



Aplicaciones Matemáticas en Grids de e-Ciencias

“Son las Grid el Nuevo Paradigma
de la Computación”



GRIDS

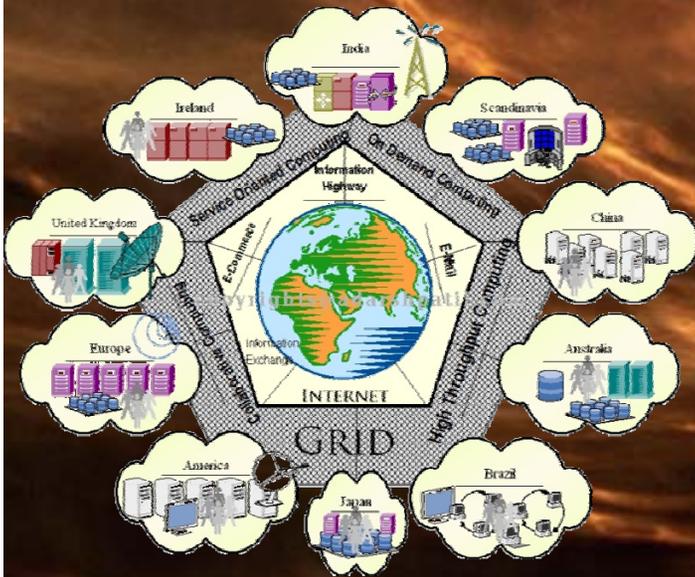
Las Grid han alcanzado un nivel de madurez tal que se están convirtiendo en la infraestructura de cómputo idónea para la investigación científica y el desarrollo de soluciones de ingeniería de alto nivel





GRIDS

- La computación distribuida ha sido diseñada para resolver problemas que exceden la capacidad de cualquier supercomputadora, mientras se mantiene la flexibilidad de trabajar en múltiples problemas más pequeños.
- En particular la computación de ciclos redundantes aprovecha las capacidades no utilizadas de las máquinas para realizar procesamiento de muy altos requerimientos.





GRIDS

El constante incremento de la capacidad del hardware y el decremento de los precios de los componentes informáticos, ha permitido que los clusters se hayan convertido en una alternativa muy atractiva en el campo de la computación paralela y distribuida. No obstante, la gran demanda tanto de computación como de espacio de almacenamiento requeridos por un diversas aplicaciones que manejan grandes cantidades de datos y deben hacerlo de forma eficiente, exige el uso de nuevas tecnologías, como es el caso de la computación *GRID*.



GRIDS



¿Qué es la e-Ciencia?

- ¿La e-Ciencia requiere del cómputo *GRID*?
- ¿En qué tecnologías se apoyan las *GRID*?
- ¿Qué es el middleware *GRID*?
- ¿Cómo es infraestructura física de la *GRID*?
- ¿Hacia dónde se dirigen las *GRID*?
- ¿Qué son los servicios *GRID*?
- ¿Qué aplicaciones hay en los servicios *GRID*?



Los Sistemas más Poderosos



- [Folding@Home](#) opera a 8.1 PetaFlops.
- [BOINC](#) promedia 1.5 PetaFlops
- [SETI@Home](#) promedia 528 TeraFlops
- [Einstein@Home](#) trabaja a 150 TeraFlops
- [GIMPS](#) a 27 TeraFlops
- Chip [multi-core POLARIS](#) logra 1 TeraFlop
- Chip de 80 cores aumenta a 2 TeraFlops a 6.26 GHz
- Procesadores quad core logran 70GigaFlops
- [GeForce 200 Series](#) 933 GigaFlops con 240 elementos de procesamiento



Proyectos GRID

Experimentos Notables

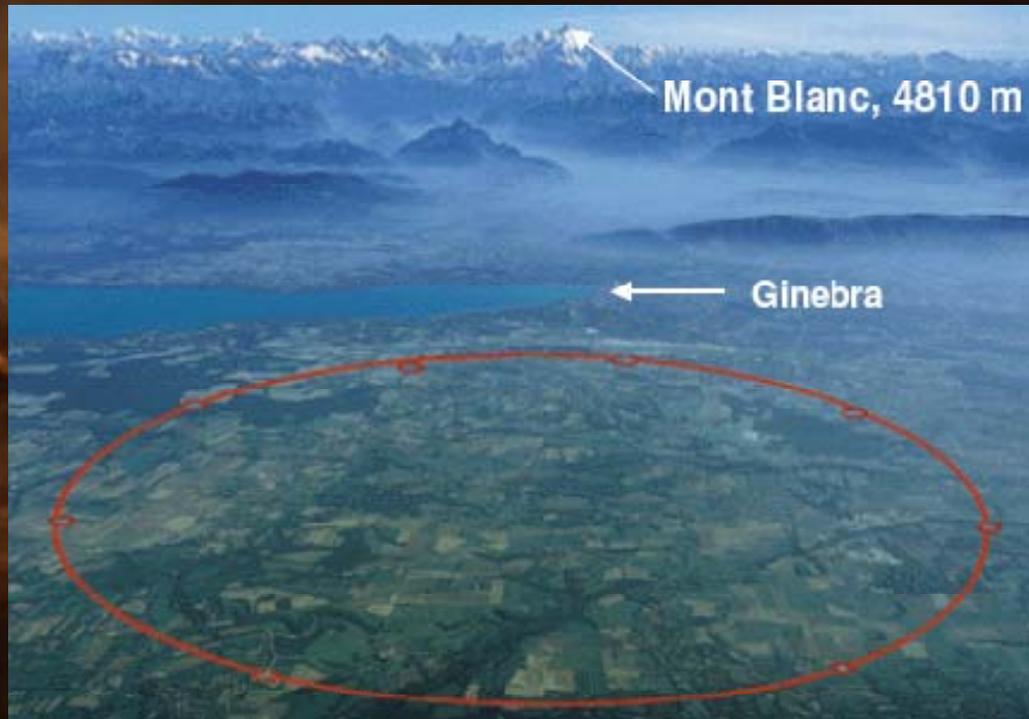


Grids Activas

Project	Start	Affiliation	Area	Peak_#hosts	Current status	Computing power
GIMPS	1996	?	mathematics	10,000	active	27 TFLOPS
distributed.net	1997	U.S. non-profit organization	cryptography	100,000	active	?
SETI@home	1999	University of California, Berkeley	SETI	362,000	active	528 TFLOPS
Electric Sheep	1999	?	art	57,000	active	?
Folding@home	2000	Stanford University	biology	406,000	active	8.1 PFLOPS
BOINC	2002	University of California, Berkeley	biomedicine, other	550,000	active	1.5 PFLOPS
Grid.org	2002	philanthropic by United Devices	biomedicine, other	3,734,000 ^[50]	closed	?
Climateprediction.net	2003	University of Oxford	climate change	150,000	active	?
LHC@home	2004	CERN	physics	60,000	active	?
World Community Grid	2004	philanthropic by IBM	biomedicine, other	700,000 ^[50]	active	?
Einstein@home	2005	LIGO	astrophysics	200,000	active	150 TFLOPS
Rosetta@home	2005	University of Washington	biology	100,000	active	?



LHC (Large Hadron Collider)



Es un acelerador/colisionador de partículas de 27km de circunferencia que fue construido por el CERN.

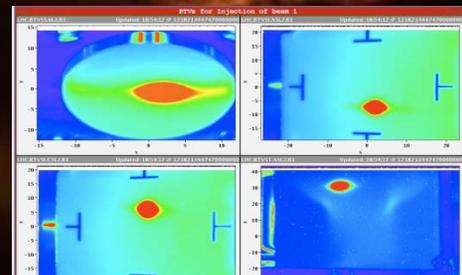
Su objetivo: encontrar el misterioso *bosón de Higgs*.

Empezó a funcionar en el 2007, momento en el cual empezó a producir cantidades desorbitadas de datos.



LHC (Large Hadron Collider)

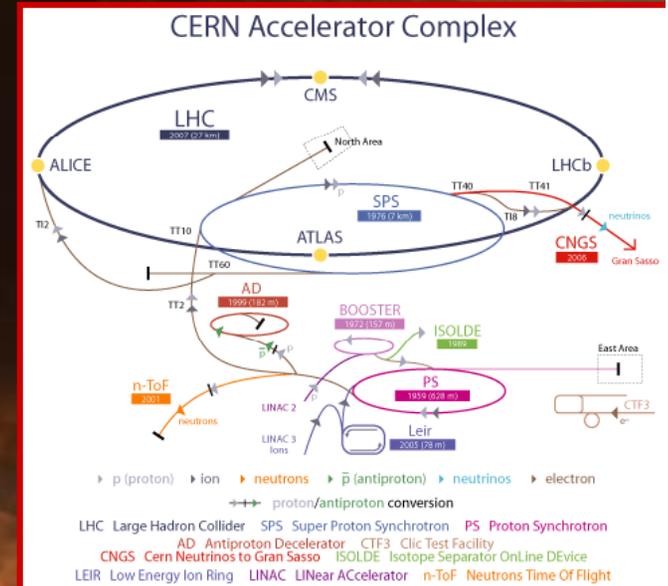
- Registra 40 millones de colisiones por segundo.
- Sólo 100 colisiones por segundo son de interés. Por ello se hace un filtrado inicial de los datos.
- Cada colisión genera 1 Megabyte de información.
- Deben ser procesados (de manera no-trivial) 100 Megabytes por segundo de datos, además de almacenarse para su estudio posterior.
- El disco duro más grande, disponible actualmente, puede almacenar 500GB aprox. 1 hora de colisiones del LHC.

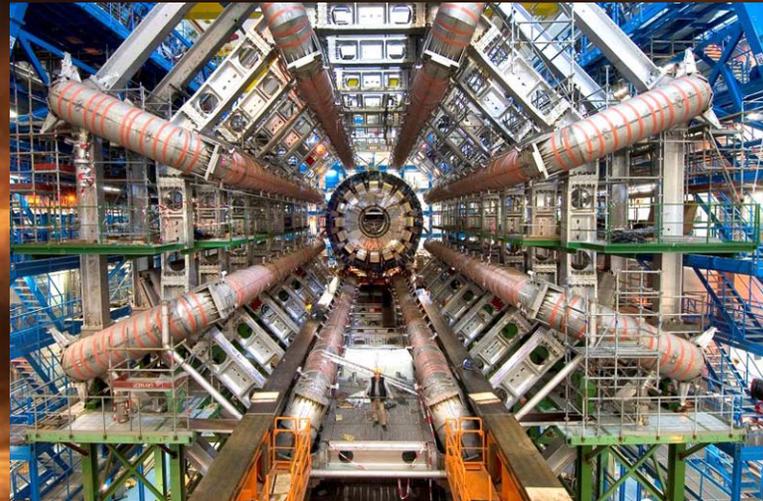
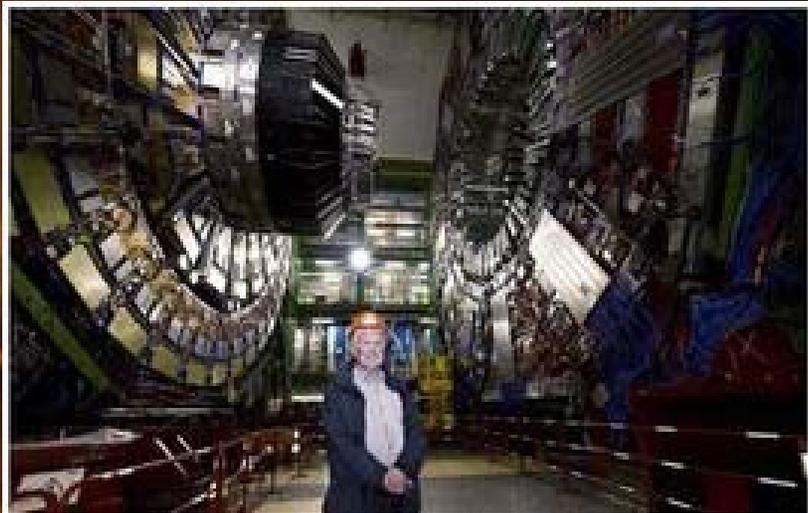
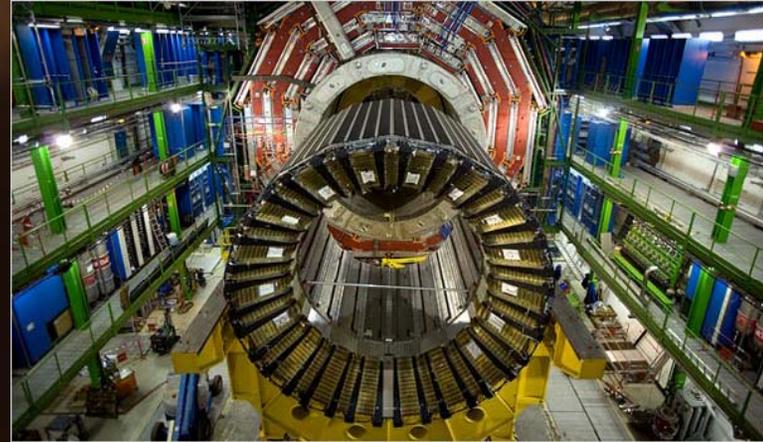




LHC (Large Hadron Collider)

- El LHC producirá 1010 colisiones por año.
- 10 Petabytes de información cada año.
- Para tener la idea:
 - 1 MB = Una foto digital
 - 1 GB = 1024 MB = Una película digital
 - 1 TB = 1024 GB = La producción anual total de libros en todo el mundo
 - 1 PB = 1024 TB = Información que produce uno de los experimentos del LHC.





Felipe Rolando Menchaca García
CIC-IPN



Berkeley Open Infrastructure for Network Computing (BOINC)

- Es una de las plataformas de cómputo grid más populares de la actualidad.
- Tiene proyectos activos en los campos de:
 - Biología
 - Ciencias de la Tierra
 - Física y Astronomía
 - Matemáticas



BOINC – Proyectos Activos

Biología

- [Cels@home](#) – Estudio de cómo se propaga el deceso en el cuerpo.
- [Malaria Control](#) — Modelado estocástico de la epidemiología de la malaria.
- [Rosetta@home](#) — Pruebas de ensamblajes de proteínas.
- [SIMAP](#) — compilación de una base de datos de similitudes de proteínas usando el algoritmo FASTA

Matemáticas

[ABC@Home](#) — Intento de solución del problema de la conjetura ABC .

[SZTAKE Desktop Grid](#) —
Búsquedas en sistemas binarios generalizados

Física, astronomía y ciencias de la tierra

- [BRaTS@Home](#) — Estudios de deflexión gravitacional.
- [Einstein@Home](#) — Detección de ondas gravitacionales.
- [LHC@home](#) — Simulación de partículas aceleradas
- [QMC@Home](#) — Geometría molecular
- [SETI@home](#) — Búsqueda de inteligencia extraterrestre.
- [Climateprediction.net](#) — Pronósticos de clima en el siglo XXI



BOINC – Proyectos en Puerta

Matemáticas

- Distributed Exact Cover Solver - resuelve problemas de Cobertura Exacta mediante el algoritmo Alpha
- [PrimeGrid](#) – Búsqueda de megaprimos
- Ramsey@Home – búsqueda de cotas inferiores para los números de Ramsey
- Rectilinear Crossing Number – Encuentra los números de cruce menores para un arreglo de puntos de cruce en un grafo.
- [Riesel Sieve](#) — Intenta resolver el problema de Riesel.
- WEP- M+2 – Investiga la factorización de primos de Mersenne.
- SHA-1 Collision Search – Busca colisiones en la función hash SHA-1
- RenderFarm@Home – Sistema distribuido de “rendering” de animaciones tridimensionales.
- BURP - Sistema distribuido de “rendering” de animaciones tridimensionales.

Física

- AQUA@home – predice el comportamiento de computadoras cuánticas adiabáticas superconductoras.
- [Leiden Classical](#) — Grid de mecánica clásica para estudiantes de ciencias.
- LHC@home alpha – Proyecto oficial de prueba del proyecto [LHC@Home](#).
- Magnetism@home – Nanotecnología de magnetización de materiales.
- [Nano-Hive@Home](#) — simula sistemas nanotecnológicos de gran escala.
- [Pirates@home](#) — Prueba del software del foro BOINC para usarlo en otro proyecto.
- RND@home – Calcula el arreglo más eficiente de radio antenas.
- [SpinHenge@Home](#) – Modela el espín cuántico de partículas elementales.
- [µFluids@Home](#) — Simula problemas de flujos de dos fases, microgravedad y microfluídica.
- [Quake-Catcher Network](#) — Usa acelerómetros conectados a computadoras para detectar terremotos
- APS@home examina los efectos de dispersión atmosférica para predicción de clima.



BOINC – “Performance”

- 1,300,000 participantes
- 2,800,000 computadoras
- 1.2 PetaFLOPS
- 12 Petabytes de espacio libre en disco
- 3.4 millones de años de tiempo de cómputo (A enero del 2008) del proyecto SETI@home



World Community Grid

- Es una plataforma de cómputo grid impulsada por la IBM.
- Tiene proyectos activos en los campos de:
 - Biología
 - Ciencias de la Tierra
 - Física
 - Matemáticas



World Community Grid - Proyectos



[Human Proteome Folding Project - Fase 2](#) — Funciones predictivas de proteínas. .

- [FightAIDS@Home](#) — Identificación de drogas candidatas con características del bloque de proteasa HIV.
- [Discovering Dengue Drugs – Together](#) — Descubrimiento de drogas para la cura del dengue hemorrágico, hepatitis C, encefalitis y fiebre amarilla.
- [Help Conquer Cancer](#) — Mejorar los resultados de la cristalografía de rayos X para entender el cancer y su tratamiento.
- [Nutritious Rice for the World](#) — Predecir estructuras de proteínas del arroz para crear plantas más nutritivas, abundantes y resistentes.
- The Clean Energy Project — Encontrar las mejores sustancias orgánicas para producir celdas solares y almacenes de energía.
- Help Fight Childhood Cancer – El proyecto de la ucha contra el cancer en niños consiste en encontrar drogas que puedan deshabilitar las tres proteínas asociadas con la neuroblastoma

- AfricanClimate@Home — Desarrollar modelos precisos del clima en regiones específicas de Africa.
- Cuboid Simulation Project — (Importante en industria, biofísica y estadística) dados de seis lados y bordes no homogéneos, probabilidades de caer en cada superficie.
- [Genome Comparison](#) — Encontrar similitudes entre proteínas de las secuencias genómicas decodificadas a la fecha.
- Help Cure Muscular Dystrophy— Fase I
- [Help Defeat Cancer](#) — Análisis de piel de pecho, cuello y cabeza.
- [Smallpox Researcht](#) —



Proyectos de Cómputo Distribuido que no hacen uso de Infraestructura de Terceros

- [GIMPS](#) — Búsqueda de primos de Mersenne muy grandes.
- [NFSNET](#) — Uso de campos numéricos de Sieve para factorizar enteros grandes
- [Seventeen or Bust](#) — Intenta encontrar número primos en secuencias de 17 para resolver el problema de Sierpinski
- [Wieferich@Home](#) — Búsqueda de nuevos primos de Wieferich
- [BREIN](#) — Usa la Web semántica y sistemas multiagente para construir sistemas grid de negocios
- [DIMES](#) — Proyecto de mapeo y evolución de la infraestructura de Internet
- [Enabling Grids for E-science](#)
- [JHDC](#) — Sistema de computación abierto de programación en Java
- [OMII-Europe](#) – Desarrollo de componentes de interoperabilidad entre plataformas grid heterogéneas.
- [OurGrid](#) — Desarrollo de tecnología grid para resolver problemas actuales

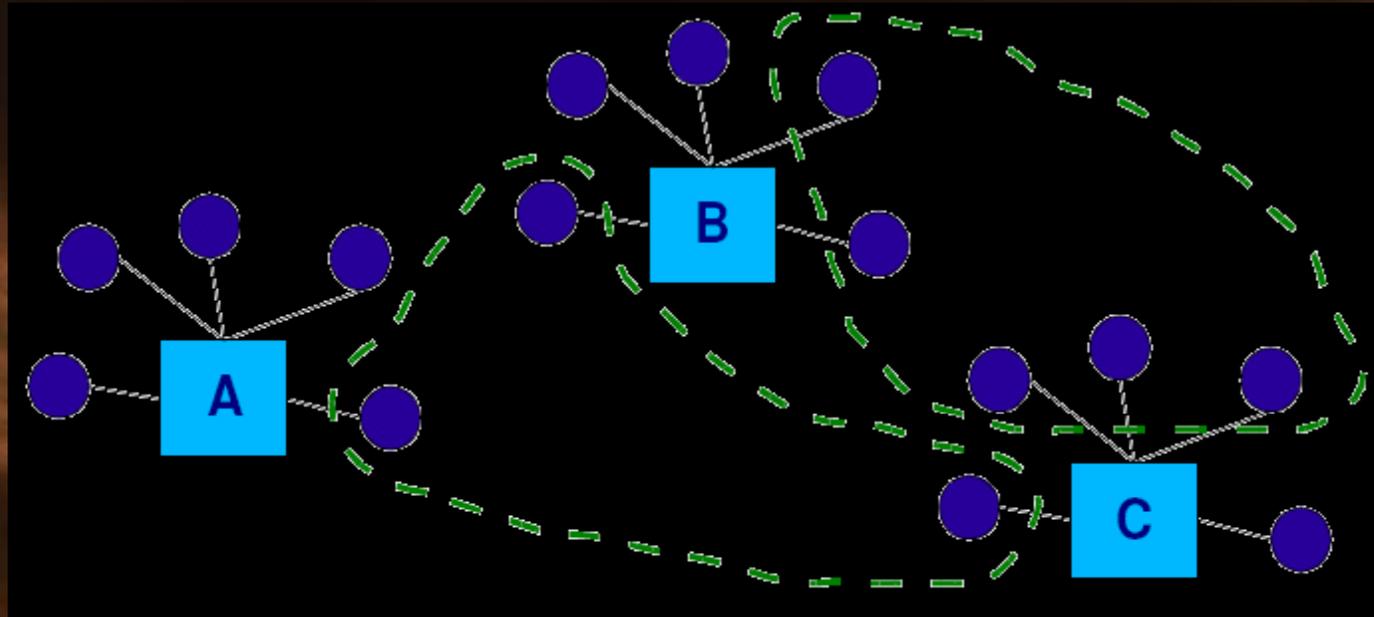


Proyectos de Cómputo Distribuido que no hacen uso de Infraestructura de Terceros

- [CommunityTSC](#) – Diseño de drogas para tratamiento de pacientes con esclerosis tuberosa.
- [D2OL](#) — Descubrimiento de drogas contra el Anthrax, Viruela, Ebola, y SARS.
- [Folding@Home](#) — Entender el proceso de desdoblamiento de proteínas.
- [Chemomentum](#) – Evaluación de químicos de riesgo.
- [Climateprediction.net](#) — Predicción de climas en el siglo XXI
- [Galaxy Zoo](#) – Clasificación de galaxias del [Sloan Digital Sky Survey](#).
- [BEinGRID](#) — Experimentos de negocios en grid
- [Gstock](#) — Búsquedas de estrategias de inversión.
- [MoneyBee](#) — Genera pronósticos de reservas.
- [BBC Climate Change Experiment](#)
- [HashClash@home](#) — Resultados experimentales de colisiones generadas con funciones hash MD5 y SHA-1
- [XtremLab](#) — Medición de recursos disponibles en PCs



Estructura de la GRID



Los recursos se agrupan dinámicamente para resolver problemas concretos, formando *organizaciones virtuales*, Paralelismo (high throughput) y/o balanceo de carga.



GRIDS Operación

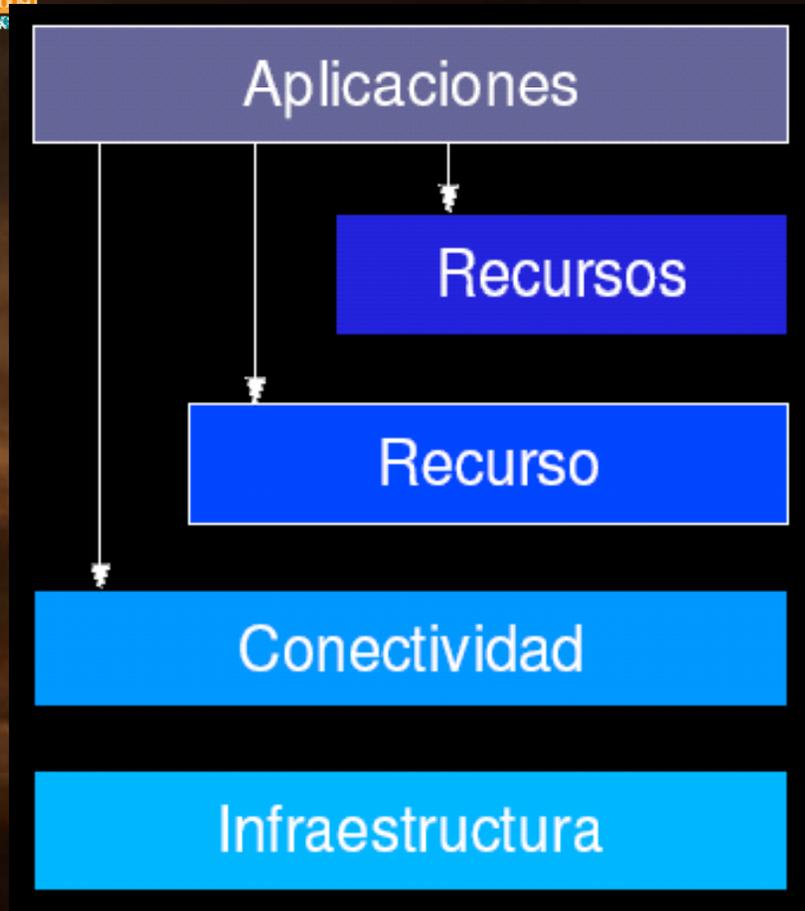


¿Cómo se decide qué recursos forman parte de la VO?

- Balanceo de carga: ¿Cómo se decide a qué problema se aplica un recurso concreto?
- No se trata de un cluster, son recursos *heterogéneos* de organizaciones *distintas*, separados por *grandes* distancias.
- Paralelismo: ¿Cómo fragmento el programa que va a ejecutarse en estas condiciones?
- ¿Cómo sabe una organización que sus recursos están siendo usados abusivamente, o de manera maliciosa?



GRIDS Operación



Gestión de N Recursos
Directorios de recursos, scheduling, monitorización, contabilidad, ...

Gestión de 1 Recurso
Información sobre el recurso
Control sobre el recurso

Comunicación y Seguridad
TCP/IP, certificados X.509, ...

Recursos computacionales
Ordenadores, clusters, supercomputadores, sistemas de almacenamiento en red, bases de datos, ...



GRIDS Operación

- Es muy sencillo decir “La interfaz de usuario (UI) envía el JOB al RB (“Resource Broker”)” pero...
- ¿Cómo se comunica la UI con el RB? ¿RPC? ¿CORBA? ¿RMI? ¿Protocolo ad - hoc?
- ¿Cómo se describe el JOB solicitado?
- ¿Cómo indico que necesito n CPU's?
- ¿Cómo indico los requerimientos de memoria? etc.
- ¿Cómo se transfieren los archivos? ¿Mediante invocación de métodos? etc.



GRIDS Operación

Al inicio de la computación GRID, cada operación se construía ad - hoc

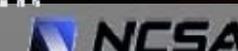
Ahora existe un gran esfuerzo de estandarización, para definir comportamientos e interfaces estándares para todos los servicios que podemos encontrar en una *GRID*:

- Gestión de recursos
 - Gestión de JOBs
 - Seguridad
- Cobro por uso de recursos
- Etc.



Ámbitos de Desarrollo de las *GRID*

OGF, GGF



Felipe Rolando Menchaca García
CIC-IPN



Open Grid Forum

El estándar unificador :

Open Grid Services Architecture (OGSA)

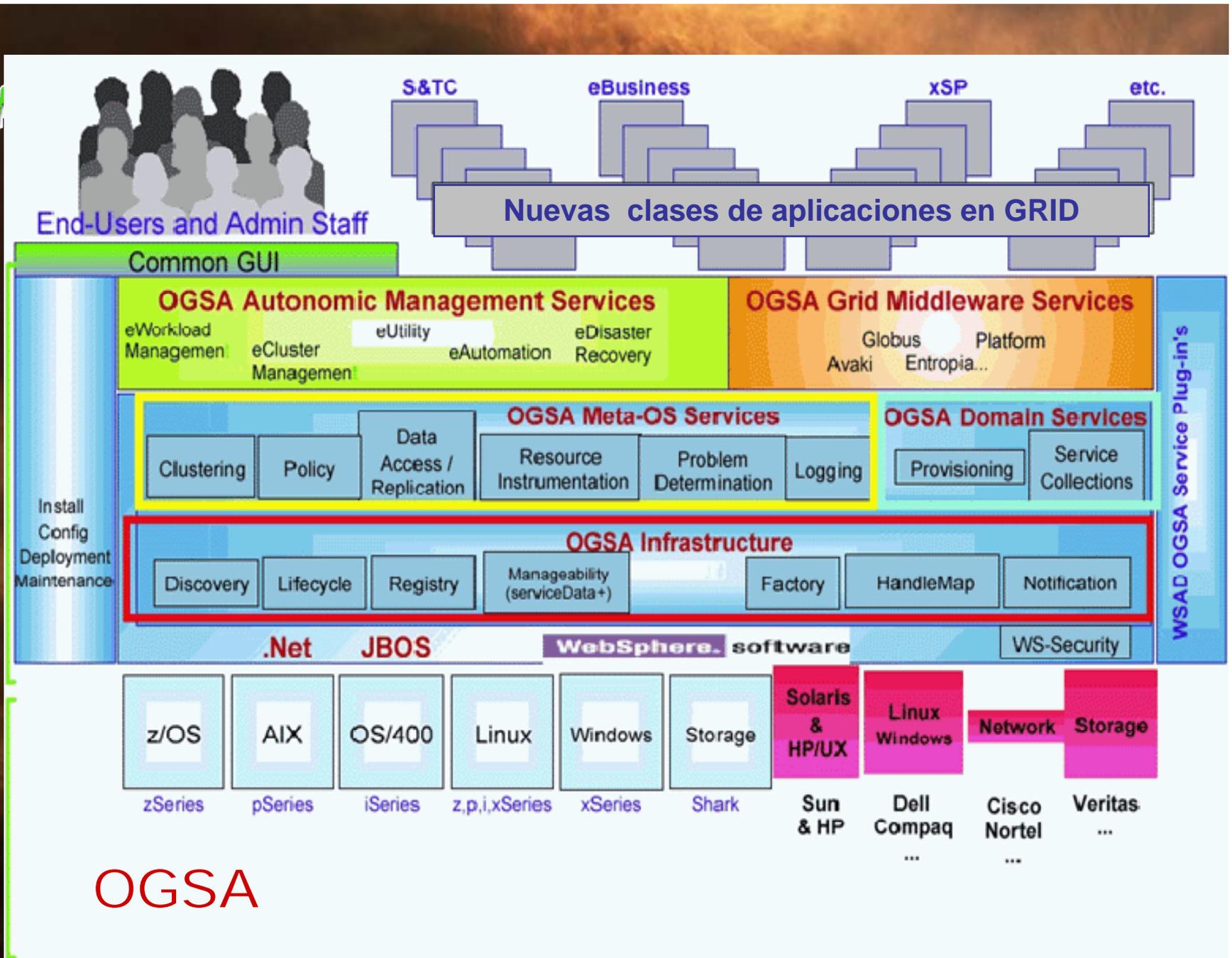
Está desarrollándose en el Global Grid Forum

(<http://www.ggf.org>)



Modelo de Servicios GRID





OGSA



Open Grid Services Infrastructure (OGSI)

En la especificación de OGSA, se decidió adoptar un concepto análogo a los *web services*, como middleware distribuido particularmente apto para sistemas débilmente acoplados.

El GGF definió así el concepto de *grid services* (web services *ampliados*)
El estándar que especifica los grid services es:
Open Grid Services Infrastructure (OGSI)



OGSI



- Introduce elementos de datos de servicio (SDEs)
- Herencia de tipos de puerto
- Manejador de servicio GRID (GSH)
- Referencia a servicio GRID (GSR)
- Notificación
- Grupos de servicio
- Introduce semánticas y capacidades a los servicios Web
 - Estados
 - Interacciones entre estados
 - Instancias transitorias
 - Administración del tiempo de vida
 - Introspección
 - Notificación de cambios de estado



Datos y Servicios Web/GRID

- Servicios Web
 - Intercambio de mensajes
 - Documentos
 - No tiene la noción de apuntador
 - ¿Es orientado a servicios?
- Servicios GRID
 - La arquitectura contempla que todo se maneje a través de interfaces y no por el envío de documentos
 - GSH es el apuntador
 - ¿Es orientado a objetos?



Globus Toolkit 3

La primera implementación estable de OGSi (y de algunas partes de OGSA) fue el **Globus Toolkit 3 (GT3)**.

Está implementado principalmente en Java.

GT3 no es *solo* una implementación de OGSi. Incluye muchos otros servicios basados en OGSi.

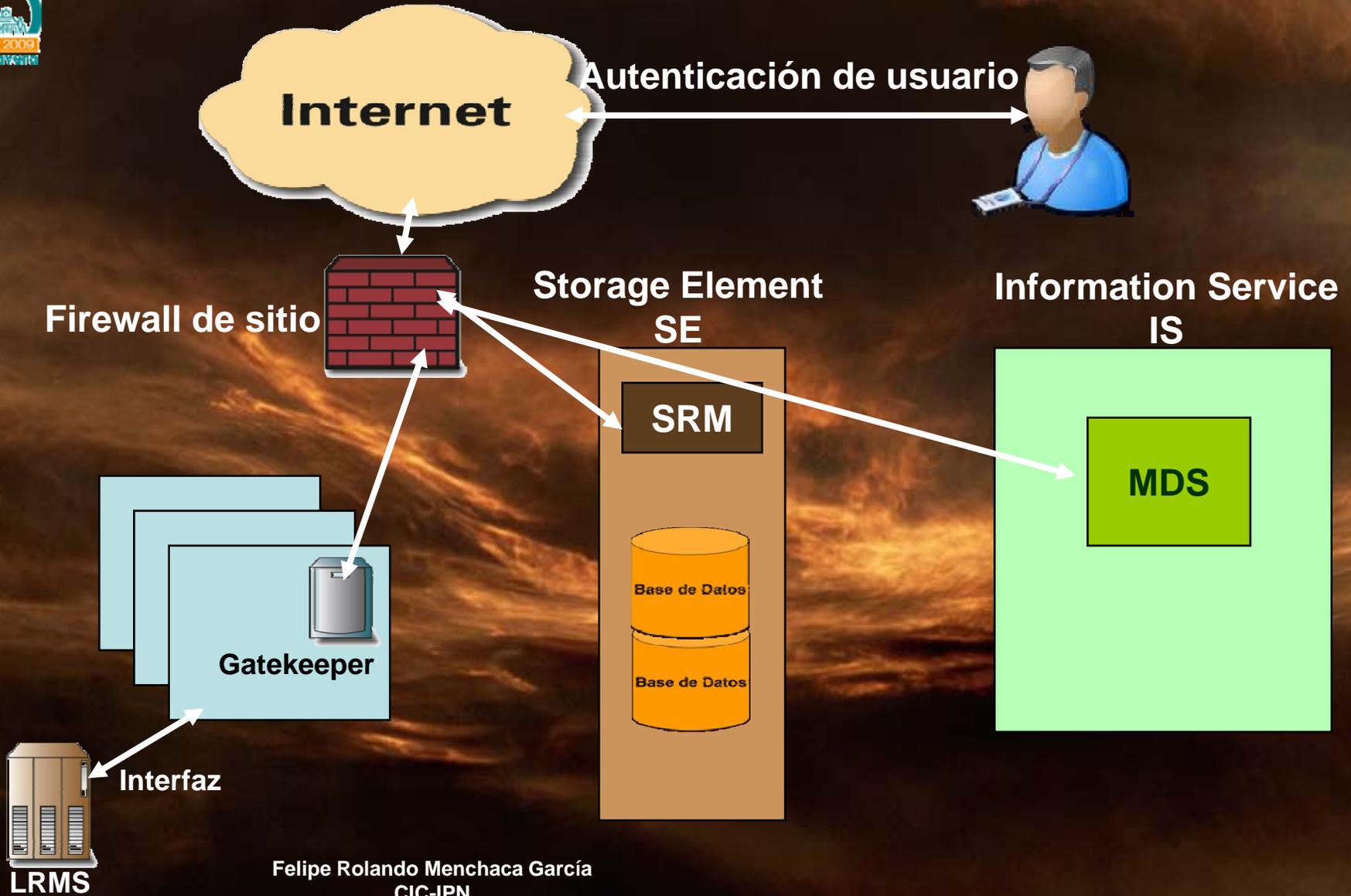
Es open source (licencia BSD - derivative)

Hoy en día ya está disponible **Globus Toolkit 4 (GT4)**.

<http://www.globus.org/>



Arquitectura Globus



Felipe Rolando Menchaca García
CIC-IPN



Proyecto EELA-2

- Es un proyecto FP7
- Proporciona facilidades GRID poderosas, con servicios versátiles que llenan los requerimientos de las aplicaciones.
- Asegura servicios de calidad
- Asegura la sustentabilidad de la e-infraestructura, más allá de la duración del proyecto
- Expande la infraestructura del proyecto EELA
- Busca nuevas comunidades de usuarios: industria, comercio.



Proyecto EELA-2



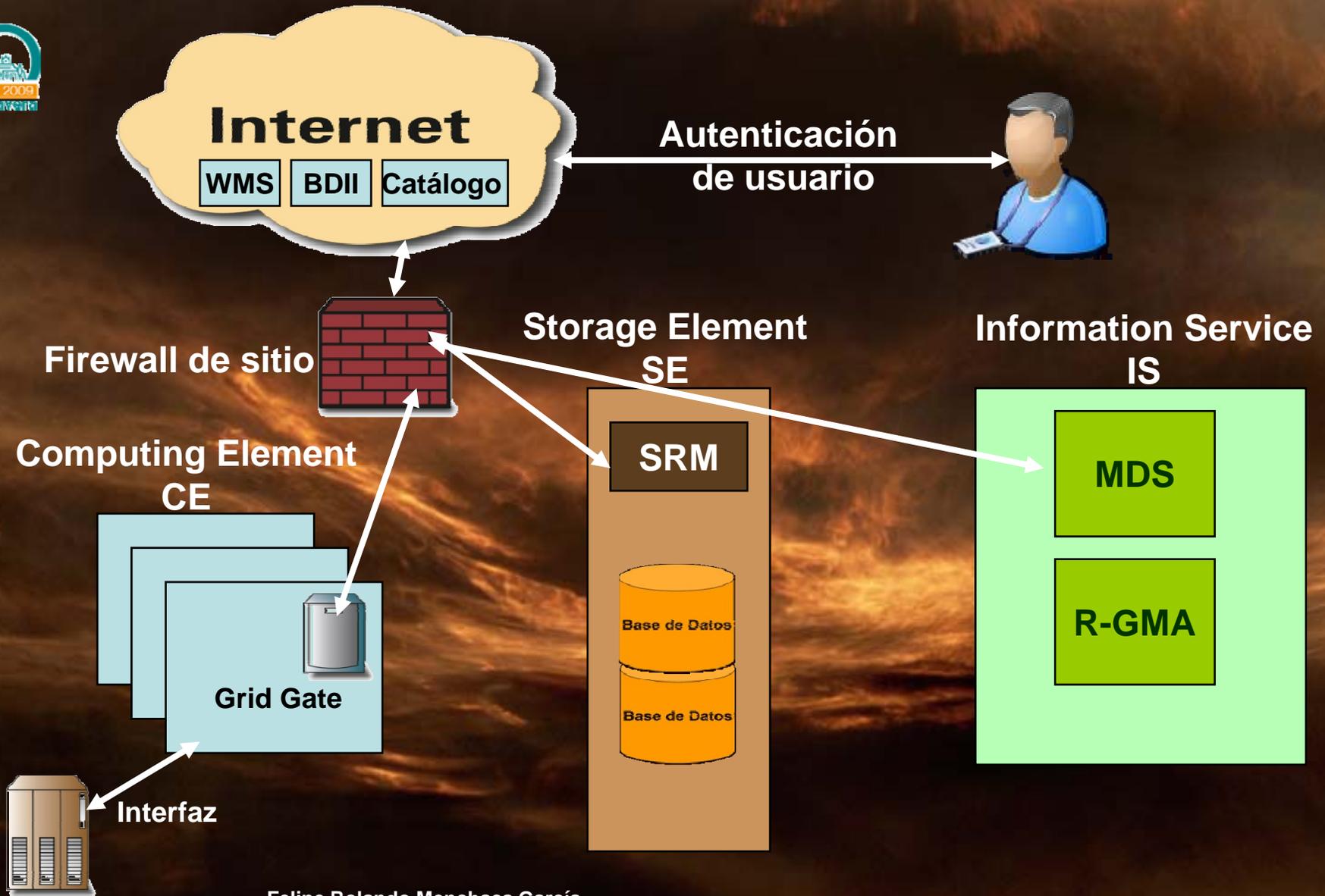


Infraestructura EELA-2

- EELA-2 está soportada por el middleware GLite
- Cuenta con un portal de aplicaciones, donde se puede tener acceso a 51 diferentes sistemas probados

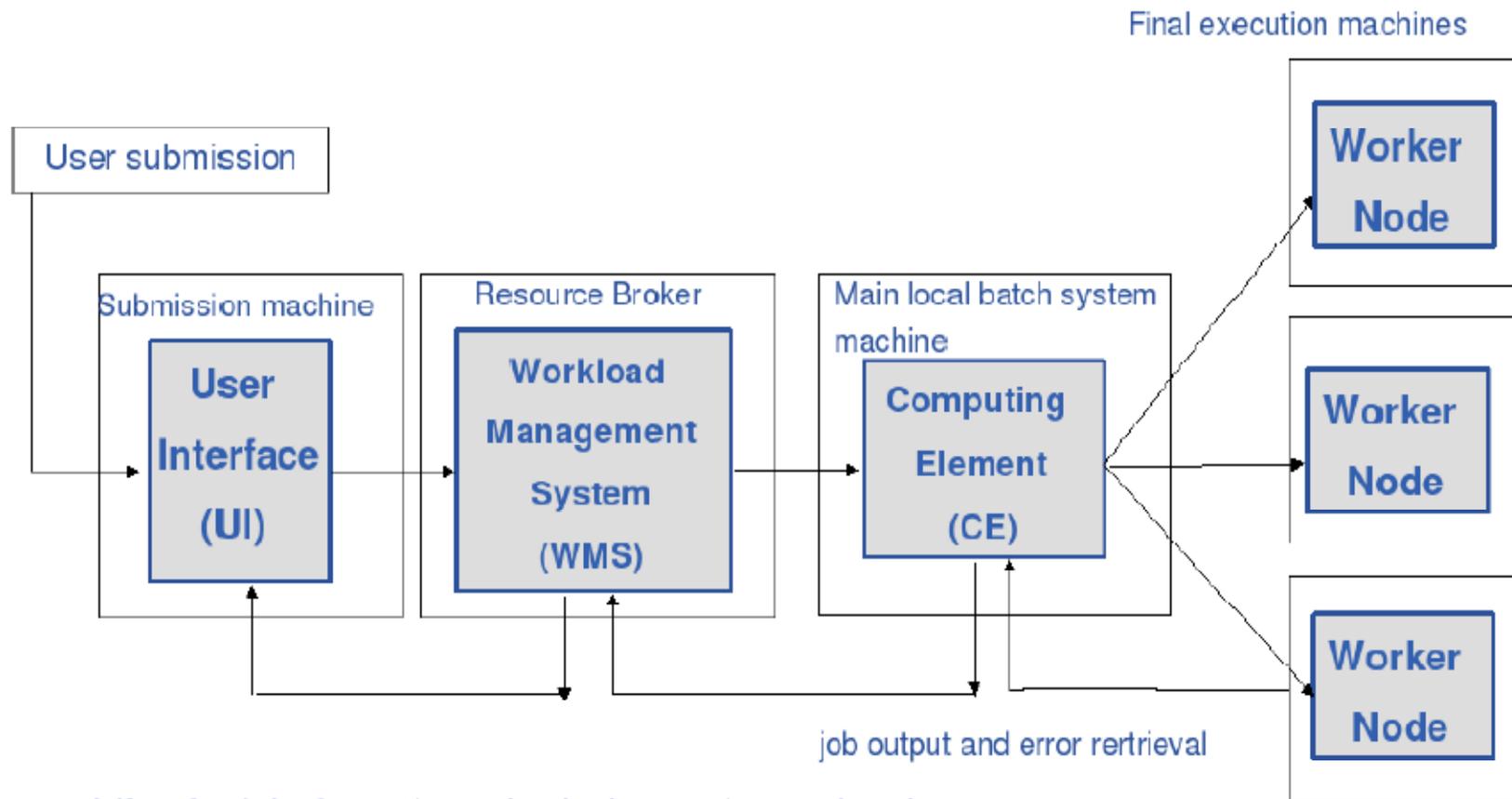


Arquitectura GLite





Arquitectura GLite



Life of a job, from the submission to the retrieval



Portal de Aplicaciones



The screenshot displays a web browser window titled "EELA-2 - Application Support" with the URL "http://applications.eu-eela.eu/". The page content includes the EELA logo and the heading "NA3 - Application Support". A sidebar on the left lists various applications, with "AERMOD" highlighted in orange. The main content area shows the "AERMOD" application details, including a general info section with a description of the model, an abstract, and a map of a region with a yellow and red pollution plume. Below the map, there are sections for "contacts" and "references".

APPLICATION

- AERMOD
- Alu
- ALICE
- AlarSeq
- ATLAS
- BIG
- BioMD
- BRAMS
- CAM-WBE
- CardoGrid Portal
- CMS-SW
- CROSS-fice
- Data_Grid_M3
- DistBlast
- DKESG
- DRACON
- eIMRT
- EM3
- qCSMT
- GenocodaGrid
- Gridpedag
- GP-GRID

GENERAL INFO

AERMOD - AERMOD (American Meteorological Service) / EPA (Environmental Protection Agency) Regulatory Model

Domain: Environmental Sciences

Status:

Abstract

AERMOD is a model for the assessment of local dispersion of atmospheric pollutants (less than 20 km from emission sources). It replaced ISC3 as the USEPA's "preferred general-purpose" model since November 9th, 2006. In Cuba it is expected to include this model in regulatory dispersion assessment and it has been already applied in the assessment of local atmospheric pollution due to power plants, generation sets and industrial sources. It also has been used in other countries in Latin America like Mexico and Argentina.

contacts

TURTOS, Lester	lturtos@subserniga.com	CUBANONIA	Key contact
SANCHEZ, Madeline	msanchez@subserniga.com	CUBANONIA	
ROSA, Henry Ricardo	hrosa@subserniga.com	CUBANONIA	
SUAREZ, Francisco	fsuarez@maat.mt.cu	CITIMAT	
SORSA, Orestis	osorsa@subserniga.com	CUBANONIA	
GARCIA, Rafael Javier	rsorcer@subserniga.com	CUBANONIA	

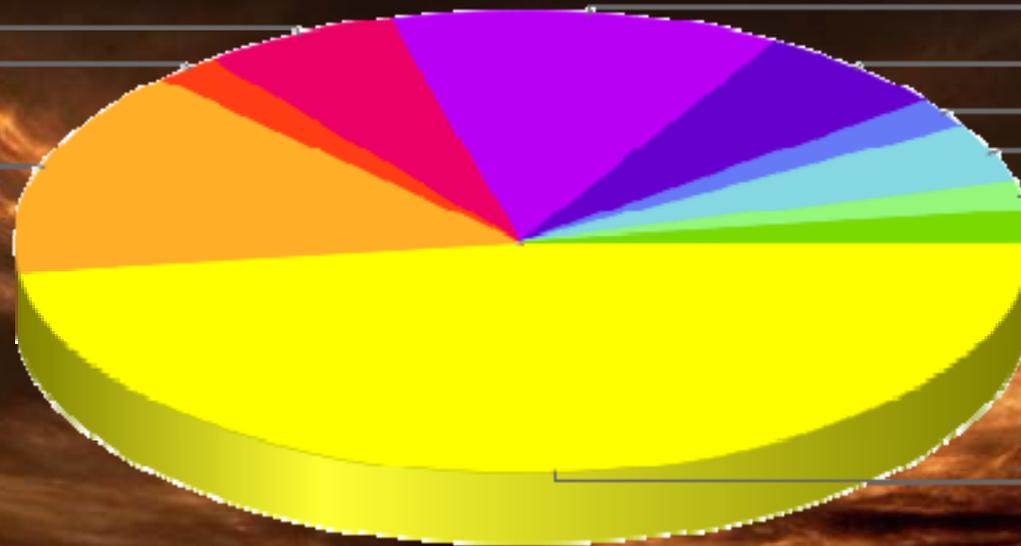
REFERENCES

- e-Learning (MAAT) No
- At / Optimization (CIC) Yes



Aplicaciones EELA-2

AI / Optimization
Civil Protection
HEP



Earth sciences

Engineering

Environmental

Data mining

Physics

e-Learning

Biomedicine (23)

<http://applications.eu-eela.eu/>



Aplicaciones EELA-2

APPLICATION	DOMAIN	INSTITUTION (s)
<u>Aïuri</u>	AI / Optimization	(UFRJ)
<u>Cinefilia</u>	AI / Optimization	(INFN)
<u>gRREEMM</u>	Engineering	(CUBAENERGIA)
<u>gSATyrus</u>	AI / Optimization	(UFRJ)
<u>Industry@Grid</u>	Engineering	(CEFET-RJ)
<u>LEMDistFE</u>	Engineering	(UNAM)
<u>PILP</u>	AI / Optimization	(UP)
<u>Portal de Pórticos</u>	Engineering	(ULA)



Otras Plataformas GRID

- [Berkeley NOW Project](#)
- [Institut Ruđer Bošković \(IRB\) Debian Cluster Components](#)
- [Open Science Grid](#)
- [SARA Computing and Networking Services in Netherlands](#)
- [Teragrid](#)
- [VirginiaTech](#)



Aplicaciones Matemáticas

- [gridMathematica](#)
 - Incluye aplicaciones a bioinformática, análisis y procesamiento de grandes volúmenes de datos, minería de datos, física, matemáticas y ciencias de la vida.
 - Incluye comandos de alto nivel para ejecución paralela de operaciones con matrices y otras funciones



Wolfram
Lightweight Grid Manager





Aplicaciones Matemáticas

- Maplesoft GRID Computing Toolbox
 - Se utiliza para:
 - Diseño de ingeniería
 - Investigación de operaciones
 - Análisis financiero
 - Modelado y simulación
 - Computación simbólica
 - Funciones holomórficas en espacios n-dimensionales
 - Álgebra lineal





Otros Recursos GRID

- **Simuladores *GRID***
 - Bricks
 - SimGrid
 - GridSim
 - GangSim
 - OptorSim
- **Emuladores *GRID***
 - MicroGrid
 - Grid eXplorer



INFOGRAFÍA

GridCafé

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

Libros “The Grid” y “The Grid 2”

Editados por Ian Foster y Carl Kesselman, 'padres' de la Grid

The Globus Toolkit 3 Programmer's Tutorial

<http://www.casa-sotomayor.net/gt3-tutorial/>



INFOGRAFÍA

- Maple Profesional Toolbox.
<http://www.maplesoft.com/products/toolboxes/GridComputing/index.aspx#>
- Wolram Research.
WolframgrdMathematica7.
<http://www.wolfram.com/products/gridmathematica/>



INFOGRAFÍA

- [^ \http://www.mactech.com/news/?p=1011816 Wolfram Research introduces gridMathematica 7] MacTech
- [^ Felix Grant tests out some of Wolfram's spin-off products, which aim to push the company into new markets](#), Scientific Computing World, June 2003.
- [^ gridMathematica offers parallel computing solution](#), Dennis Sellers, MacWorld, November 20, 2002
- [^ Proceedings of the 2006 ACM/IEEE conference on Supercomputing](#) Roman Maeder
- [^ Wolfram Research Announces the release of gridMathematica](#) GRIDToday Dec 2002
- [^ gridMathematica 2.2](#) Scientific Computing World September 2007
- [^ GridMathematika / Benchmark Tests](#) by M.Karbalai, H.Schönau, 9/14/2006.
- [^ Mathematica 7: A Review](#) Mike Riley, Dr Dobb's
- [^ gridMathematica 7 enhances parallel computing](#) MacNN



INFOGRAFÍA

- [An interview with Schoeller Porter, Senior Software Engineer at Wolfram Research](#) by Ken Farmer, WinHPC.org, Monday October 30 2006.
- [Exploration of the capabilities of gridMathematica on the Altix ia64 HPC machines](#), [University of Queensland](#) VisLab.
- [Special Projects: An Evaluation of Modified Newtonian Dynamics by Simulation in a Parallel Computing Environment](#), [Furman University](#) Physics Department.
- [Getting on the Grid](#), by Jean Thilmany, Mechanical Engineering, 2003.
- [a real-time monitor of gridMathematica's usage on the TSUBAME supercomputer in Japan](#)
- [online recorded seminar](#)
- [King's College London Centre for Financial Grid Computing](#)