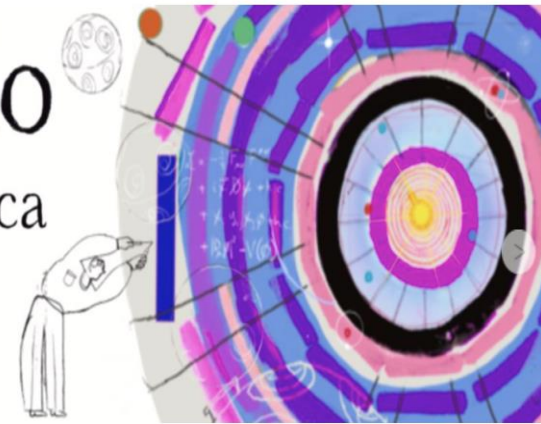


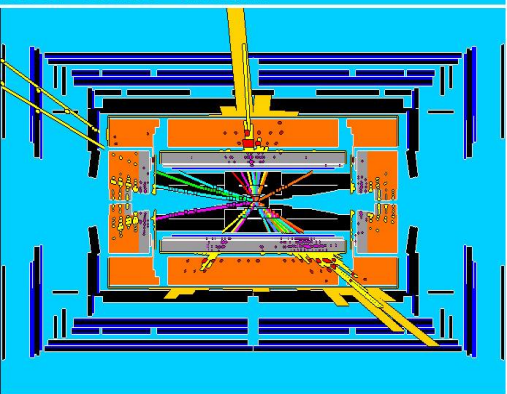
La Computación (¡ y el software !) en la era del LHC



75 Aniversario
del Instituto de Física
Corpuscular



ATLAS Atlantis Event susyevent



Dr. José F. SALT CAIROLS

Profesor de Investigación CSIC

Instituto de Física Corpuscular

(Jose.Salt@ific.uv.es)



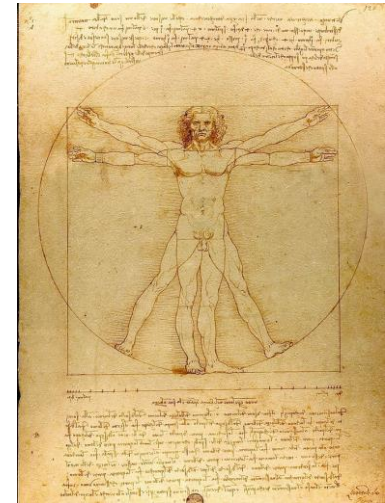
Contenido

- 1) **Introducción: Física y Computing**
=> Un poco de historia
- 2) **El Problema de la Computación y los Datos en el LHC**
- 3) **La Solución al problema: las Tecnologías GRID**
- 4) **El Modelo (Dinámico) de Computación de ATLAS**
- 5) **El Tier-2 Español de ATLAS**
- 6) **El GRID y el Higgs**
- 7) **Retos de Computación en el futuro.**
- 8) **Mensajes a retener/Conclusiones**

1- Introducción: Física y Computing

- **Físico Experimental (perfil amplio de un físico)**

- Hardware,
- constructor de detectores, R&D
- Calibración, Performance, Análisis de datos de los detectores
- Software
- Análisis de datos (Physics Analysis) y su interpretación
- Especialista en MC /Simulación
- Computing
-



- **La computación:** La tecnologías de la Información y las comunicaciones son pieza fundamental en muchas actividades. En la Ciencia su papel es esencial para la obtención de resultados



El origen (?)

- Física y Computación: confluyen en los años 30 y 40 del pasado siglo
- Un aspecto muy importante: la aparición de las Técnicas de Simulación (Monte Carlo) a partir de la concentración de científicos en Los Álamos cálculo de secciones eficaces, etc



Modelo Matemático de Implosión

John Von Neumann



Stanislaw Ulam

- Los físicos y matemáticos pioneros y visionarios:
 - Von Neuman, Ulam, Metropoli--- Fermi, Wigner, ...



Nicholas Metropolis



E. Fermi



R.P. Feynmann

- Feynmann y la Informática Cuántica:
 - Charla en 1981, 'Simulating Physics with Computers' -> Propone el uso de fenómenos cuánticos para realizar cálculos computacionales (dada la naturaleza de complejidad de los cálculos)- Ordenador Cuántico

Evolución Computing Distribuido

- **Agregación de recursos:**
 - Sistema SHIFT (“granjas” RISC Unix, L.Robertson, CERN)
 - Clusters de PCs (Sistemas Beowulf, “fábricas” de PC...)
- **Recursos distribuidos compartidos:**
 - Sistema Condor (M.Livny)
 - gestión de tiempo inactivo en sistemas Linux de la red local
 - Red Entropía, Programa SETI@home
 - Sistemas “Peer to Peer”
- **Aparición de la Web:**
 - 1989: primera propuesta, CERN, Tim Berners-Lee y Robert Cailliau
 - Primeros servidores web en laboratorios de Física europeos
 - 1991: un prototipo del sistema WWW suministrado para la comunidad de HEP



Tim Berners-Lee

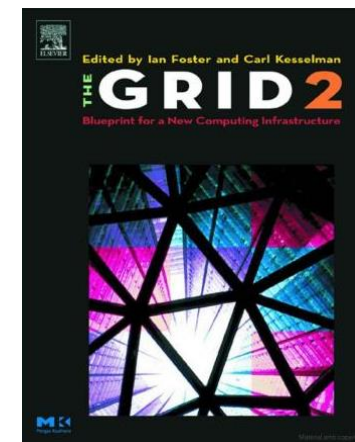
Robert Cailliau

- **1995: Supercomputing '95**

- Experiencia I-WAY: 17 centros USA conectados a 155Mbps
- Primeras iniciativas Grid:
 - NASA Information Power Grid
 - Iniciativa de la NSF con los centros NCSA y SDC
 - Advanced Strategic Computing Initiative (DoE)

- **La era de las los clusters/granjas/fabricas de PCs**

- Nodos con CPUs (duales)
- Disco local + acceso a servidores (NFS-NAS, AFS)
- Limitación:
 - Interconexión de red: Gigabit casi popular, pero Latencia baja requiere Myrinet o similar, solución mas costosa
 - Perfectos para HTC (High Throughput Computing)
 - Las aplicaciones HPC (High Performance Computing) necesitan adaptarse:
 - La memoria no está compartida
 - Las herramientas: PVM, MPI
 - Instalación y control: funciona bien para cientos de ordenadores “homogeneos”



La 'Biblia' del GRID



'Granjas' de ordenadores



- ... Y además de forma distribuida mediante la conexión de clusters
 - Gracias a la interconexión de la Red
 - Llegar a obtener la sensación de que trabajamos con un Superordenador formado por varios clusters de ordenadores (Metacomputación)



2- El Problema de la Computación y los Datos en el LHC

Almacenamiento-

Ratio de registro de datos 0.1 - 1 GBytes/sec

Acumulando a 10-15 PetaBytes/año

15 PetaBytes de disco

Procesamiento

100,000 de los PC's más rápidos actuales



- La infraestructura computacional del CERN **NO** era suficiente para procesar todos los datos

- Se necesitaba un sistema 'robusto': que haya redundancia, sin 'únicos puntos de fallo'



Los 4 experimentos del LHC:
Europa: 267 institutos, 4603 usuarios
Resto mundo: 208 institutos, 1632 usuarios

Tecnologías Computación GRID



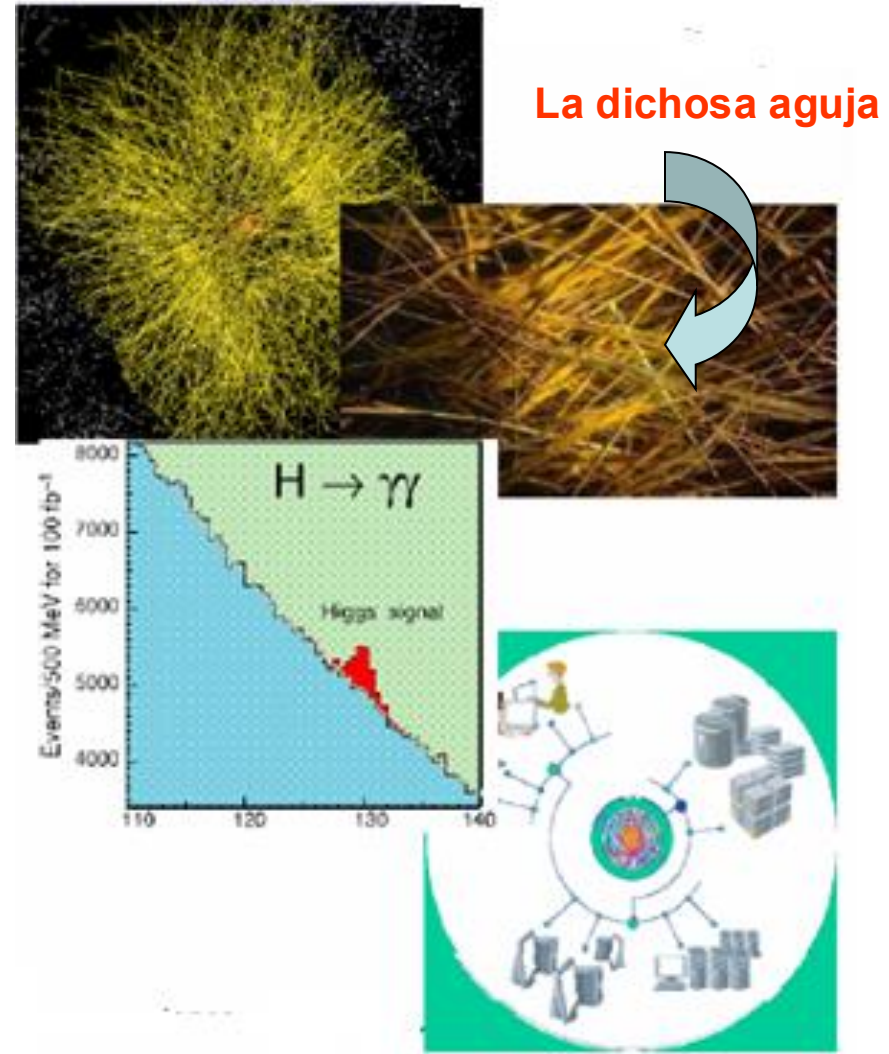
3.- La Solución al Problema: las Tecnologías GRID

¿Qué es la Computación GRID?

- **Definición:** la computación GRID es una evolución de la computación distribuida: su base está en tecnologías que permiten a las organizaciones compartir recursos informáticos para hacer frente a diferentes tipos de necesidades . En general, dichos recursos están dispersos geográficamente y conectados por Internet pero dicha red proporciona la impresión de estar trabajando con un único sistema informático virtual
- **intenta resolver problemas actuales de la Sociedad de la Información:**
 - Acceso rápido a bases de datos/almacenamiento...
 - Proporcionar su procesamiento y análisis utilizando potencia de cálculo distribuida y potentes facilidades de visualización...
 - Mediante la utilización de la red (Internet)

Las Tecnologías de la Información nos ayudan a ‘buscar una aguja en un pajar’

- La Señal es extremadamente baja (**1 sobre 10 billones de colisiones**)
- Volumen de datos:
 - (Frecuencia alta) *(gran número de canales) * (4 experimentos) = **15 PB de nuevos datos/año**
- Potencia de cálculo:
 - (Complejidad de sucesos) * (número de sucesos) *(miles de usuarios)=
100 K de los CPUs más rápidos actuales
45 PB de almacenamiento de disco (Simulación +Análisis)
- Análisis y financiación mundial:
 - Financiación local de recursos de computación en los principales países y regiones



Simil del GRID Computacional con la Red Eléctrica

Red Eléctrica

- Plantas de producción de electricidad



- Distribución jerárquica del flujo eléctrico



Subestación Eléctrica

- Tendidos eléctricos



- Usuario final: acceso a las prestaciones de la electricidad



• GRID Computacional

- Grandes centros de almacenamiento de datos/potencia de cálculo
(Supercomputador/GranCentro de Cálculo)

- Distribución jerárquica
(Centros de cálculo medianos)

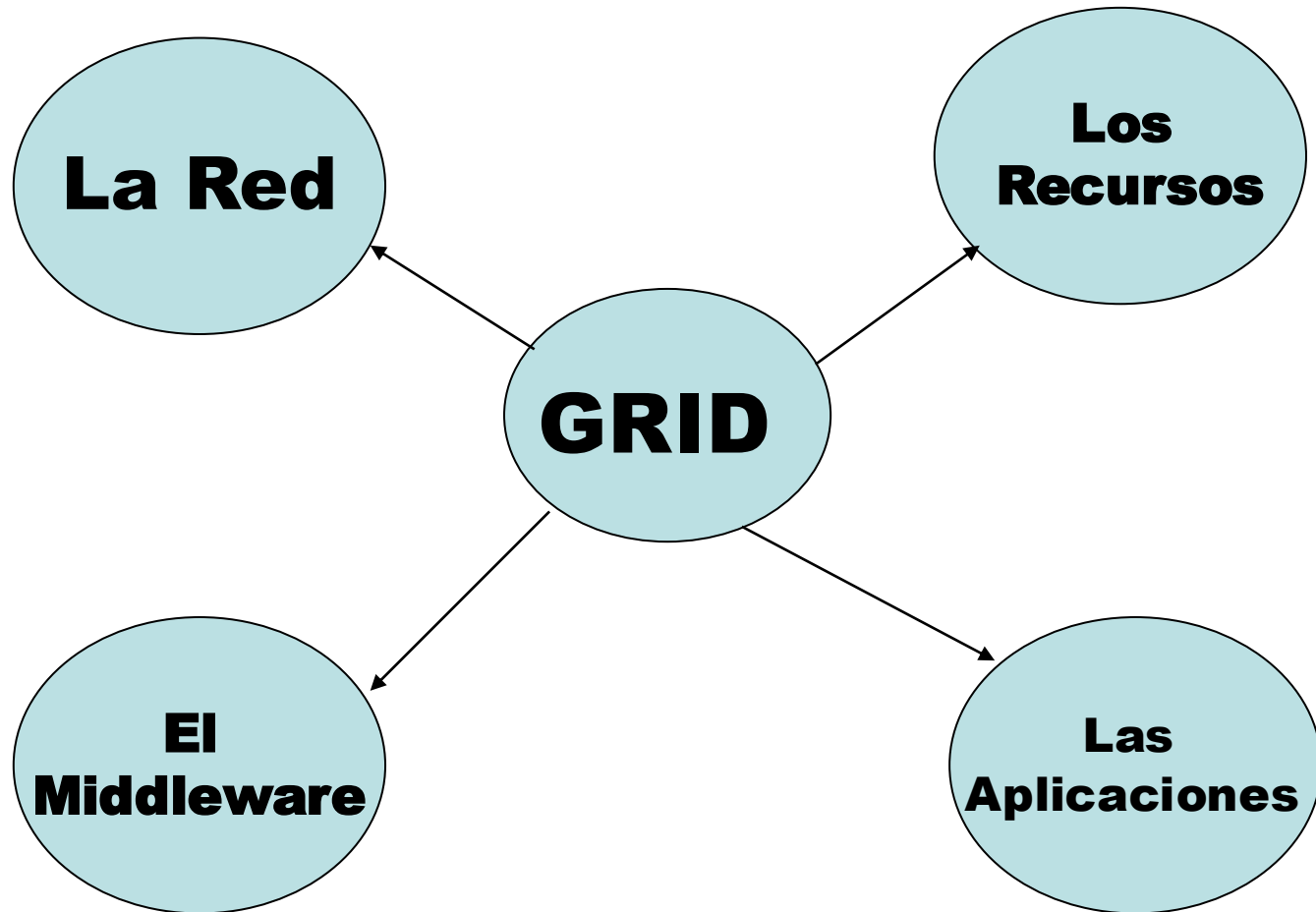
- Red Internet

← Ramo de fibras ópticas

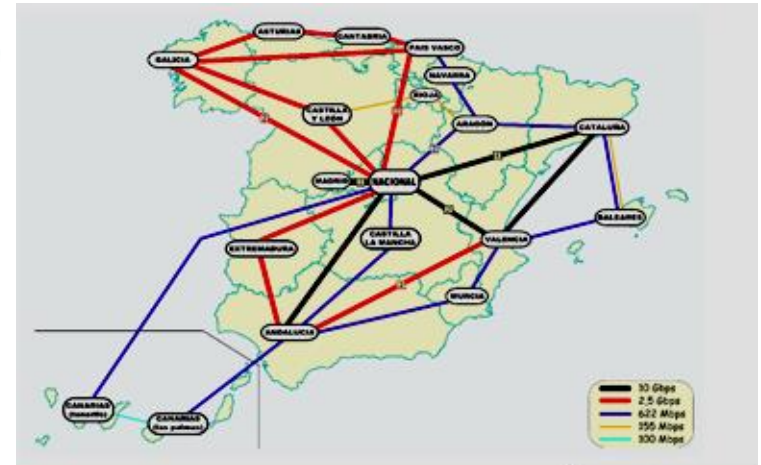


- Usuario Final: acceso a las prestaciones informáticas

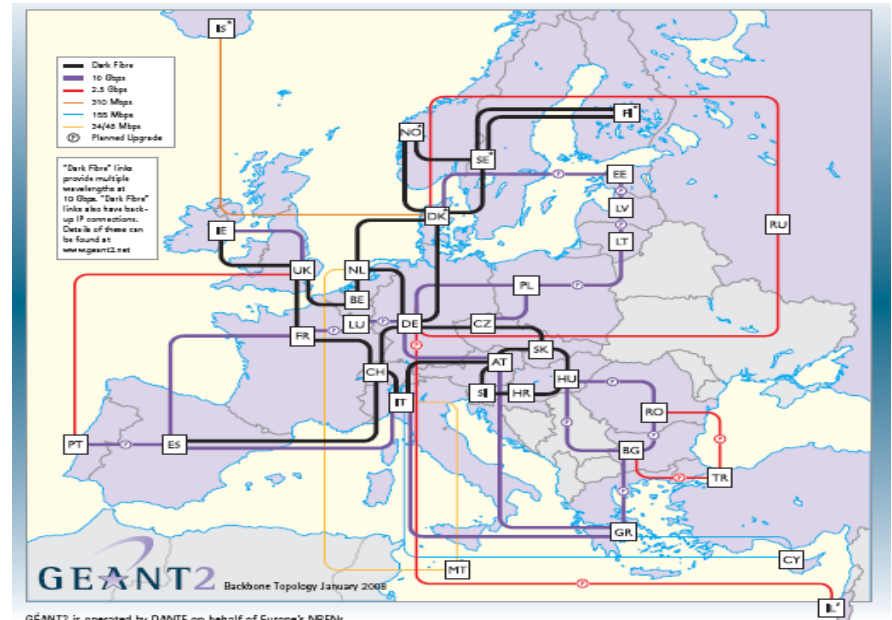




La Red: las autopistas de la Información



Mapa de la Infraestructura de la Red en España (RedIris)



La RED Académica y de Investigación Pan-europea

- El GRID no podría desarrollarse sin una Red apropiada.
- Responsable de asegurar los recursos que forman el GRID
- Comunicación
 - Protocolos de Internet: IP, DNS, routing, etc.
- Gran esfuerzo a nivel Europeo (DANTE y NRENs) y Español (RedIris)
- Constituyen **las Autopistas de la Información**

Los recursos: Las Infraestructuras

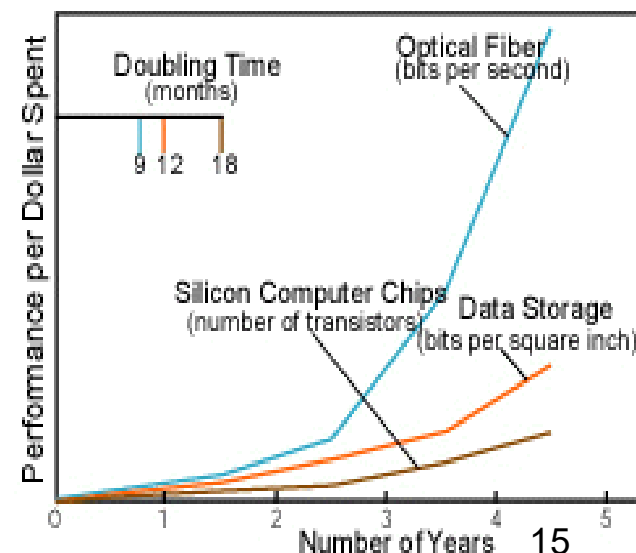
Evolución tecnológica

- Hardware
 - Procesadores...cumplen la ley de Moore: se **duplica cada 16-18 meses !**
 - Nodos de computación:
 - PCs y portátiles, Estaciones, Servidores, Clusters, Blades
 - Almacenamiento: **se duplica cada 12 meses !!**
- Mejora de la Red:
 - la capacidad de la red **se duplica cada 9 meses !!**
 - 10 GB ethernet
- Coste tiene en cuenta muchos factores:
 - Espacio, alimentación eléctrica, mantenimiento

- Tendencia al 'Commodity Computing' : adquisición de sistemas de diferentes compañías (vendedores), incorporan componentes basados en standards 'abiertos'



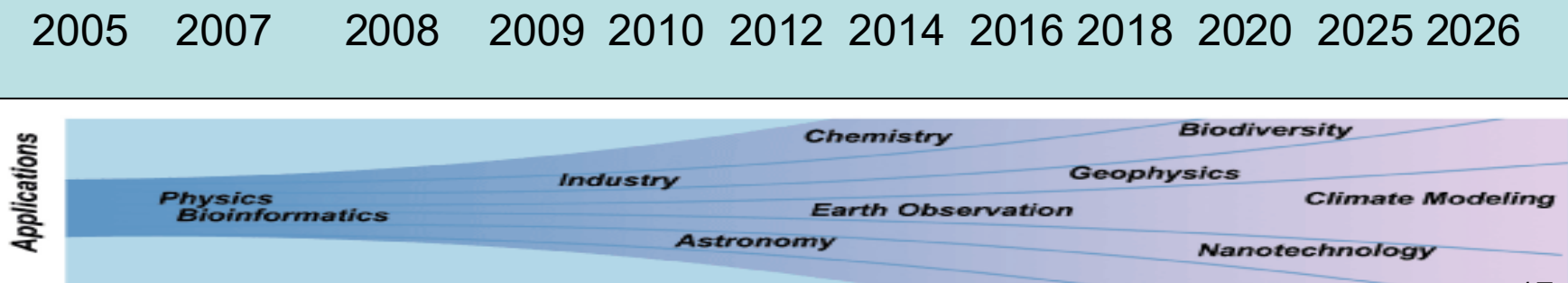
Gordon Moore



¿Qué es el Middleware?

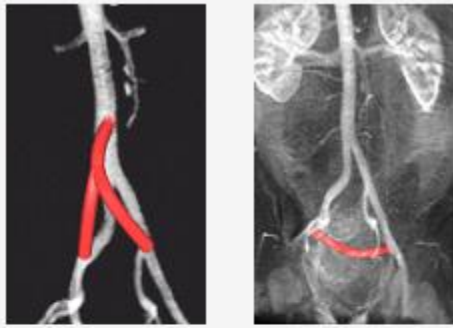
- **es un software de conectividad que ofrece un conjunto de servicios que hacen posible el funcionamiento de aplicaciones distribuidas sobre plataformas heterogeneas.**
- **nos abstrae de la complejidad y heterogeneidad de las redes de comunicaciones subyacentes, de los sistemas operativos y lenguajes de programación,**
- **Finalidad: ‘virtualizar’ los recursos de computación’**

- Algunas disciplinas científicas se han organizado durante las décadas pasadas en grandes colaboraciones científicas que agrupan una gran cantidad de científicos al ser la única manera de alcanzar logros de alto nivel que no sería viable con grupos de trabajo reducidos
- resolución de retos tecnológicos importantes (ejemplo: experimentos del LEP, Proyecto Genoma, etc)
- En aspectos computacionales y tecnologías de la información, se ha evolucionado y ha desembocado en el paradigma GRID.

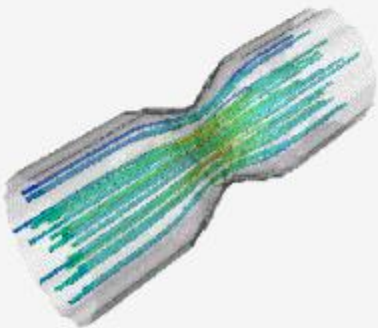


El GRID aplicado a enfermedades vasculares

- Dos procedimientos: stent y bypass
- El cirujano se ayuda con imágenes 3D obtenidas con scanners CT y MRI
- Las simulaciones de una posible intervención ayuda a la toma de decisión por parte del médico

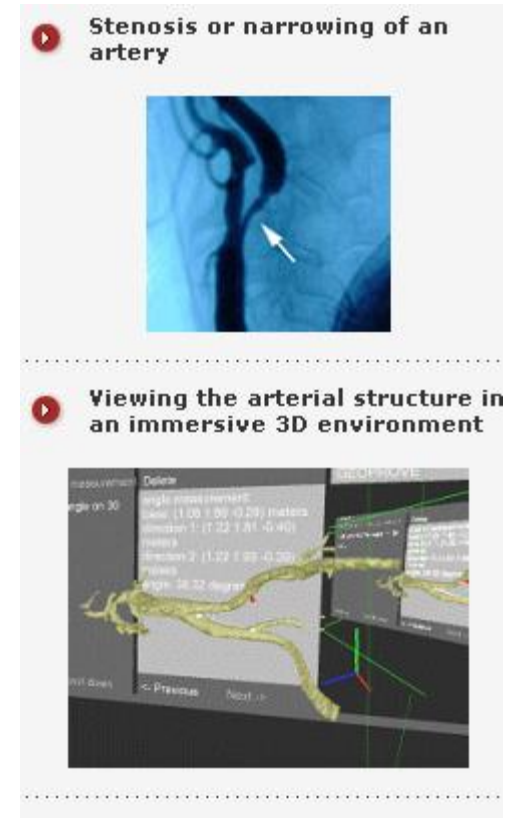


▶ Simulated flows



El GRID permite lanzar los cálculos pertinentes. Se utilizan los recursos de los centros que participan en el proyecto (CROSSGRID)

Investigación coordinada por la Universidad de Amsterdam y con la colaboración del CSIC y de todos los centros de CROSSGRID



4.- El Modelo Dinámico de Computación de ATLAS



- El proyecto de Computación GRID para el LHC (WLCG) comenzó en 2002:
 - Fase I (2002-2005): test y desarrollo, construcción y prototipo de servicios
 - Fase II (2006-2010): despliegue inicial de los servicios GRID
 - Fase III (2011-2024): consolidación de servicios, ampliación a otros paradigmas de computación
- Propósito: Suministrar los recursos de computación necesarios para procesar y analizar los datos recolectados por los experimentos del LHC

Enormes volúmenes de datos y capacidad de computación

- 15 PB/año (Raw) -> 50 PB/año (overall)
- Vida Util: 10-15 años -> escala del Exabyte

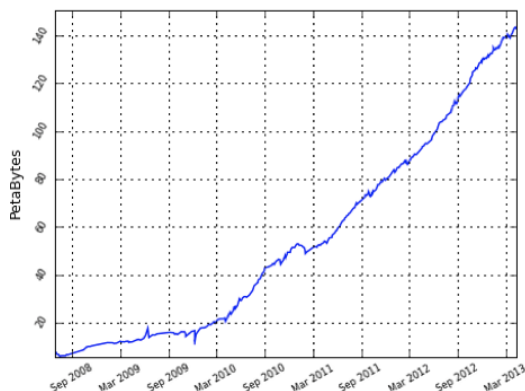
Reto de la Escalabilidad

- Resulta evidente que se necesita una infraestructura distribuida

Logros:

- CMS: ha transferido 100-200 TB/día por el GRID en los últimos dos años
- ATLAS: se han acumulado 140 PB durante el procesamiento de datos del Run I

Total GRID space usage according to DQ2



Organización de centros de recursos en ATLAS: Tiers



desktops
portables

Tier-3

small
centres

RAL

IN2P3

FNAL

CNAF

FZK

Tier-1

PIC

BNL

ICEPP

Tier-2

UAM

IFIC

IFAE

Legnaro

Krakow

NIKHEF

USC

MSU

CIEMAT

IFCA

UB

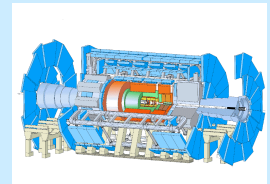
Cambridge

Budapest

Prague

Taipei

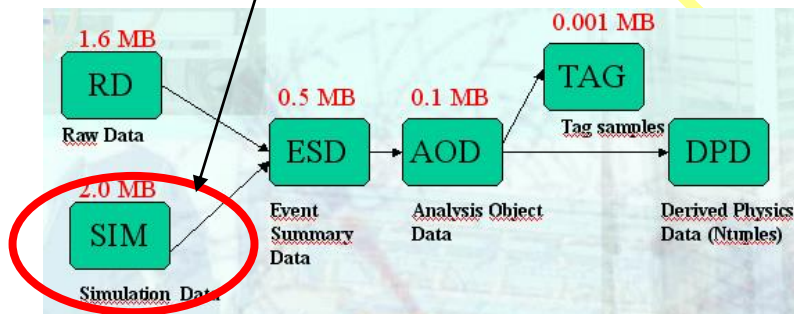
TRIUMF



Tier-0

asegurar un crecimiento sostenible
de la infraestructura Tier-2 entre
estos centros y su operación de
forma estable

- Los datos sufren transformaciones encaminadas a la reducción en tamaño y la extracción de la información relevante
- Con el fin de prepararnos para los análisis con Datos Reales y testear el sistema, trabajamos con Datos Simulados (Monte Carlo)

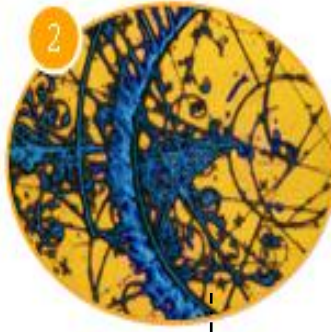


Proceso de transformación de los datos



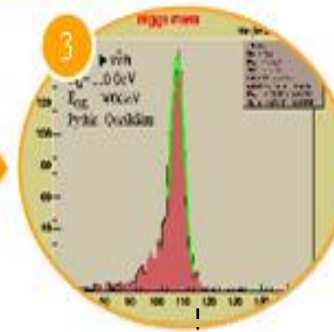
Datos de ATLAS 'brutos'

Proceden directamente del experimento (después del "trigger") ubicados en el CERN, serán procesados y pasados a los Tier-1 (1.6MB/colisión)



Datos procesados

Proporcionan información clara de trazas y valores de energía registrada por los detectores del experimento ATLAS. (500KB/colisión)



Datos listos para analizar

Provenientes de los datos procesados, servirán para descubrir nuevas partículas por los científicos de ATLAS. (100KB/colisión).

- Los datos sufren transformaciones encaminadas a la reducción en tamaño y la extracción de la información relevante

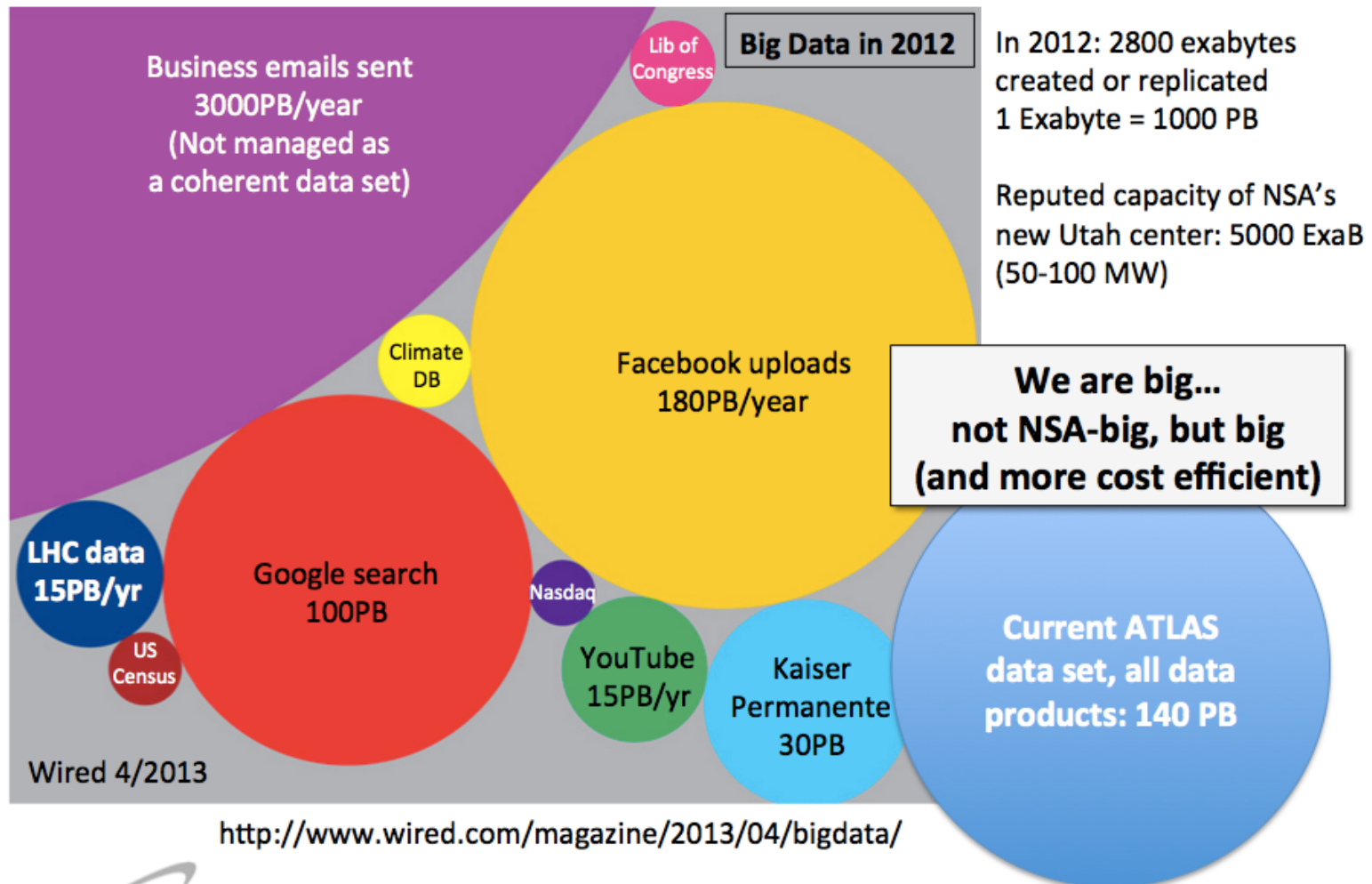




El GRID: “Un despliegue colaborativo donde no se pone nunca el Sol”
(‘round the clock’ : 24 h / 24 h)

Data Management

Where is HEP in Big Data Terms?



5.- El Tier-2 Español para ATLAS

Proyecto Coordinado del Plan Nacional de FAE de 3 centros:



IFAE (25%)

UAM (15%)

IFIC (60%)

4-5% del total

Tier-2

Equipamiento:

IFIC

CPU = 54 kHEPSpec06

Disk = 5.1 PB

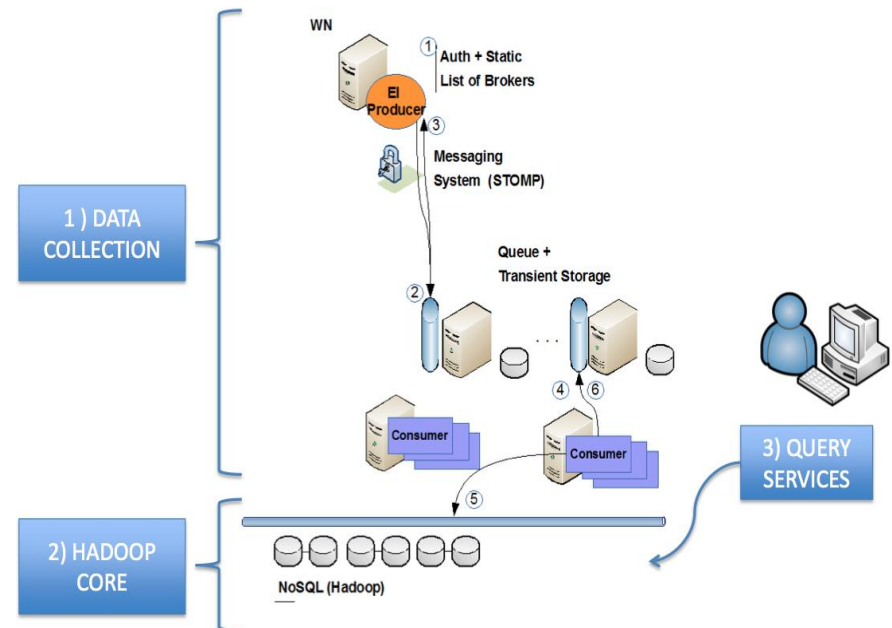
Recursos Humanos:

4.5 FTE

Julio 2026

Desarrollo del Event Index de ATLAS (I+D)

- Desarrollo y despliegue de un catálogo completo de todos los sucesos de física (Real Data y MC) para todas las etapas de procesamiento
- Uso de nuevas tecnologías de Big Data y NoSQL (Hadoop)
- Alta participación del grupo IFIC:
 - Responsable of Data Collection Task.
 - Responsabilidad en la tarea de Data Collection
 - Ayuda en las verificaciones de consistencia en la producción global de datos



Evolución del Crecimiento de los recursos de TIER-2 para el experimento ATLAS



Primera Granja de PC's



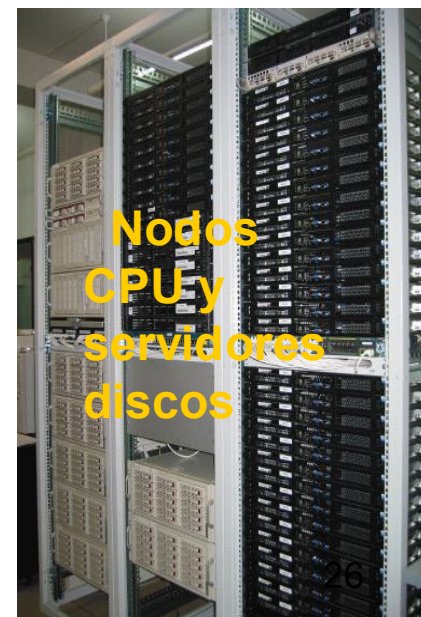
**Actualización
Infraestructura
Computación del IFIC:
GLUON**



Fuerte Crecimiento en recursos



**Robot de
Almacenamiento
en Cintas**



**Nodos
CPU y
servidores
discos**

**ACTUALIZACIÓN
DEL CENTRO DE
CALCULO DEL IFIC
(Propósito General/
Grupos Investigacion
IFIC)**

GLUON

**INFRAESTRUCTURA
PARA INTELIGENCIA
ARTIFICIAL**

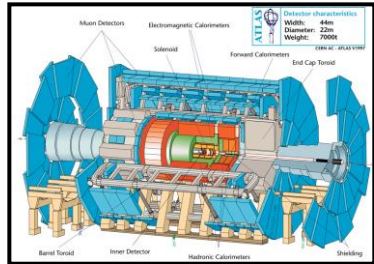
ARTEMISA

6- El GRID y el Higgs

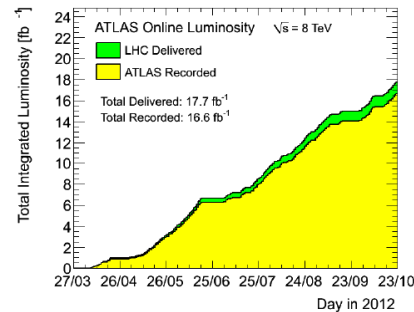
- Los resultados de Física son comparaciones de observables con sus predicciones teóricas/modelo
- Hay un proceso complejo entre la toma de datos y los resultados de física: Trigger/adquisición de datos, reconstrucción , análisis y simulación, grupos de análisis, publicación de resultados.
- Se necesitan algoritmos sofisticados para reconstrucción y análisis
- Uno de los factores más esenciales es disponer de una buena simulación detallada
- Un solo suceso tiene un determinado tamaño y 'simple', pero la escala de computación es enorme
- El software y el computing deben funcionar muy bien si se quiere lograr los resultados de primer nivel en el LHC

La computación GRID posibilita la obtención rápida de resultados de física : Un Caso Exitoso

1) Desde el experimento a 100 m bajo tierra...



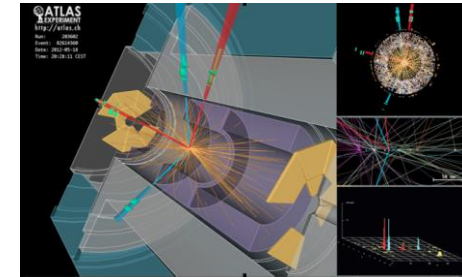
2) ... una ingente cantidad de datos es almacenada, procesada y distribuida...



3) ... por todo el mundo mediante la red y usando el GRID; en aproximadamente 2 horas los sucesos están disponibles en los centros de GRID



4) ... y el análisis final de los sucesos seleccionados (como el suceso Higgs representado en la figura) se realiza por los grupos de física ...



5) ... y (a veces) ocurre un final feliz de la historia (2012)

Higgs boson-like particle discovery claimed at LHC

COMMENTS (1665)

By Paul Rincon

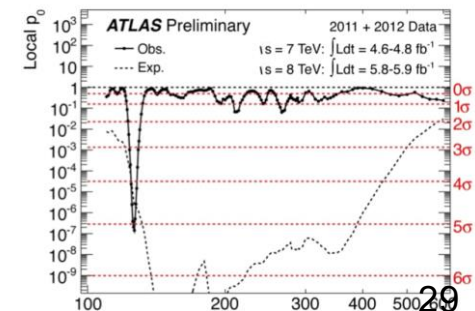
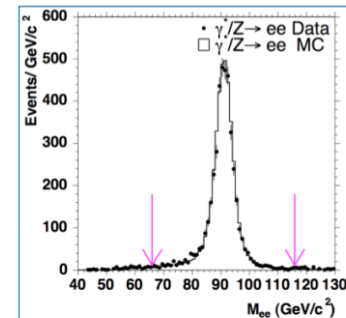
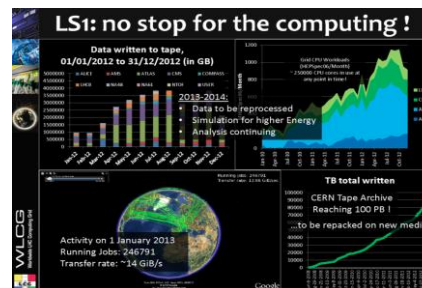
Science editor, BBC News website, Geneva



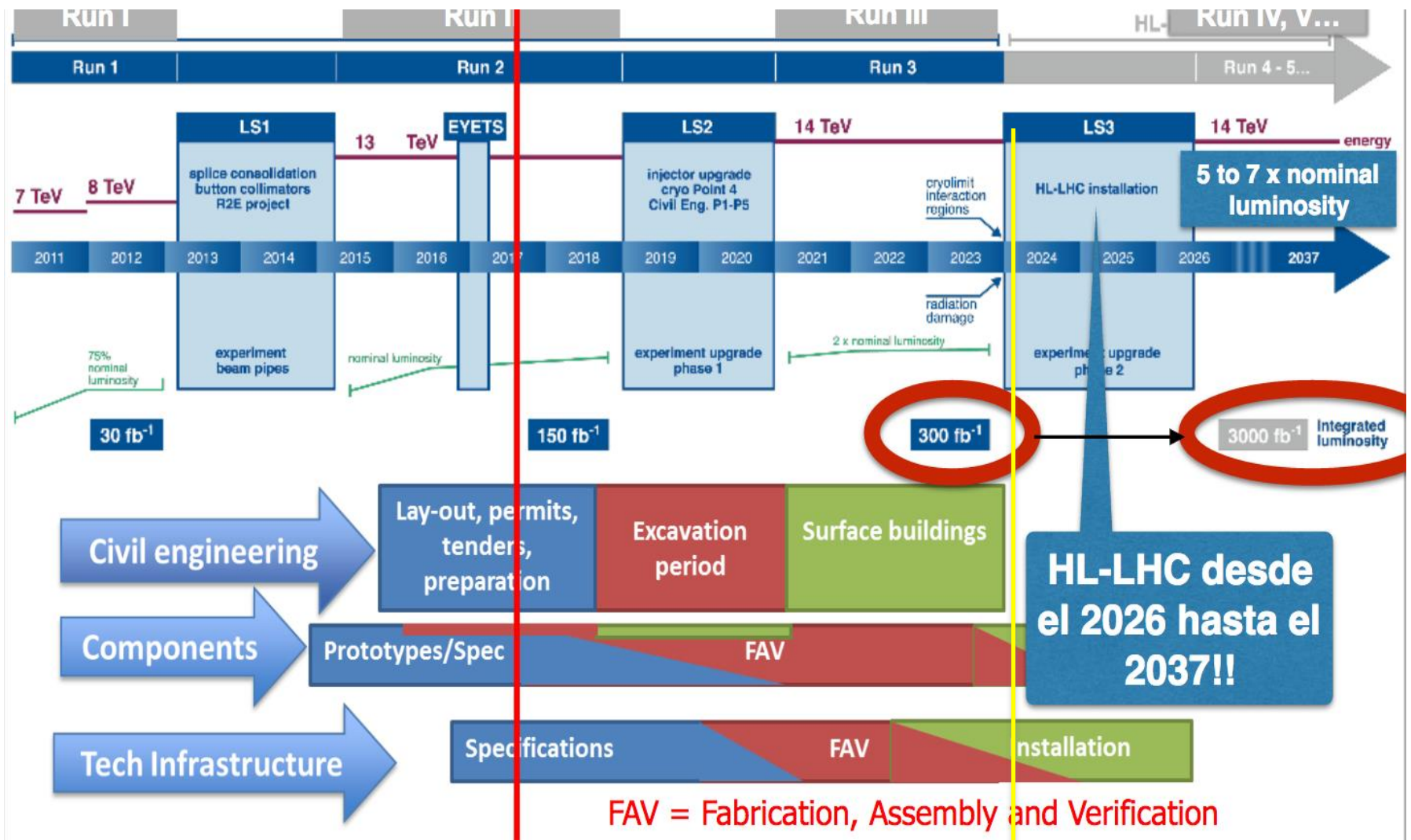
The moment when Cern director Rolf Heuer confirmed the Higgs results

Cern scientists reporting from the Large Hadron Collider (LHC) have claimed the discovery of a new particle consistent with the Higgs boson.

6) ...en **2026**, ha acabado el Run 3 (con una energía de 13.6 TeV) y hemos empezado el LS3 y una cantidad ingente de datos recolectados pero NO hay STOP para el computing



7- Retos de Computación en el futuro



Estamos aquí :
Julio 2026

El escenario actual: un 'ecosistema' de Computación

Estamos en un nuevo escenario debido a:

- las muy buenas prestaciones de **la red** en términos de latencia y ancho de banda
- la **mejora de la ratio calidad /coste** en los equipos de ordenadores
- Los **supercomputers (HPC)** han seguido una tendencia de crecimiento continuo durante las dos pasadas décadas
- **Cloud Computing** ha parecido hace unos 8-9 años como otra posible solución para resolver problemas actuales de computación
- **Sostenibilidad de la Infraestructura:** mayor rendimiento con menos energía, huella de Carbono, green computing

Computer cluster



Cloud Computing

GRID Computing



Supercomputers

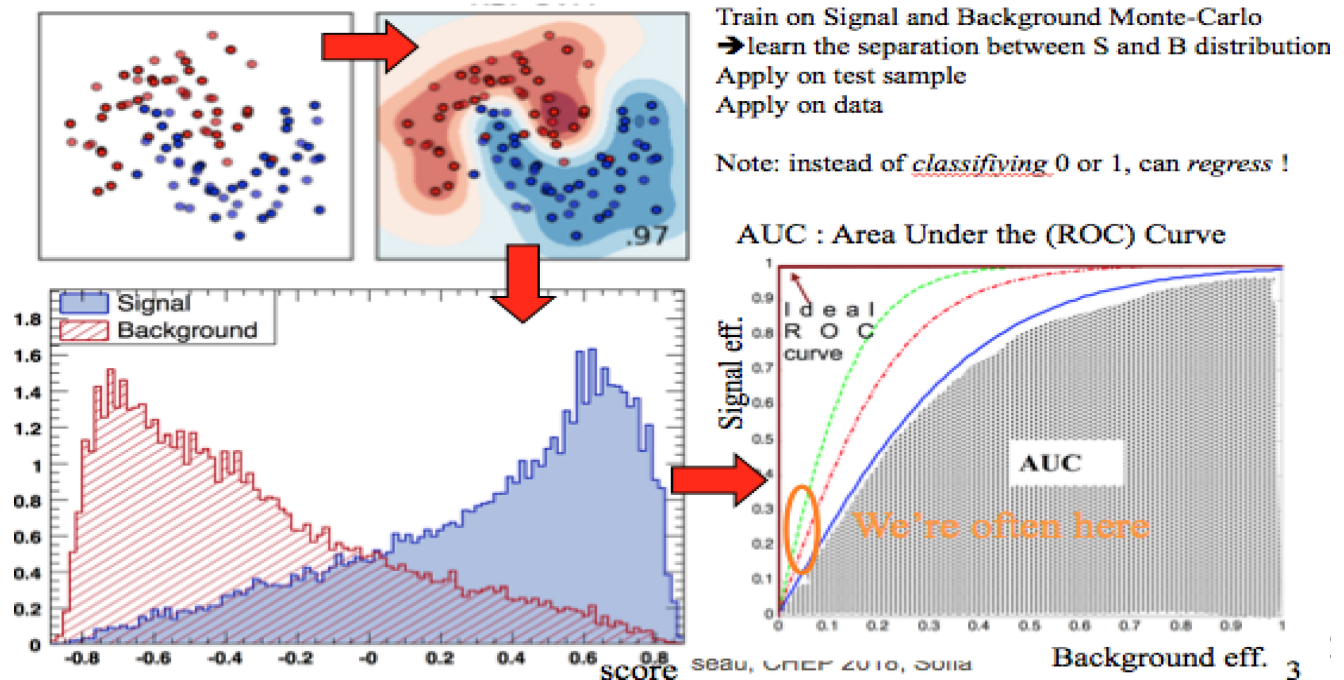
Machine Learning

- Utilización del Machine Learning en diferentes aspectos:

- ML en simulación
- ML en reconstrucción
- ML en análisis



Clasificación Señal- Background



Premio Nobel 2024: la Revolución del Machine Learning



POLÉMICA: ¿ La ML , puede considerarse Física?

Fijémonos en la explicación de abajo:

El Premio Nobel de Física 2024 ha sido otorgado a dos figuras clave en el desarrollo de las redes neuronales artificiales: **John Hopfield y Geoffrey Hinton**. Sus trabajos han revolucionado el campo de la inteligencia artificial, proporcionando las bases sobre las que se construyen muchas de las tecnologías avanzadas que utilizamos hoy en día. En la década de 1980, **Hopfield introdujo un modelo que conectaba la neurociencia con la física teórica**, desarrollando una red que podía almacenar y recuperar patrones de información de manera similar a como lo hace el cerebro humano al recordar. Este enfoque, conocido como **memoria asociativa**, permitió por primera vez simular aspectos de la memoria utilizando un modelo basado en el modelo de Ising del espín atómico. **Hinton, por su parte, utilizando principios de la física estadística, expandió este concepto a redes más profundas y complejas**, lo que dio lugar a que **los sistemas computacionales aprendieran de datos** en lugar de seguir instrucciones predeterminadas. Su enfoque hizo posible que las máquinas no solo clasificaran imágenes, sino también que generaran nuevas a partir de las que ya conocían, lo que abrió un mundo de aplicaciones, desde el reconocimiento de patrones hasta la predicción de estructuras biológicas.

Lo notable de los avances generados a partir del trabajo de Hopfield y Hinton no es solo su aplicabilidad tecnológica, sino la influencia transversal que han tenido en varias disciplinas, como la física de partículas, la astrofísica, la ciencia de materiales y la neurociencia. Estos descubrimientos han sido fundamentales para el crecimiento exponencial de la inteligencia artificial en la última década, demostrando cómo el aprendizaje automático puede abordar problemas que antes parecían intratables. En este punto, Hinton al recibir la noticia del premio, destacó que **"los avances en este campo tendrán un impacto similar al de la revolución industrial, aunque en lugar de reemplazar el trabajo físico humano, superarán nuestras capacidades intelectuales"**. Sus palabras reflejan la magnitud del cambio que estos descubrimientos están produciendo, tanto en la ciencia como en la sociedad.

En la Facultad de Ciencias Exactas y en la UNCPBA existe la sinergia entre diversas disciplinas como la física, la informática y la ingeniería en diferentes proyectos de investigación, que resuena con los principios detrás de estos avances, con aplicaciones a la Neuro-Inteligencia Artificial (NeuroAI), Educación, las Ciencias Veterinarias, Ambientales y Sociales. En ello los avances en inteligencia artificial pueden integrarse en investigaciones locales, enfatizando la importancia de combinar ciencia básica y aplicada para abordar problemas complejos de manera innovadora. Usualmente, **tales avances implican estudiar directamente este tipo de sistemas desde una perspectiva multidisciplinaria** o aplicar las herramientas que se han desarrollado a partir de ellos en áreas de la mayor diversidad.

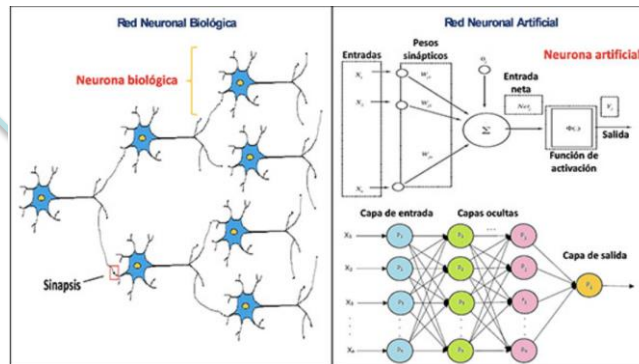
Así, mientras celebramos este merecido reconocimiento a Hopfield y Hinton, también **vemos cómo sus descubrimientos están dejando huella en nuestro propio entorno académico**, impulsando la investigación interdisciplinaria y expandiendo los horizontes de lo que es posible en la ciencia.

Departamento de Ciencias Físicas y Ambientales y Departamento de Computación y Sistemas

Machine Learning/Deep learning

NEUROCIENCIAS

- modelo biológico
- De redes neuronales
- el cerebro como sistema Físico



Física y Matemáticas

HEP

- Resolución de problemas Complejos
- Algoritmos
- Estadística
- Análisis Matemático

- disposición de tecnologías Informáticas
- aspectos de ingeniería para Organización de sistemas

Ingeniería/Informática

Salidas Profesionales:

A destacar dos (pero hay más!):

•En Investigación en Física Experimental de Partículas, Nuclear y Astropartículas (estáis viendo qué hacemos):

- trabajando en líneas de investigación que ya estáis conociendo en el IFIC Summer 2026

•Data Scientist:

- Profesión con muy buenas perspectivas y con una demanda cada vez mayor por parte de empresas y proyectos**
- Big Data: Obtención de conocimiento a partir de los datos**
 - El experimento ATLAS y otros experimentos (AGATA, DUNE, por ejemplo) son claros ejemplos de Big Data**
- Data Mining, Machine learning, Inteligencia Artificial, etc**

8.- Mensajes a retener/Conclusiones

- **La Física subatómica (HEP & Nuclear & Astroparticle) y Computación están estrechamente relacionadas**
- **La Computación en la era del LHC se ha desarrollado para resolver problemas originados por las características de los experimentos. La solución inicial ha sido el GRID**
- **El Computing es como un detector más : colaboración de sites, upgrades, organización de servicios y de 'shifts', R&D , etc**
- **El I + D + i asociado a estas tecnologías GRID está dando lugar a una 'segunda revolución' del mismo nivel que el que supuso la web hace 25 años**
- **Esto está generando un Modelo Dinámico de Computación**
- **Aplicaciones en otras disciplinas científicas. Ya se están obteniendo retornos en Biomedicina, Sector Energético, Química, Biotecnología, etc**
- **El GRID y otras contribuciones de las Tecnologías de la Información y de la comunicación han dado lugar a la e-Ciencia**
- **En Investigación es necesario ' Data Scientists' y son también muy apreciados en las empresas en el sector privado y público.**

- GRACIAS por vuestra atención
- PREGUNTAS / COMENTARIOS