

ALICE GRID en México

Primer nodo de ALICE en América Latina

21 de junio de 2007

Luciano Díaz

Lukas Nellen, Pedro Podesta, Patricia Méndez, Kitzia
Lubreras, Hans Reyes, Latchezar Betev, Pedro
Rausch

Introducción

ALICE, al igual que los otros tres experimentos (ATLAS, CMS y LHCb) del acelerador LHC, ejemplifica perfectamente la aplicación de la tecnología Grid.

El experimento ALICE tiene las siguientes características:

- Generará una enorme cantidad de datos.
- Requerirá una gran cantidad de recursos de procesamiento y almacenamiento.
- Gran consumo de energía por el equipo de cómputo.

Dada la gran cantidad de recursos que necesita, estos no pueden concentrarse en un solo lugar.

Estimación de recursos para ALICE

1 MSI2K = 10^6 SPECint2000

La máquina de referencia de los SPEC CPU2000 es la UltraSPARC10 con un procesador UltraSPARC Iii, a 300 MHz y 256 MB de memoria.

1 Xeon 3Ghz con 4GB de memoria tiene aproximadamente un rendimiento de 1.5KSI2K

Podemos estimar que 1MSI2K son aproximadamente 700 procesadores Xeon

1PB = 10^{15} Bytes

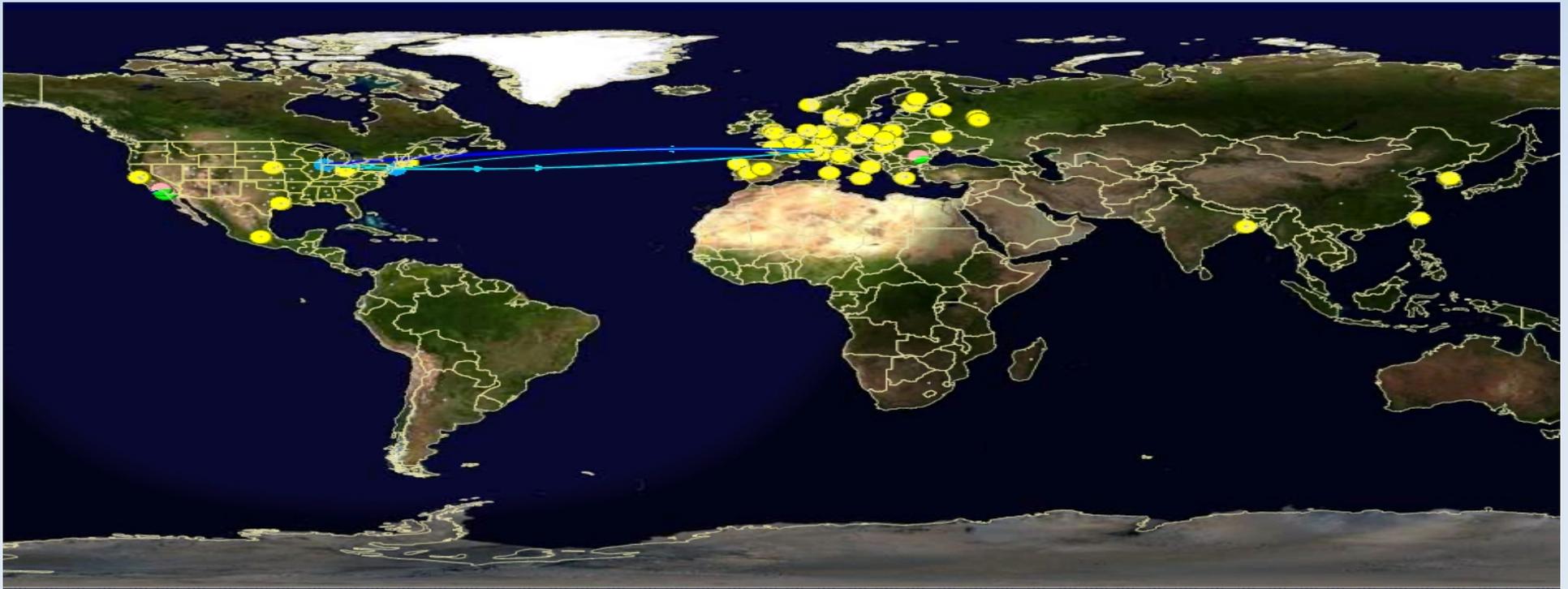
New 2007	CERN				External			Total
	Tier0	CAF	Tier1	Total	Tier1s	Tier2s	Total	
CPU(MSI2K)	0.053	0.026	0.90	0.90	3.63	5.79	9.41	10.3
DISK(PB)	0.014	0.051	0.26	0.32	0.86	0.77	1.63	1.95
MS (PB)	0.066	-	0.63	0.70	1.66	-	1.66	2.36

New 2008	CERN				External			Total
	Tier0	CAF	Tier1	Total	Tier1s	Tier2s	Total	
CPU(MSI2K)	3.31	1.04	2.90	3.58	11.43	12.87	24.30	27.87
DISK(PB)	0.095	0.11	1.54	1.74	3.34	1.62	4.96	6.70
MS (PB)	1.06	-	1.79	2.85	6.41	-	6.41	9.26

New 2009	CERN				External			Total
	Tier0	CAF	Tier1	Total	Tier1s	Tier2s	Total	
CPU(MSI2K)	8.29	2.61	4.60	8.94	18.88	19.96	38.84	47.78
DISK(PB)	0.24	0.27	2.87	3.38	6.29	4.05	10.34	13.72
MS (PB)	3.53	-	2.92	6.45	13.39	-	13.39	19.84

GRID geográficamente distribuida

Por tal motivo se acordó implementar una Grid geográficamente distribuida.



Los recursos de cómputo y almacenamiento se encuentran agrupados en sitios (SITES) en los diferentes países e interconectados por redes de alta velocidad.

Conceptos

Para hablar del cómputo en el experimento ALICE es necesario definir cuatro conceptos muy importantes:

- Grid
- Middleware
- La aplicación
- Los usuarios

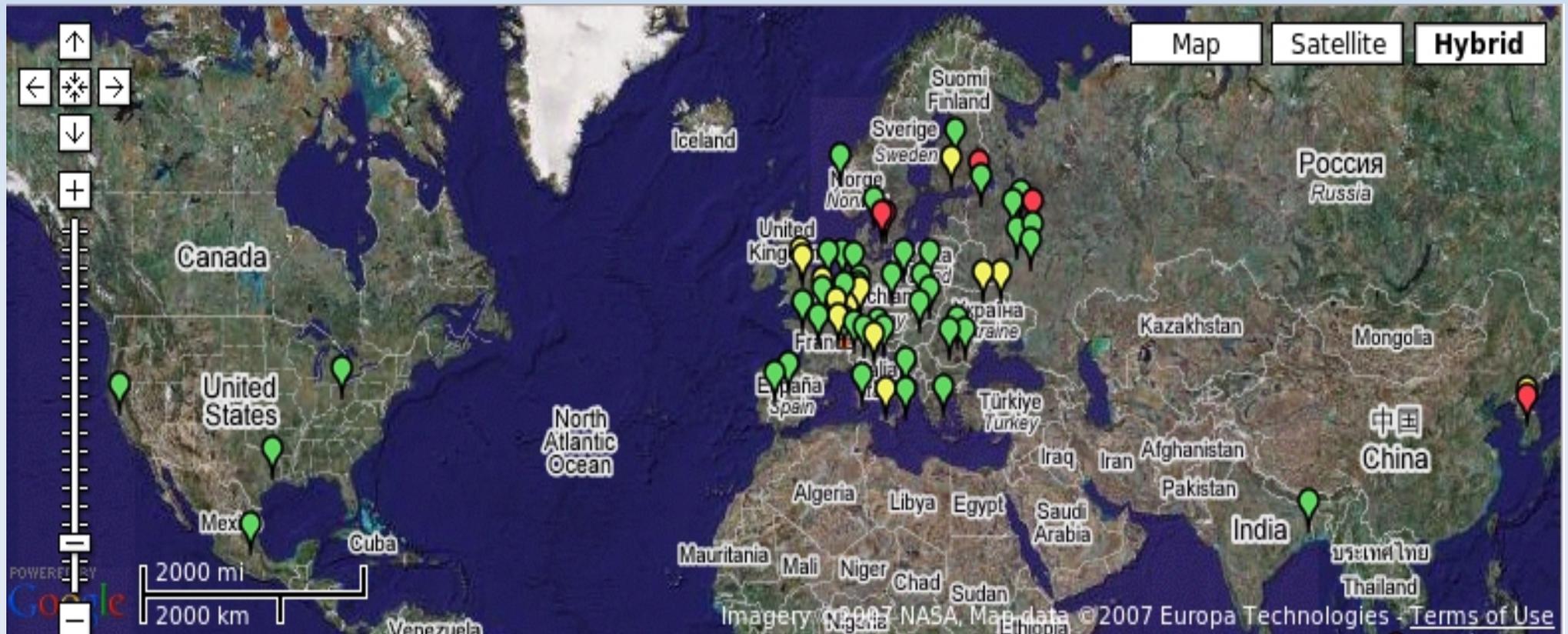
Conceptos- La tecnología

- **GRID**

Es una tecnología que permite utilizar de forma coordinada y segura recursos de procesamiento, almacenamiento y aplicaciones, que no están sujetos a un control centralizado. Los recursos pertenecen y son administrados por las instituciones que forman la infraestructura de Grid.

El software que permite realizar la integración e interacción de los recursos es el middleware.

GRID de ALICE



- Grid de ALICE

Conceptos-El software

- **Middleware**

Es un software de conectividad que ofrece un conjunto de servicios que hacen posible el funcionamiento de aplicaciones distribuidas sobre plataformas heterogeneas.

Crea una capa de software que se situa entre la capa de aplicación y las capas inferiores (Sistema operativo y de red).

Middleware para ALICE

El middleware para ALICE es gLite y está desarrollado por el EGEE; usa componentes de varios proyectos Grid como:

- DataGrid (EDG)
- DataTag (EDT)
- INFN-GRID
- Globus
- Condor

En algunos casos los componentes son modificados y el software final no es exactamente igual al distribuido por el proyecto original.

Tabla de componentes

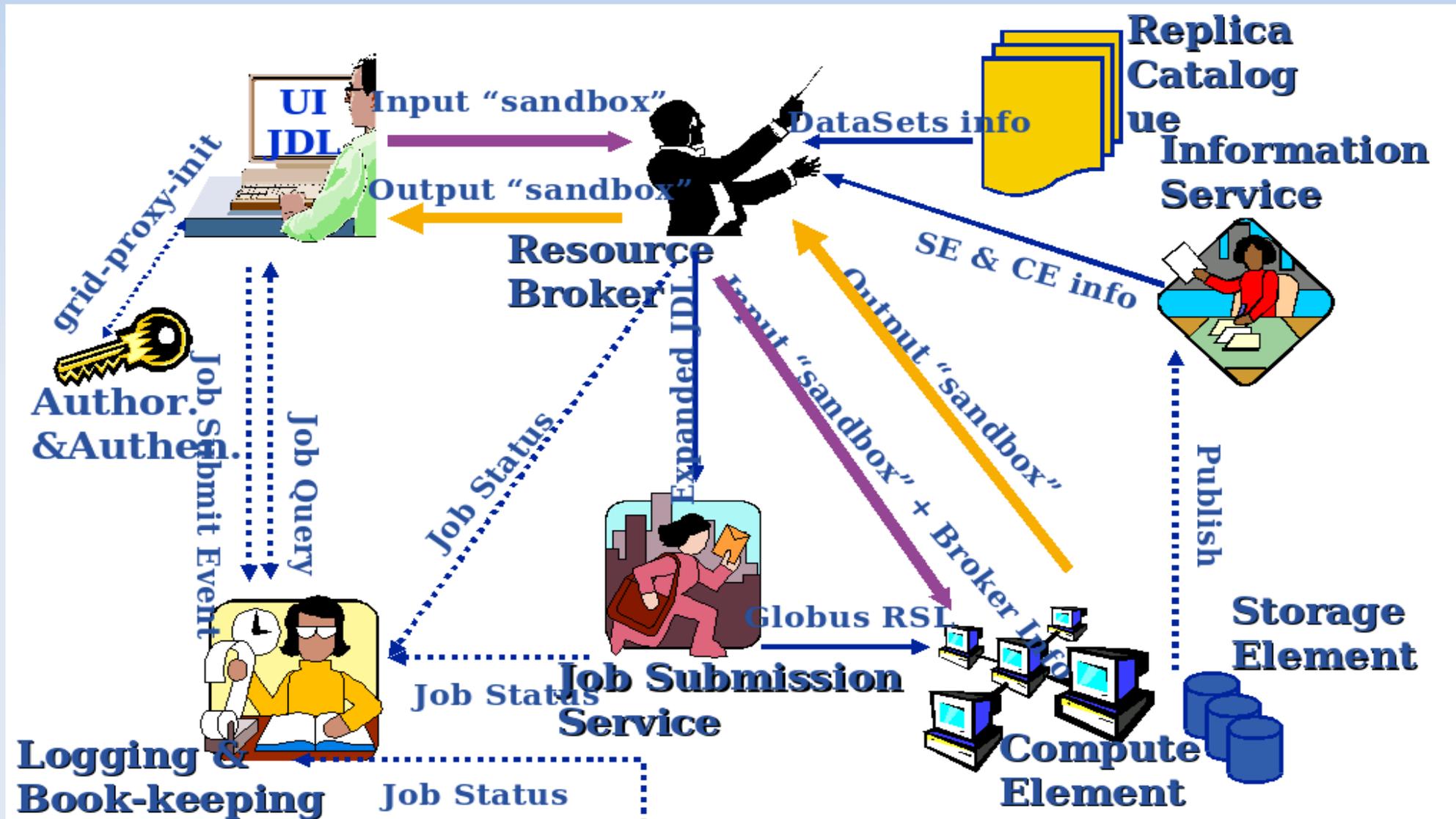
Component	EGEE	EDG	EDT	INFN Grid	Globus	Condor	Other
Basic middleware							
Globus Toolkit 2.4.3					✓		
ClassAds 0.9.7						✓	
Authentication and Authorisation							
MyProxy 0.6.1							✓
VOMS	✓	✓					
VOMRS							✓
LCAS/LCMAPS		✓					
Workload management							
Condor-G 6.6.7						✓	
EDG WMS		✓					
gLite WMS	✓						
Data management							
LFC	✓						
DPM	✓						
FTS	✓						
GFAL	✓						
LCG DM tools	✓						
Fabric management							
Quattor							✓
YAIM	✓						
Monitoring							
GridICE				✓			
Information system							
MDS					✓		
Glue Schema			✓				
BDII	✓						
R-GMA	✓	✓					
LCG Information tools	✓						

Servicios de GRID de gLite

Los principales servicios Grid que se crean a partir de los componentes de software son:

- **CE** Computer Element
- **WN** Worker node
- **SE** Storage Element
- **LFC** Local File Catalog
- **RB/WMS-LB** Resource Broker
- **VOBOX** Virtual Organization Box
- **BDII** Berkeley Database Information Index (Information System)
- **MON-BOX** Monitor Box
- **UI** User Interface

Servicios Grid



- Conjunto de servicios Grid

Conceptos- La aplicación

- **AliEn (ALICE-Environment)**

Es el marco de trabajo desarrollado por la colaboración de ALICE basado en componentes “Open Source” bajo el modelo de servicios Web. Es ocupado para la simulación, reconstrucción y análisis de datos en un ambiente distribuido.

La version actual de AliEn (v2-13) tiene mayor tolerancia a problemas de red como la latencia o reducido ancho de banda.

Problemas que presentan sitios ubicados en México, India y Corea, y cuyo desempeño mejoró con esta última versión.

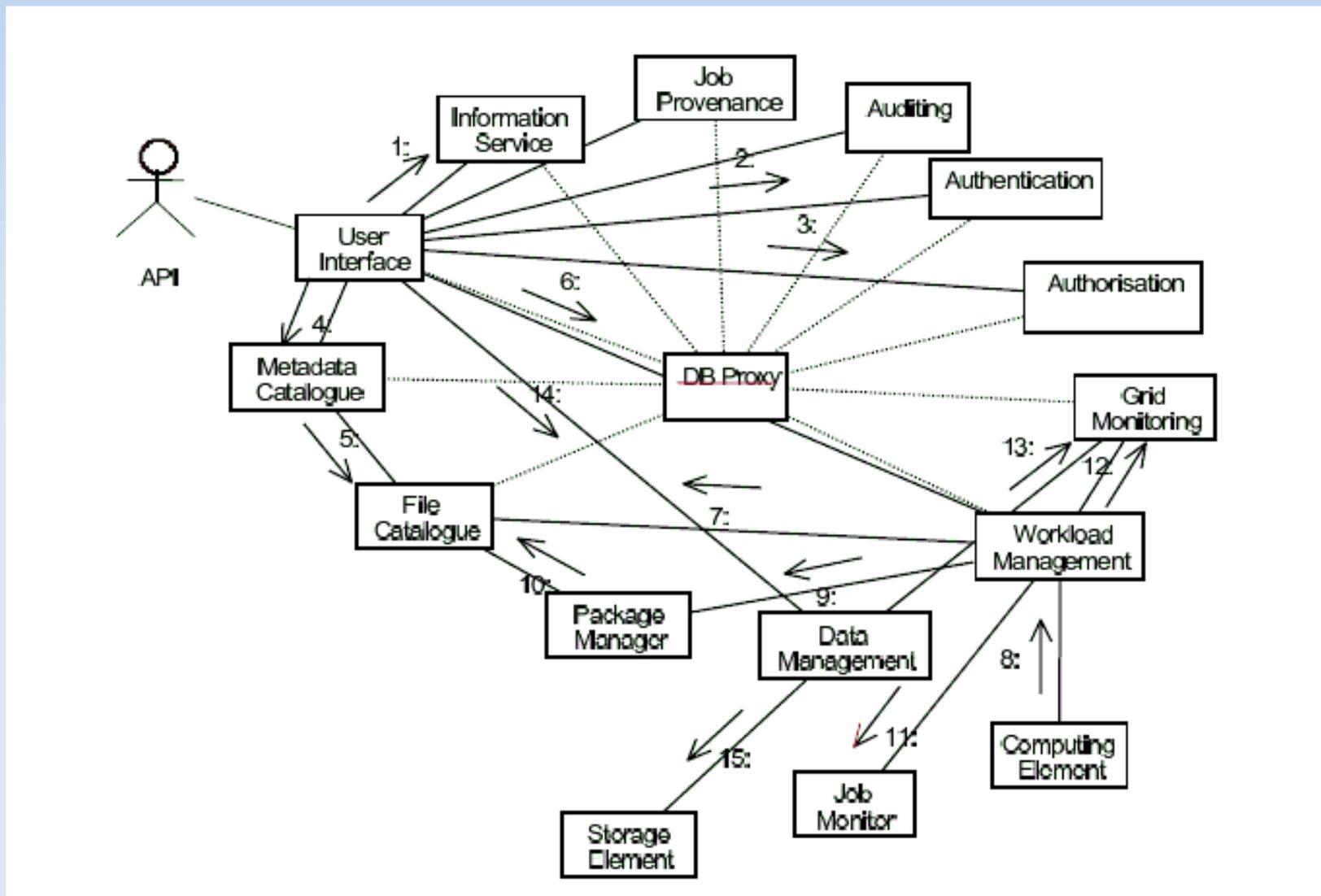


Diagrama de interacción entre componentes de AliEn, análisis de caso de uso.

Conceptos-Usuarios

El manejo de usuarios en la infraestructura de Grid se hace a través de **Organizaciones Virtuales (VOs)**.

Una VO es una entidad abstracta que agrupa usuarios, institutos y recursos en un mismo dominio administrativo.

A cada experimento del LHC le corresponde una VO.
En el caso de ALICE la VO correspondiente es:ALICE

Instalación de un SITIO

Para realizar la instalación básica de un sitio Grid sólo hace falta instalar 2 tipos de servicios:

- UN CE (Computer Element) y al menos un WN (Worker Node).

Para ALICE la instalación de AliEn se realiza sobre un VOBOX-UI.

Requerimientos

- Todos los servicios se instalan sobre Scientific Linux 3.0.x
- Solicitar un certificado personal y registrarlo en las VOs. Se recomienda un certificado por VO.
- Cada servidor necesita una IP homologada y su nombre registrado en el DNS.
- Se recomienda 1GB de memoria.
- Algunos servicios corren desde un PIII a 800 Mhz.
- Cada servidor debe contar con su certificado.

Requerimientos para ALICE

Adicionalmente para ALICE:

- El administrador del VOBOX, además de pertenecer al VO de ALICE, debe registrarse como usuario de AliEn y recibir permisos de administrador de sitio.
- Se debe registrar el certificado del VOBOX en los servicios centrales de ALICE.
- Instalar AliEn en el VOBOX y exportar el software en todos los WNs.

Instalación del sitio en el ICN

Como se mencionó, los recursos se agrupan por sitios y un sitio puede pertenecer a más de una VO.

Los servicios de Grid que se instalaron en el ICN comparten los recursos para dos VOs:

- ALICE
- EELA (E-Infraestructure shared between Europe and Latin America)

SITIO de ALICE en México

Se instaló primero un:

CE (Computer Element) y 14 WN's (Worker Nodes)

Por las condiciones geográficas y el rendimiento de la red fue necesario instalar otros servicios:

- SE (Storage Element)
- LFC (Local File Catalog)
- RB (Resource Broker)

Para ALICE específicamente se instaló:

- VOBOX-UI

Adicionalmente se instaló el servicio de monitoreo:

- Mon-BOX.

Configuración final

Así el SITIO queda configurado de la siguiente manera:

Nombre del sitio: **EELA-UNAM**

Equipo: Cluster **Tochtli** en ICN

Servicios:

- **CE**:tochtli.nucleares.unam.mx con 14 WN's
- **SE**:tlapiacalli.nucleares.unam.mx
- **RB**:cuauhtli.nucleares.unam.mx
- **MONBOX**:nauyaca.nucleares.unam.mx
- **VOBOX-UI**:calpolli.nucleares.unam.mx (sólo para ALICE)
- **LFC**:lfc.unam.mx (Instalado en DGSCA)

Instalación de AliEn

Un vez instalado AliEn en el VOBOX este crea los siguientes servicios que encapsulan los servicios del middleware:

- AliEn CE
- AliEn SE
- AliEn ClusterMonitor
- AliEn PackMan
- AliEn Monalisa

El VOBOX debe ser probado y certificado por los expertos del CERN antes de integrarse a esquema de producción.

Los jobs de producción son enviados por el sistema de colas central administrada desde el CERN.

Funcionamiento de Tochtli

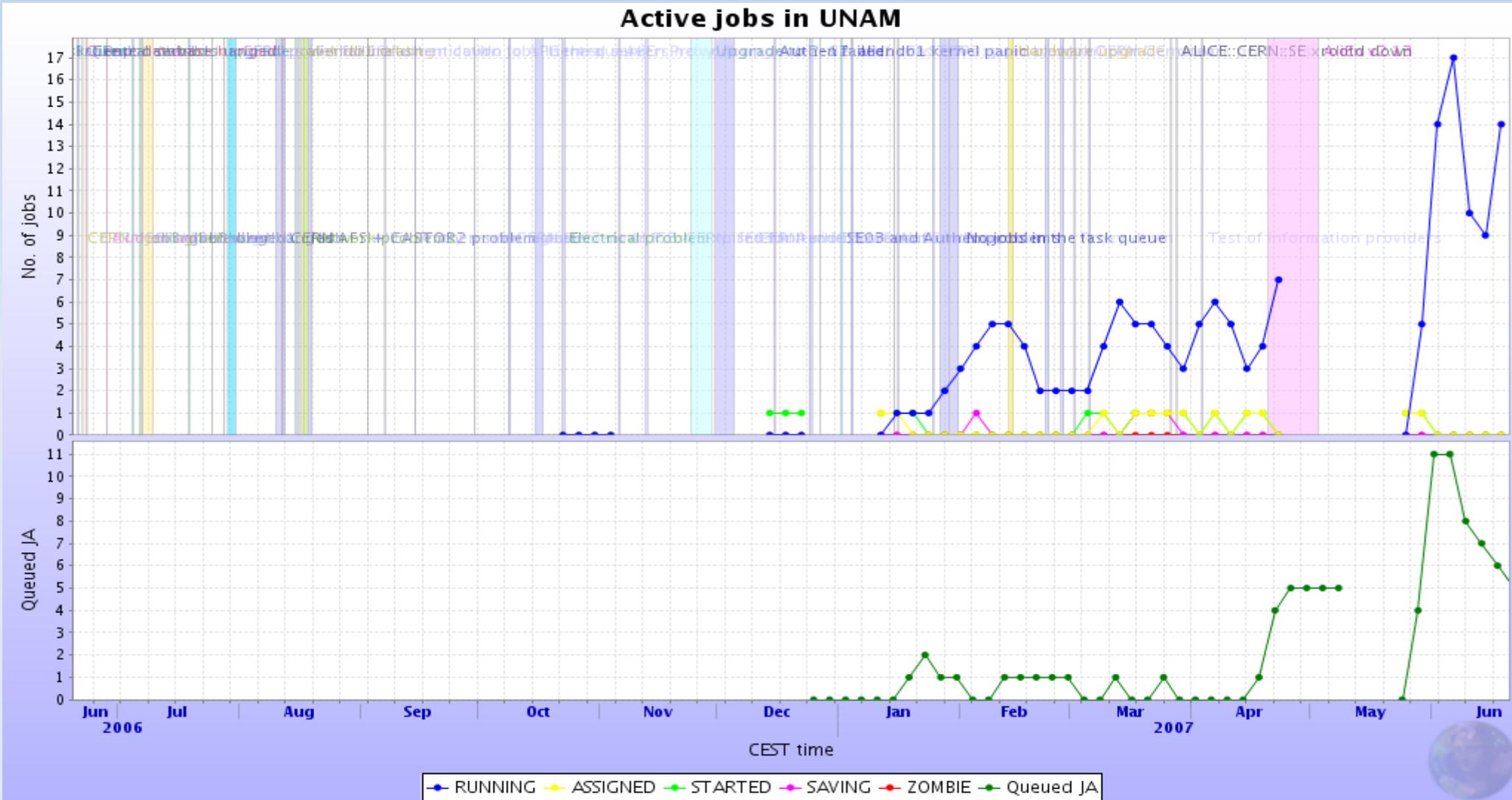
Inicio de producción: **febrero de 2007**

Capacidad

Procesamiento: **28 jobs, promedio 20**

Almacenamiento : **1TB**

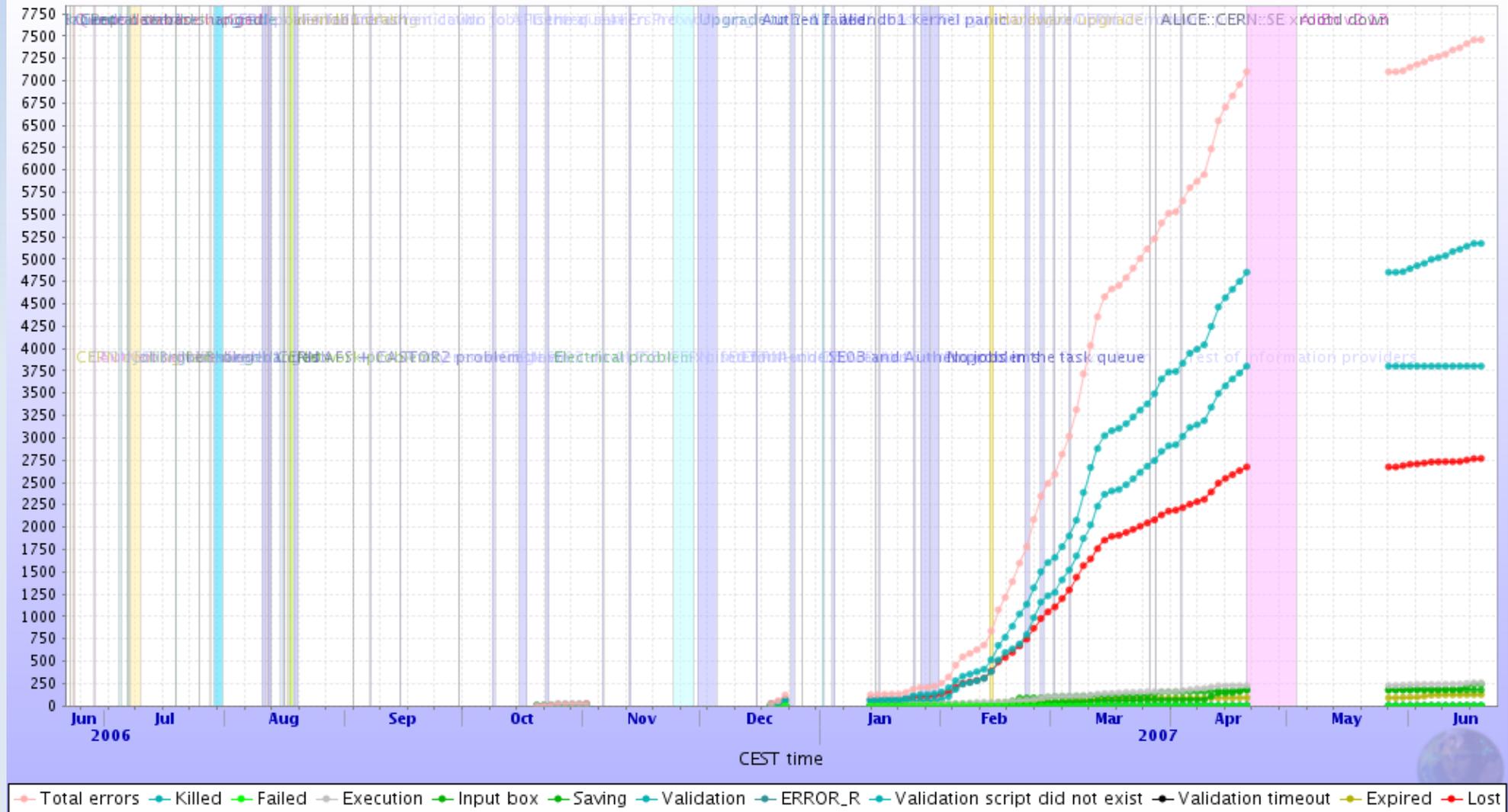
Actividad de Tochtli en ALICE



- Historial de Jobs en Tochtli

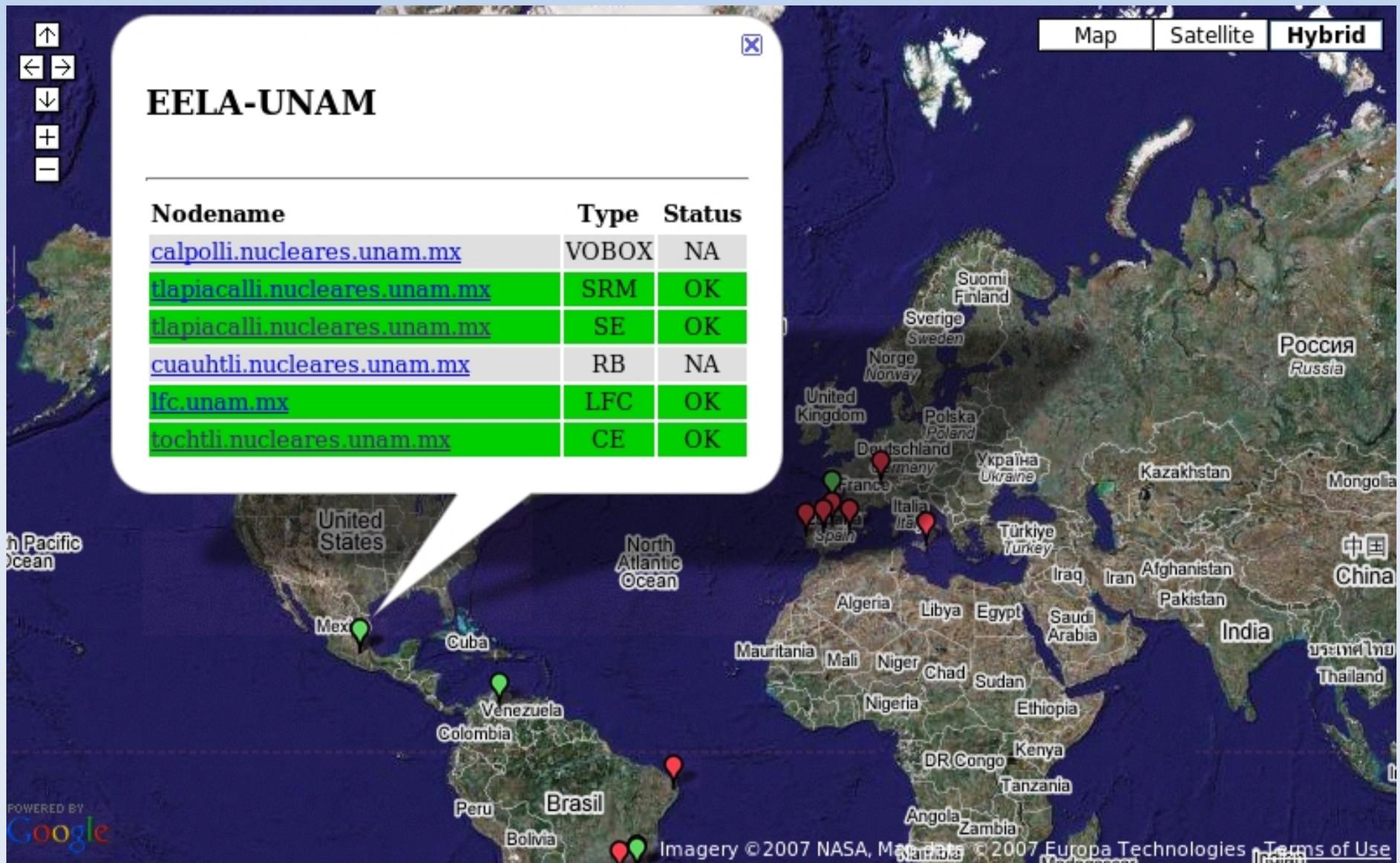
Actividad de Tochtli en ALICE

Error jobs in UNAM



- Acumulativo de errores

Estado de los servicios Grid



- Monitoreo de EELA de los servicios Grid

Siguientes pasos

El primer paso fué instalar los servicios de Grid y lograr el estaus de “producción” en ambas VOs, los siguientes pasos son:

- Aumentar la capacidad de procesamiento y almacenamiento.
- Mejorar el rendimiento de la red

Mayor capacidad de procesamiento



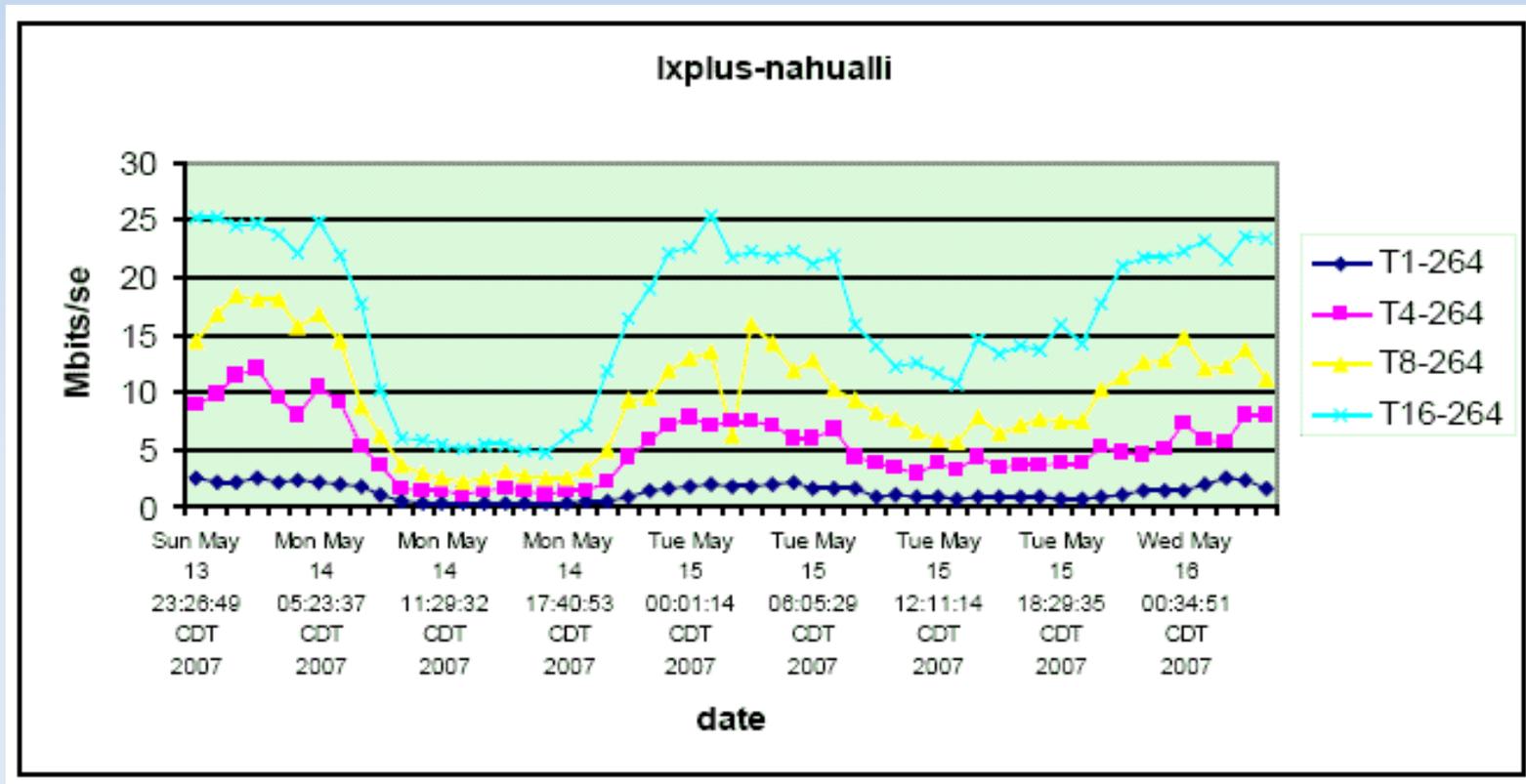
- Se han realizado pruebas de instalación de middleware para WN y AliEn en Kan Balam y no se encontraron grandes problemas. Se espera incluir algunos nodos de este cluster en la Grid de ALICE.

Rendimiento de la red

Se han realizados pruebas de rendimiento de red entre el ICN y el CERN encontrando los siguientes resultados:

- En BW máximo disponible es de 25Mbits/sec por Internet2
- El BW no es constante, es decir, depende del día y hora de la semana:
 - Durante la noche, fines de semana y días feriados es posible alcanzar el máximo BW.
 - En días laborales en las horas de oficina se alcanza el el BW mínimo que puede ser de 0.2Mbits/s (200Kbits/sec)
- El ruteo en dirección CERN-ICN es por el Internet comercial. Existe un ruteo asimétrico.

Gráfica de tráfico



- Trafico de red CERN-ICN

Actualmente se esta trabajando con los expertos de CUDI y el CERN para mejorar esta situación.

Conclusiones

- El ICN-UNAM logró incluir exitosamente a Tochtli en la Grid de ALICE.
- Tochtli se encuentra actualmente en “producción”.
- La experiencia con sitios como el de México, India y Corea ayuda a robustecer el software por las condiciones especiales en las que corre.
- La experiencia será muy útil para instalar más fácil y rápidamente nuevos sitios en condiciones similares.