Is the Internet becoming the Computer

www.science.uva.nl/~delaat

Cees de Laat

GigaPort EU

University of Amsterdam
Talk contents

• Just wait 20 minutes
Four LHC Experiments: The Petabyte to Exabyte Challenge

ATLAS, CMS, ALICE, LHCB

Higgs + New particles; Quark-Gluon Plasma; CP Violation

6000+ Physicists & Engineers; 60+ Countries; 250 Institutions

Tens of PB 2008; To 1 EB by ~2015

Hundreds of TFlops To PetaFlops
LHC Data Grid Hierarchy

CERN/Outside Resource Ratio \( \sim 1:2 \)
Tier0/(\(\sum\) Tier1)/(\(\sum\) Tier2) \( \sim 1:1:1 \)

Online System

~100-1500 MBytes/sec

Tier 0 +1

2.5-10 Gbps

Tier 1

~10 Gbps

~2.5-10 Gbps

Tier 2

Tier 3

~2.5-10 Gbps

Referencing Workstations

Tier 4

An Exabyte \( \sim 5-7 \) Years later.

Institutes

2.5-10 Gbps

Tier 2 Center

Center

Center

Center

Institute

Start

IN2P3 Center

RAL Center

INFN Center

FNAL Center

Emerging Vision: A Richly Structured, Global Dynamic System
The term VLBI is easily capable of generating many Gb of data per day. The sensitivity of the VLBI array scales with the data rate and there is a strong push to increase the data rate. Rates of 8Gb/s or more are entirely feasible under development. It is expected that parallel correlation will remain the most efficient approach. Distributed processing may have an application if multi-gigabit data streams will aggregate into larger correlator and the capacity of the final link to the data correlator.

Westerbork Synthesis Radio Telescope - Netherlands
Lambdas as part of instruments

www.lofar.org
OptIPuter Project Goal: Scaling to 100 Million Pixels

• JuxtaView (UIC EVL) for PerspecTile LCD Wall
  – Digital Montage Viewer
  – 8000x3600 Pixel Resolution ~30M Pixels

• Display Is Powered By
  – 16 PCs with Graphics Cards
  – 2 Gigabit Networking per PC

Source: Jason Leigh, EVL, UIC; USGS EROS
Grids

Showed you:

• Computational Grids
  – HEP and LOFAR analysis requires massive CPU capacity

• Data Grid
  – Storing and moving HEP, Bio and Health data sets is major challenge

• Instrumentation Grids
  – Several massive data sources are coming online

• Visualization Grids
  – Data object (TByte sized) inspection, anywhere, anytime
iGrid 2002
September 24-26, 2002, Amsterdam, The Netherlands

• 28 demonstrations from 16 countries: Australia, Canada, CERN, France, Finland, Germany, Greece, Italy, Japan, The Netherlands, Singapore, Spain, Sweden, Taiwan, United Kingdom, United States

• Applications demonstrated: art, bioinformatics, chemistry, cosmology, cultural heritage, education, high-definition media streaming, manufacturing, medicine, neuroscience, physics, tele-science

• Grid technologies demonstrated: Major emphasis on grid middleware, data management grids, data replication grids, visualization grids, data/visualization grids, computational grids, access grids, grid portals

• 25Gb transatlantic bandwidth (100Mb/attendee, 250x iGrid2000!)

www.igrid2002.org

Note: iGrid2005 @ San Diego sept 2005
Internal versus external bandwidth

Mbit/s

Computer busses

LAN

WAN

Problem Solving Environment
Applications and Supporting Tools
Application Development Support

Collective Grid Services
- Brokering
- Global Queuing
- Co-Scheduling
- Data Cataloguing
- Auditing
- Authorization
- Monitoring
- Fault Management

Common Grid Services
- Grid Information Service
- Uniform Resource Access
- Global Event Services
- Uniform Data Access
- Communication Services

Grid Security Infrastructure (authentication, proxy, secure transport)

Communication

Grid access (proxy authentication, authorization, initiation)

Grid task initiation

Local Resources
- Resource Manager
  - CPUs
- Resource Manager
  - Monitors
- Resource Manager
  - On-Line Storage
- Resource Manager
  - Scientific Instruments
- Resource Manager
  - Tertiary Storage
- Resource Manager
  - Highspeed Data Transport
- Resource Manager
  - net QoS

layers of increasing abstraction taxonomy
Grid - a Vision

Researchers perform their activities regardless geographical location, interact with colleagues, share and access data.

The GRID: networked data processing centres and “middleware” software as the “glue” of resources.

Scientific instruments and experiments provide huge amounts of data.

Ref: EDG
The END

(2 more ggf slides)

Partially complete list:
Caas
Chase
Cess
Kess
Case
What is Global Grid Forum?

Specifications and Best Practices

Identify where / what specs are critical

Frameworks / Architectures (how specs fit together)

Research and Technology Directions

Application Requirements

Partnerships

Education

Working Groups

Research Groups and Workshops

Conferences, Tutorials & Partnerships
GGF Structure

GGF Corp.
- Secretariat: Operations
  - Board of Directors (Catlett, Messina)
    - Financial, Legal Responsibility
    - Appoints GGF Chair
  - Conference Mgmt
  - Sponsor Programs
  - Finance/Legal
  - IT (Website, etc.)
  - Staffing & Services

Hold non-exclusive copyright for document series

GGF
- Document and Standards Work
  - Steering Group
    - GGF Management
    - Document Series Review
  - Chair
  - A-Large Committee
  - AD
  - AL
  - Area

GGF Advisory Com.
- Advise on strategic direction, industry, government
- Advise on governance

Grid Research Oversight Council
- RG/WG Workshop series
- Advise on Research Issues

= working group
= research group

November 2003 (catlett@mcs.anl.gov)
www.ggf.org