CLUSTER AND GRID COMPUTING

O que é um cluster?

- De forma geral, é um aglomerado de máquinas conectadas em uma rede local ou dedicadas
- NOWs (Network of Workstations) algumas vezes não são consideradas clusters
- No nosso contexto, consideraremos cluster como qq aglomerado de máquinas em rede local com serviços básicos de rede (ssh, etc)

Clusters

- Nível de usuário:
  - políticas de utilização
  - gerenciadores de recursos
  - interface com o usuário
- Nível de administração:
  - gerência de hardware
  - gerência de software

O que é um grid?

- Um conjunto de clusters?
- Mais do que isso:
  - Organização virtual que permite a aglomeração de recursos que estão distantes geograficamente
  - Recursos podem ser: máquinas, dados, instrumentos etc

Grids

- Nível de usuário:
  - políticas de utilização
  - gerenciadores de recursos locais
  - Gerenciadores de recursos globais
  - Monitoração
  - Autenticação
  - Certificação
  - interface com o usuário
- Nível de administração (local e global):
  - gerência de hardware
  - gerência de software

Diferenças

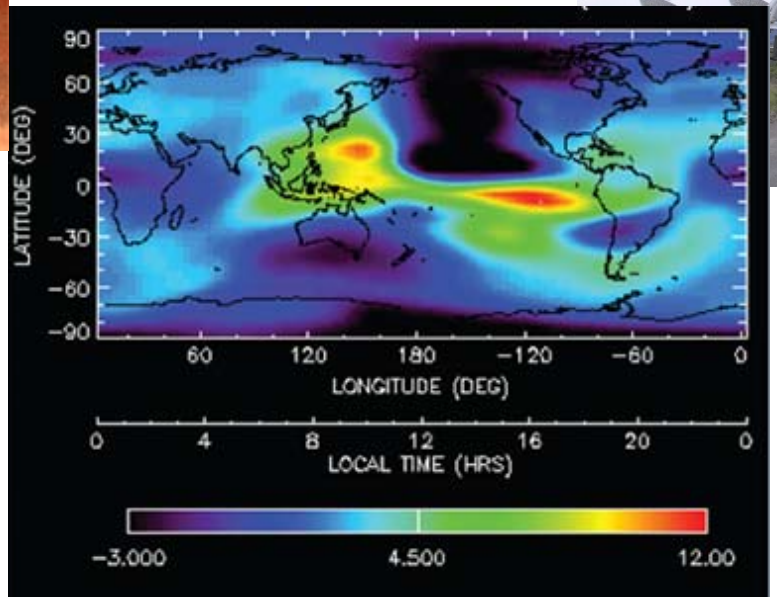
	Ambiente distribuído convencional	Grid
1	um conjunto virtual de nodos computacionais	um conjunto virtual de recursos
2	um usuário tem acesso a todos os nodos do conjunto	um usuário tem acesso ao conjunto mas não aos sítios individuais
3	acesso a um nodo significa acesso a todos os recursos do nodo	acesso a um recurso pode ser restrito
4	um usuário tem conhecimento das características do nodo	um usuário tem pouco conhecimento sobre cada sítio
5	nodos pertencem a um mesmo domínio administrativo	recursos se espalham por múltiplos domínios administrativos
6	elementos no conjunto: 10-100, praticamente estático	elementos no conjunto: 1000-10000, dinâmico

Por que Grid?

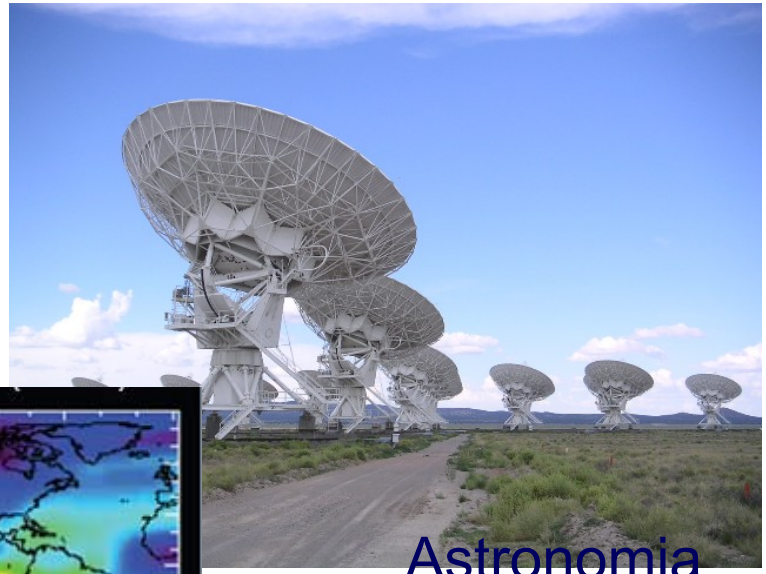
- Várias razões:
  - Científicas
  - Políticas
  - Econômicas
  - Sociais

O Problema

Bioinformática



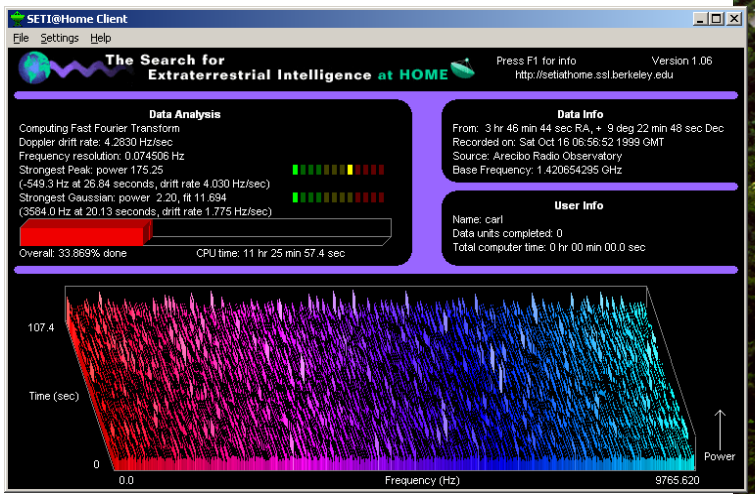
Clima / previsões



Astronomia

Seti@Home (1999)

- *Search for Extraterrestrial Intelligence*
- http://setiathome.berkeley.edu/
- Screensaver
  - Ciclos ociosos
- “volunteer computing“



O Problema

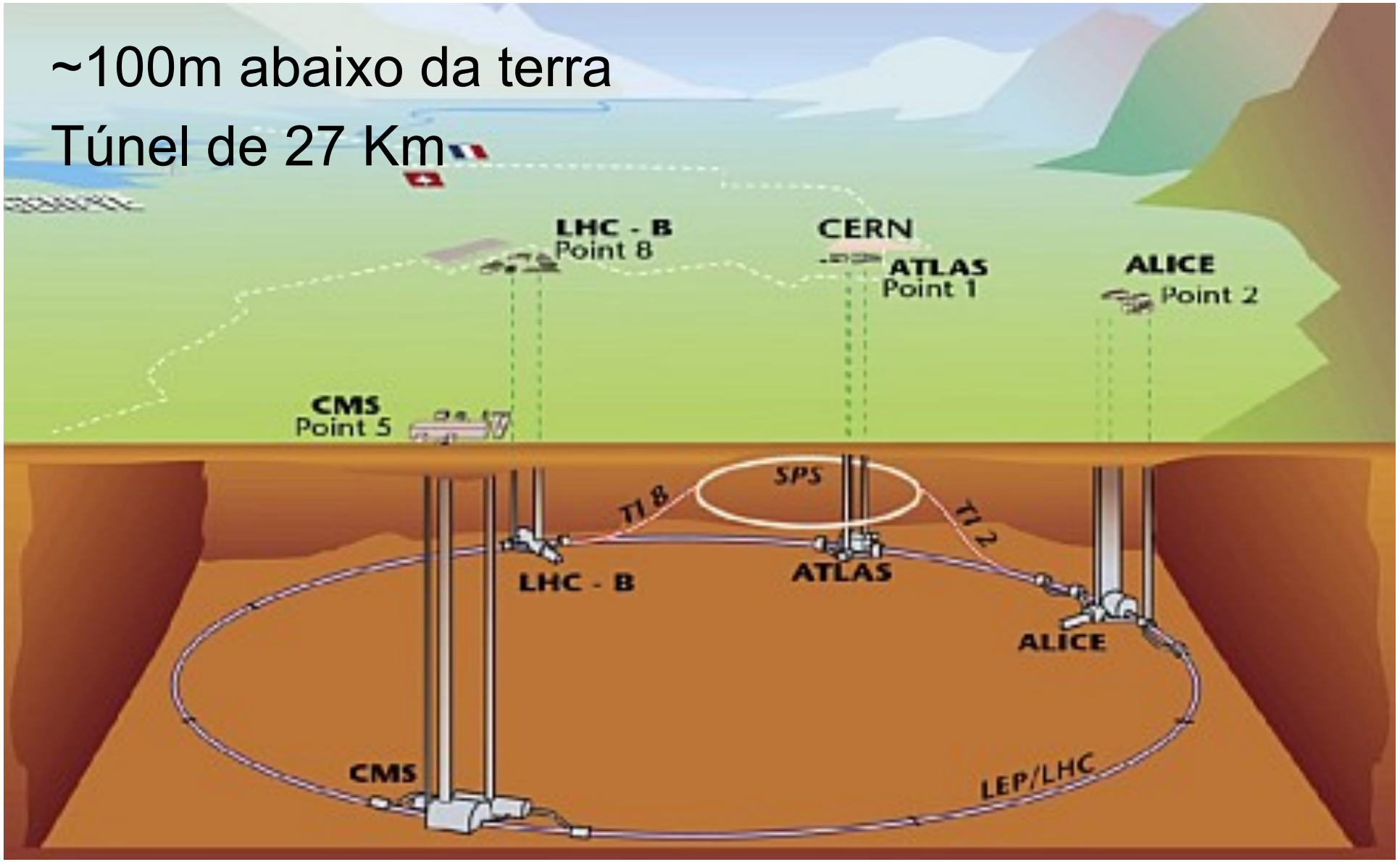
- Frequentemente, um único computador ou mesmo supercomputadores não são suficientes para esses tipos de cálculos,
  tornando muito difícil, caro e às vezes impossível alcançar determinados objetivos

Um problema maior ainda!

E-infrastructure shared between Europe and Latin America

- O maior experimento científico do mundo

LHC - Large Hadron Collider



LHC - Large Hadron Collider

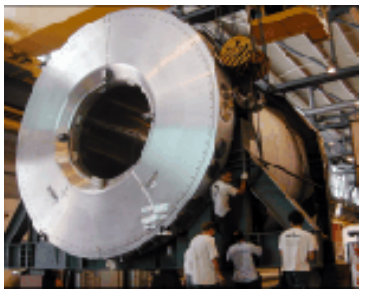
- 40.000.000 de colisões por segundo em cada detector

- 15 Petabytes de dados por ano (~15.000.000 GB)
  - ~ 21 milhões de CDROMs
  - 41TB por dia
  - 150 vezes todo conteúdo publicado anualmente na WWW *

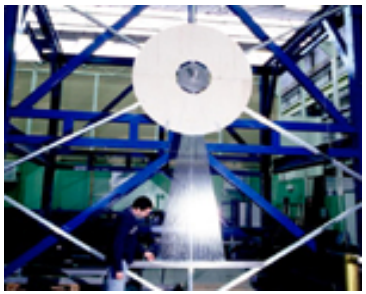
CMS

LHCb

ATLAS



ALICE



(*) Baseado em uma extimativa do vice-presidene de operações do Google

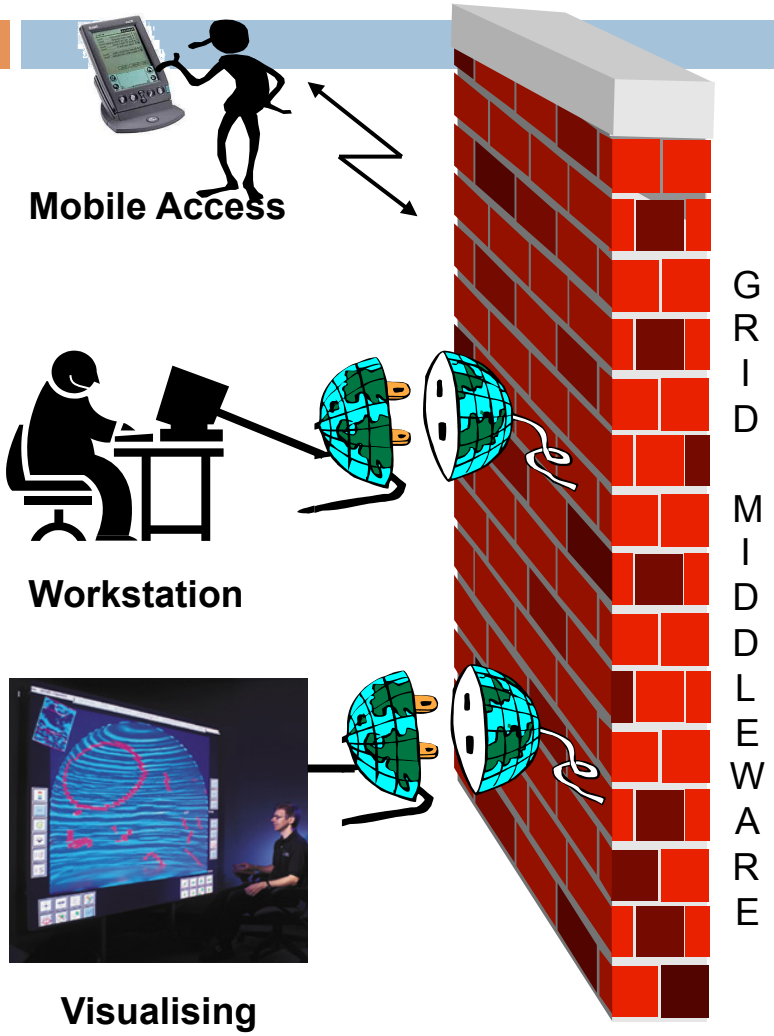
LHC - Large Hadron Collider

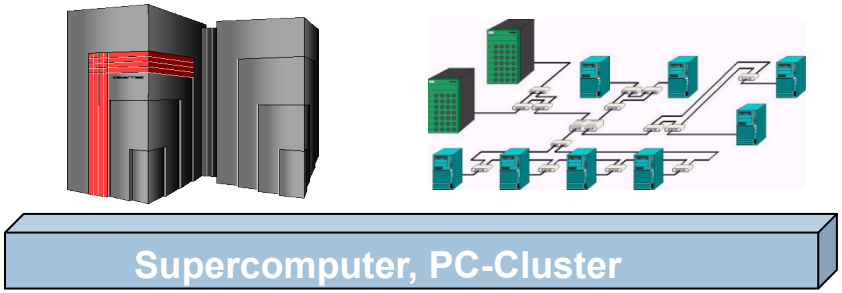
- Seria necessário um cluster com ~100.000 CPUs
- Os dados precisam estar disponíveis para milhares de cientistas, independete da sua localização

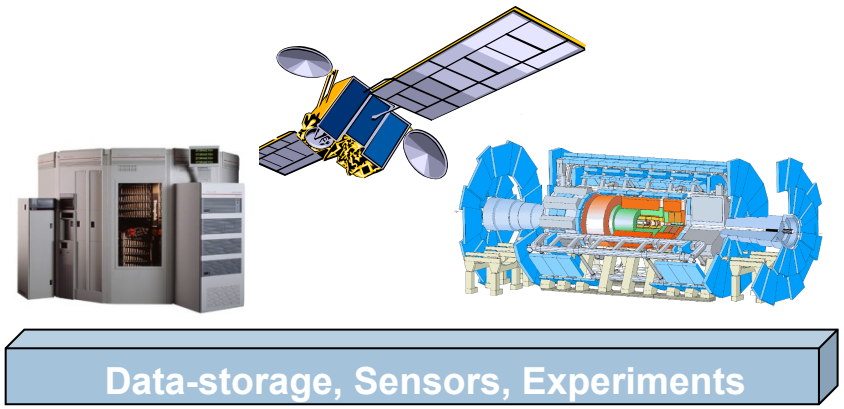
A Solução

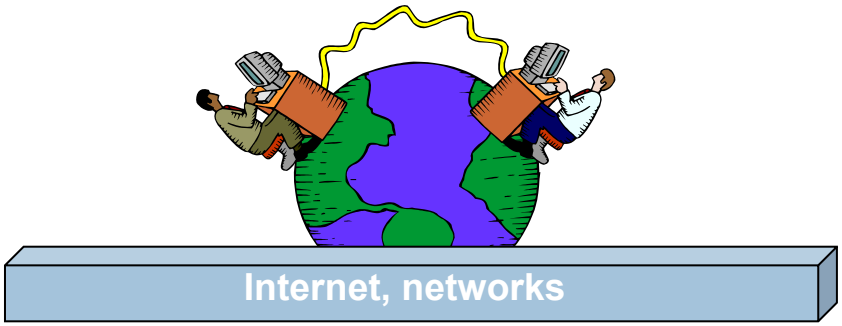
Grid computing - Analogia à rede elétrica (*electrical power grid*)

A metáfora do Grid





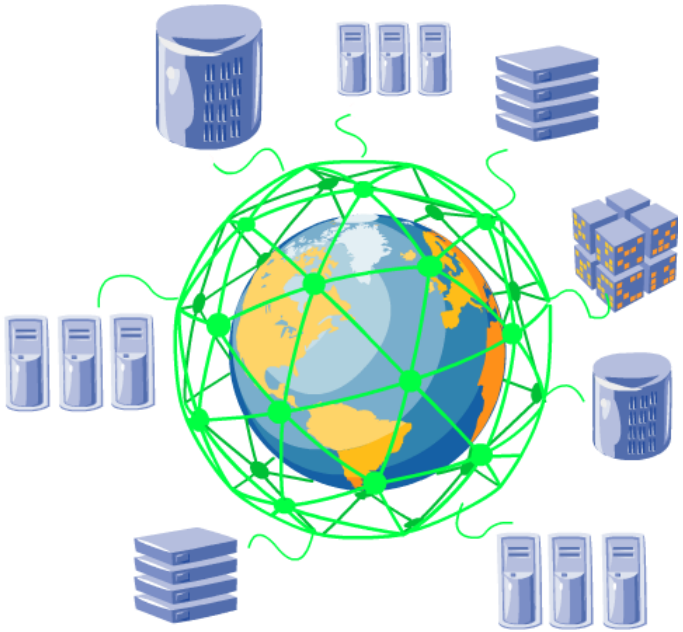




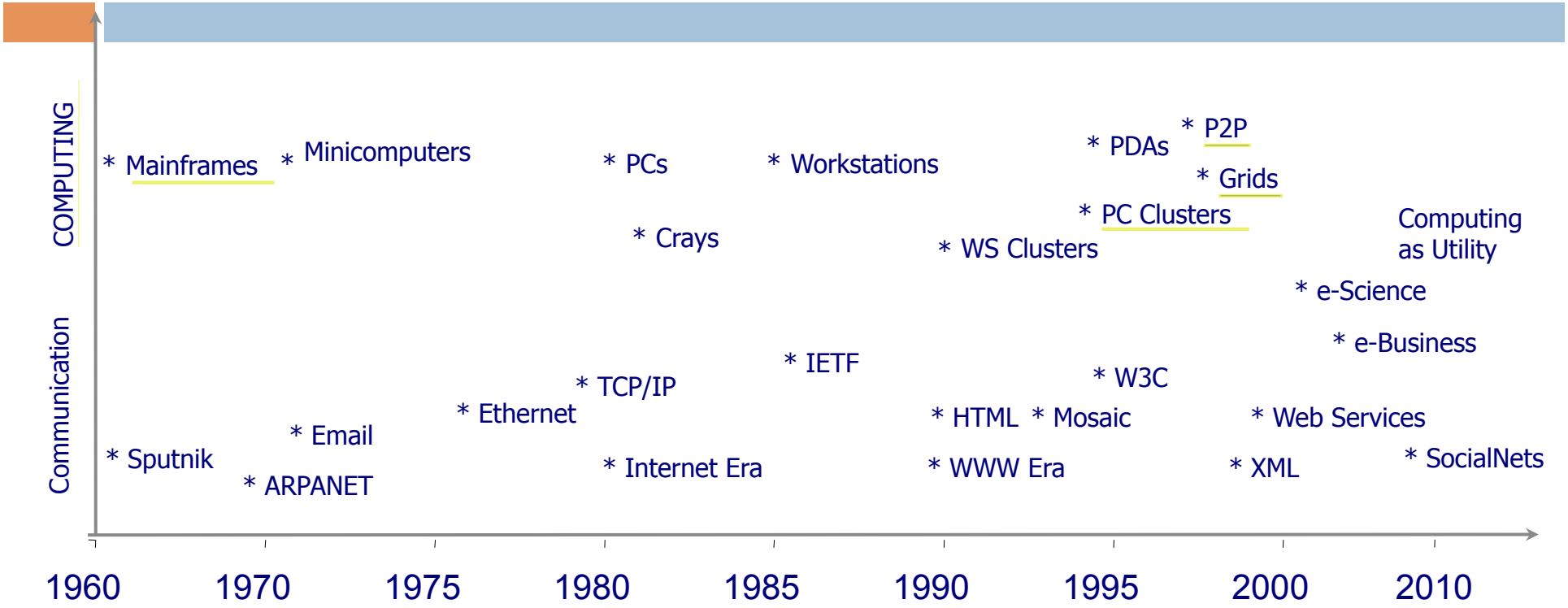
Características

- Espaço de armazenamento abundante
- Altíssimo poder de computação
- Colaboração com colegas distantes, compartilhando recursos, dados e resultados

e-Ciência



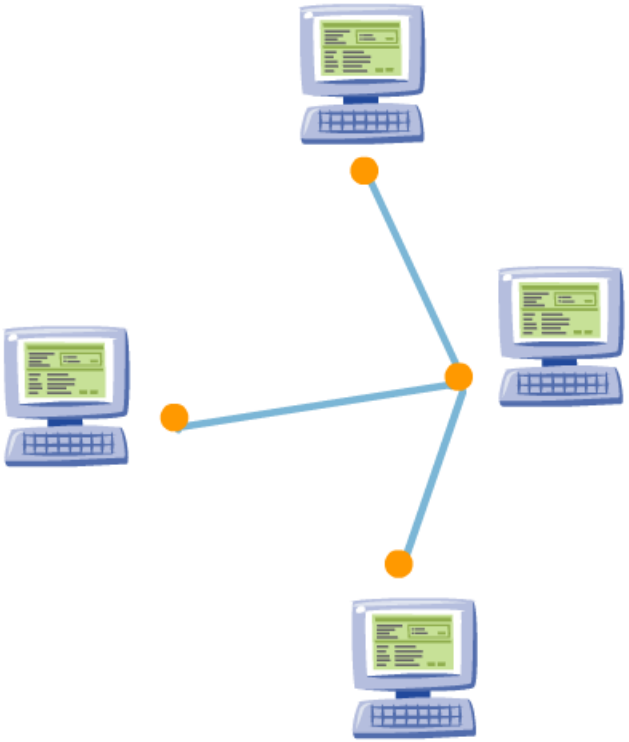
Evolução tecnológica



Controle Centralizado ⇢ **Controle Descentralizado**

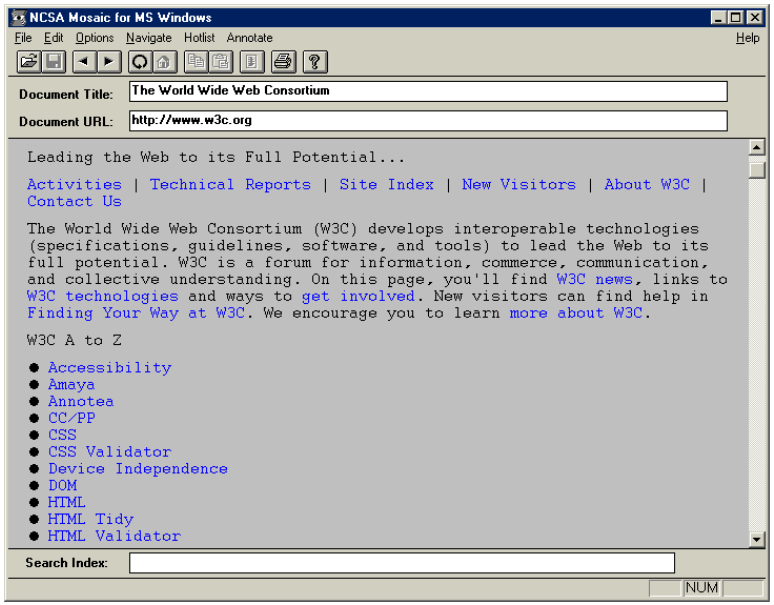
Internet X Web X Grid

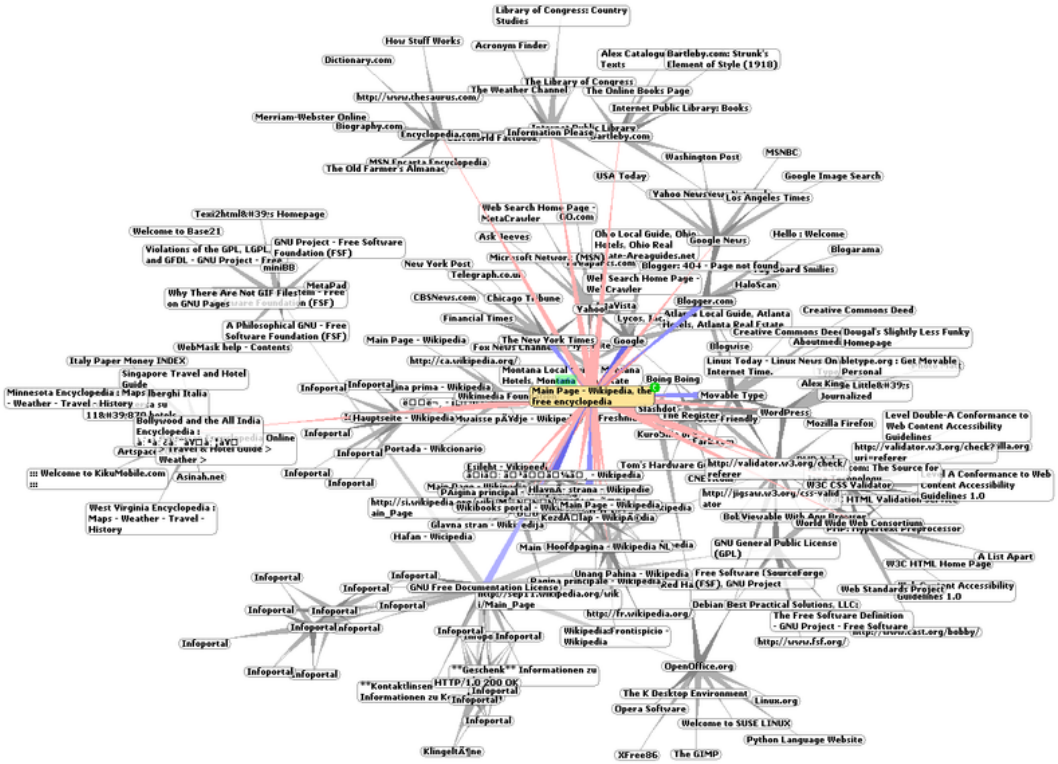
- INTERNET
  - Infra-estrutura de rede que conecta milhões de computadores ao redor do mundo
  - TCP/IP
  - Década de 1970



Internet X Web X Grid

- World Wide Web
  - Serviço de compartilhamento de informações/documentos construído no topo da Internet
  - Início década de 1990





Internet X Web X Grid

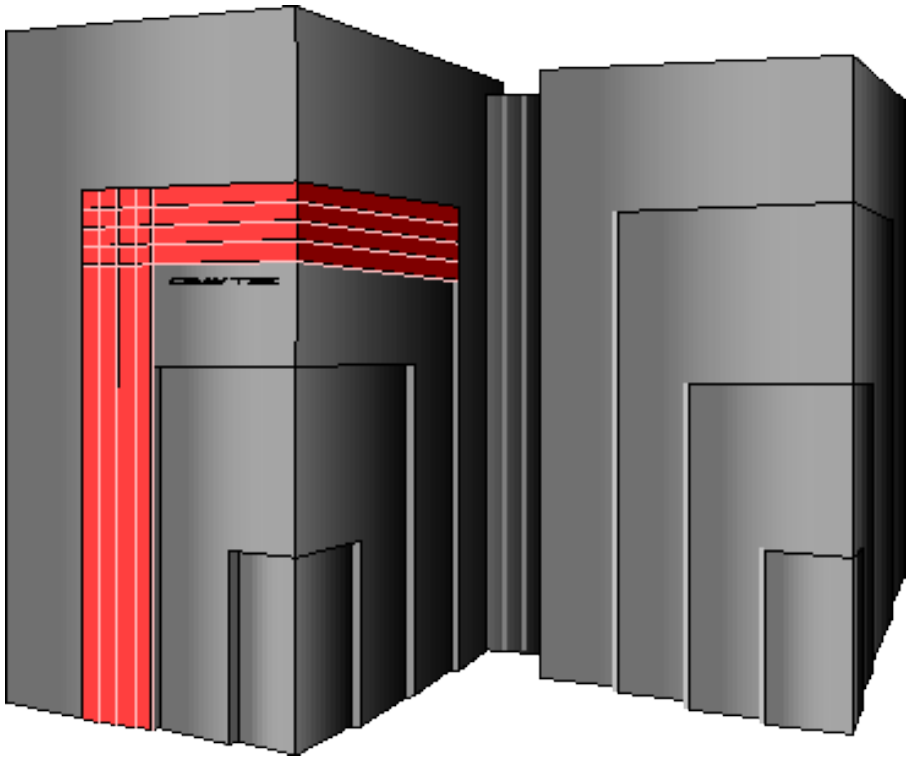
- Grid
  - Também é um serviço construído no topo da Internet, mas vai um passo a diante…
  - Na WWW, os computadores compartilham informações
  - Graças ao Grid, computadores serão capazes de compartilhar poder computacional e recursos como armazenamento de disco, aplicativos e base de dados.

A revolução do Grid

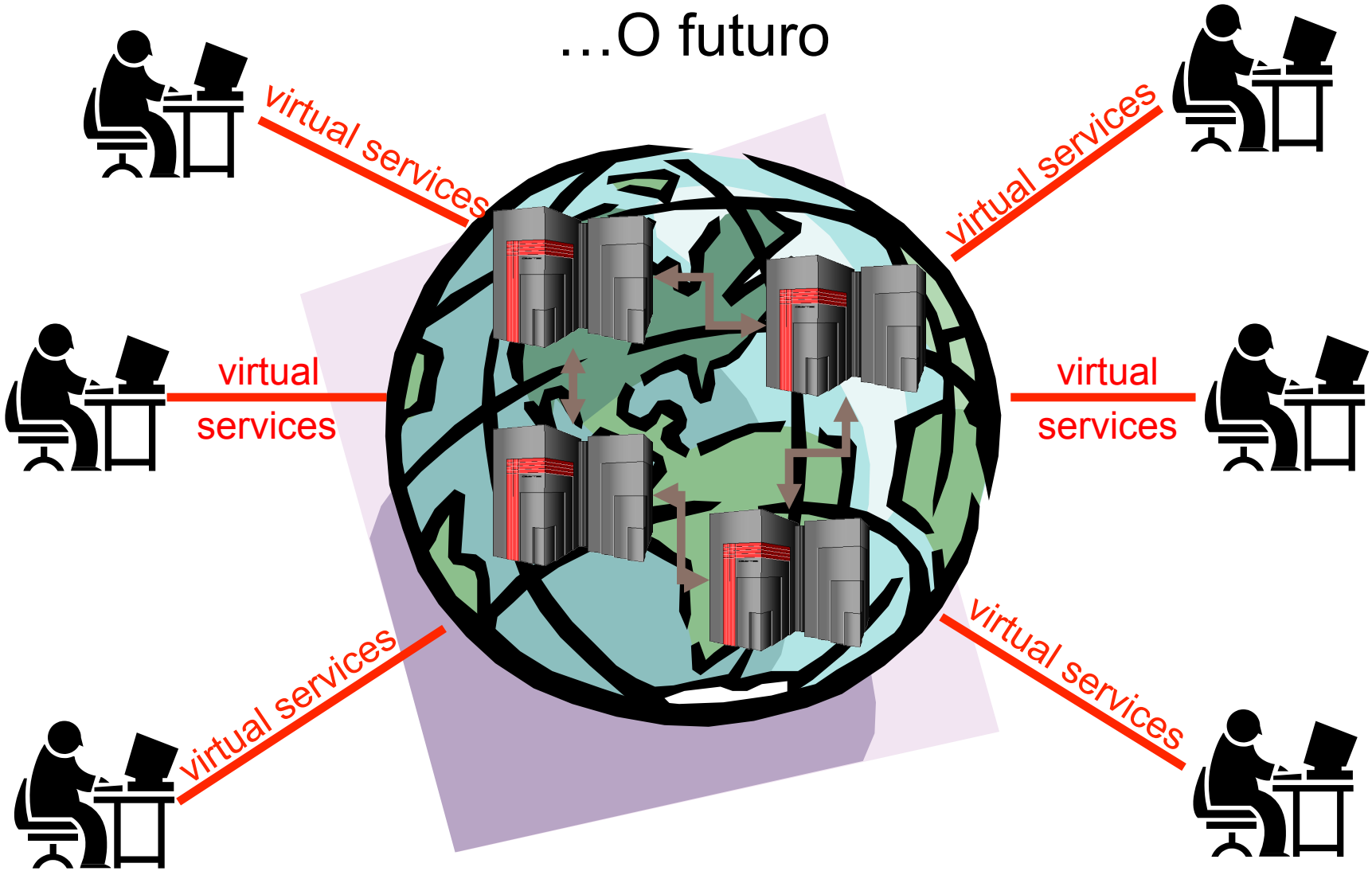
O passado , o presente …

- **CPU**
- **Memory**
- **Disc**
- **Input/Output**

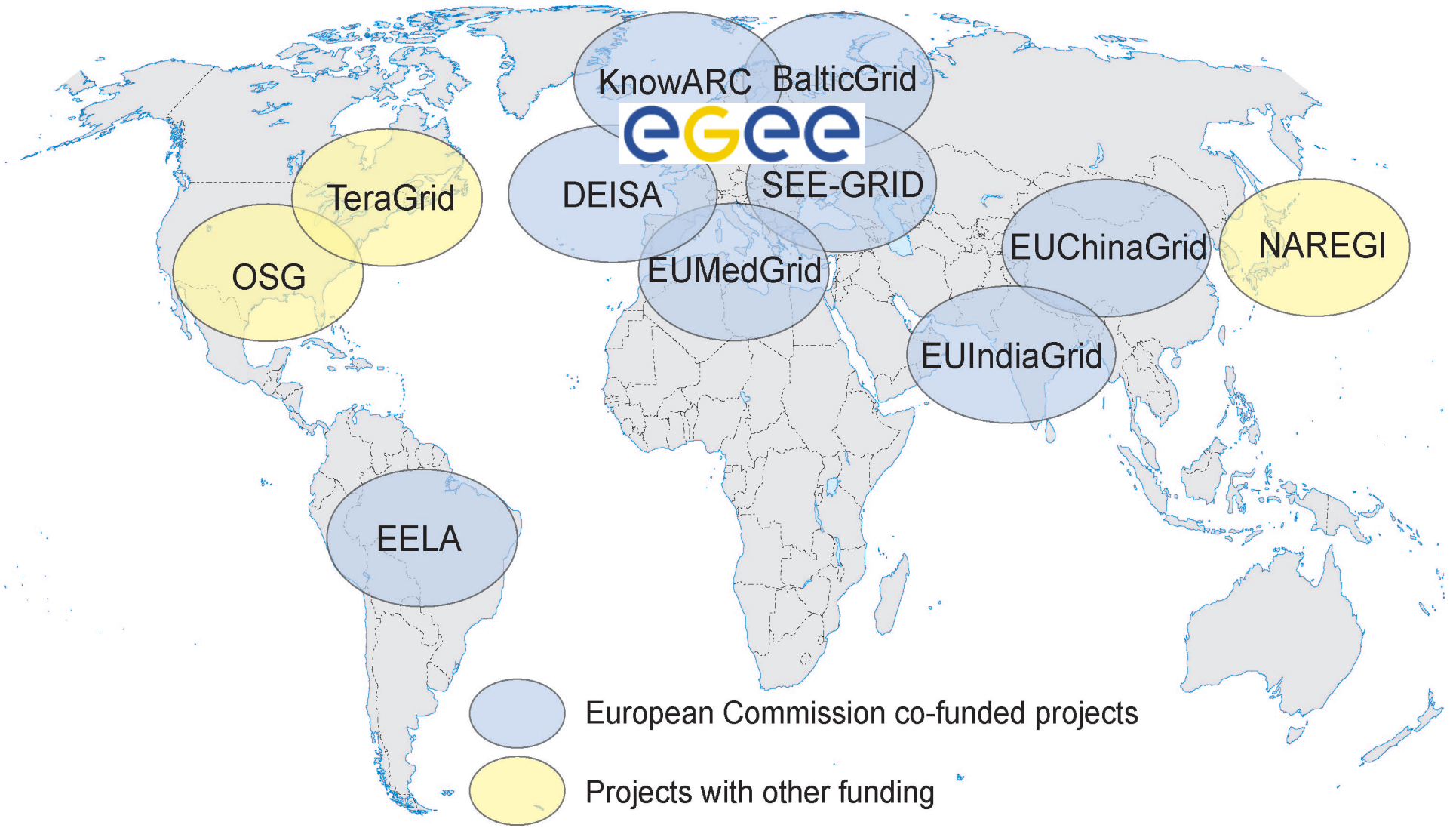
01011010110



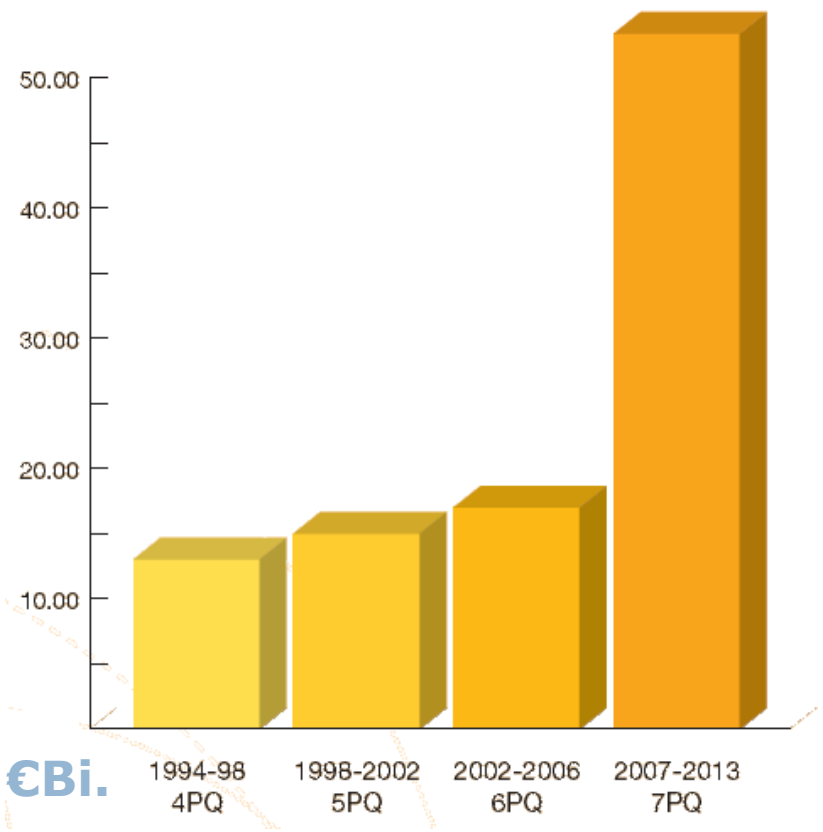
A revolução do Grid

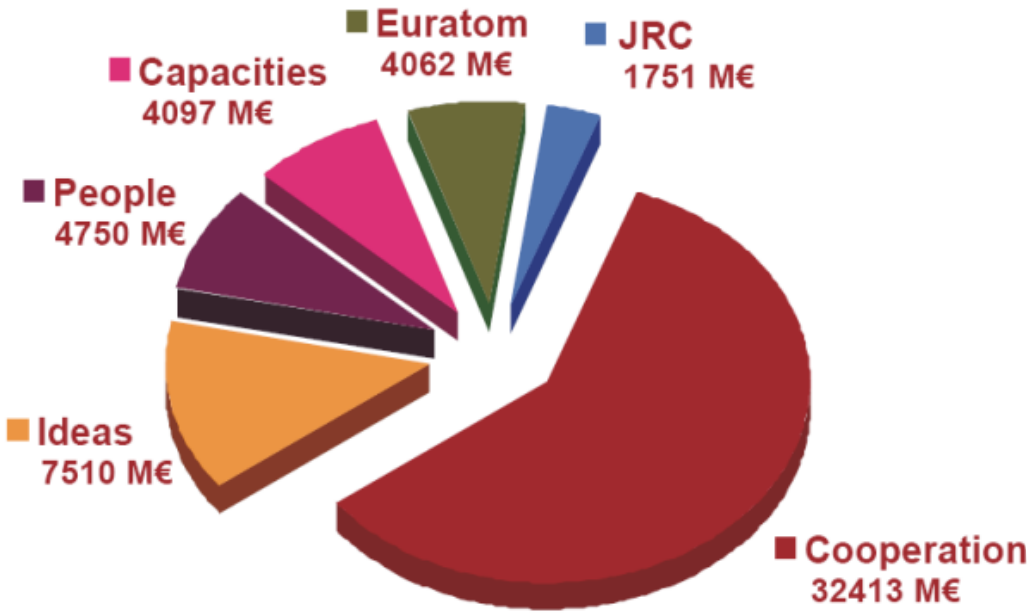


Cenário em 2007



Investimentos da UE



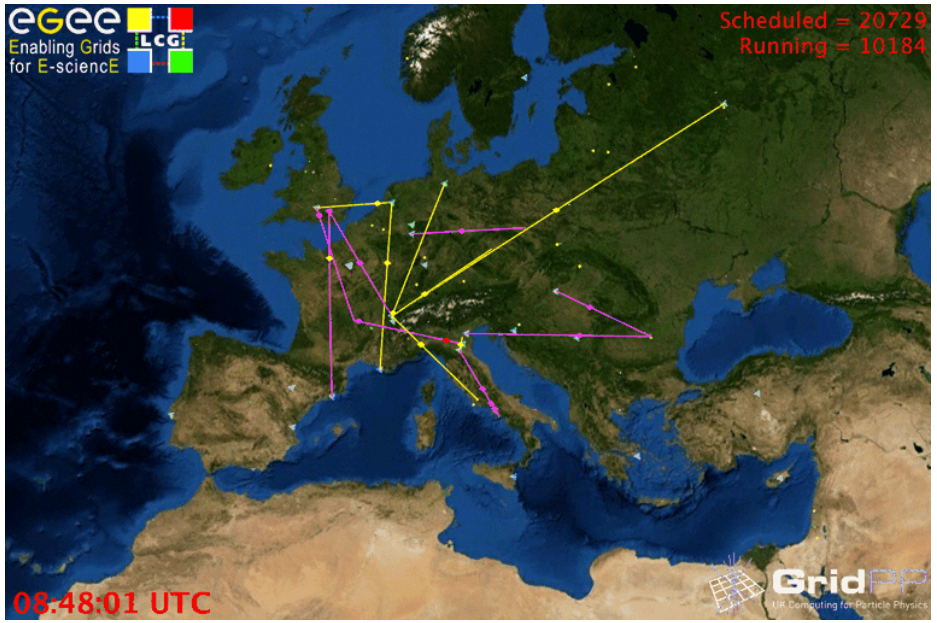


Projeto EGEE

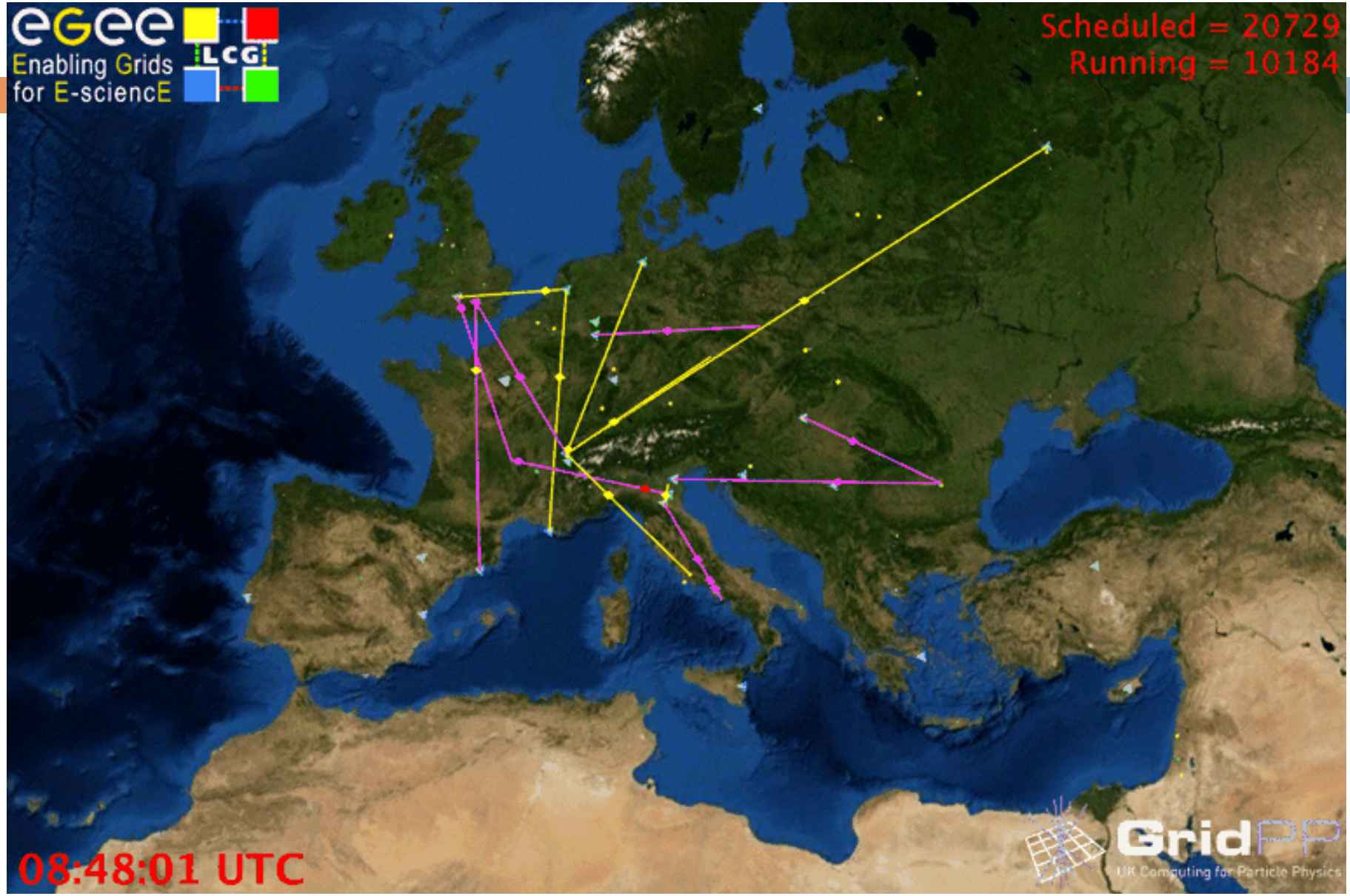
- Coordenado pelo CERN
- 32 países
- 91 instituições
- Orçamento de 35+ M€

- > 35.000 CPUs
- ~ 2.500 TB storage
- > 50.000 jobs per day

www.eu-egee.org

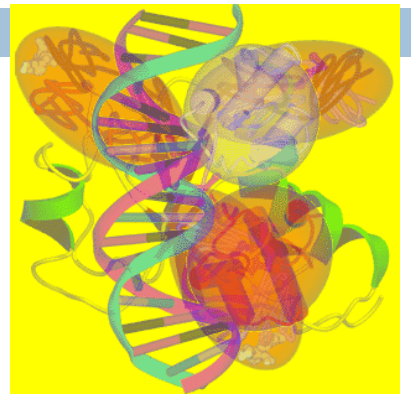


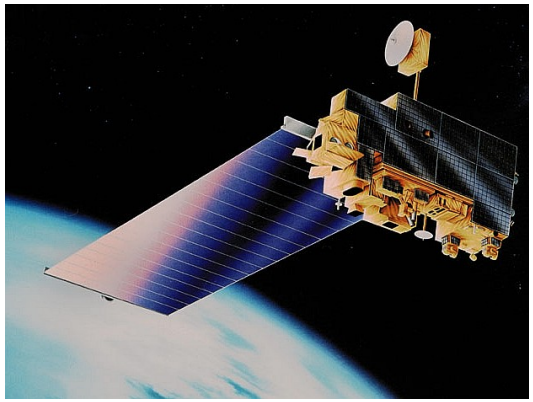
Projeto EGEE



Projeto EGEE - Aplicações

- Variado domínios científicos
  - Arqueologia
  - Astrofísica
  - Química
  - Geofísica
  - Física de Altas Energias
  - Engenharia
  - Simulações Financeiras
  - Biologia
  - Genética





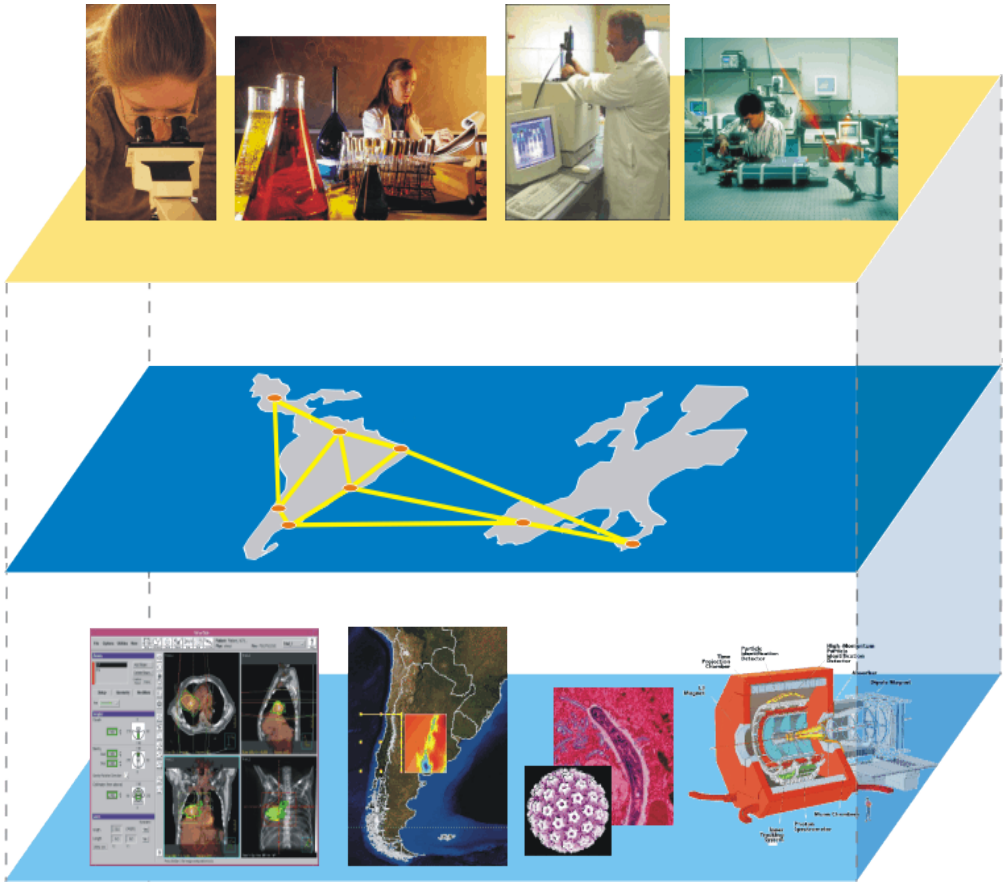
http://indico.cern.ch/conferenceTimeTable.py?confId=22351

O Projeto EELA

O Projeto EELA

- Argentina (*JRU*)
  - 3 members (coord. LINTI-UNLP)
- Brasil (*JRU*)
  - 15 members (coord. UFRJ)
- Chile (*JRU*)
  - 7 members (coord. REUNA)
- Colombia (*JRU*)
  - 2 members (coord. UNIANDES)
- Cuba (CUBAENERGIA)
- Equador (UTPL)
- França (*JRU*)
  - 2 members (coord. CNRS)
- Irlanda (UCC-CMRC)
- Italia (INFN)
- Mexico (UNAM)
- Peru (*JRU*)
  - 4 members (coord. SENAMHI)
- Portugal (*JRU*)
  - 3 members (coord. UPORTO)
- Espanha (*JRU*)
  - 8 members (coord. CIEMAT)
- Venezuela (*JRU*)
  - 2 members (coord. ULA)
- Internacional (CLARA)

EELA em 1 slide



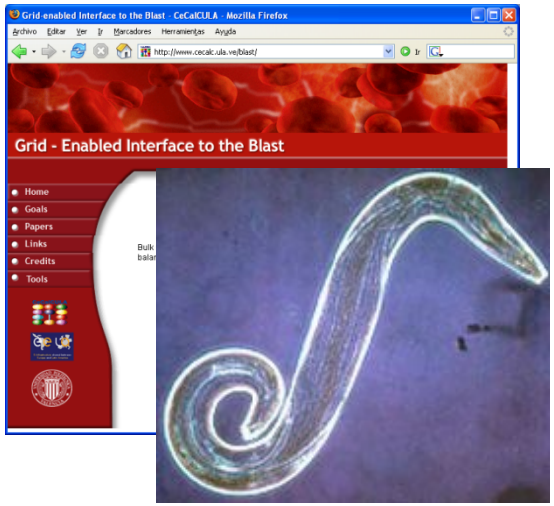
- Pesquisadores querendo realizar pesquisas em conjunto com outros colegas/ instituições

- Infra-estrutura: centros de computação na UE e na América Latina (Biomed, e-Learning, Clima)

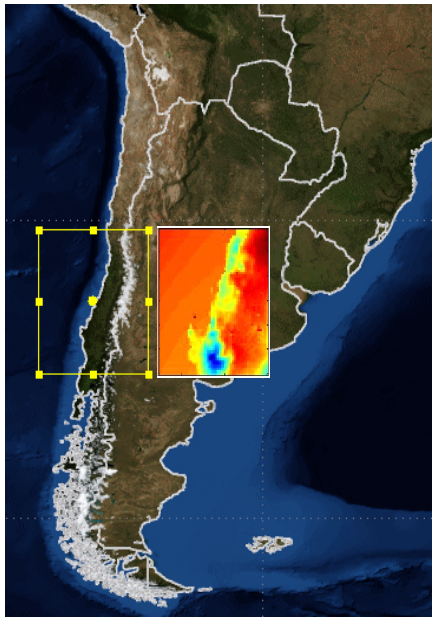
www.eu-eela.eu

Aplicações

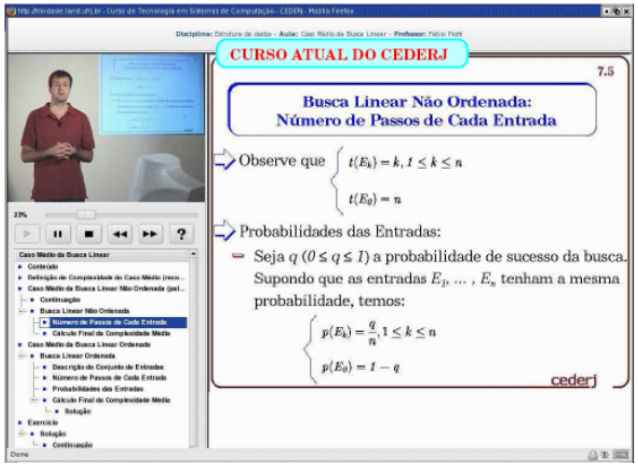
- 47 aplicações (13 no projeto precedente)
- Ap. voltadas para problemas da América Latina



Malaria



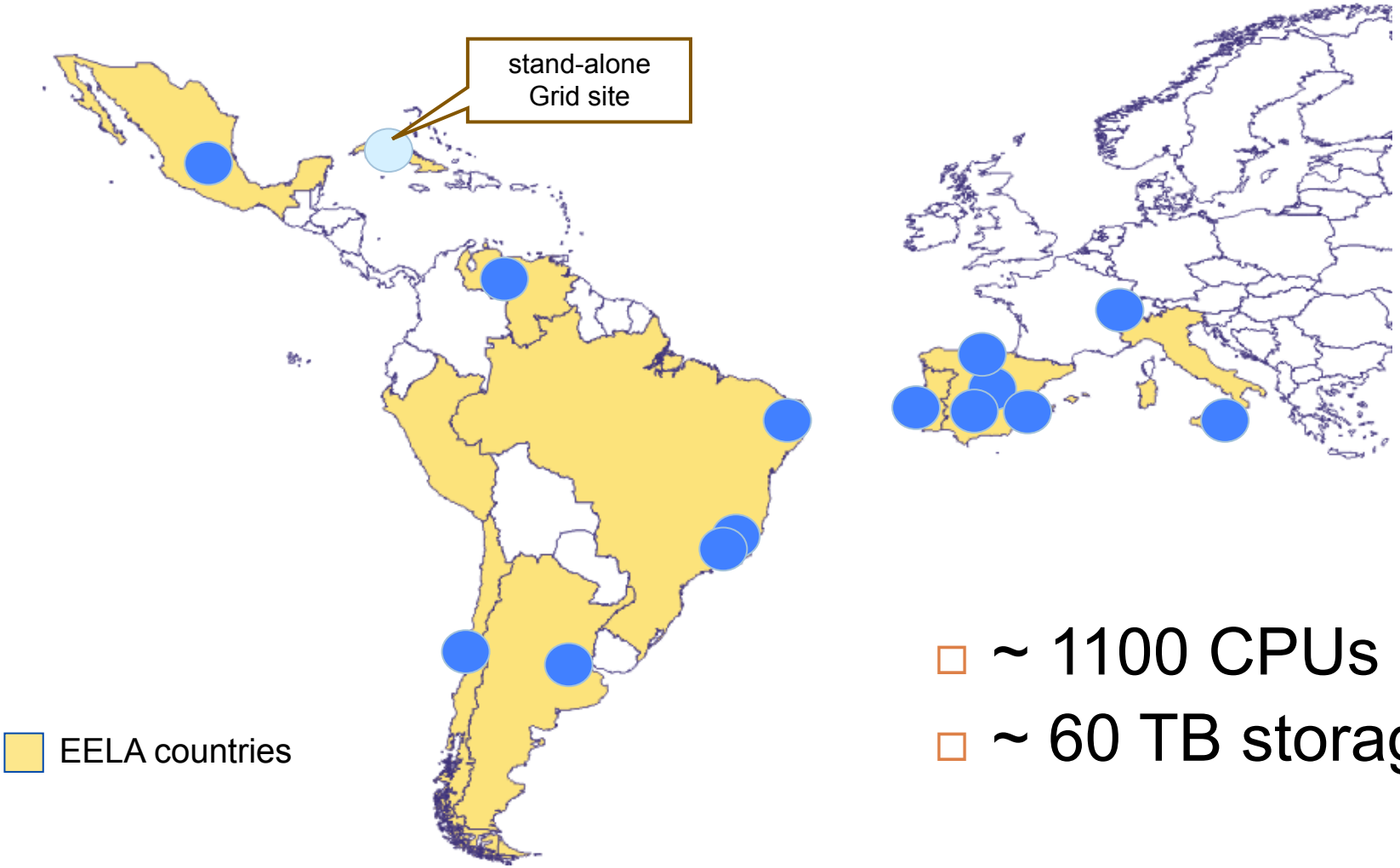
El Niño



E-learning

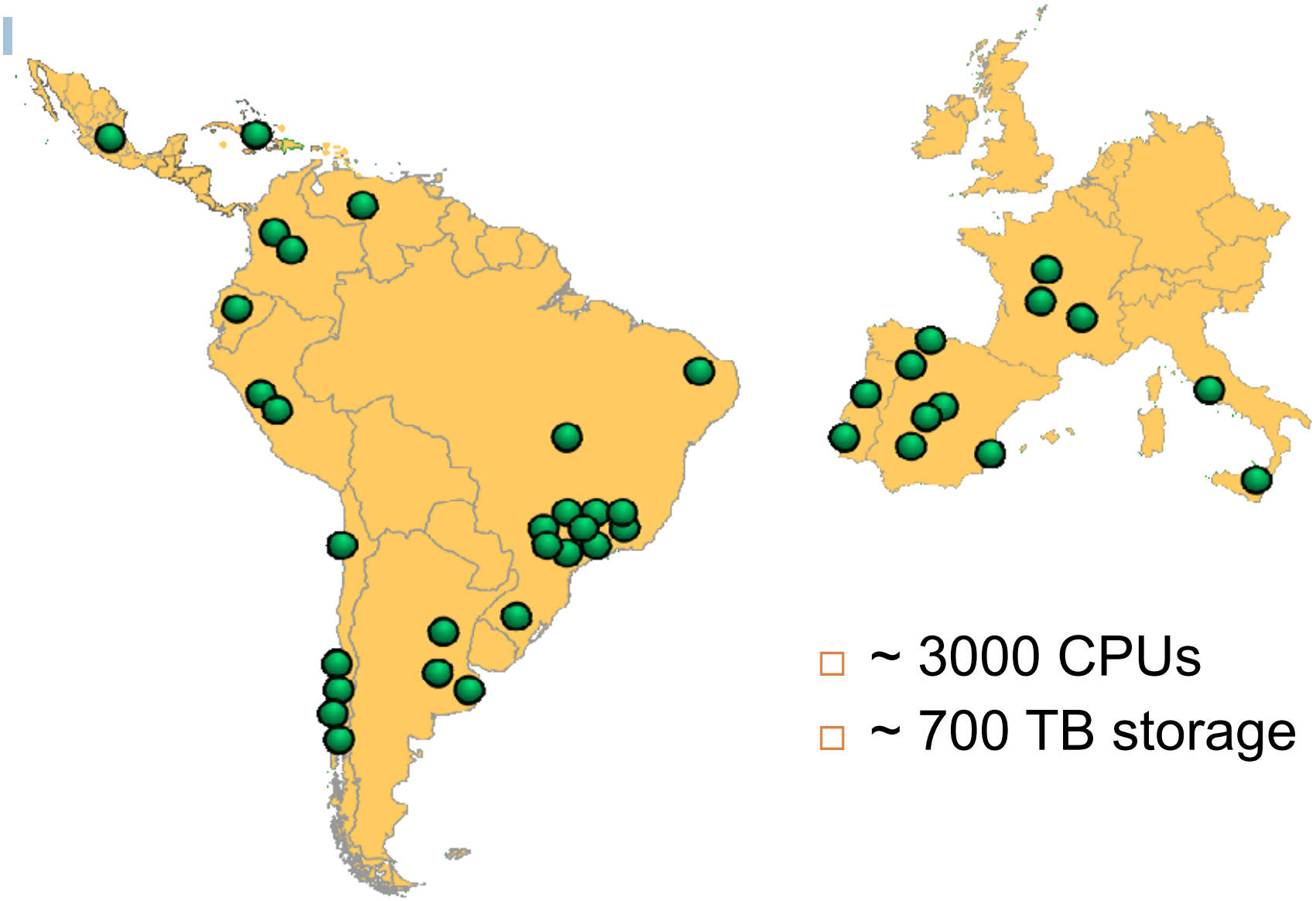
http://applications.eu-eela.eu

Centros de computação - EELA



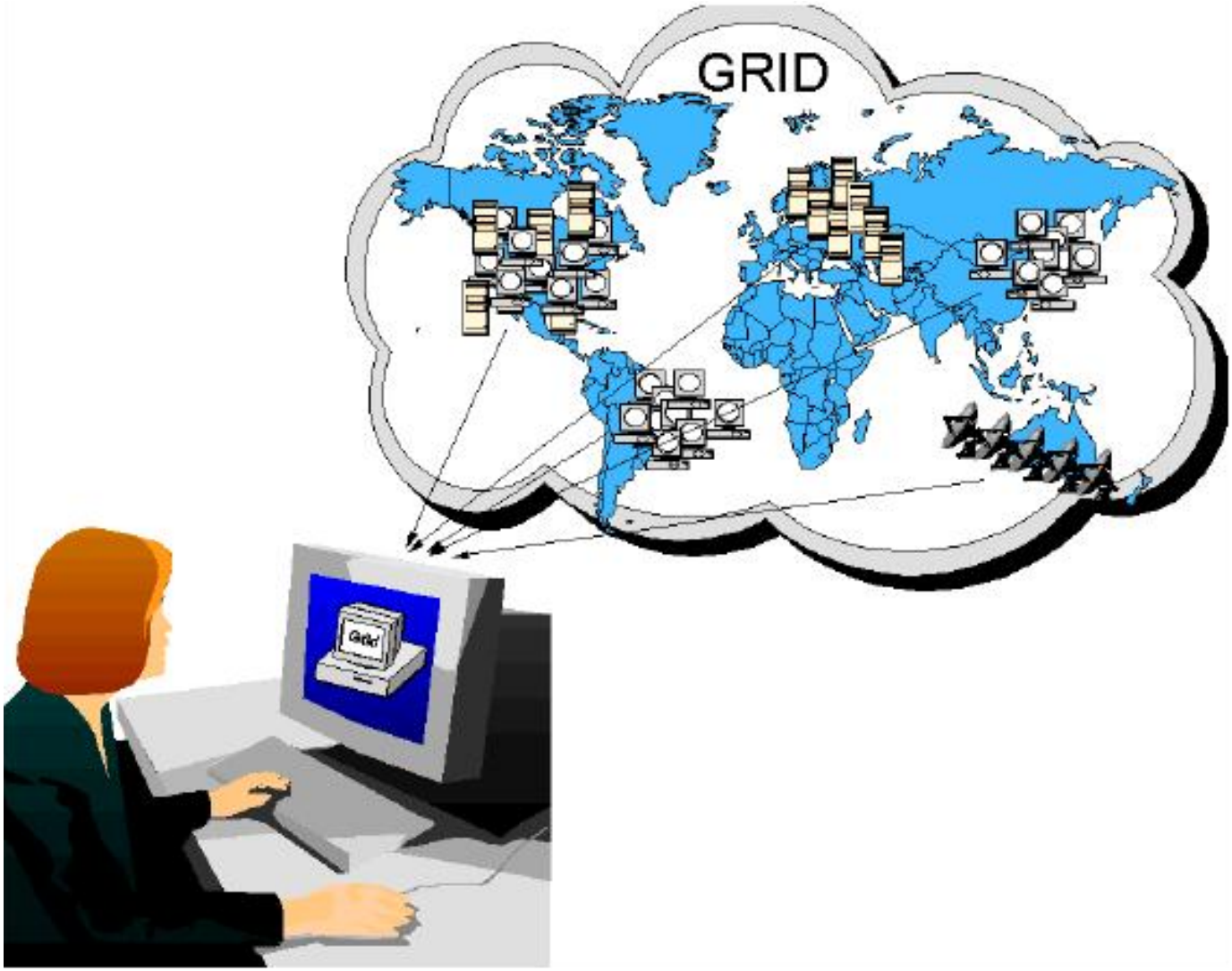
- ~ 1100 CPUs
- ~ 60 TB storage

Centros de computação – EELA-2



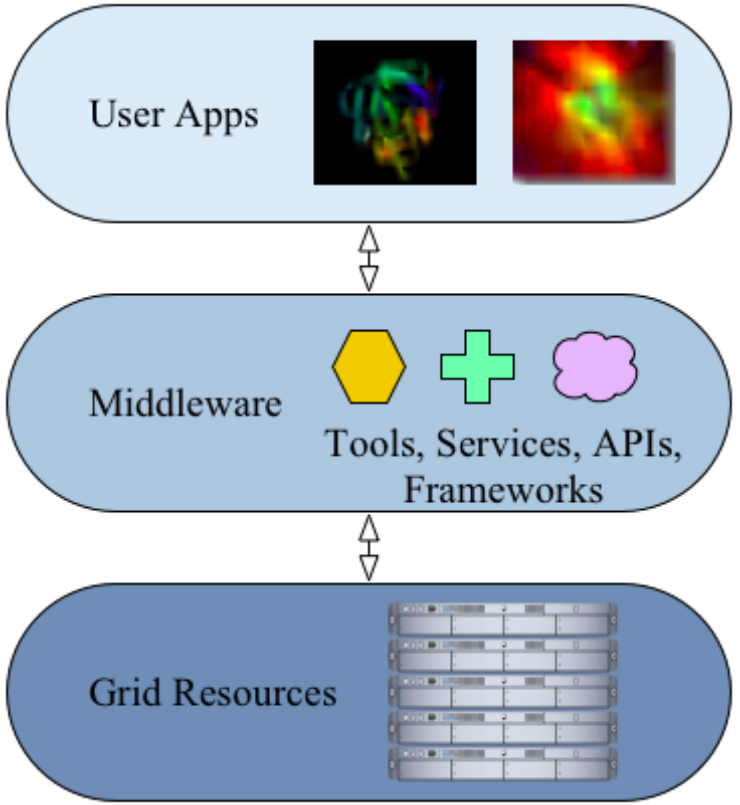
- ~ 3000 CPUs
- ~ 700 TB storage

Conceitos básicos

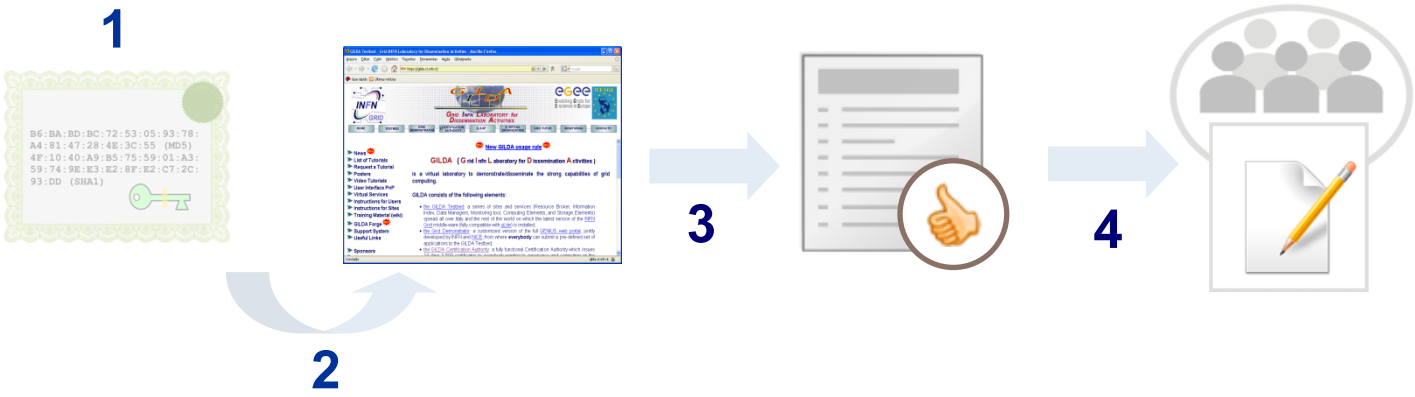


Middleware

- Software de mediação”
- Grid Middlewares
  - gLite
    http://glite.web.cern.ch/glite/
  - Globus Toolkit
    http://www.globus.org
  - Condor
    http://www.cs.wisc.edu/condor/
  - UNICORE
    http://www.unicore.eu
  - OMII-UK
    http://www.omii.ac.uk
  - Etc…



Preocupação com a segurança



1. Solicitar seu certificado digital à um CA (*Certification Authority*) certificado pelo IGTF (*International Grid Trust Federation*)
2. Carregar o certificado no browser
3. Aceitar os “Termos de Uso” do Grid
4. Solicitar sua inscrição em uma das VOs (*Virtual Organization*) disponíveis para o Grid em questão

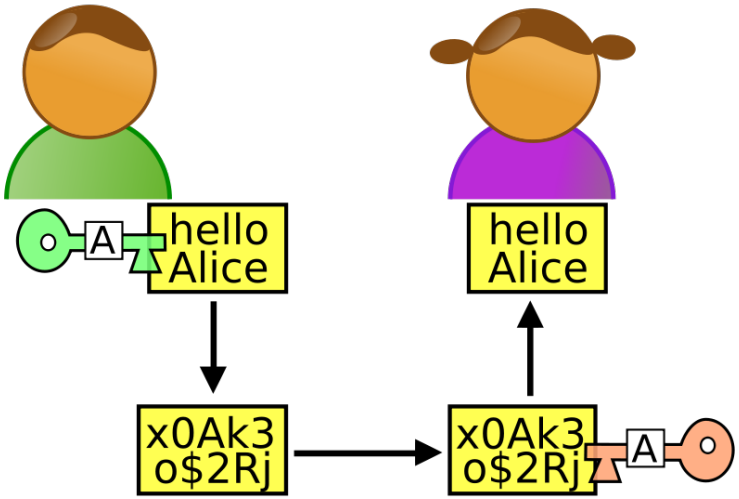
Preocupação com a segurança

- Seu certificado é sua chave para acessar o Grid
- Certificado X.509
- Criptografia de chaves pública e privada

OBS: todos os atores em um Grid (usuários, PCs, instrumentos...) precisam de um certificado

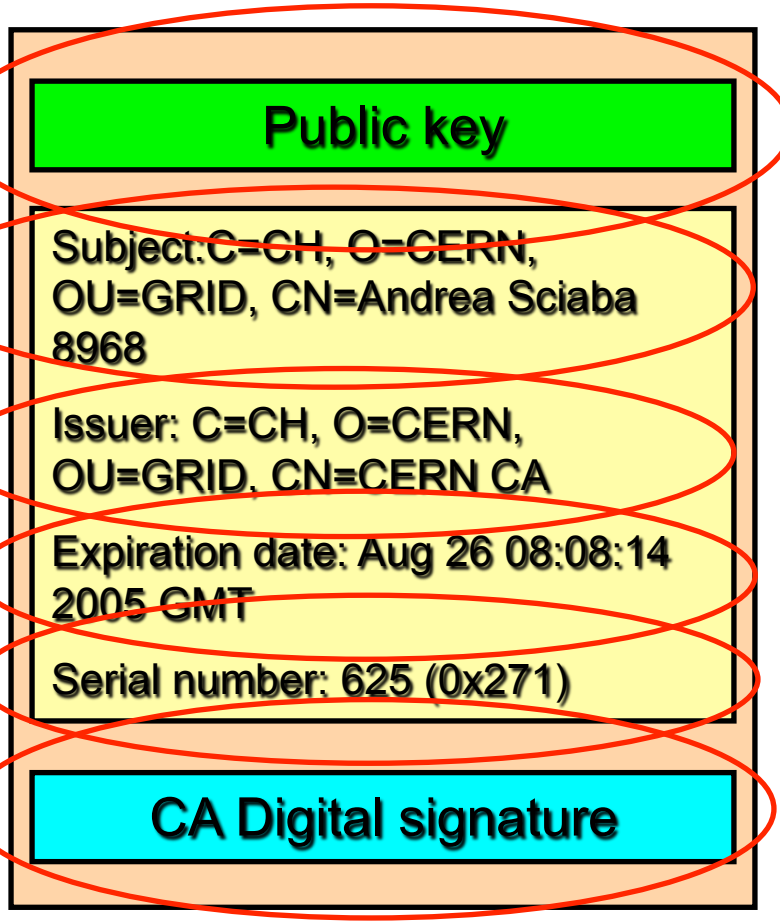
- **Chave pública: distribuída livremente**

- **Chave privada: apenas o dono a possui**

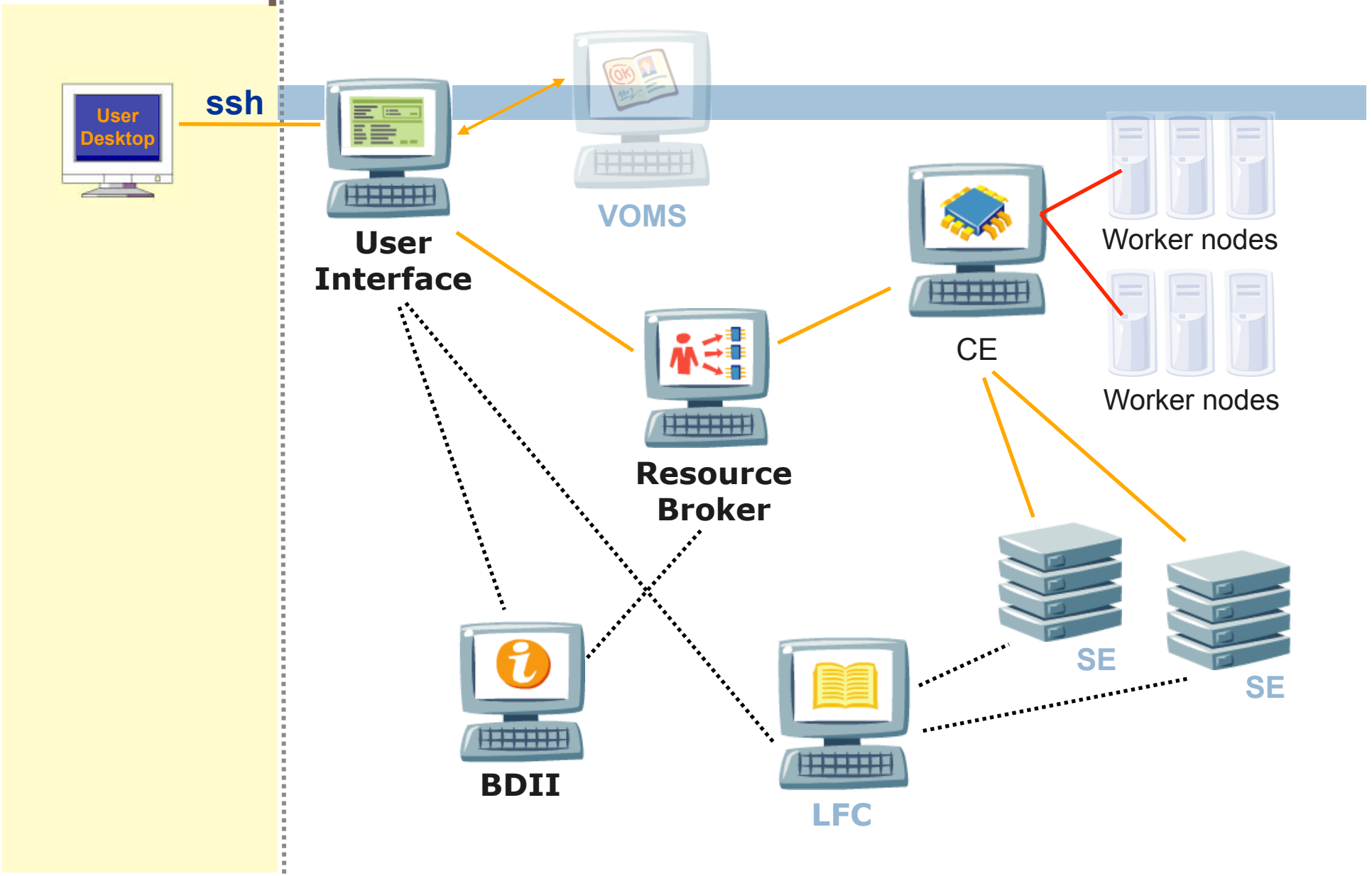


Preocupação com a segurança

- Conteúdo de um certificado X.509:
  - Chave pública do usuário
  - Identidade do usuário
  - Informações sobre o CA
  - Validade
  - Número de serie
  - Assinatura Digital do CA



Arquitetura básica



Concluindo…

- A fase atual do desenvolvimento do Grid pode ser comparada à da Web de 10 anos atrás
- Acredita-se que Grid Computing revolucionará a TI da mesma forma que a Web fez (e está fazendo)
- Atualmente empresas como HP, Sun, Oracle, IBM e Microsoft também estão investindo em pesquisas relacionadas ao Grid
- Instituições que antes eram privadas de pesquisas que exigiam muito poder computacional, agora podem tirar proveito do Grid
- NGIs (Iniciativas Nacionais de Grid) estão sendo criadas em vários países
- A chamada “e-Science” representa um ativo que contribuí para o desenvolvimento de um país

Concluindo…

Alessandro Volta apresenta em Paris,
na presença de Napoleão, a primeira bateria (1801).

Afresco de Nicola Cianfanelli – Museu de Hist. Natural de Florença

Concluindo…


O quê ?!?! Esse cara é louco...
Oh, mon Dieu !
…e no futuro, haverá redes elétricas em todo o mundo…
Power (Electrical) Grid

Prorrogação

Para saber mais…

Grid Café

http://gridcafe.web.cern.ch/gridcafe/

Ian Foster

"Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations" (I.Foster)

Links e contato

- **Slides sobre gLite**
  https://grid.ct.infn.it/twiki/bin/view/EELA2/TrainingOnGLite
- **gLite tutorial – GILDA Wiki**
  https://grid.ct.infn.it/twiki/bin/view/GILDA/UserTutorials
- **What is the Grid?**
  http://access.ncsa.uiuc.edu/witg/
- **iSGTW**
  http://www.isgtw.org/?pid=1000550
- **Open Grid Forum**
  http://www.ogf.org

Plataformas de computação paralela e distribuída

- Execução eficiente de aplicações intensivas em dados ou computação
- Tipos de ambientes:
  - HPC (High Performance Computing)
  - HTC (High Throughput Computing)
- Exs de apps HPC: meteorologia, processamento matemático em geral
- Exs de apps HTC: bioinformática, finanças etc

Cluster - Definição

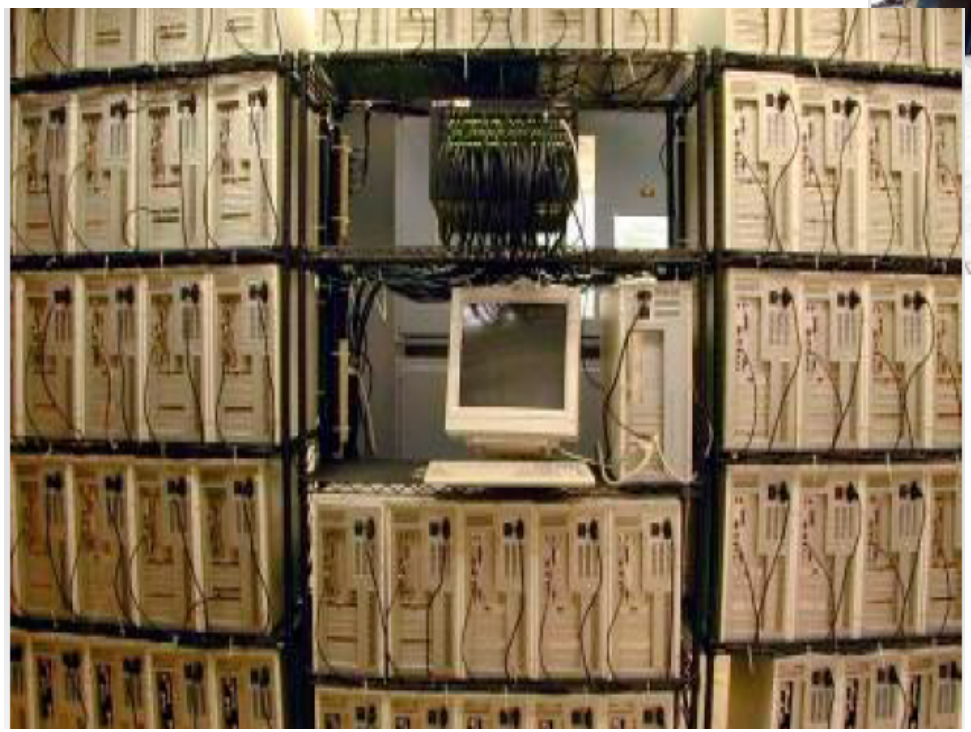
"Cluster is a widely-used term meaning independent computers combined into a unified system through software and networking. At the most fundamental level, when two or more computers are used together to solve a problem, it is considered a cluster"

- http://www.beowulf.org

"Construído a partir de computadores convencionais, os quais são ligados em rede e comunicam-se através do sistema, trabalhando como se fossem uma única máquina de grande porte"

- http://pt.wikipedia.org/wiki/Cluster

Ex



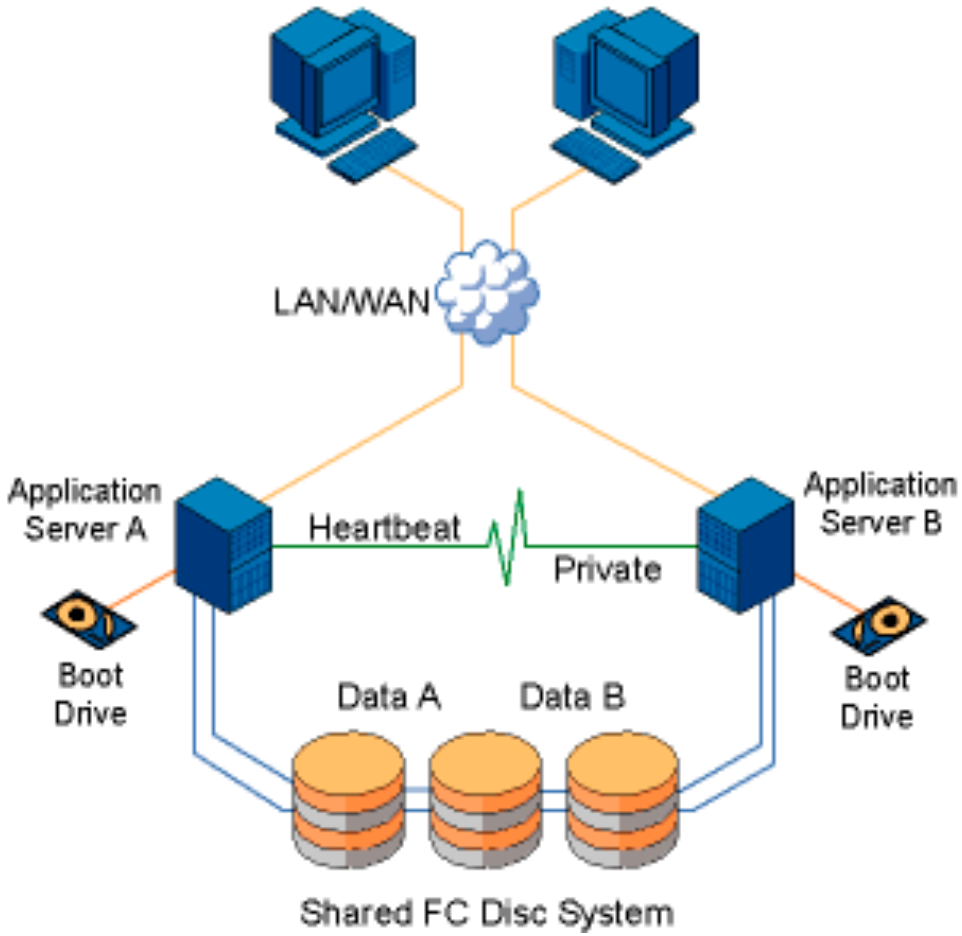
Razões para usar clusters

- Os clusters ou combinações de clusters são utilizados a fim de processar conteúdos críticos ou disponibilização de serviços durante a maior parte do tempo.
  - **Clusters de Alta Disponibilidade e Balanceamento de Carga** geralmente são utilizados por serviços críticos, como aplicações web, vídeo streaming, servidores de email entre outras.
  - **Clusters paralelos** normalmente são utilizados pela indústria cinematográfica a fim de renderizar gráficos de altíssima qualidade e animações.
  - **Clusters Beowulf** são utilizados na pesquisa cientifica, pelo seu poder de processamento e custo de implementação

Tipos de Clusters

- Alta Disponibilidade (*High Availability (HA) and Failover*)
  - Construídos para prover uma disponibilidade de serviços e recursos de forma ininterruptas
  - Se um nó do cluster vier a falhar (failover) as aplicações/ serviços estarão disponíveis em um outro nó.
  - Utilizados para base de dados de missões críticas, correio, servidores de arquivos e aplicações.
  - Replicação de Serviços e Servidores.
  - Tolerância a falha através de: Raid, fontes, placas e links redundantes
  - Exemplos:
    - Linux HA - http://www.linux-ha.org
    - DRBD - http://www.drbd.org/

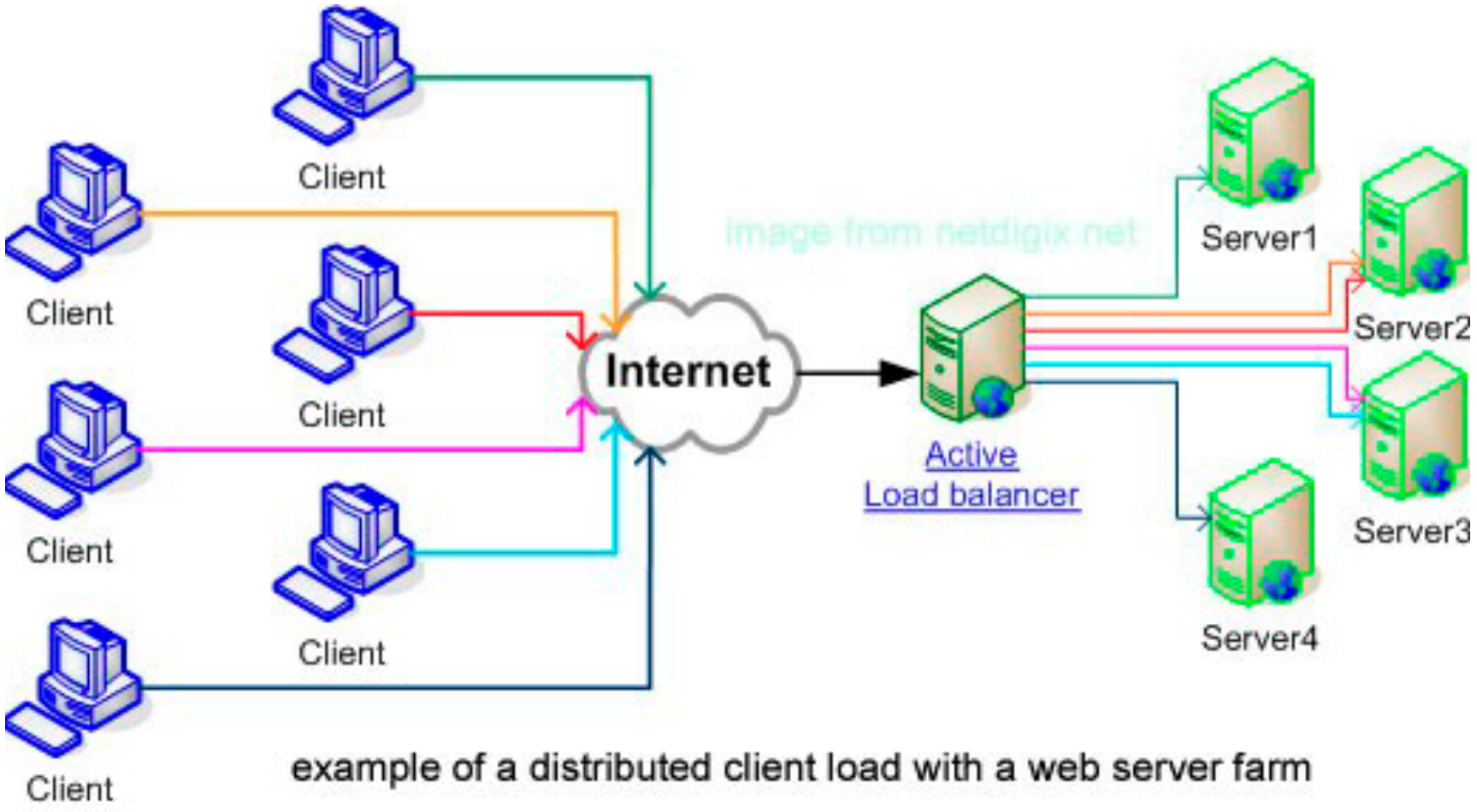
Alta Disponibilidade HA



Tipos de Clusters

- Balanceamento de carga (Load Balancing)
  - Distribui o tráfego ou requisições entre as máquinas que compõem o cluster
  - Se um nó falhar, as requisições são redistribuídas entre os nós disponíveis no momento.
  - Os sistemas não trabalham junto em um único processo, mas redirecionando as requisições de forma independente, baseados em um escalonador e um algoritmo próprio
  - Utilizados para fazenda de servidores web (web farms)

Balanceamento de carga



Tipos de Clusters

Alguns exemplos de soluções para Balanceamento de carga:

- Linux Virtual Server -
  - http://www.linuxvirtualserver.org/
- Zeus Load Balancer -
  - http://www.zeus.com/products/zlb/
- .vantronix Load Balancer -
  - http://www.vantronix.com/
- Kemp Technologies -
  - http://www.kemptechnologies.com/
- Barracuda -
  - http://www.barracudanetworks.com/
- F5 Load Balancer –
  - http://www.f5.com/

Tipo de Clusters

- Processamento Distribuído ou Processamento Paralelo (HPC - High Performance Computing)
  - Aumenta a disponibilidade e performance para as aplicações, particularmente as grandes tarefas computacionais
  - Uma grande tarefa computacional pode ser dividida em pequenas tarefas que são distribuídas ao redor dos nodos, como se fosse um supercomputador massivamente paralelo
  - Utilizados para computação cientifica ou análises financeiras, tarefas típicas para exigência de alto poder de processamento.
  - Exemplos:
    - Beowulf Cluster - http://www.beowulf.org/
    - LinuxHPC - http://www.linuxhpc.org

Tipo de Clusters

- Beowulf Cluster
  - É o nome de um projeto para aglomerados de computadores (ou Clusters) para computação paralela, usando computadores pessoais, não especializados e portanto mais baratos
  - O projeto foi criado por Donald Becker da NASA
  - Possui desempenho escalável.Baseados numa infraestrutura de hardware comum, rede privada e software 'open source' (Linux)
  - Existe um servidor responsável por controlar todo ocluster, principalmente quanto à distribuição de tarefas e processamento.

Beowulf Cluster

Alta Discponibilidade

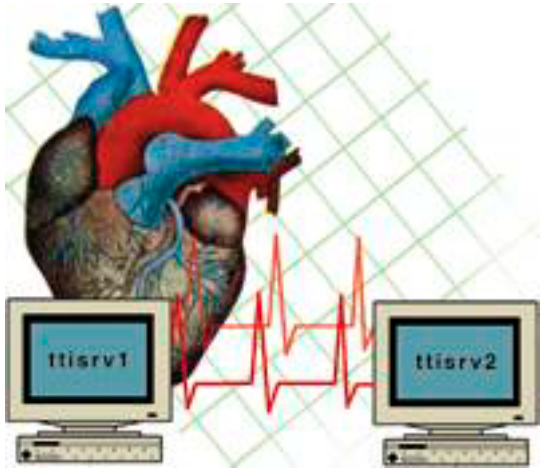
- “Um sistema de alta disponibilidade é aquele que utiliza mecanismos de detecção, recuperação e mascaramento de falhas, visando manter o funcionamento dos serviços durante o máximo de tempo possível, inclusive no decurso de manutenções programadas”
- “Disponibilidade refere-se a capacidade de um usuário de determinado sistema acessar, incluir ou modificar os dados existentes em qualquer intervalo de tempo. Caso, por qualquer que seja o motivo, um usuário não tenha acesso, é dito então que ele está indisponível, sendo o tempo total de indisponibilidade conhecido pelo termo downtime.”

Heartbeat

- “Heartbeat é um deamon que provê uma infraestrutura de serviço de cluster (comunicação e associação de membros) para seus clientes. Ele permite que os clientes tomem conhecimento sobre a presença (ou desaparecimento) dos processos em outras máquinas (peers/nodes) e de forma fácil, trocar mensagens com ele.
  - http://www.linux-ha.org/doc/ch-fundamentals.html

Heartbeat

- O Heartbeat é um dos componentes do
- projeto Linux-HA (*High-Availability* Linux);
  - Roda nas plataformas Linux, FreeBSD e Solaris;
  - Detecta a morte de um 'host' e gerencia cluster.

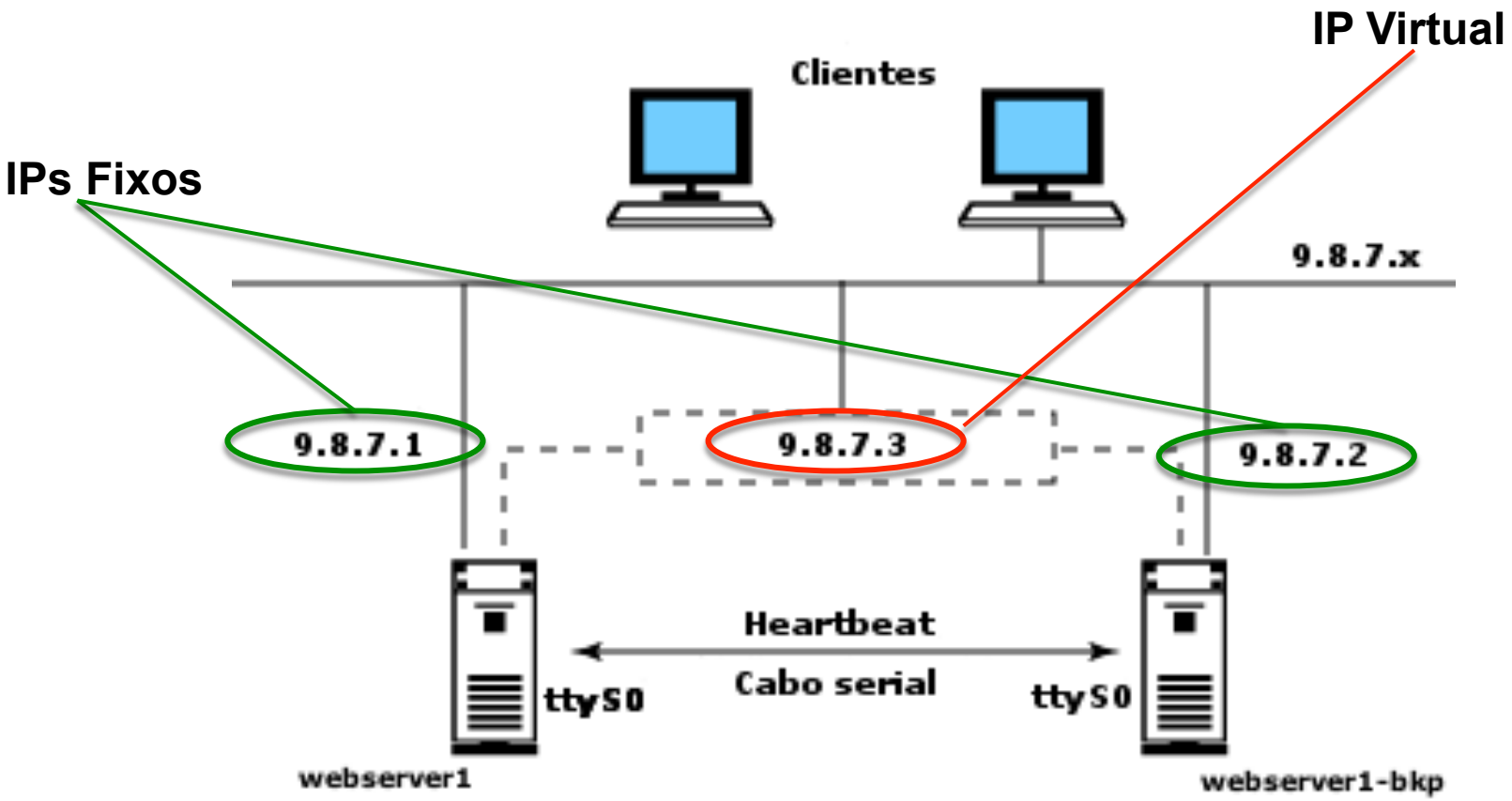


Heartbeat - Funcionamento

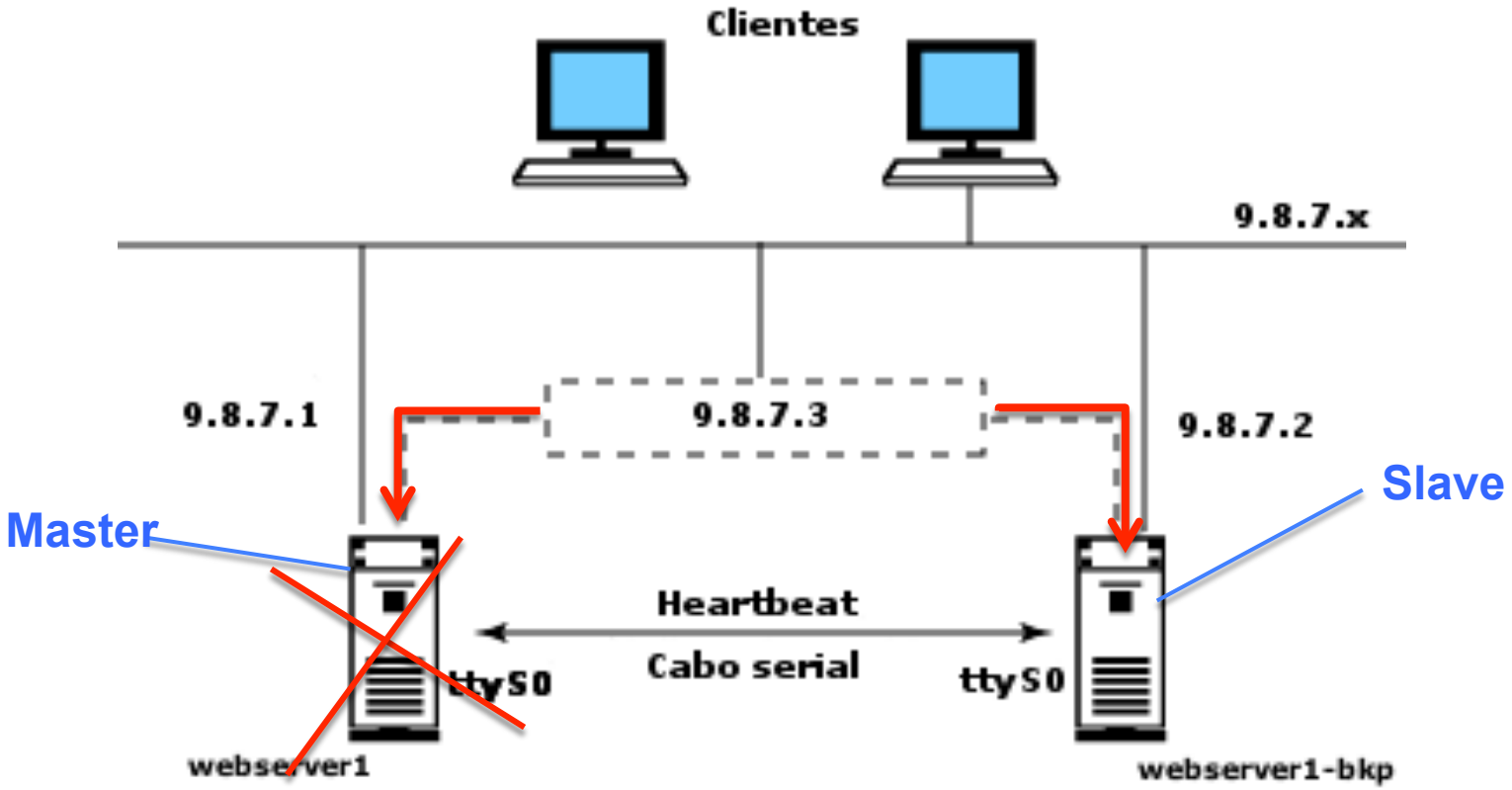
- Através de um meio de comunicação, que pode ser Ethernet ou Serial, um servidor redundante verifica a disponibilidade do servidor em produção. Essa checagem é feita entre as duas instâncias do Heartbeat instaladas nos dois servidores. Se o servidor em produção não responder, ele será considerado indisponível, e então o Heartbeat do servidor redundante providencia a configuração e inicialização dos serviços locais, além de outros recursos, como o endereço IP, partições de disco, etc.

- Segmentos UDP são enviados regulamente entre os hosts;
- Se o segmento não for recebido;
- Será detectado que um host está com problema;
- E é tomada uma ação;
- Quando o serviço HeartBeat é iniciado em um host uma Interface Virtual sobe;
- Essa Interface Virtual será acessada pelos clientes;
- Se esse host falhar, então será detectado e a interface do outro host subirá como o mesmo IP;

Heartbeat - Funcionamento



Heartbeat - Funcionamento



Considerações

- Evita qualquer tipo de conflito que possa afetar o
- correto funcionamento do sistema.
  - Não é seu objetivo garantir a sincronia e a integridade dos dados entre os servidores.
  - Necessário atuar em conjunto com algum software que se encarregue de manter os mesmos arquivos do servidor em produção também no servidor redundante.

Instalação

- Para instalar o heartbeat utilizando o yum, basta executar o comando abaixo:

  yum install heartbeat

  Irá instalar os pacotes abaixo:

  - heartbeat-2.1.3-3.el5.centos
  - heartbeat-pils-2.1.3-3.el5.centos
  - heartbeat-stonith-2.1.3-3.el5.centos

Configuração

Toda instalação do heartbeat deve conter os seguintes arquivos de configuração:

- **/etc/ha.d/ha.cf** — Arquivo global de configuração do cluster
- **/etc/ha.d/authkeys** — Arquivo que contem chaves para autenticação mútua entre os nodos da rede
- **/etc/ha.d/haresources** — Arquivo que contem os recursos que queremos habilitar no cluster

Configuração

- **Uma configuração básica para ha.cf é:**

```
logfile /var/log/ha-log
logfacility local0
Keepalive 2  # Intervalo entre os heartbeats
Deadtime 30 # Define quando um nó está offline
Initdead 120 # Declara que o node está offline após
             #o startup. Deve ser alto
bcast bond0 # Qual interface os heartbeats serão
            # enviados
Udpport 694 # porta UDP utilizada para intra-cluster
            # communication
auto_failback on  # Retorna serviço para master
node server3 # nome das máquinas do cluster
node server4 # nome das máquinas do cluster
```

Configuração

Configuração authkeys

- Este arquivo possui as chaves de autenticação a serem utilizadas pelos nodes. Abaixo está um exemplo deste arquivo:

```
auth 1
1 sha1 8499ffe31ca6edc6998ec54ac99c009b
```

- Este arquivo deverá ser legível apenas pelo root, para tanto:

```
chmod 600 /etc/ha.d/authkeys
```

Configuração

Configuração haresources

- Este arquivo possui a lista dos recursos que serão movidos de um nó para o outro quando um nó entra no status de falha ou quando ele se recupera
- Este arquivo deve ser igual para todos os nós do cluster
- Cada linha indica um grupo de recursos que estará ativo.

Exemplo:

server3	192.168.15.50	httpd
Server4	192.168.15.51	vsftpd

Testando

- **No Nó 1:**
  - echo "Apache ativo no node01" > /var/www/html/index.html
- **No Nó 2:**
  - echo "Apache ativo no node02" > /var/www/html/index.html\
- **Forçando a Falha do nó 1:**
  - /etc/init.d/heartbeat start ou ainda:
  - [root@server3 ha.d]# /usr/share/heartbeat/hb_standby 2010/04/18_14:09:14 Going standby [all].
  - Automaticamente será detectada a falha do nó 1 e o heartbeat ativará os serviços no nó2.
- **Recuperando a Falha do nó 1:**
  - [root@server3 ha.d]# /usr/share/heartbeat/hb_takeover