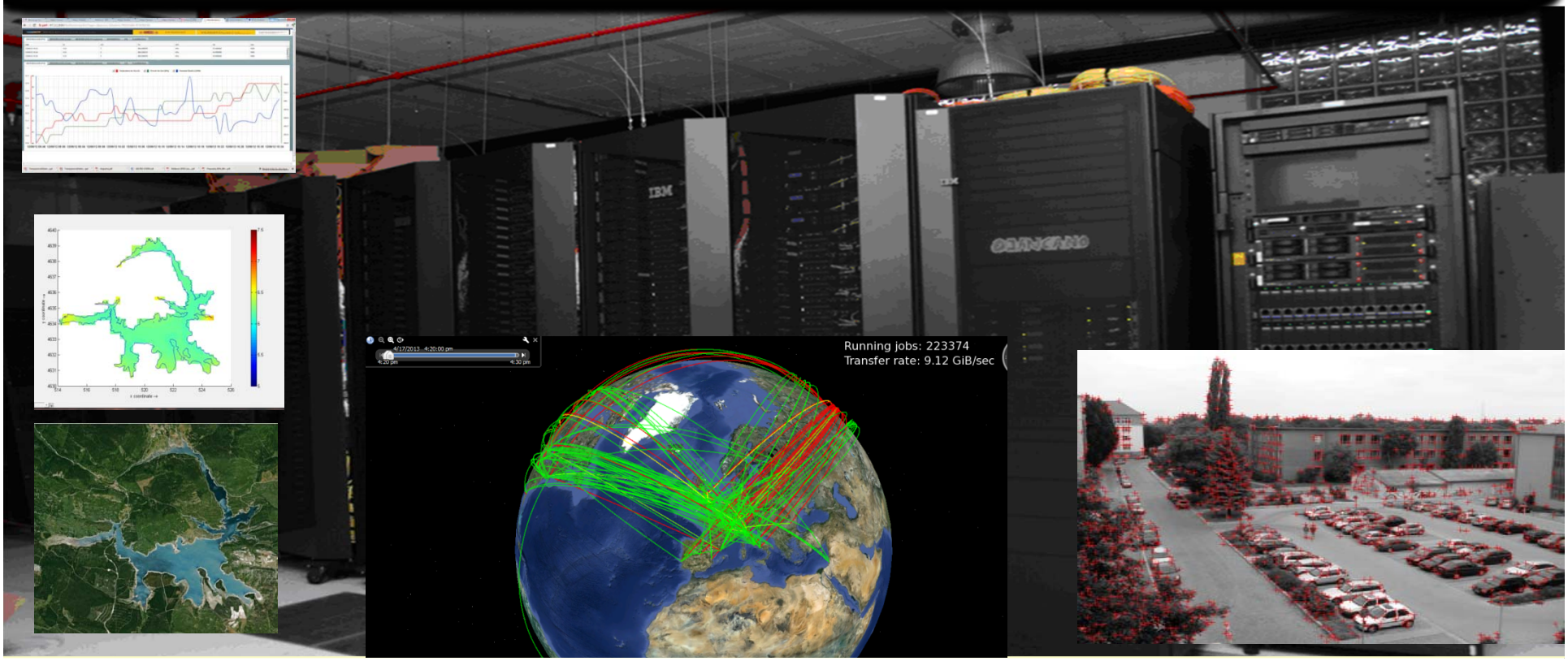
BIG DATA
desde el punto de vista tecnológico



JORNADA "BIGDATA APLICADO EN DEFENSA Y SEGURIDAD"

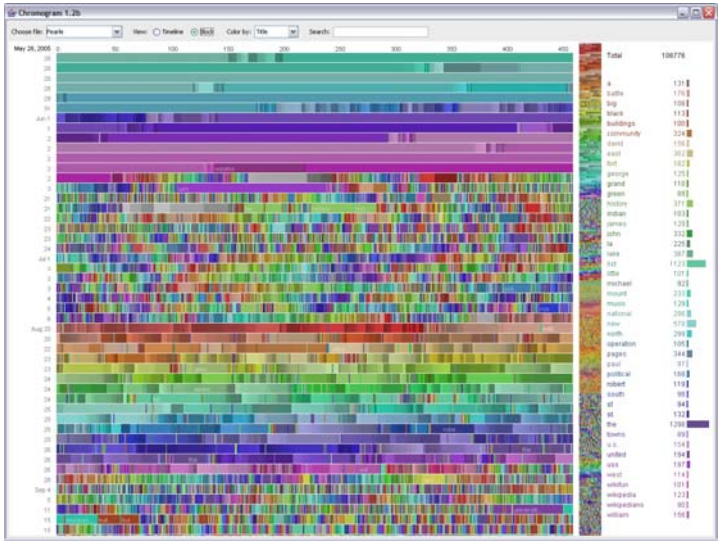
RESULTADOS DEL GRUPO DE TRABAJO, Presentado por Jesús Marco de Lucas

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

Una visión de BIG DATA



- Big Data según Wikipedia:

"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."

- Un poco de historia y evolución:

"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

GOBIERNO DE ESPAÑA | MINISTERIO DE DEFENSA | SECRETARÍA DE ESTADO DE DEFENSA | DIRECCIÓN GENERAL DE ARMAMENTO Y MATERIAL

Portal de Tecnología e Innovación del Ministerio de Defensa

MAPA WEB | CONTACTO | BUSCAR

Presentación | Estrategia de Tecnología e Innovación | Contenidos del Portal | Contacto y Participación

Inicio ETID > Contenidos del Portal > Referencia de interés

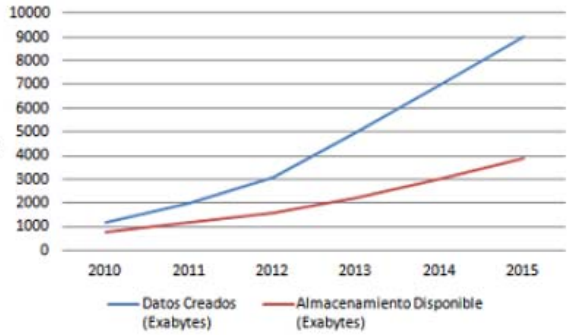
Referencias de Interés

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

« 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

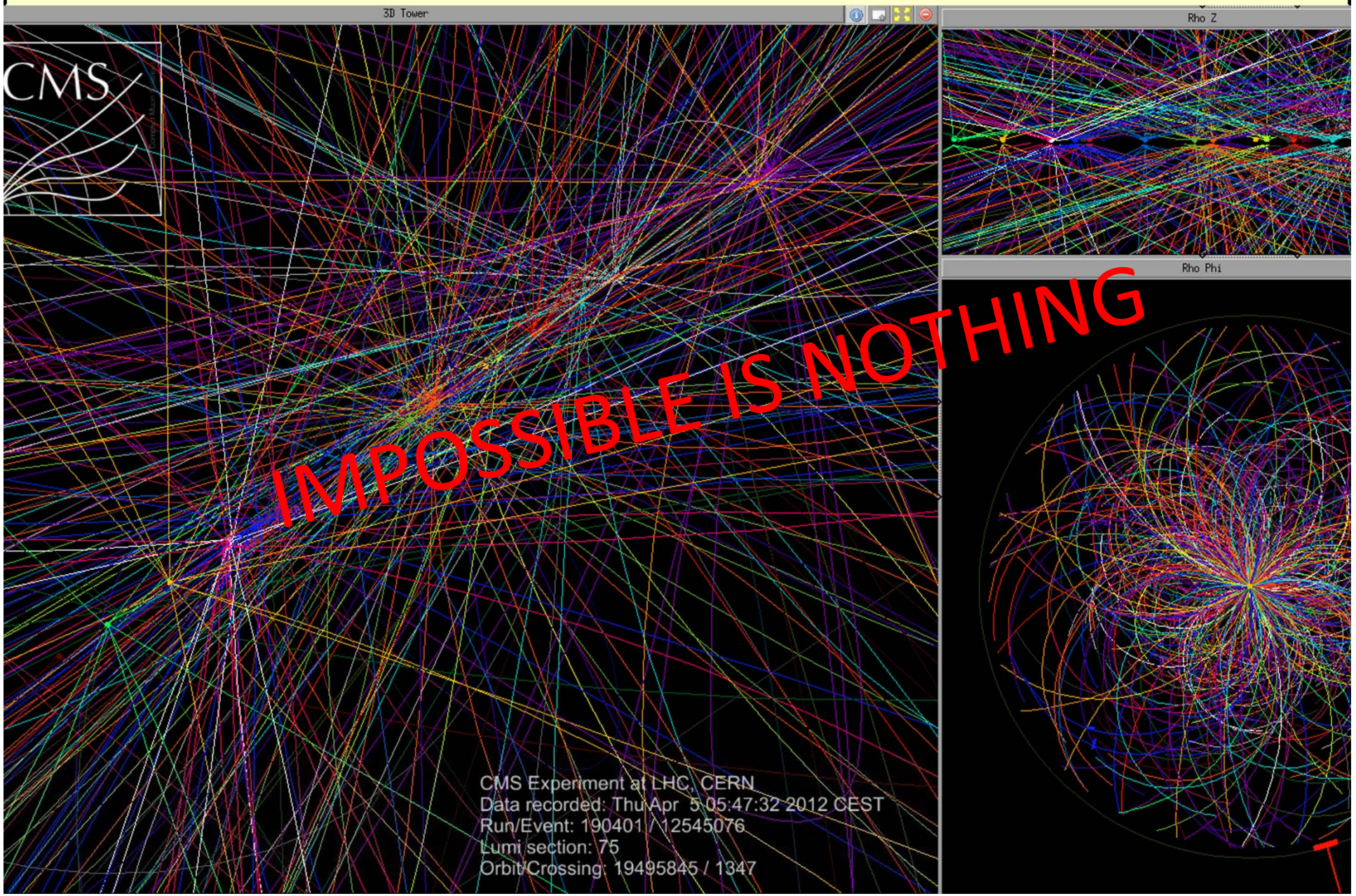
Concepto de BIG DATA: Introducción

- 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)*

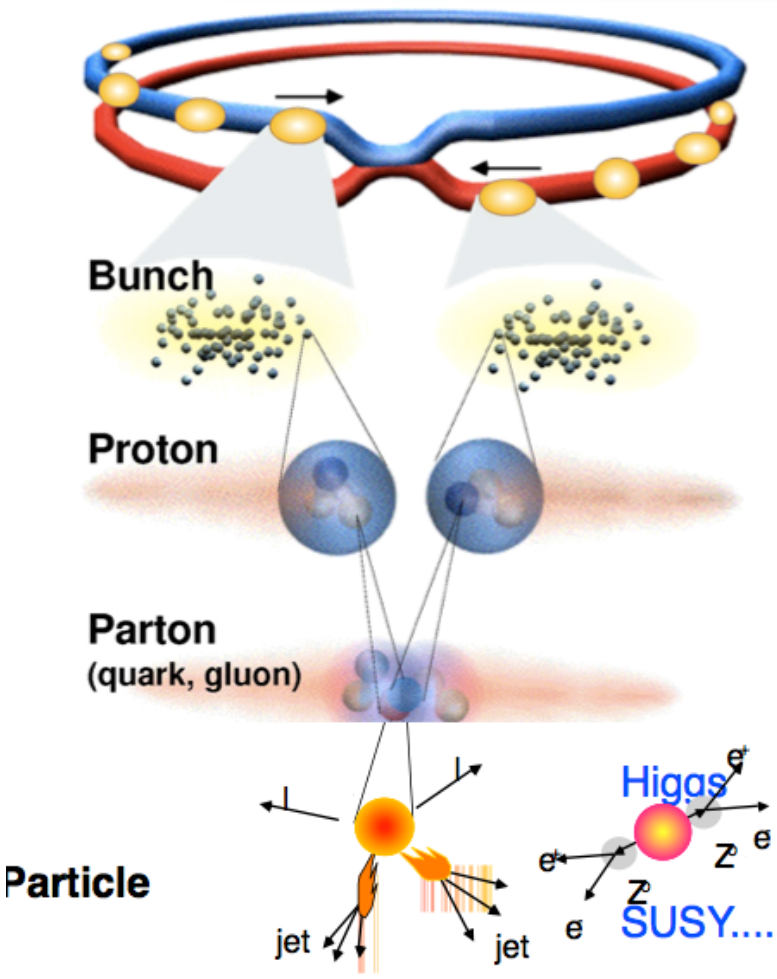
Una visión de BIG DATA: Introducción

- **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 procesado 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…

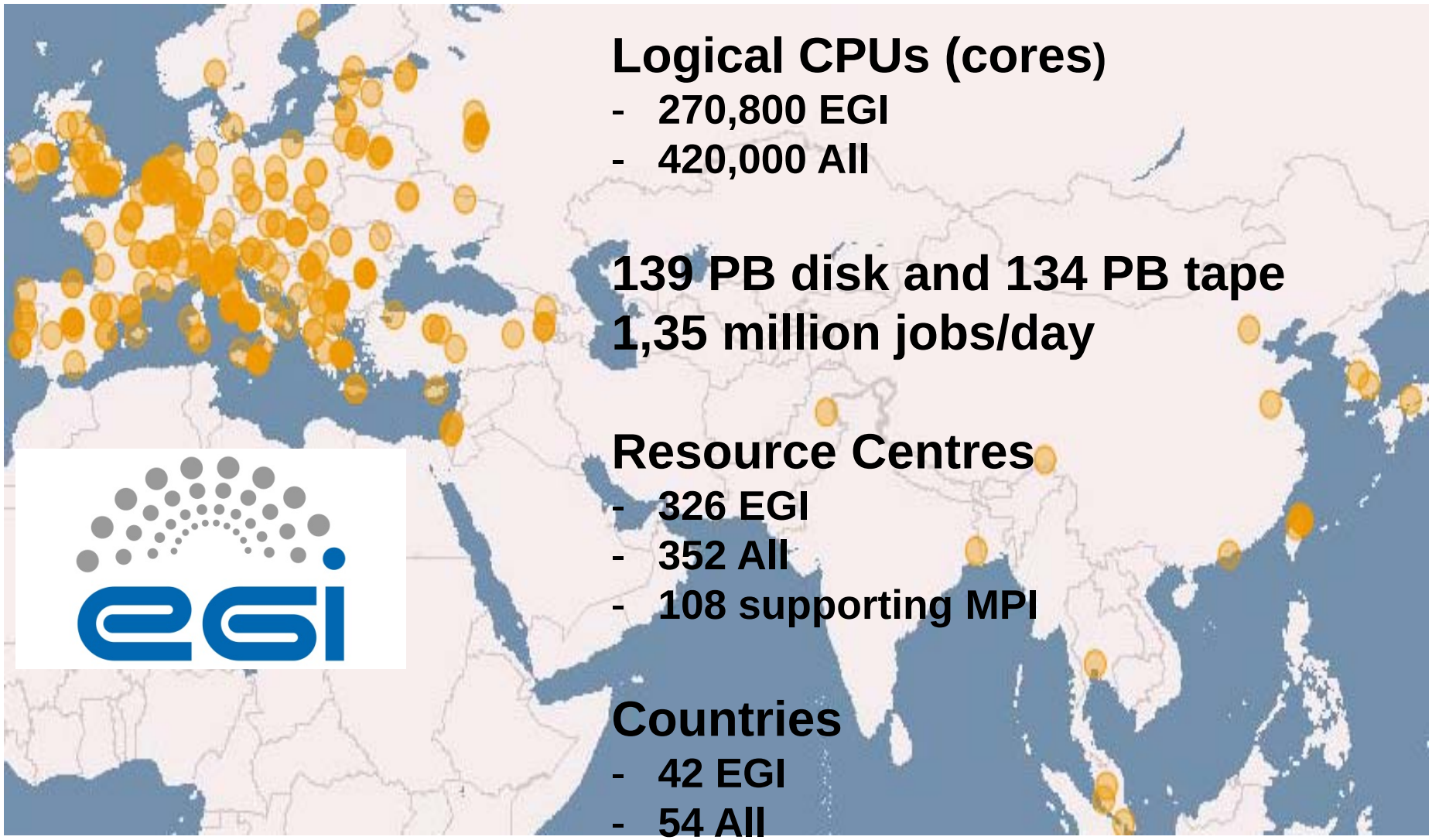


Proton - Proton	2808 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	$10^{34} cm^{-2} s^{-1}$
Crossing rate	40 MHz
Collision rate ≈	10^{7}-10^{9}

New physics rate ≈ .00001 Hz

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

European Grid Infrastructure



Logical CPUs (cores)
- **270,800 EGI**
- **420,000 All**

139 PB disk and 134 PB tape
1,35 million jobs/day

Resource Centres
- **326 EGI**
- **352 All**
- **108 supporting MPI**

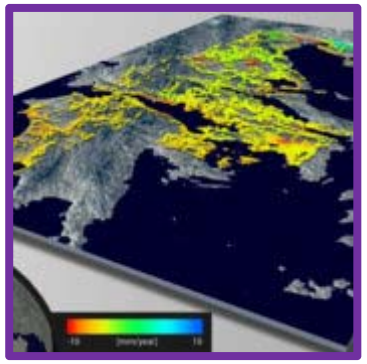
Countries
- **42 EGI**
- **54 All**

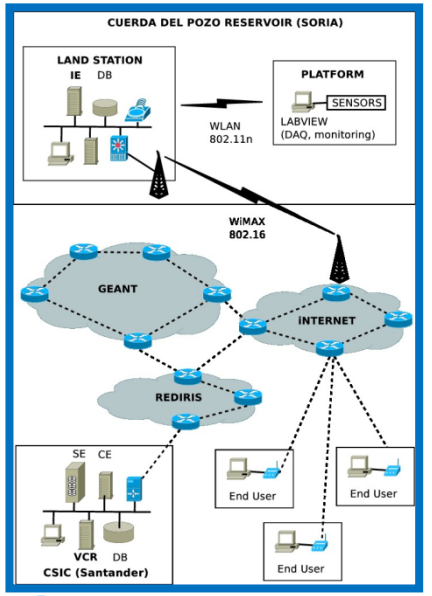
Capacidades Técnicas

IMPOSSIBLE IS NOTHING

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 procesado 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

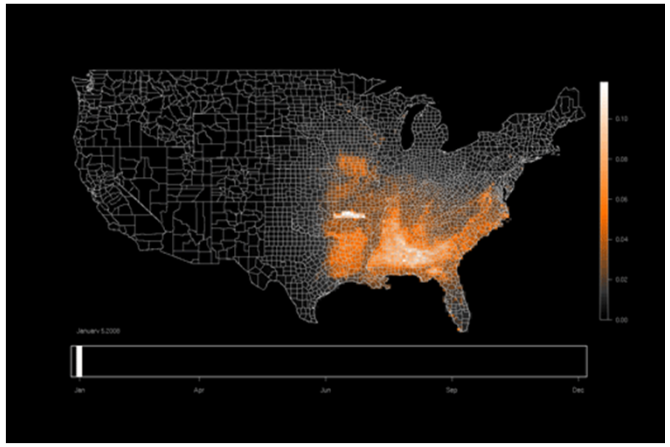
- Redes de comunicación
  - España: RedIris-Nova (fibra oscura, n x 10Gb/s)

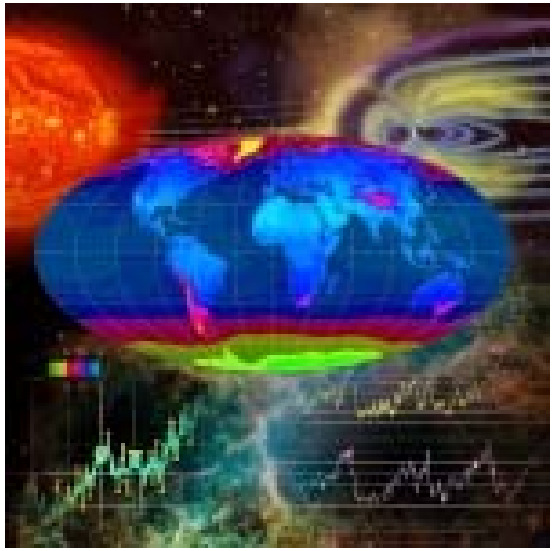
- Cloud
  - Amazon , IBM, BT, Verizon, …, Arsys, INDRA,
  - *IBERCLOUD*

Presentación

- **“Hardware”: Dashboards, Walls, Visores personales 3D**

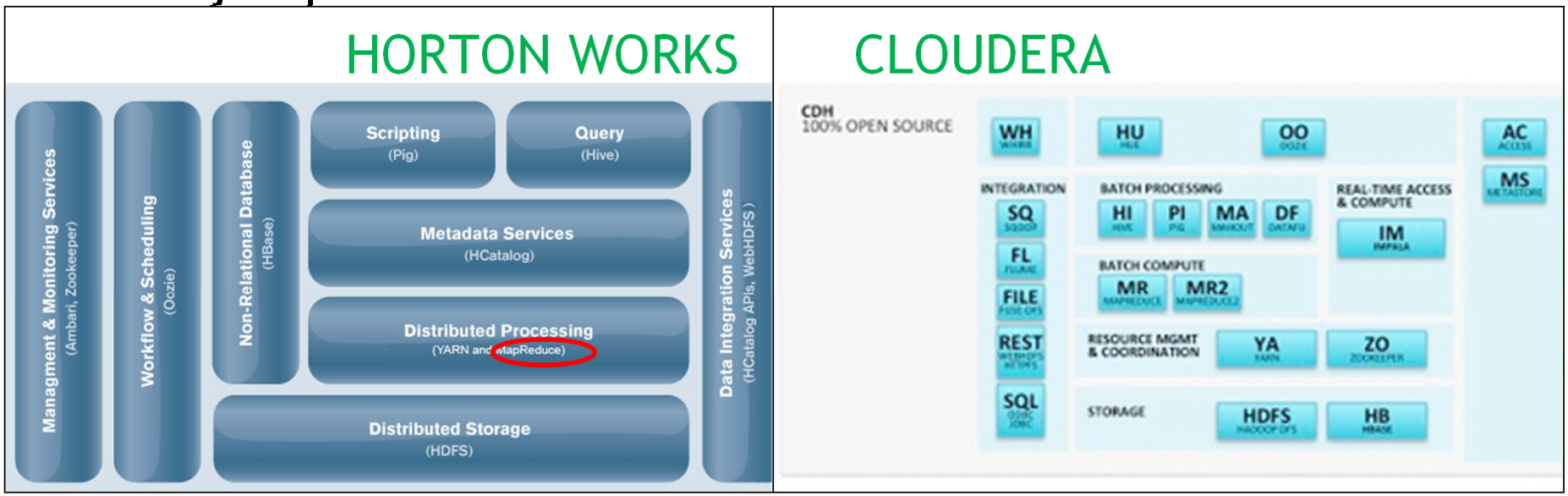
- **Software/Gráficos:**





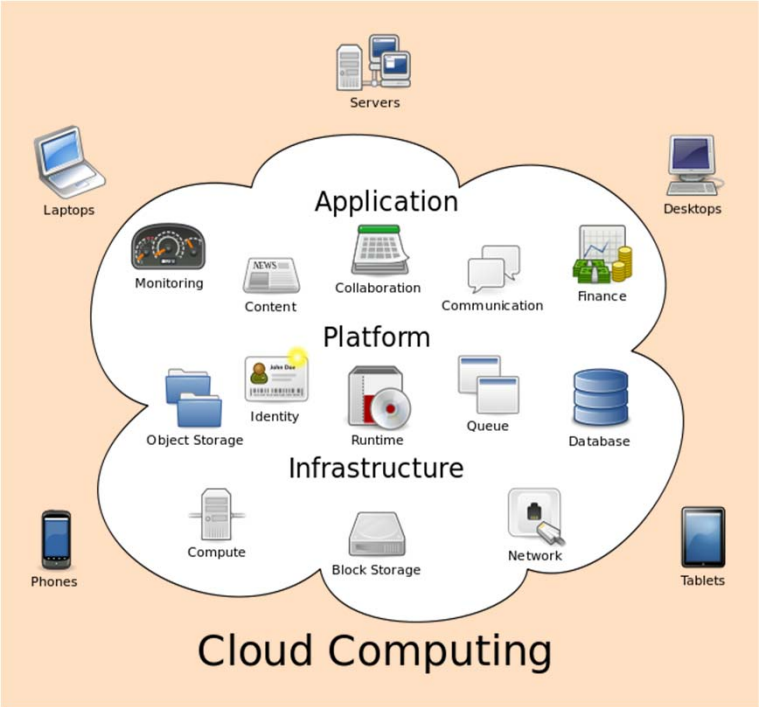
Capacidades Técnicas: Tratamiento Automático

- 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

- OBJETIVO:

de los datos

a la información

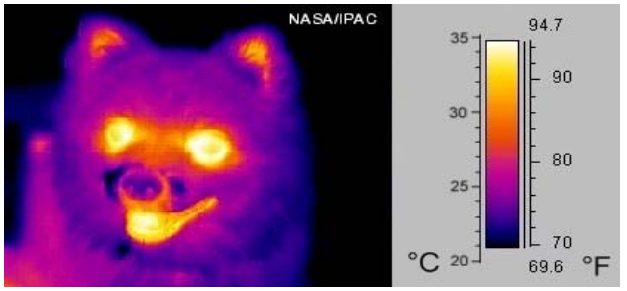
- Técnicas "conocidas"

- Implementación en un marco BIG DATA?
  - Programación ad-hoc, incluyendo "paralelización", en algunos casos en un lenguaje ad-hoc (ej. R)
  - Suites/plataformas analíticas

ÍNDICE

Casos de Uso: Procesado de Imágenes

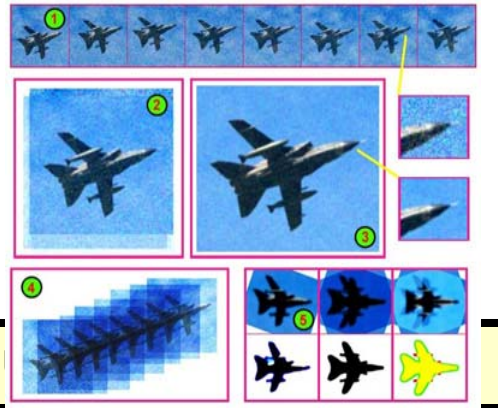
- Las **imágenes** y **vídeos** (secuencias de imágenes): **~80 %** de los grandes volúmenes de datos **no estructurados** corporativos y públicos.
- Ejemplos de sistemas de captura:
  - cámaras CCD sensibles al espectro visible
  - cámaras infrarrojas
  - sistemas LIDAR ("iluminación" láser)
  - sistemas SAR ("iluminación" radar)



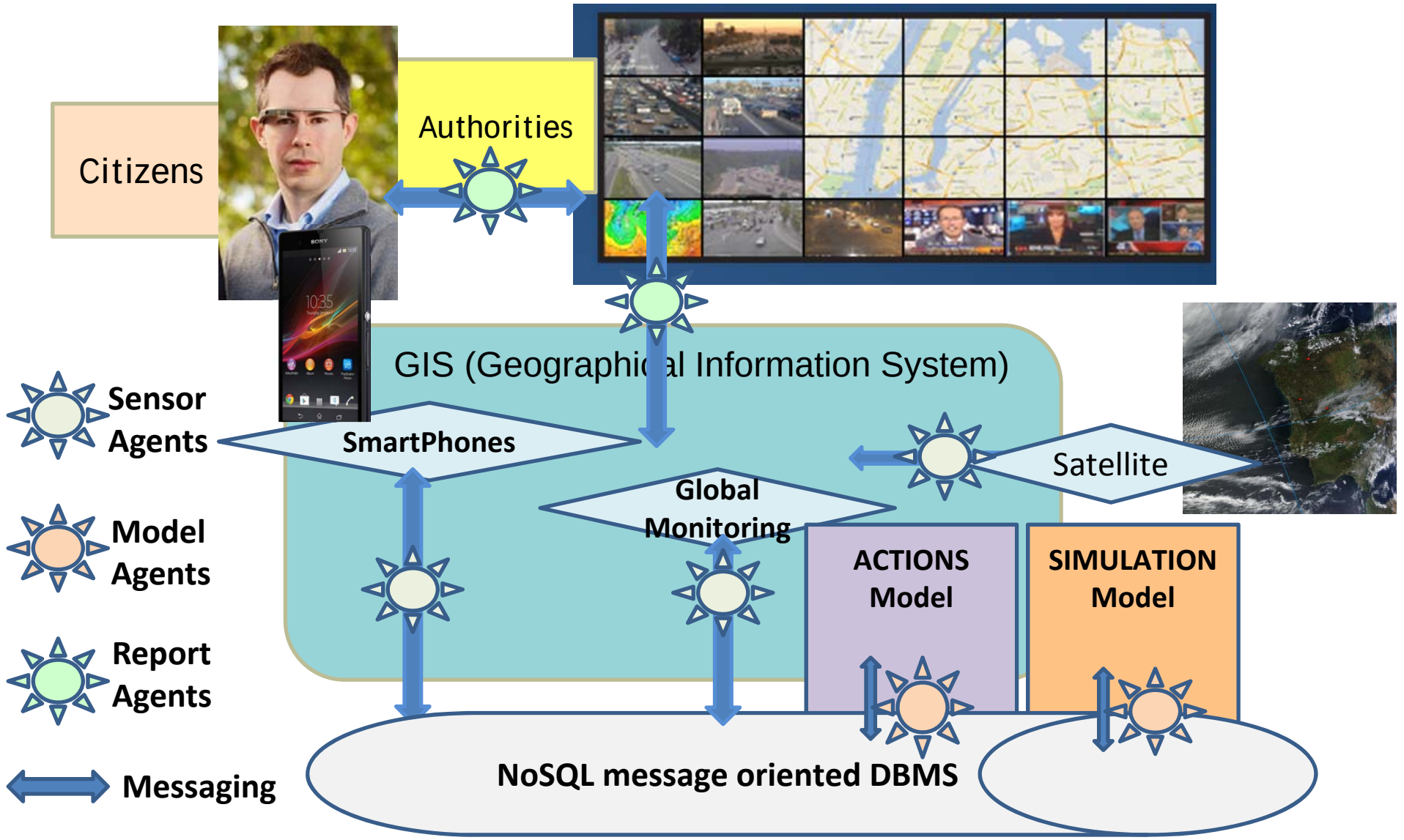
- Sistemas captura + procesado "sencillos"

(ej. Plataforma Arduino) permiten integrar imágenes o vídeo en "Internet of Things"

- Claves del procesado:
  - Preprocesado, segmentación y etiquetado
  - Identificación de patrones/variables objetivo
- Procesado (paralelizable) se puede beneficiar de clusters, supercomputadores, Grid, Cloud, etc.



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)

OK

- **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 ?