

Eventos IBM España, Madrid - 19 Septiembre 2012

Sistema de Supercomputación ALTAMIRA

Abierto a la Innovación



Jesús Marco de Lucas

Profesor de Investigación del CSIC

Instituto de Física de Cantabria (IFCA)

CENTRO MIXTO CSIC-UNIVERSIDAD DE CANTABRIA

marco@ifca.unican.es

(agradecimiento especial a Luis Cabellos)

El sistema ALTAMIRA

- ✚ El contexto técnico: el sistema HPC
- ✚ El contexto científico:
 - ✚ El Instituto de Física de Cantabria
 - ✚ La Red Española de Supercomputación
 - ✚ e-Ciencia
 - ✚ Cantabria Campus Internacional
- ✚ El contexto de innovación:
 - ✚ USO **NO CONVENCIONAL** DE UN SUPERCOMPUTADOR
- ✚ **Business model**

ALTAMIRA como sistema HPC

Cluster HPC (High Performance Computing):

- Cluster HPC IBM idataplex con 240 nodos dx360m4
 - 2x SandyBridge-EP E5-2670 2.6GHz/1600 20M 8-core 115W
 - 64GB RAM, 16x4G DDR3-1600 DIMMs (4GB/core)
 - 500GB 7200 rpm SATA II local HDD
 - 332.8Gflop/node
- Red HPC Infiniband FDR10 (40 Gbps)
 - FDR10 IB HCA Mellanox
 - Switches 36 bocas, leafs+ core layers : FAT TREE non-blocking
- Red de gestion avanzada

Almacenamiento HPC:

- Sistema GPFS conectado por IB y por 10G
- Capacidad total >2 Petabytes en disco, + archivado LTO5

Otros: cluster GPU, cluster Power7

ALTAMIRA como sistema HPC

RACKS NODOS COMPUTACION

RACK GESTIÓN + GPFS



ALTAMIRA como sistema HPC

Factores clave

Rendimiento de los nodos en todos los entornos

- Grupos de investigadores que requieren x86
- Última generación de procesadores
- x2 flops en Sandy Bridge, ok para top500...



Coste y consumo energético

- Nodos ajustados a requerimientos de computación
- Consumo total: reducción de un 40% frente a sistema anterior
- Gestión energética
 - A pesar de usar la configuración “HPC”

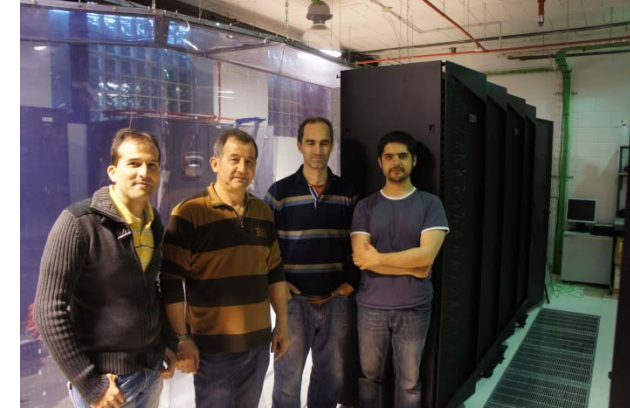
Garantía

- Confianza en soluciones IBM e INTEL
 - Primer cluster en 2001 (“hibernando...”) 80 nodos x220 PIII
 - 2005: 90 nodos x336 Xeon, 2008: 180 nodos HS21 E5420 (en producción)

ALTAMIRA como sistema HPC

Software:

- ❑ XCAT (gestión de la instalación)
- ❑ Scientific Linux
- ❑ SLURM (gestor de colas)
- ❑ MPI (mvapich2-x86_64 openmpi-x86_64)
- ❑ gcc (compiladores)



Rendimiento:

- ❑ Sistema instalado en Junio 2012, acepta jobs desde Julio
- ❑ un ejemplo típico, mail de un investigador:

“os mando algunos datos en relación a los tiempos de un programa que se usa para la producción y análisis de datos de XXXXX. Tiempo de ejecución:

*(i) Procesador Intel SB: de 45 min a 60 min **

(ii) Procesador IBM Power 7: 100 min como mínimo

*(iii) Procesador AMD: de 80 min a 180 min **

* Dependiendo de la carga de trabajo de cada nodo. Comprobamos que los AMD son bastante sensibles al número de jobs corriendo en cada nodo. Con respecto al nuevo sistema Altamira, parte de la variación podría explicarse por ralentización de I/O cuando todos los jobs (del orden de 1000) quieren escribir/leer en disco 1 GB de datos.

El contexto científico: IFCA

✦ Instituto de Física de **CANTABRIA**

- Centro Mixto Universidad de Cantabria-CSIC para la investigación en “Ciencia Básica”:

Understand the basic components and structures in the Nature:
from the elementary ones (Particle Physics)
to the most distant structures in the Universe (Astronomy and Space Sciences)
as well as the collective behaviour of the matter (Statistical & Non-Linear Physics)

✦ **Datos Básicos:**

- Creado en 1995, en el Campus de la Universidad de Cantabria (Santander)
 - Ed. Juan Jordá: ~2000 m², labs de metrología & electronica, sala limpia, CPD
- Personal: > 100 (30 investigadores en plantilla, 20 postdocs +contratos, becarios, etc.)
- Proyectos de investigación competitivos (nacionales y europeos)
 - Participación en colaboraciones internacionales (CERN, ESA, ESO, FNAL, EGI.eu...)
 - Investigadores del IFCA en comites de máximo nivel: ESO Council, CERN SPC, EGI Exec.Board
- ~ 150 artículos anuales en revistas de primera línea
 - Ej._ artículo reciente en Science: descubrimiento del bosón de Higgs

Centro relevante a nivel nacional con clara visibilidad internacional!
Primer grupo de investigación de la Universidad de Cantabria
Centro relevante dentro del área de Física en el CSIC

Recursos del IFCA



- ❁ Laboratorio de Electrónica
- ❁ Laboratorio de Metrología
- ❁ Sala Limpia
- ❁ Salas de Computación
- ❁ Telescopio
- ❁ Sala de Ensayos
- ❁ Sala de Ensayos Térmicos
- ❁ Sala de Mecanizado



IFCA: Sala de computación / CPD

- ✧ Sala en sótano, 120 m2, acondicionada CPD (suelo, refrigeración, UPS...)
 - ✦ Hasta ~160 KW con “direct external cooling” (entrada aire externo + 2 Hiross UA66 + expulsión aire caliente al exterior)
- ✧ Sección “GRID”:
 - ✦ 250 servidores, todos con virtualización
 - ✦ 25% tiempo asignado a Ibergrid y a la European Grid Initiative
 - ✦ Almacenamiento: GPFS (>2 Petabytes)
 - ✦ Red: backbone 10Gb + salida directa a RedIris Nova (hasta 10Gbps)
 - ✦ Servicio cluster, servicios GRID a diferentes comunidades: HEP, Astro, Medio Ambiente, Ingeniería, Bio, Medicina, etc. , + acuerdo CNRS (grid/cloud)

uCloud
(UC-CSIC)

Cluster Tier-2
CMS (UC)

Almacenamiento
Tier-2 CMS (UC)

Petabyte
IFCA-CSIC

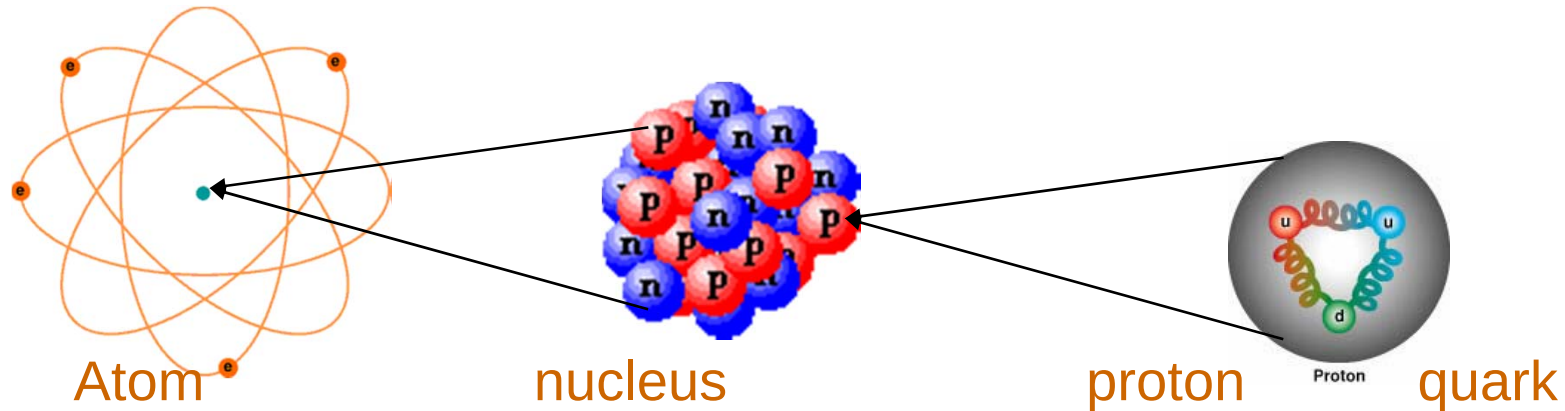
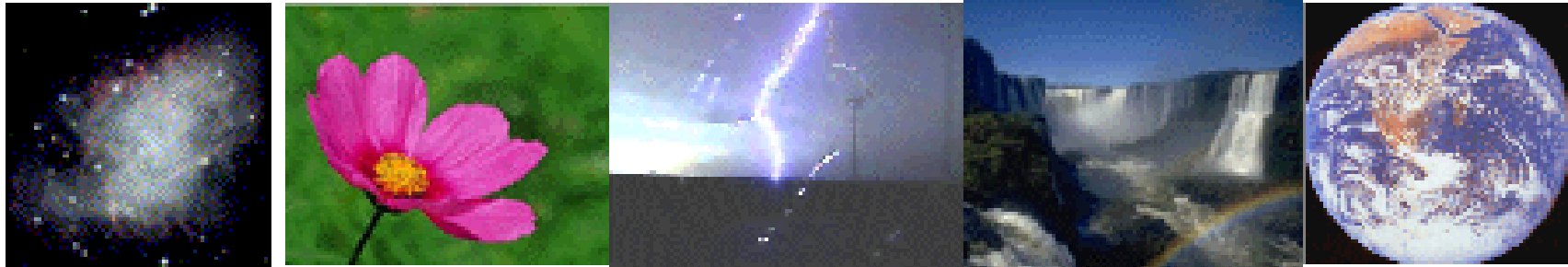
Red ifca.es
CSIC

GRID-CSIC



Searches in Particle Physics

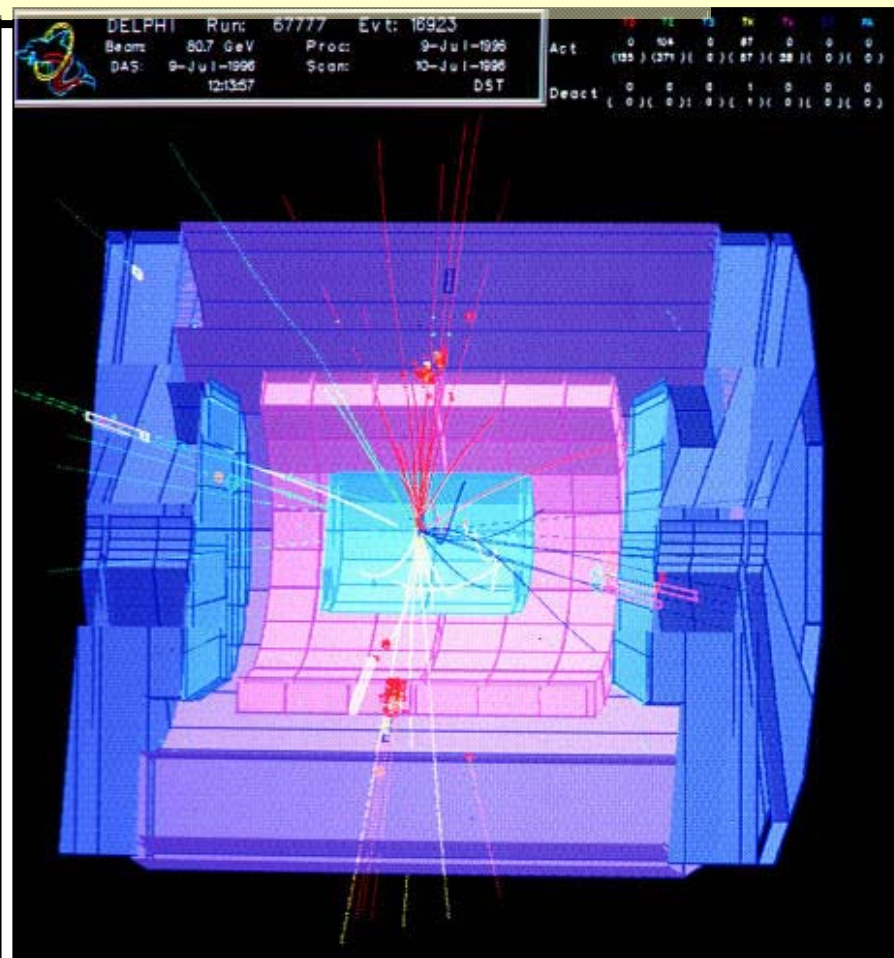
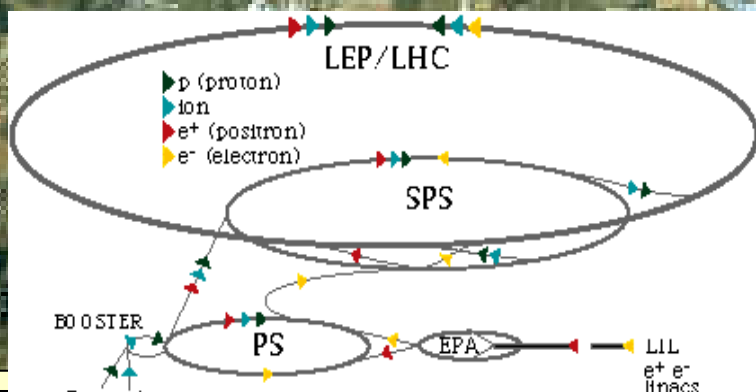
- Particle physics: studying the basic constituents of all matter around!



The origin of the “*mass*” of all particles is linked to a fundamental particle predicted, and now discovered: **the Higgs boson**

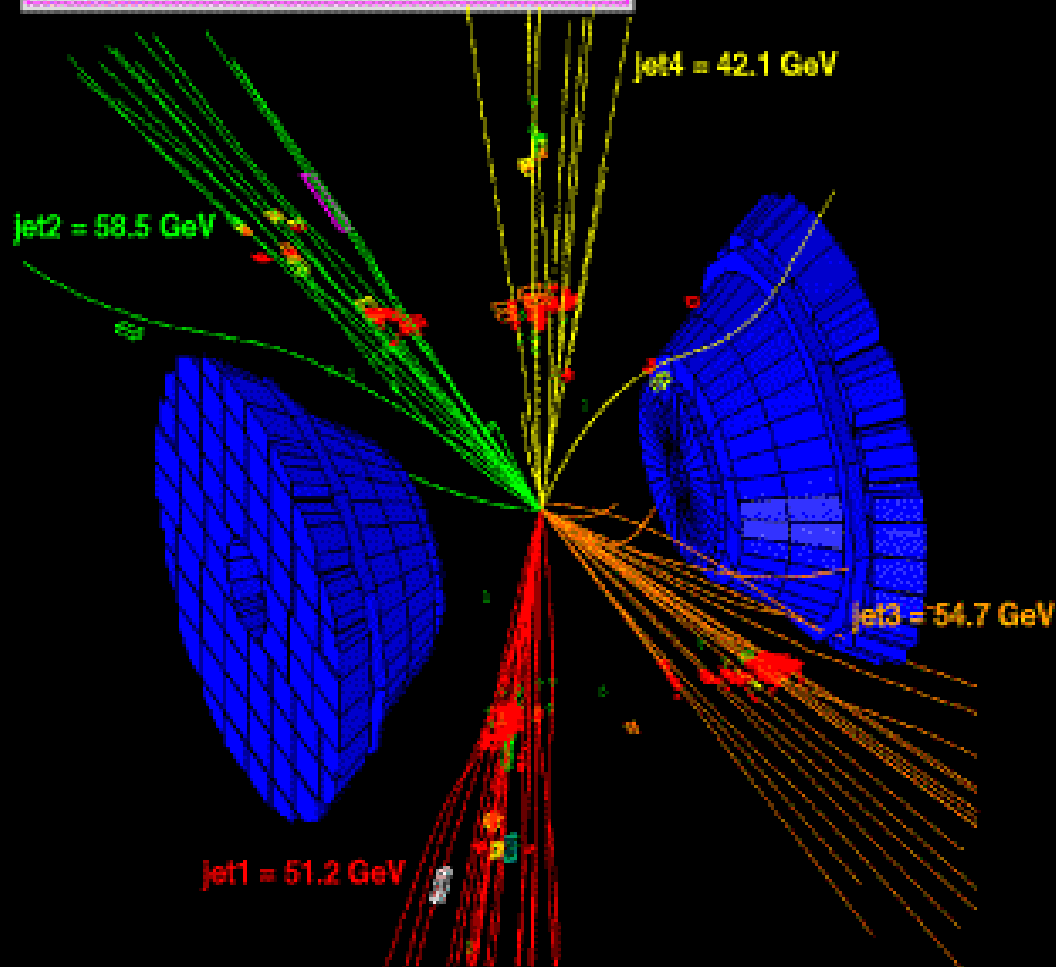
Accelerators and detectors

CERN Lab (Geneve, Switzerland)



LEP (ended in 2000)
Large Electron Positron Collider
 $E_{cm} = 200 \text{ GeV}$, e⁺e⁻ collisions
Search for Higgs up to $M=115 \text{ GeV}$

DELPHI Run: 104574 Evt: 5797
 Beam: 103.3 GeV PBeam: 24-Aug-2000
 DAQ: 24-Aug-2000 Run: 28-Aug-2000
 1700000 Tau+10T



4C fit :

1st pairing hypothesis

$$M_{j1j2} = 101.7 \text{ GeV}/c^2$$

$$M_{j3j4} = 86.4 \text{ GeV}/c^2$$

2nd pairing hypothesis

$$M_{j1j4} = 98.9 \text{ GeV}/c^2$$

$$M_{j2j3} = 105.9 \text{ GeV}/c^2$$

5C fit Z mass :

$$b\text{-tag} (j_1, j_2) = +7.26 \rightarrow M_{j1j2} = 97.4 \text{ GeV}/c^2$$

$$b\text{-tag} (j_3, j_4) = -0.16 \rightarrow M_{j3j4} = M_Z$$

$$b\text{-tag} (j_1, j_4) = +1.43 \rightarrow M_{j1j4} = M_Z$$

$$b\text{-tag} (j_2, j_3) = +5.67 \rightarrow M_{j2j3} = 113.4 \text{ GeV}/c^2$$

⊗ Higgs Boson decays into two heavy particles:

- ⊠ B quarks
- ⊠ Z or W bosons

⊗ Complex event characteristics:

- ⊠ Shape
- ⊠ Particles lifetime
- ⊠ Masses

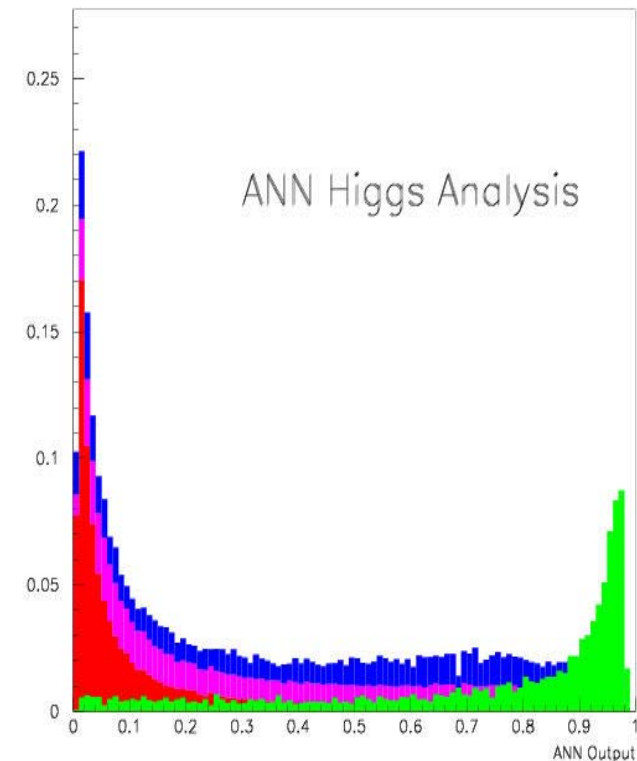
⊗ S/B ratio is extremely low:

- ⊠ LEP: 10 in 10^6
- ⊠ LHC: 10 in 10^9

NEURAL NETWORKS
are used to
optimize the search

How did we select that “collision”?

- ANN: example of architecture 16-10-10-1
 - 16 input variables
 - 2 hidden layers with 10 nodes each
 - 1 output layer, 1=signal, 0=background
- Trained on MC sample
 - Higgs generated at a given mass value
 - All types of Background
 - 10x real data statistics
- Applied on real collected data to order in S/B the candidates to Higgs boson
- Training process:
 - Minimize classification “error”
 - Iterative process
 - No clear “best strategy”
- Computing Intensive: hours to days for each “try”



Distributed Analysis

✧ Distributed Configuration:

- ✧ Master node and N worker nodes.
 - Scan to filter events & select variables
 - ResultSet in XML, split according to N (number of slave nodes)

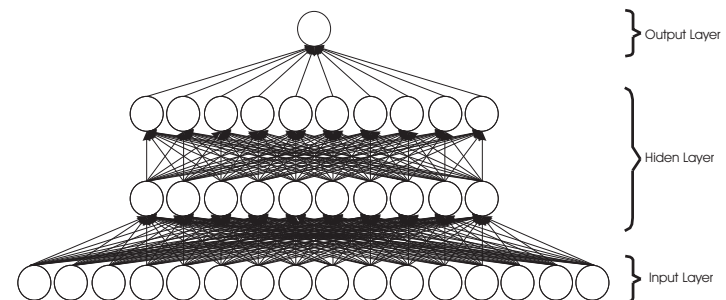
✧ Training procedure:

- ✧ Master reads input parameters and sets the initial weights to random values.
- ✧ The training data is distributed to the workers.
- ✧ At each step:
 - The master sends the weights to the workers.
 - The workers compute the error and the gradient and return them to the master.

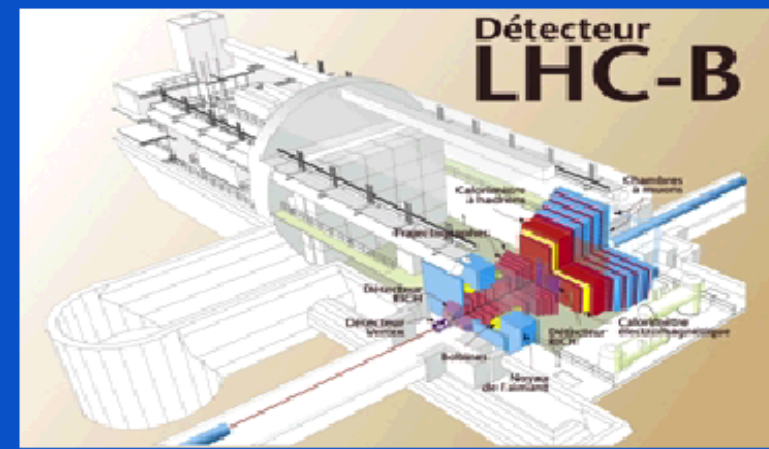
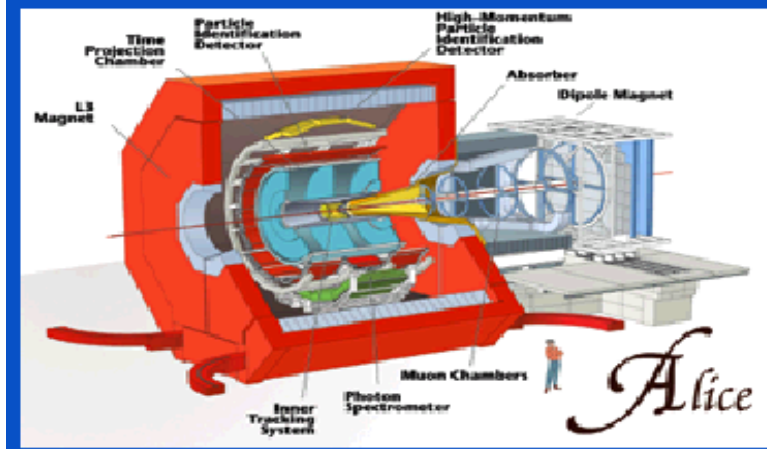
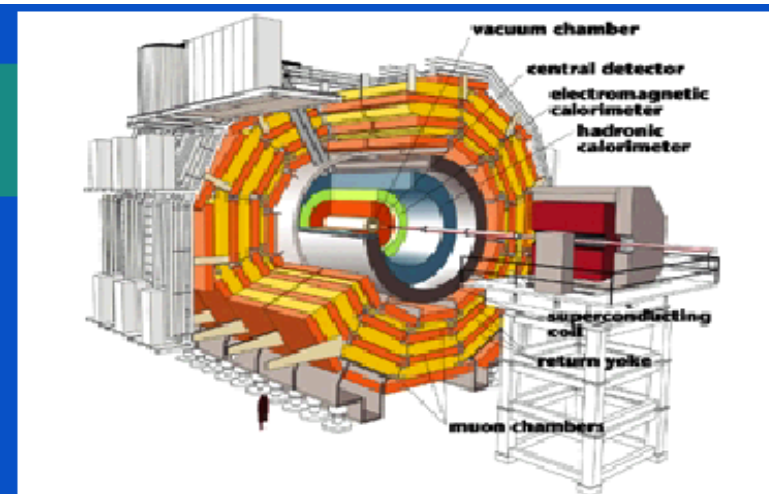
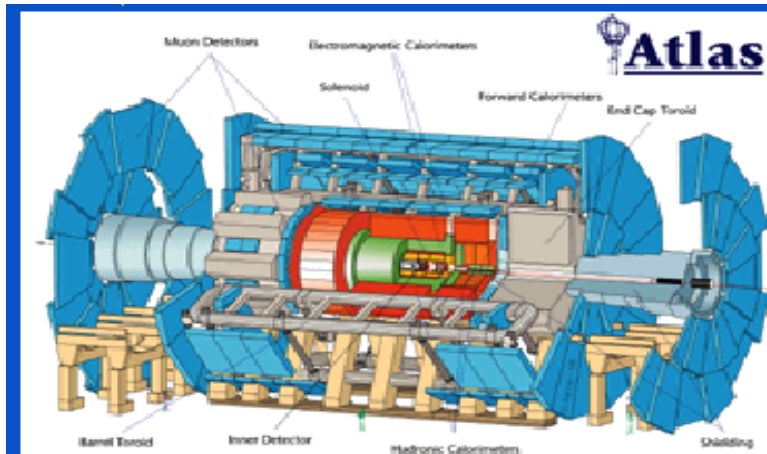
✧ This training procedure has been implemented using **MPI** and adapting the MLP-fit package.

✧ Conditions:

- ✧ train an ANN with 650.000 simulated realistic LEP events, 20000 of them corresponding to signal events.
- ✧ Use a 16-10-10-1 architecture (270 weights)
- ✧ Need 1000 epochs training.
- ✧ Similar sized samples for the test.
- ✧ BFGS training method.

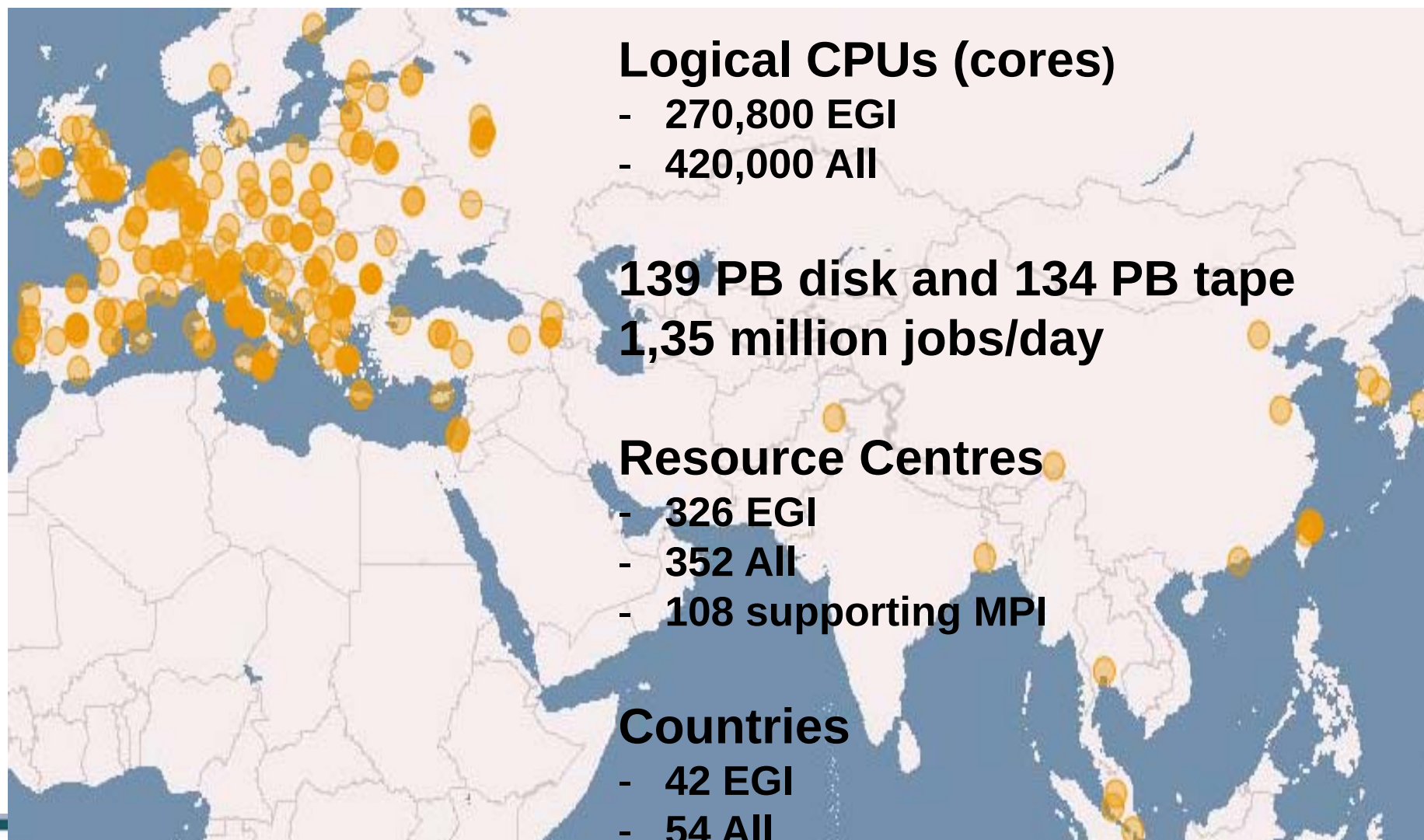


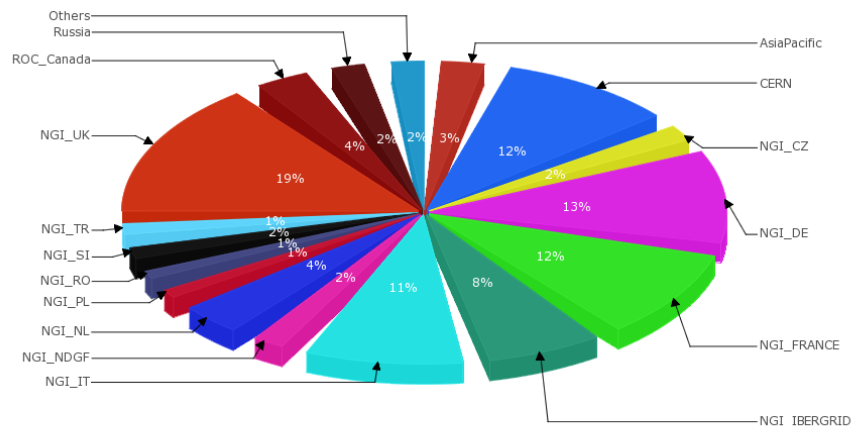
LHC Experiments



European Grid Infrastructure

(2012)



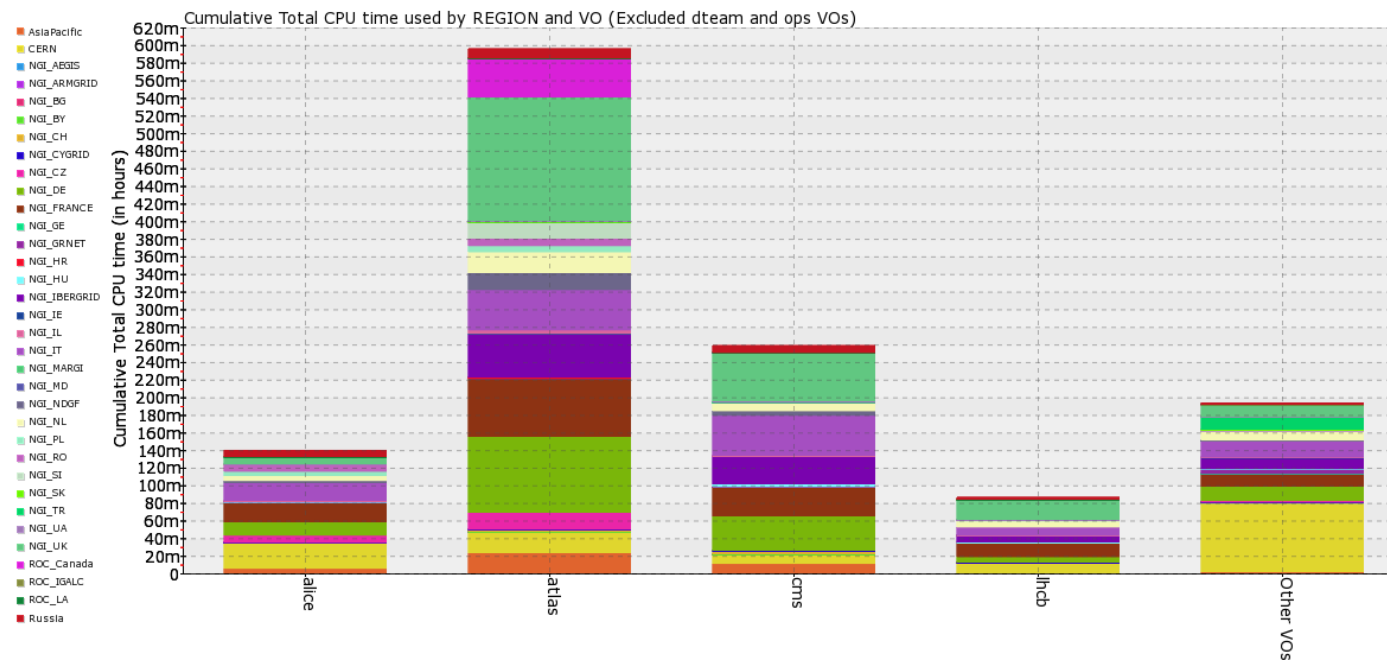


The GRID success!

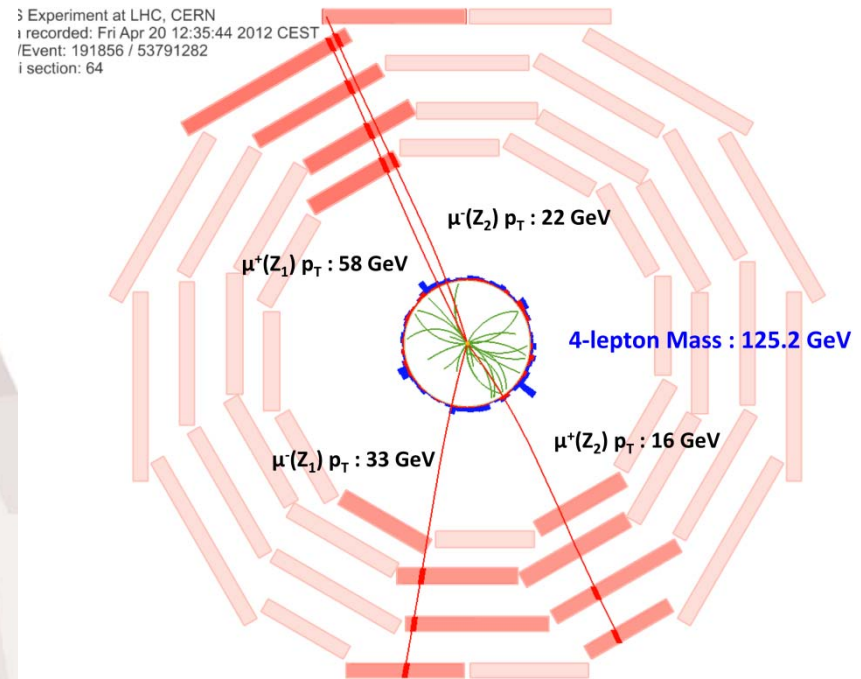
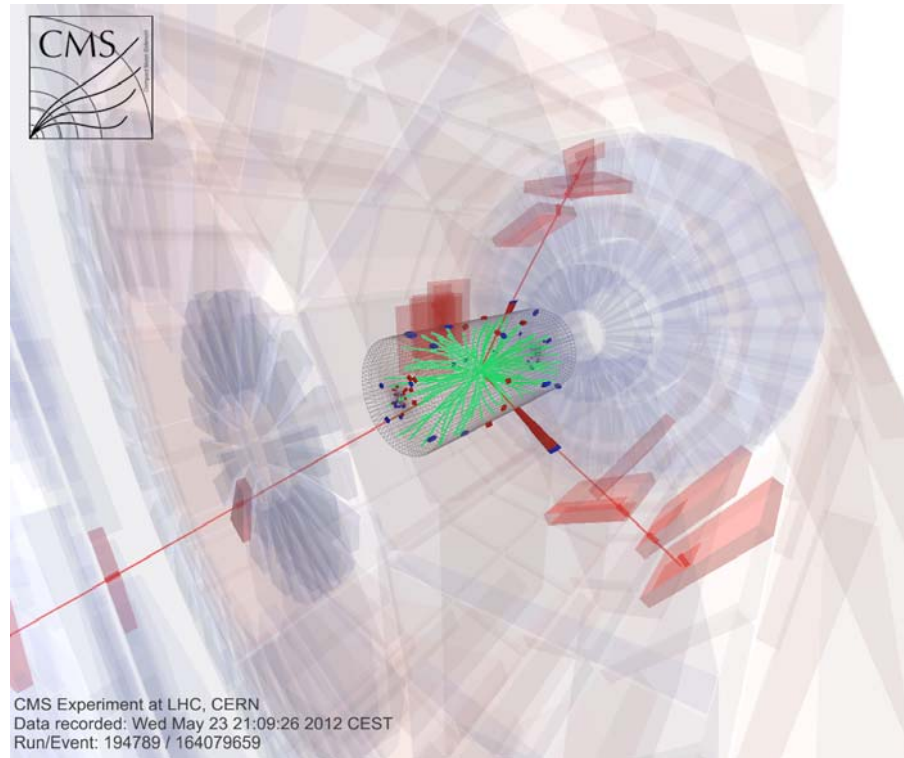
**>1.000 MILLION HOURS
IN 1 YEAR
(80 MILLION IN IBERGRID)**

Developed by CESGA EGI View: / sumcpu / 2011:10-20129 / REGION-VO / lhc (i) / ACCBAR-LIN / x

2012-09-18 00:11

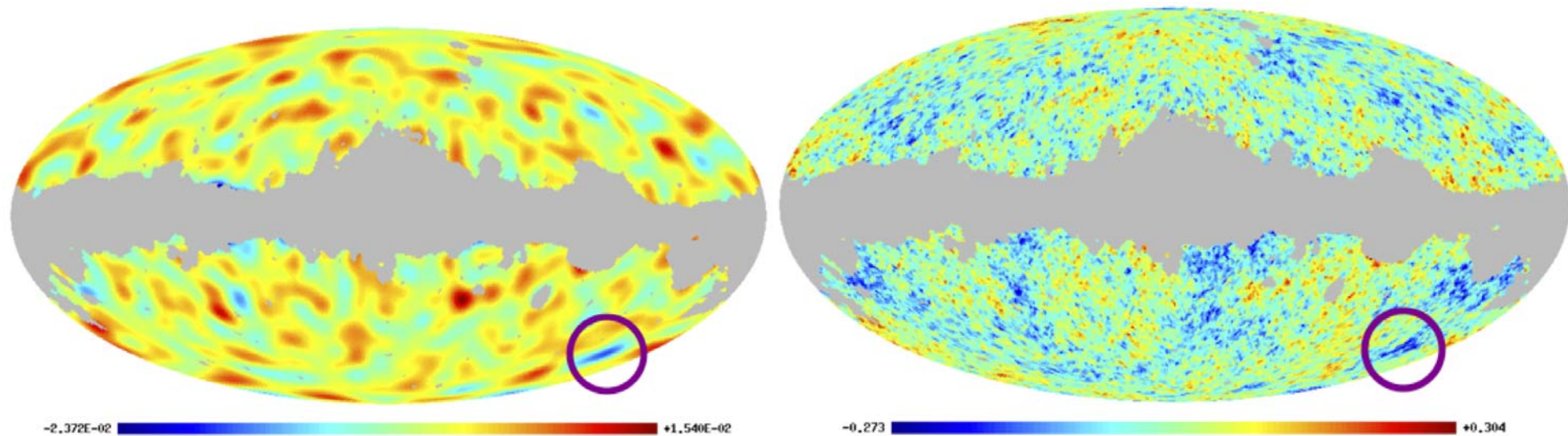


LHC collisions: a Higgs event?

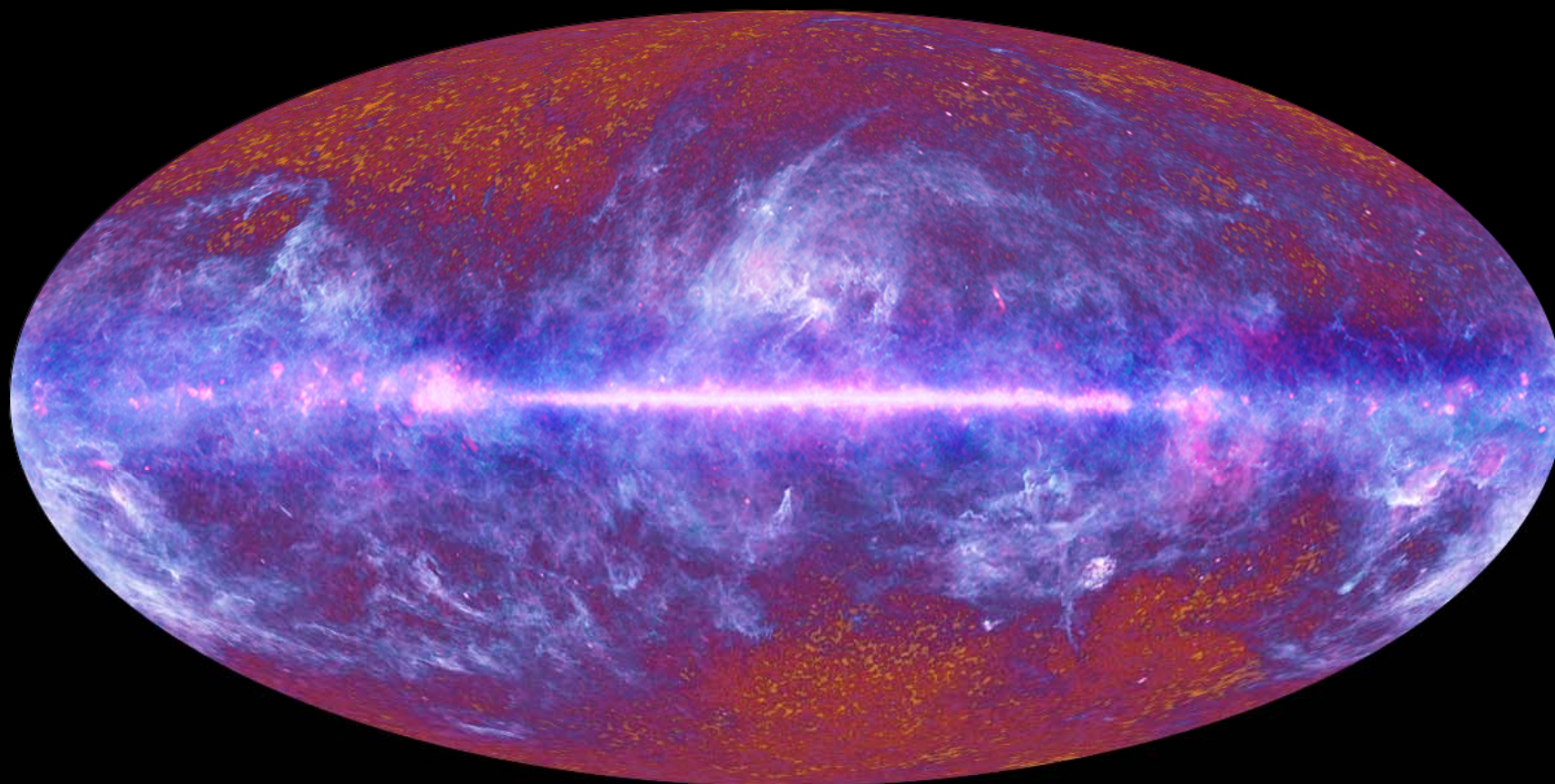


Find 1 event out of 10 million of similar events!
-SIMULATION OF SIGNAL, BACKGROUND (HTC hobs)
-PATTERN RECOGNITION (ANN, PARALLEL JOBS)

Analysis: CMB non-gaussianity sources



- ✚ Example of analysis on WMAP (P.Vielva, E.Martinez, et al. 2004)
- ✚ Skewness and the kurtosis of the SMHW (spherical Mexican hat wavelet) coefficients are calculated at different scales (ranging from a few arc minutes to tens of degrees).
- ✚ A non-Gaussian signal is detected at scales of the SMHW around 4° (size in the sky of around 10°)



The Planck one-year all-sky survey



(c) ESA, HFI and LFI consortia, July 2013



J.Marco, Madrid 19 Sep. 2012

RES (Red Española de Supercomputación)

- « La Red Española de Supercomputación o RES, es una infraestructura virtual distribuida al servicio de la I+D en España consistente en la interconexión de un conjunto de supercomputadores que en régimen de trabajo compartido, gestionan de forma conjunta y eficiente su capacidad de cálculo.
- « La RES ofrece a la comunidad científica:
 - Recursos de supercomputación
 - Servicios de disco
 - Servicios de movilidad
 - Soporte técnico
 - Formación técnica para usuarios
 - Conferencias dirigidas a un área concreta de la ciencia
 - Disseminación de la investigación realizada
- « Desde su creación en julio de 2006, la RES ha contribuido al desarrollo de más de 1950 proyectos científicos a través de los recursos de supercomputación que gestiona.



Instituciones y equipos que forman parte de la RES



BSC-CNS (MareNostrum)

Processor: 10240 PowerPC 970 2.3 GHz
Memory: 20 TB
Disk: 280 + 90 TB
Network: Myrinet

BSC-CNS (MinoTauro)

Processor: 128 nodes: 2 Nehalem + 2 M2090
Memory: 3 TB
Network: IB QDR

BSC-CNS (Altix)

Processor: SMP 128 cores
Memory: 1,5 TB

UPM (Magerit II - renovado en 2011)

Processor: 3.920 (245x16) Power7 3.3 GHz
Memory: 8700 GB
Disk: 190 TB
Network: IB QDR

IAC, UMA, UC, UZ, UV (LaPalma, Picasso, Altamira, Caesaraugusta, Tirant)

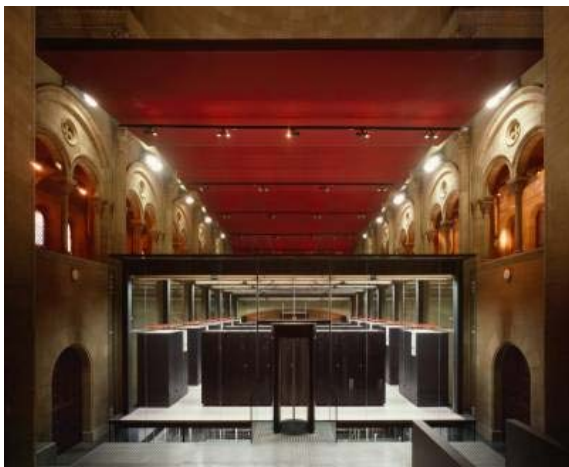
Processor: 512 PowerPC 970 2.2 GHz
Memory: 1 TB
Disk: 14 + 10 TB
Network: Myrinet

Gobierno de Islas Canarias - ITC (Atlante)

Processor: 336 PowerPC 970 2.3 GHz
Memory: 672 GB
Disk: 3 TB
Network: Myrinet



Supercomputadores de la RES



MareNostrum
BSC-CNS



CaesarAugusta
Universidad de Zaragoza



Magerit
Universidad Politécnica de Madrid



Tirant
Universitat de València



Atlante
Instituto Tecnológico de Canarias



La Palma
Instituto Astrofísico de Canarias



Picasso
Universidad de Málaga



Altamira
Universidad de Cantabria

Nuevo nodo Altamira

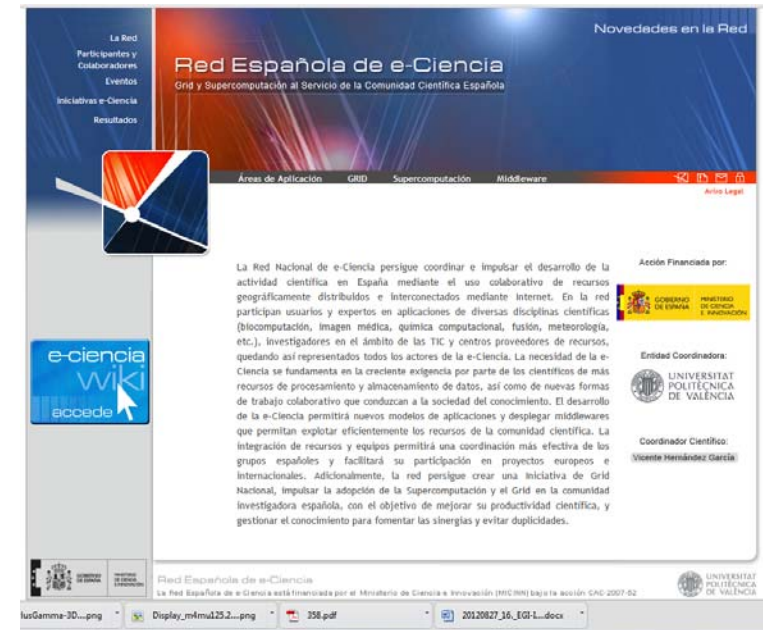
- « Acción Innocampus Campus Excelencia Universidad de Cantabria
 - Nuevo nodo RES [+ nuevo cluster Instituto de Hidráulica Ambiental]
- « Nuevo nodo es heterogeneo:
 - 158 nodos SMP de 16 cores (Intel E5-2670), 64GB RAM, IB FDR
 - 5 nodos 2xGPU (Nvidia M2090), 64 GB RAM, IB FDR
 - 11 nodos SMP de 16 cores (Power7 3GHz)
- « Total: >60 Tflops + >10 Tb RAM
- « Acceso a > 1 Pbyte de disco on line bajo GPFS (IFCA-CSIC)



List	Rank	System	Vendors	Cores	Rmax (Gflop/s)	Rpeak (Gflop/s)
06/2012	358	ALTAMIRA - iDataPlex DX360M4, Xeon E5-2670 8C 2.600GHz, Infiniband FDR	IBM	3840	74389	79872
11/2006	415	eServer BladeCenter JS20 (PowerPC970 2.2 GHz), Myrinet	IBM	512	2994.0	4505.6

e-Ciencia

- ❖ e-Science = enhanced Science
 - ❖ Gracias a una colaboración basada en el uso de la red
- ❖ Libro blanco e-Ciencia: 2005
- ❖ Red Española de e-Ciencia:
 - ❖ Infraestructura Grid
 - ❖ Infraestructura de Supercomputación
 - ❖ Middleware
 - ❖ APLICACIONES



J.Marco, Madrid 19 Sep. 2012

Cantabria Campus Internacional

www.cantabriacampusinternacional.com/Paginas/Cantabria-Campus-de-Excelencia-Internacional.aspx



UIMP
Universidad Internacional
Menéndez Pelayo



CANTABRIA
CAMPUS
INTERNACIONAL

Inicio de sesión

Castellano

English



Buscar en este sitio...

Buscar

Sobre CCI

Instituciones Agregadas

Acciones

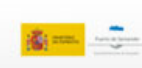
Documentación CCI

Convocatorias y Destacados

Docencia

I+D+i

Comunicación



CAMPUS DE EXCELENCIA
INTERNACIONAL:

UNA REGIÓN DE CONOCIMIENTO

1 2 3 4 5 6 7 8 9 10

Ámbitos de mejora

**Mejora Docente
y Adaptación al EEES**

**Mejora Científica y
Transferencia del
Conocimiento**

**Transformación del
Campus e Interacción
con el entorno**

Campus Nobel

**¡Disfruta
de todas las oportunidades!**

Noticias

- Máster en Educación Emocional, Social y de la Creatividad, últimos días de preinscripción
- Arqueólogos de los cinco continentes se dan cita esta semana en la Universidad de Cantabria
- Últimos días para matricularse en el nuevo Máster Internacional en Dirección de DataCenters
- La Universidad de Cantabria aspira a liderar en España el emprendimiento universitario
- La Universidad de Cantabria participa en un proyecto europeo para impulsar el emprendimiento desde las universidades
- 40 investigadores de 11 países analizan los avances logrados en geometría en el CIEM
- Las redes de comunicación de la UC y del Ayuntamiento de Santander se unen para mejorar su servicio
- Investigaciones de la UC sobre evacuación en caso de incendio en la agencia SINC

La fuerza de la agregación



Áreas de Excelencia

Agua y Energía

Biomedicina y Biotecnología

Banca, Finanzas

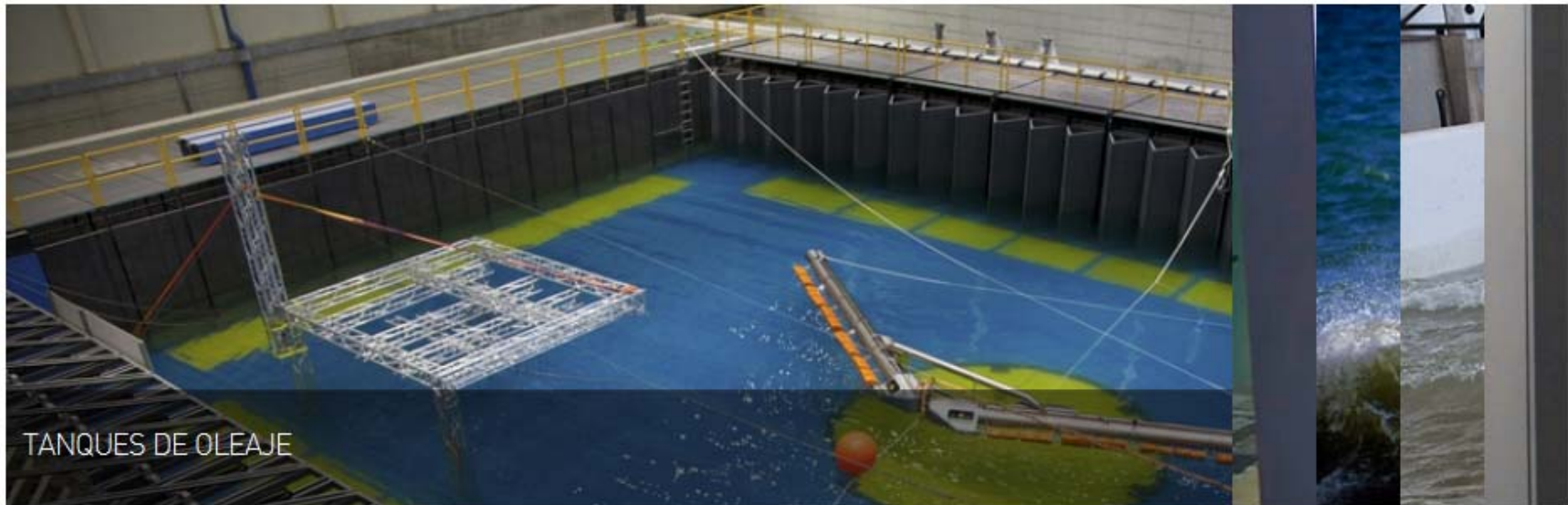
y Actividad Empresarial

Cantabria Campus Internacional

IH cantabria

FORMACIÓN | INVESTIGACIÓN | SERVICIOS | IHLAB

Usted está aquí: Inicio ▶ IHLab ▶ IHLab Hidro



TANQUES DE OLEAJE

« || »

IHLAB HIDRO

El Laboratorio de Ingeniería de Costas, Oceanografía e Hidráulica (IHLab Hidro) realiza ensayos en modelo físico donde se estudian fenómenos relacionados con la generación y propagación del oleaje, interacción ola-ola y ola-estructura, estabilidad y comportamiento de estructuras de protección costera, diques de abrigo y estructuras marinas, comportamiento de estructuras flotantes, funcionamiento de válvulas y máquina hidráulicas, así como el ensayo de dispositivos de generación de energía marina. El IH LAB HIDRO cuenta con 2 tanques de oleaje direccional con capacidad de generación de corrientes y viento, 3 canales de oleaje- corriente, 1 canal de corriente de pendiente variable, 1 tanque para el estudio de vertidos hipersalinos, así como un área de construcción de modelos fluviales.



Las capacidades de las instalaciones, la experiencia de los miembros del equipo y el empleo de software de modelado numérico de última generación e instrumentación avanzada, permiten realizar ensayos de investigación básica del comportamiento de

CONSOLIDACIÓN E INTERNACIONALIZACIÓN DEL CENTRO DE CÁLCULO CIENTÍFICO Y POTENCIACIÓN DE LINEAS DE INVESTIGACIÓN E INNOVACIÓN ASOCIADAS.

Objetivo de la actuación:

- mejorar y ampliar la capacidad del centro de cálculo científico existente mediante el *desarrollo en colaboración con la empresa de un centro de hosting de servicios avanzados de computación* (supercomputación, grid, *cloud computing* y soporte para *big data*), *con vocación de soporte a nivel internacional*, y que se instalaría en el PCTCAN.

Innovación en Innocampus

✿ Acercar HPC a las PYMES:

- ✿ Instalación de un nodo HPC en el Parque Científico Tecnológico de Cantabria, gestionado por una empresa (CIC-SL)
- ✿ Dar acceso para el desarrollo de aplicaciones piloto
- ✿ Posibilitar la adaptación de aplicaciones
- ✿ Uso del sistema ALTAMIRA en proyectos de innovación:
 - Ej: desarrollo de un nuevo fármaco
 - Ej: diseño de un nuevo material

✿ “Convencional”:

- ✿ **Paralelizar** programas/aplicaciones (open o propias)
- ✿ Simulaciones paramétricas

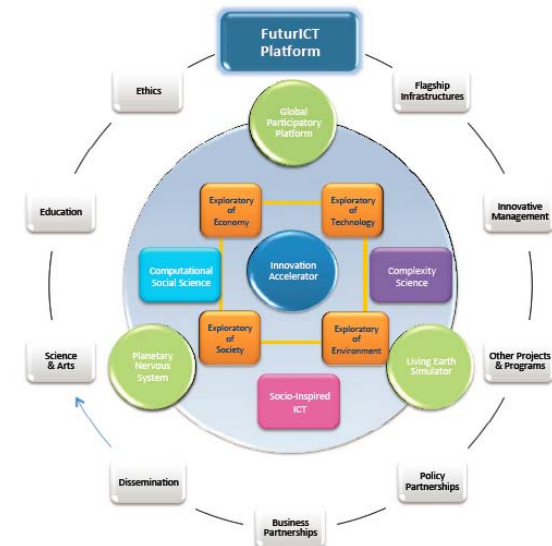
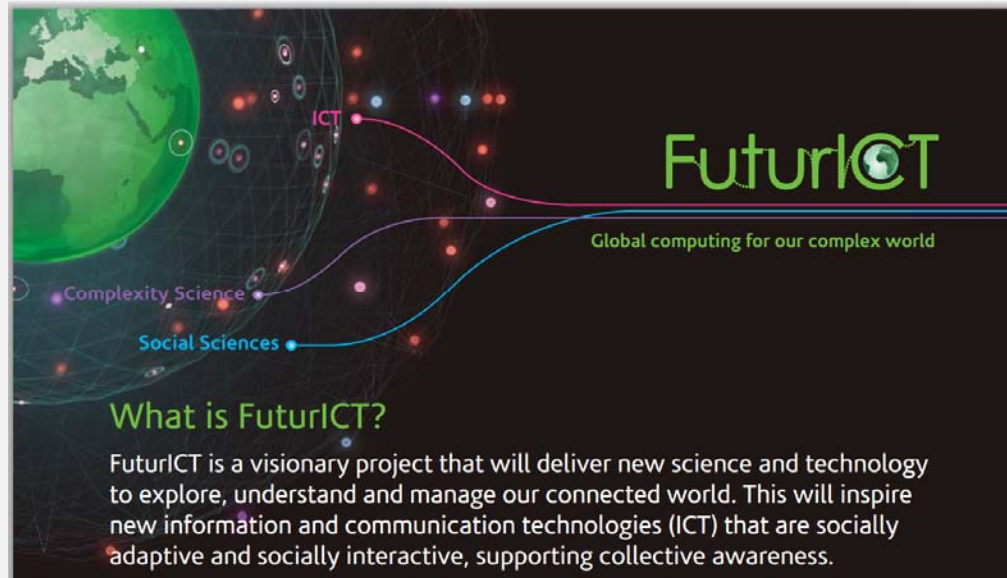
Innovación en ALTAMIRA

🌀 Uso “no convencional” de un supercomputador:

- ❑ Asignación priorizada e inmediata de recursos
 - Packs TEST
- ❑ Posibilidad de acceso interactivo (INTERACTIVE SUPERCOMPUTING)
 - Incluyendo visualización
- ❑ Virtualización
 - Posibilidad de usar diferentes SO
 - Oferta de imagenes “cluster” de diferente tamaño
- ❑ Combinación con almacenamiento HPC para
 - Procesado distribuido de grandes volúmenes de datos
 - Despliegue de bases de datos noSQL
- ❑ Oferta (propia o por terceros) de entornos integrados:
 - Ej: simulación basada en agentes
 - Ej: análisis de datos con redes neuronales
 - Ej: Open GIS + procesamiento de información (R-paralelo)

Oferta de colaboración en proyectos

FUTUR-ICT



CITIMON

-  Integrar y procesar la información de redes de sensores
-  Acceso directo y transparente a los ciudadanos