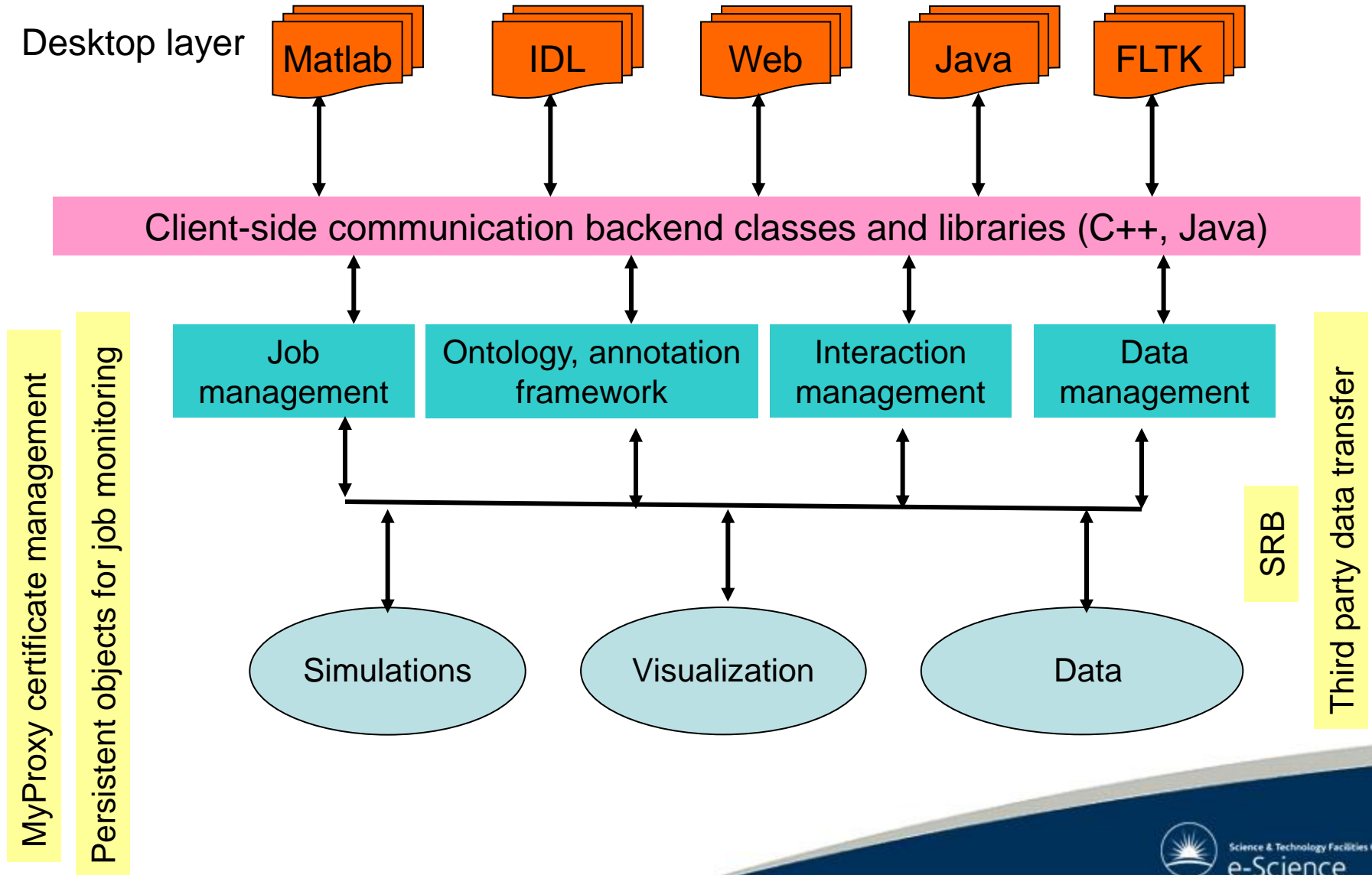
Data Analysis Grid
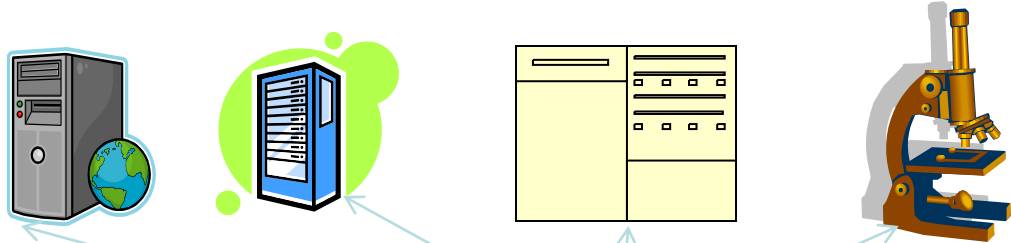
Data and compute resources, data from field and data from photos, camera, simulations

Seamlessly and securely interchanged to support computation, visualization and analysis using grid protocols

Interface for simulation, steering, collaborative visualization and data analysis, embedded into a variety of tools that users already have

Field worker

Researcher in office

Collaborative visualization & steering

Applications

SXD 2002

· Analysis and data fitting for Single Crystal Difractometer

· Fits experimental data to a Gaussian

· Serial code and GUI in IDL

· Performance profiled

· One module grid enabled with parallel Fortran

· Job submission

· Interface through IDL

· Directly uses CoG toolkit
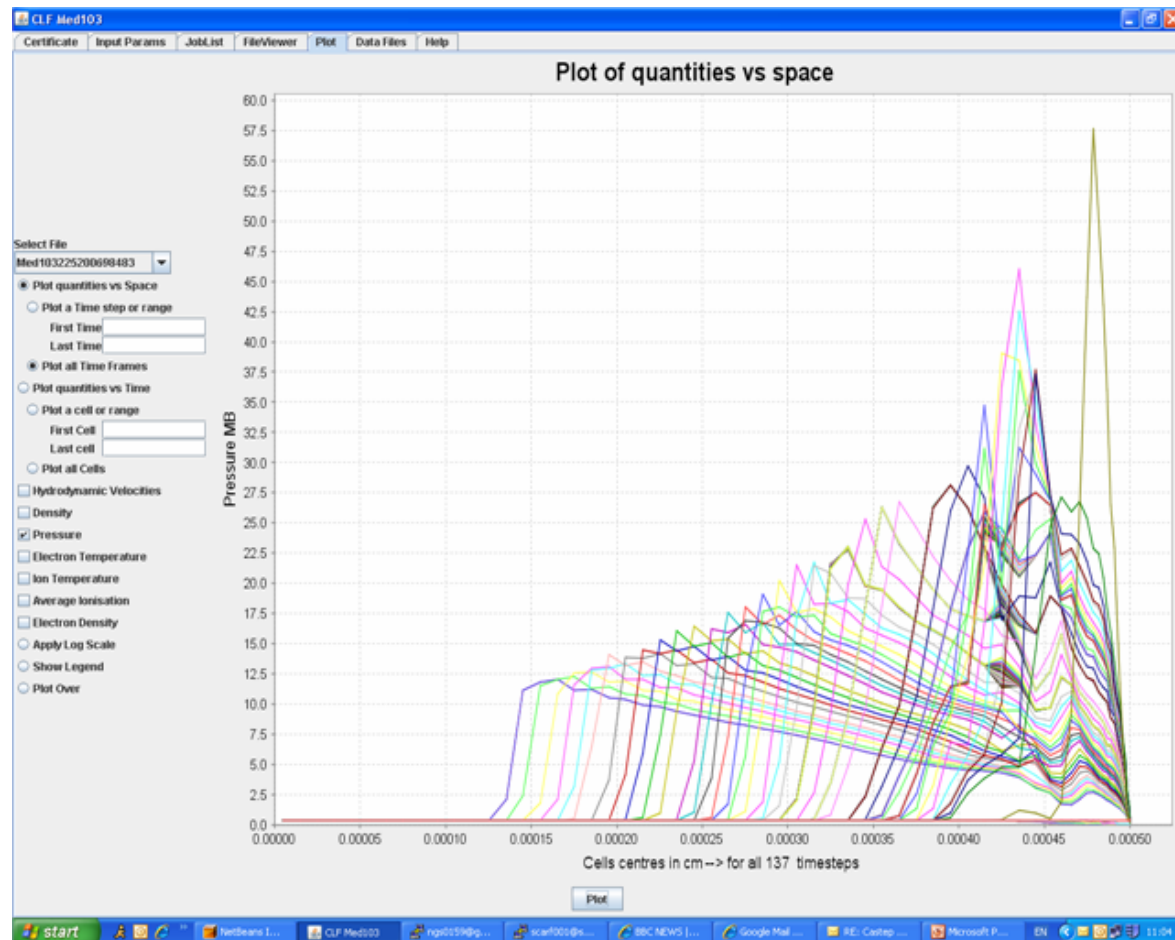
· Scales very well, nearly linear.
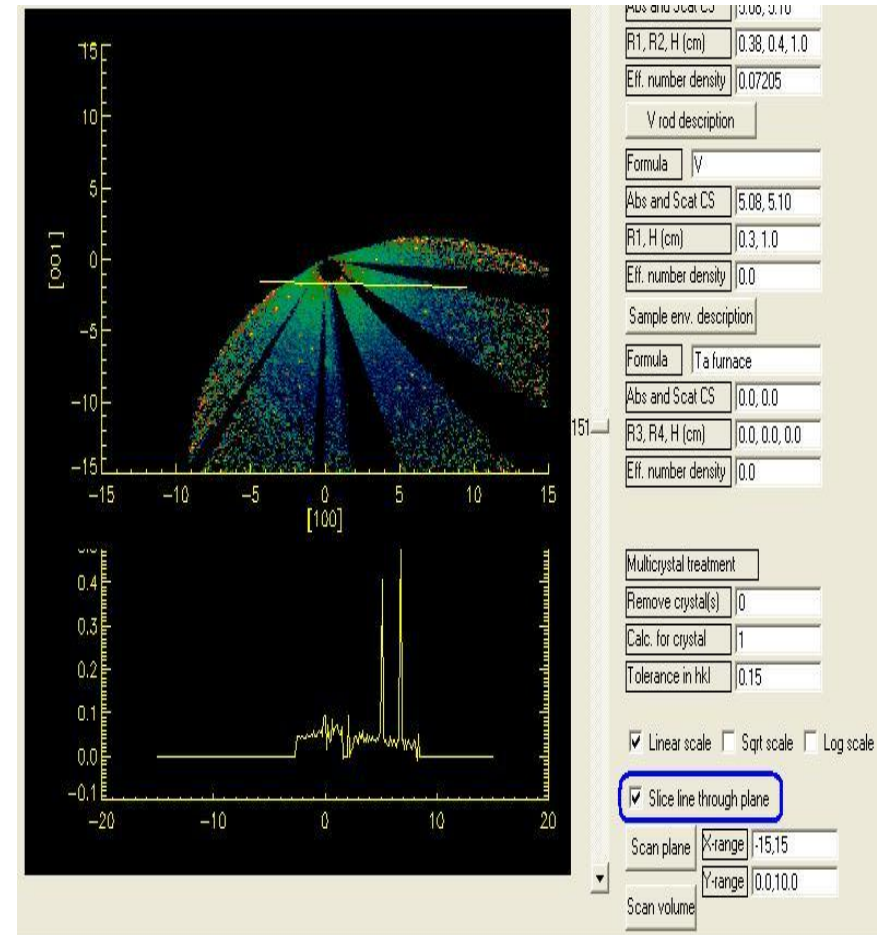
Applications

MED103

· 1D hydrodynamic simulation of laser beam interaction with target

· Grid enabled

· Java Swing GUI

· Access to all parameters for parameter sweep

· Data management with SRB

· Interface to SRB via the GUI to browse results

· Job management

Integrated Services - Integrative Biology Interface

- An integrated grid framework for computational and experimental biology to study disease processes
- EPSRC funded, international
- Services include:
  - job management
  - data management
  - workflow
  - visualisation and image processing
  - Server side data handling
  - computational steering
  - FLTK interface
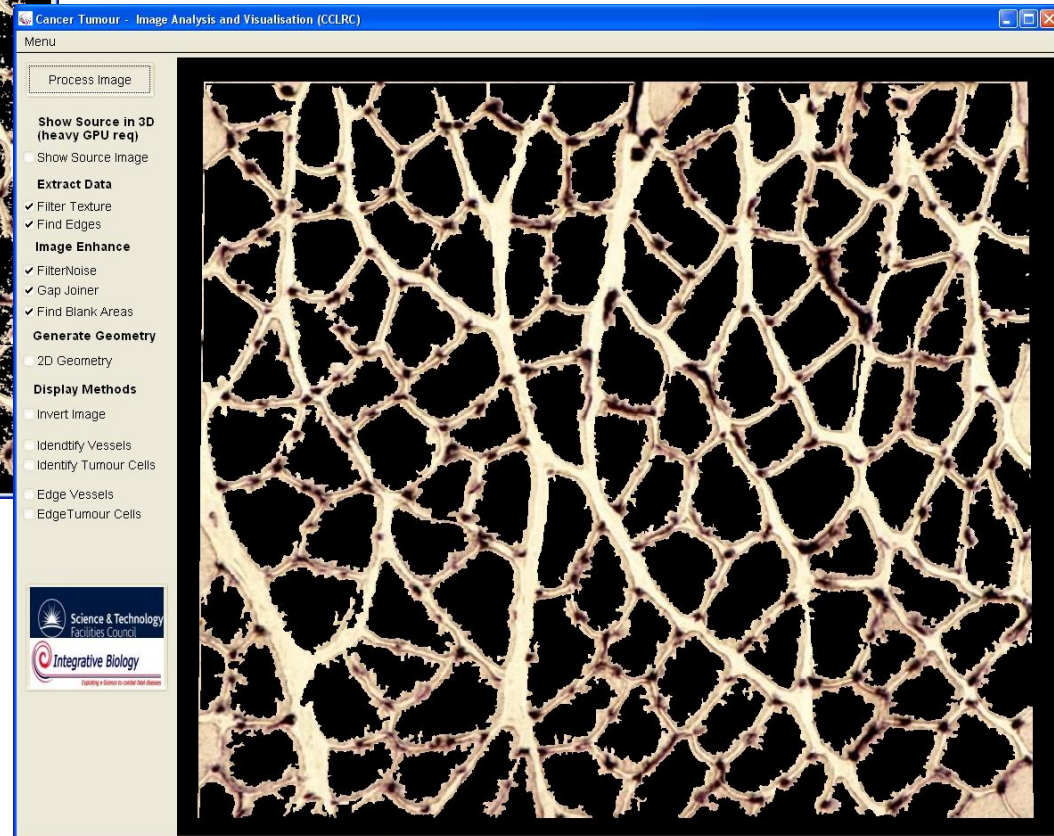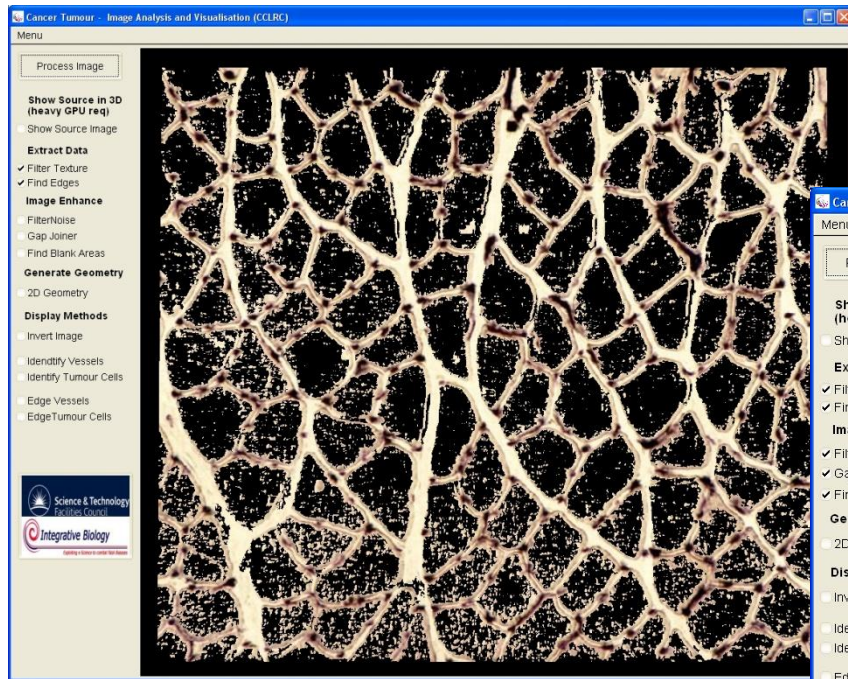 - Integration of application specific toolkits
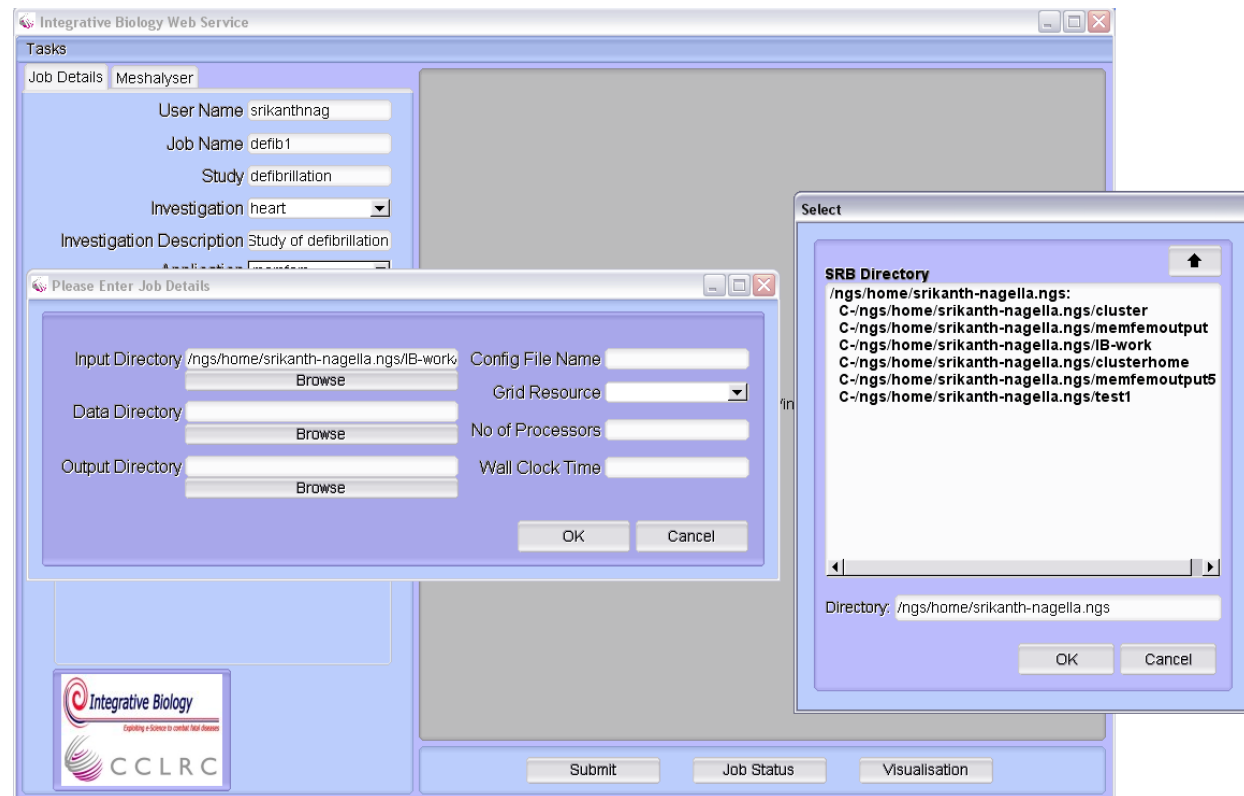
Advanced visualisation & Image processing
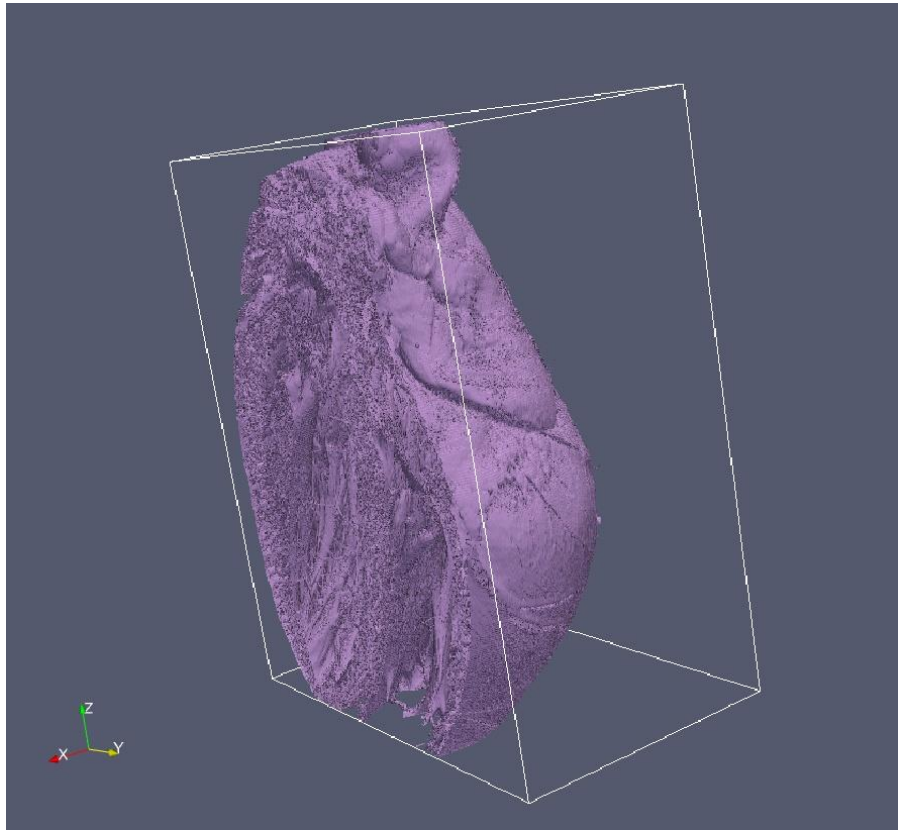
Advanced visualisation & Image processing

vizNET & ENVS

vizNET (**http://www.viznet.ac.uk**) is a JISC funded project
- to promote advanced visualization tools and techniques among user communities,
- to create awareness of best practices,
- to directly support users visualise their data through surgeries.
- Loughborough, Cardiff, Leeds, Manchester and STFC are partners
- STFC is conducting an experimental national visualization service
  - Setting up VirtualGL, VNC and scripting interface fr port forwarding for interactive access
  - Get affiliate status
  - First set of NGS users from Oxford, Reading & Manchester universities
  - Serious technical issues of performance (ParaView), compatibility (IRIS Explorer/IDL with VirtualGL)
  - One to one user support

Simulation & fitting with Matlab interface
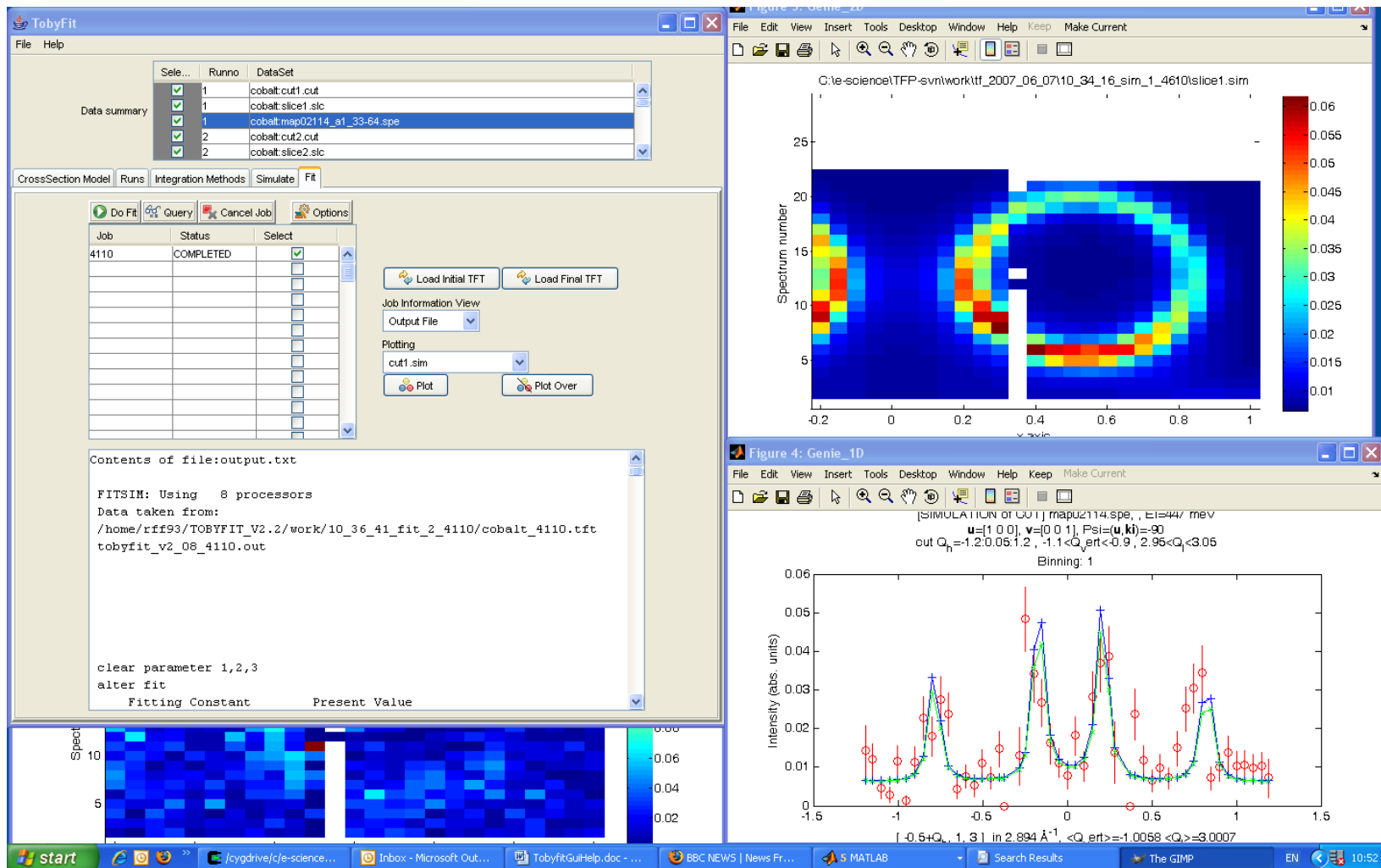
XtreemOS

*Enabling Linux
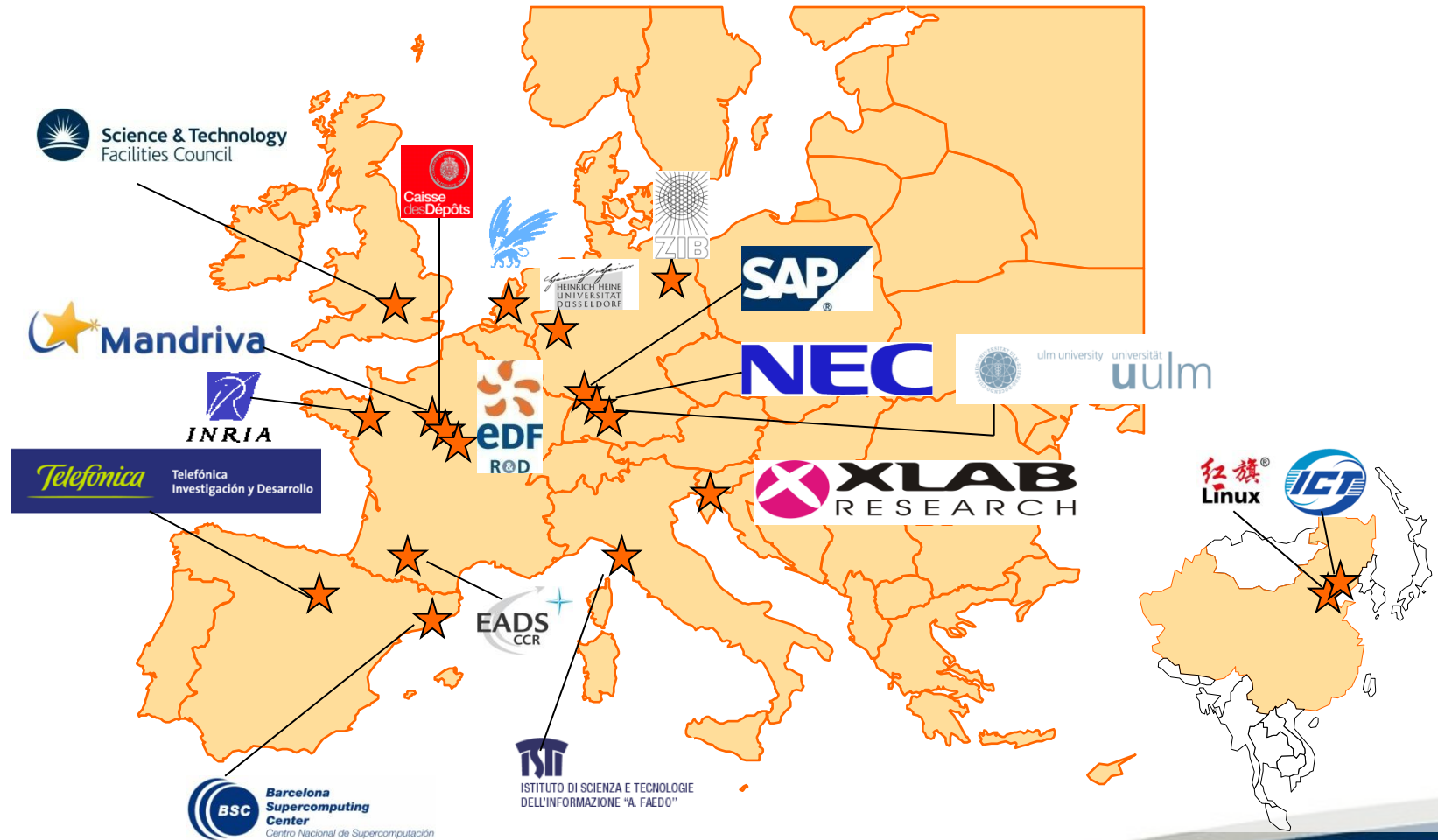for the Grid*

Information Society
Technologies

XtreemOS

4-year R&D project started in June 2006 in the FP6 framework

30 M€ budget, 14.2 M€ EC grant

18 academic & industrial partners from Europe & China

XtreemOS Consortium

XtreemOS Objectives

Design & implementation of an open source Linux-based Grid Operating System with native VO support

Grid Operating System

A comprehensive set of cooperating system services providing a stable interface for a large-scale wide-area dynamic distributed infrastructure

- – Abstraction
- – Sharing of heterogeneous resources in multiple administrative domains

Two fundamental properties: transparency & scalability

- – Bring the Grid to standard users
- – Scale with the number of entities and adapt to evolving system composition

Science & Technology Facilities Council
e-Science

XtreemOS Software Architecture
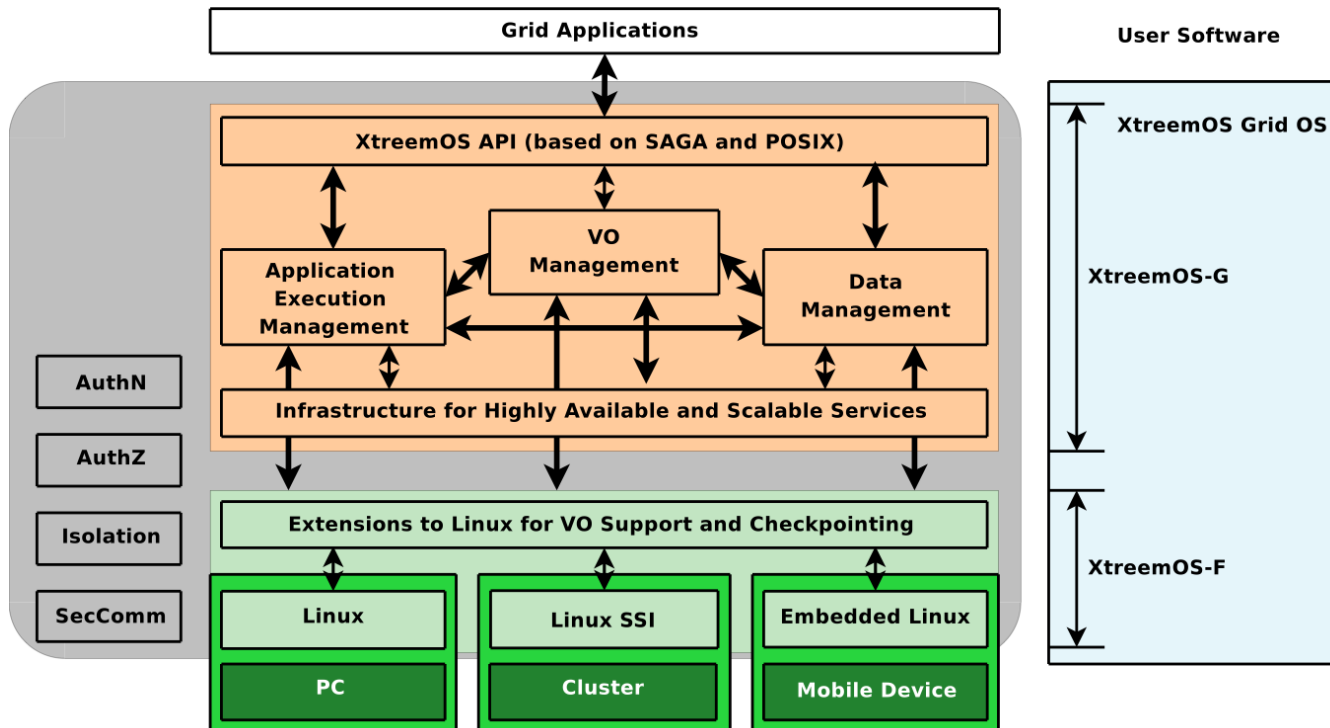
XtreemOS IP project
 is funded by the European Commission under
contract IST-FP6-033576

Science & Technology Facilities Council
e-Science

Future

- Currently seeking to implement XtreemOS in a set of Xboxes to build a affordable visualization cluster for local dedicated data analysis
- Exploit Xbox graphics capability to provide local cluster computing with visualization

Conclusion

- Supporting visualization and interactive data analysis on the grid is more complex than installing a library and provide it to users
- It is quite labour intensive in the initial phase
- The layers of toolkits go out of sync.
- System is too complex for non-computing scientists
- However, it is worth the while and needs more community work