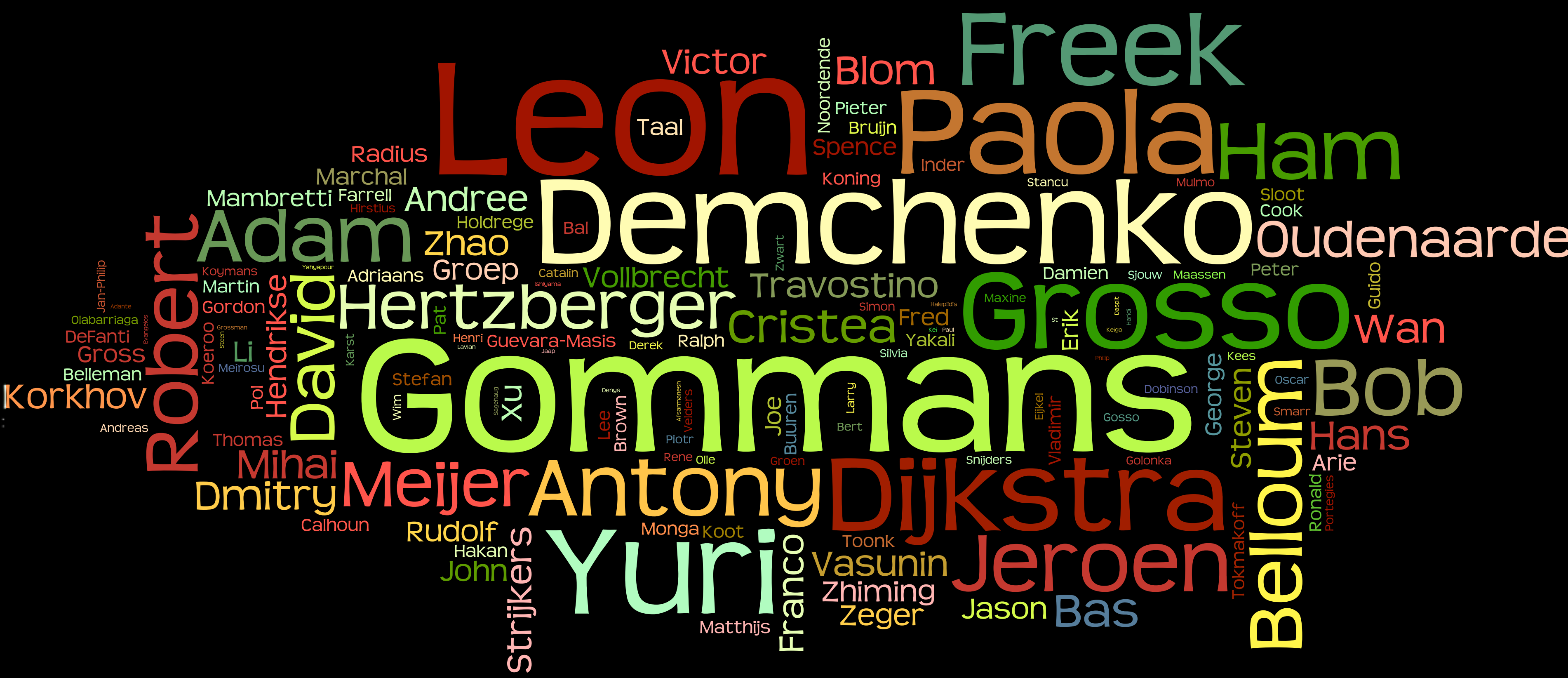
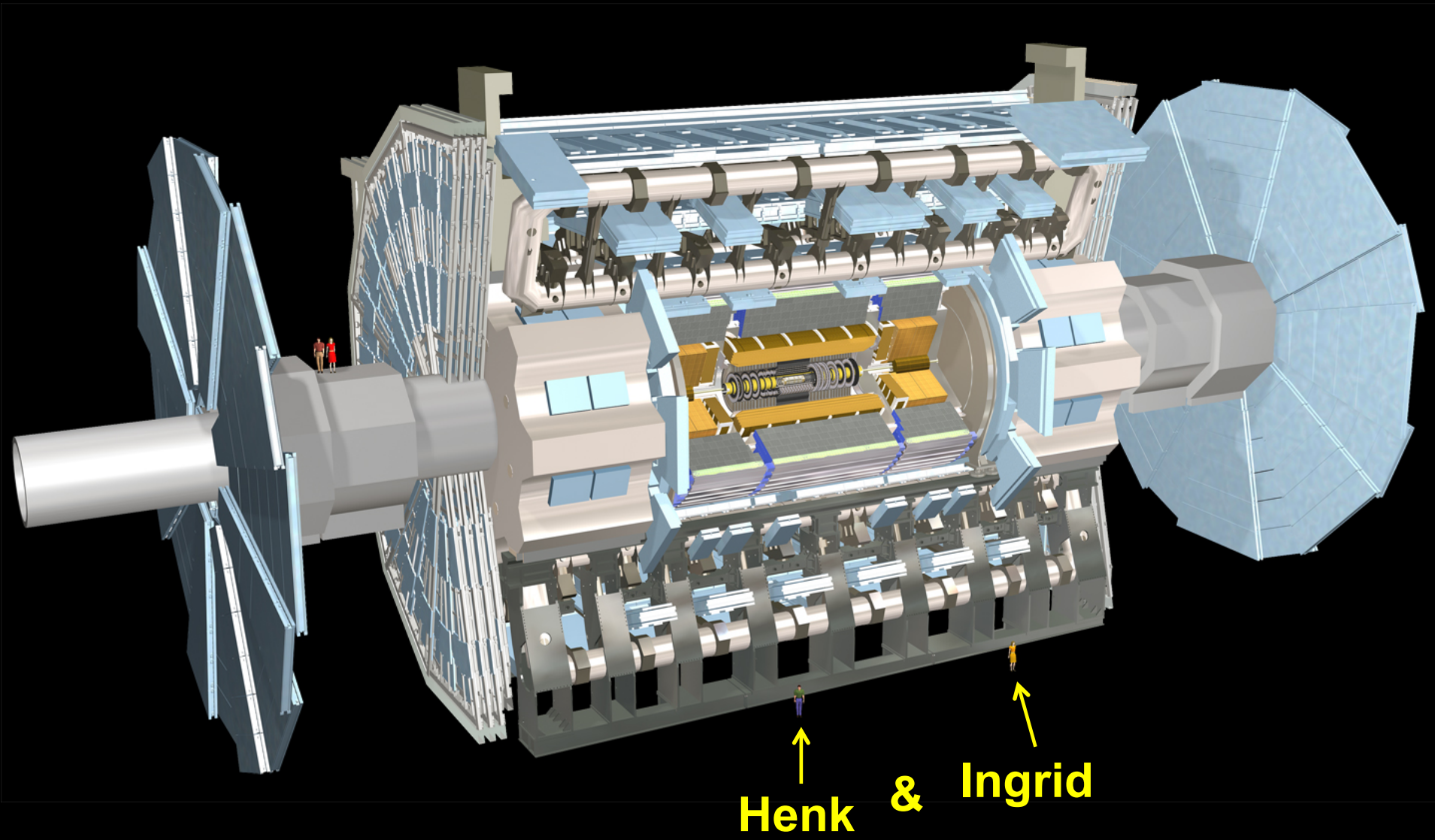
Smart Cyber Infrastructure for Big Data Processing

Cees de Laat



ATLAS detector @ CERN Geneve



15 Pbyte/year

What Happens in an **Internet Minute?**

And **Future Growth** is **Staggering**

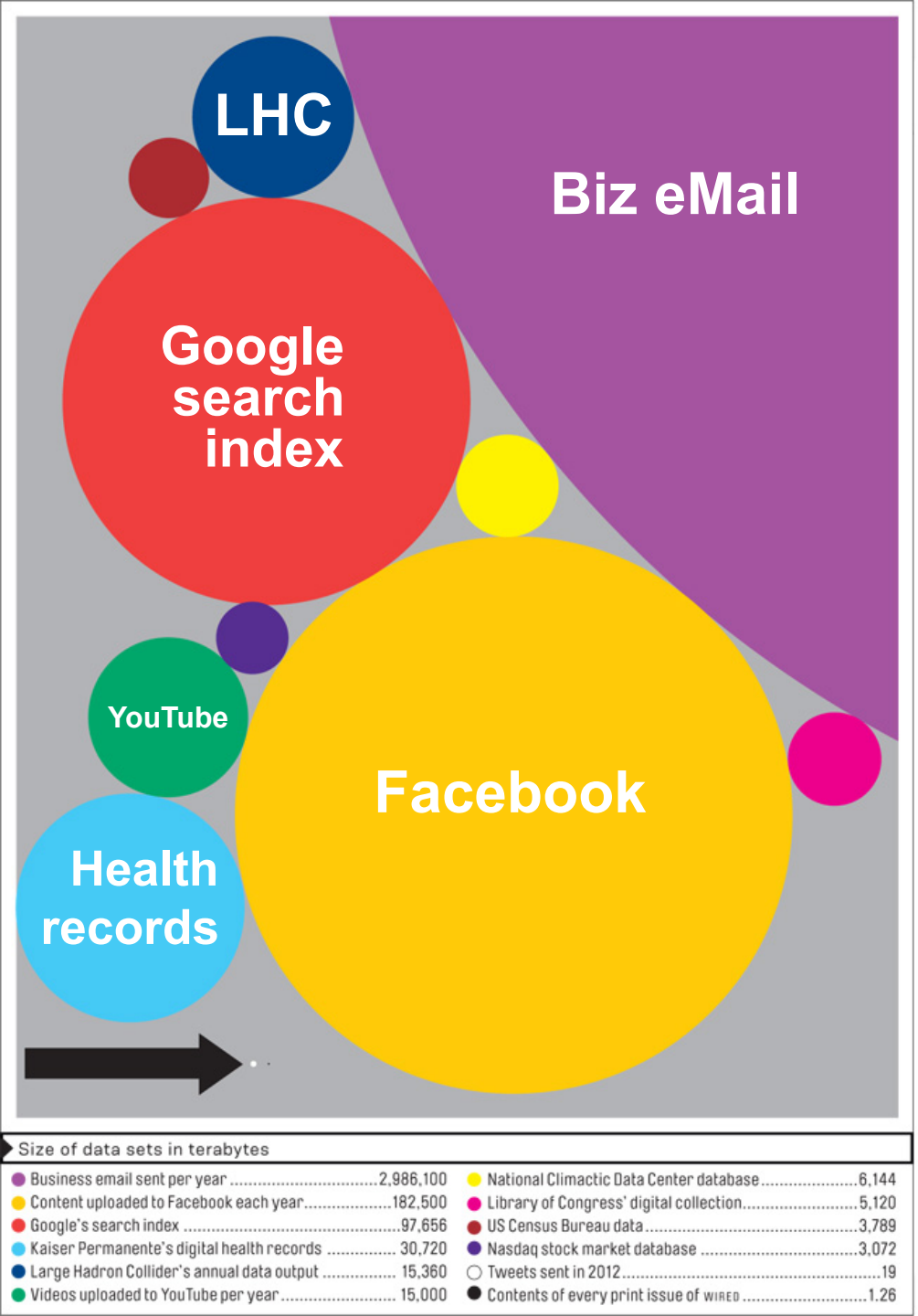
Today, the number of networked devices = the global population

By **2015**, the number of networked devices = 2x the global population

In **2015**, it would take you **5 years**

to view all video crossing IP networks each **second**

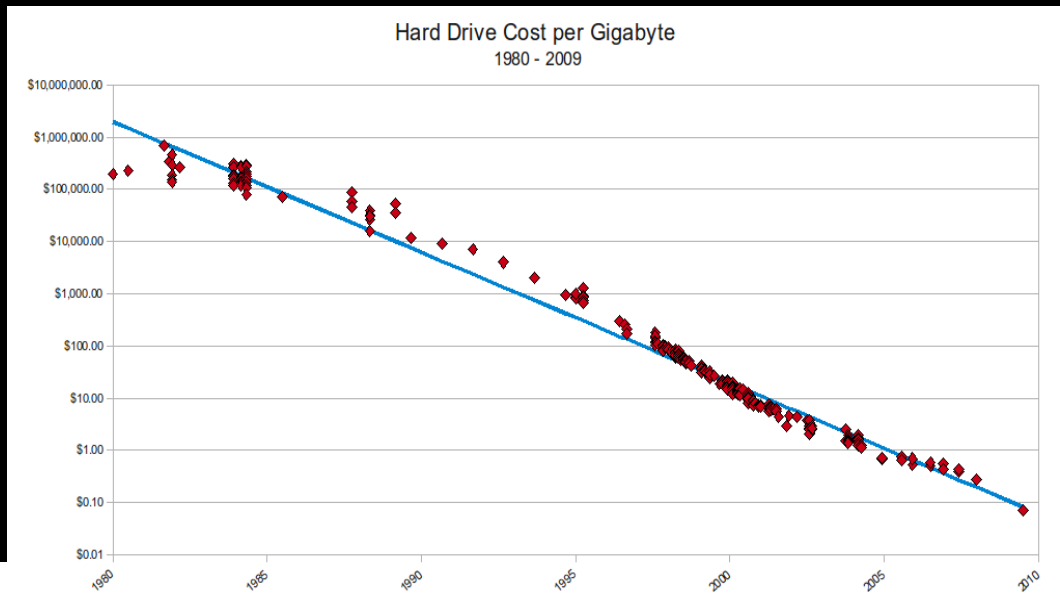
intel

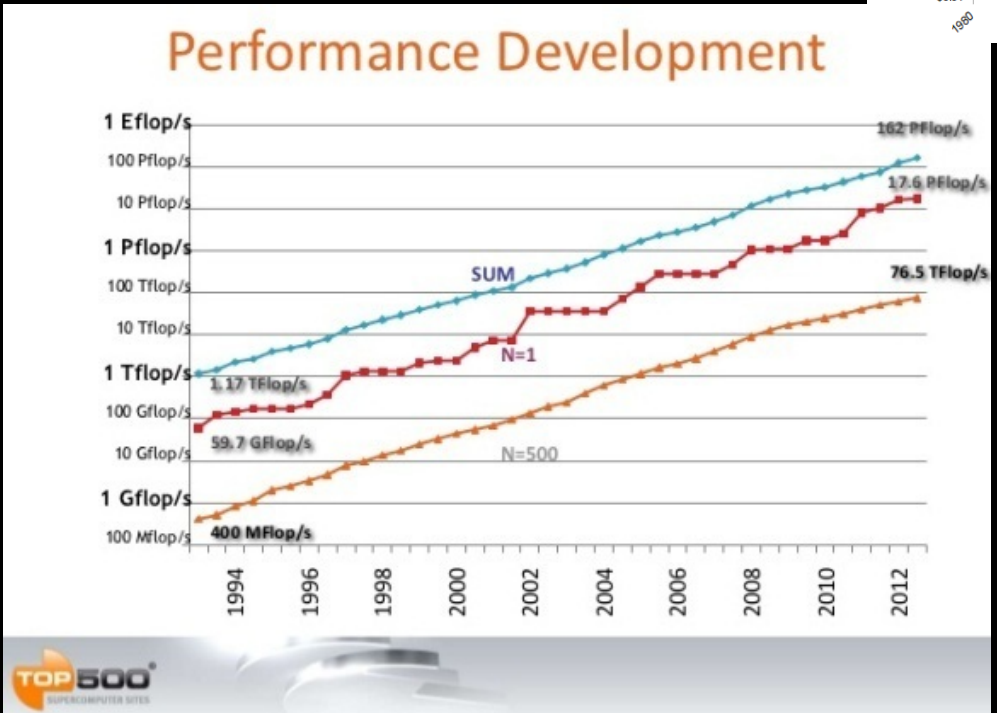


There is always a bigger fish

Computing vs Data

Computing per unit cost has doubled roughly every 18 months.



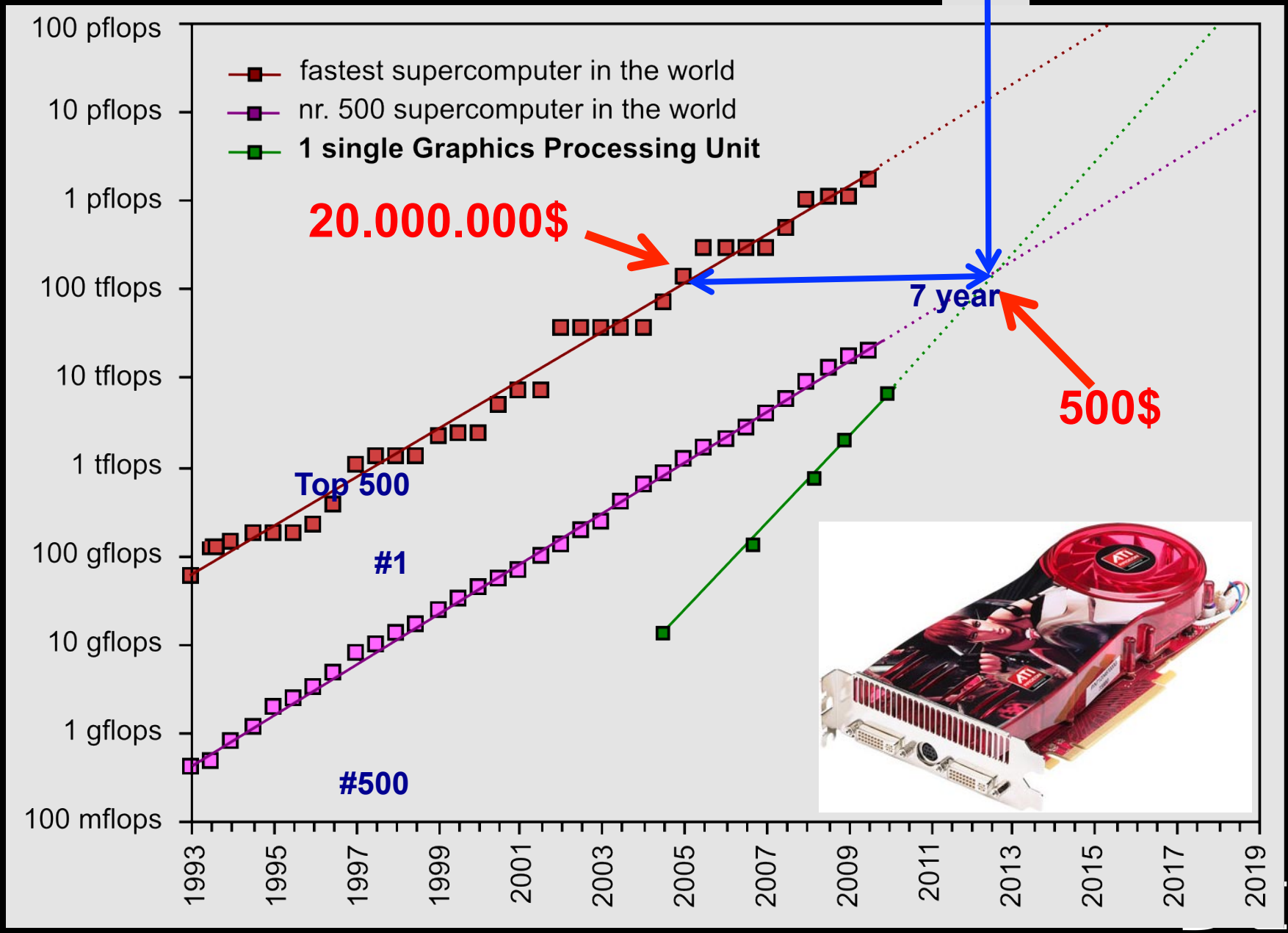


Space per unit cost has doubled roughly every 14 months.

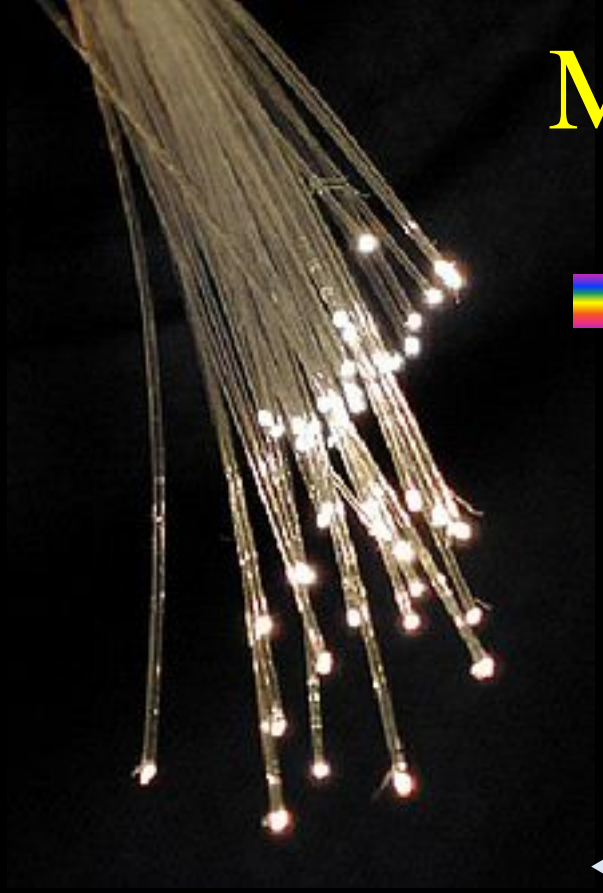
So: data becomes exponentially uncomputable.

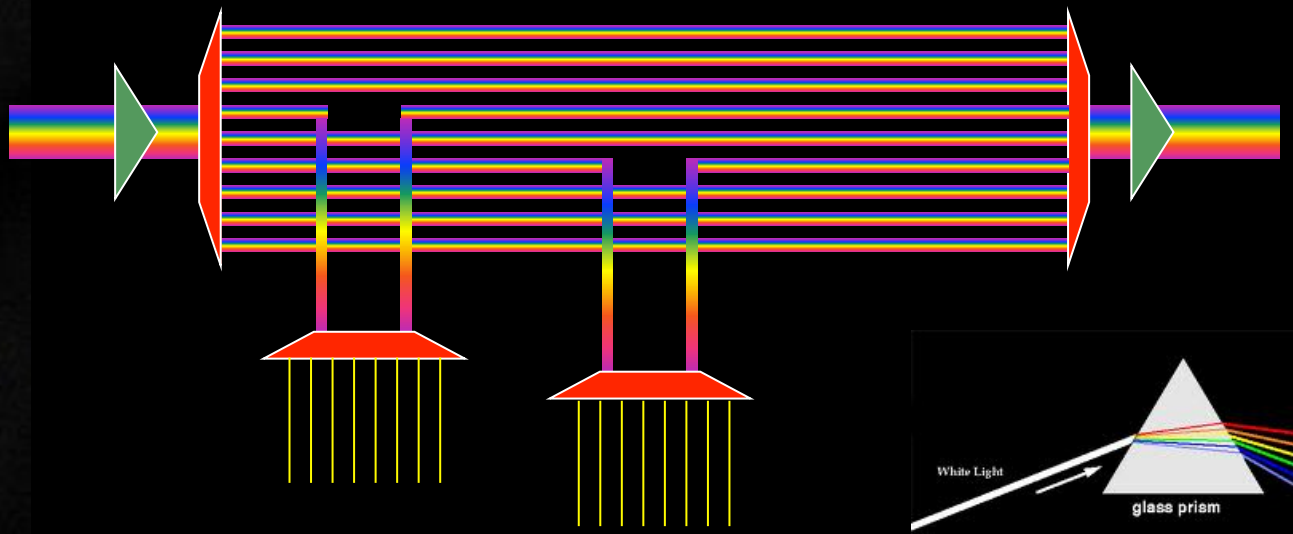
http://www.mkomo.com/cost-per-gigabyte

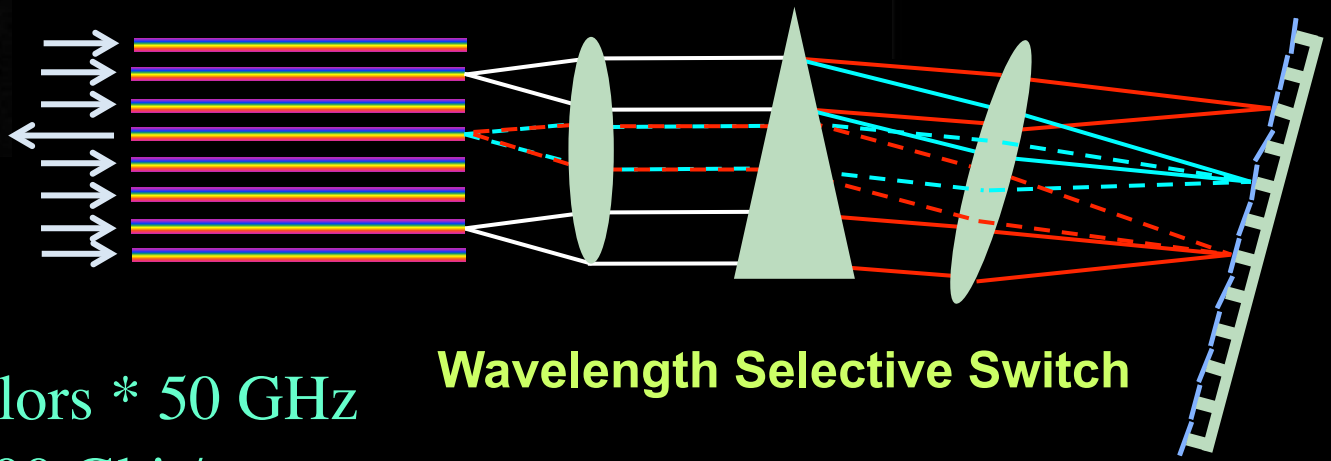
GPU cards are distruptive!



Multiple colors / Fiber







Wavelength Selective Switch

Per fiber: ~ 80-100 colors * 50 GHz

Per color: 10 – 40 – 100 Gbit/s

About 10 Tbit/s per fiber long dist.

BW * Distance ~ $2*10^{17}$ bm/s

New: Hollow Fiber!

➔ less RTT!

Mission

Can we create smart and safe data processing systems that can be tailored to diverse application needs?

- *Capacity*
- *Capability*
- *Security*
- *Sustainability*
- *Resilience*

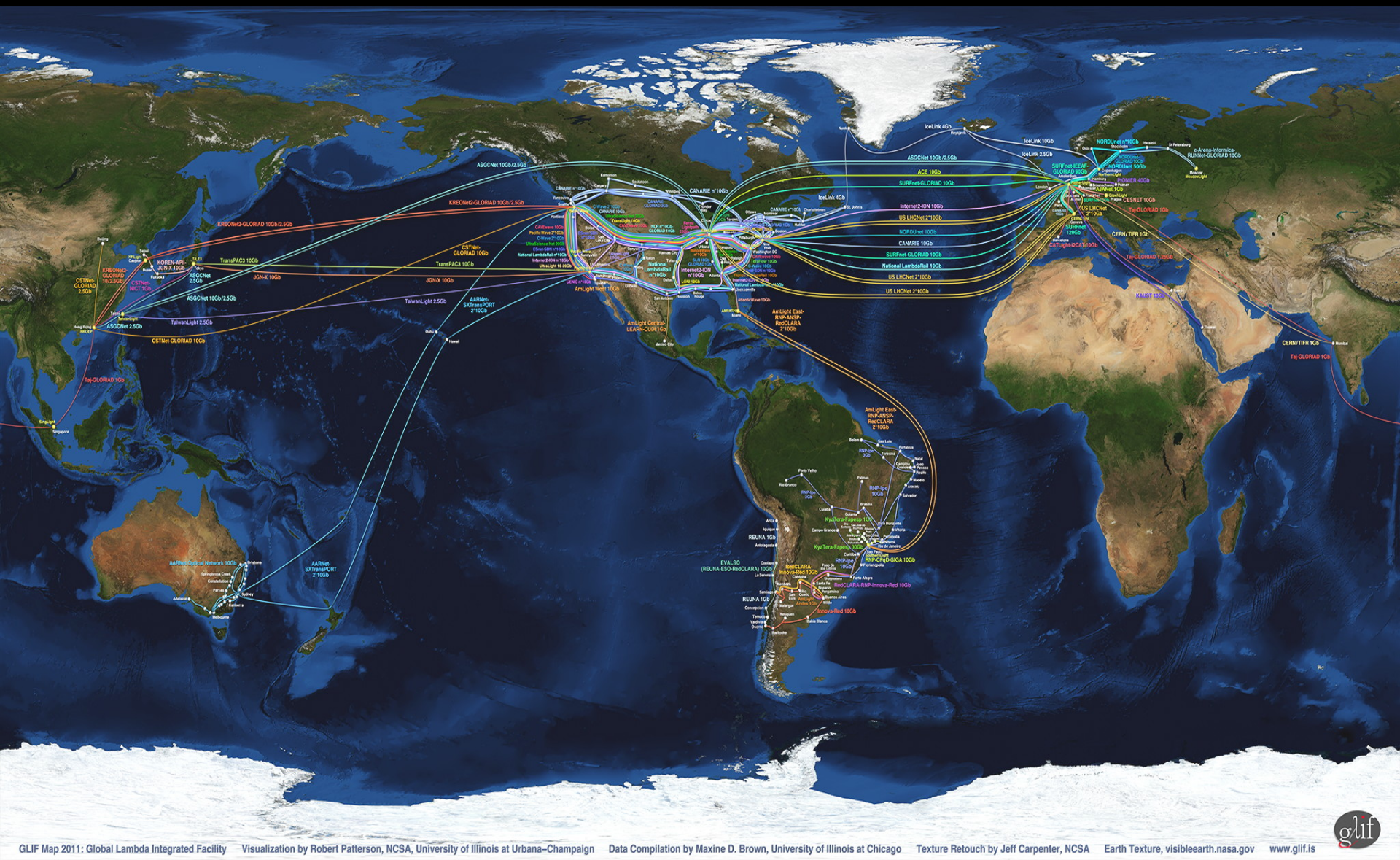
Mission

Can we create smart and safe data processing systems that can be tailored to diverse application needs?

- *Capacity*
  - *Bandwidth on demand, QoS, architectures, photonics, GPU, performance*
- *Capability*
  - *Programmability, virtualization, complexity, semantics, workflows*
- *Security*
  - *Anonymity, integrity of data in distributed data processing*
- *Sustainability*
  - *Greening infrastructure, awareness*
- *Resilience*
  - *Systems under attack, failures, disasters*

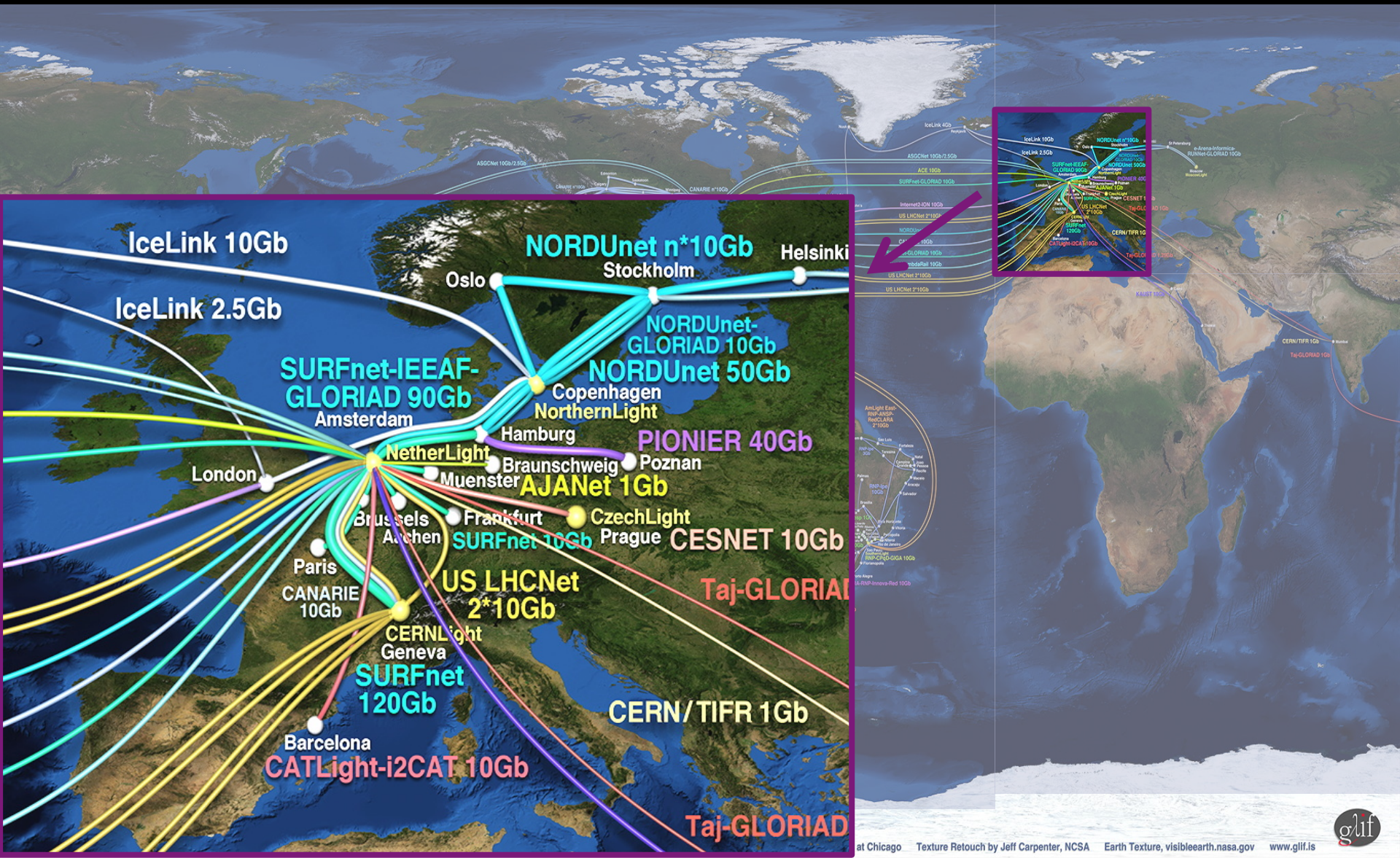
The GLIF – LightPaths around the World

F Dijkstra, J van der Ham, P Grosso, C de Laat, “A path finding implementation for multi-layer networks”, Future Generation Computer Systems 25 (2), 142-146.



Amsterdam is a major hub in The GLIF

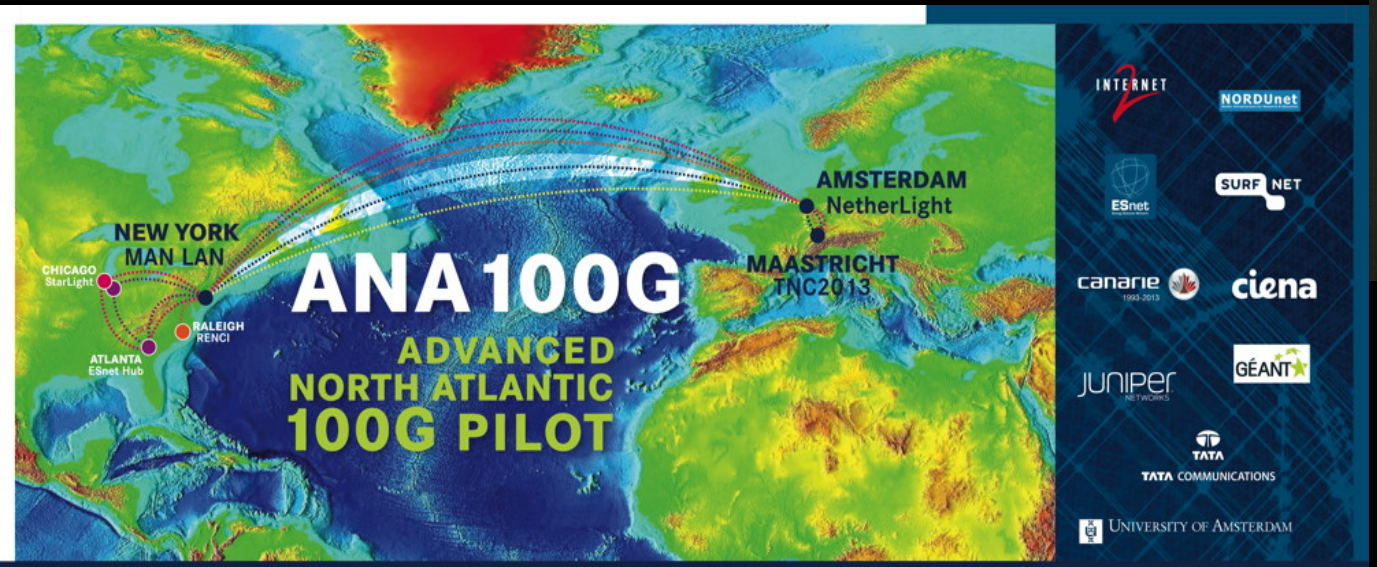
F Dijkstra, J van der Ham, P Grosso, C de Laat, “A path finding implementation for multi-layer networks”, Future Generation Computer Systems 25 (2), 142-146.



ExoGeni @ OpenLab – UvA

http://sne.science.uva.nl/openlab/

Installed and up June 3th 2013



TNC2013 DEMOS JUNE, 2013

DEMO	TITLE	OWNER	AFFILIATION	E-MAIL	A-SIDE	Z-SIDE	PORTS(S) MAN LAN	PORTS(S) TNC2013	DETAILS
1	Big data transfers with multipathing, OpenFlow and MPTCP	Ronald van der Pol	SURFnet	ronald.vanderpol@surfnet.nl	TNC/MECC, Maastricht NL	Chicago, IL	Existing 100G link between internet2 and ESnet	2x40GE (Juniper)+ 2x10GE (OME6500)	In this demonstration we show how multipathing, OpenFlow and Multipath TCP (MPTCP) can help in large file transfers between data centres (Maastricht and Chicago). An OpenFlow application provisions multiple paths between the servers and MPTCP will be used on the servers to simultaneously send traffic across all those paths. This demo uses 2x40G on the transatlantic 100G link. ESnet provides 2x40G between MAN LAN and StarLight, ACE and USLHCnet provide additional 10GEs.
2	Visualize 100G traffic	Inder Monga	ESnet	imonga@es.net					Using an SNMP feed from the Juniper switch at TNC2013,and/or Brocade AL2S node in MANLAN, this demo would visualize the total traffic on the link, of all demos aggregated. The network diagram will show the transatlantic topology and some of the demo topologies.
3	How many modern servers can fill a 100Gbps Transatlantic Circuit?	Inder Monga	ESnet	imonga@es.net	Chicago, Ill	TNC showfloor	1x 100GE	8x 10GE	In this demonstration, we show that with the proper tuning and tool, only 2 hosts on each continent can generate almost 80Gbps of traffic. Each server has 4 10G NICS connected to a 40G virtual circuit, and has iperf3 running to generate traffic. ESnet's new "iperf3" throughput measurement tool, still in 'beta', combines the best features from other tools such as iperf, nuttcp, and netperf. See: https://my.es.net/demos/tnc2013/
4	First European ExoGENI at Work	Jeroen van der Ham	UvA	vdham@uva.nl	RENCI, NC	UvA, Amsterdam, NL	1x 10GE	1x 10GE	The ExoGENI racks at RENCI and UvA will be interconnected over a 10G pipe and be on continuously, showing GENI connectivity between Amsterdam and the rest of the GENI nodes in the USA.
5	Up and down North Atlantic @ 100G	Michael Enrico	DANTE	michael.enrico@dante.net	TNC showfloor	TNC showfloor	1x 100GE	1x 100GE	The DANTE 100GE test set will be placed at the TNC2013 showfloor and connected to the Juniper at 100G. When this demo is running a loop @ MAN LAN's Brocade switch will ensure that the traffic sent to MAN LAN returns to the showfloor. On display is the throughput and RTT (to show the traffic traveled the Atlantic twice)

Connected via
the new 100 Gb/s
transatlantic
To US-GENI

Alien light From idea to realisation!

40Gb/s alien wavelength transmission via a multi-vendor 10Gb/s DWDM infrastructure

NCF

Alien wavelength advantages

- Direct connection of customer equipment[1] → cost savings
- Avoid OEO regeneration → power savings
- Faster time to service[2] → time savings
- Support of different modulation formats[3] → extend network lifetime

Alien wavelength challenges

- Complex end-to-end optical path engineering in terms of linear (i.e. OSNR, dispersion) and non-linear (FWM, SPM, XPM, Raman) transmission effects for different modulation formats.
- Complex interoperability testing.
- End-to-end monitoring, fault isolation and resolution.
- End-to-end service activation.

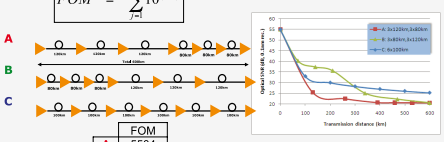
In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

New method to present fiber link quality, FoM (Figure of Merit)

In order to quantify optical link grade, we propose a new method of representing system quality: the FOM (Figure of Merit) for concatenated fiber spans.

$$FOM = \sum_{j=1}^{N} 10^{\left(\frac{L_j}{10}\right)}$$

Lj, span losses in dB
N, number of spans

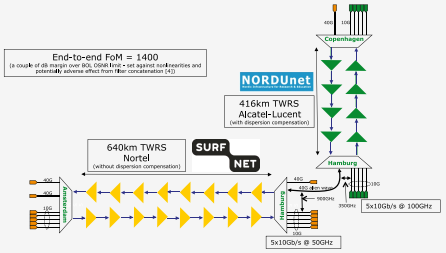


	FOM
A	5504
B	5504
C	1897

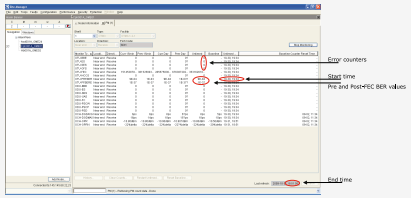
Easy-to-use formula that accurately quantifies transmission system performance

Transmission system setup

JOINT SURFnet/NORDUnet 40Gb/s PM-QPSK alien wavelength DEMONSTRATION.



Test results



Error-free transmission for 23 hours, 17 minutes → BER < 3.0 10^{-16}

Conclusions

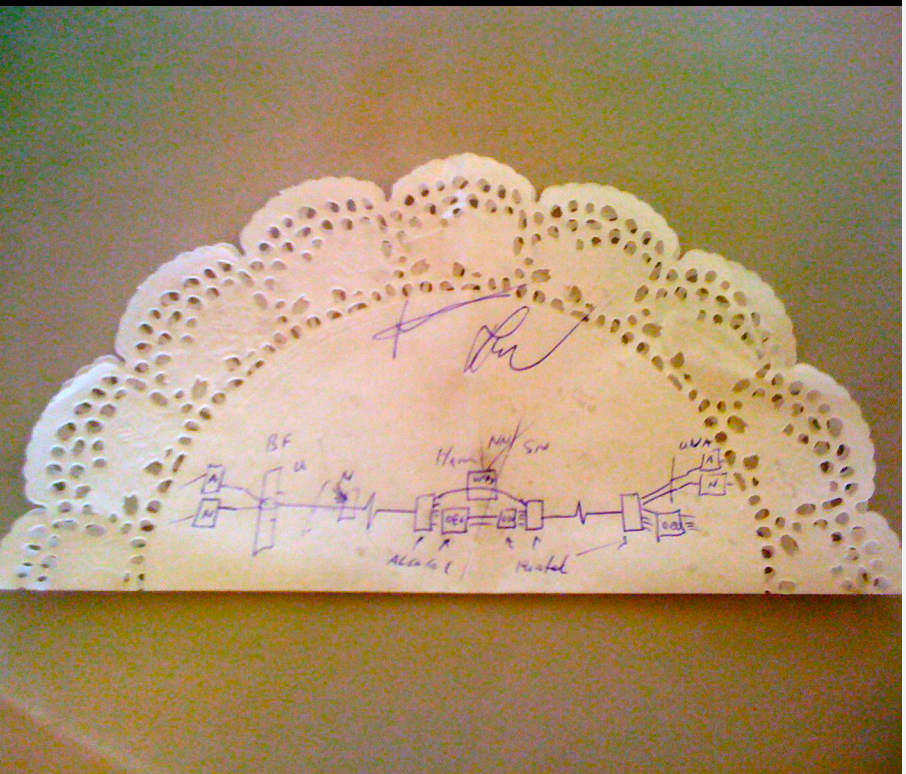
- We have investigated experimentally the all-optical transmission of a 40Gb/s PM-QPSK alien wavelength via a concatenated native and third party DWDM system that both were carrying live 10Gb/s wavelengths.
- The end-to-end transmission system consisted of 1056 km of TWRS (TrueWave Reduced Slope) transmission fiber.
- We demonstrated error-free transmission (i.e. BER below 10-15) during a 23 hour period.
- More detailed system performance analysis will be presented in an upcoming paper.

NORTEL NORDUnet telindus together with ISTY SURF NET

REFERENCES [1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", O. GERSTEL ET AL, OFC'2009 | [2] "AT&T OPTICAL TRANSPORT SERVICES", BARBARA E. SMITH, OFC'09
[3] "OPEX SAVINGS OF ALL-OPTICAL CORE NETWORKS", ANDREW LORD AND CARL ENGINEER, ECOC2009 | [4] NORTEL/SURFNET INTERNAL COMMUNICATION

ACKNOWLEDGEMENTS WE ARE GRATEFUL TO NORDUNET FOR PROVIDING US WITH BANDWIDTH ON THEIR DWDM LINK FOR THIS EXPERIMENT AND ALSO FOR THEIR SUPPORT AND ASSISTANCE DURING THE EXPERIMENTS. WE ALSO ACKNOWLEDGE TELINDUS AND NORTEL FOR THEIR INTEGRATION WORK AND SIMULATION SUPPORT

Alien light From idea to realisation!



40Gb/s alien wavelength transmission via a multi-vendor 10Gb/s DWDM infrastructure

NCF

Alien wavelength advantages

- Direct connection of customer equipment[1] → cost savings
- Avoid OEO regeneration → power savings
- Faster time to service[2] → time savings
- Support of different modulation formats[3] → extend network lifetime

Alien wavelength challenges

- Complex end-to-end optical path engineering in terms of linear (i.e. OSNR, dispersion) and non-linear (FWM, SPM, XPM, Raman) transmission effects for different modulation formats.
- Complex interoperability testing.
- End-to-end monitoring, fault isolation and resolution.
- End-to-end service activation.

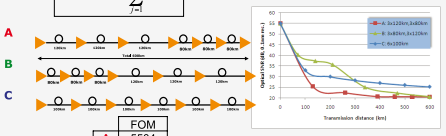
In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

New method to present fiber link quality, FoM (Figure of Merit)

In order to quantify optical link grade, we propose a new method of representing system quality: the FOM (Figure of Merit) for concatenated fiber spans.

$$FOM = \sum_{j=1}^{N} 10^{\left(\frac{L_j}{10}\right)}$$

Lj, span losses in dB
N, number of spans

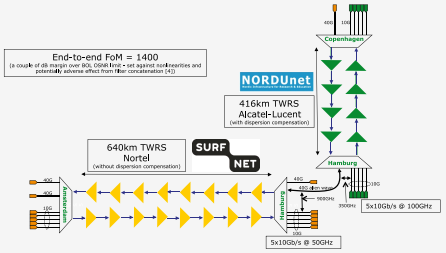


	FOM
A	5504
B	5504
C	1897

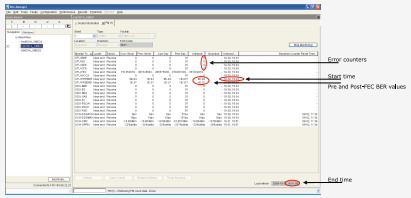
Easy-to-use formula that accurately quantifies transmission system performance

Transmission system setup

JOINT SURFnet/NORDUnet 40Gb/s PM-QPSK alien wavelength DEMONSTRATION.



Test results



Error-free transmission for 23 hours, 17 minutes → BER < 3.0 10^{-16}

Conclusions

- We have investigated experimentally the all-optical transmission of a 40Gb/s PM-QPSK alien wavelength via a concatenated native and third party DWDM system that both were carrying live 10Gb/s wavelengths.
- The end-to-end transmission system consisted of 1056 km of TWRS (TrueWave Reduced Slope) transmission fiber.
- We demonstrated error-free transmission (i.e. BER below 10-15) during a 23 hour period.
- More detailed system performance analysis will be presented in an upcoming paper.

NORTEL NORDUnet telindus together with ISTY SURF NET

REFERENCES [1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", O. GERSTEL ET AL, OFC'2009 | [2] "AT&T OPTICAL TRANSPORT SERVICES", BARBARA E. SMITH, OFC'09
[3] "OPEX SAVINGS OF ALL-OPTICAL CORE NETWORKS", ANDREW LORD AND CARL ENGINEER, ECOC2009 | [4] NORTEL/SURFNET INTERNAL COMMUNICATION

ACKNOWLEDGEMENTS WE ARE GRATEFUL TO NORDUNET FOR PROVIDING US WITH BANDWIDTH ON THEIR DWDM LINK FOR THIS EXPERIMENT AND ALSO FOR THEIR SUPPORT AND ASSISTANCE DURING THE EXPERIMENTS. WE ALSO ACKNOWLEDGE TELINDUS AND NORTEL FOR THEIR INTEGRATION WORK AND SIMULATION SUPPORT

ClearStream @ TNC2011

Setup codename: FlightCees

UvA

iPerf
I7 3.2 GHz Q-core

iPerf
Amd Ph II 3.6 GHz HexC

Mellanox

40G E

Copenhagen

iPerf
2* dual 2.8 GHz Q-core

iPerf

Mellanox

CIENA OME 6500

CERN

Hamburg

CIENA DWDM

Alcatel DWDM

CIENA OME 6500

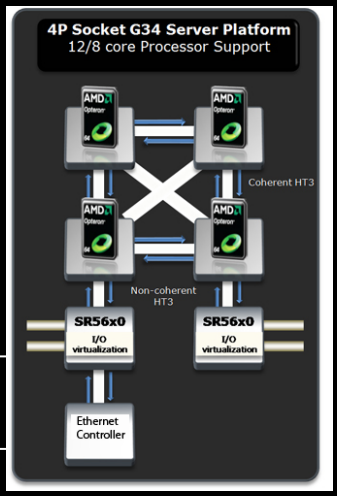
CIENA OME 6500

17 ms RTT

27 ms RTT

Amsterdam – Geneva (CERN) – Copenhagen – 4400 km (2700 km alien light)

CPU Topology benchmark



cpu3

cpu0

Bandwidth (bps)

Bandwidth (bps) for single iperf thread

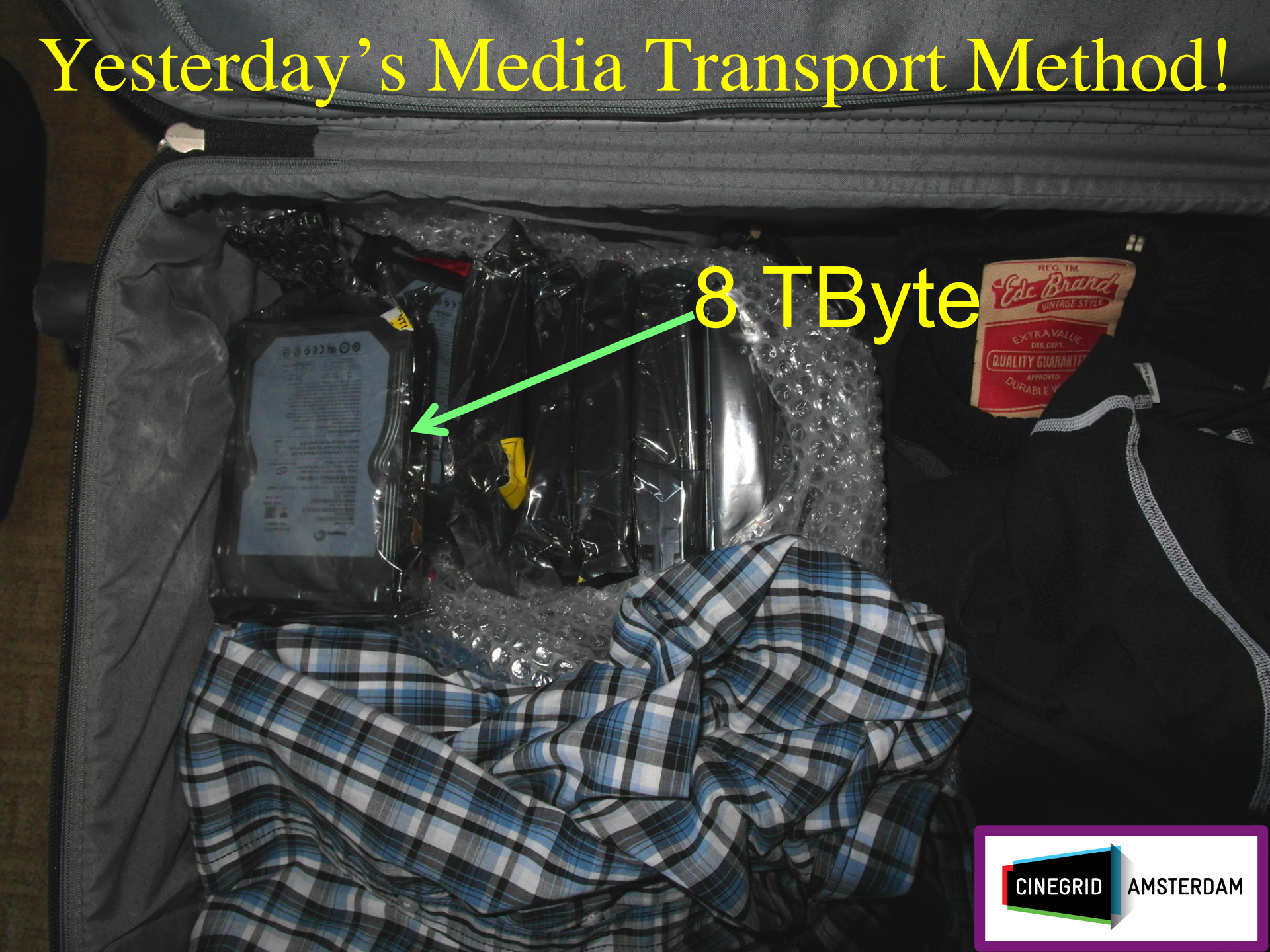
1,8E+10
1,6E+10
1,4E+10
1,2E+10
1E+10
8E+09
6E+09
4E+09
2E+09
0

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47

CoreID

We used numactl to bind iperf to cores

Yesterday's Media Transport Method!



CINEGRID AMSTERDAM

Moving Cinegrid Objects Globally

- **Digital Motion Picture for Audio Post-Production**
  - 1 TV Episode Dubbing Reference ~ 1 GB
  - 1 Theatrical 5.1 Final Mix ~ 8 GB
  - 1 Theatrical Feature Dubbing reference ~ 30 GB
- **Digital Motion Picture Acquisition**
  - 4K RGB x 24 FPS x 10bit color: ~ 48MB/Frame uncompressed *(ideal)*
  - 6:1 ~ 20:1 shooting ratios => 48TB ~ 160TB digital camera originals
- **Digital Dailies**
  - HD compressed MPEG-2 @ 25 ~ 50 Mb/s
- **Digital Post-production and Visual Effects**
  - Gigabytes - Terabytes to Select Sites Depending on Project
- **Digital Motion Picture Distribution**
  - Film Printing in Regions
    - Features ~ 8TB
    - Trailers ~ 200GB
  - Digital Cinema Package to Theatres
    - Features ~ 100 - 300GB per DCP
    - Trailers ~ 2 - 4GB per DCP

CineGrid

CINEGRID AMSTERDAM

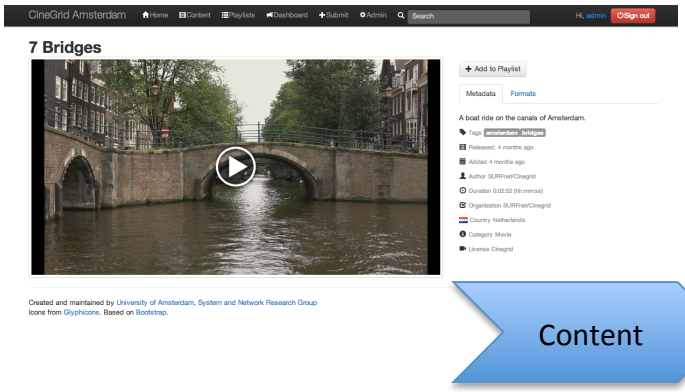
Content Metadata

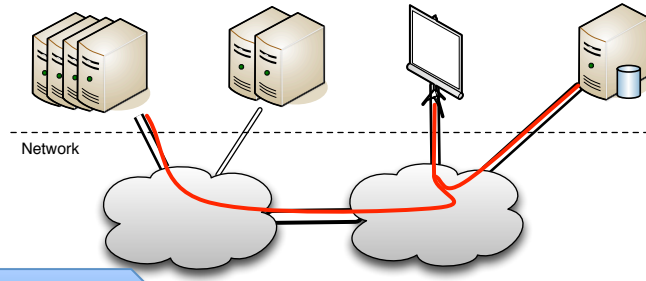
Resource Descriptions

Programmable Infrastructure

CineGrid Portal

Unified orchestration of distributed CineGrid resources

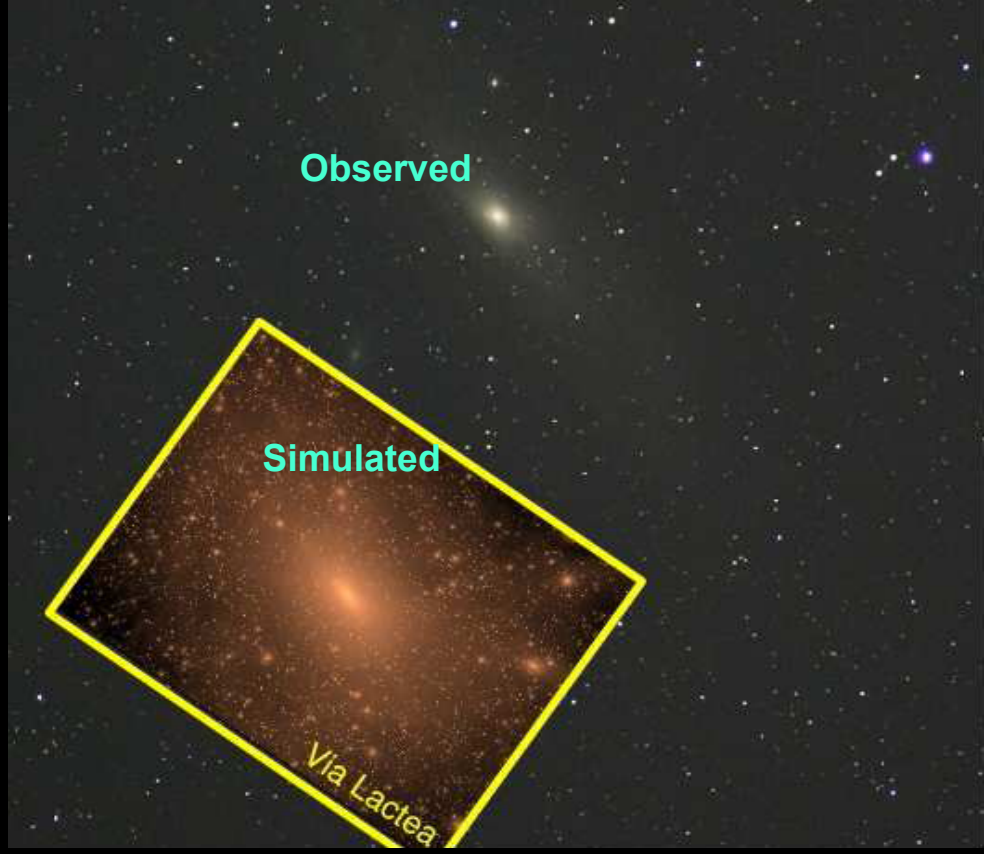




CosmoGrid

Simon Portegies Zwart et al.

- Motivation:
  - **previous simulations found >100 times more substructure than is observed!**
- Simulate large structure formation in the Universe
- Method: Cosmological *N*-body code
- Computation: Intercontinental SuperComputer Grid
- Current (2013) problem:
  - 2 PByte data in Oak Ridge!



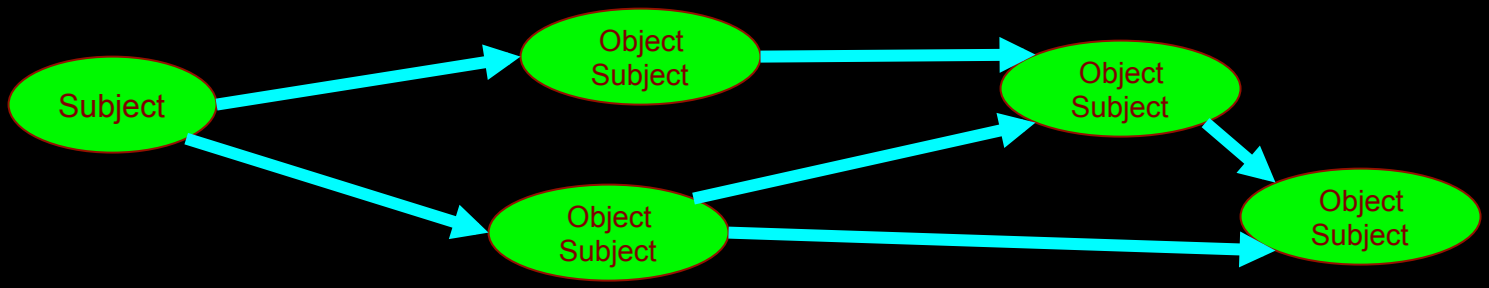
10 Gb/s dedicated network

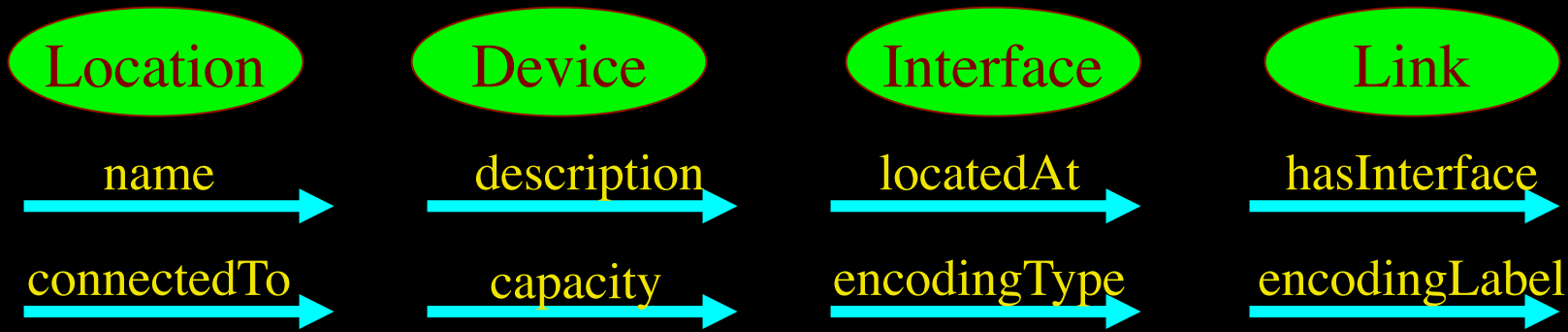
270 ms RTT

We investigate: for complex networks!

LinkedIN for Infrastructure

- From semantic Web / Resource Description Framework.
- The RDF uses XML as an interchange syntax.
- Data is described by triplets (Friend of a Friend):

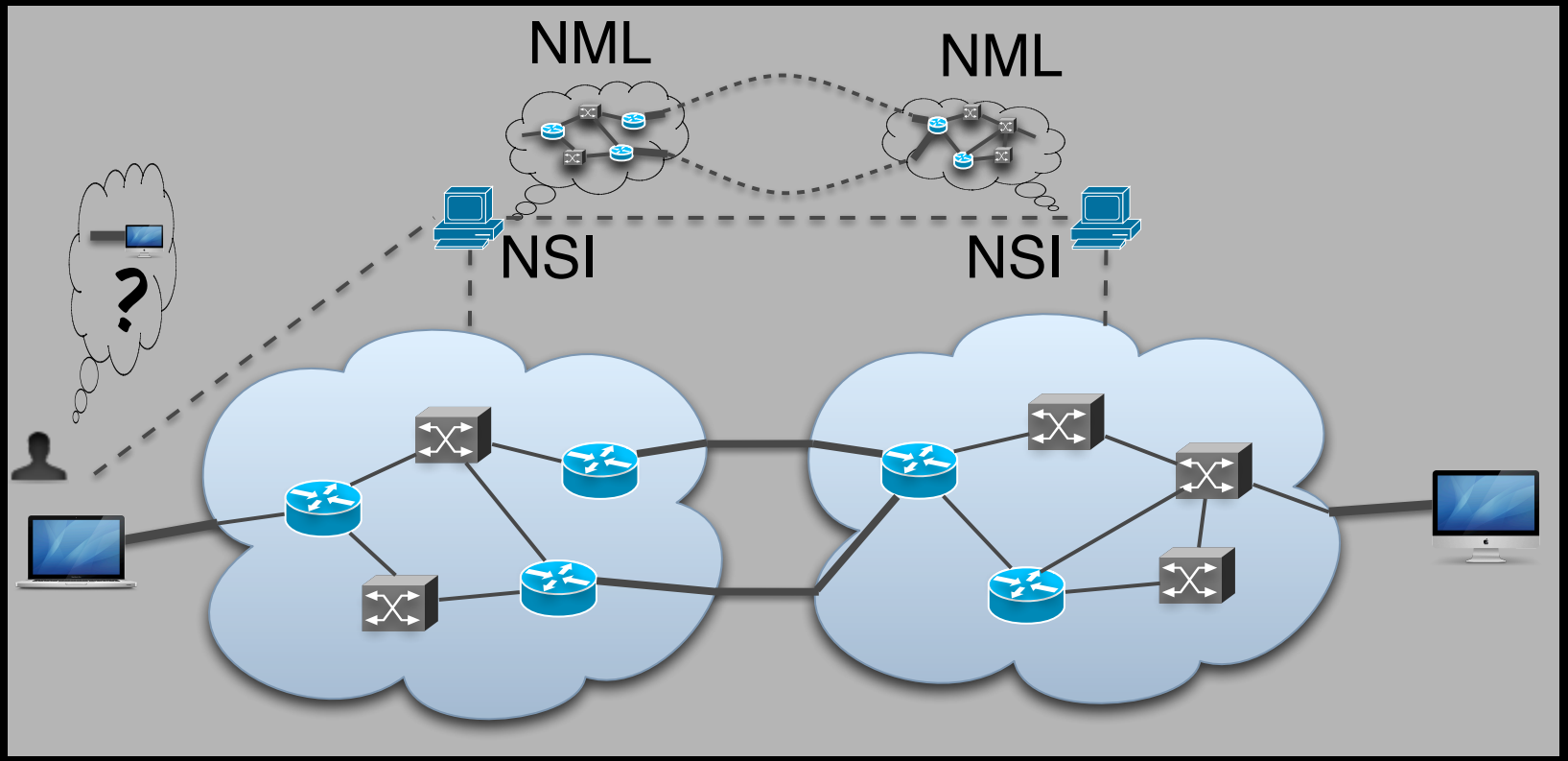




Network Topology Description

Network topology research supporting automatic network provisioning

- Inter-domain networks
- Multiple technologies
- Generalized to ExoGeni & FED4FIRE

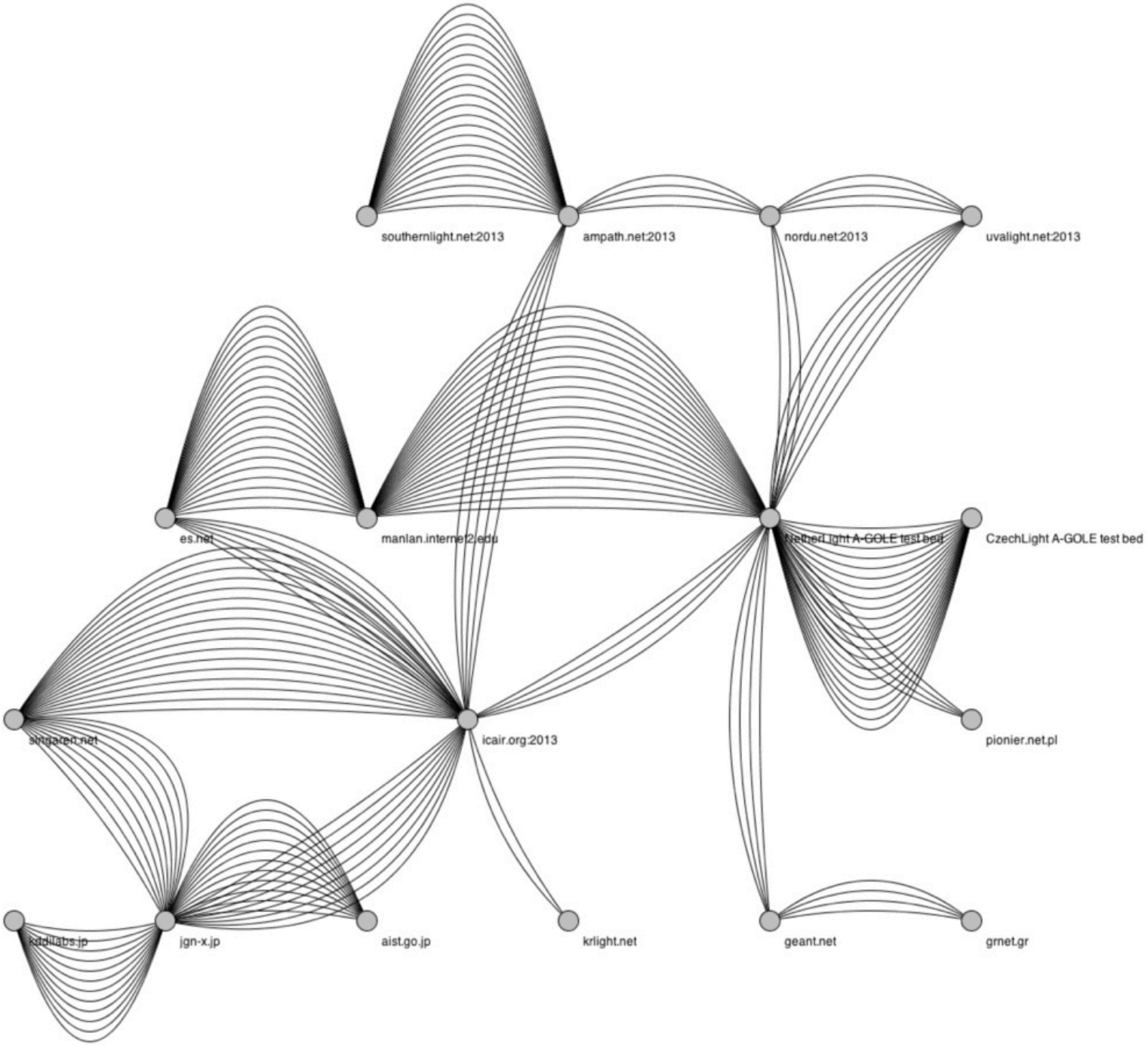


OpenGridForum
OPEN FORUM | OPEN STANDARDS

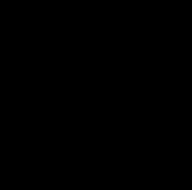
http://redmine.ogf.org/projects/nml-wg
http://redmine.ogf.org/projects/nsi-wg

http://sne.science.uva.nl/ndl

GLIF NSI 2013 in NML



CdL



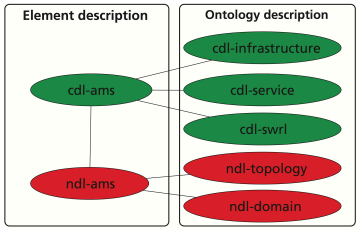
Applications, Data and Networks become aware of each other!

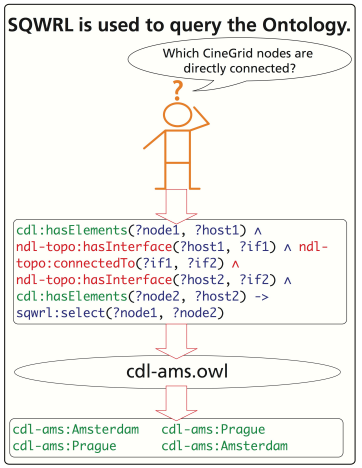
CineGrid Description Language

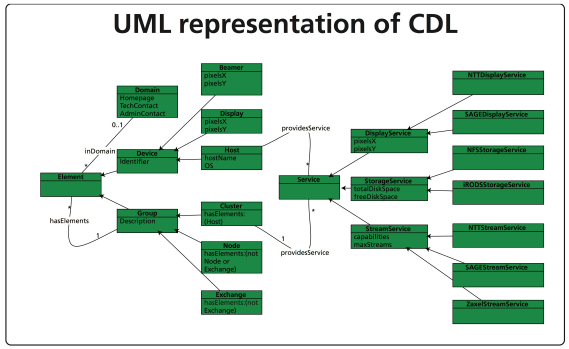
NCF

CineGrid is an initiative to facilitate the exchange, storage and display of high-quality digital media.

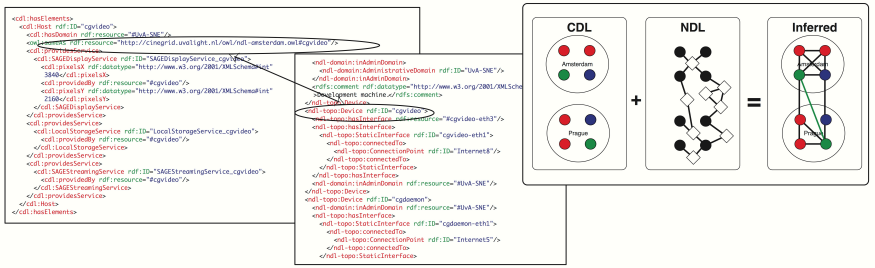
The CineGrid Description Language (CDL) describes CineGrid resources. Streaming, display and storage components are organized in a hierarchical way. CDL has bindings to the NDL ontology that enables descriptions of network components and their interconnections. With CDL we can reason on the CineGrid infrastructure and its services.







CDL links to ***NDL*** using the ***owl:SameAs*** property. ***CDL*** defines the services, ***NDL*** the network interfaces and links. The combination of the two ontologies identities the host pairs that support matching services via existing network connections.



UNIVERSITEIT VAN AMSTERDAM

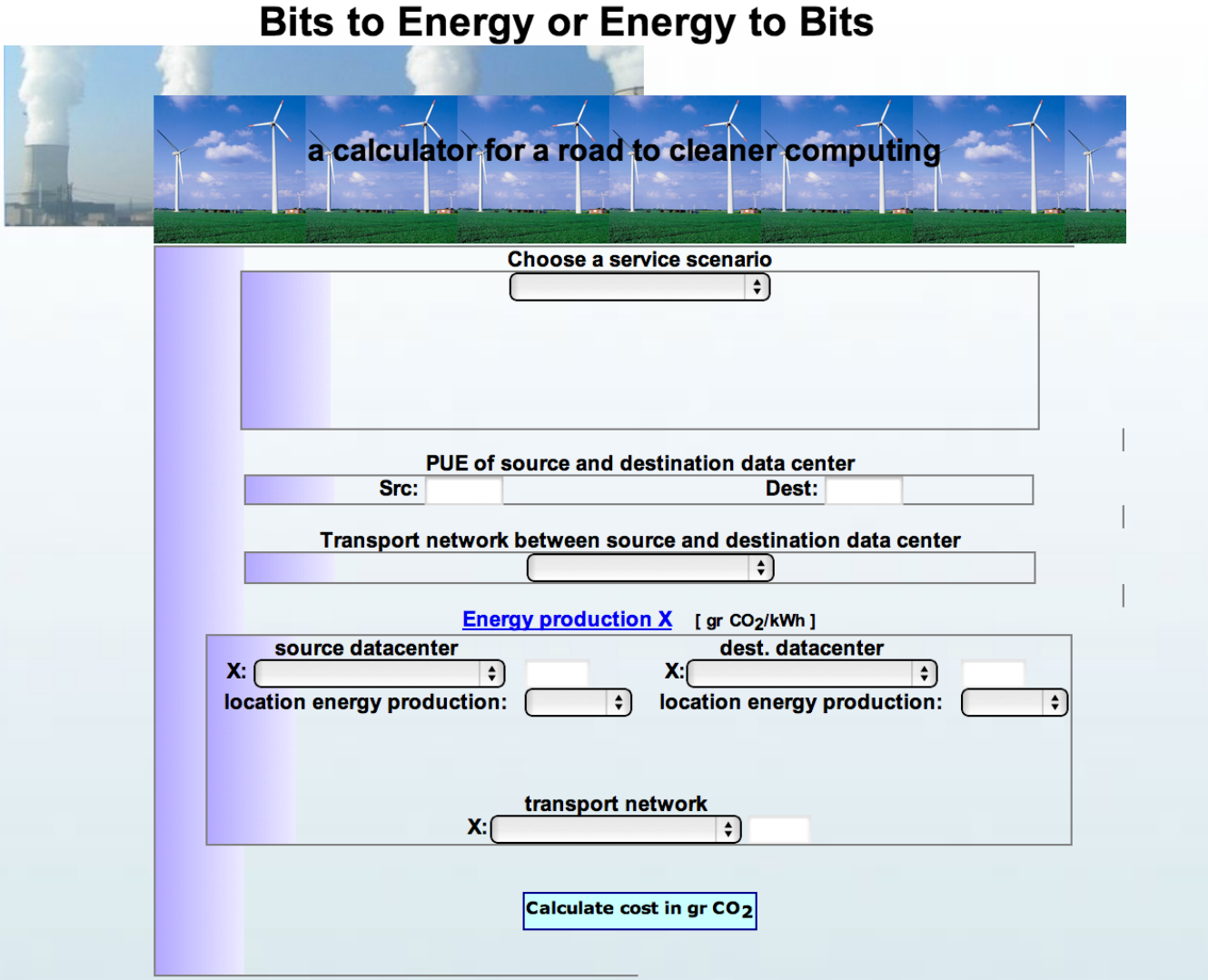
CineGrid

GigaPort

CREDITS CINEGRID.UVALIGHT.NL | WWW.CINEGRID.NL | WWW.CINEGRID.ORG
FACULTY OF SCIENCE UNIVERSITY OF AMSTERDAM | RALPH KONING: RALPH@SCIENCE.UVA.NL | PAOLA GROSSO: P.GROSSO@UVA.NL

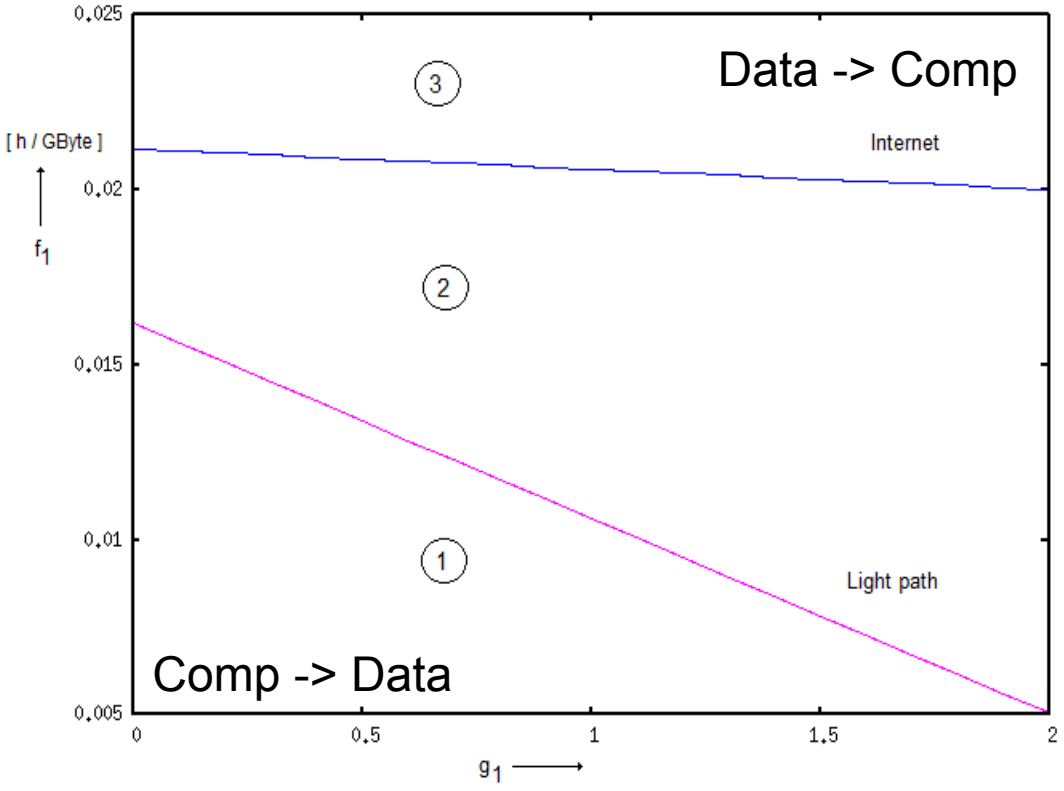
Bits-Nets-Energy

http://sne.science.uva.nl/bits2energy/

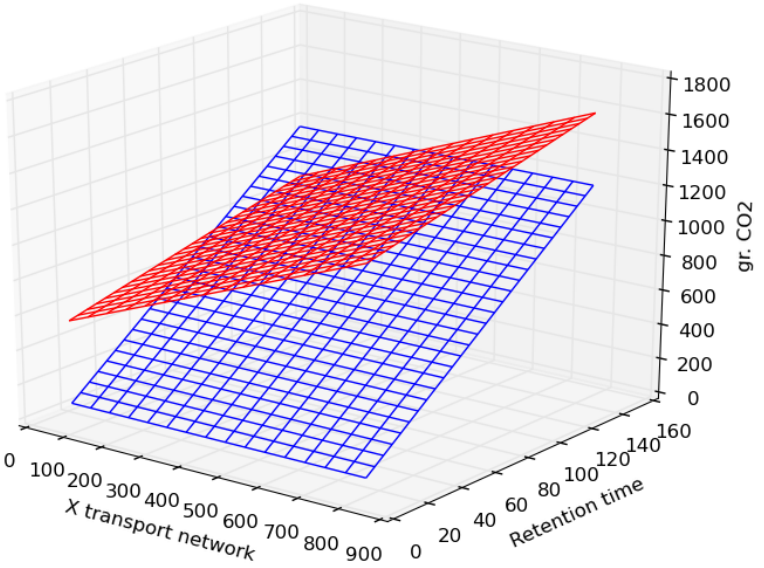


Taal & Grosso

“Data to Computing or vice versa?”



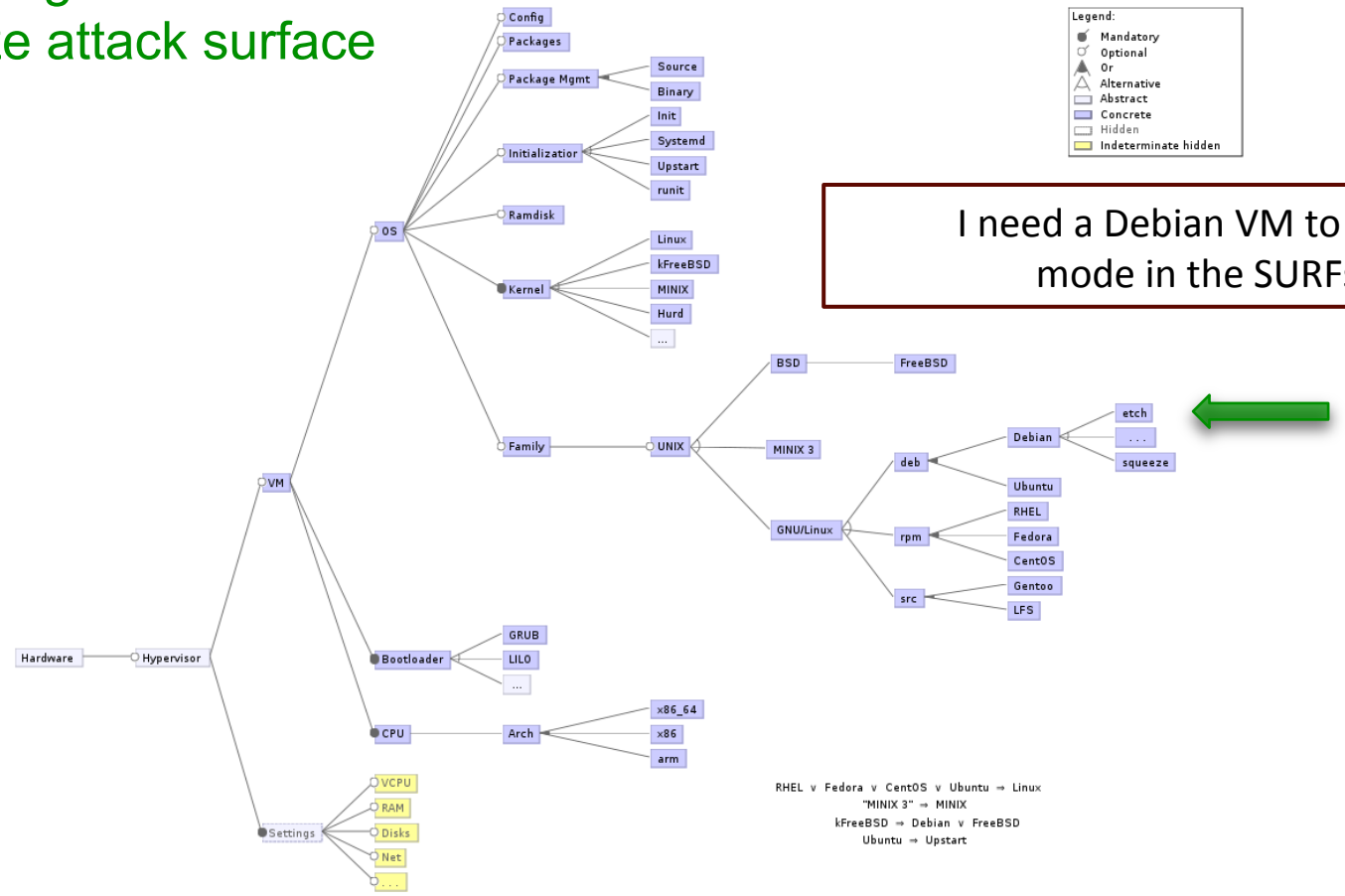
Storage to energy: when should you move hot or cold data to a green remote data centra for storage?



Given different network paths what are the decision boundaries as function of the task complexity.

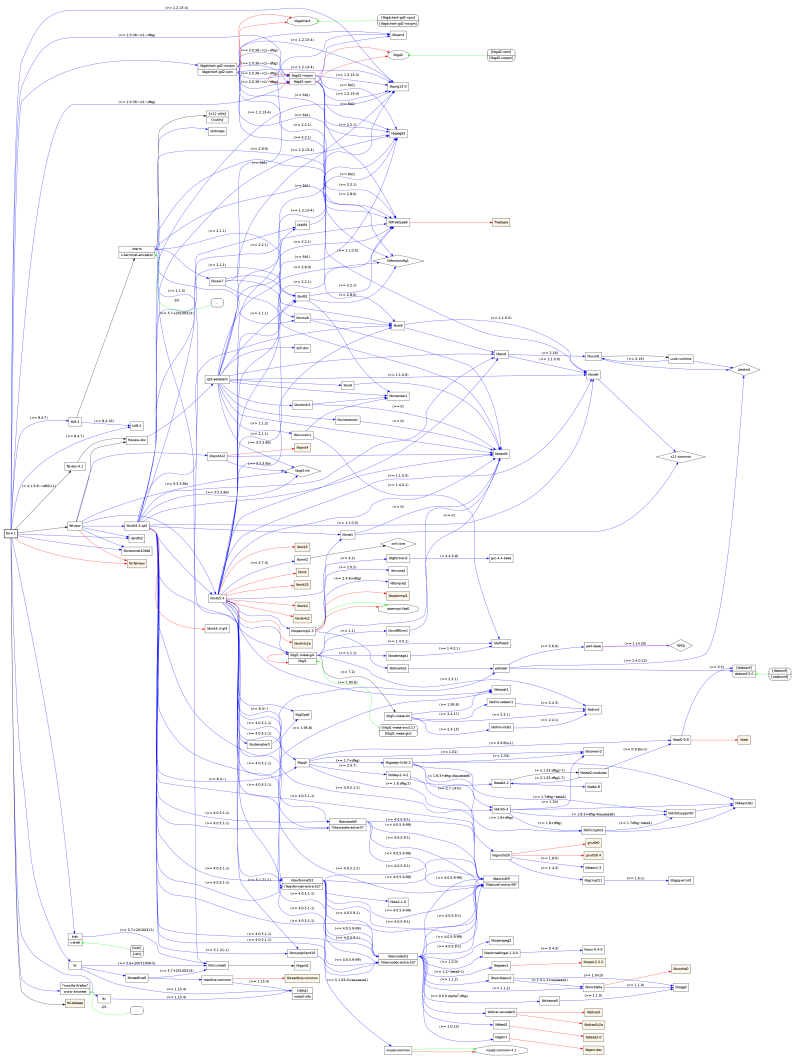
Security of Data in Purpose-Driven Virtual Machines

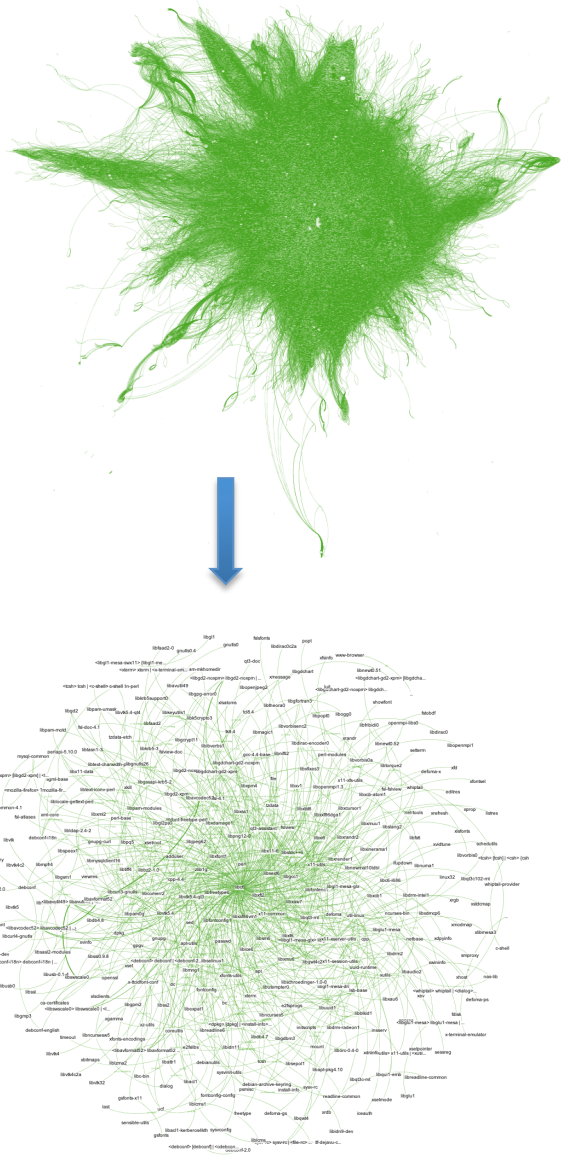
⇒ Cloud VM's contain lots of lib's and features not needed by application
⇒ Everything is a risk for hacks
⇒ Minimize attack surface



I need a Debian VM to run `fsl` in text mode in the SURFsara HPC Cloud.

Dependencies of an application



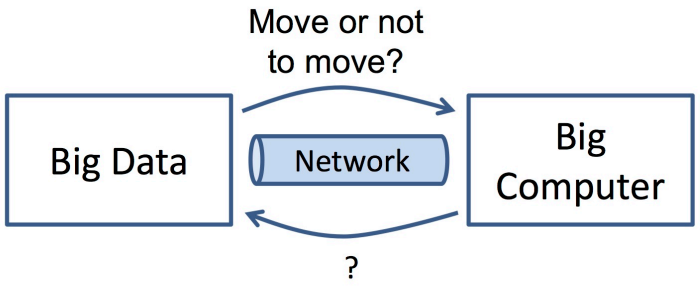


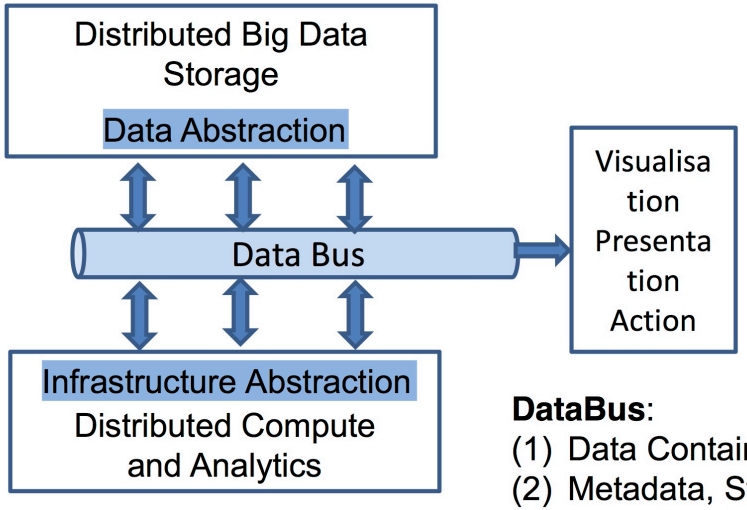
The application is `fsl-4.1`. LEFT: before dependency resolution, with all dependency constraints shown and RIGHT: resolved dependencies in a particular setup, `libc` is the center node

Towards Defining Big Data Architecture Framework

Yuri Demchenko, Marcel Worring, Wouter Los, Cees de Laat

Big Data Paradigm Change: Moving to Data-Centric Models





DataBus:
(1) Data Container
(2) Metadata, State
(3) Data Transfer Protocol

Current IT and communication technologies are host based or host centric (service/message centric)

- Any communication or processing are bound to host/computer that runs software
- For security: all security models are host/client based

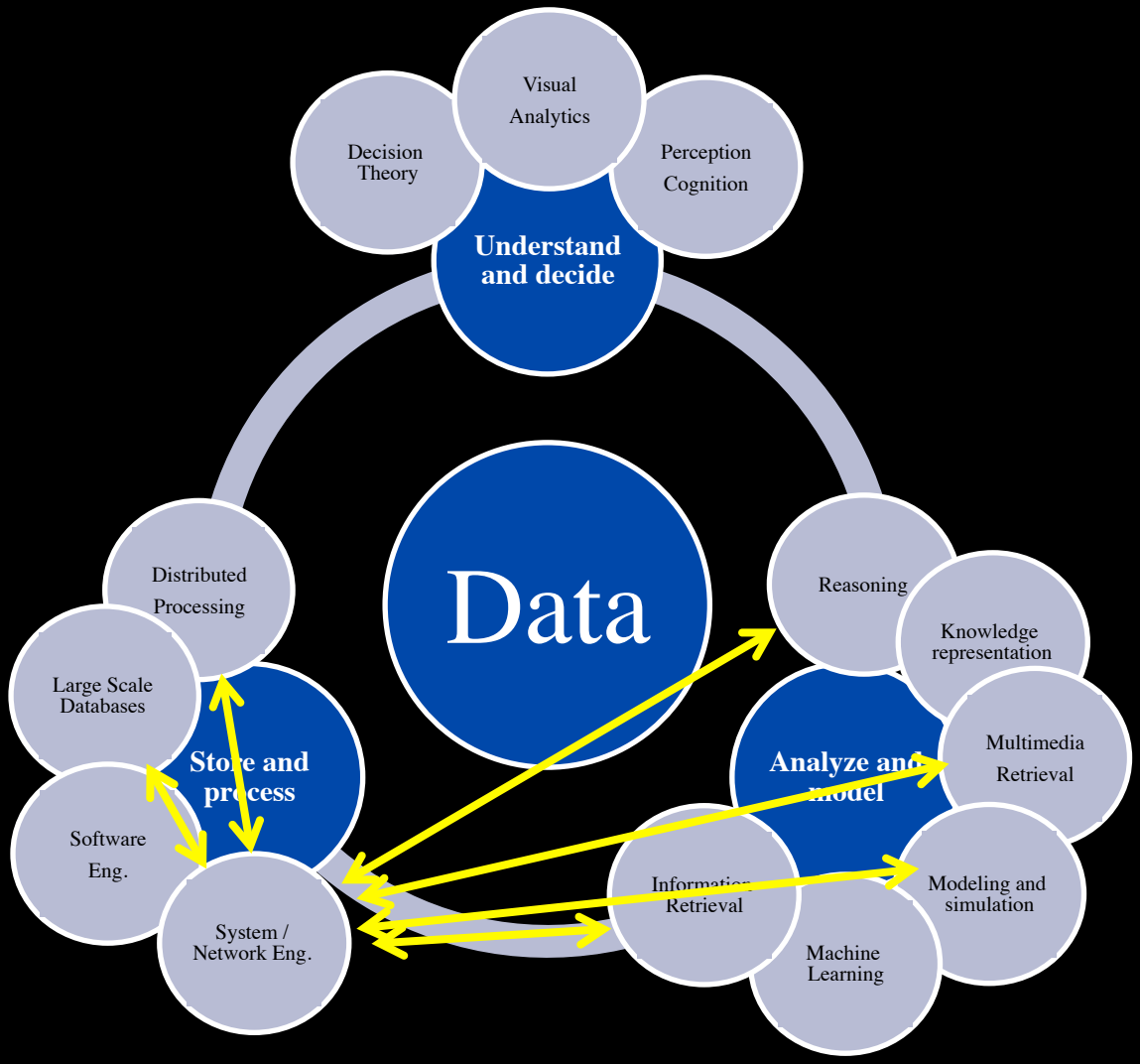
Big Data requires new data-centric models

- Data location, replication, search, access
- Data lifecycle, transformation, variability
- Data integrity, identification, ownership
- Data centric security and access control

Paradigm changing factors

- **Big Data properties: 5+1 V's**
- **Data aggregation:** multi-domain, multi-format, variability, linkage, referral integrity
- **Policy granularity:** variety and complex structure, for their access control processing
- **Virtualization:** Can improve security of data processing environment but cannot solve data security "in rest"
- **Mobility** of the different components of the typical data infrastructure: data, sensors or data source, data consumer

SNE in Data Science



Research direction

- Information on Infrastructure
- Control & programmability of Infrastructure
- Virtualization
- Networked data processing
- Sustainability & Complexity

Events on the horizon

- I4DW & DSRC
  - Launch Nov 13
- PIRE & OpenScienceDataCloud.org
  - Workshop June 2014 @ UvA
- Research Data Alliance
  - Conference in Amsterdam Sept 2014

Announcement June 2014
PIRE Workshop Amsterdam

- OpenScienceDataCloud.org
- PIRE Fellowship Application (+/- 15)
- The OSDC PIRE Program is six to eight week fully funded fellowship for US graduate student researchers with an information technology background.
- Format:
  - 1 week tutorials and hands on training
  - 2 months research at a participating institute
  - Results in science/tools and papers/posters/

NSF Partnerships for International Research and Education

OSDC OPEN SCIENCE DATA CLOUD

cisa

CIARA Center for Internet Augmented Research & Assessment research • collaboration • scholarship

AMPATH PATHWAY TO THE AMERICAS

FIU FLORIDA INTERNATIONAL UNIVERSITY

OCC OPEN CLOUD CONSORTIUM

THE UNIVERSITY OF CHICAGO

COMMIT/

sicsa* The Scottish Informatics & Computer Science Alliance

edinburgh data-intensive research

UNIVERSITY OF AMSTERDAM

PIRE - OpenScienceDataCloud.org

1000 Genomes Project
Human sequence data from populations around the world with the goal of cataloging human genetic variation.
Total Size: 383.5TB Categories: genomics, biology

ASTER
ASTER Level-1B Registered Radiance at the Sensor
Total Size: 12.7TB Categories: earth science

Complete Genomics Public Data
Whole human genome sequence data sets provided by Complete Genomics, containing 69 standard, non-diseased samples as well as two matched tumor and normal sample pairs.
Total Size: 47.1TB Categories: genomics, biology

Earth Observing-1 Mission
Data gathered by the Advanced Land Imager (ALI) Hyperspectral Imager (Hyperion) instruments on NASA's Earth Observing-1 Mission (EO-1) satellite.
Total Size: 45.2TB Categories: earth science, satellite imagery

City of Chicago Public Datasets
Data set from the City of Chicago Data Portal in JSON format for tabular data and the raw files for "blob" data.
Total Size: 9.7GB Categories: social science

EMDataBank
Unified Data Resource for 3-Dimensional Electron Microscopy
Total Size: 91.3GB Categories: biology

Enron Emails
Data sets based on the original Enron emails released to the public by the Federal Energy Regulatory Commission as part of their investigation.
Total Size: 155.9GB Categories: social science

FlyBase
FlyBase is the leading database and web portal for genetic and genomic information on the fruit fly Drosophila melanogaster and related fly species.
Total Size: 614.8GB Categories: biology, genomics

.....

OSDC OPEN SCIENCE DATA CLOUD

ENVRI

Research Data Alliance

- https://rd-alliance.org
- The Research Data Alliance implements the technology, practice, and connections that make Data Work across barriers.
- The Research Data Alliance aims to accelerate and facilitate research data sharing and exchange.
  - Working groups and interest groups
  - Joining groups and attendance at the twice-yearly plenary meetings is open.
- Plenary Sep 2014 hosted by the Netherlands - Amsterdam
  - Conference Management Team (CMT) Chair: Peter Doorn (DANS)
  - Program Committee (PC): chair Cees de Laat (UvA)
  - Satellite Events Committee (SEC): Jeroen Rombouts (TUD)

Questions?

http://sne.science.uva.nl

http://www.os3.nl/

http://i4dw.nl/

http://dsrc.nl/

http://sne.science.uva.nl/openlab/

http://pire.opensciencedatacloud.org

http://staff.science.uva.nl/~delaat/pire/

https://rd-alliance.org