Collaborative research using eScience infrastructure and high speed networks
Network Enabled Collaborative Science (NECS12)

Paola Grosso (UvA), Inder Monga (Esnet), Scott Lusher (NLeSC), Gera Pronk & Peter Hinrich (SURFnet)
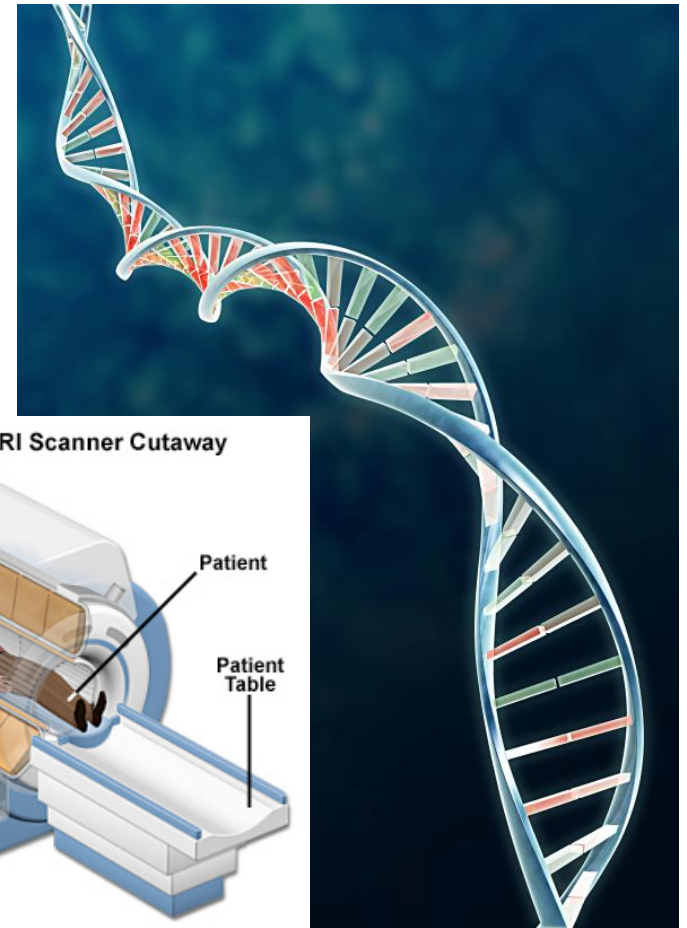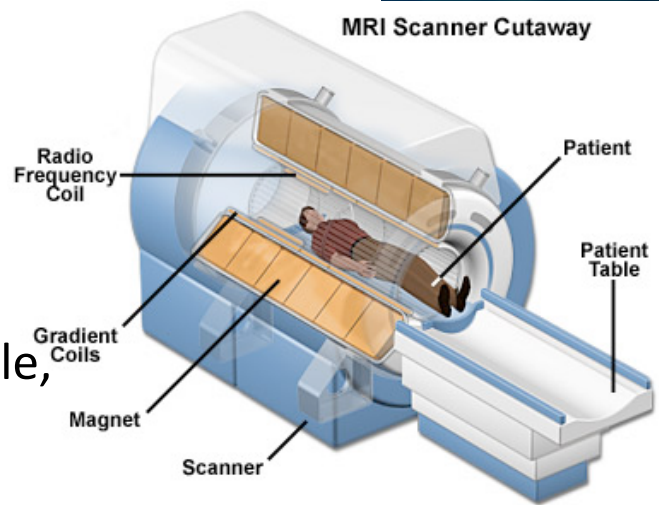
e-Science Infrastructure & high speed networks

science becomes data driven

Instruments, Storage,

Compute are indispensable

resources.

Networks become Programmable,

On-Demand Resources as well

MRI Scanner Cutaway

SURF NET

Bringing together worlds of eScience and networking

Special section in "Future Generation Computer Systems

http://www.journals.elsevier.com/future-generation-computer-systems/calls-for-papers/collaborative-research

Paper submission: October 31 2012

SURF NET

Contributions from various disciplines

- Bioinformatics / Genomics
- Radio Astronomy/eVLBI
- Astronomy
- Chemistry
- Climatology

Computer- / Network Sciences

Programme

http://www.surfnet.nl/en/Hybride_netwerk/SURFlichtpaden/Pages/
CollaborativeresearchusingeScienceinfrastructureandhighspeednetworks.aspx

- Enabling large genomic data transfers using nationwide and international dynamic lightpaths **(Jan Bot)**

- Pilot Abstractions for Compute, Data, and Network **(Mark Santcroos)**

- Verification and User Experience of High Data Rate Bandwidth-on-Demand (**Richard Hughes-Jones**)

- Scientific Workflows using Dynamically Provisioned High-Speed Circuits (**Paul Ruth)**

- Distributed Multi-Model / Multi-Kernel Simulations: A Case Study in Jungle Computing using eScience Infrastructure and High Speed Networks **(Frank Seinstra**)

- Efficient Data Transfer Protocols for Big Data **(Brian Tierney**)

Enabling large genomic data transfers using nation-wide and international dynamic lightpaths

Jan Bot, Migiel de Vos, Sander Boele, Marcel Reinders, Joost Kok

- NGS data explosion. Data transfer is major bottleneck
- Currently Genome of the Netherlands (750 persons)
- Upcoming: RNA seq & methylation data (3000 samples)
- sequencing at external locations (*e.g.* BGI & Complete Genomics
- 10 institutes with BoD connection + Complete genomics/BGI)
- Problems due to firewalls and heterogeneity can be overcome
- Future: more (international) partners; Additional bandwidth; BoD awareness to data transfer tools

Universiteit Leiden

The Delft Bioinformatics Lab

Pilot Abstractions for Compute, Data (and Network?)

Mark Santcroos, Silvia Delgado Olabarriaga,

Daniel S. Katz, Shantenu Jha

- Compute resources heterogeneous, fluctuating capacity
- Decoupling payload management and resource assignment important for efficiently utilizing distributed resources
- P* a model for pilot abstraction
- Pilot-Job is an <u>abstraction</u> that uses a <u>placeholder job</u> as a container for a set of compute tasks
- Abstractions for <u>Data</u> and <u>Compute</u> elements, now also for <u>Network</u> elements
- Dynamic lightpaths offer new degree of freedom in optimizing use of e-Infrastructure

Verification and User Experience of High Data Rate Bandwidth-on-Demand networks

Jimmy Cullen, <u>Richard Hughes-Jones </u>(presenter), Ralph Spencer

- eVLBI enables new science. Targets of opportunity can be observed, latest example: Nova Observed with 8 eVLBI telescopes on 18 Sep 12
- BoD very closely matches the requirements of eVLBI observations
- PCs are suitable for network characteristic verification up to 10 Gbps
- BoD links are stable, reliable and suitable for eVLBI operations
- However, it is important to verify the links before experiments

MANCHESTER 1824
The University of Manchester

GÉANT

Dynamic Network Provisioning for Data Intensive Applications in the Cloud

Paul Ruth, Anirban Mandal, Yufeng Xin, Ilia Baldine, , Jeff Chase,Chris Heermann

- Scientific workflows using networked clouds
- Dynamic Network Circuit Provisioning for Workflows
- IaaS: Virtualization of Compute and Storage Infrastructure & Network Infrastructure
- ORCA is a "wrapper" for off-the-shelf cloud and circuit networks etc., enabling federated orchestration
- ExoGENI: 14 GPO-funded racks, each rack is a networked cloud
- Workflow Management Systems (Pegasus)
- Examples from Chemistry (Solar fuels, Catalysis)

Distributed Multi-Model / Multi-Kernel Simulations: A Case Study in Jungle Computing using eScience Infrastructure and High Speed Networks

Frank Seinstra N. Drost, J. Maassen, M.A.J. van Meersbergen, B. van Werkhoven, H.E. Bal, F.I. Pelupessy, S. Portegies Zwart, M. Kliphuis, H.A. Dijkstra

- Several disciplines require Jungle computing: Distributed, Heterogeneous, Hierarchical (incl. multi- / many-cores)

- Ibis Software Framework Offers functionality to efficiently & transparently implement & run Jungle Computing applications

- Designed for dynamic / hostile environments

- Modular and flexible, Open Source : http://www.cs.vu.nl/ibis/

- Examples: Computational Astrophysics & Climate Modeling

netherlands
eScience center

GLIF Demo: Thursday, 19:00-21:00

Collaboratorium

The Collaboratorium is a joint initiative by SARA and the Netherlands eScience Center (NLeSC). This advanced and very practical visualization room is a key element in collaborative activities, and real-time and improvised experiments in eScience. By visualizing large data sets and by combining multiple data sources, scientific discussions are enabled and new insights may be born.

Interactive Visualization Wall

The Collaboratorium consists of a video wall equipped with 4x2 Full HD screens with a total resolution of 7.680 x 2.160 pixels (16.6 MegaPixel). In combination with the multi-touch overlay this creates a gigantic smart tablet. The interactive room provides excellent opportunities to strengthen (informal) working relationships and technology transfer between science, governments and industry.

Award Nominee

The Collaboratorium is nominated for the AV Awards 2012 in the category 'Collaborative Communications Project of the Year'.

https://www.sara.nl/systems/collaboratorium
http://esciencecenter.com/about-the-center/collaboratorium/

Direct contact:

Paul Wielinga, SARA, email: wielinga@sara.nl
Prof. Henk Dijkstra, IMAU, email: h.a.dijkstra@uu.nl
Dr. René van Schaik, NLeSC, email: r.vanschaik@esciencecenter.nl

Collaborative Analysis of Climate Models using Remote Scientific Visualization

As climate modeling is an international science, a remote collaboration and visualization environment is essential for leading climate researchers worldwide to discuss breakthrough simulation results as soon as they become available.

Intercontinental Collaboration

Climate researchers from IMAU and Los Alamos located in North America, Africa, and Europe discuss results of recent simulations by using a new visualization tool for climate data, combined with video conferencing technologies and using SAGE middleware to share the visual data.

http://www.sara.nl
http://www.imau.nl
http://www.esciencecenter.com

Efficient Data Transfer Protocols for Big Data

Brian Tierney, Eric Pouyoul: Berkeley National Lab / Esnet, Ezra Kissel, Martin Swany: Indiana University

- Motivation:
  - Increasing demand on/capacity of networks
  - End of CPU frequency scaling
  - Where does TCP hit a wall?
- Zero copy TCP & RDMA over Converged Ethernet (RoCE)
- RoCE can work very well over the WAN
  - ➤ 39.2 Gbps with less than 2% CPU load
- Reduced overhead, consistently good performance
- Requires NIC support, clean paths, proper buffer management

Conclusions

- Workflow management and tools are important
- Complex and heterogeneous infrastructure are necessary to several disciplines
- Networks are becoming programmable and on-demand resources as well
- Virtualization of infrastructure
- Decoupling of payload management and resource assignment
- Interdomain issues
- Solutions for TCP overhead at very high speeds

NECS12 social output and input

Very happy with

- 6 speakers
- 22 participants

Sorry for

- 21°C++++ in the room
- too much powerpoint-beamer challenges

Let's keep in **touch**

Footprints

past tweets & pics ->

http://**storify**.com/gera_p/necs12-workshop-at-ieee

future discussions ->

http://www.**linkedin**.com/groups?home=&gid=4664366&trk=anet_ug_hm

and Join us at **NECS13**

NECS12

See you in 2013 ?

✉ **peter.hinrich[at]surfnet.nl**

W **www.surfnet.nl**

☎ **+31 30 2 305 305**

SURF NET