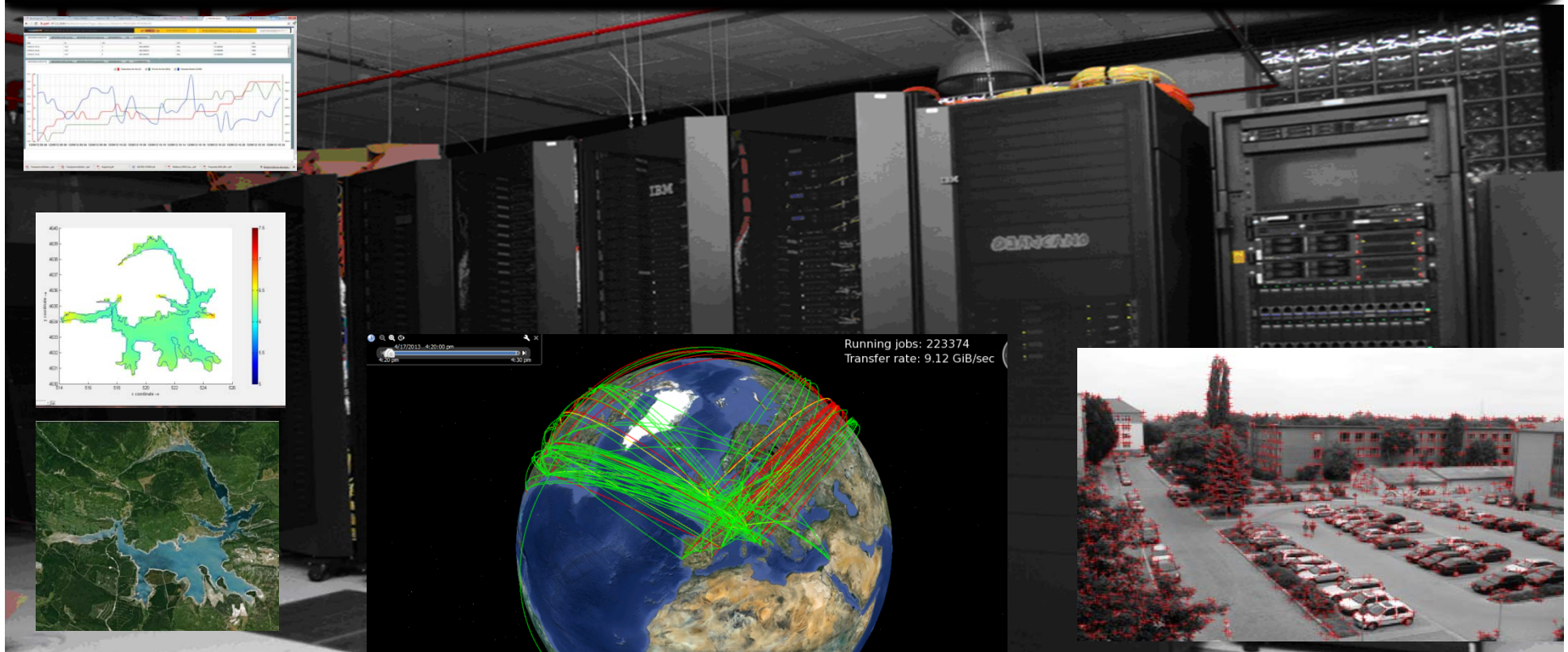


BIG DATA en Defensa y Seguridad desde el punto de vista tecnológico



DESEi+d 2013

Jesús Marco de Lucas, Instituto de Física de Cantabria, CSIC

Big Data en los entornos de Defensa y Seguridad

Grupo de trabajo:

COORDINADOR: Tcol. D. Luis Fernando Pereda Laredo

División de Sistemas de Información y telecomunicaciones, Estado Mayor Conjunto de la Defensa

Tcol. D. José Antonio Carrillo Ruiz

Jefe del Área de Información y Ayuda a la Decisión (AIAD) Vicesecretaría General Técnica
Profesor de la asignatura *Estadística* (Grado en Ingeniería en Tecnologías de la Información)-UNED,
Profesor de *Investigación Militar Operativa* en la Escuela Logística del Ejército de Tierra.

D. Juan Carlos Dueñas López

Catedrático de Ingeniería telemática ETSI Telecomunicación
Director del Centro de “Open Middleware” Universidad Politécnica de Madrid

D. Fernando Cases Vega

Gerencia de Planeamiento Tecnológico, Área de Prospectiva. ISDEFE

D. José Cristino Fernández,

Ingeniero Jefe del Área de Desarrollo. Future Space S.A.

D. Guillermo González Muñoz de Morales

Unidad de Observatorios Tecnológicos . Sistema de Observación y Prospectiva Tecnológica
Subdirección General de Tecnología e innovación Dirección General de Armamento y Material

Dr. D. Jesús E. Marco De Lucas

Instituto de Física de Cantabria (IFCA)

BIG DATA desde el punto de vista tecnológico

✚ ¿ Tenemos una visión propia de que es *Big Data* ?

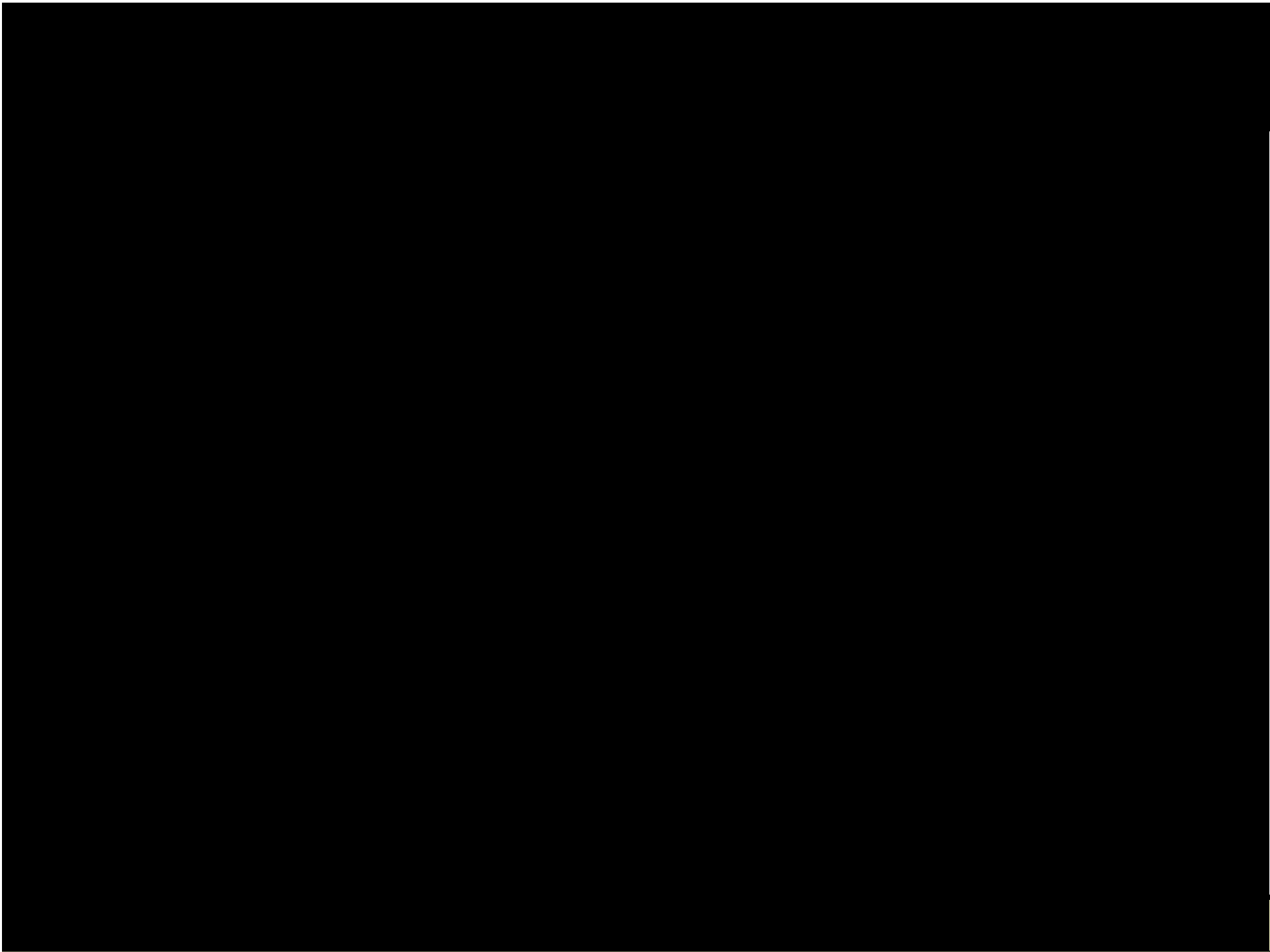
✚ Capacidades técnicas

- ✚ Adquisición
- ✚ Infraestructuras
- ✚ Presentación
- ✚ Tratamiento Automático

✚ Casos de Uso

- ✚ **Aprendizaje automático**
- ✚ **Procesamiento de Imágenes**
- ✚ **¿Un caso piloto?**

Misiones de Emergencia: Gestión de Incendios



However, as time goes by...

I have suffered a "brain phase change"

1-Finished a short course on genetics (MIT)

2-Did an intro course named "Justice" (Harvard)

3-Visited US labs

4-Participated in EGI-TF, CHEP 2013, LHC requests (-2017)

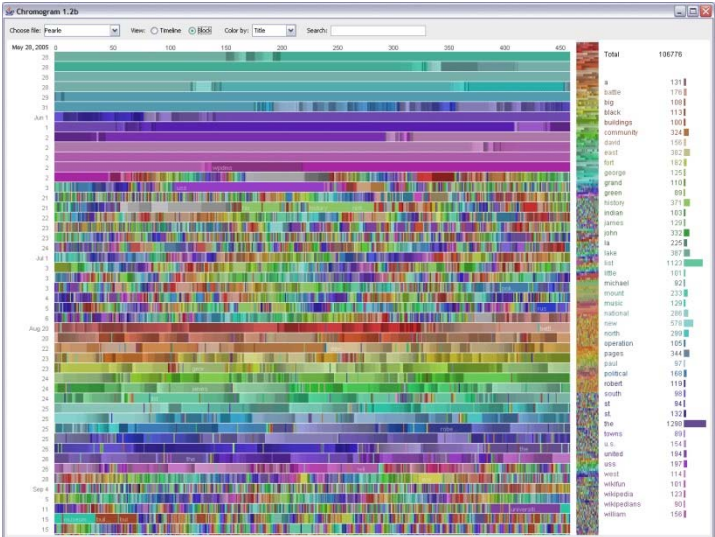
5-This week: e-IRG, ICT2013 Vilnius

After I read a book in Stansted Airport last Sunday,
I started to prepare these B&W slides

All opinions/facts presented in them are on my own

Una visión de BIG DATA

“Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.”



*"In a 2001 research report, META Group (now Gartner) analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional, i.e. increasing **volume** (amount of data), **velocity** (speed of data in and out), and **variety** (range of data types and sources) (**3V**). In 2012, Gartner updated its definition as follows: "Big data are high volume, high velocity, and/or high variety information assets that require new forms of processing **to enable enhanced decision making, insight discovery and process optimization.**"*

Concepto de BIG DATA: Introducción



Inicio ETID > Contenidos del Portal > Referencia de interés

Referencias de Interés

28/11/2012

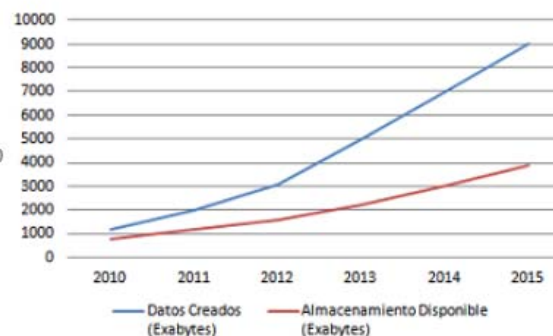
Concepto de BIG DATA y su aplicación a defensa y seguridad

Ámbito: OTROS

En la película "Cortocircuito", de 1986, el robot Número 5 expresaba su necesidad de conocimientos con la frase: ¡datos, necesito más datos!. Hoy en día alimentar esa necesidad no sería ningún problema, solamente en 2010 se generaron más de 1200 exabytes de información (un exabyte equivale a 1000 millones de gigabytes). Enmarcando el ejemplo en el ámbito de la defensa, baste decir que durante 2009 los UAVs de EE.UU. remitieron a sus mandos de control el equivalente a 24 años de grabación.

Sin embargo, hoy en día "Número 5" tendría problemas para:

- Almacenar la información que recibiera.
- Procesar y analizar esa información en forma y en tiempo. El 95 por ciento de esos 1200 exabytes creados en 2010, eran datos no estructurados. Este volumen de datos, conjuntamente con la falta de estructura, supone un desafío técnico para los sistemas de bases de datos relacionales existentes hoy en día.



Eventos

May 2013

May 2013						
today						
M	T	W	T	F	S	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

Próximos eventos

Seminario sobre Aplicaciones de Big Data en los entornos de Defensa y Seguridad

Del 30/05/2013 al 30/05/2013

El CESEDEN/IEEE en colaboración con la Fundación Circulo e Isdefe organizan un seminario sobre "Las tecnologías de Big Data y sus aplicaciones a los entornos de Defensa y Seguridad". El evento tendrá lugar el próximo día 30 de mayo en las instalaciones de Isdefe a las 9:30 de la mañana. [Mas información](#)

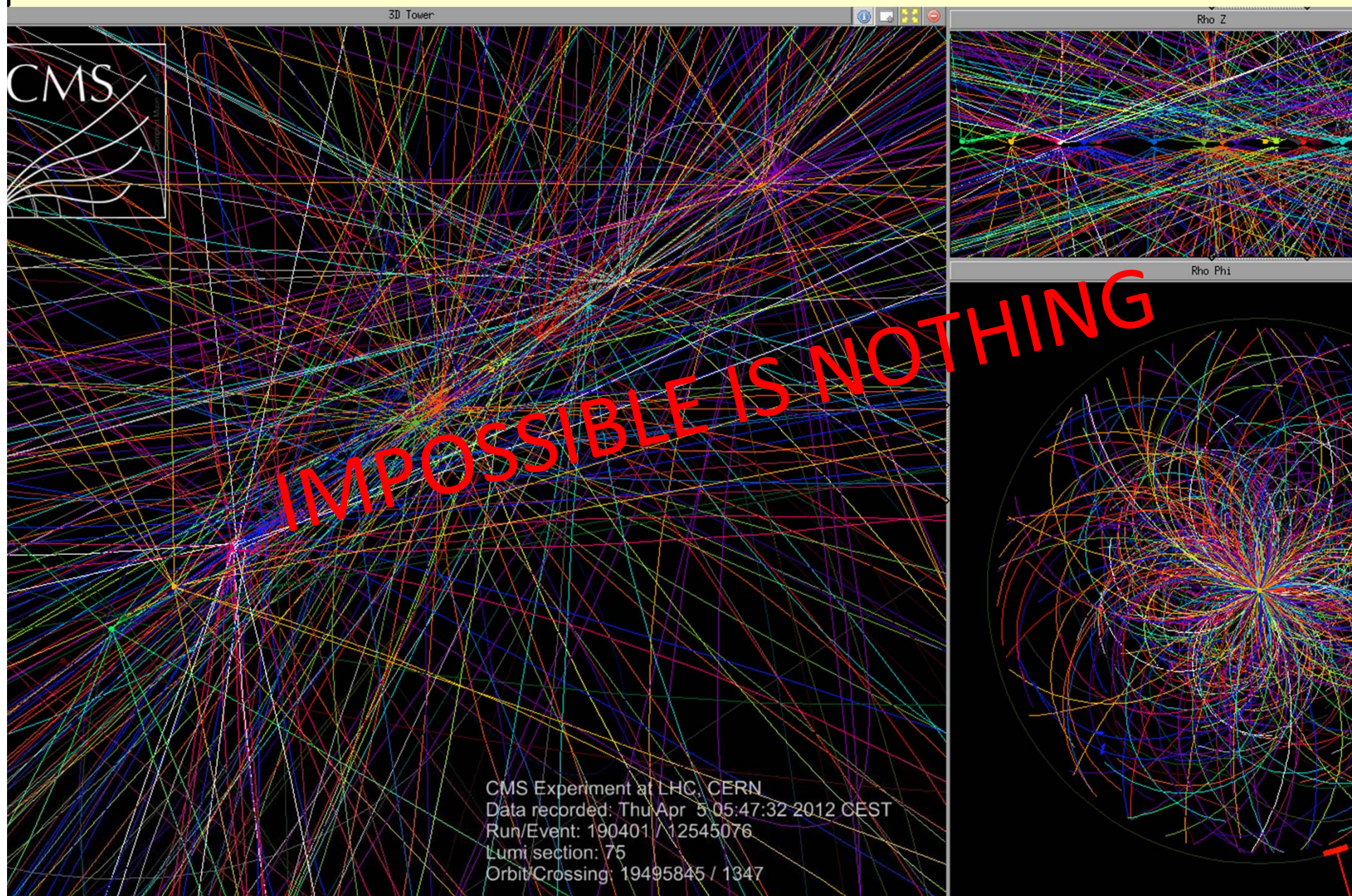
Concepto de BIG DATA: Introducción

- ✦ En estos últimos años, los ámbitos **empresarial**, **académico**, **investigador** y de la **administración** han estado haciendo frente a la avalancha de datos, con la ayuda de un nuevo término, el Big Data.
- ✦ ¿Cómo podemos definir Big Data?
 - ✦ “es el término que describe grandes volúmenes de datos (de terabytes pasamos a **zetabytes**) que se generan a gran velocidad (pasamos de datos en lotes/archivos a datos en “**streaming**”), con una posible componente de complejidad y variabilidad en el formato de esos datos (pasamos de datos estructurados a datos semi-estructurados o **no estructurados**) y que requieren de técnicas y tecnologías específicas para su captura, almacenamiento, distribución, gestión, y **análisis** de la información”.
- ✦ Otras “definiciones”:
 - ✦ “Se considera Big Data cuando **el volumen de los datos se convierte en sí mismo parte del problema** a solventar” (O’Reilly Radar).
 - ✦ “Las tecnologías de Big Data describen un nuevo conjunto de tecnologías y arquitecturas, diseñadas para **extraer valor** y beneficio de grandes volúmenes de datos con una amplia variedad en su naturaleza, mediante procesos que permitan capturar, descubrir y analizar información a alta velocidad y con un **coste reducido**”. (EMC/IDC)

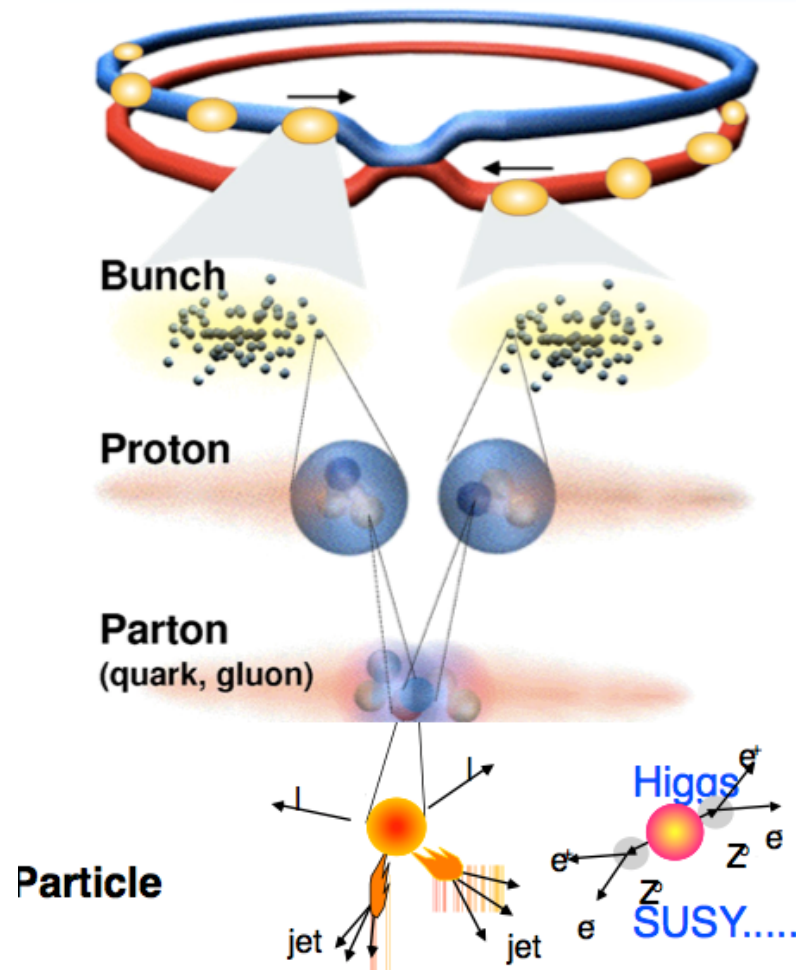
Una visión de BIG DATA: Introducción

- ✦ **Big Data no es una tecnología en sí misma**, si no más bien un planteamiento de trabajo para la obtención de valor y beneficios de los grandes volúmenes de datos que se están generando hoy en día.
- ✦ Se deben contemplar aspectos como los siguientes:
 - Cómo capturar, gestionar y explotar todos estos datos.
 - Cómo asegurar estos datos y sus derivados, así como su validez y fiabilidad.
 - Cómo disponer la compartición de estos datos y sus derivados en la organización para la obtener mejoras y beneficios.
 - Cómo comunicar estos datos y sus derivados (técnicas de visualización, herramientas, y formatos) para facilitar la toma de decisión y posteriores análisis.
- ✦ **¡DEBEMOS CONSTRUIR UNA “VISIÓN” PROPIA DE BIG DATA!**
 - Ejemplo: en investigación la tecnología GRID nos ha permitido resolver el reto del procesamiento de datos de LHC, que “era” un problema Big Data.
 - Para construir esta “visión” **necesitamos conocer la tecnología disponible**
 - Estar “al tanto” de los desarrollos tecnológicos es un reto en sí:
 - Evolución muy rápida de técnicas y capacidades
 - Dificultad de separar interés real y el interés profesional/comercial

Capacidades Técnicas



Una extraordinaria maquina muy compleja...



Proton - Proton	2808 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	$10^{34} \text{cm}^{-2} \text{s}^{-1}$

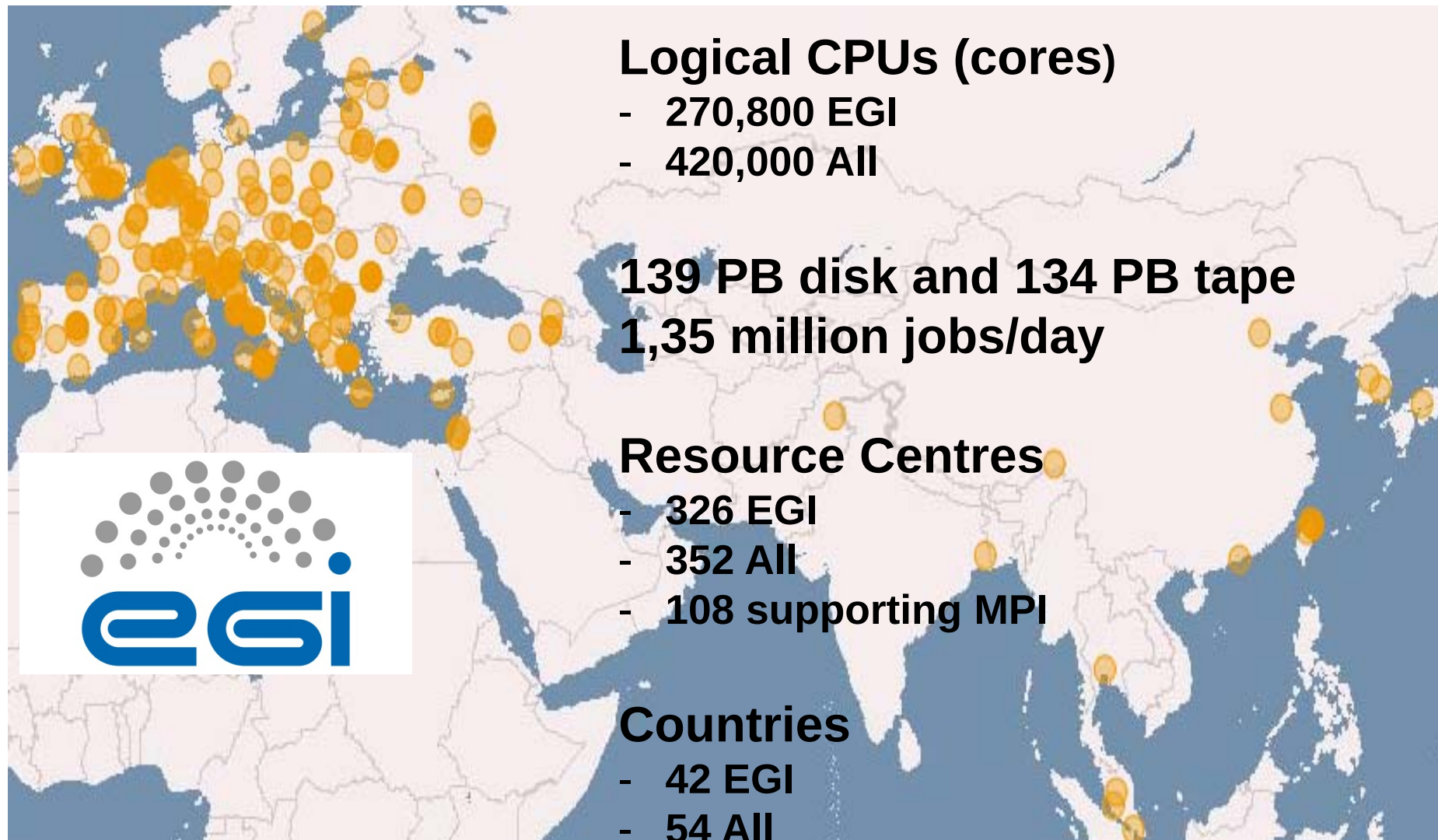
Crossing rate	40 MHz
----------------------	--------

Collision rate $\approx$	10^7 - 10^9
--	-----------------

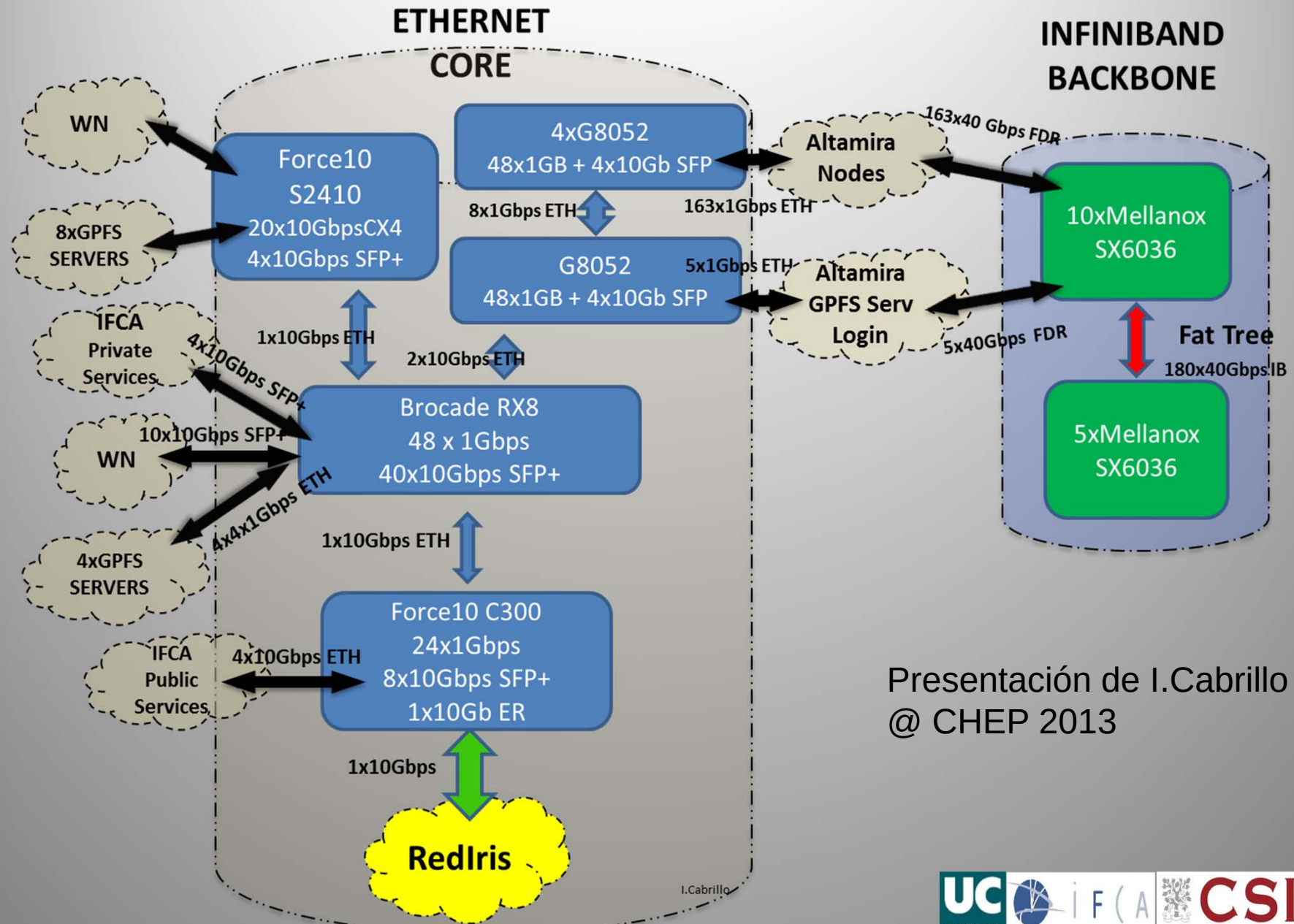
New physics rate $\approx$.00001 Hz

Event selection:
1 in 10,000,000,000,000

European Grid Infrastructure



ALTAMIRA INTEGRATION: Network View



Presentación de I.Cabrillo
@ CHEP 2013

Capacidades Técnicas

Branchiootorenal syndromes
Branchiootic syndrome
Adrenal hyperplasia, congenital
Aldosteronism
Nijmegen breakage syndrome
Giant cell hepatitis, neonatal
Renal tubular acidosis-osteopetrosis syndrome
Segmentation syndrome
Spastic paraplegia
Brain-specific angiogenesis inhibitor
Papillomavirus type 18 integration site
Muscular dystrophy with epidermolysis bullosa
Macular dystrophy, atypical vitelliform
Renal cell carcinoma
Langer-Giedion syndrome
Burkitt lymphoma
Hypothyroidism, hereditary congenital
Goiter, adolescent multinodular and nonendemic

Fructose intolerance
Basal cell carcinoma, sporadic
Muscular dystrophy, Fukuyama congenital
Basal cell nevus syndrome
Dysautonomia (Riley-Day syndrome)
Esophageal cancer
Endotoxin hyporesponsiveness
Amyotrophic lateral sclerosis, juvenile dominant
Berardinelli-Seip congenital lipodystrophy
Dystonia, torsion, autosomal dominant
Lethal congenital contracture syndrome
Leukemia, acute undifferentiated
Tuberous sclerosis
Hemolytic anemia
Telangiectasia, hereditary hemorrhagic
Ehlers-Danlos syndrome, types I and II
Joubert syndrome
Leukemia, T-cell acute lymphoblastic

Epithelioma, self-healing, squamous
Leukemia, T-cell acute lymphoblastic
Muscular dystrophy, limb-girdle, type 2H
Bladder cancer
Sex reversal, XY, with adrenal failure
Leukemia transcription factor, pre-B-cell
Porphyria, acute hepatic
Lead poisoning, susceptibility to
Citrullinemia
Dopamine-beta-hydroxylase deficiency
Amyloidosis, Finnish type
Microcephaly, primary autosomal recessive
Leigh syndrome
Leukemia
Nail-patella syndrome
Prostaglandin D2 synthase (brain)
Pituitary hormone deficiency

Polycystic juvenile osteoarthritis
Prostate cancer
Progressive external ophthalmoplegia
Corneal dystrophy, Thiel-Behnke type
Leukemia, T-cell acute lymphocytic
Spinocerebellar ataxia, infantile-onset
Split hand/foot malformation, type 3
Polycystic kidney disease
Meningioma-expressed antigen
Adrenal hyperplasia, congenital
Diabetes mellitus, insulin-dependent
Anterior segment mesenchymal dysgenesis
Cataract, congenital
Malignant brain tumors
Glioblastoma multiforme
Medulloblastoma
Crouzon syndrome
Jackson-Weiss syndrome
Beare-Stevenson cutis gyrata syndrome

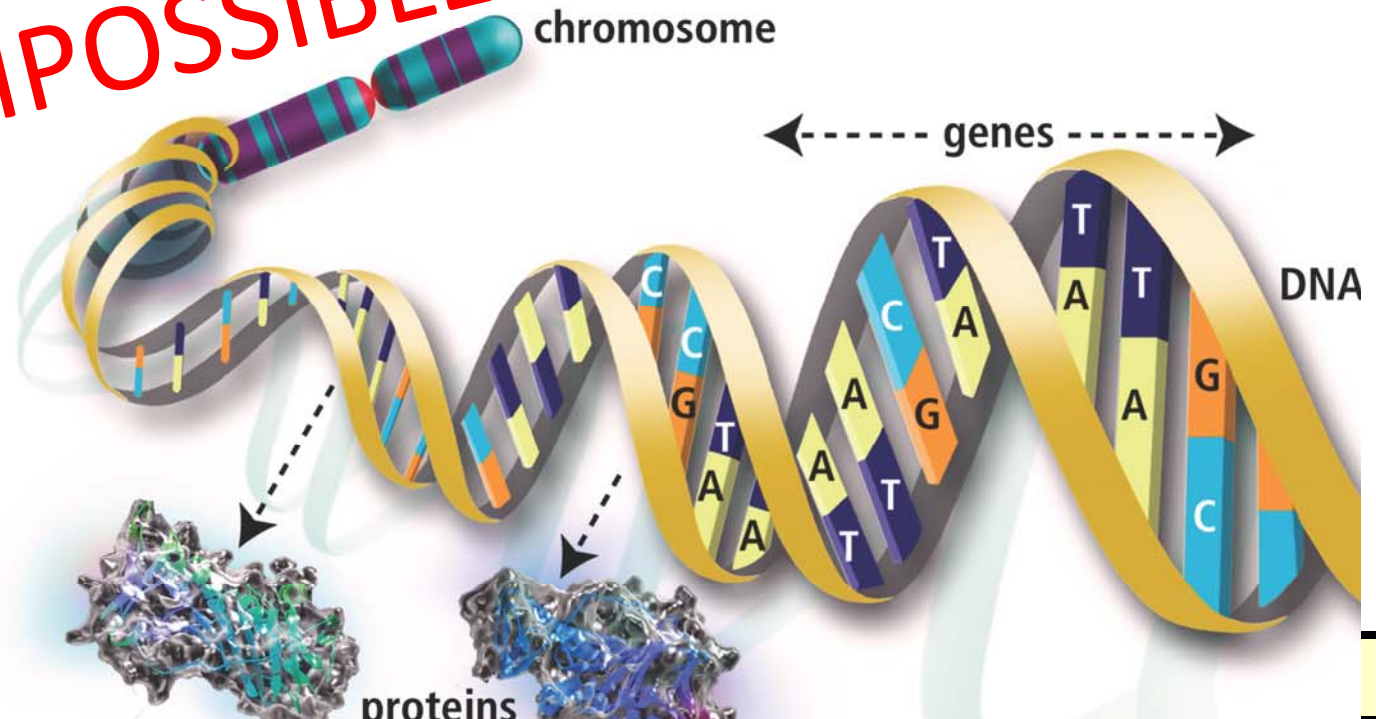
Cholesteryl ester storage disease
Tumor necrosis factor receptor superfamily
Autoimmune lymphoproliferative syndrome
Epidermolysis bullosa, generalized atroph
Optic nerve coloboma with renal disease
Prostate cancer
Neurofibrosarcoma
Porphyria, congenital erythropoietic
Endometrial carcinoma
Gyrate atrophy of choroid and retina
Pancreatic lipase deficiency
Glaucoma
Pfeiffer syndrome
Apert syndrome
Saethre-Chatzen syndrome
Schizencephaly
Polykaryocytosis inducer (promoter)
Usher syndrome, autosomal recessive, sev

105 million base pairs

Basal ganglia calcification (Fahr disease)
Multinodular goiter
Retinitis pigmentosa, autosomal dominant
Leukemia/lymphoma, T-cell
Oculopharyngeal muscular dystrophy, autosomal recessive
APEX nuclease (multifunctional DNA repair enzyme)
Cardiomyopathy, familial hypertrophic
Oligodontia
Goiter, familial
Carbohydrate-deficient glycoprotein syndrome, type II
Elliptocytosis
Spherocytosis
Anemia, neonatal hemolytic, fatal and near-fatal
Arrhythmogenic right ventricular dysplasia
Marfan syndrome, atypical
DNA mismatch repair gene MLH3
Diabetes mellitus, insulin-dependent
Krabbe disease
Hypothyroidism, congenital
Thyroid adenoma, hyperfunctioning
Graves disease
Hyperthyroidism, congenital
Usher syndrome, autosomal recessive
Emphysema-cirrhosis
Hemorrhagic diathesis
X-ray repair

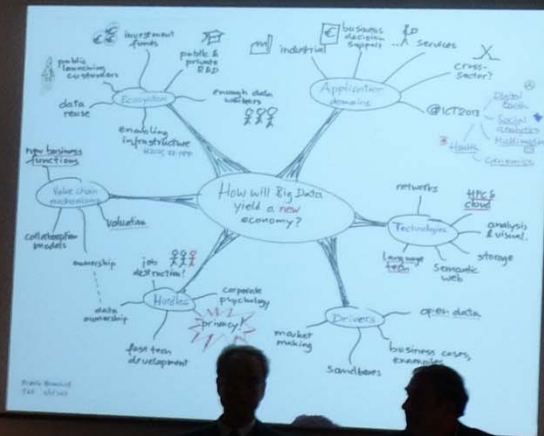
76 million base pairs

IMPOSSIBLE IS NOTHING



The \$1,000 genome is a term that predicts a new era of predictive and personalized medicine during which the cost of FULL genome sequencing an individual or patient drops to roughly USD\$1,000

Privacy of individuals undergoing genetic testing must be protected under all circumstances



corporate
psychology

rdles

privacy!

- legal
- technological

Capacidades Técnicas: Adquisición y Transmisión de Datos

Fuentes de Información (Sensores):

Instrumentación

- Redes de sensores
- Cámaras
- **Satélites**



Uso personal

- **Smartphones**
- Automóviles
- Instrumentación personal (DNA chips)

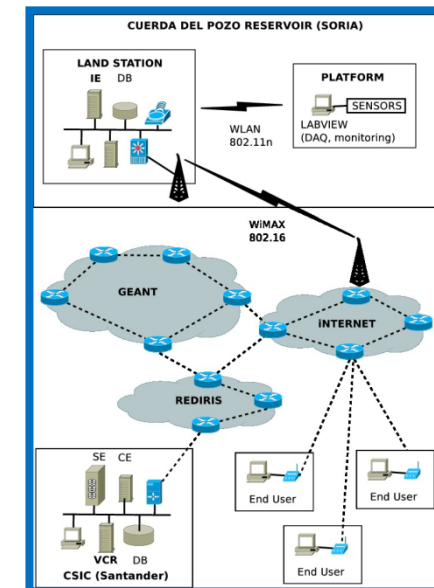
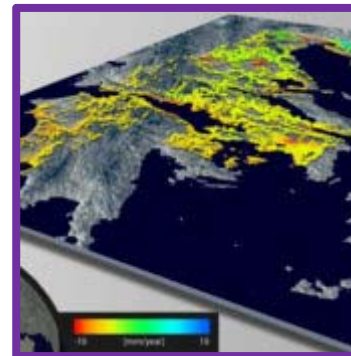


Mensajería en la red

OPEN DATA

Ejemplos Globales:

- **ESA's GMES (Observing the Earth)**
- SmartCities



Integración: Estándares: Sensor Web Enablement/ Web Services

Capacidades Técnicas: Infraestructura

Centros de procesamiento de datos

- **Supercomputadores** top500: hasta 20 Petaflops
 - España: Red Española de Supercomputación (RES)
- **GRID**: hasta 400.000 cores (WLCG), >100 Petabytes
 - España: IberGrid
- Componentes:
 - Almacenamiento: HADOOP, GPFS, Lustre...
 - Clusters: Redes Infiniband



Redes de comunicación

- España: **RedIris-Nova** (fibra oscura, n x 10Gb/s)



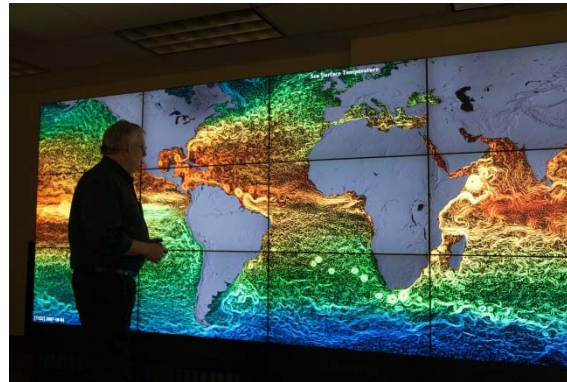
Cloud

- Amazon , IBM, BT, Verizon, ..., Arsys, INDRA,
- **IBERCLOUD**

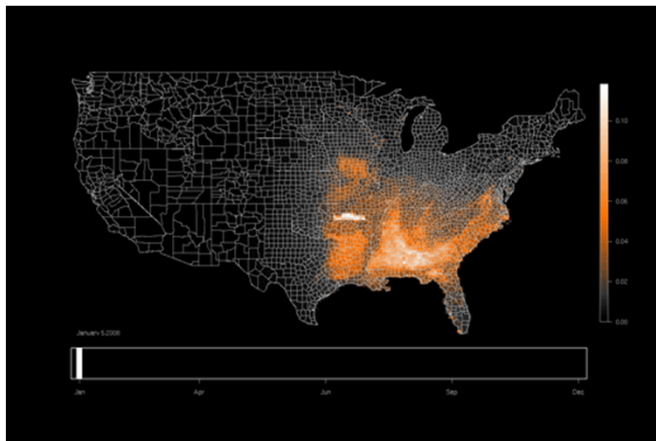


Presentación

☉ “Hardware”: Dashboards, Walls, Visores personales 3D

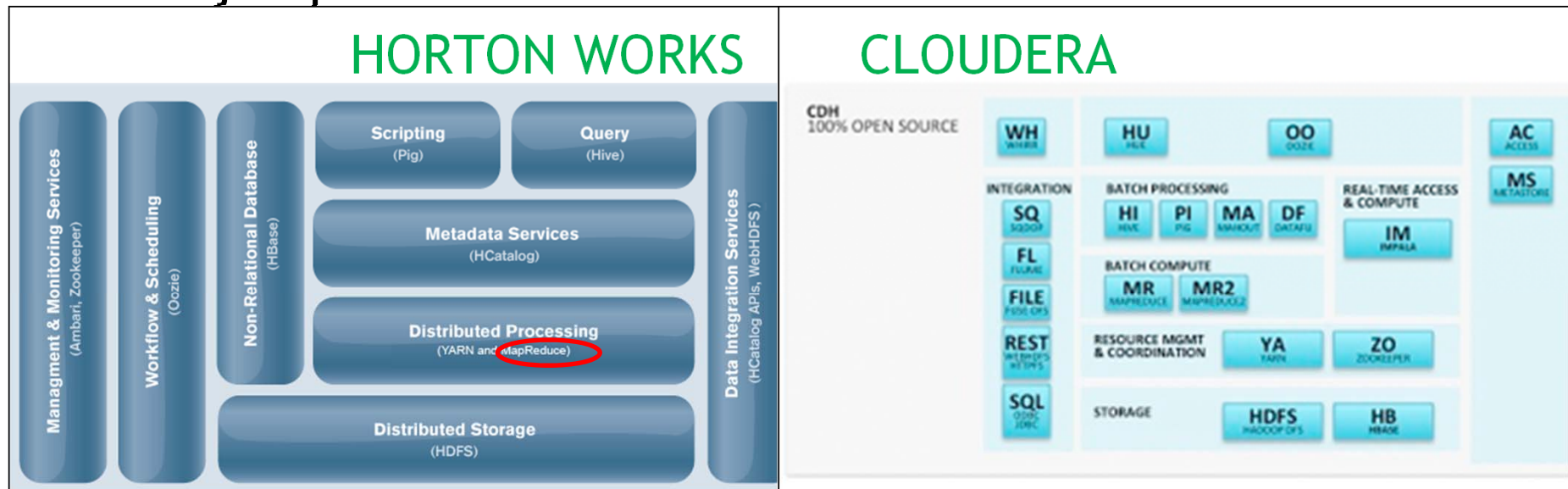


☉ Software/Gráficos:



Capacidades Técnicas: Tratamiento Automático

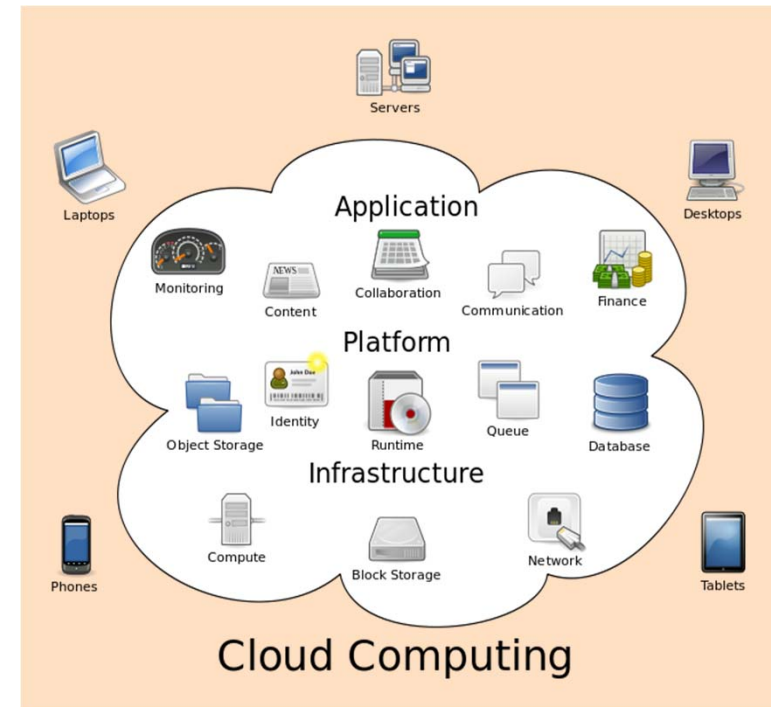
- ⊕ Reto: **3V** (Volumen, Velocidad, Variabilidad)
- ⊕ Arquitectura de una aplicación:
 - ⊕ Nivel de adquisición/almacenamiento/persistencia
 - ⊕ Nivel de análisis
 - ⊕ Nivel de presentación
- ⊕ Dos ejemplos:



Capacidades Técnicas: Big Data y CLOUD

Cloud Computing

- Un paradigma que permite ofrecer **servicios** de computación a través de Internet.
- Se apoya técnicamente en la virtualización
- Tecnología madura, pero no totalmente lista para grandes colaboraciones



¿Por qué Big Data “encaja” bien con Cloud Computing?

e-IRG [white paper](#) 15 may 2013:

“These two major challenges, Big Data and cloud computing, are not totally independent: not only because Big Data may require a huge computing power but also because Big Data could represent the killer application for clouds”.

Casos de Uso: Aprendizaje Automático

ÍNDICE

OBJETIVO:

*de los datos
a la información*

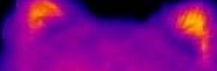
Técnicas “conocidas”

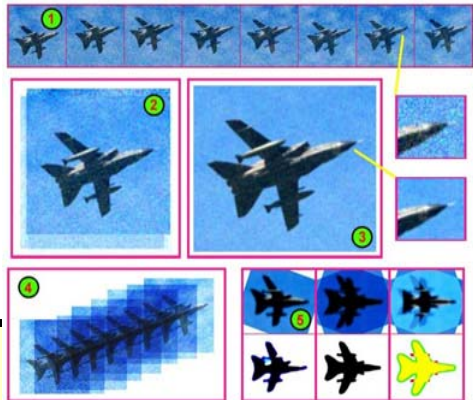
Implementación en un marco BIG DATA?

- Programación ad-hoc, incluyendo “paralelización”, en algunos casos en un lenguaje ad-hoc (ej. R)
- Suites/plataformas analíticas

1	Aprendizaje semiautomático
1.1	Depuración, exploración y visualización
1.1.1	Análisis de componentes principales
1.1.2	Análisis factorial
1.2	Aprendizaje supervisado
1.2.1	Modelos de Clasificación
1.2.1.1	Árboles de clasificación
1.2.1.2	Clasificadores basados en reglas
1.2.1.3	Clasificadores basados en el vecino más cercano .
1.2.1.4	Clasificadores bayesianos.....
1.2.1.5	Redes neuronales
1.2.1.6	Maquinas de soporte vectorial
1.2.2	Modelos de regresión
1.3	Aprendizaje no supervisado
1.3.1	Asociación.....
1.3.2	Análisis clúster.....

Casos de Uso: Procesado de Imágenes

- 
- NASA/PAC
- 35 94.7
30 90
25 80
20 69.6 °C °F



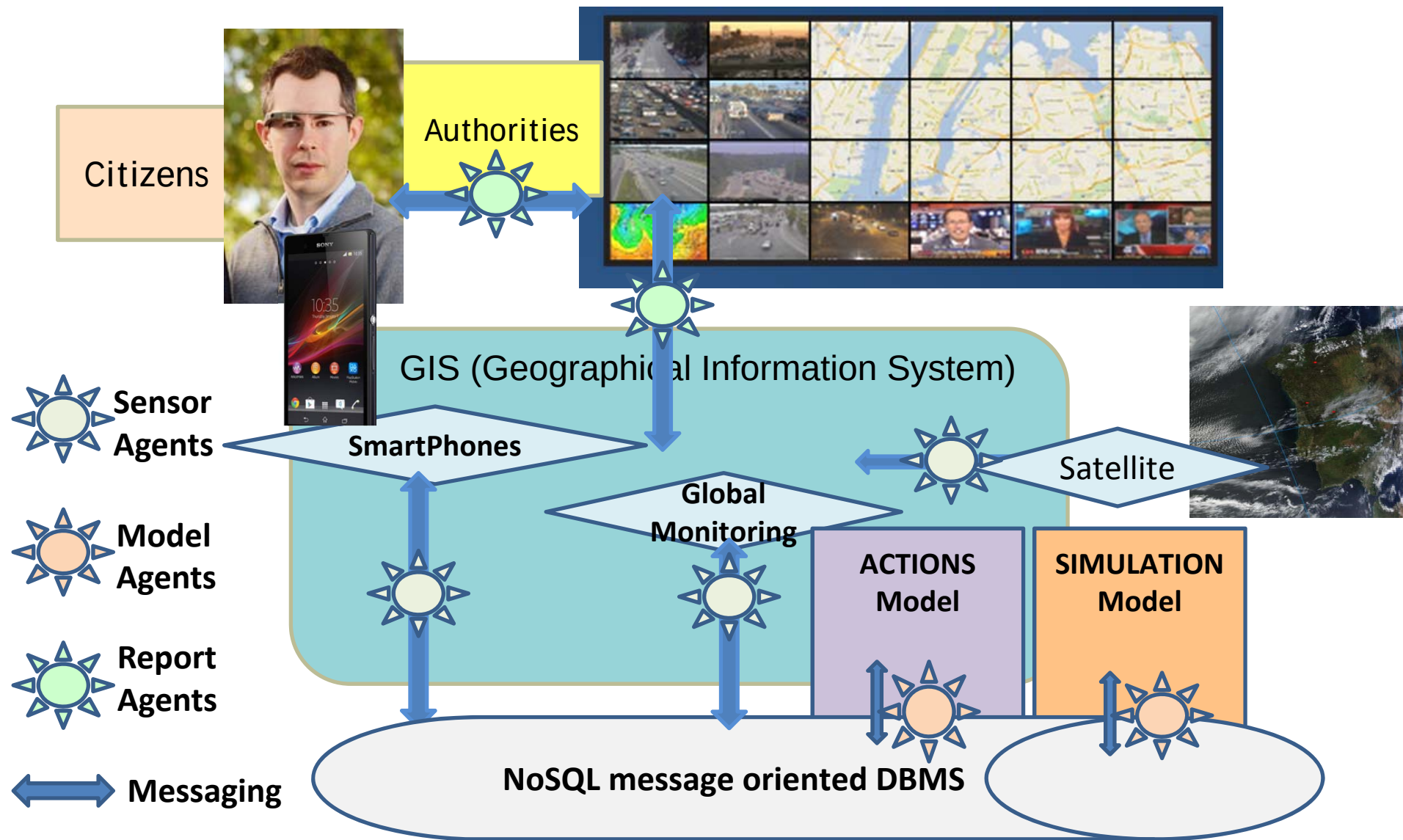
My favorite digital images of last months:

- my digital fingerprints plus an updated photo
- an accurate 3D profile of my face (TLS)
- an accurate 3D profile of my body (AMW)
- my country, my city, my street
- my car parked in front of my home
- my photo crossing the finish line in Amsterdam

Joining my favorite digital info of last years:

- my "private" gmail
- my searches in Internet
- my "private" wifi information (keys included)

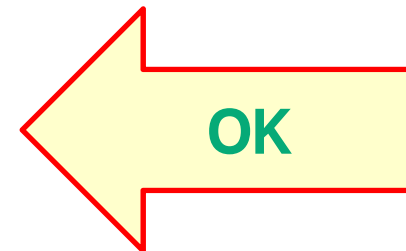
Caso de uso: gestión de emergencias



Casos de Uso Real: Problemas Prácticos

Recursos de computación:

- ❑ Cloud computing (p.ej. 100Tb, 5M horas = 100Keuros)
- ❑ Sistemas GIS, bases de datos, etc.
- ❑ Sistemas de predicción/monitorización
 - Sistema Europeo de Información de Incendios Forestales de la UE (EFFIS) (accesible via GEOSS)
- ❑ Software específico (agentes, noSQL, monitorización)



COORDINACIÓN

- ❑ Autoridades y Expertos
- ❑ Equipos de intervención
- ❑ Ciudadanos

Adquisición de datos en tiempo real y transmisión (**alerta** e intervención)

- ❑ Terminales móviles, posiciones **[PRIVACIDAD]**
- ❑ Información de Satélites (?)
- ❑ Aviones/UAV/drones de intervención (?)
- ❑ Equipos y vehículos de intervención in-situ

Toma de decisiones ???

Ejemplo: INCENDIOS 2012

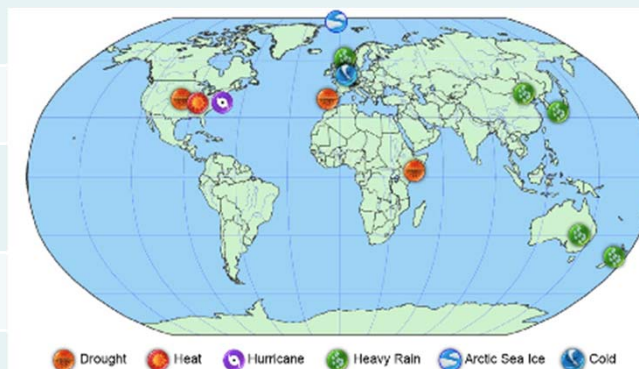
Informe MAGRAMA

210.000 Ha **(+83%)**

~ 75 M euros invertidos

IMPOSSIBLE IS NOTHING ?

Extreme event	Evidence for anthropogenic influence ?
US heatwave	YES
Hurricane Sandy storm surge	YES
September Arctic sea ice minimum	YES
February European cold spell	NO
Wet UK summer	NO
Iberian winter drought	YES
Rainfall deficits in Eastern Kenya and Southern Somalia	NO
North China floods, July	NO
Heavy rainfall in Southwestern Japan, July	NO
Extreme rainfall over Eastern Australia, March	YES
Extreme rainfall, Goldan Bay, New Zealand, December 2011	YES
US drought	NO



6onCORE IPv6 INTELLIGENT NETWORK ENABLER

1 Set up on-site intervention command center using 6onCORE

- Compact mobile network node
- Installed in first responder vehicle or fully portable
- Provides communication services for intervention management by selecting best available networks for the required services
- Connects to 6onFIRE sensor deployments on site



2 Deploy 6onFIRE sensor system on site of the intervention

- Ad-hoc or permanent deployment
- Various supported sensors – temperature, avalanche level, radiation sensor, water level, butane sensor ...
- 6onMOBILE application for manual triage reporting
- 6onFIRE gateway connects sensors to the command

6onDASHBOARD IoT-DRIVEN INTERVENTION MANAGEMENT

3 Manage the intervention with 6onDASHBOARD

- Central web-based intervention management and monitoring portal
- Tailor-made for intervention site monitoring
- Customizable views and access control for strategic, command, lead and operational levels

6onMOBILE TRIAGE APPLICATION

4 Use a one-touch triage application to report conditions and events from the intervention site

- Triage application for Android and iOS
- Simple one-touch reporting from the intervention site
- Alerts transmitted to 6onDASHBOARD

6onMOBILE features

- Instant one-touch alerting – Fire, Earthquake, Tree down, Avalanche, Heavy snow, Line down, No service, Unknown
- Triage reporting of on-site conditions – Fire, Flood, Earthquake; with optional reporting on Facility, Person, Animal, Hazardous
- Report procedures customizable to match standardized protocols
- Reports and alerts transmitted to 6onDASHBOARD in real time



WHAT NEXT?

WHAT NEXT is PRECISELY... the KEY in BIG DATA...

WHAT NEXT is KNOWLEDGE

NEXT is what IBM would name WATSON

NEXT IS Knowledgebases (building the semantic web) like DBpedia using techniques like YAGO

NEXT IS Connecting IA with the robotic field to empower any "unmanned" object

Fiat justitia

« Je vis comme je peux dans ce pays malheureux, riche de son peuple et de sa jeunesse, provisoirement pauvre dans ses élites, lancé à la recherche d'un ordre nouveau et d'une renaissance à laquelle je crois. Sans liberté vraie et sans un certain honneur, je ne puis vivre. »

« J'ai toujours pensé que si l'homme qui espérait dans la condition humaine était un fou, celui qui désespérait des événements était un lâche. »