The National e-Science Centre

and

UK e-Science Programme

Muffy Calder
University of Glasgow

http://www.nesc.ac.uk
http://www.dcs.gla.ac.uk/~muffy
An overview

- E-Science
  - challenges and opportunities

- Grid computing
  - challenges and opportunities

- Activities
  - UK and international
  - Scotland and National e-Science Centre
e-Science

‘e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it.’

‘e-Science will change the dynamic of the way science is undertaken.’

John Taylor
Director General of Research Councils
Office of Science and Technology
The drivers for e-Science

- More data
  - instrument resolution and laboratory automation
  - storage capacity and data sources

- More computation
  - computations available, simulations doubling every year

- Faster networks
  - bandwidth
  - need to schedule

- More interplay and collaboration
  - between scientists, engineers, computer scientists etc.
  - between computation and data
The drivers for e-Science

- Exploration of data and models
  - in silico discovery

- Floods of public data
  - gene sequence doubling every 9 months
  - Searches required doubling every 4-5 months

In summary

Shared data, information and computation by geographically dispersed communities.
Current model: ad-hoc client-server

[Diagram showing relationships between Scientist, Experiment, Storage, Analysis, Computing, and HPC]
The vision: computation & information utility
e-Science Examples

- Bioinformatics/Functional genomics
- Collaborative Engineering
- Medical/Healthcare informatics
- Earth Observation Systems (flood monitoring)
- TeleMicroscopy
- Virtual Observatories
- Robotic Telescopes
- Particle Physics at the LHC
  - EU DataGrid: particle physics, biology & medical imaging, Earth observation
  - GridPP, ScotGrid
  - AstroGrid
Multi-disciplinary Simulations

- Lift Capabilities
- Drag Capabilities
- Responsiveness
- Deflection capabilities
- Responsiveness
- Braking performance
- Steering capabilities
- Traction
- Dampening capabilities
- Thrust performance
- Reverse Thrust performance
- Responsiveness
- Fuel Consumption

Whole system simulations are produced by coupling all of the sub-system simulations.
Many aircraft, flight paths, airport operations, and the environment are combined to get a virtual national airspace.
Global in-flight engine diagnostics

airline

ground station

in-flight data

global network eg SITA

DS&S Engine Health Center

data centre

internet, e-mail, pager

maintenance centre

Distributed Aircraft Maintenance Environment: Universities of Leeds, Oxford, Sheffield & York
LHC computing

Online System

~PByte/sec

Tier 0

CERN Computer Centre >20,000 PCs

Tier 1

US Regional Centre
Italian Regional Centre
French Regional Centre
RAL Regional Centre

~Gbit/sec or Air Freight

Tier 2

Tier 2 Centre
~1000 PCs

ScotGRID++

~Gbit/sec

~1000 PCs

Tier 3

Institute
~200 PCs

Physicists work on analysis “channels”
each institute has ~10 physicists working on
one or more channels
data for these channels is cached by the
institute server

Tier 4

Workstations

100 - 1000 Mbit/sec

Physical data cache

~100 MByte/sec

~100 MByte/sec

assumes PC = ~ 25 SpecInt95

one bunch crossing per 25 ns
100 triggers per second
each event is ~1 MByte

~Gbit/sec

physicists work on analysis “channels”
each institute has ~10 physicists working on
one or more channels
data for these channels is cached by the
institute server

Tier 2 Centre
~200 PCs

Centre
~100 PCs

Centre
~100 PCs
Emergency response teams

- bring sensors, data, simulations and experts together
  - wildfire: predict movement of fire & direct fire-fighters
  - also earthquakes, peacekeeping forces, battlefields, …
Earth observation

- ENVISAT
  - € 3.5 billion
  - 400 terabytes/year
  - 700 users

- ground deformation prior to a volcano
To achieve e-Science we need...
To achieve e-Science we need...
The Grid

• Computing cycles, data storage, bandwidth and facilities viewed as commodities.
  – like the electricity grid

• Software and hardware infrastructure to support model of computation and information utilities on demand.
  – middleware

• An emergent infrastructure
  – delivering dependable, pervasive and uniform access to a set of globally distributed, dynamic and heterogeneous resources.
The Grid

Supporting computations with differing characteristics:

- **High throughput**
  - Unbounded, robust, scalable, use otherwise idle machines

- **On demand**
  - Short-term bounded, hard timing constraints

- **Data intensive**
  - Computed, measured, stored, recalled, e.g. particle physics

- **Distributed supercomputing**
  - Classical CPU and memory intensive

- **Collaborative**
  - Mainly in support of hhi, e.g. access grid
Grid Data

- Generated by sensor
- From a database
- Computed on request
- Measured on request
Grid Services

• Deliver bandwidth, data, computation cycles

• Architecture and basic services for
  – authentication
  – authorisation
  – resource scheduling and coscheduling
  – monitoring, accounting and payment
  – protection and security
  – quality of service guarantees
  – secondary storage
  – directories
  – interoperability
  – fault-tolerance
  – reliability
But...has the emperor got any clothes on?

- Maybe *string vest* and *pants*:
  - semantic web
    - machine understandable information on the web
  - virtual organisations
  - pervasive computing
    "... a billion people interacting with a million e-businesses with a trillion intelligent devices interconnected"
  
Lou Gerstner, IBM (2000)
National and International Activities

• USA
  – NASA Information Power Grid
  – DOE Science Grid
  – NSF National Virtual Observatory etc.

• UK e-Science Programme

• Japan – Grid Data Farm, ITBL

• Netherlands – VLAM, PolderGrid

• Germany – UNICORE, Grid proposal

• France – Grid funding approved

• Italy – INFN Grid

• Eire – Grid proposals

• Switzerland - Network/Grid proposal

• Hungary – DemoGrid, Grid proposal

• ...

UK e-science centres

AccessGrid
always-on video
walls
National e-Science Centre

- Edinburgh + Glasgow Universities
  - Physics & Astronomy × 2
  - Informatics, Computing Science
  - EPCC
- £6M EPSRC/DTI + £2M SHEFC over 3 years

- e-Science Institute
  - visitors, workshops, co-ordination, outreach
- middleware development
  - 50 : 50 industry : academia
- UK representation at GGF
- coordinate regional centres
Generic Grid Middleware

- All e-Science Centres will donate resources to form a UK ‘national’ Grid
- All Centres will run same Grid Software
  - based on Globus*, SRB and Condor
- Work with Global Grid Forum (GGF) and major computing companies.

*Globus – project to develop open architecture, open source Grid software, Globus toolkit 2.0 now available.
How can you participate?

• Informal collaborations

• Formal collaborative LINK funded projects

• Attend seminars at eSI

• Suggest visitors for eSI

• Suggest themes and topics for eSI.
eSI forthcoming events

Friday, March 15
Monday, March 18
Tuesday, March 19
Thursday, March 21
Monday, April 8
Monday, April 15
Wednesday, April 17
Monday, April 29
Sunday, July 21

Blue Gene
IWOX: Introductory Workshop on XML Schema
AWOX: Advanced Workshop on XML: XML Schema, Web Services and Tools
GOGO: Getting OGSA Going 1
Software Management for Grid Projects
DAI: Database Access and Integration
DBFT Meeting
The Information Grid
GGF5/ HPDC11
Watch the web pages.... www.nesc.ac.uk

Thank you!