



# Data Deluge

twitter



LinkedIn

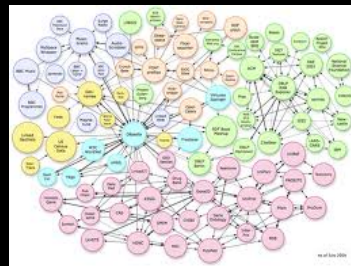
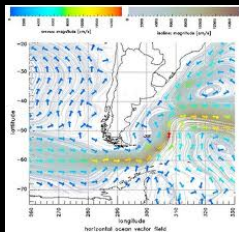


SchoolBANK

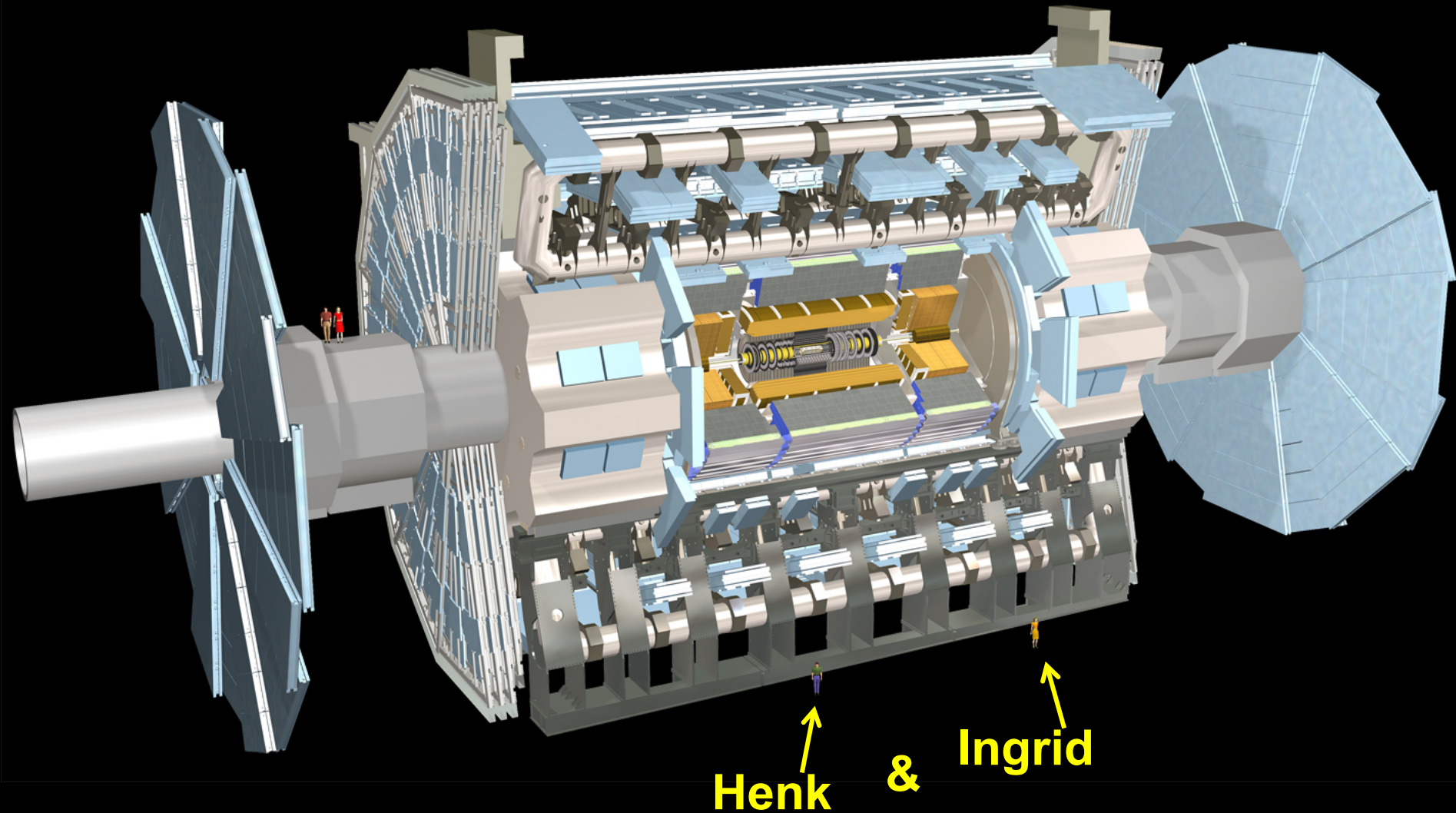
Hyves



Google



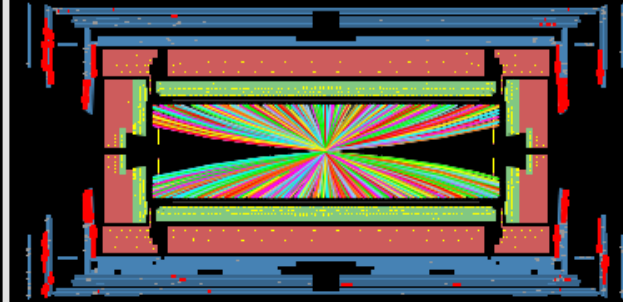
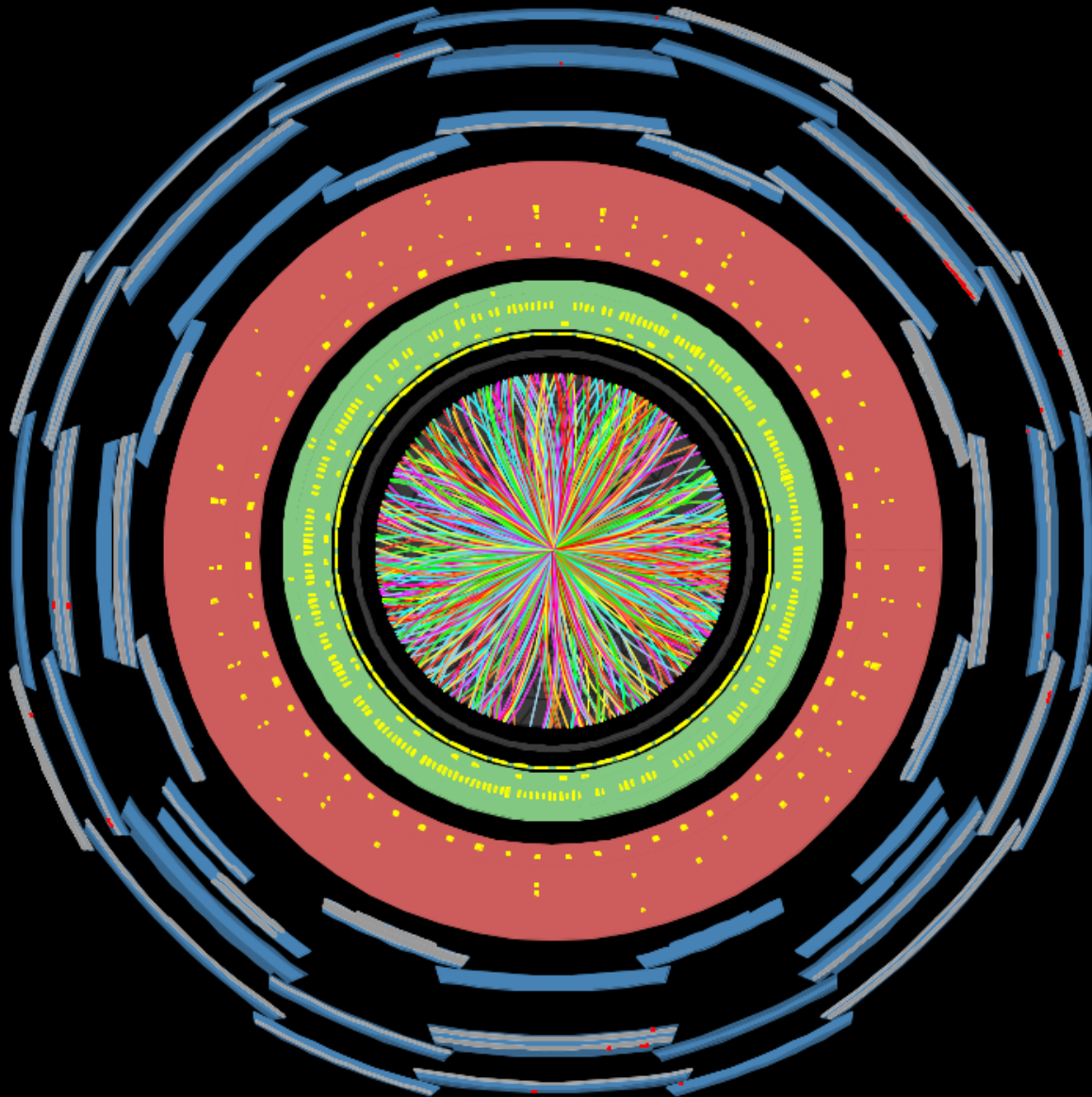
# ATLAS detector @ CERN Geneve



# ATLAS detector @ CERN Geneve



# Een gebeurtenis



Run Number: 170482, Event Number: 3936308

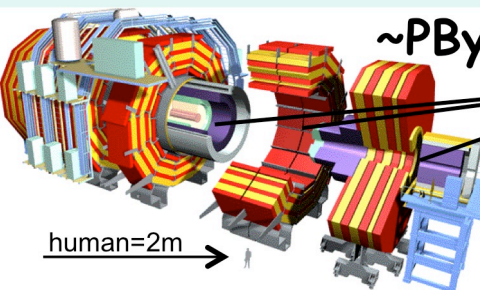
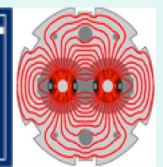
Date: 2010-12-06 17:21:31 CET

Snapshot of a heavy ion collision  
directly from the ATLAS experiment



# LHC Data Grid Hierarchy

CMS as example, Atlas is similar



human=2m →

**CMS detector: 15m X 15m X 22m**  
12,500 tons, \$700M.

Online System

Tier 0 + 1

~100 MBytes/sec

100000 flops/byte

10 Pflops/s

event simulation

event reconstruction



~2.5 Gbits/sec

Italian Regional Center

German Regional Center

NIKHEF Dutch Regional Center

FermiLab, USA Regional Center

...

Tier2 Center

~0.6-2.5 Gbps

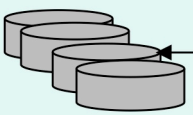
Tier 2

analysis

~0.6-2.5 Gbps

Tier 3

Institute ~0.25TIPS



Physics data cache

100 - 1000 Mbits/sec



Workstations

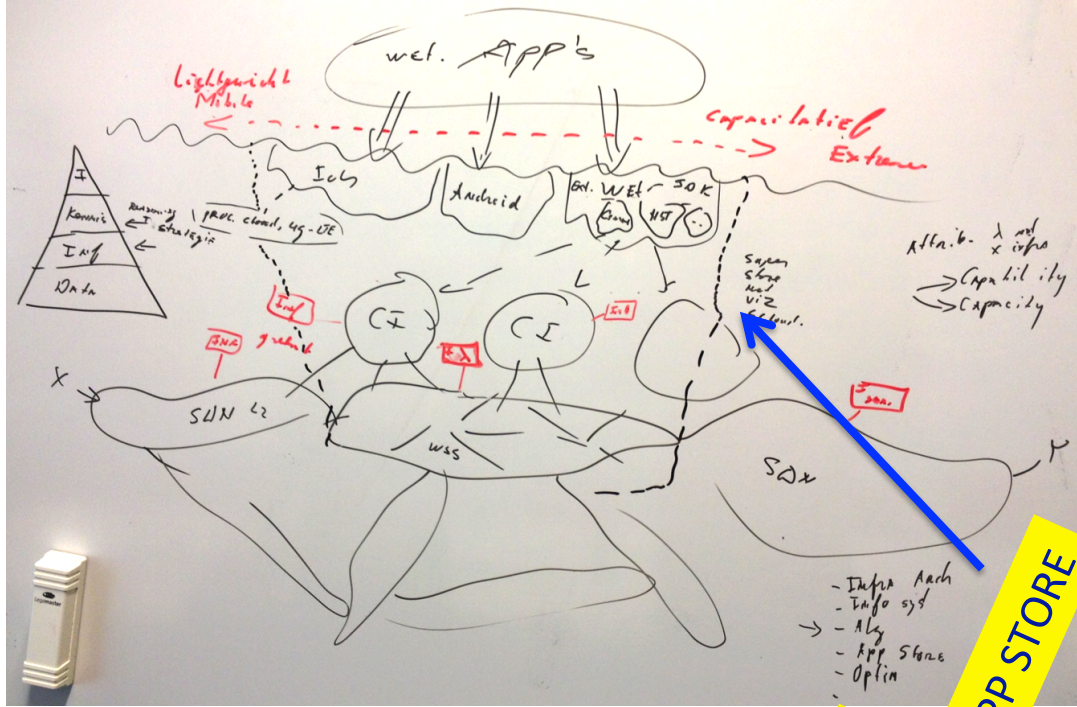
Tier 4

CERN/CMS data goes to 6-8 Tier 1 regional centers, and from each of these to 6-10 Tier 2 centers.

Physicists work on analysis "channels" at 135 institutes. Each institute has ~10 physicists working on one or more channels.

2000 physicists in 31 countries are involved in this 20-year experiment in which DOE is a major player.

# TimeLine



Remote Procedure Call

Distributed Computing (MPI)

GRID

CLOUD

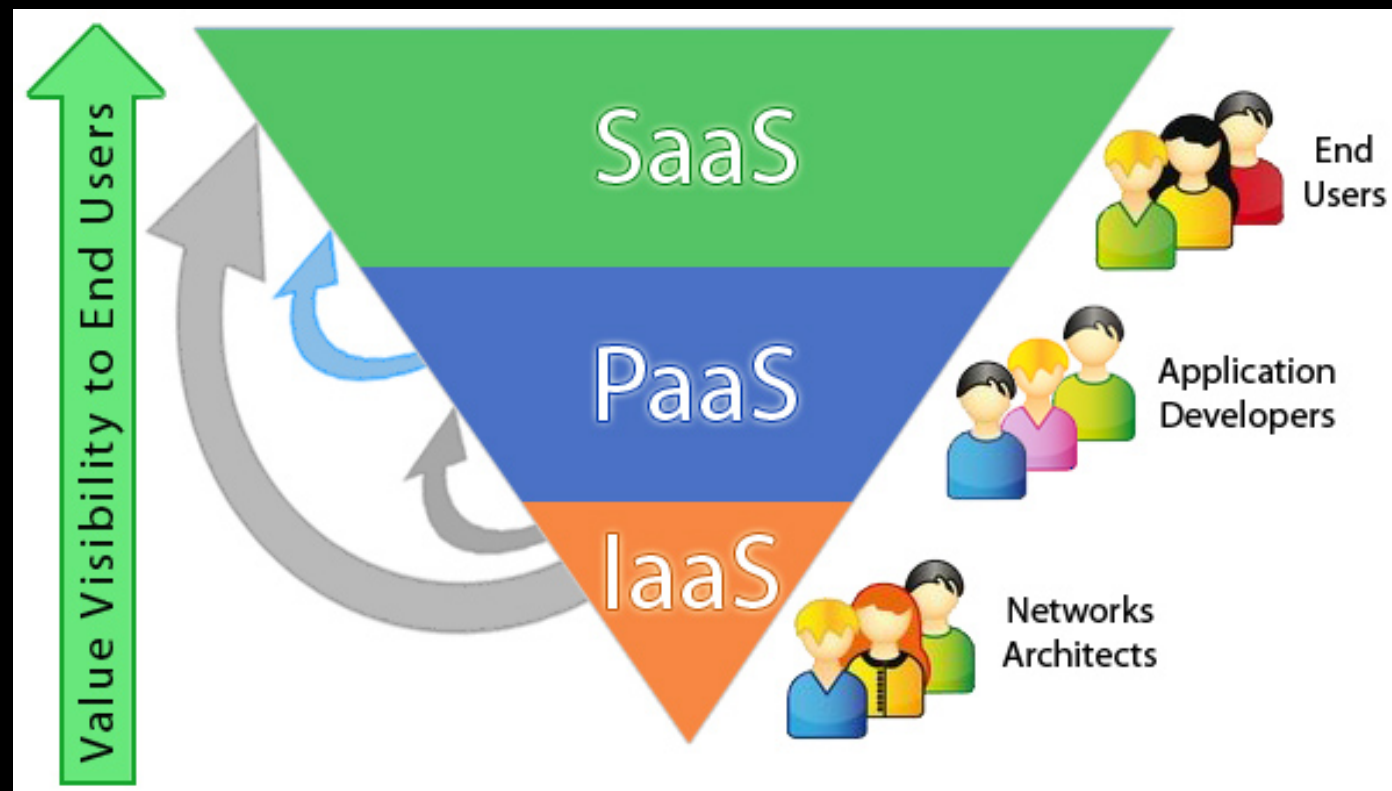
BIG DATA

WET. APP STORE

1980 1990 2000 2005 2013

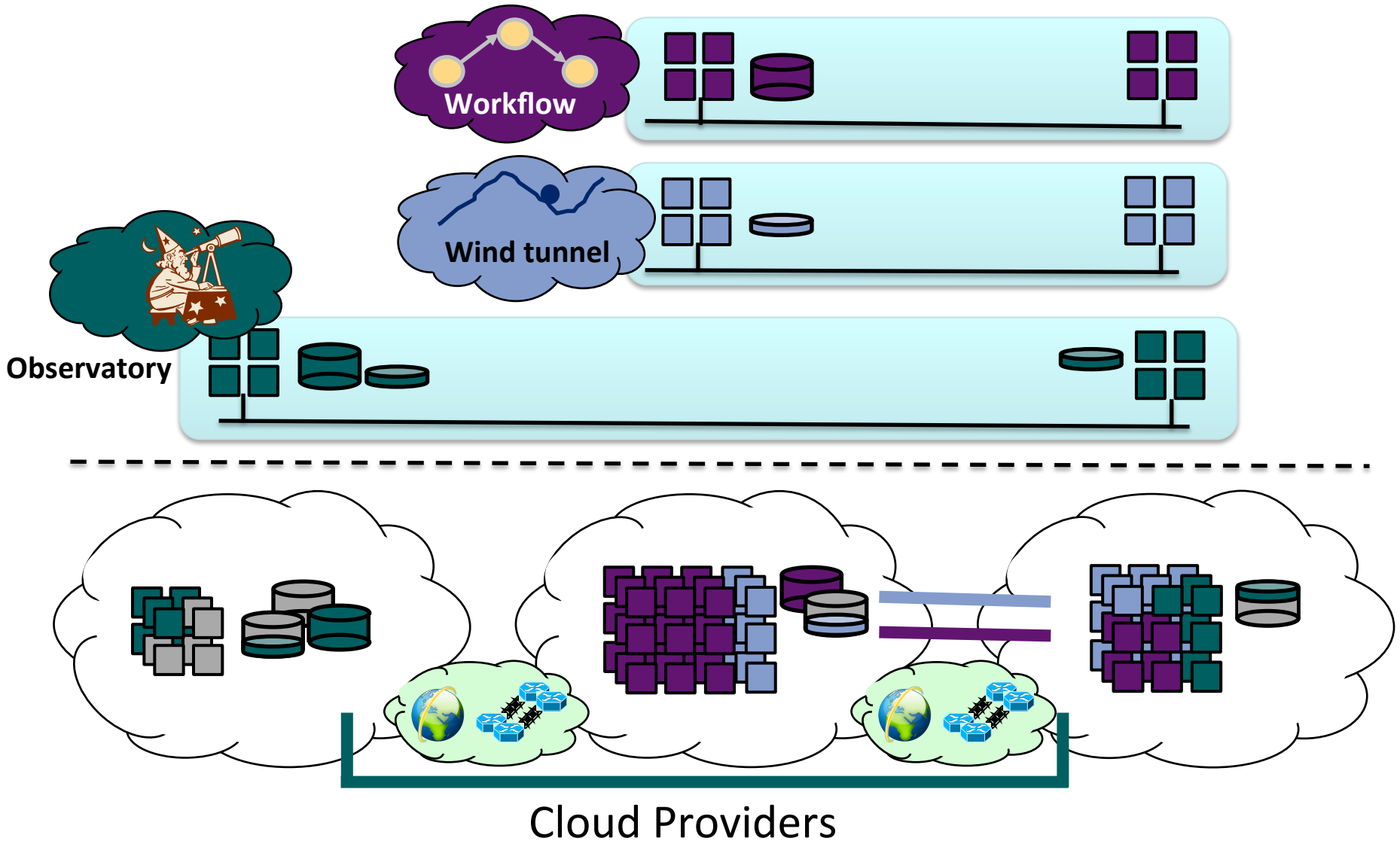
# Techniek

- IAAS - *Infrastructure-as-a-Service*
- PAAS - *Platform-as-a-Service*
- SAAS - *Software-as-a-Service*

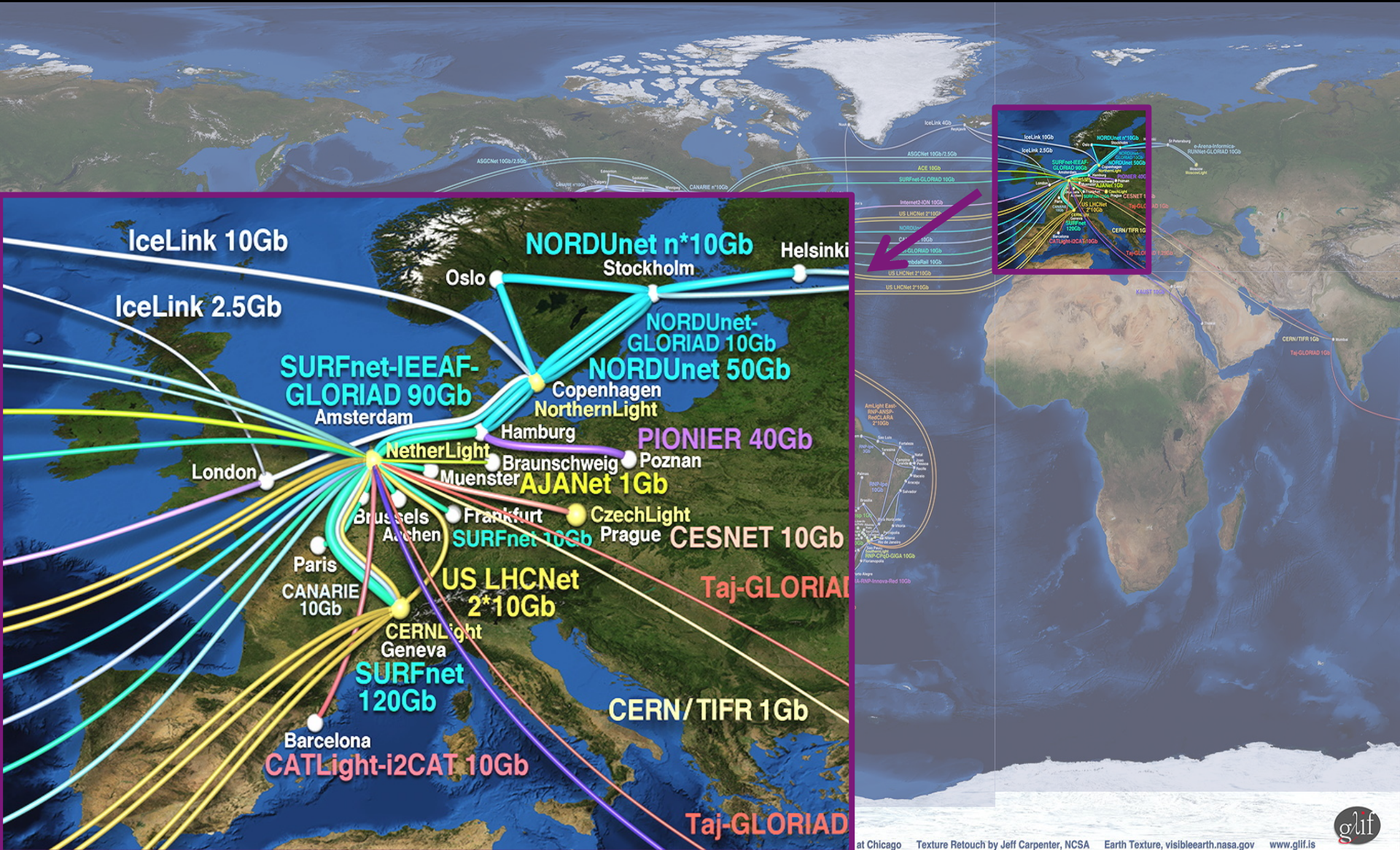




# IaaS: Networked Clouds







# VIZUALIZATION

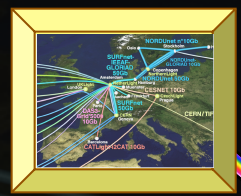
DataExploration

RemoteControl

TV

Medical

CineGrid



Gaming

Conference

Workflow

Clouds



Distributed

EventProcessing

# GRID&CLOUD

Management

Mining

Web2.0



Meta

NetherLight

# DATACENTER

Backup

Media

Visualisation

Security

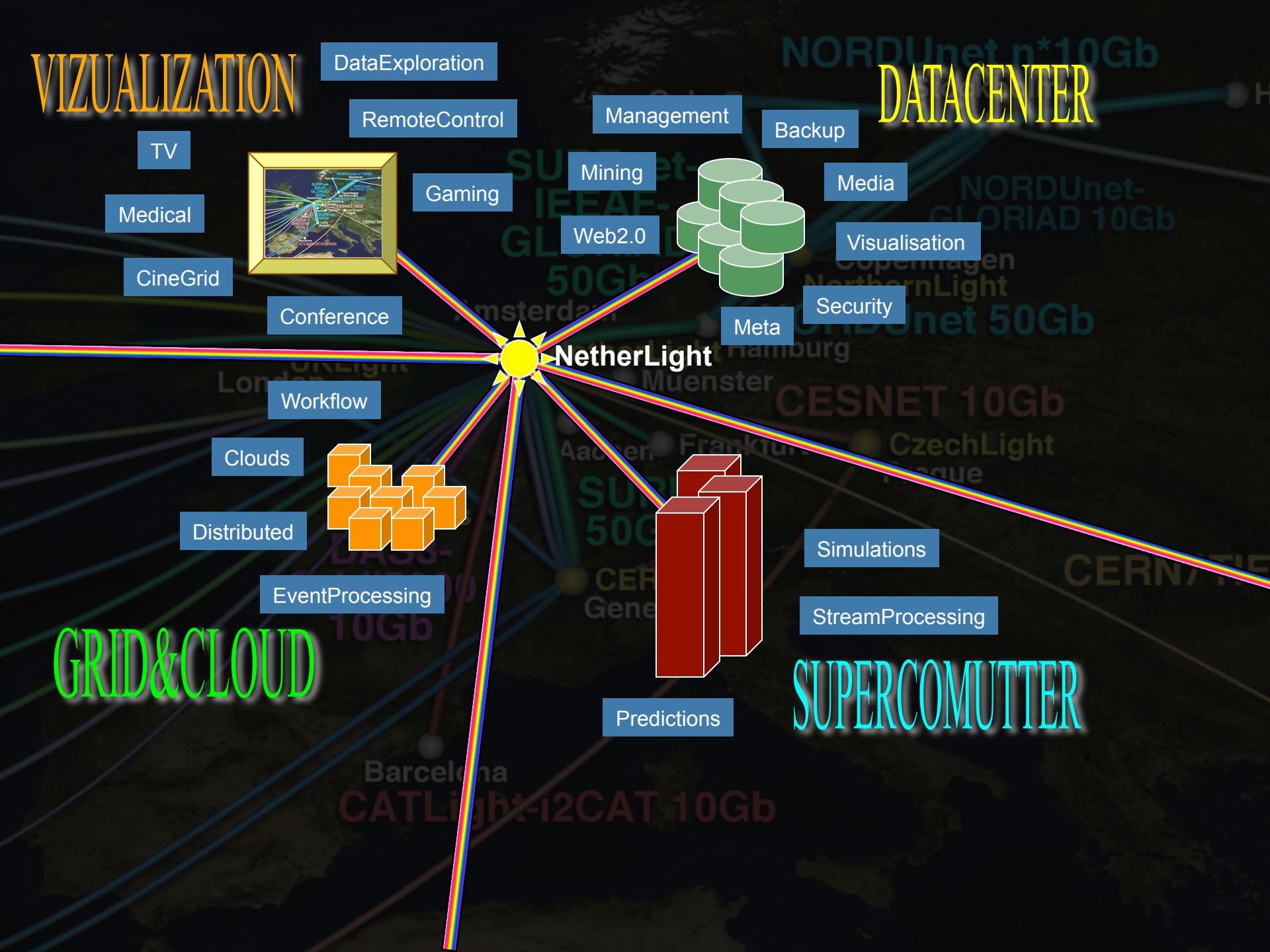
Simulations

StreamProcessing

# SUPERCOMUTTER

Predictions

Barcelona  
CATLight-i2CAT 10Gb

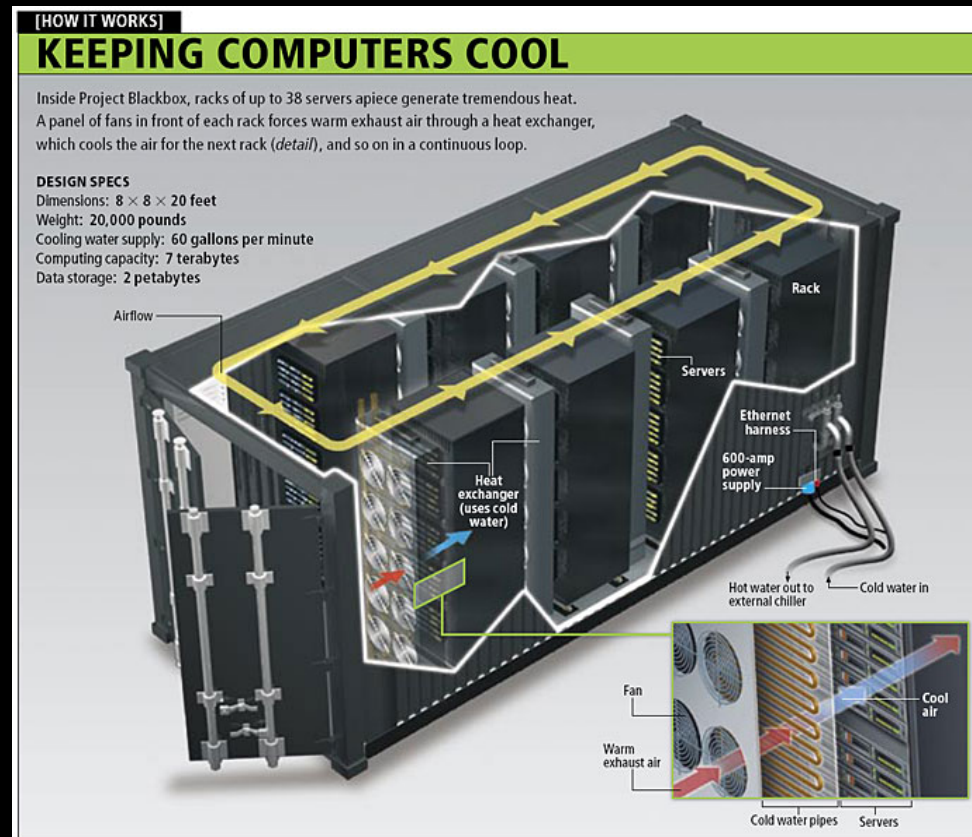


# Techniek

- **Publiek cloud:** commerciële cloud diensten in grote datacenters zoals Amazon, Google, Microsoft, Rackspace and Eduserv.
- **Private cloud:** lokale infrastructuur op basis van cloud technologie, in bezit van de eigen organisatie of een groep van organisaties (community).
- **Hybride cloud:** een private cloud met de mogelijkheid om naadloos te integreren met een of meerdere publieke clouds als de workload en gebruikerseisen dat nodig of wenselijk maken.

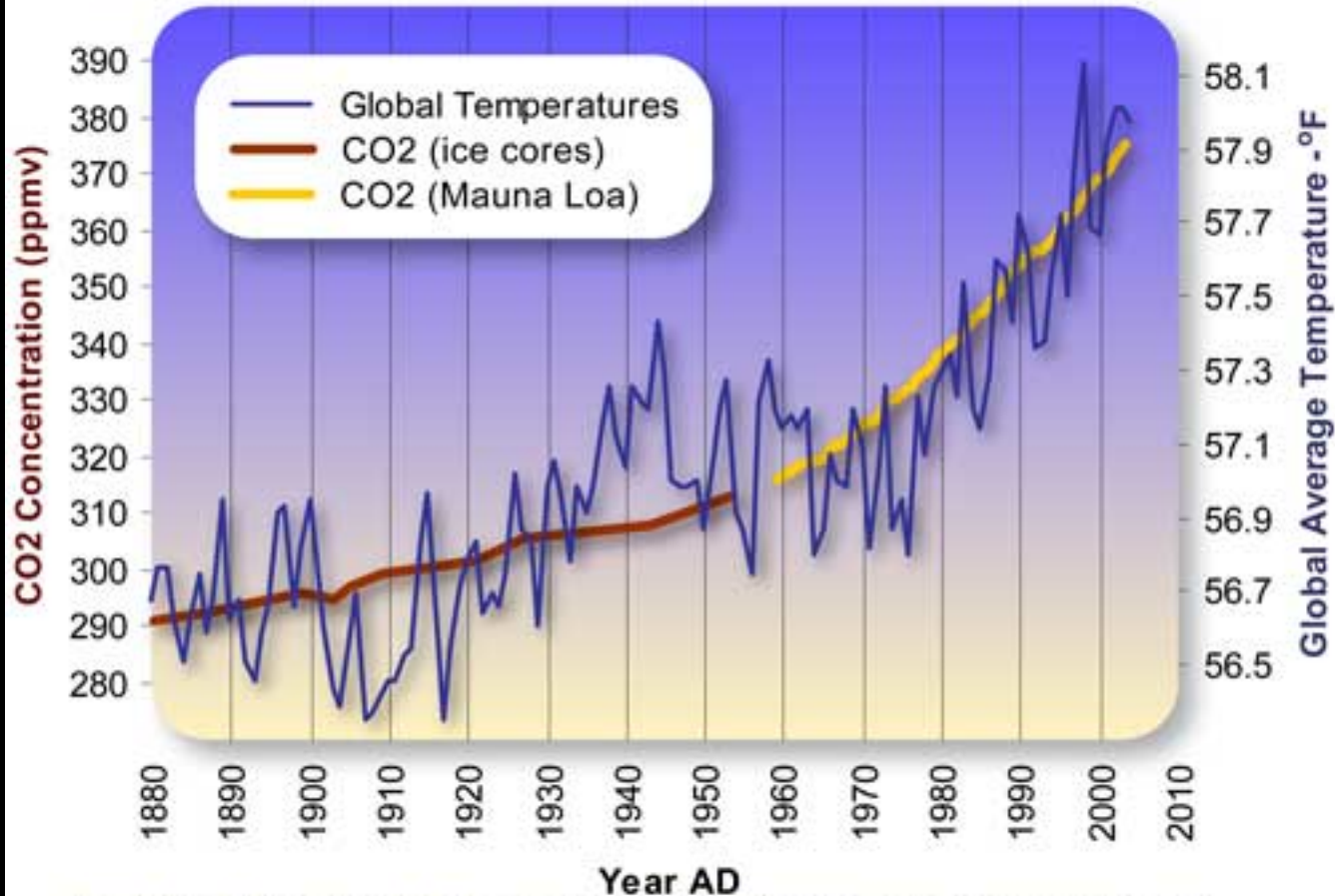
# Techniek

- Capacity computing ->
  - Grid/cloud – embarrassingly parallel
- Capability computing
  - Blue Gene, Top500
- GreenIT
  - >>>Megawatts!



# Need for GreenIT

## Global Average Temperature and Carbon Dioxide Concentrations, 1880 - 2004



Data Source Temperature: [ftp://ftp.ncdc.noaa.gov/pub/data/anomalies/annual\\_land\\_and\\_ocean.ts](ftp://ftp.ncdc.noaa.gov/pub/data/anomalies/annual_land_and_ocean.ts)

Data Source CO2 (Siple Ice Cores): <http://cdiac.esd.ornl.gov/ftp/trends/co2/siple2.013>

Data Source CO2 (Mauna Loa): <http://cdiac.esd.ornl.gov/ftp/trends/co2/maunaloa.co2>

Graphic Design: Michael Ernst, The Woods Hole Research Center



# Greening the Processing System

***Positive proof of global warming.***







# Turn Green Tech into Greenbacks

IT Certifications for Jobs That Make a Difference



## Uptime Institute Accredited Tier Designer

The Uptime Institute has long been a proponent for green data center design and implementation. Its certification course on data center design embeds green principles into the curriculum.



SUSTAINABILITY  
**Your Career**



# ECO-Scheduling



# Opportunities

- ICT als utility (water, stroom)
- Schaalbaarheid & elasticiteit
- Kosten
- Modernisering
- Spreiding & resilience

the new Facebook data center in Lulea, Sweden  
27,000 square meter



# Verschuiving

Computing

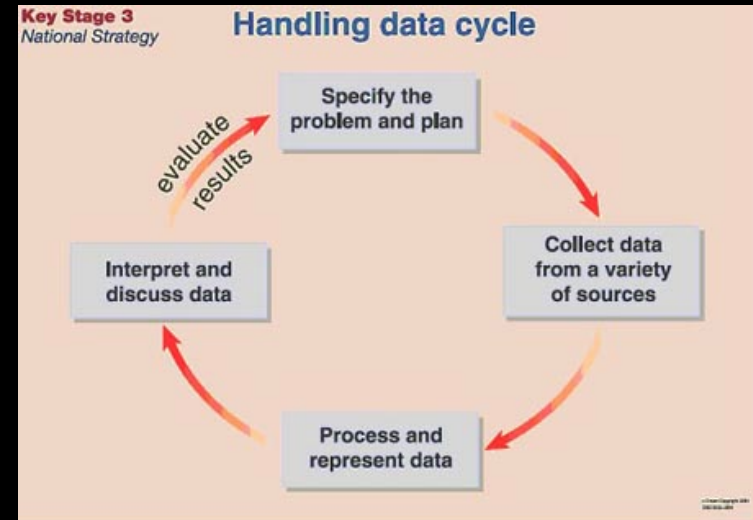
Storage

Data

Meaning/analyse

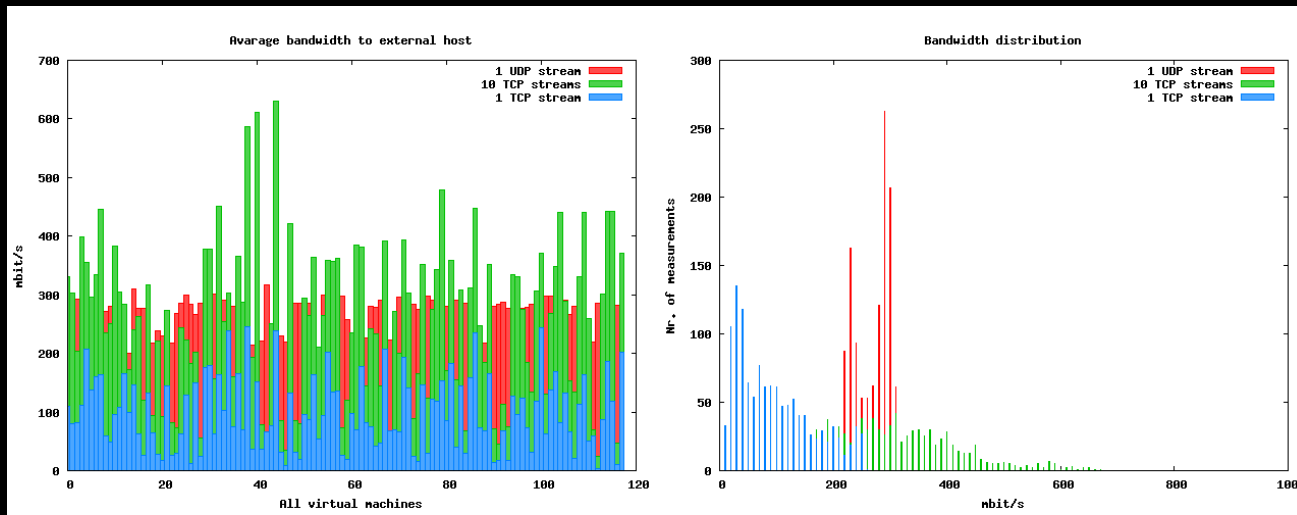
Model/predict/simulate

Control/decision support



# Pitfalls

- Nog weinig gestandaardiseerd
- Data in makkelijker dan data out
- Quality of Service van complete dienst
- Grote variëteit van performance
- Specifieke capability computing blijft nodig (SKA)



# Security & privacy in cloud

- Gevoelige data en jurisdictie
- Techniek geeft geen garanties
- Cloud onder Nederlands recht (SURF taak)
- Policy firewall naar schaalbaar publiek cloud

# Vragen

- Wat te doen bij faillissement provider?
- Wat te doen bij (on)voorziene destructieve rampen?
- Wat indien data echt compleet in verkeerde handen valt?
- Hoe lang kan een organisatie bedrijfsmatig overleven als specifieke diensten uitvallen?

<http://delaat.net/>

