

GRID TALK....



...by Vijay Sahota

Outline...

- What's the difference between Grids & Cluster
 - Current trends in computing
 - What relevant skill will I gain?
 - What does MDX have?
 - What is the NGS?
 - Applications on the NGS
-

Definition...

“Grid computing is applying the resources of many computers in a network to solve a single problem at the same time - usually to a scientific or technical problem that requires a great number of computer processing cycles or access to large amounts of data.”

HPC/ Super Computer race...

- Super computers are super in every aspect
- Very expensive to run 2.5 + MW
- China has just recently won the CROWN
- SUPER COMPUTING IS MAINSTREAM
- YOU MAY BE USING ONE SOON!



▣ Oil Exploration

▣ Bioengineering and Drug

▣ Construction Engineering

▣ Meteorological



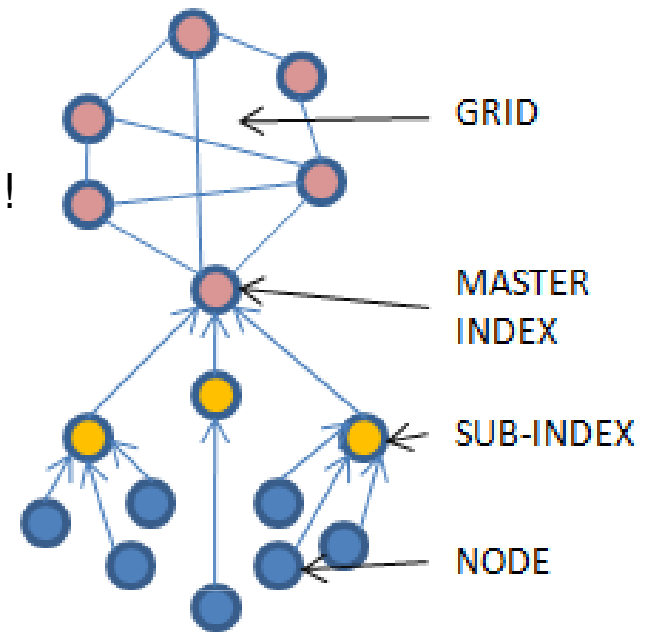
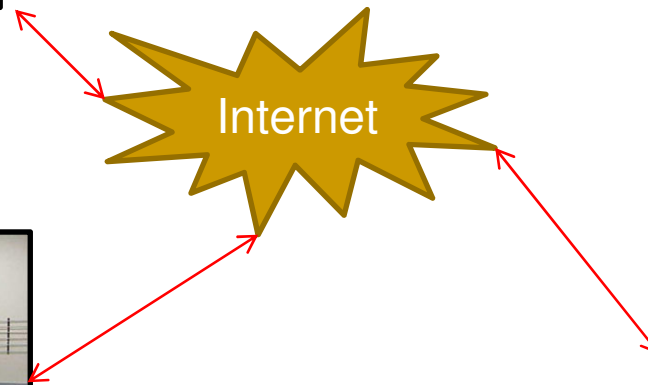
Clustering...

- Basically a collection of cheap... standard computers networked together to perform a common task.
- There are basically 2 kind of computing....
- HPC & HTC!
- Respectively one mainly aims to solve problems as quickly as possible while the other aims to process as much data as possible.
- The latter being more relative to grid computing.

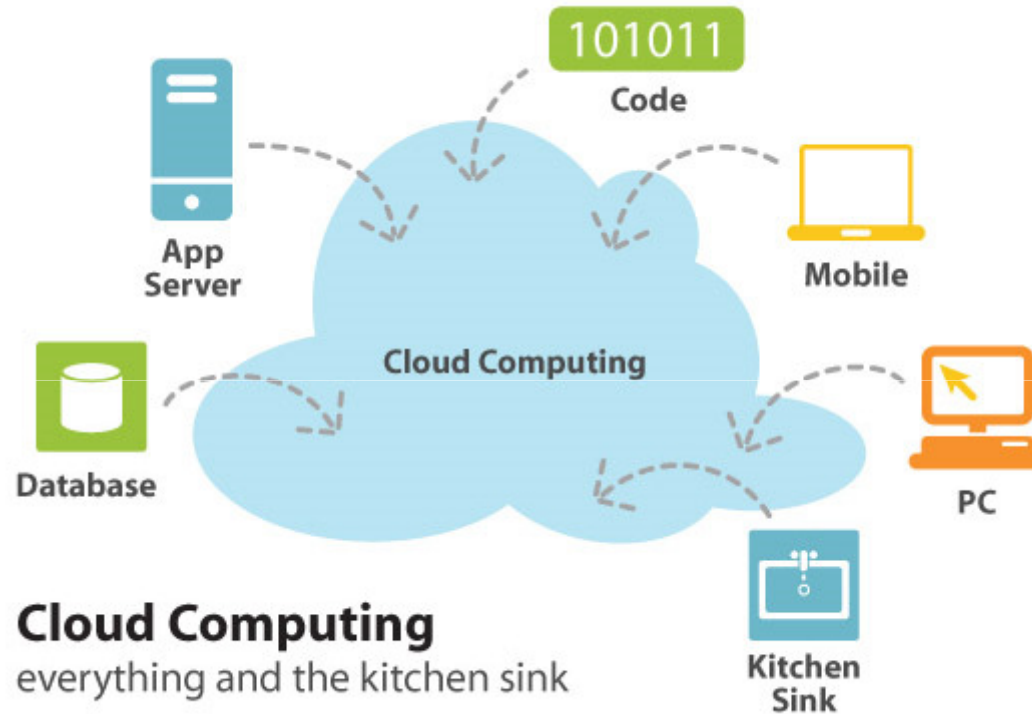


Grid...

Clusters daisy chained via the internet for HTC!

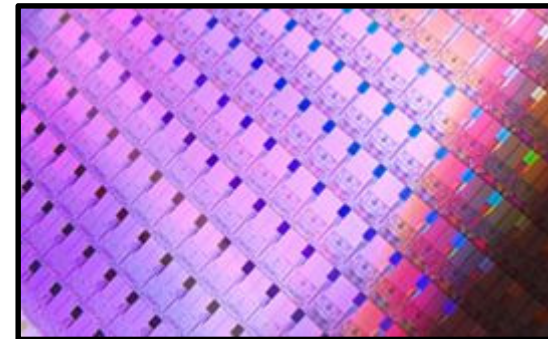
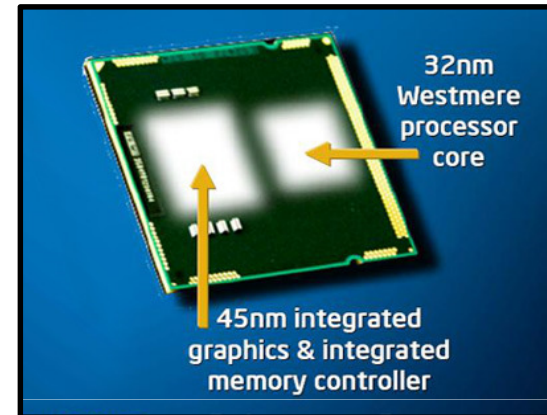


Cloud Computing...



Parallelisation is Key part...

- we have reached maximum clock speeds
 - we are increasing core counts
 - we are integrating GPUs into CPUs
 - we also have Intel MIC
-
- We will see a shift to exploit this tech
 - GGPU and legacy x86 code will merge
-
- Any high performance software will be implemented in parallel code!



Skills to be gained...

If you start using any form of distributed computing, you will gain key skill in the areas of.....

- Computer hardware and computer networking skills
- Concepts of distributed systems
- Concepts of distributed software
- Handling & processing large data set
- Experience in running large simulations

This is standard what isn't is what follows!

Inherent skills ...

Better Impact Factor!

- Better social/ networking skills
- Leader/ management skills
- Better **CV** outlook
- Willingness to learn
- **Better equipped to answer interview questions.**

SAVING MONEY IS KING!

What is a **Bio-information** ???

- More a computer Scientist
- With a bit of a biologist



Bio-informatics Applications...

- **Pattern Finding** – Interpreting DNA & lab results
 - **Data mining/ processing** – Discover/ & Relate data
 - **Image processing** – Reconstruction (& pattern)
 - **Modelling** – Molecular & folding
-

Trends in computation usage...

- Mass Storage
 - Mass processing
 - Visualisation
 - Statistical Analysis
 - Data mining
 - Pattern matching
 - Modelling & Simulation
-

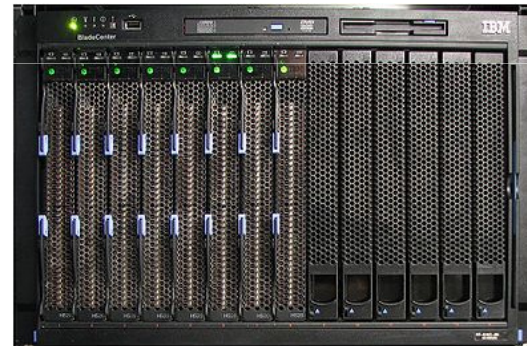
What do we have MDX...

Currently 2 resources, 1 Linux Condor pool cluster, and access to the NGS.

Linux Cluster is actually a test bed to entice new user, currently it only have Matlab installed on it.

Just email me for a user account.

Tutorial will be online soon,



Condor
High Throughput Computing



NGS

The NGS used to stand for the National Grid Services, but are undergoing a rebranding and can't shake of this term!

It is a free Grid for academic use consisting of many institutes....

**Imperial College
London**



 **University of
Reading**

MANCHESTER
1824



Brunel
UNIVERSITY
WEST LONDON



Science & Technology Facilities Council
Rutherford Appleton Laboratory

Royal Holloway
University of London

CARDIFF
UNIVERSITY
PRIFYSGOL
CAERDYDD



LANCASTER
UNIVERSITY
University
of Southampton





Because the NGS is an existing professional, entity – an executive decision has been made to off load all the management side of Grid computing at Middlesex to them.

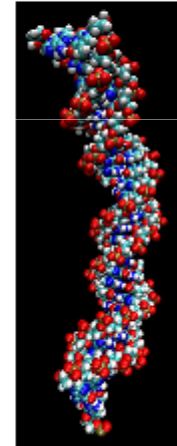
This means that there is **always some one to support** your queries.

Also while we are formulating our resources at Middlesex, **we can use the existing facilities** already up and running provided by the NGS members.

Applications...

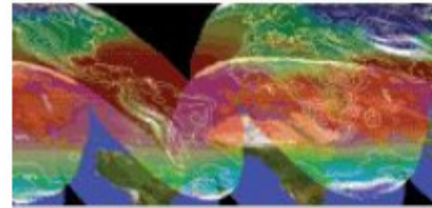
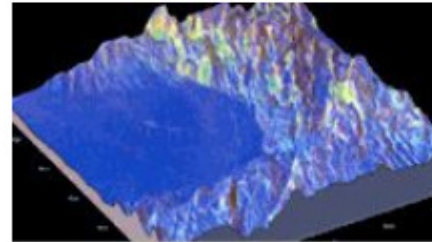
There are many applications in use, I however will go thorough just a few...

- **MrBayes** Bayesian estimation of phylogeny
- **Beast** Bayesian MCMC analysis of molecular sequences
- **NCBI BLAST Toolbox**
- **NCBI BLAST** - Rapid searching of nucleotide and protein databases
- **mpiBlast** - Parallelization of NCBI BLAST (deprecated)
- **Emboss** - Suite of bioinformatics applications e.g. for sequence analysis, enzyme kinetics
- **EXONERATE** - Pairwise sequence comparison
- **FASTA** - Sequence homology search
- **GROMACS** - Molecular dynamics
- **NAMD** - Molecular dynamics
- **SIESTA** - Electronic structure calculations

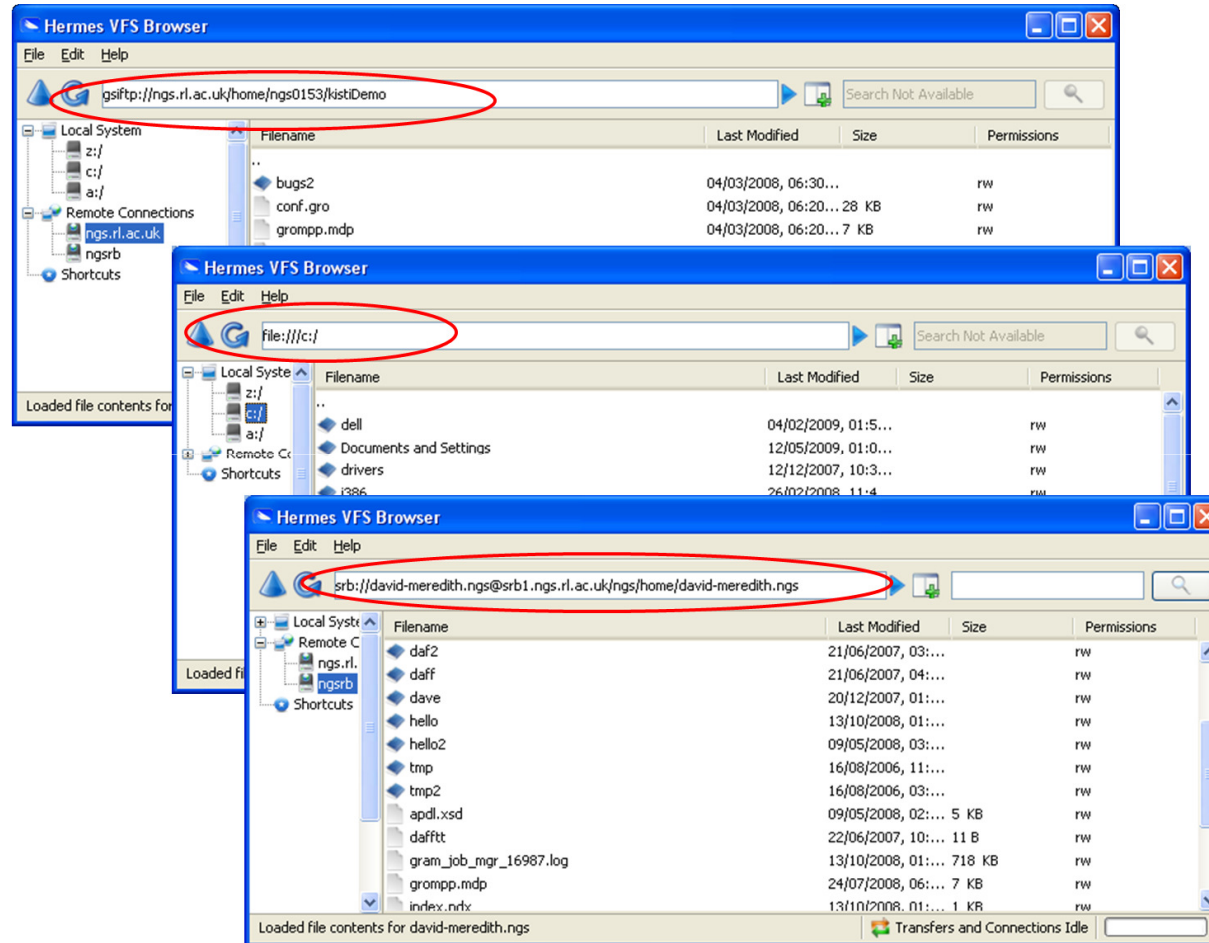


Others...


- **GNU Plot**
- **IDL**
- **Octave** - (a Matlab-alike language)
- **Matlab Compiler Runtime**
- **MGL Tools**
- **POV-Ray**
- **R**
- **Root**
- **Sabre**
- **SciPy**
- **Weka**




Hermes...



NGS Portal...



JSDL Application Repository / Job Submission Portal [\(JSR-168 Compliant\)](#)



Science & Technology
Facilities Council

My Grid Identity: *None* [Help / Get Started](#) | [Logout](#)

- [Authenticate](#)
- [Applications](#)
- [Job_Categories](#)
- [Browse_Host](#)
- [Data_Transfer](#)
- [SrbMdasEnv](#)
- [Admin](#)
- [Info](#)

My Job:

PageLayout
 JsdILayout

- [Detail](#)
- [Candidate_Hosts](#)
- [Description](#)
- [Executable](#)
- [Arguments](#)
- [Environment](#)
- [File_Systems](#)
- [DataStaging](#)
- [Files/Links](#)
- [<JSDL/>](#)
- [Submit/Run](#)

Applications Repository

My Job : *JobProfile name* [Save](#) [SaveAsNew](#) [New](#)

▶ Load an application from the table below to update My Job (Active Job).
▶ Permission may be required to run selected jobs - please contact [NGS helpdesk](#) for access.

List Applications

In Category: [\(Create/Edit\)](#) **Engineering**
With Status: **all**
Search / List Applications:

Details	Application Name	Version	Job Name/Detail	Modified	Delete	MyJob
<input type="button" value="+"/>	Abaqus	6.6-1	Abaqus Flexing wire	May 12, 2009	<input type="checkbox"/>	load
<input type="button" value="+"/>	Abaqus	6.9	Abaqus Flexing wire	Apr 28, 2010	<input type="checkbox"/>	load

Page: 1 of 1 (2 NGS Applications Found in Category 'Engineering')

[|<](#) [<](#) [>](#) [|>](#)

NGS - National Grid Service

- [Authenticate](#)
- [Applications](#)
- [Job Categories](#)
- [Browse Host](#)
- [Data Transfer](#)
- [SrbMdasEnv](#)
- [Admin](#)
- [Info](#)

My Job:

- PageLayout
- JsdlLayout
- [Detail](#)
- [Candidate Hosts](#)
- [Description](#)
- [Executable](#)
- [Arguments](#)
- [Environment](#)
- [File Systems](#)
- [DataStaging](#)
- [Files/Links](#)
- [<JSDL/>](#)
- [Submit/Run](#)

Submit My Job

My Job : *Abaqus Flexing wire*

[Save](#) [SaveAsNew](#) [New](#)

Review and submit My Job (the active job) to the selected candidate host.

My Job Summary

Application Name: (Browse / Select New Application)	Abaqus	Edit
Application Version:	6.6-1	Edit
Job Name:	Abaqus Flexing wire	Edit
Job Category:	Engineering	Edit
Selected Execute / Submission Endpoint:	ngs.rl.ac.uk:2119/lsf	Edit
Candidates:	<Candidate Hosts> <input type="button" value="Update"/>	Edit
Port:	2119	Edit
Path (JobManager):	/lsf	Edit
Executable:	/usr/ngs/ABAQUS_6_6_1	Edit
Standard Input File:		Edit
Standard Output File:	StdOut.txt	Edit
Standard Error File:	StdErr.txt	Edit
Working Directory:		Edit
Job Status:	UNSUBMITTED	<Check Job Status
JobHandle:		<Kill Job Process <Force Kill Job
Process Count:	1	Edit
Job Type:	single	Edit
Node Count:		Edit
Min Mem (MB):		Edit
Max Mem (MB):		Edit
Arguments:	"job=flexwire" "input=std_inst"	Edit

Job Description: [Edit](#)

```
Abaqus v6.6-1
-----
template configured by A.Maniopoulou (HPCSG,RAL,STFC)
-----

This is an simple example-job of a flexing 1D wire. It is a serial job, although Abaqus has MPI like capabilities.

The input files required for this examaple can be retrieved and staged into your home (or working) directory from ngs.rl.ac.uk:/apps/abaqus/examples/std_inst.inp This template is setup to stage the example input files from this directory. Before running this example, change and/or create the working directory on the "Detail" page.

To now run this example job (or another job based on this template) go to the "Submit/Run" page and check the box next to "Ok to overwrite the job status when re-submitting the job" and "Stage all data when submitting job" then click the "Submit My Job" button. The status of the job should appear in red. First it will say submitted. To update the status, click on "Check Job Status". When completed, go to the "Data Transfer" or "Browse Host" page and download your output.

For more information about Abaqus on the NGS (RAL node) refer to:
```

Some still need CMD...

```
Type = "Job";
JobType = "normal";
Executable = "/usr/ngs/FLUENT";
Environment = { "FLUENT_LICENSE_FILE=license-server-info" };
Arguments = "solver solver-options";
CpuNumber = 1;
StdOutput = "std.out";
StdError = "std.err";
Myproxyserver= "myproxy.ngs.ac.uk";
InputSandbox = {"input"};
InputSandboxBaseURI = "gsiftp://ngsui03.ngs.ac.uk:2811/home/ngsxxxx/fluents";
OutputSandbox = {"std.err", "output", "std.out"};
OutputSandboxBaseDestURI = "gsiftp://ngsui03.ngs.ac.uk:2811/home/ngsxxxx/fluents";
Requirements = (
    member("NGS-UEE-FLUENT", other.GlueHostApplicationSoftwareRunTimeEnvironment) #
);
ShallowRetryCount = -1;
```

e-Science website...

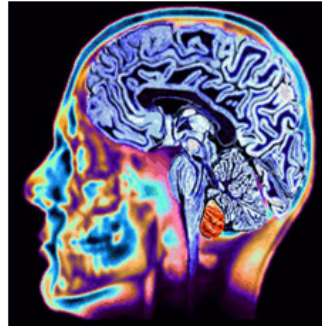
Search for

e-Science
MDX site..

You are here: [Home](#) > [Research & innovation](#) > [Research areas](#) > [Biomedical and Molecular Science](#) > [e-Science centre](#)

E-SCIENCE CENTRE

The e-Science centre is jointly headed by Prof Richard Bayford (HSSC) and Prof Mark Workman (EIS).



Scientific research is increasingly carried out by communities of researchers that span multiple disciplines who exist across many entities that operate in the global arena. International collaborations have given rise to both the availability and production of masses amounts of data. Understanding and processing this data have lead to many research discoveries, shifting future research towards a data driven model. The recent surge in resources such as the masses of available databases, sensor instrumentation, trial data, computational simulations and geographically distributed computational resources have all contributed to the progress towards this data centric model.

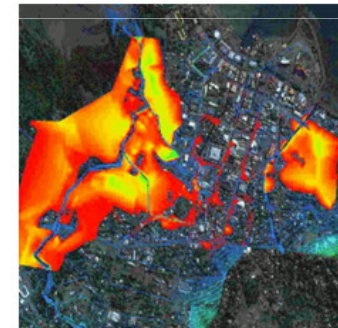
It is our aim to apply such ideals across all fields in our university environment. Bio modelling, informatics, cognitive modelling, e-Social science and environmental simulations are just a sample of applications which can benefit from the expertise presented by the centre of e-Science. e-Science is not only a tool to increases the ability to understand and systematically process all this information, but has forced collaborations to become a global affair. Collaborations to date far exceed the wildest predictions made only several years ago. Such collaborations are formally known as e-Science (EU) or Cyber Infrastructure (USA).

The two key technologies that have had significant impact in e-Science are primarily...

Grid Computing – A distributed computing paradigm which uses the internet to enable the existence of a global heterogeneous computational resource

WEB 2.0 – A set of web technologies which are centred on user generated content. Web 2 enables groups to collaborate in a media rich environment over geographically disperse locations

Considering the significant investment large corporations and countries have poured into investments to build their own grid styled infrastructure, confirms that research through data analysis is the next stage in the evolution for research and discovery. This is a view shared by almost all figure heads in academia.



We are constantly updating our resources and expertise to maintain a complete edge in the ever dynamic and increasing field of e-Science.

Please visit our temporary external home page for more information. [Centre for e-Science, Collaboration and Application](#)

Thank you for your time!
