

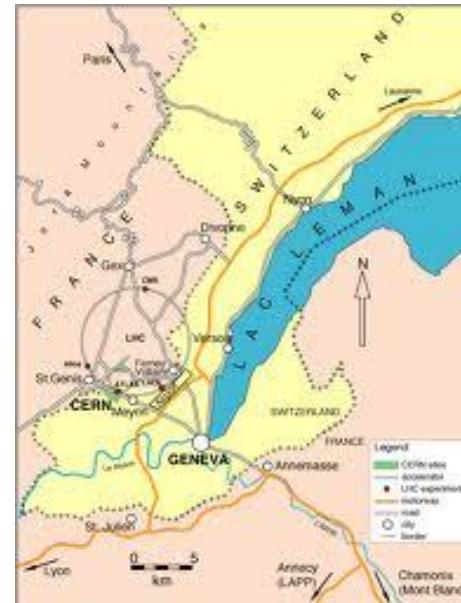
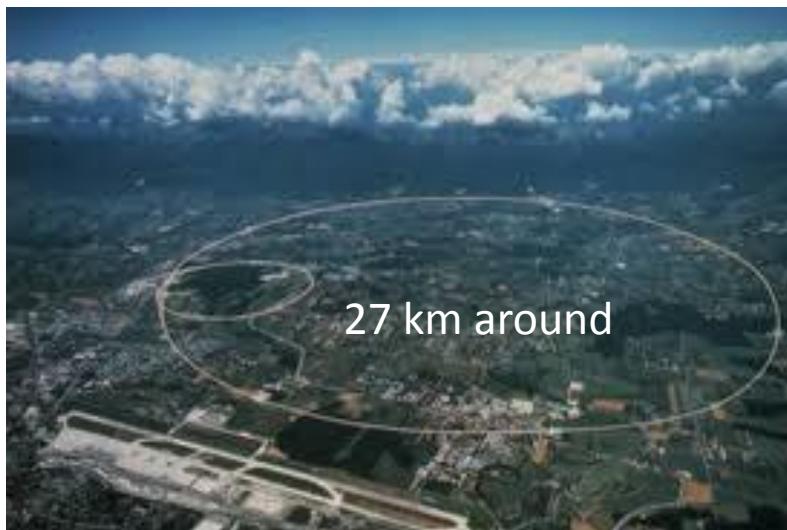
Proyectos de e-ciencia con la red CUDI en la UNAM: El proyecto T1 ALICE-UNAM

Alejandro Ayala

Instituto de Ciencias Nucleares
UNAM

El Gran Colisionador de Hadrones (LHC)

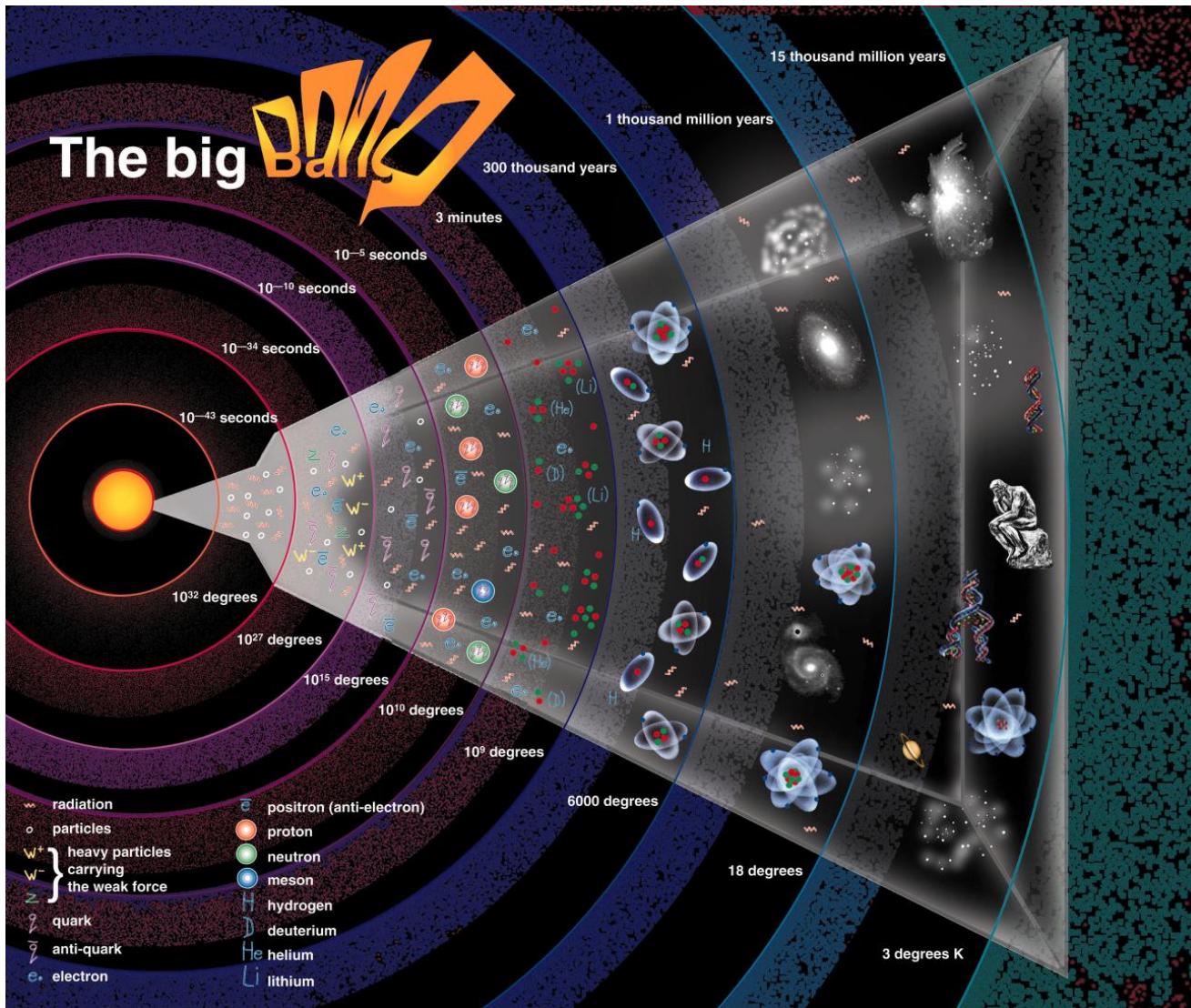
- LHC es un colisionador de protones y de núcleos pesados
- $p + p$ Energía en el centro de masa 14 TeV
- $Pb + Pb$ Energía en el centro de masa 2.7 TeV



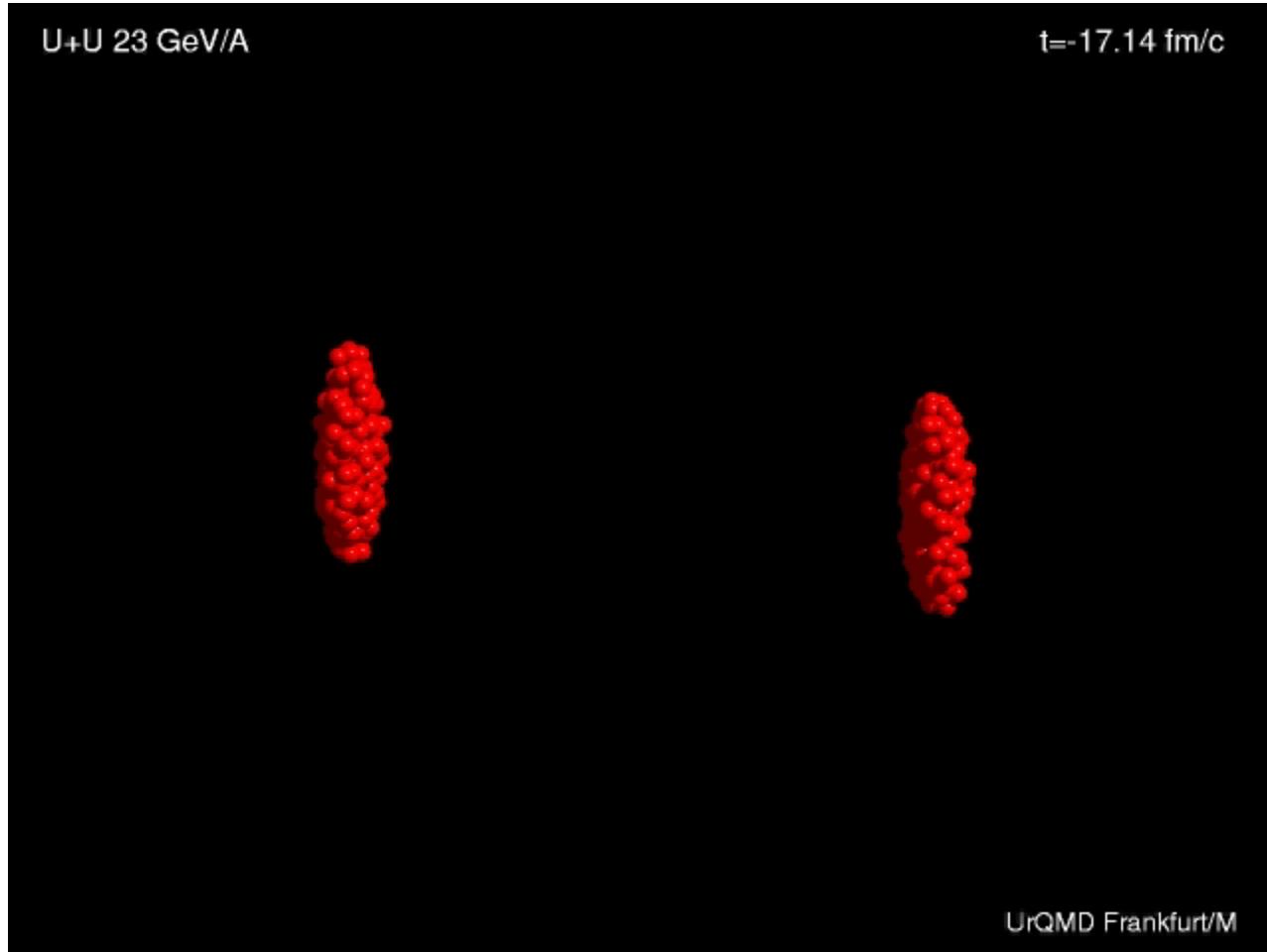
Metas científicas del LHC

- Búsqueda de los grados de libertad e interacciones fundamentales
- Recreación de los primeros instantes del universo mediante la producción de un estado denso y de alta temperatura, el plasma de quarks y gluones

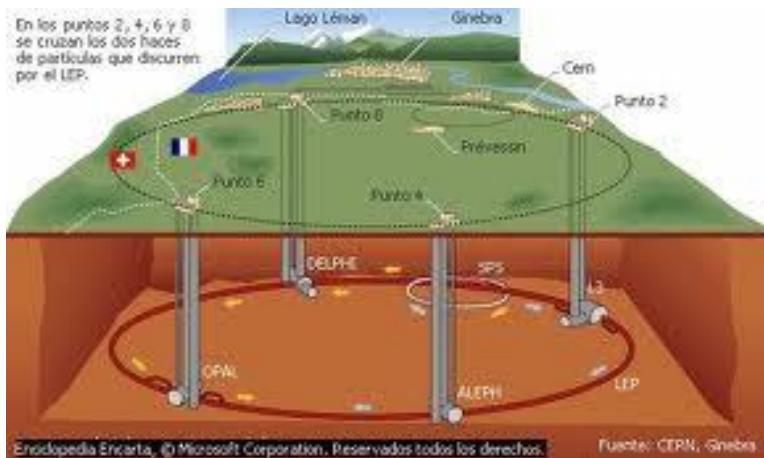
Breve historia del universo



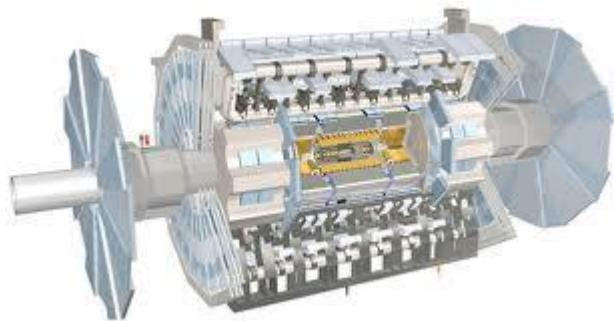
Método: colisiones de hadrones



Cuatro grandes experimentos



Cuatro grandes experimentos

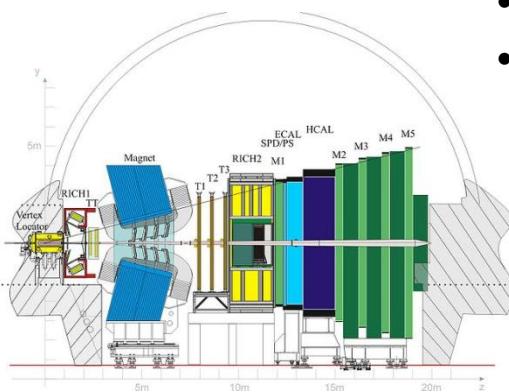


ATLAS

- 8 – 10 PetaBytes / año
- 10^{10} eventos / año
- 10^3 usuarios interactivos y en colas



CMS



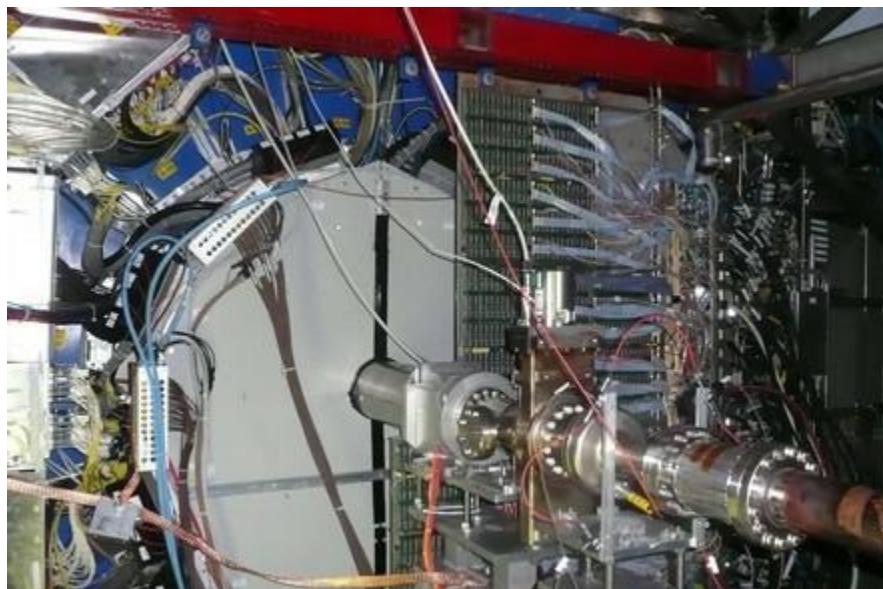
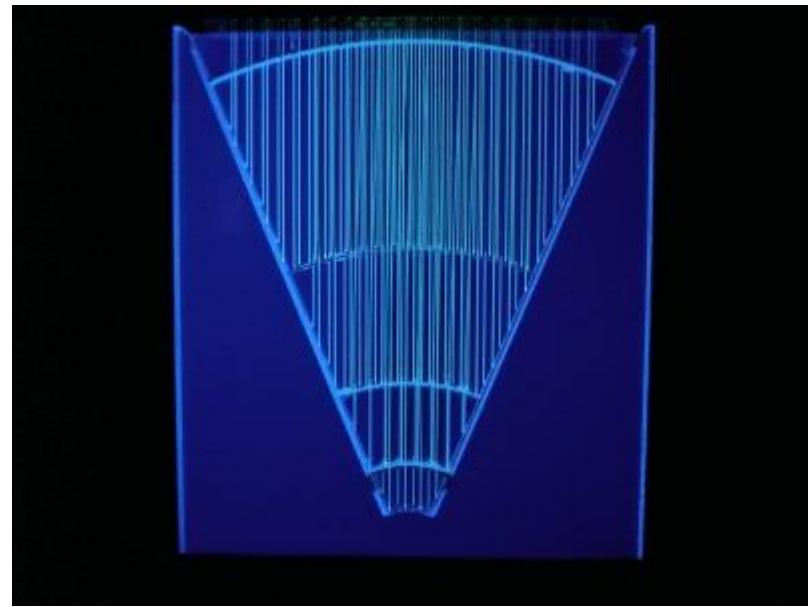
LHCb



ALICE

La comunidad mexicana de altas energías forma parte de esta empresa

- Una de las mas grandes colaboraciones que aglutina teóricos y experimentales es la colaboración ALICE.
- El equipo mexicano responsable de la construcción y operación de dos detectores de disparo: V0A y ACORDE
- Fenomenología, análisis de datos y simulaciones Monte Carlo



Gran cantidad de datos

20,000 m

Pila de CDs con un año de toma de datos

10,000 m

Altura de vuelo

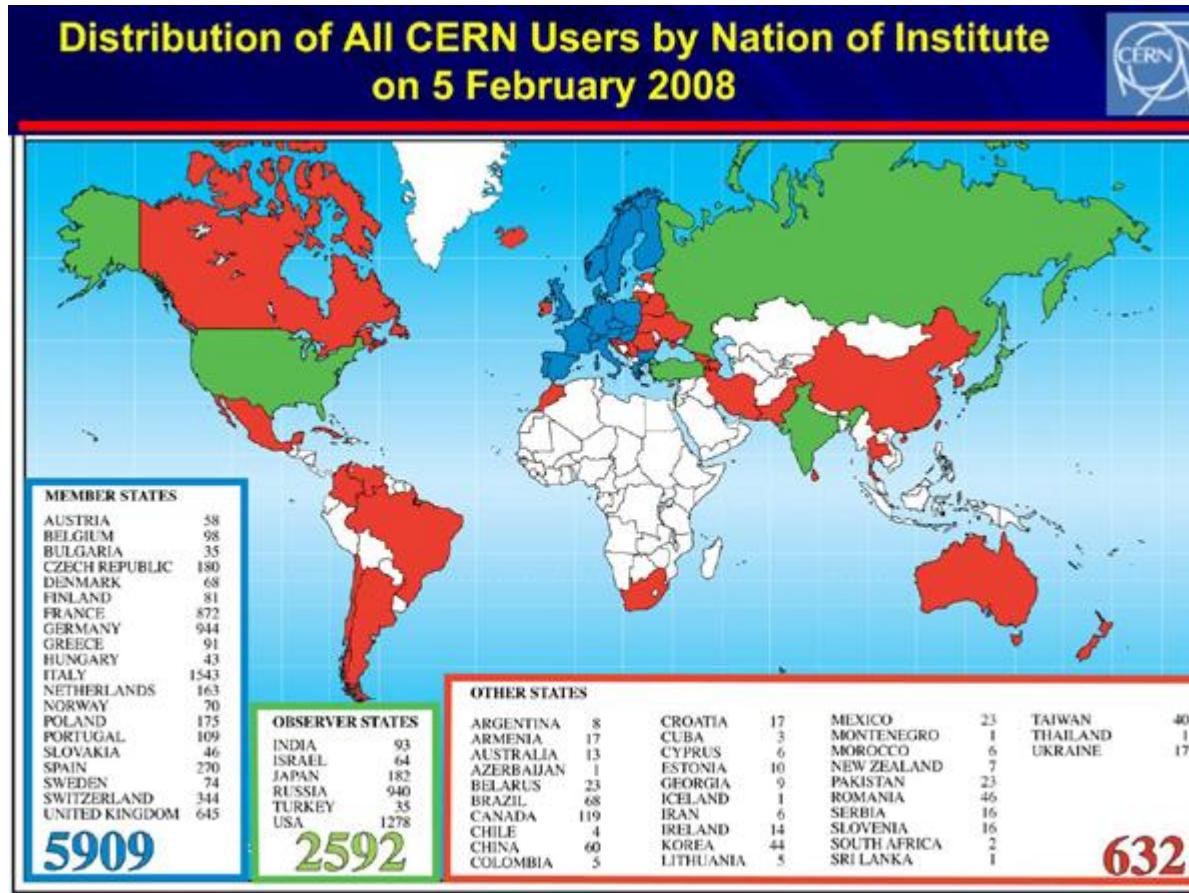


5,500 m

Popocatepetl



(vieja) distribución de usuarios



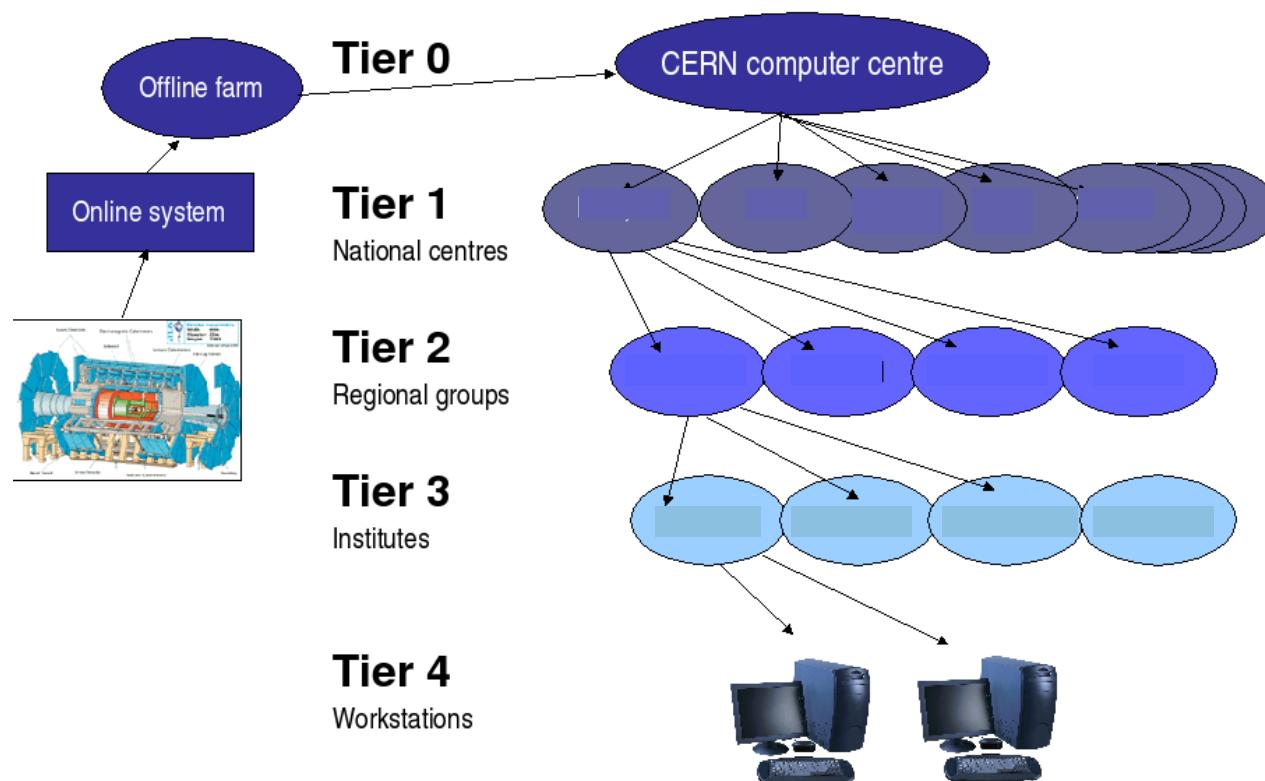
Los experimentos del LHC comparten características

- Generan grandes cantidades de datos
- Requieren grandes volúmenes de almacenamiento
- Equipo de cómputo consume grandes cantidades de energía
- Solución: Cómputo distribuido

Solución: GRID

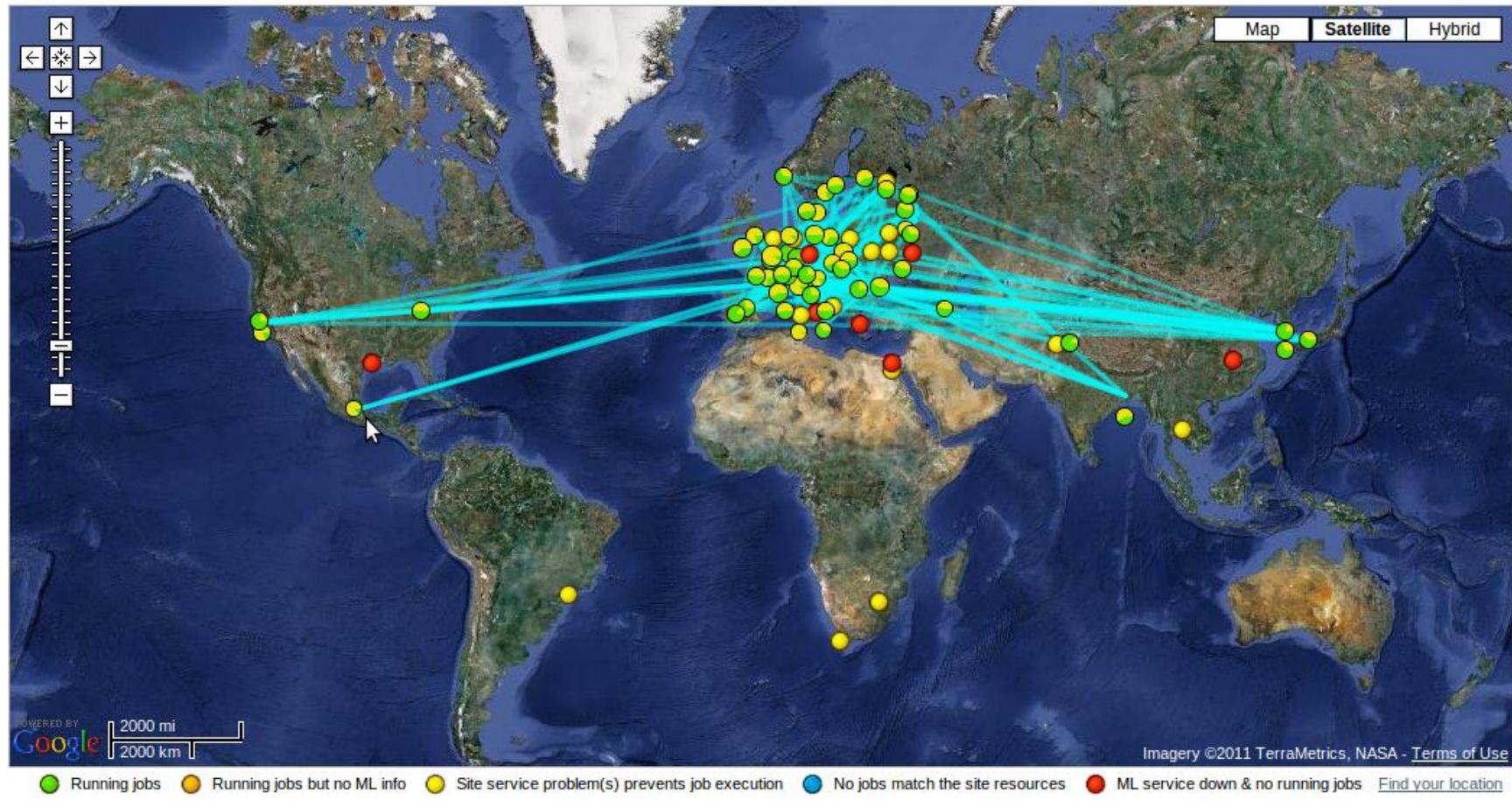
- La idea de la GRID se origino en una conferencia de computo en HEP en 2001
- La GRID ha sido uno de los mayores éxitos del LHC
- Al principio no era claro como hacer que diferentes sitios trabajasen juntos
- Hoy en día es una realidad que se esta volviendo simplemente una herramienta mas en Física

Tier Structure



ALICE GRID

México es parte de la GRID de ALICE desde el 2007 cuando nos unimos a su infraestructura de producción a través de nodo ICN UNAM.



Sitios T1 ALICE

1. NORDUGRID (Europa del Norte)
2. NIKHEF/SARA (Holanda)
3. RAL (Inglaterra)
4. FZK (Francia)
5. CERN
6. CC-IN2P3 GRID (Equipo Pan-Europeo basado en Francia)
7. INFN-CNAF (Italia)
8. KISTI (Corea del Sur)
9. UNAM (¿Mexico?)

Service	Maximum delay in responding to operational problems			Average availability measured on an annual basis	
	Service interruption	Degradation of the capacity of the service by more than 50%	Degradation of the capacity of the service by more than 20%	During accelerator operation	At all other times
Acceptance of data from the T0 center during accelerator operation	12 hours	12 hours	24 hours	99%	n/a
Networking service to the T0 center during accelerator operation	12 hours	24 hours	48 hours	98%	n/a
Data-intensive analysis services, including networking to T0, T1 centers beyond accelerator operation	24 hours	48 hours	48 hours	n/a	98%
All other services – prime service hours	2 hours	2 hours	4 hours	98%	98%
All other services – beyond prime service hours	24 hours	48 hours	48 hours	97%	97%

La instalación y mantenimiento de un sitio GRID en producción **definitivamente no** es un proceso de un solo click



Double click to install a site

¡Mucho menos la de un T1!

Metas

- Construir un centro de datos con al menos 2000 núcleos (cores) y un petabyte de datos (un millón de gigabytes).
- Configuración del centro de datos como nodo en la infraestructura de GRID EGI, ejecutando trabajos del experimento ALICE.
- Expansión del centro para dar servicio a otros clientes.
- Abrir la posibilidad para futuras colaboraciones para el desarrollo de nuevas herramientas para la GRID mundial de cómputo. Integración de tecnologías GRID con otras tecnologías de manejo de recursos de cómputo, como nubes de cómputo.
- Consolidar los recursos de cómputo de alto rendimiento en la UNAM y desarrollo de esquemas para asignación de recursos en función de demanda con garantías fuertes a proyectos prioritarios.
- Formación de expertos en temas de GRID, manejo de grandes volúmenes de datos y en el diseño de centros de *cómputo verde* a escalas medianas y grandes.

Respecto a la comunidad nacional

- Centros de operación regional están evolucionando hacia iniciativas nacionales de GRID
- La generación de profesionales con experiencia y con habilidades es una fuente deseable de “know how” que puede ser explotada en diversas aplicaciones

¿Recursos?

- Un T1 no es acerca de recursos financieros sino mas bien del desarrollo de experiencia; se trata mas de gente que de máquinas
- Aun así, los servicios de un T1 hacen sentido solo si alcanzan un cierto grado de eficiencia
 - Un T1 debe contar con unos cuantos miles de cores

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

SECRETARÍA GENERAL

DIRECCIÓN GENERAL DE CÓMPUTO Y DE TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

Estimación de inversiones

Proyecto: TIER-1. GRID de las Américas

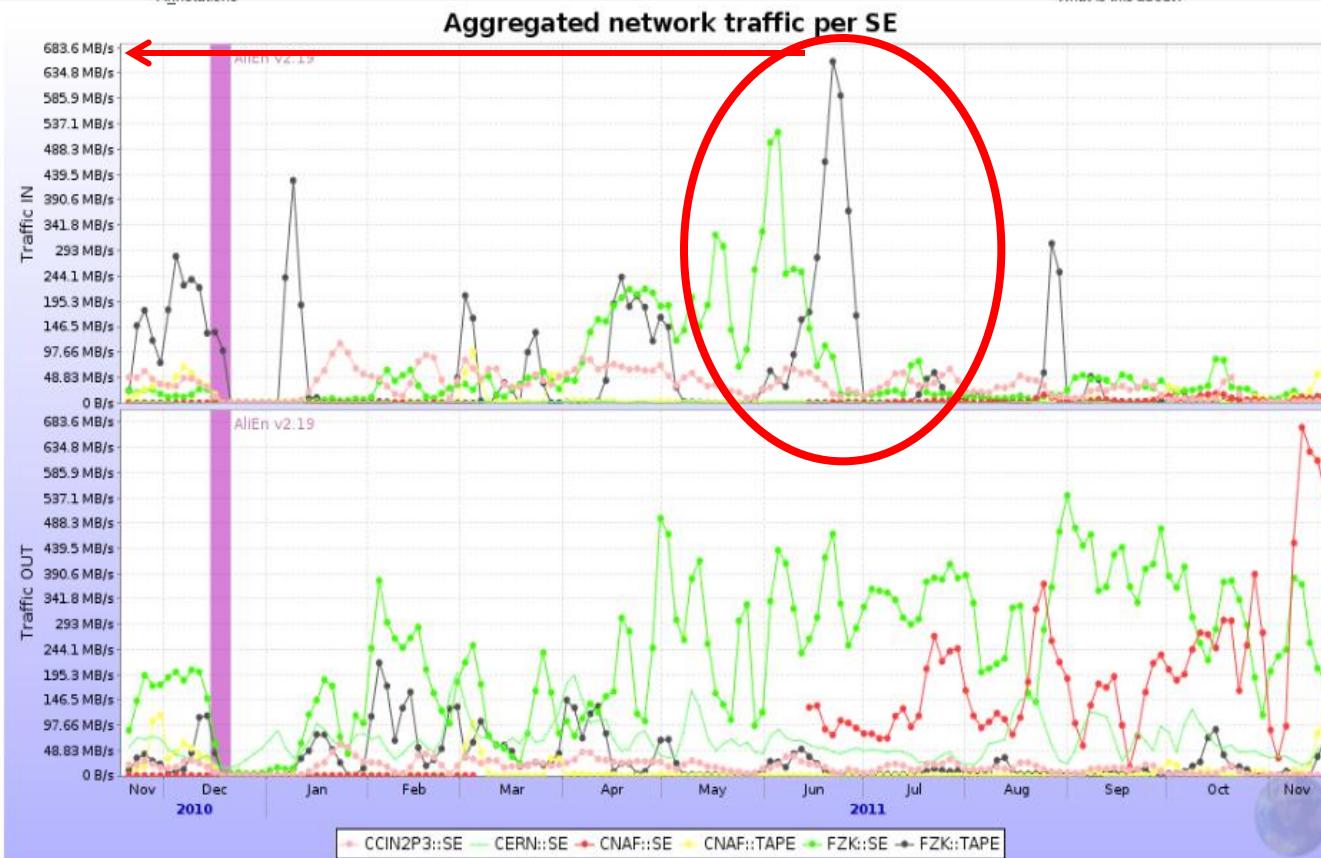
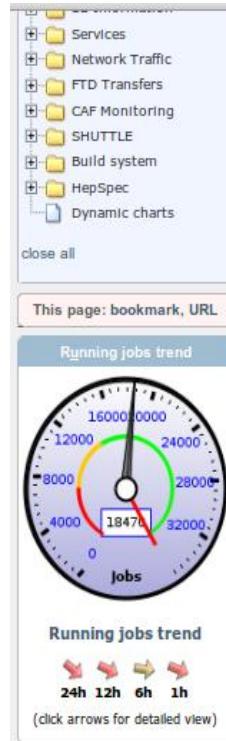
Última actualización 21 de febrero de 2011.

Cantidades expresadas en dólares americanos (USD)

Cantidad	Concepto	Costo unitario	Costo total
Infraestructura de almacenamiento (Datacenter Tier-1)			
1	Sistema de almacenamiento tipo SAN - NAS 1.3 PB (20% SAN discos de alta velocidad 15KRPM - 80% NAS discos de mediana velocidad 7.2 KRPM) Expandible hasta 2.6 PB utilizables	\$ 250,000.00	\$ 250,000.00
Infraestructura de procesamiento (Tier-1)			
32	Nodo de procesamiento con dos procesadores (CPU) y 16 núcleos por procesador. Tecnología de 32nm. Total 32 núcleos (cores) por nodo @ 2.4 Ghz. Bus de 64 bits. Interconexión Infiniband. 166.4 GigaFLOPS por nodo. 5.31 TeraFLOPS por los 32 nodos	\$ 15,600.00	\$ 499,200.00
Infraestructura de operación (Tier-1)			
2	Acondicionamiento de espacio y red para Racks 2U. Nodos de Tier-1 procesamiento	\$ 12,500.00	\$ 25,000.00
3	Acondicionamiento de espacio y red para Racks 2U. Nodos de Tier-1 almacenamiento	\$ 12,500.00	\$ 37,500.00
1	Sistema de aire acondicionado	\$ 52,000.00	\$ 52,000.00
1	Sistema de suministro ininterrumpible de energía eléctrica (UPS)	\$ 137,000.00	\$ 137,000.00
Mantenimiento y soporte técnico (Anuales)			
4	Salario anual de Técnico Académico Asociado C tiempo completo. Especialista en cómputo de alto rendimiento (HPC)	\$ 9,375.00	\$ 37,500.00
1	Contrato de mantenimiento y soporte en sitio de infraestructura instalada. Anual	\$ 50,035.00	\$ 50,035.00
TOTAL			\$ 1,088,235.00

¿Recursos?

- Un T1 no es acerca de recursos financieros sino mas bien del desarrollo de experiencia; se trata mas de gente que de máquinas
- Aun así, los servicios de un T1 hacen sentido solo si alcanzan un cierto grado de eficiencia
 - Un T1 debe contar con unos cuantos miles de cores
 - Un T1 debe contar con una red confiable, ancha y de calidad



Statistics

Traffic IN

Series	Last value	Min	Avg	Max	Total
1. CCIN2P3::SE	4.427 MB/s	0 B/s	35.31 MB/s	333.9 MB/s	1.03 PB
2. CERN::SE	786.2 KB/s	45.77 B/s	1.184 MB/s	73.59 MB/s	35.39 TB
3. CNAF::SE	12.12 MB/s	0 B/s	2.841 MB/s	781.3 MB/s	84.9 TB
4. CNAF::TAPE	64.35 MB/s	0.286 KB/s	7.371 MB/s	357.4 MB/s	220.2 TB
5. FZK::SE	8.212 MB/s	0 B/s	61.73 MB/s	73.48 GB/s	1.801 PB
6. FZK::TAPE	20.26 MB/s	0 B/s	61.69 MB/s	1.165 GB/s	1.8 PB
Total	110.1 MB/s		170.1 MB/s		4.964 PB

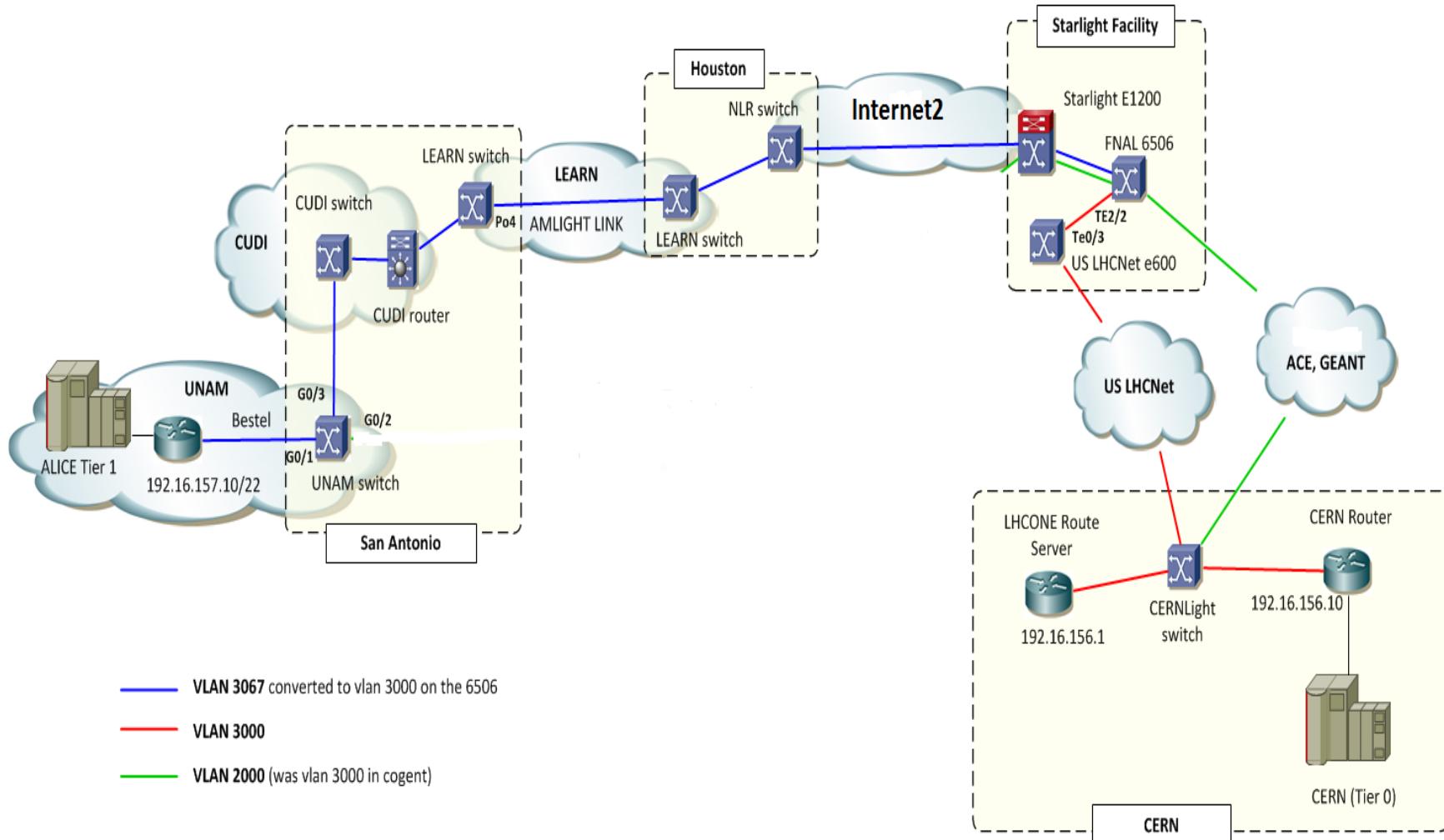
Traffic OUT

Series	Last value	Min	Avg	Max	Total
1. CCIN2P3::SE	1.664 MB/s	0 B/s	16.75 MB/s	158.3 MB/s	500.5 TB
2. CERN::SE	68.49 MB/s	60.23 B/s	66.83 MB/s	588.4 MB/s	1.95 PB
3. CNAF::SE	539.6 MB/s	0 B/s	117.7 MB/s	2.25 GB/s	3.434 PB
4. CNAF::TAPE	117.6 MB/s	94.59 B/s	7.925 MB/s	473.6 MB/s	236.8 TB
5. FZK::SE	178.6 MB/s	0 B/s	240.1 MB/s	54.05 GB/s	7.006 PB
6. FZK::TAPE	72.9 MB/s	0 B/s	30.1 MB/s	2.288 GB/s	899.5 TB
Total	978.8 MB/s		479.4 MB/s		13.99 PB

Infraestructura

- Espacio para la instalación del equipo de cómputo del centro de datos (site).
- Conexión de 1Gbps a las redes académicos en los EEUU y del mundo (**ALICE no mantiene flujo constante pero en toma de datos picos de 400 Mbps. Conexión al CERN deben garantizarse**) (**Sumando HAWC con 1 TB de datos diarios se requiere 100 Mbps continuo constante**)
- Conexión a la red metropolitana del LANCAD (Delta metropolitana) con enlaces al CINVESTAV y a la UAM-Iztapalapa.
- RedUNAM del campus CU con dorsal de 10Gbps y conexión de 1Gbps al ICN-UNAM.

Conexión UNAM-CUDI CERN

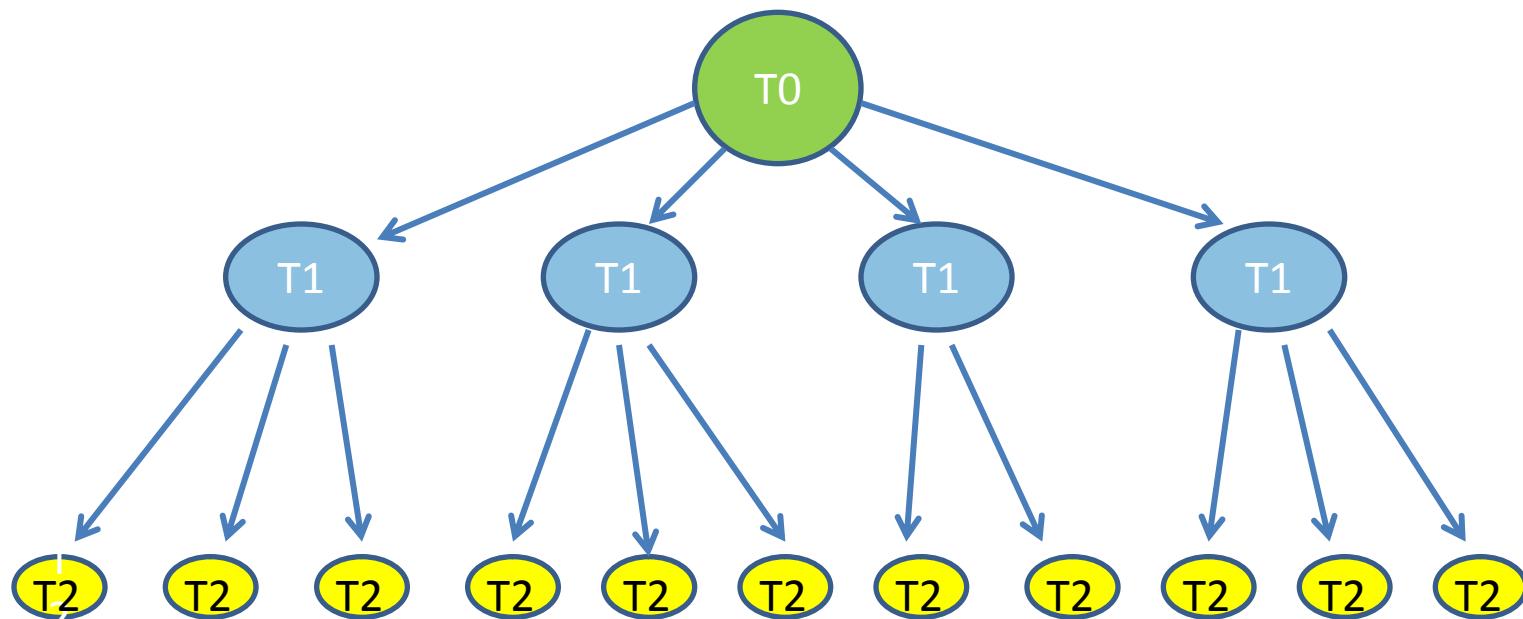


Capacidad de extremo a
Extremo de 1Gb/s en L2

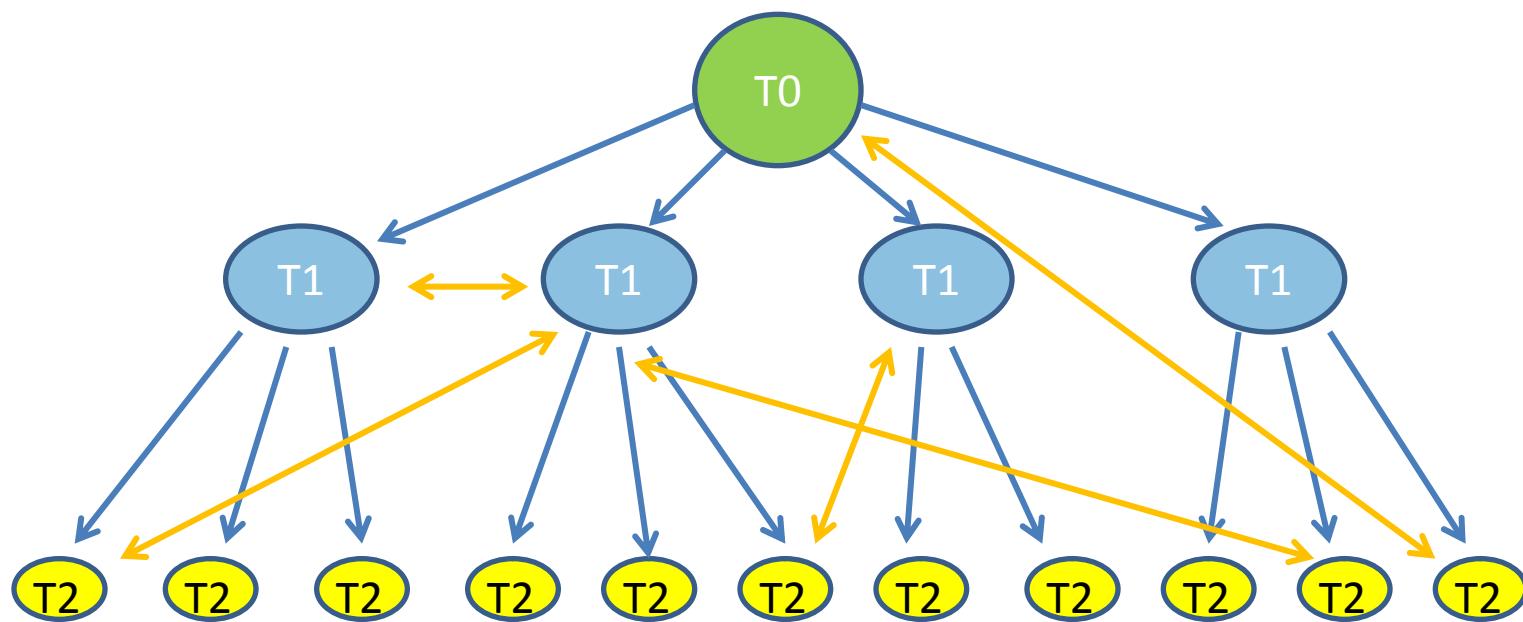
Redes participantes

- RedUNAM
- CUDI
- Amlight-LEARN
- Internet2
- StartLight
- CERN-NET

Dado el buen comportamiento de la red el modelo de computo esta evolucionando al de la nube



Dado el buen comportamiento de la red el modelo de computo esta evolucionando al de la nube



Una nube sobre la GRID

- Jerarquía T0-1-2-3 se suaviza debido a que el usuario decide la transferencia de datos y su almacenamiento
- T1 y T2 se están volviendo equivalentes en la red
- Ya no será el espacio de disco sino el **ancho de banda** lo que escalará con el numero de usuarios y la cantidad de datos

¿Y para que un T1?

- La administración de un T1 significa enfrentar retos importantes concentrados en el desarrollo de la experiencia para proveer servicios confiables a la comunidad mundial
- Los T1, en su calidad de centros bien mantenidos y con alta disponibilidad se encuentran al frente de la evolución de la GRID hacia la nube y mas allá
- Operar un T1 significa estar al frente de esta evolución y en una posición ideal para influenciarla

Conclusiones

- México cuenta con una comunidad de Física de Altas Energías madura. Dentro de esta existe un grupo sólido de física de iones pesados en la frontera de los desarrollos. Somos miembros de ALICE-CERN desde hace 10 años
- Para ser un protagonista en el análisis de datos se requieren de recursos modestos pero la experiencia se puede escalar, desarrollar mas “know how” y compartirla con otras comunidades GRID
- Un T1 dentro de la colaboración donde esta experiencia se desarrolle contribuirá a aumentar el ritmo de este proceso.

GRACIAS