



# Introduction to Grid Computing

May 2013

Marian Babik

CERN



# Outline

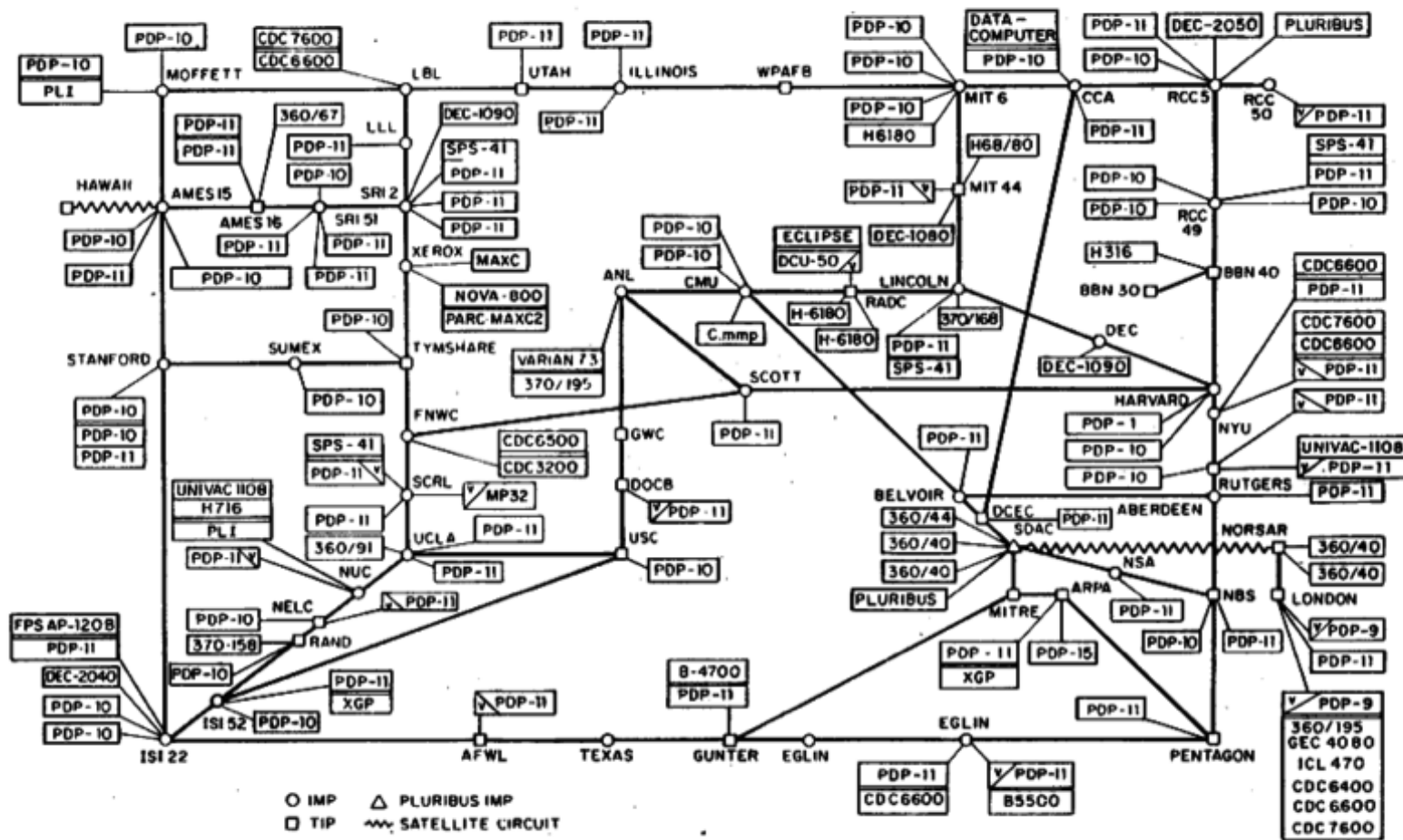
- Networking
- Web
- Web 2.0
- Distributed systems
- Grid
- Cloud

# NETWORKING



# Networking

ARPANET LOGICAL MAP, MARCH 1977



(PLEASE NOTE THAT WHILE THIS MAP SHOWS THE HOST POPULATION OF THE NETWORK ACCORDING TO THE BEST INFORMATION OBTAINABLE, NO CLAIM CAN BE MADE FOR ITS ACCURACY)

NAMES SHOWN ARE IMP NAMES, NOT (NECESSARILY) HOST NAMES



# Networking

- ARPAnet (Advanced Research Projects Agency Network, 1962)
  - MIT, DARPA, US Dept. of Defense,
  - first operational packet switching network
  - data system could use one communications link to communicate with more than one machine by disassembling data into datagrams, then gather these as packets.

29 OCT 67	2100	LOADED OP. PROGRAM	CSK
		FOR BEN BARKER	
		BBV	
	22:30	Talked to SRS	CSK
		Host to Host	
		Left op. program	CSK
		running after sending	
		a host dead message	
		to imp.	

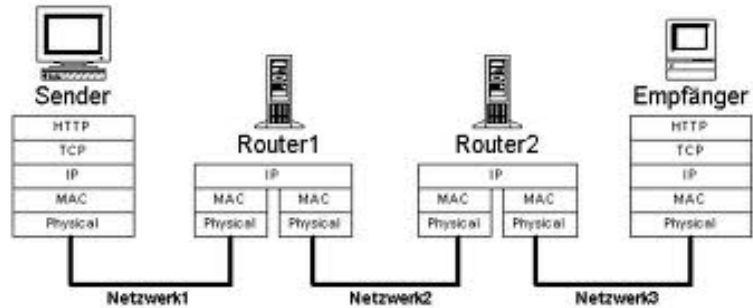
# Networking

- + many others (UUCP, NPL, X.25, etc.)
- TCP/IP network stack
- IP(Internet protocol) – datagrams, packets, addressing
  - IP Address
    - 32bit (4 bytes) (e.g. 209.85.229.147)
    - 4,294,967,296 ( $2^{32}$ )
  - Domain Name System (DNS)
    - [www.google.com](http://www.google.com) -> 209.85.229.14
- TCP (Transmission Control Protocol)
  - provides the service of exchanging data reliably directly between two network hosts

# Networking

- Routing

- how to deliver packets from one host to another
  - Graph theory
  - Minimum spanning tree
  - Shortest path



- Throughput/Latency

- average rate of successful message delivery over a communication channel
  - 10Mbit/100Mbit/1Gbit
- measure of time delay experienced in a system

**WEB**





# Web of Today!

Mittwoch, 10.9.2008 Schlagzeilen | TV-Programm | RSS | Newsletter | Mobil | Dienste | Startseite | Wetter | DER SPIEGEL

Startuj z Gazeta.pl

Web Bilder M Display O

B

+ Cu

melden

Mon **Le Matin Online** Inscription | Login

Rechercher

Shopping | Le Matin Mobile | Services abonnés | Edition électronique | Aide | Contact | Plan

## Le Matin Online

Le Matin Video | BleuBlog | Dossiers | Photos | Dessins | Concours | Jeux | Annonces

Accueil | Actu | Sports | People | Tendances | Multimédia | Loisirs | Dépêches | Femina | Bilan

**Imiter le chef**



**Toutes les vidéos et les recettes sur le blog**

- > Dom
- > Eduk
- > Praca
- > Turys
- > Auto
- > Co Jek
- > Dzieck
- > Film
- > Gazet
- > Gielda
- > Gospo
- > Gry
- > Kobiet
- > Komó
- > Kuchn
- > Miłośc
- > Moda

**Golfe du Mexique: l'ouragan Ike menace les USA**



L'ouragan Ike, balayant les eaux du Golfe du Mexique, poursuivait lentement sa route mercredi vers les Etats-Unis et les côtes du Texas. A Cuba, les intempéries ont entraîné l'évacuation de 2,6 millions de personnes et fait quatre morts. | [Suite](#)

**Edipresse: bénéfice en hausse de 10% au 1er semestre**



Edipresse a vu son bénéfice net progresser de 10% à 24 millions de francs au premier semestre. En revanche, les revenus de l'éditeur vaudois ont reculé dans les mêmes proportions, à 385,3 millions de francs. | [Suite](#)

**Collissionneur du CERN**

**Mise en route réussie du LHC: ce n'est pas la fin du monde!**



Le CERN a procédé avec succès ce matin au démarrage du plus grand accélérateur de particules au monde, le Grand collisionneur de hadrons (LHC). A 09h33, le faisceau de protons était injecté dans l'anneau de 27 kilomètres et à 10h25, le tour complet était réalisé pour la première fois. Suivez l'expérience en direct sur [le matin.ch](#) | [Suite](#)

Le Matin Online 69 commentaires

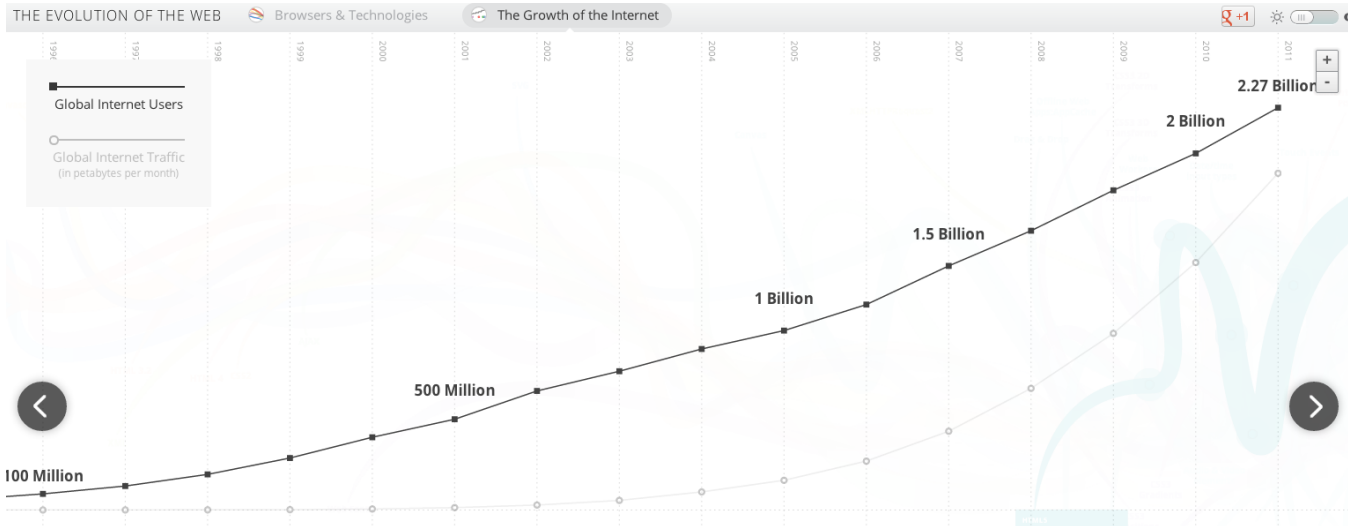
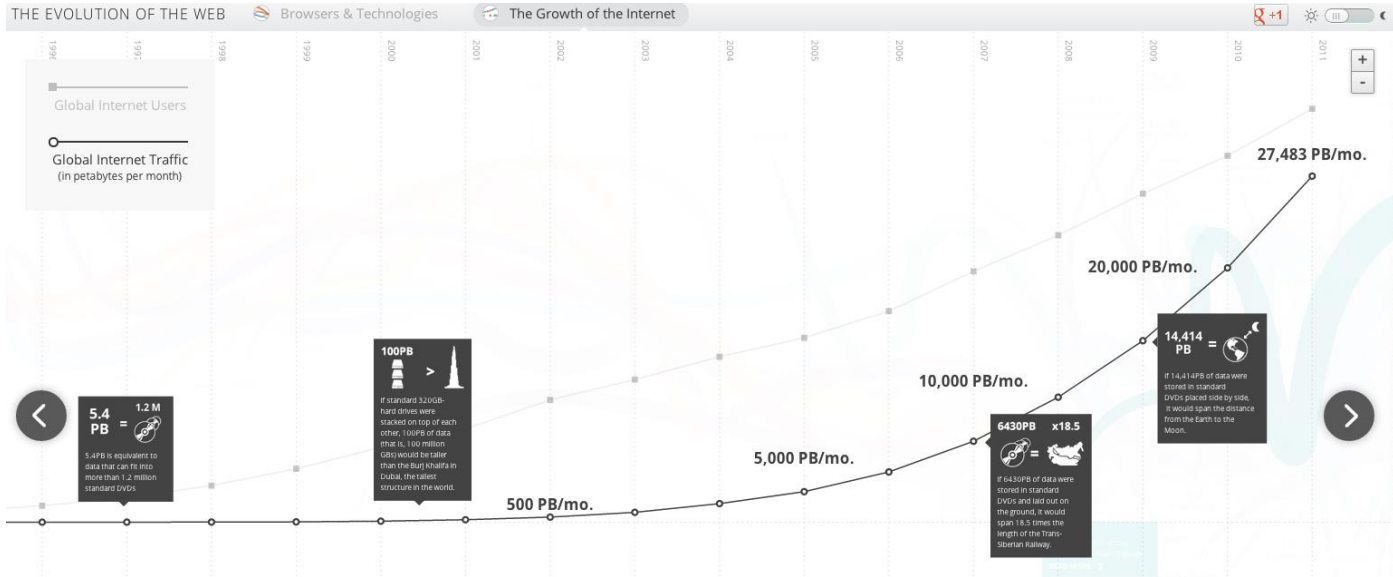
**Dossier**

> Le CERN et la fin du monde

© Keystone



# Web of Today



source: <http://www.evolutionoftheweb.com/>



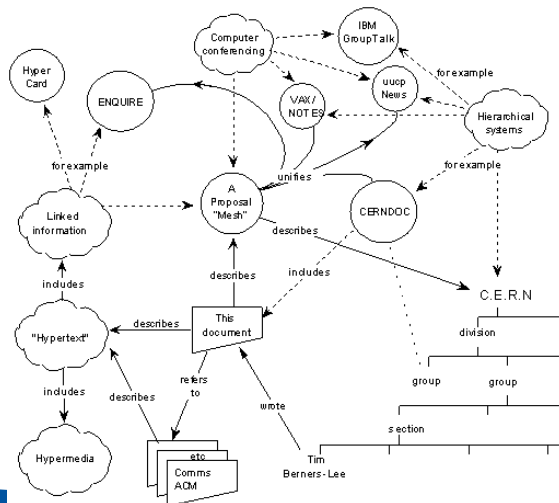
# What is the Web ?



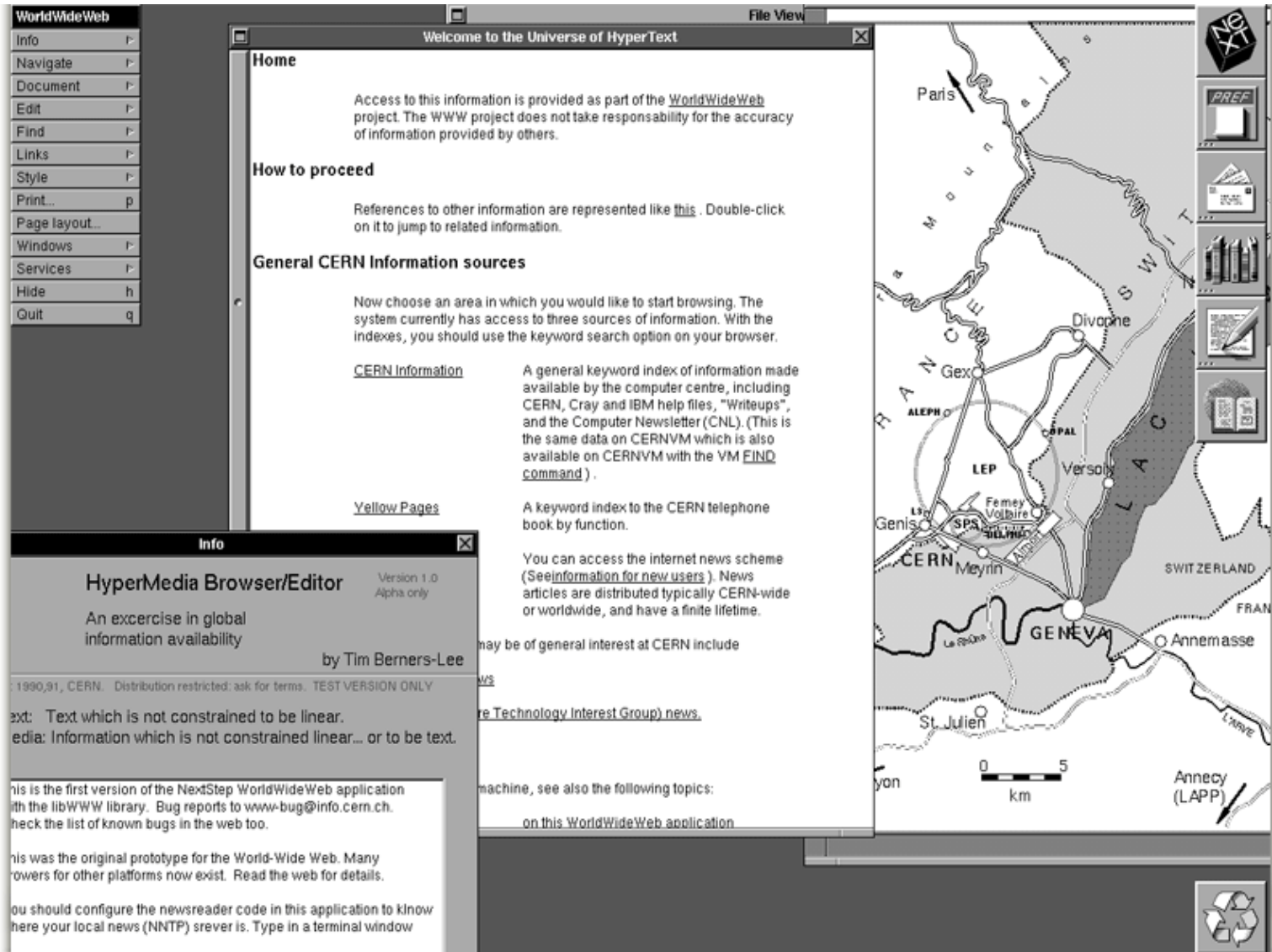
- Internet + Hypertext
- Hypertext ?
  - text displayed on a computer or other electronic device with references

1991: Early **www** system released to HEP via the CERN program library. First web servers located in European physics laboratories.

1993: First Mosaic browser; web reaches 500 servers and 1% of Internet traffic; CERN places **WWW** in the public domain.



# Web browser



# What happens when you visit a web page

- HTTP (HyperText Transfer Protocol) request

```
GET http://facebook.com/ HTTP/1.1
Accept: application/x-ms-application, image/jpeg, application/xaml+xml, [...]
User-Agent: Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; WOW64; [...])
Accept-Encoding: gzip, deflate
Connection: Keep-Alive
Host: facebook.com
Cookie: datr=1265876274-[...]; locale=en_US; lsd=WW[...]; c_user=2101[...]
```

- DNS
  - facebook.com -> 66.220.156.25
- TCP/IP
  - Message -> packets -> datagrams
  - Routed to 66.220.156.25 port 80
- HTTP



# What happens when you visit a web page

- HTTP response
  - DNS, TCP/IP
  - HTML (HyperText Markup Language)



```
HTTP/1.1 200 OK
Cache-Control: private, no-store, no-cache, must-revalidate, post-check=0,
pre-check=0
Expires: Sat, 01 Jan 2000 00:00:00 GMT
P3P: CP="DSP LAW"
Pragma: no-cache
Content-Encoding: gzip
Content-Type: text/html; charset=utf-8
X-Cnection: close
Transfer-Encoding: chunked
Date: Fri, 12 Feb 2010 09:05:55 GMT
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en"
lang="en" id="facebook" class=" no_js">
<head>
<meta http-equiv="Content-type" content="text/html; charset=utf-8" />
<meta http-equiv="Content-language" content="en" />
...
```

# What happens when you visit a web page

```
josh@blackbox: ~  
File Edit View Terminal Tabs Help  
josh@blackbox:~$ telnet en.wikipedia.org 80  
Trying 208.80.152.2...  
Connected to rr.pmtpa.wikimedia.org.  
Escape character is '^]'.  
GET /wiki/Main Page http/1.1  
Host: en.wikipedia.org  
  
HTTP/1.0 200 OK  
Date: Thu, 03 Jul 2008 11:12:06 GMT  
Server: Apache  
X-Powered-By: PHP/5.2.5  
Cache-Control: private, s-maxage=0, max-age=0, must-revalidate  
Content-Language: en  
Vary: Accept-Encoding, Cookie  
X-Vary-Options: Accept-Encoding;list-contains=gzip, Cookie;string-contains=enwikiToken;string-contains=enwikiLoggedOut;string-contains=enwiki_session;  
string-contains=centralauth Token;string-contains=centralauth_Session;string-contains=centralauth_LoggedOut  
Last-Modified: Thu, 03 Jul 2008 10:44:34 GMT  
Content-Length: 54218  
Content-Type: text/html; charset=utf-8  
X-Cache: HIT from sq39.wikimedia.org  
X-Cache-Lookup: HIT from sq39.wikimedia.org:3128  
Age: 3  
X-Cache: HIT from sq38.wikimedia.org  
X-Cache-Lookup: HIT from sq38.wikimedia.org:80  
Via: 1.0 sq39.wikimedia.org:3128 (squid/2.6.STABLE18), 1.0 sq38.wikimedia.org:80 (squid/2.6.STABLE18)  
Connection: close  
  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">  
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en" dir="ltr">  
  <head>  
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />  
    <meta name="keywords" content="Main Page,1778,1844,1863,1938,1980 Summer Olympics,2008,2008 Guizhou riot,2008 Jerusal  
    ...  
    ... This content has been removed to save space  
    ...  
    "Non-profit organization">nonprofit</a> <a href="http://en.wikipedia.org/wiki/Charitable_organization" title="Charitable organization">charity</a>.<b  
r /></li>  
    <li id="privacy"><a href="http://wikimediafoundation.org/wiki/Privacy_policy" title="wikimedia:Privacy policy">Privac  
y policy</a></li>  
    <li id="about"><a href="/wiki/Wikipedia:About" title="Wikipedia:About">About Wikipedia</a></li>  
    <li id="disclaimer"><a href="/wiki/Wikipedia:General_disclaimer" title="Wikipedia:General disclaimer">Disclaimers</a>  
</li>  
  </ul>  
</div>  
</div>  
  
  <script type="text/javascript">if (window.runOnloadHook) runOnloadHook();</script>  
<!-- Served by srv93 in 0.050 secs. --></body></html>  
Connection closed by foreign host.  
josh@blackbox:~$
```

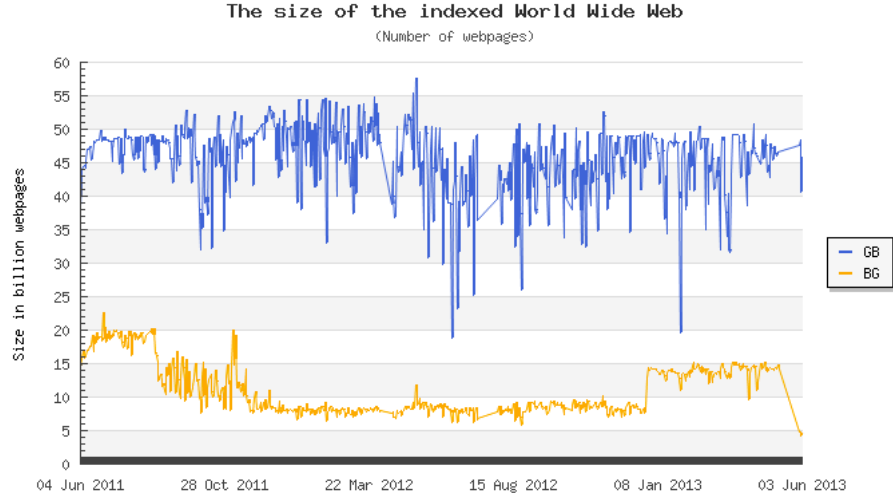
Request

Response headers

Response body



# Web



- Search

- Pagerank – Sergey Brin, Larry Page (Google)
- rank web pages that match a given search string
- assumes that web pages linked from many important pages are themselves likely to be important
- a recursive score for pages, based on the weighted sum of the PageRanks of the pages linking to them



# WEB 2.0

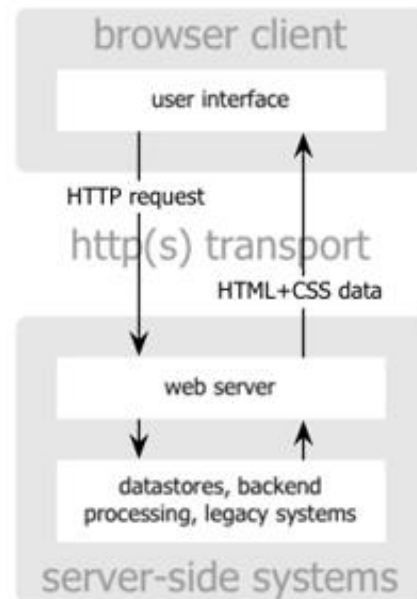


# Web 2.0

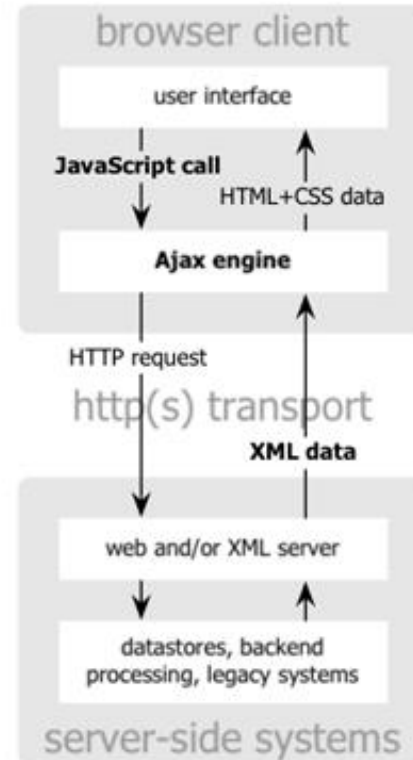


# Web 2.0

- Javascript
- AJAX



classic  
web application model



Ajax  
web application model

# Semantic Web

- group of methods and technologies to allow machines to understand the meaning - or "semantics" - of information on the Web
- Based on
  - Logic models
  - Information retrieval
  - Machine learning
- e.g. Wolfram Alpha (<http://www.wolframalpha.com/>)

# DISTRIBUTED SYSTEMS



# Distributed systems

- Distributed system
  - Web (WIKIPEDIA)
  - Torrent
  - Distributed computing system:
    - Data intensive (Hadoop, map-reduce)
    - CPU intensive (Condor)


# Distributed systems

- Common properties
  - Fault tolerance
    - When one or some nodes fails, the whole system can still work fine except performance.
    - Need to check the status of each node
    - Each node play partial role
    - Each computer has only a limited, incomplete view of the system. Each computer may know only one part of the input.
  - Resource sharing
    - Each user can share the computing power and storage resource in the system with other users
  - Load Sharing
    - Dispatching several tasks to each nodes can help share loading to the whole system.
  - Easy to expand
    - We expect to use few time when adding nodes. Hope to spend no time if possible.

# Wikipedia

- Web

WIKIPEDIA



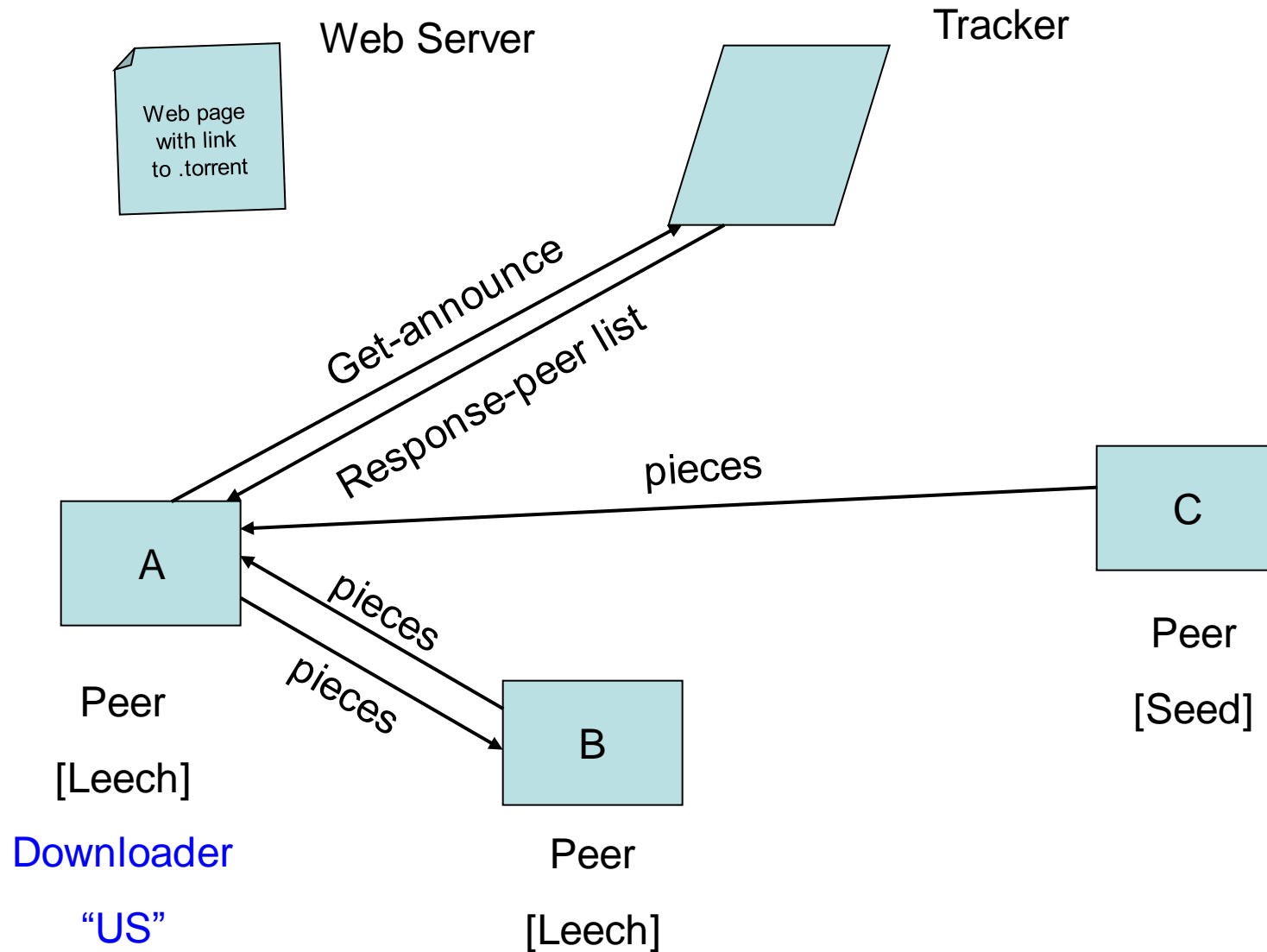
<b>English</b> <i>The Free Encyclopedia</i> 3 413 000+ articles	<b>日本語</b> フリー百科事典 703 000+ 記事
<b>Deutsch</b> <i>Die freie Enzyklopädie</i> 1 122 000+ Artikel	<b>Español</b> <i>La enciclopedia libre</i> 647 000+ artículos
<b>Français</b> <i>L'encyclopédie libre</i> 994 000+ articles	<b>Русский</b> <i>Свободная энциклопедия</i> 586 000+ статей
<b>Italiano</b> <i>L'enciclopedia libera</i> 726 000+ voci	<b>Português</b> <i>A enciclopédia livre</i> 612 000+ artigos
<b>Polski</b> <i>Wolna encyklopedia</i> 728 000+ haseł	<b>Nederlands</b> <i>De vrije encyclopedie</i> 639 000+ artikelen

search • suchen • rechercher • szukaj • ricerca • 検索 • buscar • zoeken • busca •  
поиск • sök • 搜索 • cerca • søk • haku • пошук • keresés • hledání • căutare • ara •  
찾기 • søg • serçu • بحث • cari • tìm kiếm • претрага • suk • hľadaf • paieška • חיפוש •  
търсене • جستجو • poišči • bilnga

English↕→



# Torrent

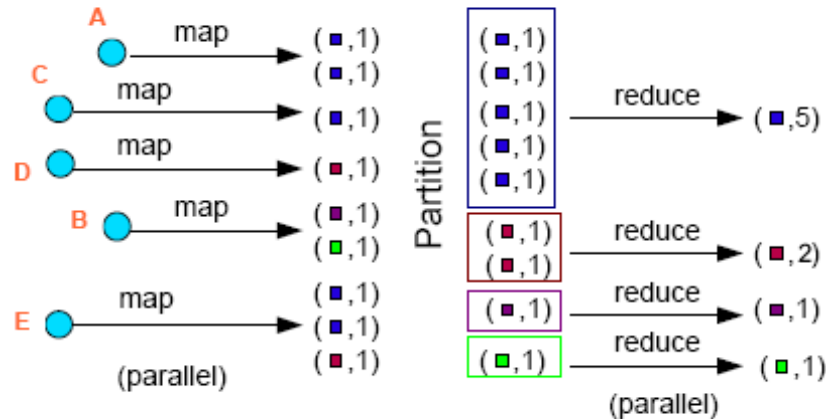


# Hadoop (MapReduce)

## MapReduce Algorithm Example

count the appearances of each different word in a set of documents

Map(A, ●) -> Foreach(■)    Foreach(■)  
                                   n x Map(■,1)    Reduce(■, list<sub>n</sub>(1)) -> Map(■, sum(n))



All documents are processed in parallel

Each document is split up into maps with key = word and value = word count

The reduce function processes the partitions in parallel and returns a result map

Legend:

- document content (text)
- A document name (title)
- single word (different color = different word)
- partition

© Markus Klemm



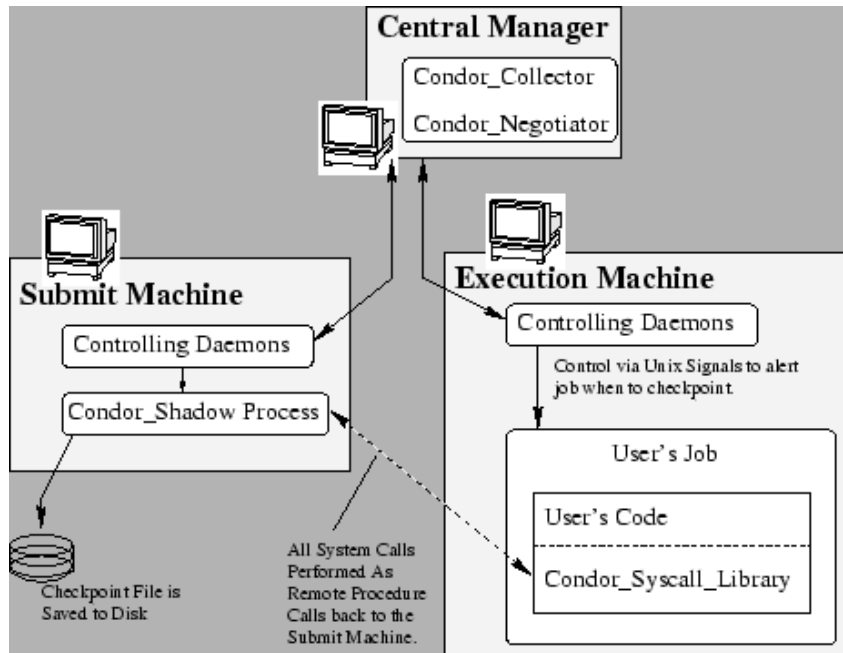
# Condor

- Distributed computing (clusters, farms)



# Condor

- 1992
- Queuing policy
  - Match task and computing nodes
- Resource Classification
  - Each resource can advertise its attributes and master can classify according to this
- 2010 (680k LOC)
  - Apache Web Server: ~60,000 LOC
  - Linux TCP/IP network stack: ~80,000 LOC
  - Windows XP (complete) : ~40 million LOC



# GRID



# What is the Grid?

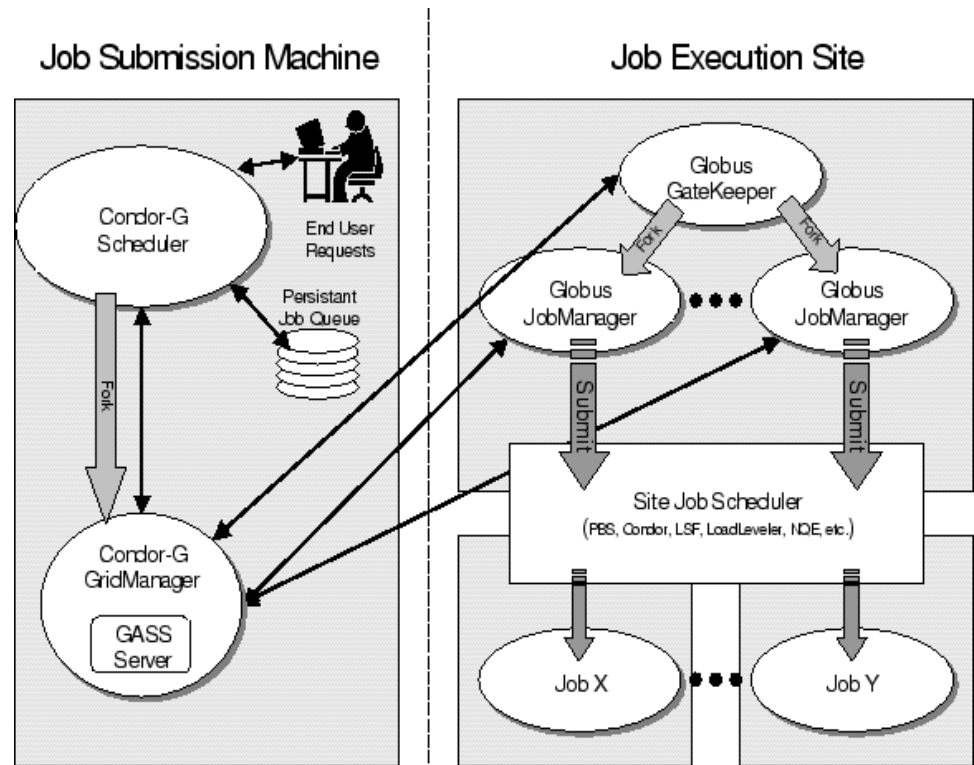
The **World Wide Web** provides seamless access to information that is stored in many millions of different geographical locations

The **Grid** is an infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe



# What is the Grid?

- Distributed computing + Web ;-)
- Request/response
  - Data resource (dataset)
  - Computational job



# Grid history

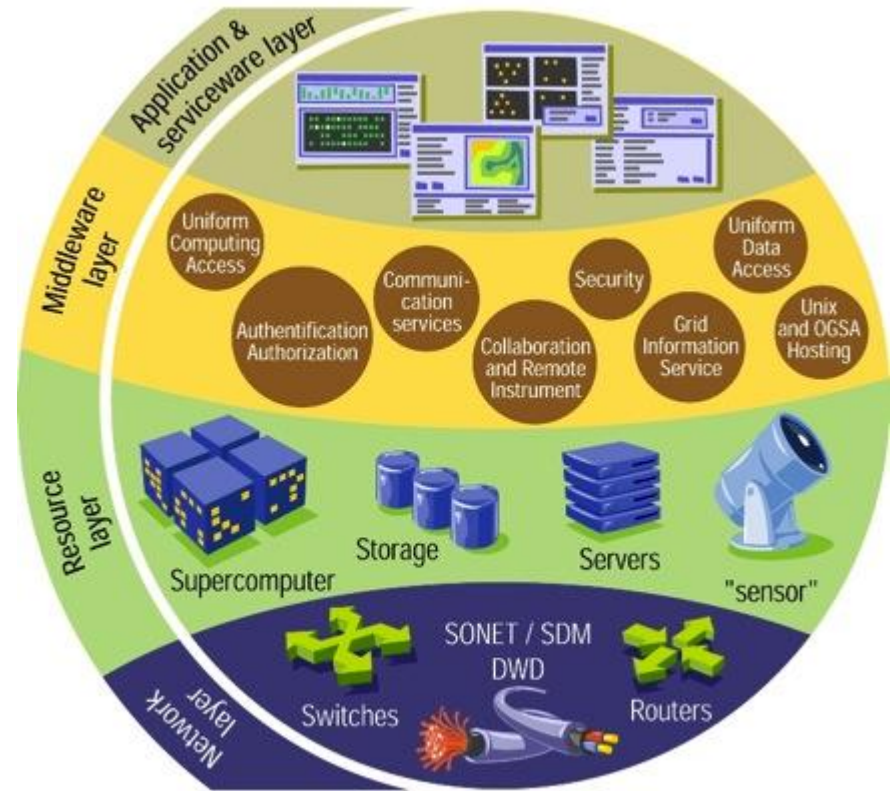
- Name “Grid” chosen by analogy with electric power grid (Foster and Kesselman 1997)
- Vision: plug-in computer for processing power just like plugging in toaster for electricity.
- Concept has been around for decades (distributed computing, metacomputing)
- Key difference with the Grid is to realise the vision on a global scale.





# How does the Grid work?

- It relies on advanced software, called **middleware**.
- Middleware automatically finds the **data** the scientist needs, and the **computing power** to analyse it.
- Middleware balances the load on different resources. It also handles **security**, **accounting**, **monitoring** and much more.



# Virtual Organization

High Energy Physics, Earth Observation and Biology are examples of communities made up of several institutions and individuals sharing the same interests and the same scientific goals. They greatly benefit from putting together their computing resources, data and scientific instruments.

Such distributed communities are called *Virtual Organisations*.

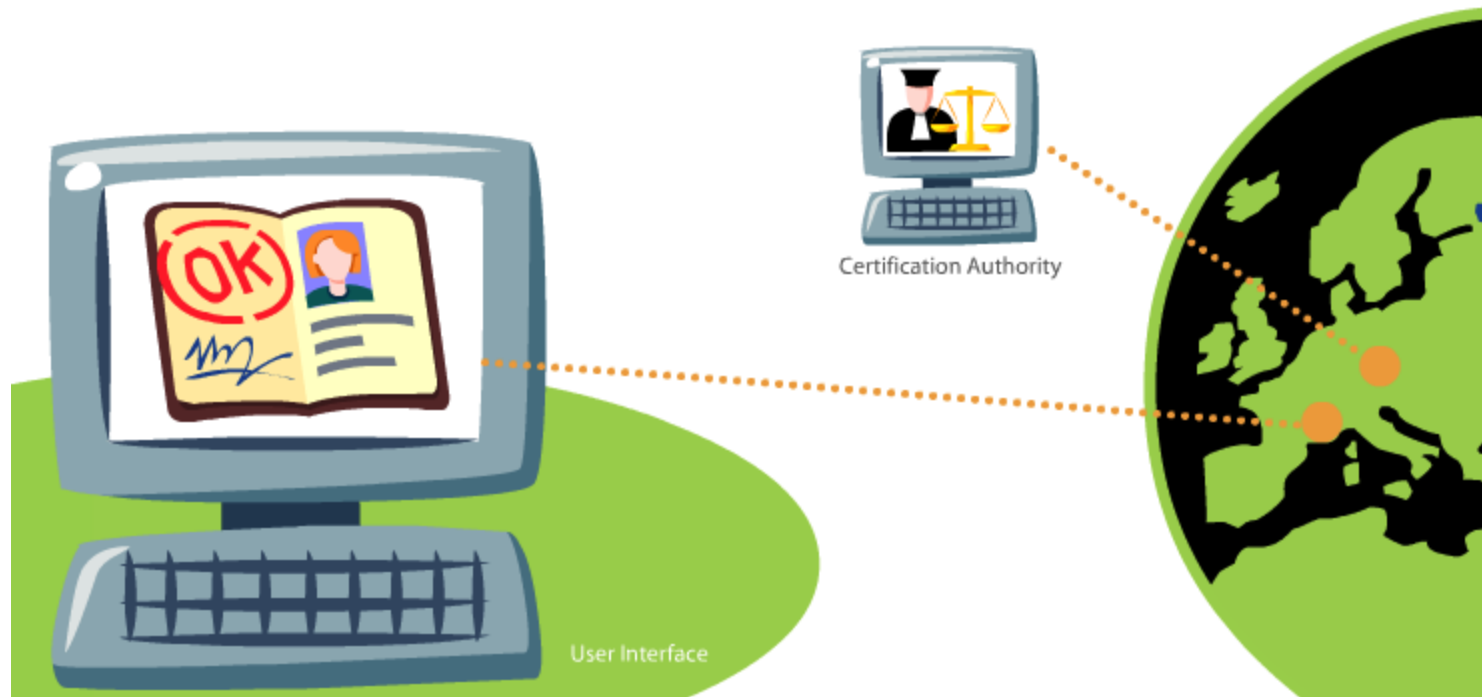


# Step 1: Join Virtual Organization

To have access to the Grid facility, you need to join a *Virtual Organisation*.



## Step 3: Get a permission



## Step 4: Write a file describing your job

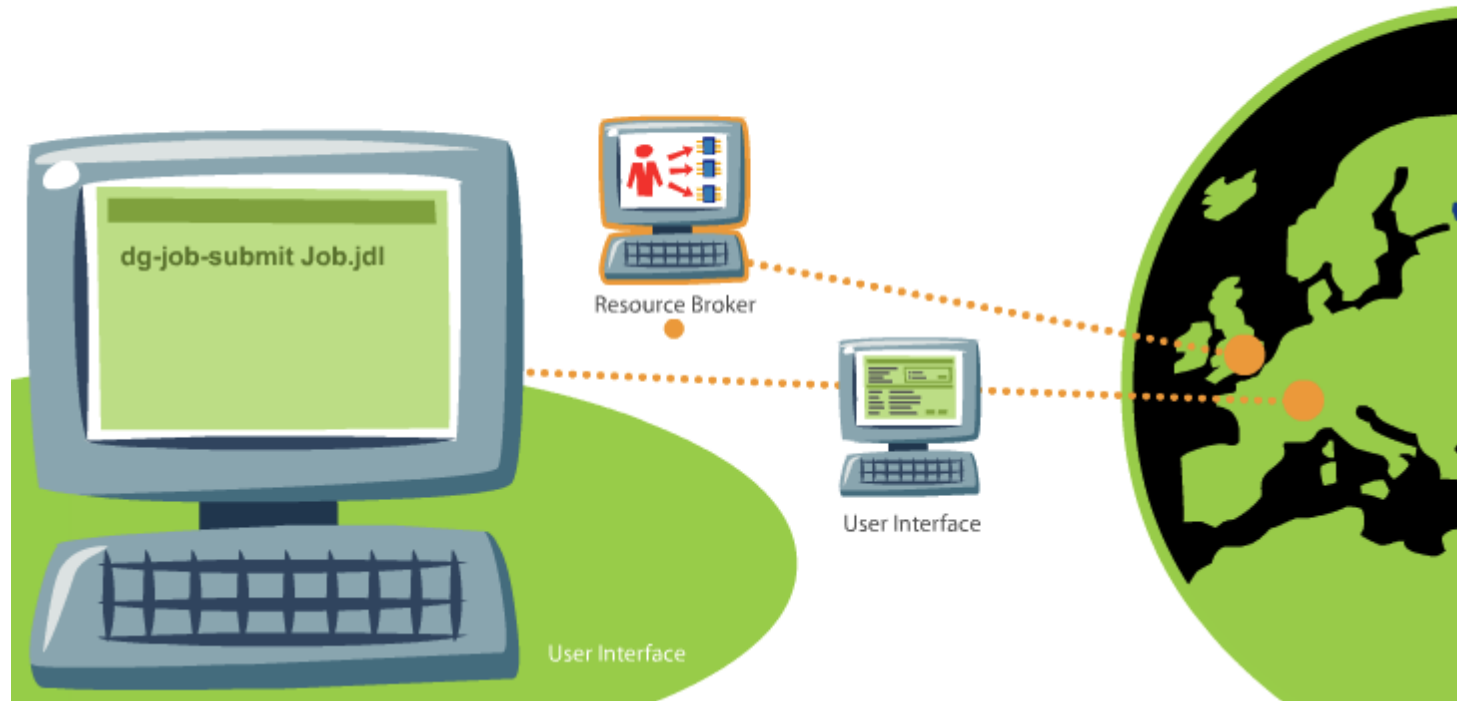
To run a job on the LCG/EGEE Grid Facility, you have to describe it in the **Job Description Language (JDL)**.

**JDL** specifies job characteristics such as the application to use, the input data, the required resources, etc.



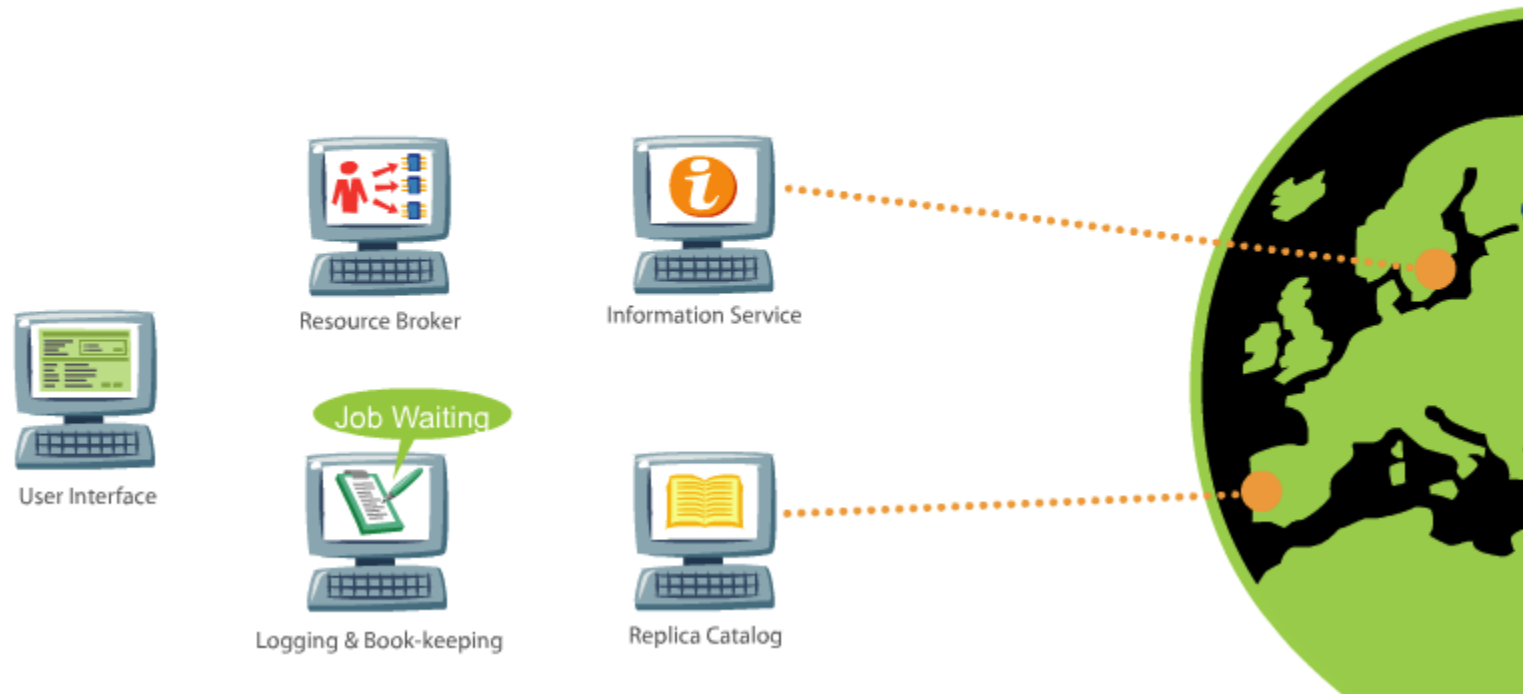
# Step 5: Submit your job to the resource broker

Once you have the jdl file for your job, you can submit it to the *Resource Broker*.



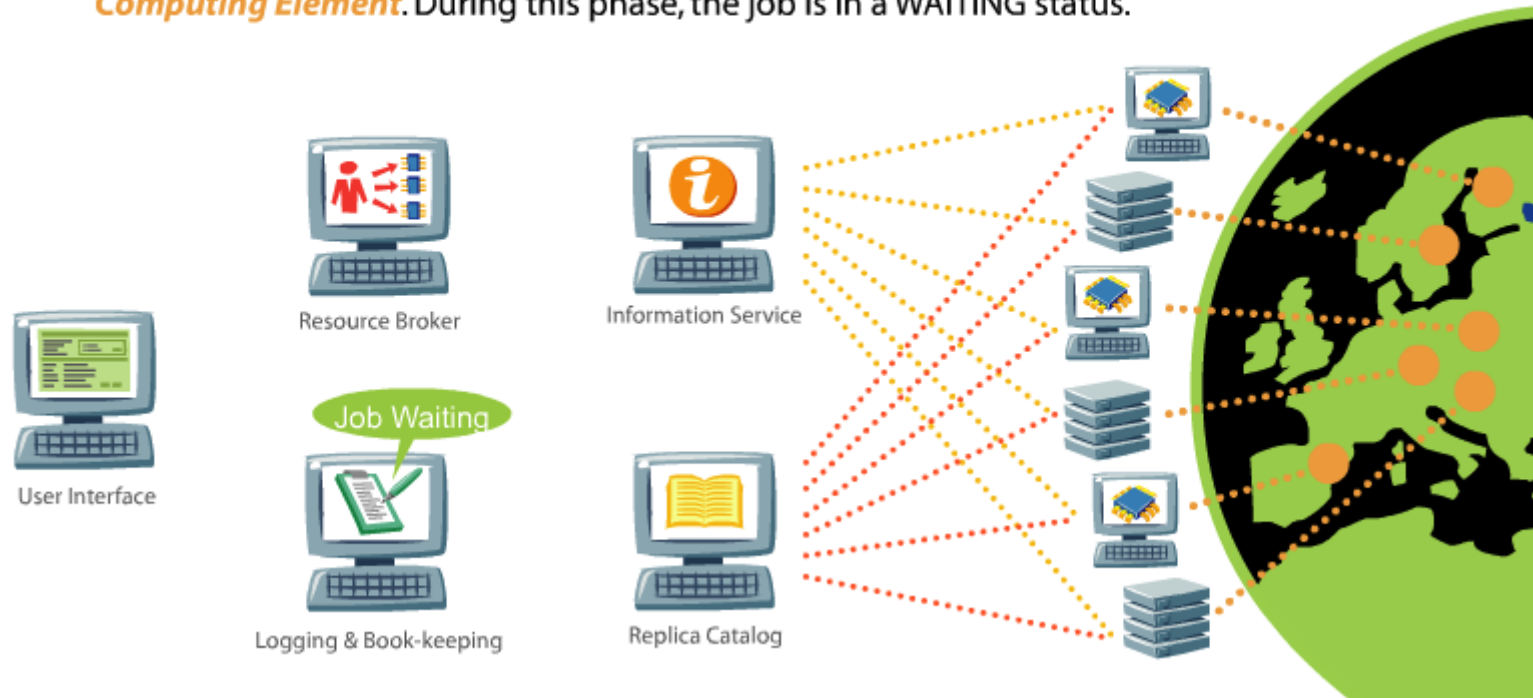
# Step 5: Submit your job to the resource broker

Based on the information given in the JDL file, the **Resource Broker** queries the **Information Service** and the **Replica Catalog** to check resources.



# Step 5: Submit your job to the resource broker

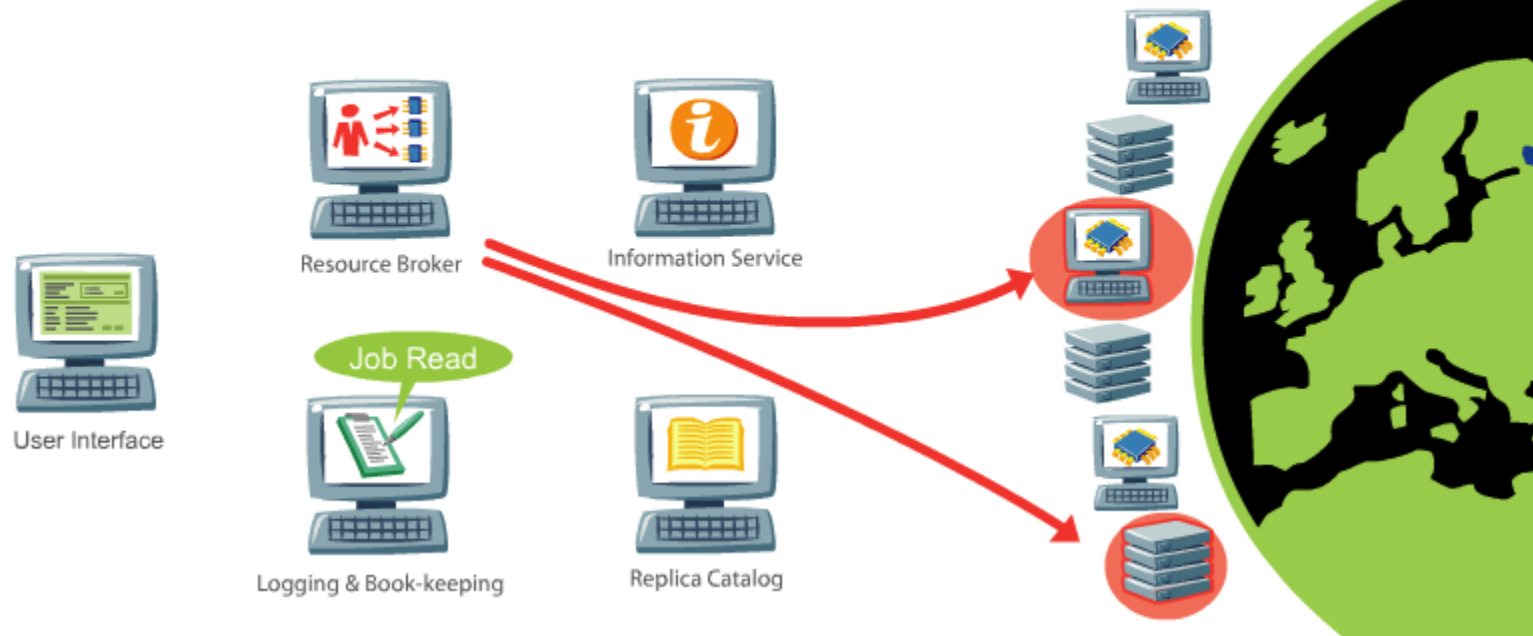
The **Replica Catalog** and the **Information Service** hold information on the current status of all the sites. The Resource Broker uses this info to match the job to a suitable **Computing Element**. During this phase, the job is in a **WAITING** status.





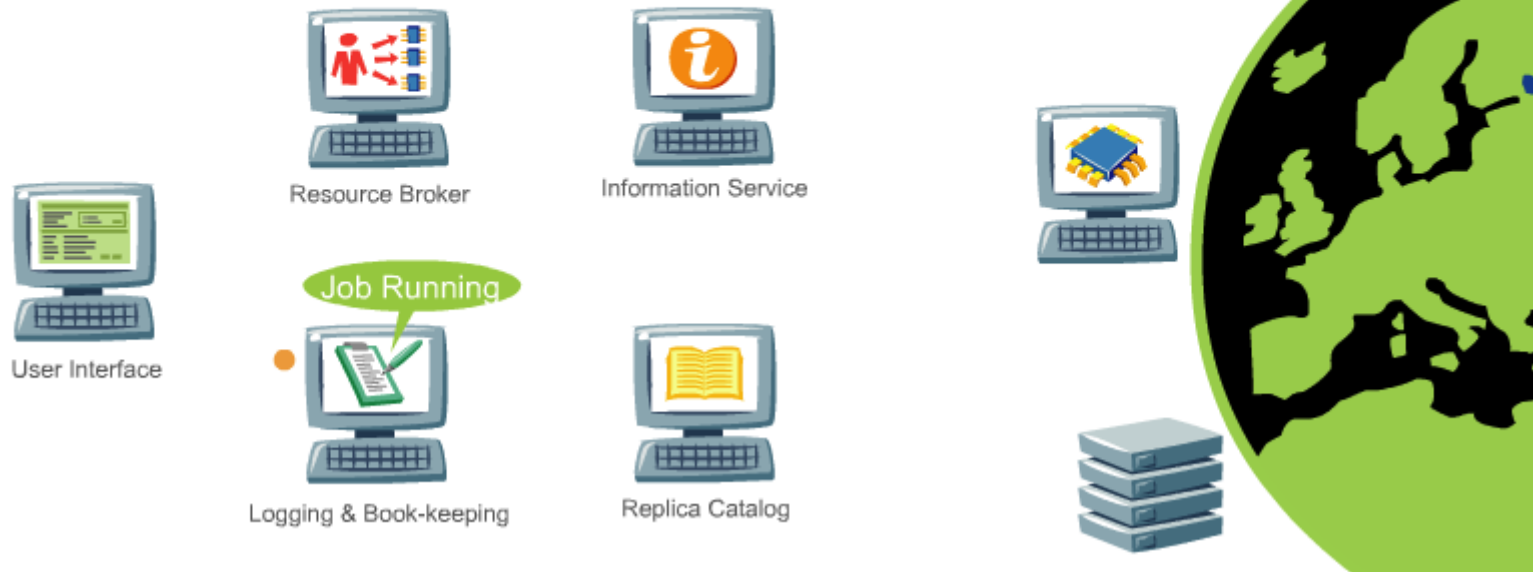
# Step 5: Submit your job to the resource broker

The Resource Broker *makes its choice*. It has found a suitable Computing Element and the Storage Element with the necessary data. It informs the logging & Book-keeping service of its decision.



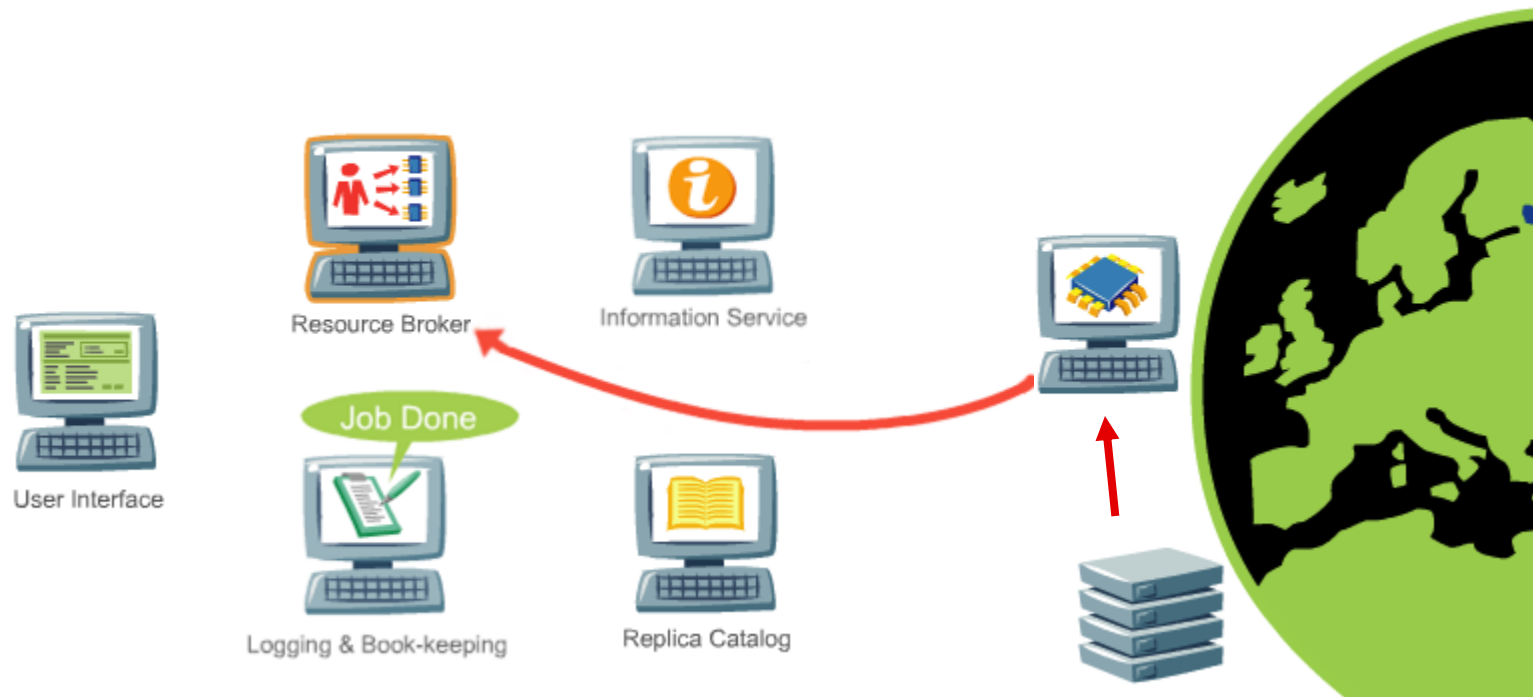
# Step 6: Check the status of your job

During all this process, you can check the status of your job by contacting the **Logging & Book-keeping Service**.



# Step 6: Check the status of your job

The execution of the job has completed on the Computing Element.  
The *Computing Element transfers the output to the Resource Broker*.



# Step 7: Get the results

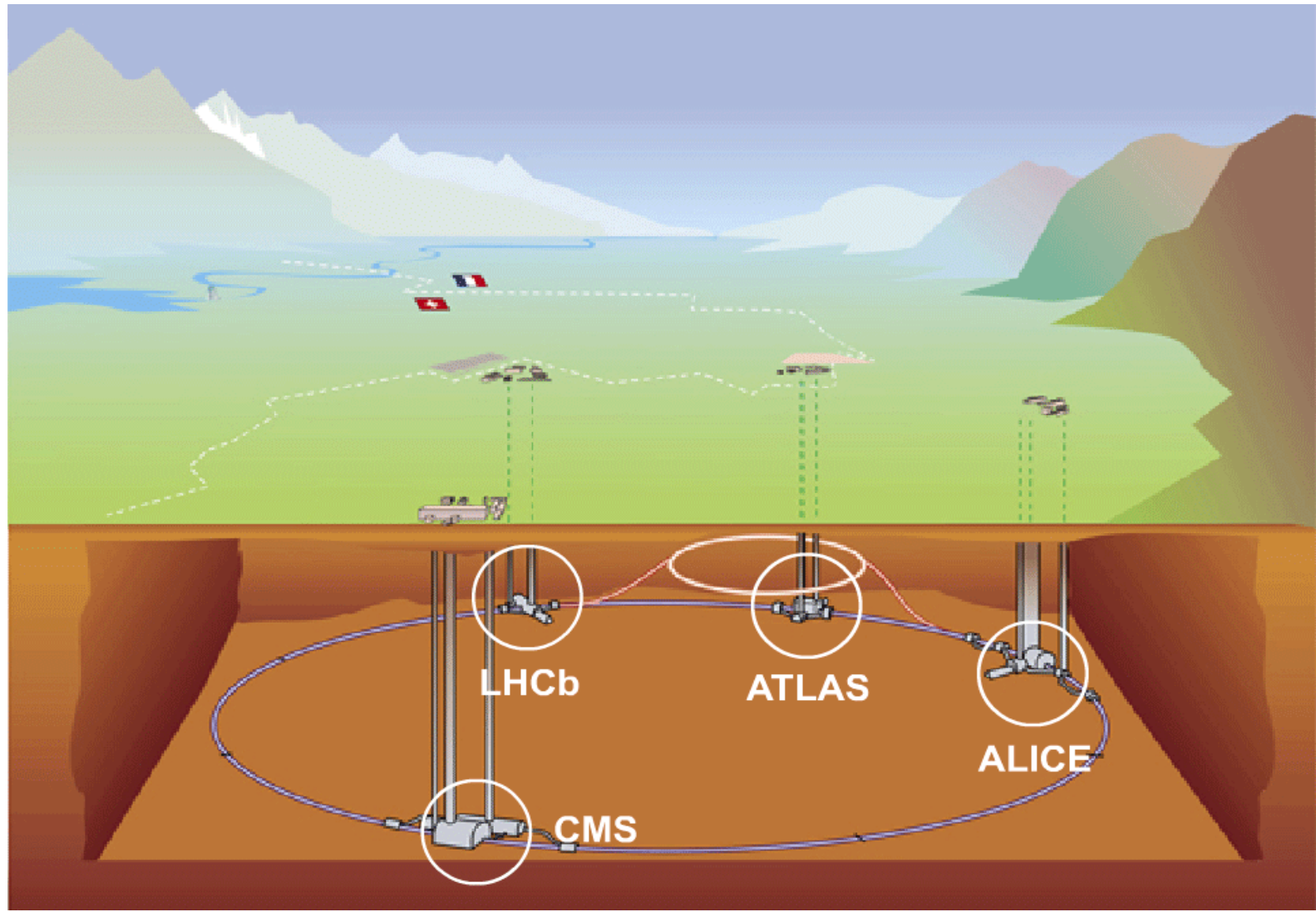
You can now retrieve your Output from the Resource Broker.  
When finished, book-keeping *information is purged*.



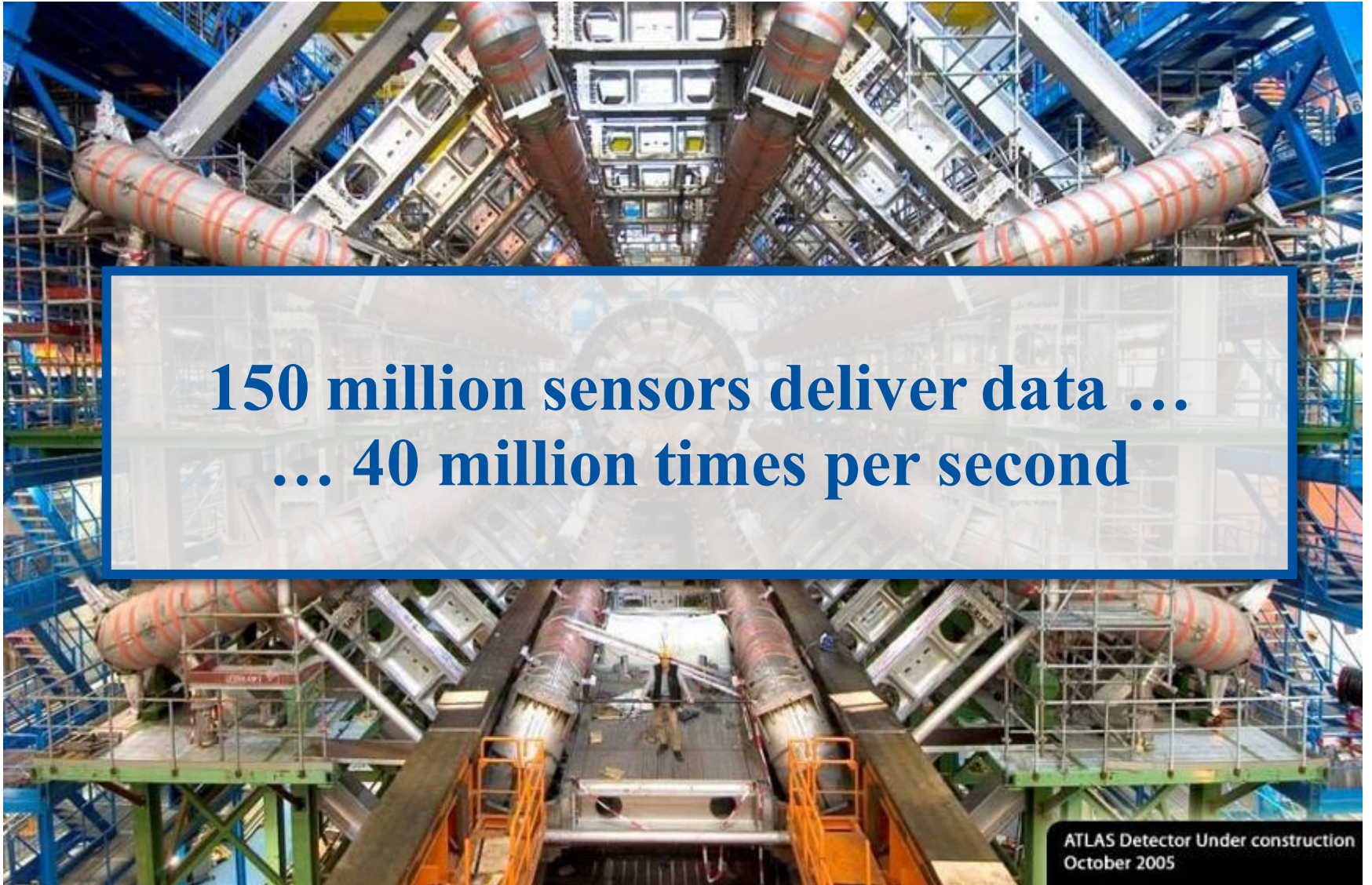
# Why the Grid at CERN?



# The LHC accelerator and the 4 experiments

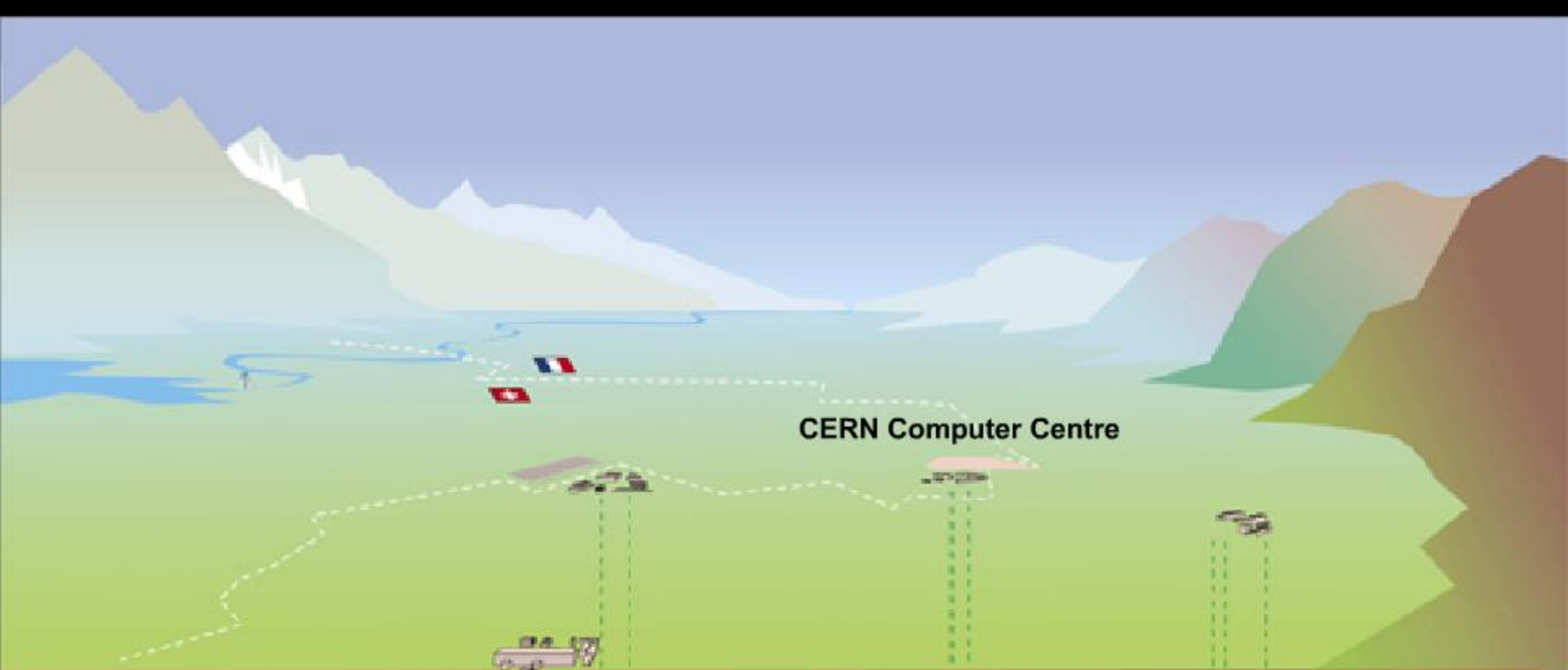


# View of the ATLAS detector

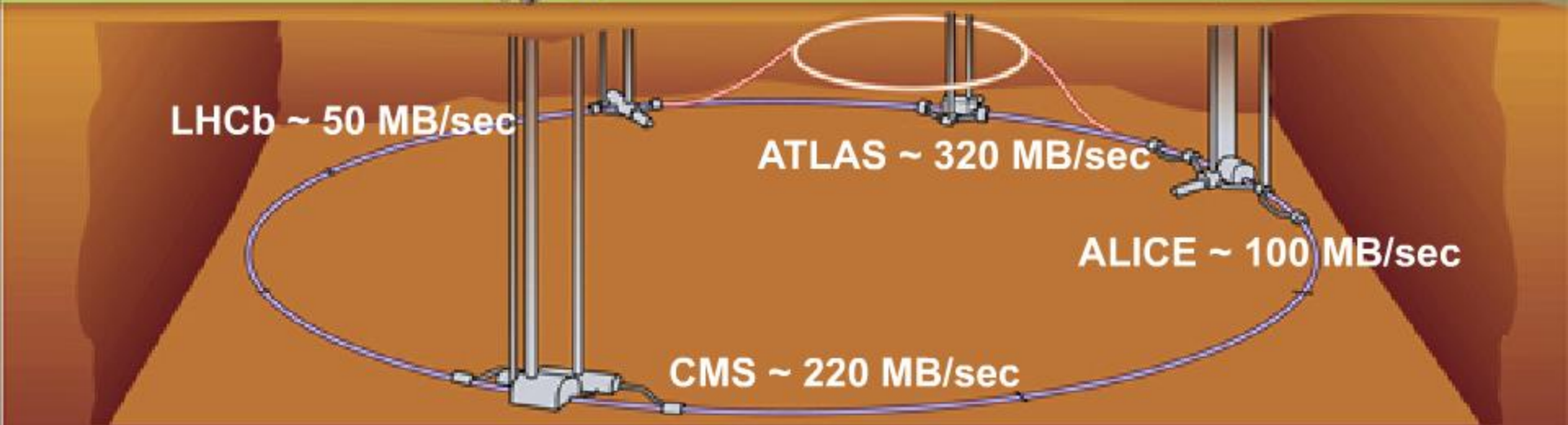


**150 million sensors deliver data ...  
... 40 million times per second**

ATLAS Detector Under construction  
October 2005



CERN Computer Centre



LHCb ~ 50 MB/sec

ATLAS ~ 320 MB/sec

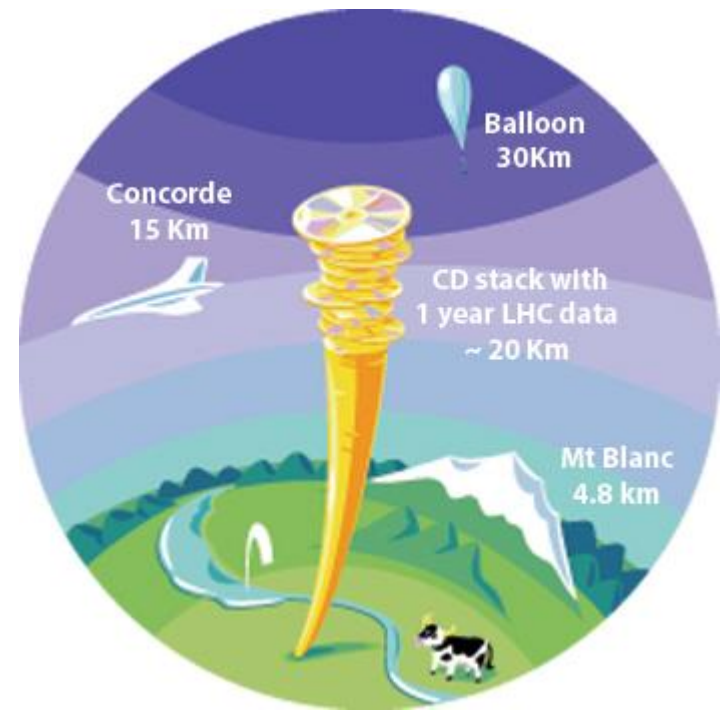
ALICE ~ 100 MB/sec

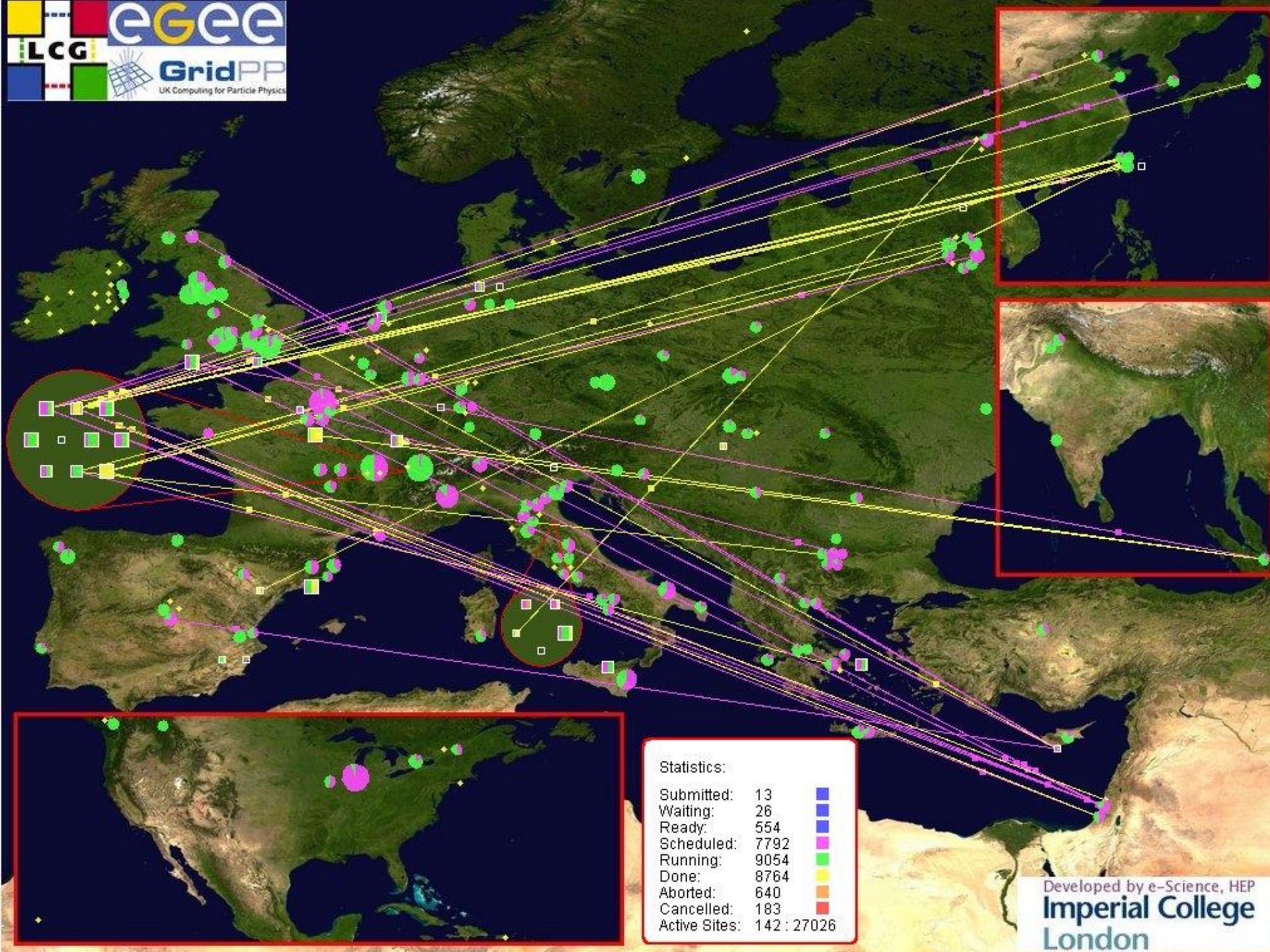
CMS ~ 220 MB/sec



# The LHC Data Challenge

- The accelerator will run for 10-15 years
- Experiments will produce about **15 Million Gigabytes** of data each year (about 20 million CDs!)
- LHC data analysis requires a computing power equivalent to **~100,000 of today's fastest PC processors**
- Requires many cooperating computer centres, as CERN can **only** provide **~20% of the capacity**

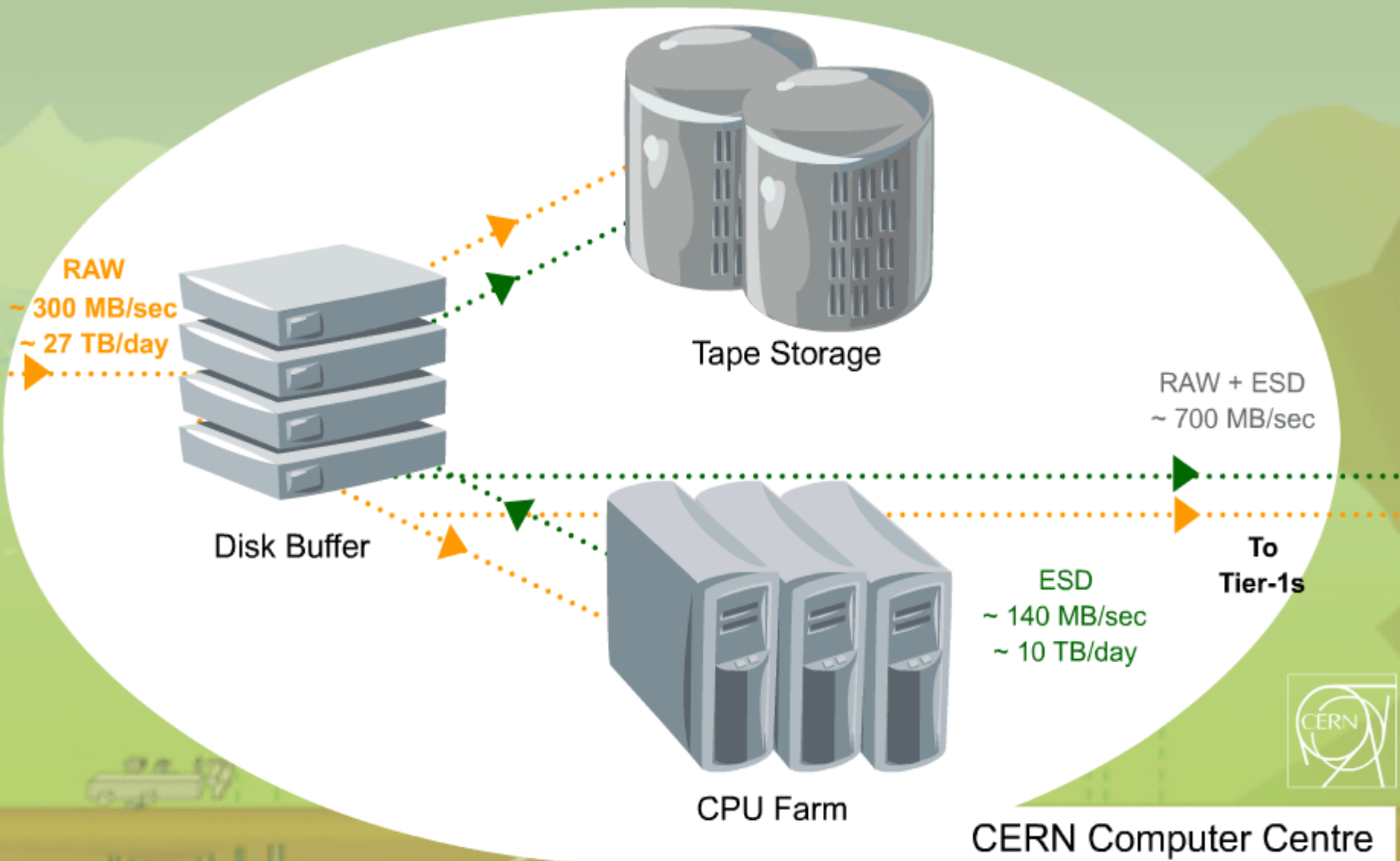




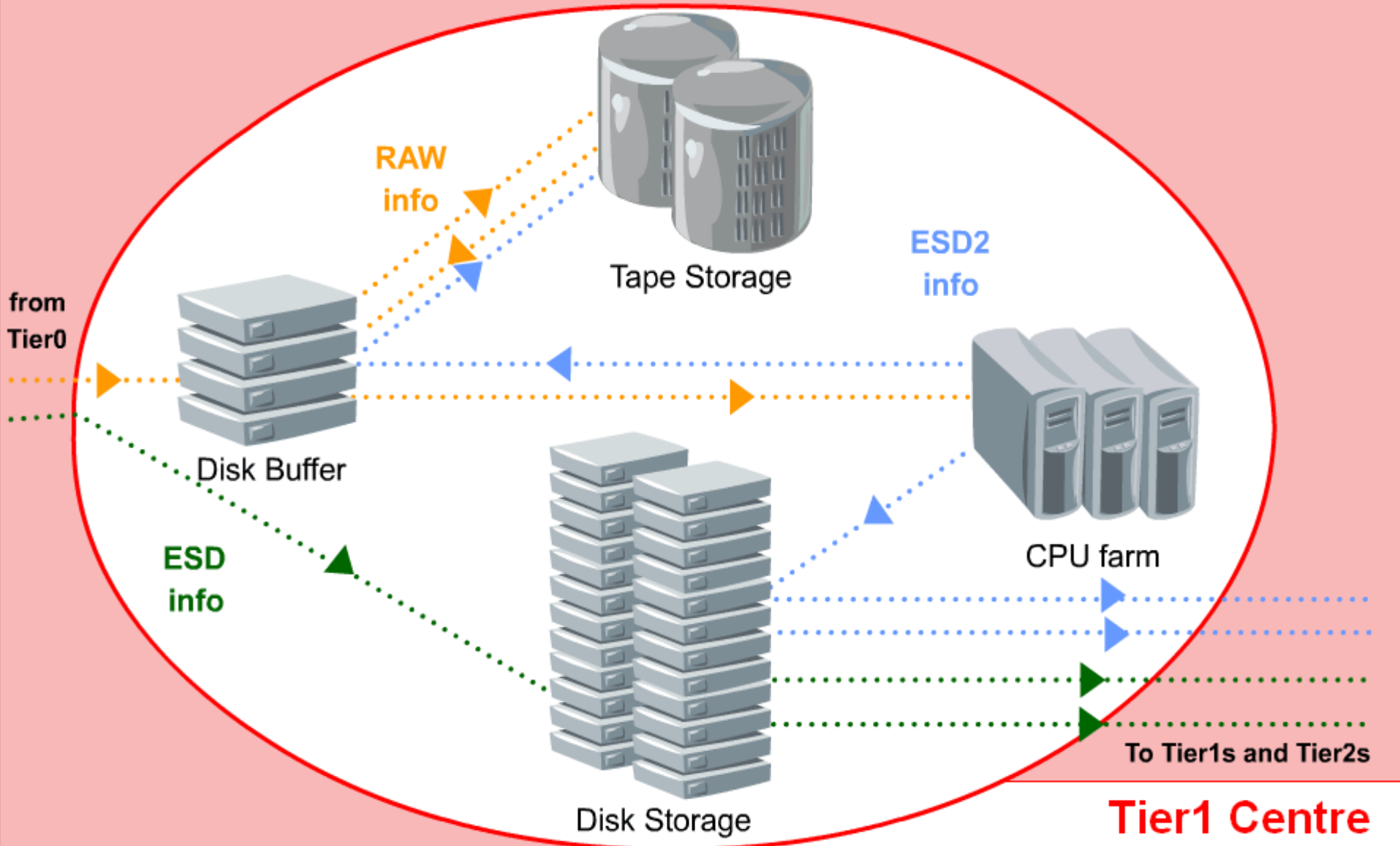
Statistics:

Submitted:	13	<span style="color: blue;">■</span>
Waiting:	26	<span style="color: purple;">■</span>
Ready:	554	<span style="color: green;">■</span>
Scheduled:	7792	<span style="color: yellow;">■</span>
Running:	9054	<span style="color: orange;">■</span>
Done:	8764	<span style="color: red;">■</span>
Aborted:	640	<span style="color: grey;">■</span>
Cancelled:	183	<span style="color: grey;">■</span>
Active Sites:	142 : 27026	

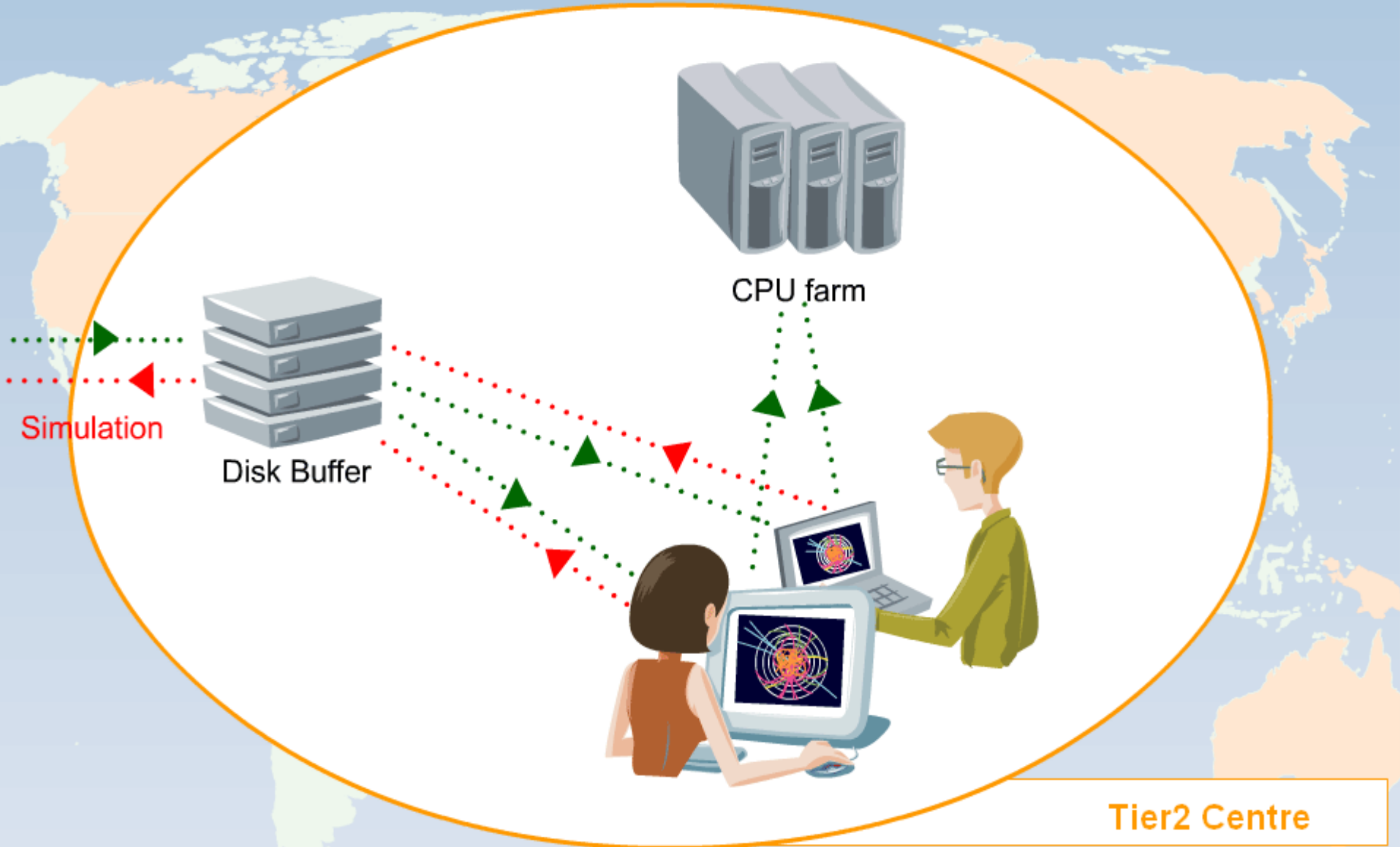
# One Tier0 Centre (CERN)



# Eleven Tier1 Centres



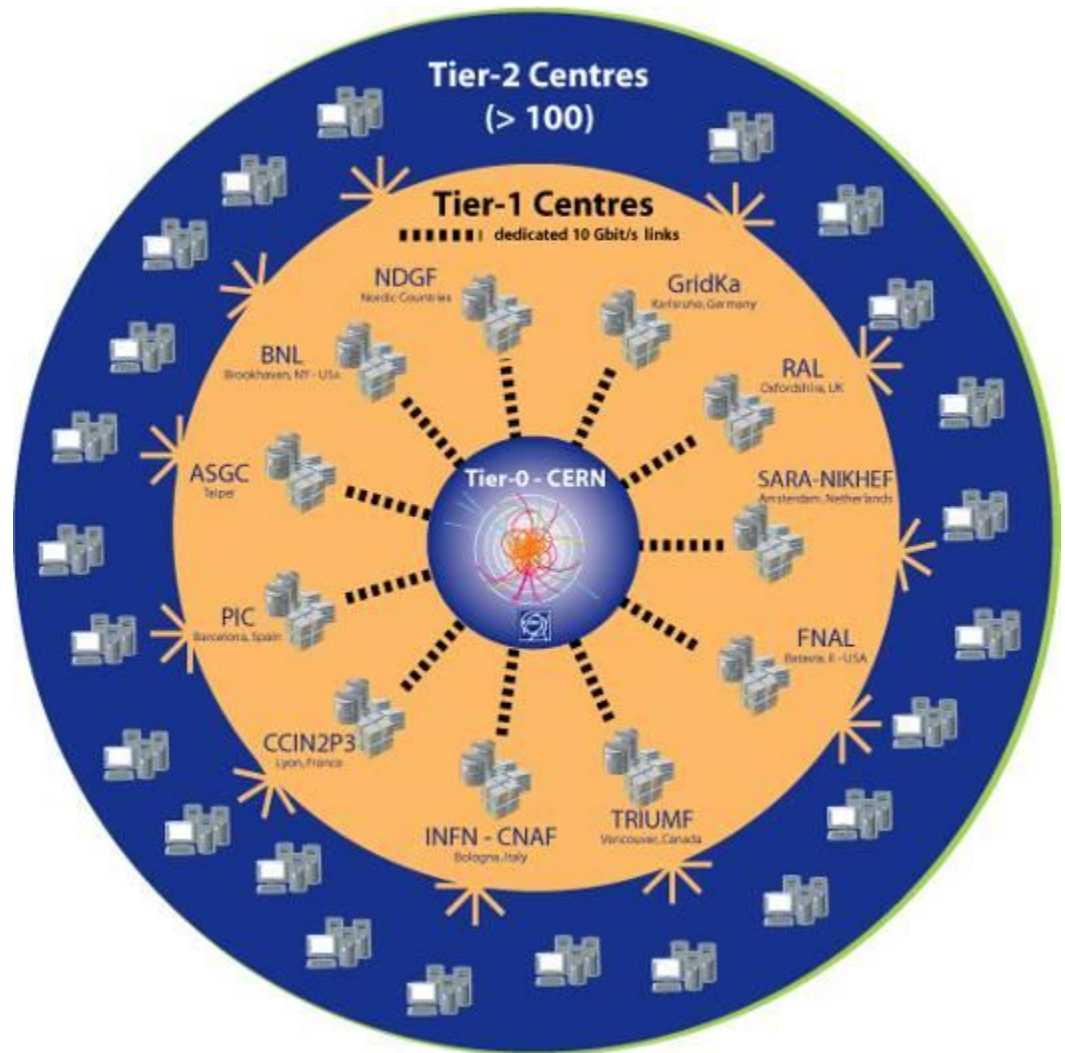
# >100 Tier2 Centres



Tier2 Centre

# LHC Computing Grid project (LCG)

- More than 170 computing centres
- 12 large centres for primary data management: CERN (Tier-0) and eleven Tier-1s
- 38 federations of smaller Tier-2 centres
- 35 countries involved



# WLCG: Worldwide LHC Computing Grid

- Project to build and maintain data storage and computing infrastructure for LHC
- Uses infrastructure of several Grid organizations where 2 the biggest ones are
  - EGI (founded by EC)
  - OSG (founded by US)



# Resources needed for the LHC computing

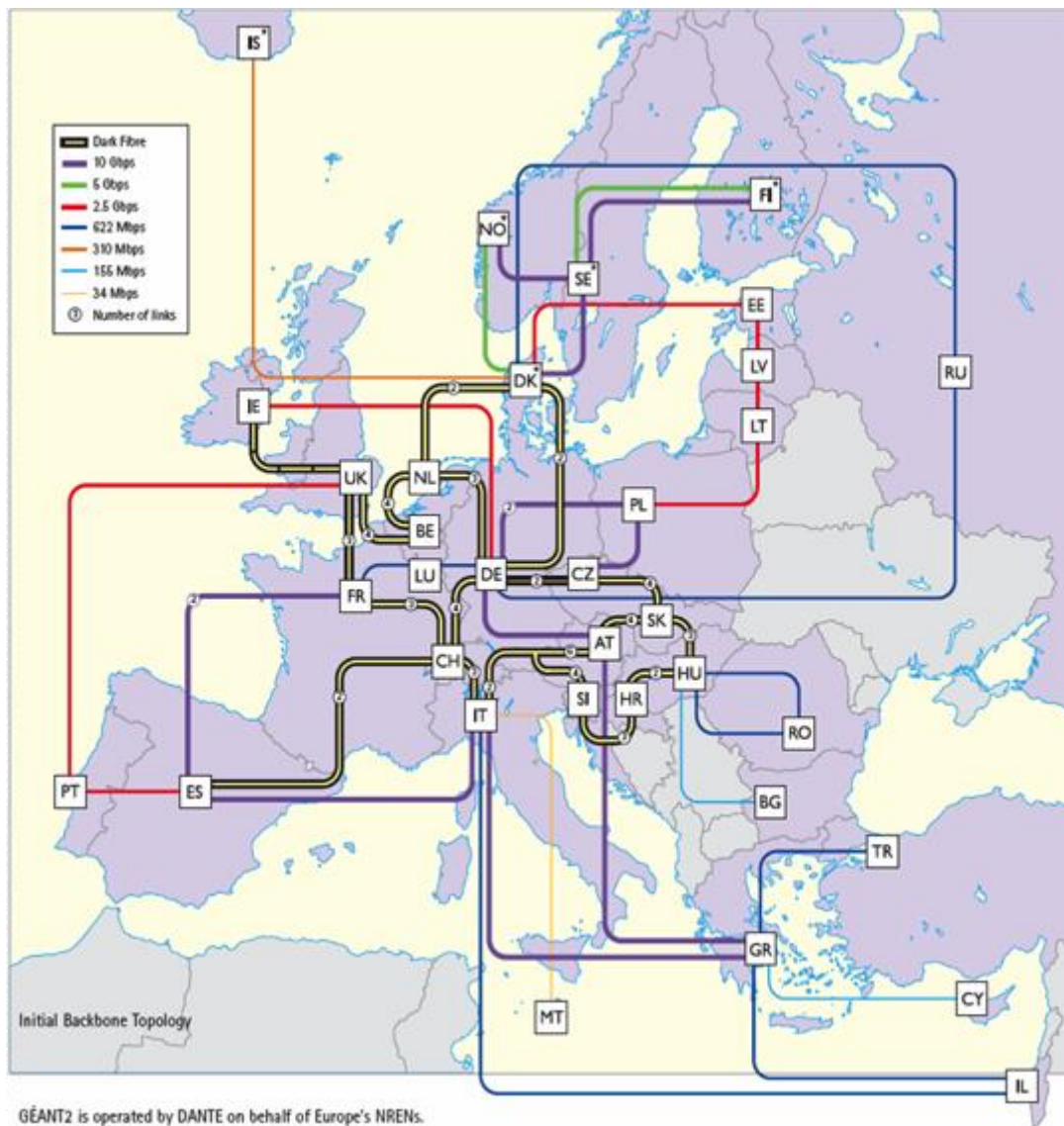
Name	Sites	CPUs			Online Storage Space (GB)		Nearline Storage Space (GB)		Grid Jobs		
		Physical	Logical	SI2000	TotalSize	UsedSize	TotalSize	UsedSize	Total	Running	Waiting
AEGIS	6	312	1,056	2,305,920	27,761	 35%	0	 0%	576	 54%	 0%
ArmGRID	4	36	144	345,600	2,179	 13%	0	 0%	12	 2%	 75%
BALTICGRID	3	56	92	102,684	0	 0%	0	 0%	11	 10%	 9%
BIGGRID	14	4,017	17,152	41,226,064	4,204,350	 50%	1,265,011	 90%	13,973	 65%	 19%
Consorzio Cometa	6	684	1,368	2,243,288	115,679	 34%	0	 0%	763	 0%	 98%
D-Grid	11	5,518	21,802	47,901,980	10,631,901	 38%	0	 0%	68,635	 250%	 20%
DE-KIT	1	108	216	491,712	167,405	 33%	0	 0%	16	 6%	 12%
DECH	1	112	448	879,424	86,995	 0%	0	 0%	174	 38%	 1%
EELA	7	115	317	515,295	7,144	 10%	0	 0%	81	 17%	 18107008%
EELA2	1	152	344	1,037,504	35,641	 0%	0	 0%	0	 0%	 0%
EGEE	248	56,492	167,714	373,956,582	60,176,251	 54%	60,083,689	 65%	274,805	 124%	 45315%
EGEE IBERGRID	1	128	512	1,143,296	73	 12%	0	 0%	3	 0%	 0%
EGI	59	13,711	48,770	106,364,766	13,836,881	 41%	40,000	 43%	81,129	 121%	 19200%
EUFORIA	1	152	344	1,037,504	35,641	 0%	0	 0%	0	 0%	 0%
EUMED	2	227	668	1,224,960	0	 0%	0	 0%	2	 0%	 399999650%
GILDA	3	86	120	137,616	545	 19%	0	 0%	78	 1%	 97%
GRID-CSIC	1	0	0	0	322	 100%	0	 0%	0	 0%	 0%
GRIDPP	15	3,569	11,309	23,138,493	4,037,221	 57%	0	 0%	15,078	 69%	 353763%
GRISU	8	755	1,894	2,978,876	120,491	 32%	0	 0%	1,174	 20%	 66%
<a href="http://www.euasiagrid.org">http://www.euasiagrid.org</a>	1	12	48	790,176	131	 8%	0	 0%	1	 2%	 0%
I2G	2	451	1,780	3,489,248	539,422	 48%	0	 0%	960	 53%	 0%
IBERGRID	7	987	3,438	6,864,770	556,196	 45%	0	 0%	1,028	 29%	 648508%
INGRID	2	317	1,258	2,153,914	281,304	 45%	0	 0%	844	 66%	 0%
LCG	1	1	2	762	72	 30%	0	 0%	0	 0%	 0%
LITGRID	2	44	44	16,764	0	 0%	0	 0%	1	 0%	 100%
<b>Total</b>	<b>599</b>	<b>158,633</b>	<b>473,428</b>	<b>1,053,365,254</b>	<b>179,214,559</b>	<b>92,301,318</b>	<b>132,426,147</b>	<b>85,373,578</b>	<b>817,785</b>	<b>612,548</b>	<b>354,444,881</b>





# The Géant Research Network

- General purpose network connecting national research and education nets (NREN)
- For LHC in addition: Optical Private Network with 10Gb/s links over dark fibres



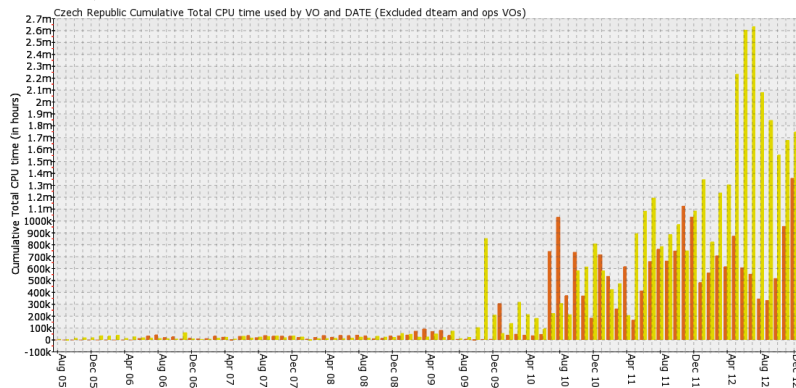
# MetaCentrum – Czech Grid Initiative

- Activity of CESNET association
- <http://www.metacentrum.cz/>
- Supporting ALICE, ATLAS, Auger, VOCE, EUAsiaVO, etc.
- Tier 2 Sites:
  - PragueLCG2
  - Prague\_CESNET\_LCG2
- Participating institutes:
  - CESNET
  - Masaryk University Brno
  - Charles University
  - West Bohemian University in Pilsen
  - Brno University of Technology
  - South Bohemian University in Ceske Budejovice

# Czech Grid Initiative Activity

Chart showing the Cumulative Total CPU time used grouped by VO and DATE (only information about LHC VOs (Excluded dteam and ops VOs) is returned).  
Developed by CESGA EGI View: / sumcpu / 2005E-2012I2 / VO-DATE / lhc (x) / GRBAR-LIN / x

2013-06-03 09:23

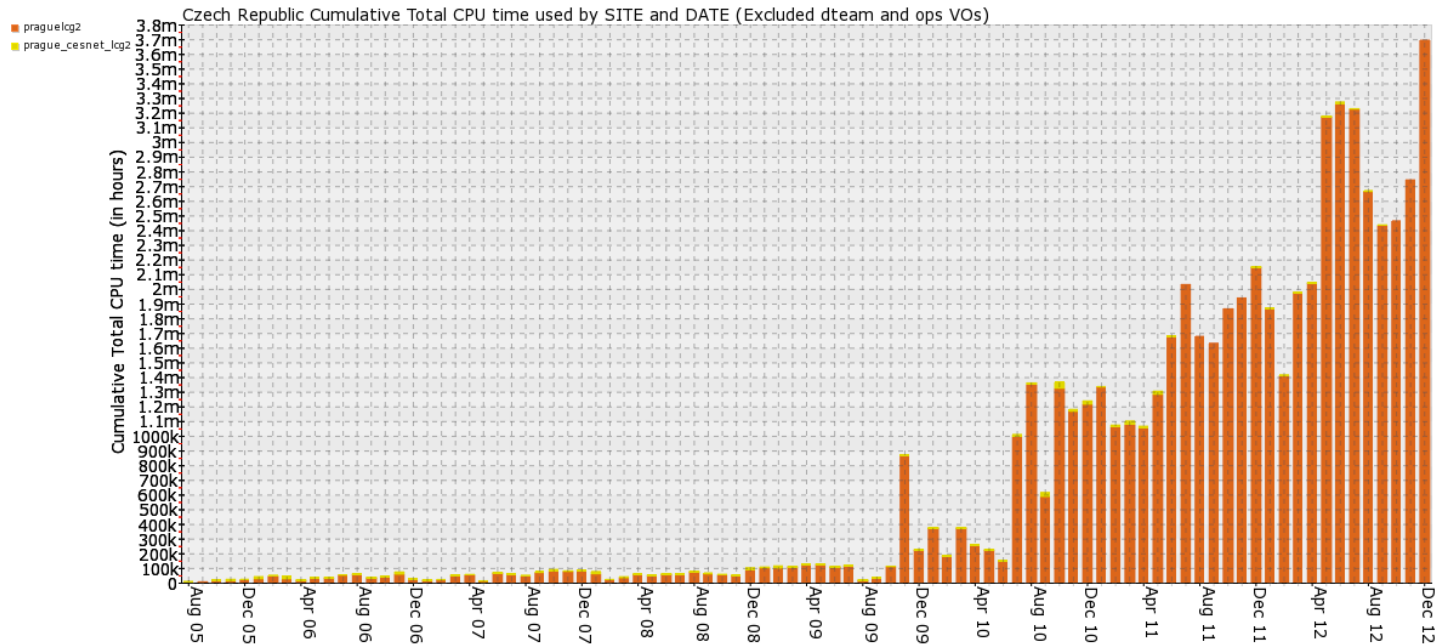


- Czech Grid ran ~ 5.7M jobs in 2012 – workload has continued to increase

Chart showing the Cumulative Total CPU time used grouped by SITE and DATE (only information about TOP 10 -ordered by CPU time- VOs (Excluded dteam and ops VOs) is returned).

Developed by CESGA EGI View: / sumcpu / 2005E-2012I2 / SITE-DATE / top10 (x) / ACCBAR-LIN / x

2013-06-03 09:23



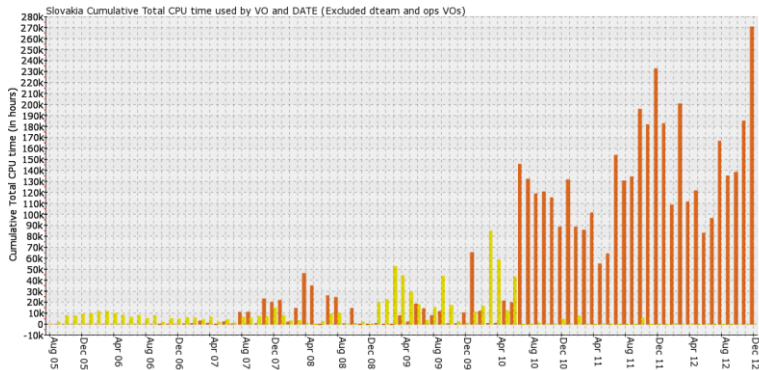
# Slovak Grid Infrastructure

- Slovakia is a federated Tier-2/3 HEP community
  - Slovak Grid Initiative / [www.slovakgrid.sk](http://www.slovakgrid.sk)
- .Tier 2 sites:
    - FMPHi-UNIBA – ATLAS, ALICE
    - IEPSAS-Kosice – ATLAS, ALICE, CDF, H1
  - .TU-Kosice – VOCE, BIOMED
  - .IISAS-Bratislava – ESR, VOCE, BIOMED



# Slovak Grid activity

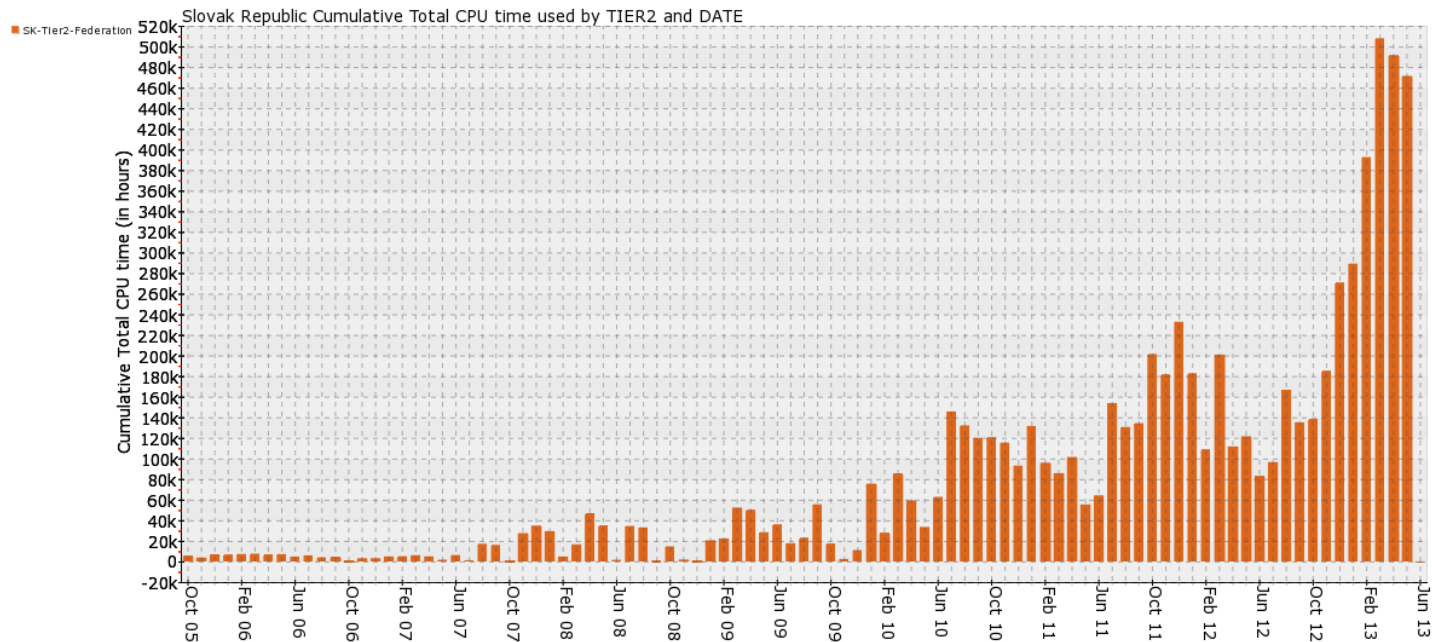
Chart showing the Cumulative Total CPU time used grouped by VO and DATE (only information about LHC VOs (Excluded dteam and ops VOs) is returned).  
Developed by CESGA EGI View: / sumcpu / 2005-2012 / VO-DATE / lhc (x) / GRBAR-LIN / x



- Slovak Grid ran ~ 927k jobs in 2012 – workload has continued to increase
- Distribution of work across Tier 2/Tier3 really illustrates the importance of the grid system

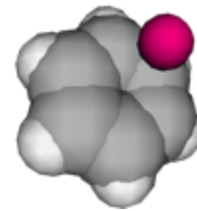
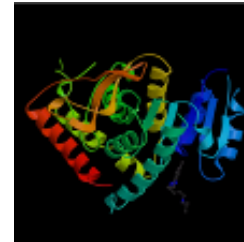
Chart showing the Cumulative Total CPU time used grouped by TIER2 and DATE (only information about TOP 10 -ordered by CPU time- VOs is returned).  
Developed by CESGA EGI View: / sumcpu / 2005-10-2013 / TIER2-DATE / top10 (x) / GRBAR-LIN / i

2013-06-01 22:28

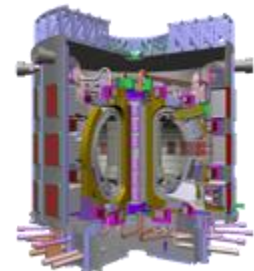
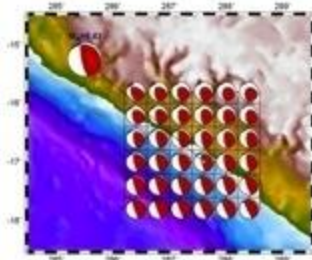


# EGEE: Scientific disciplines

- Growth and diversification of applications.
- Reported apps. only  $\Rightarrow$  *underestimate!*



	6/2006	2/2007	1/2008
<b>Astron. &amp; Astrophysics</b>	2	8	9
<b>Computational Chemistry</b>	6	27	21
<b>Earth Science</b>	16	16	18
<b>Fusion</b>	2	3	4
<b>High-Energy Physics</b>	9	11	7
<b>Life Sciences</b>	23	39	37
<b>Others</b>	4	14	21
<b>Total</b>	<b>62</b>	<b>118</b>	<b>117</b>

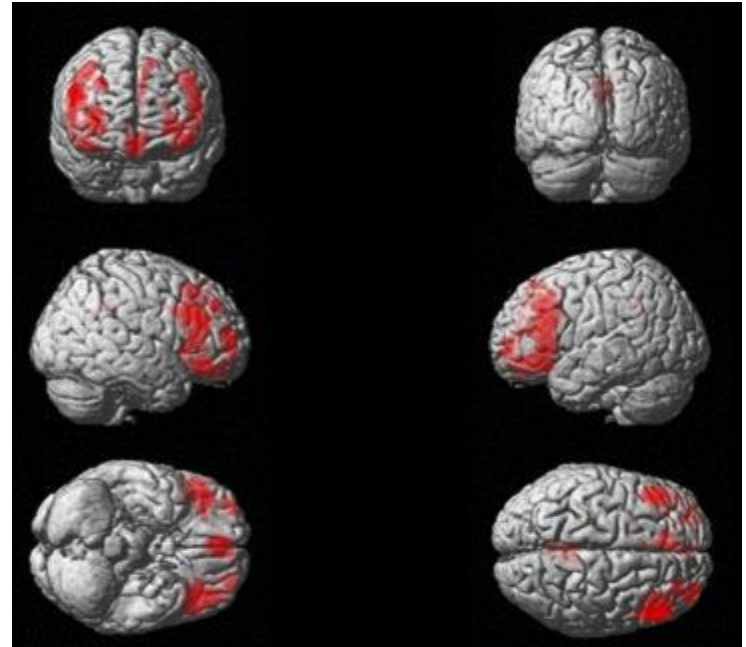


Condensed Matter Physics  
 Comp. Fluid Dynamics  
 Computer Science/Tools  
 Civil Protection  
 Finance



# SPM: Alzheimer's disease evaluation

- Statistical Parametric Mapping (SPM) – powerful but difficult to use
  - Need of huge amount of data → sharing instead of building own repositories
  - Computational power needed to achieve results fast enough

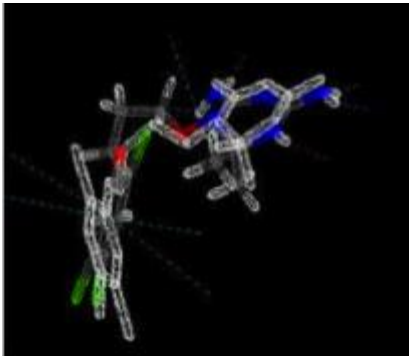
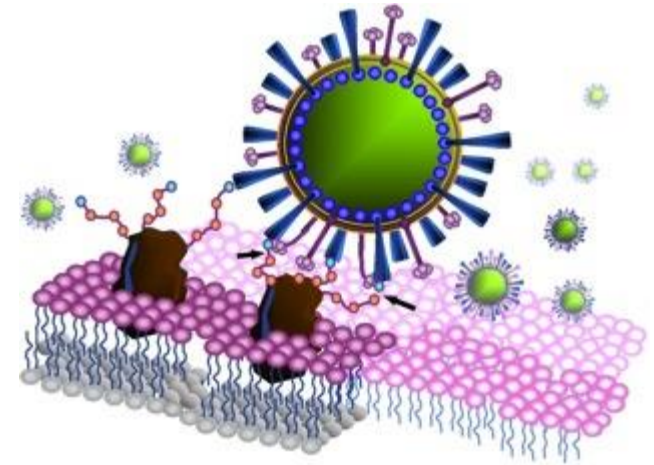


# WISDOM: *In silico* drug discovery

April 2006, Asian and European labs analysed 300K possible drug components against the H5N1 virus using EGEE infrastructure

- 2000 computers for 4 weeks
- (equivalent of 100 years on a single computer)

Preparation for the second run are ongoing



October 2006 – January 2007, the same analysis for malaria resulted in 140 million dockings (80 000 per hour)

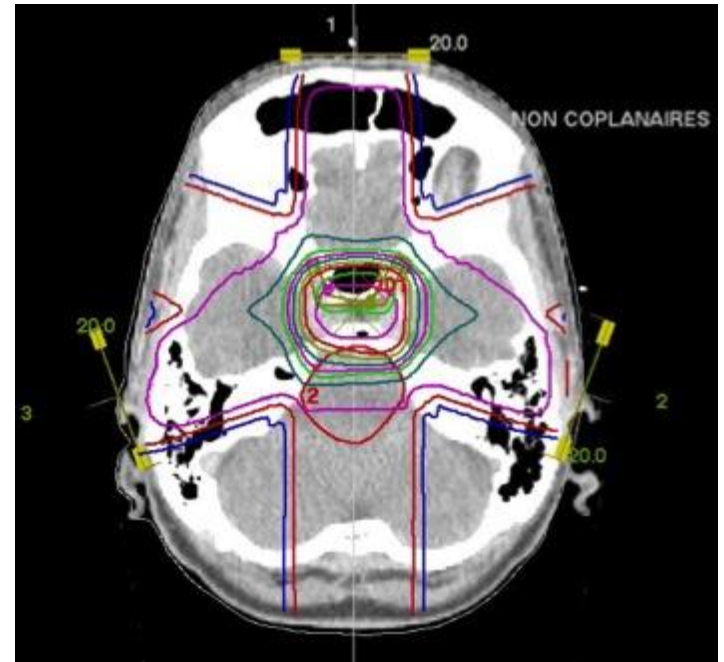
- 5000 computers for 2.5 month (equivalent of 420 years on a single computer)





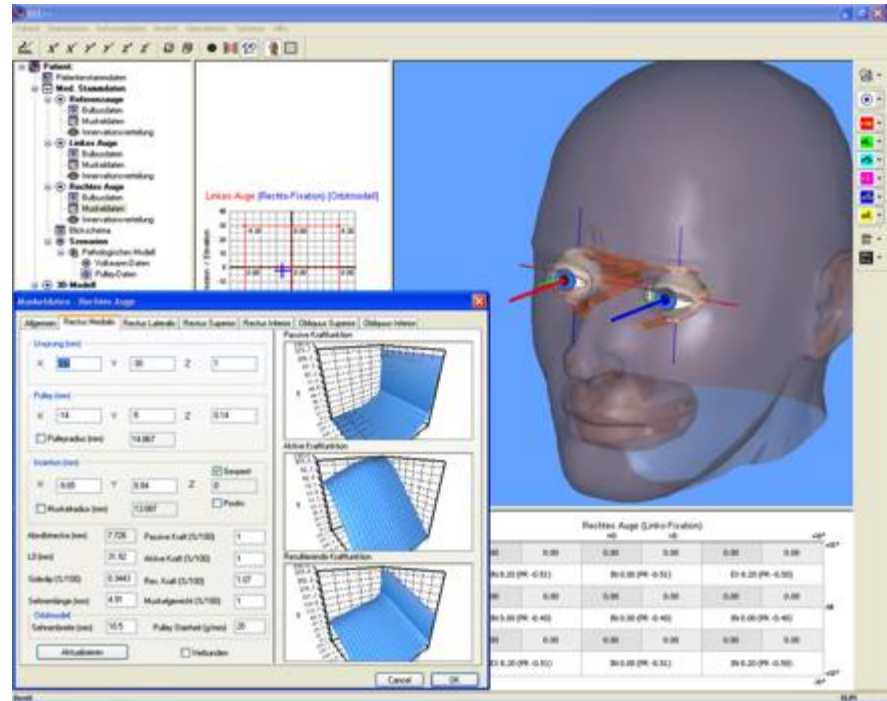
# GATE: radiotherapy planning

- Goal is to improve accuracy of the treatment of cancer by ionizing radiations of the tumours
- Therapy planning is computed from the MR (Magnetic Resonance) scans
  - Tumours are accurately located in 3D
  - Radiation doses needed to be applied are computed
- Due to the Monte Carlo approach – very easy to divide in the parallel



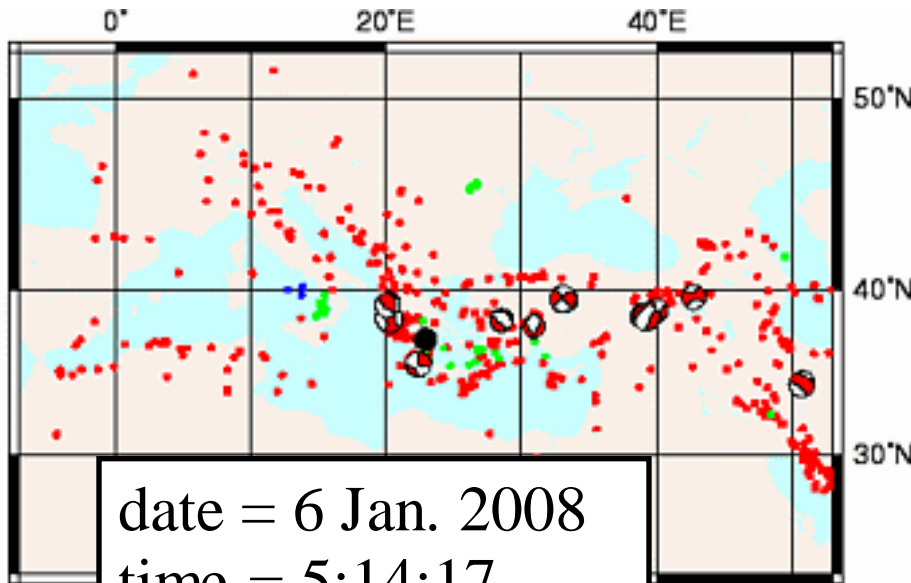
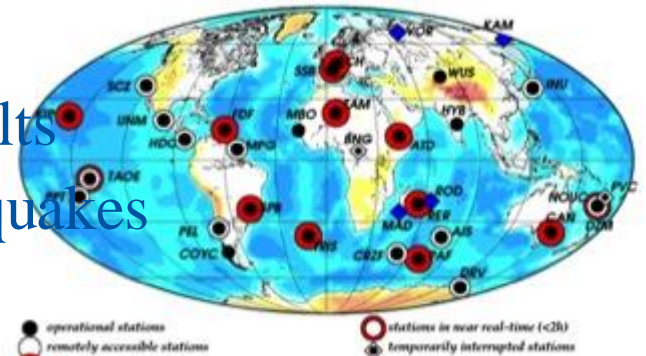
# SEE++: Virtual Eye Surgery

- Simulation system that aims at the forecast of clinical operation results
- Used to simulate pathologies and evaluate possible treatments
- User presentation on the client -> computation in the Grid



# GEOSCOPE: determine the earthquake source

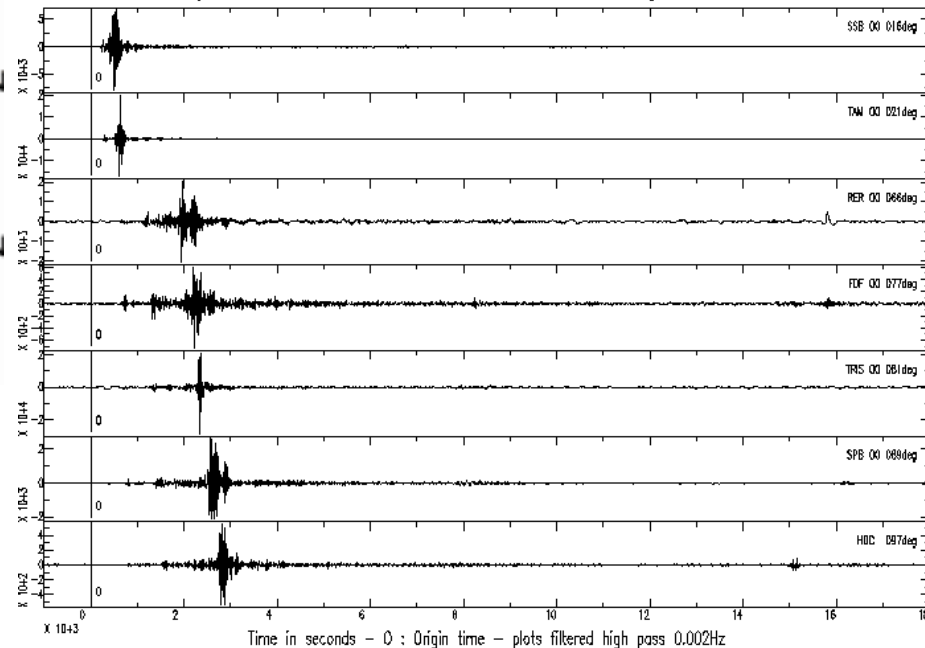
- Localization of earthquakes
- Determination of rupture modes of the faults
- Results within a few hours of major earthquakes



date = 6 Jan. 2008  
time = 5:14:17  
depth = 50.9 km  
magnitude = 6.1  
latitude =  $37.150^\circ$   
longitude =  $22.934^\circ$



January 6th 2008 – Southern Greece, M=6.1, origin time 05h14m17s



# The Grid = SETI@HOME ???

- The GRID
  - Reliable computer centres
  - Hardware and software
    - Dedicated,
    - Static,
    - Fully controlled
  - costs
- SETI@HOME
  - Home PCs
  - Hardware and software
    - Heterogeneous
    - Shared
    - Dynamic
  - “for free”



# CLOUD



# Cloud

- Web-based computing, whereby shared resources, software and/or information are provided to computers and other devices on-demand.
- Providers
  - Amazon EC2 Cloud
  - Google Compute Engine
  - Azure
- Service models
  - Computer (PaaS)
  - Web/Database (SaaS)
  - Cluster (IaaS)
  - Network (NaaS)

# References

- Wikipedia ([www.wikipedia.org](http://www.wikipedia.org))
- Web
  - Semantic Web
  - Linked Data ([linkeddata.org](http://linkeddata.org))
  - Wolfram alpha ([www.wolframalpha.com](http://www.wolframalpha.com))
- Grid
  - EGEE ([www.eu-egee.org](http://www.eu-egee.org))/ EGI
  - SlovakGrid ([www.slovakgrid.sk](http://www.slovakgrid.sk))
  - WLCG([lcg.web.cern.ch](http://lcg.web.cern.ch))
- Cloud
  - Amazon EC2 service, Google Web Apps

# Summary

- World Wide Web – born at CERN - elaborate set of interlinked documents accessible over the Internet
- Grid computing
  - Infrastructure that provides seamless access to computing power and data storage
  - Suitable for problems that need huge/shared data, can run in parallel or need to be run frequently
- EGI – largest existing European Grid infrastructure (academic)
- Slovak Grid Initiative
- Many existing applications besides high energy physics

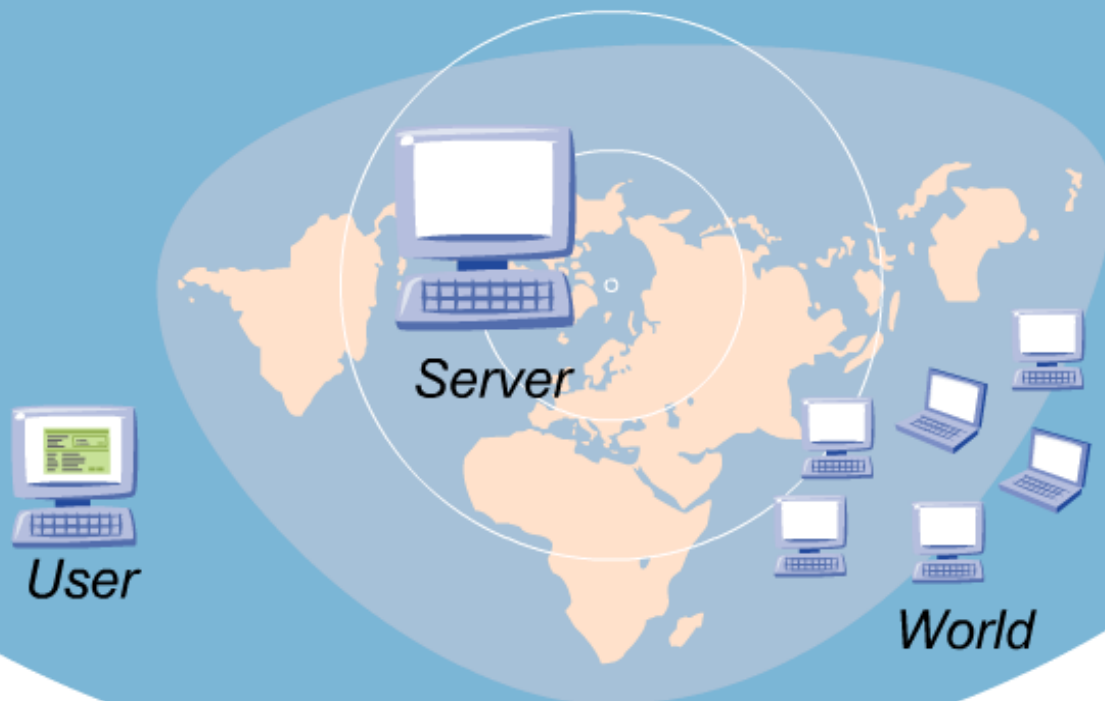




# Commercial Software

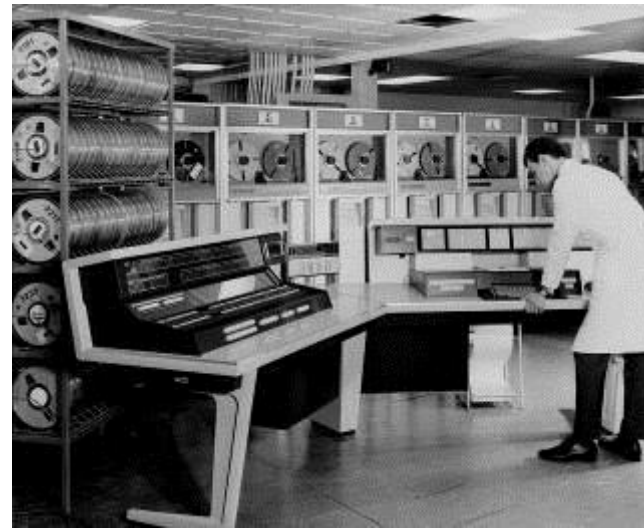
- Gaussian
  - <http://www.gaussian.com/>
  - Predicts the energies, vibrational freq., ... of molecular systems.
  - VO-based licensing model, actually in use in gaussian VO.
- MathWorks
  - <http://www.mathworks.com/>
  - Integrate MATLAB & Distributed Computing Engine with EGEE.
  - Both client and server are licensed in this model.
- Interactive Supercomputing
  - <http://www.interactivesupercomputing.com/>
  - Similar to DCE; used from multiple clients (MATLAB, Python, R)
  - Server licensed, some clients licensed

# What is the Web?



# Why was the Web invented at CERN?

- Science depends on free **access to information** and exchange of ideas. CERN is the hub of a worldwide community of **6500 scientists** in **80 countries**.
- CERN has a long history of being at the forefront of **scientific computing** and **networking** (first lab on Internet outside the US).
- During the preparation of the previous large project LEP, the need to share documents in a global way became vital.



# How did it start?

- 1989: **Tim Berners-Lee** circulates “Information Management: A proposal” to help with future Large Hadron Collider project.
- 1991: Early **www system released** to high energy physics via the CERN program library. First web servers located in European physics laboratories.
- 1993: First Mosaic browser; web reaches 500 servers and 1% of Internet traffic; CERN places **WWW in the public domain**.

