From web to grid: The future of scientific computing?

Neasan O'Neill
GridPP Dissemination Officer
In The Beginning …

CERN
27 Km Around
Answering the big question

• What actually happened at the big bang?

Answering other questions

• How does gravity work?
• How do particles have mass?
• Where is the rest of the universe?
A Data Challenge

Combined they will create 15PB of Data every year

That’s the equivalent of 22 Trillion Sheets of A4

More than any single, current, system can handle
Solution needs to be:

- Able to handle massive amounts of data
- Able to process large computing jobs
- Relatively inexpensive
- Simple to use
- Accessible 24/7
- Easily upgraded
Super Computers?

So we just build them bigger?

But:
- Expensive
- Inaccessible
- Easily outdated
The Internet?

The physical network which connects the worlds computers allowing them to communicate.
Tools built upon the Internet

- Worldwide Web
- File Sharing networks
- BOINC e.g. SETI@home
The World Wide Web

The Web specifically designed by scientists at CERN to help with their work
Build a new tool

- The computers in the institutions are already connected
- They already share files

How about sharing everything?
• Always on
• As much or as little as you need on tap
• Where/how power is generated is irrelevant to the end user
• Just plug in and go
A Computing Grid

- Always on
- As much or as little as you need on tap
- Where/how computing power is generated is irrelevant to the end user
- Just plug in and go
Distributed computing has been available to scientists for some time but:

• The use of different sites has to be negotiated by each scientist individually.
• They need a separate account on each system.
• Jobs have to be submitted and results collected back by hand.

Distributed computing can mean the user has a lot of work to do to get any work done.
Middleware lets users simply submit jobs to the Grid without having to know where the data is or where the jobs will run. The software can run the job where the data is, or move the data to where there is CPU power available.

Using the Grid and middleware, all the user has to do is submit a job and pick up the results.
Operating System - Scientific Linux CERN edition (SLC5)

Middlewares:

- **gLite** - Developed during the Enabling Grids for E-sciencE projects, used across most of Europe, Australasia and Southern and Central America.

- **ARC** - Advanced Resource Connector, run by the Nordic countries (and glasgow).

- **UNICORE** - UNiform Interface to COmputing Resources, grid middleware for supercomputers

- **Globus Toolkit** - Used by the Open Science Grid in America
Worldwide LHC Computing Grid

130 computing centres
34 countries
100,000 CPUs
More than the LHC

SuperB is a Super Flavour Factory, this means that they will be looking for new physics relating to the changes in the properties of elementary particles like quarks. It will be built outside Rome.

MICE is the foundations for a neutrino factory. The theory behind creating a steady beam of muons (which decay neutrinos) is quite well understood but controlling various parameters of the muons isn't.

T2K, in Japan, studies neutrino oscillations by generating a beam of neutrinos and directing it towards the Super-Kamiokande detector 300Km away.
European Grid Initiative

10000 users
317 sites
175 VOs
52 countries

243,020 CPUs
61PB tape
40PB disk
15m jobs/month
- **WISDOM Challenges:**
  - Avian Flu
    - 100 years work done in 4 weeks
  - Malaria
    - 50% of computing power provided by GridPP

- **Digital Libraries**
  - Using the Grid to preserve cultural heritage by digitising and storing the information. The distributed storage makes access easier.

- **ASTRA**
  - Archaeological and musical project
    - Modelling ancient instruments, such as the Epigonion, Salpinx and Kithara to make it easier for researchers and musicians to study long dead instruments and music

- **Collaboration with CityDance**
  - A Washington DC based dance troupe used a technique for converting seismic activity to sound to create music for a piece they were working on
    - The technique called volcano sonification is usually used to predict earthquakes and is very CPU intensive so the Grid is used to do the number crunching
GridPP is a collaboration of Particle Physicists and Computing Scientists.

Based at 20 UK universities, and CERN, they are building the UK arm of the wLCG.
- 20 Institutes
- 100+ Individuals
- 10,000+TB of storage
- Equivalent of ~24,000 desktop computers
The e-ScienceTalk project communicates the success stories of e-Infrastructures to policy makers, general public, scientists and students.
http://www.gridpp.ac.uk