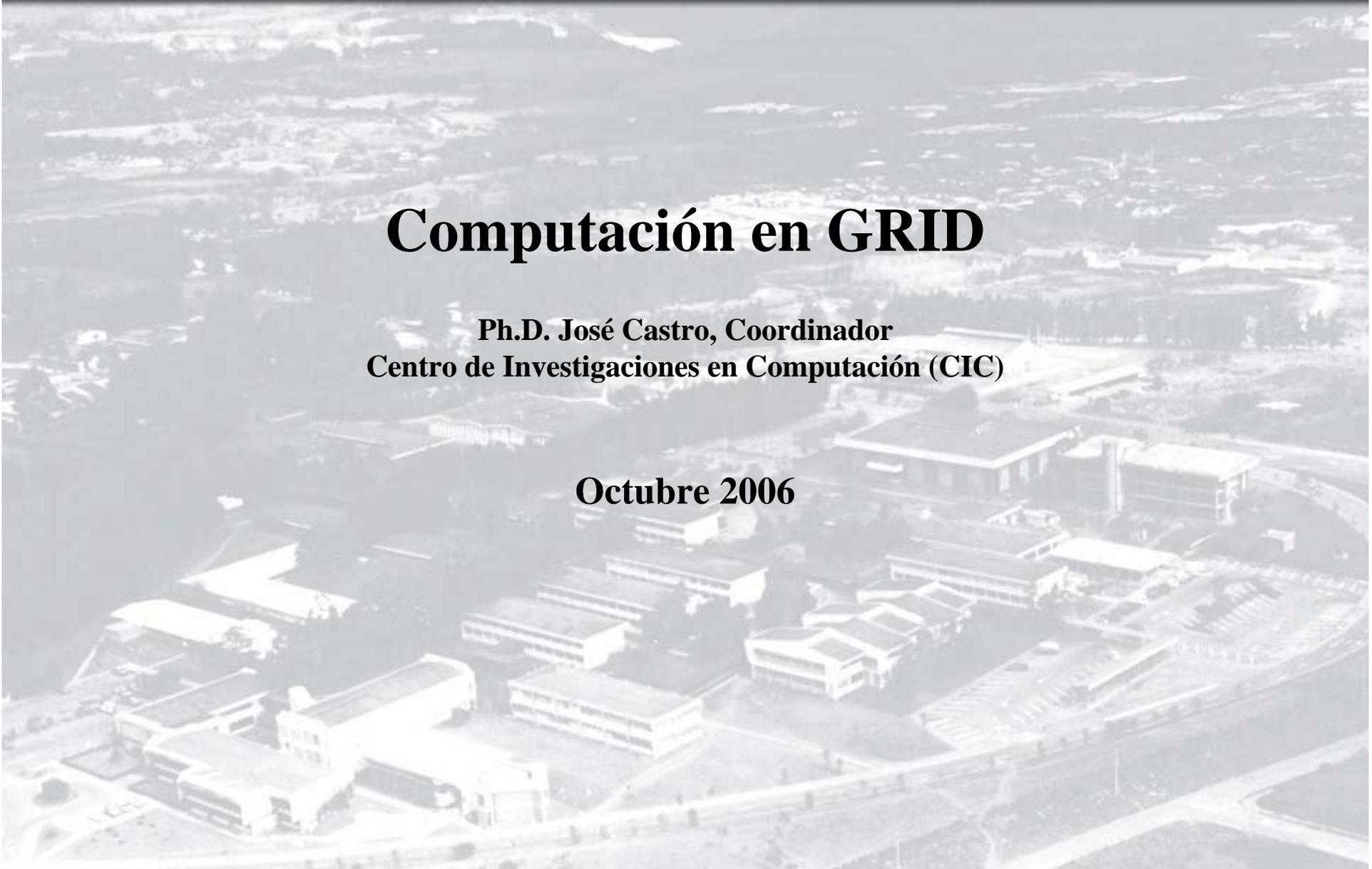


Computación en GRID

**Ph.D. José Castro, Coordinador
Centro de Investigaciones en Computación (CIC)**

Octubre 2006





Índice de presentación

- Qué es computación en Grid
 - Diferencias y similitudes con computación en clusters
 - Conceptos principales
 - Un poco (muy poco) de historia
- Estándares abiertos en Grid
- Algunas aplicaciones en computación en Grid
- El proyecto cluster/Grid interuniversitario
 - Justificación y enfoque
 - Alcances y oportunidades
- Aplicaciones posibles en el contexto nacional



Que es computación en Grid

- El concepto actualmente se ha utilizado para expresar una forma de hacer las cosas que tiene un fuerte traslape con tecnologías y aplicaciones que muchos ya conocemos y no llamaríamos Grid.
 - Parte investigación científica
 - Parte reconceptualizar herramientas existentes dentro de un nuevo contexto
 - Concepto sumamente ambicioso



Computación Grid

- Concepto Cluster
 - Aglomeración de muchas máquinas convencionales en una red que utilizando un software funcionan como un solo supercomputador.
- Ventaja de la plataforma
 - Construcción del supercomputador con componentes fácilmente adquiribles en el mercado.
 - Costo ordenes de magnitud mas bajo que otros supercomputadores de comparable capacidad computacional (ie: CRAY computers).
 - Nodos del cluster conformados solamente por la tarjeta madre, eliminando la necesidad de adquirir monitor y demás periféricos.
 - Operan con software totalmente libre sobre el sistema operativo Unix/Linux, reduciendo considerablemente el costo de la investigación.
 - Operan sobre software fácilmente adquirible, reduciendo la curva de aprendizaje de los programadores.



Que es computación en Grid

Misión Computación Grid

- El grid es un servicio para compartir poder de procesamiento y de almacenamiento sobre el internet. Su objetivo es convertir la red global en un extenso recurso computacional. (A diferencia de la web, cuyo objetivo es compartir ofrecer un servicio para compartir datos e información)



Que es computación en Grid

Visión Computación Grid

- Máquinas y clusters a lo largo y ancho de internet, conjunto a equipo especializado (laboratorios, sensores) que son accesibles remotamente y transparente por un usuario en cualquier parte del mundo sin la necesidad de preocuparse por accesos, tiempo procesamiento, instalación de software, etc.
 - Esta definición, sobra decir, es trabajo en proceso, e implica la implementación y depuración de muchos programas y estándares.



Computacion en Grid

- Alcanzar la computación en Grid es un objetivo grande de la comunidad científica internacional.
 - Porqué se busca computación en Grid?
 - Incremento exponencial en la información disponible en formato digital
 - Necesidad de integrar esa información (problema semántico)
 - Capacidad de procesamiento local no permite la obtención de resultados con premura.
 - Declaraciones de problema/visión
 - CyberInfrastructure Initiative
 - eScience project (european grid)
 - Reporte 2020 Microsoft
 - Caso de la BioInformatica, Ciencias física



Computación en Grid

- Tecnologías utilizadas para avanzar el estado de computación grid
 - Computación distribuida
 - Metacomputing
 - Computación en clusters
 - Peer to Peer computing
 - Computación por internet (SETI@home)
 - Cycle scavenging



Computación en Grid

- Historia
 - 1965 - Primeros conceptos tipo grid aparecen en el sistema operativo Multics (compartir recursos era un problema bastante importante)
 - 1995 - Metacomputadoras FAFNER y I-WAY influyen fuertemente en la arquitectura del actual OGSA
 - FAFNER (Factoring via Network Enabled Recursion) scavenger software
 - I-WAY Information Wide Area Year, conexión de computadoras utilizando infraestructura existente nada mas, genero el inicio de un resource broker computacional e influyo en GLOBUS



Computación en Grid

- 1997
 - Publicación de: *The Grid: Blueprint for a new computational infrastructure*. Ian Foster (Argonne National Labs), Carl Kesselman (University of South California)

Considerada la biblia de computación Grid

Mismos autores publican: *Globus: a Metacomputing infrastructure toolkit*



Computación Grid

- Proyectos precursores a grid (grid like)
 - Condor (1987), proyecto de la universidad de Wisconsin, su objetivo era el “cycle scavenging”, condor fue originalmente concebido para una LAN, ahora esta el condor-G que trata de integrar Globus para someter tareas en Grids
 - CODINE (Computing for Distributed Network Environments), desarrollado por una compañía alemana, adquirida por SUN en el 2000, actualmente el corazón del SUN Grid Engine. En EPCC computing center en Escocia se esta trabajando para incorporar SUN Grid Engine con Globus



Computación Grid

- Legion 1993, Universidad de Virginia. Enfoque basado en orientación a objetos. Genero compania Applied Meta -> Avaki. Produce soluciones Grid.
- Nimrod 1994, Universidad de Monash en Australia. Distribución de cálculos similares en un LAN, actualmente esta Nimrod/G
- Unicore 1997, iniciativa Alemana para conectar sus centros de supercomputación. Actualmente se trabaja en su integración con Globus.



Computación en Grid

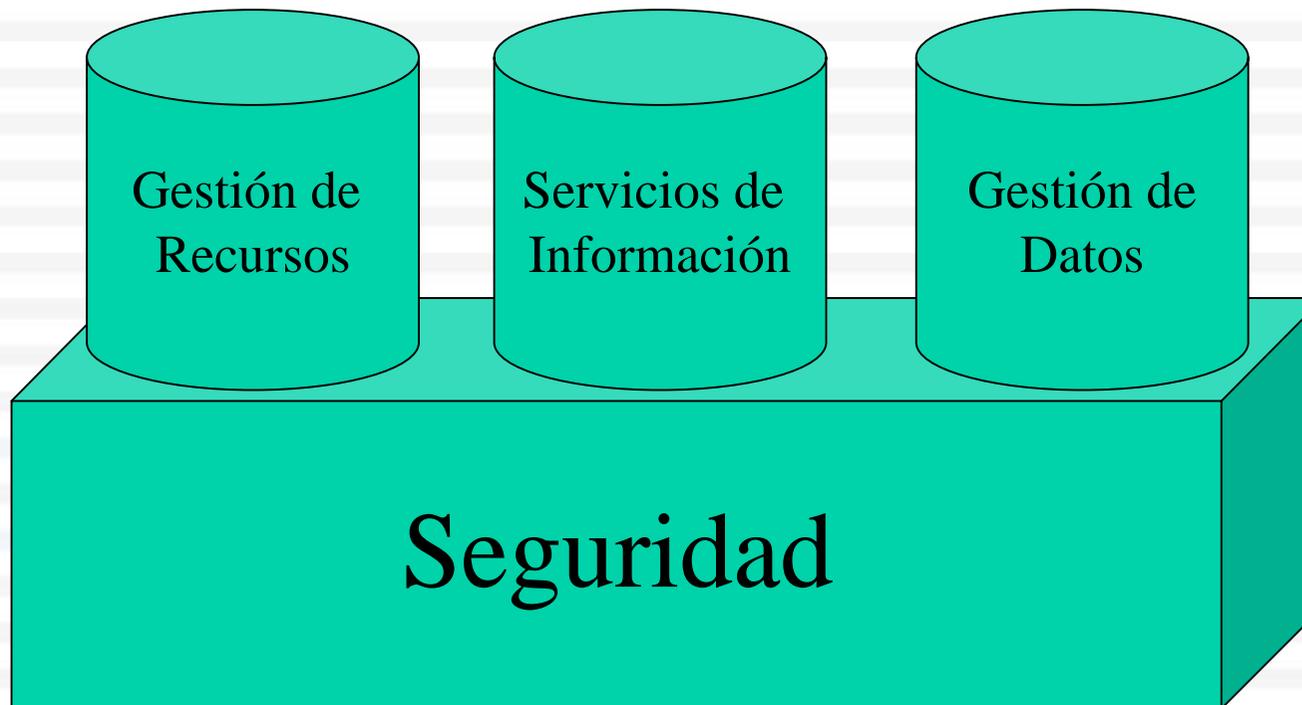
- Actualmente no existe un unico Grid, sino muchos grids, cada uno con capacidades restringidas y posibilidades de conexión limitadas. Pero también ofreciendo enormes recursos computacionales y de almacenamiento para aquellos interesados en utilizarlas.
 - TeraGrid
 - European Grid Initiative
 - LataGrid
 - PRAGMAy muchos mas...



Computación en Grid

- Un punto donde las comunidades científicas y la comercial convergen es en la definición de estándares
 - Actualmente se tiene el estándar del Open Grid Service Architecture (OGSA) que se monta sobre servicios en Web (web services).
 - La definición de este estándar conlleva a la definición del Globus Toolkit, desarrollado por un consorcio entre academia e industria en el cual uno de los participantes mas activos es IBM. Actualmente va por su versión 4.0
 - Dentro de las aplicaciones comerciales el uso de arquitectura OGSA y globus, con su concepto de middleware lo indica, se mantiene transparente para el usuario.

- Globus Toolkit (componentes fundamentales)





Computación Grid

- Gestión de Recursos = Grid Resource Allocation Manager (GRAM)
- Gestión de Información = Monitoring and Discovering Services (MDS)
- Gestión de Datos = GridFTP (FTP optimizado para redes con altos anchos de banda)
- Seguridad = Grid Security Infrastructure (GSI), esquemas de certificados con llave pública y privada



Computación Grid

- GridFTP:
 - Múltiples canales de datos para transferencias paralelas.
 - Transferencias parciales de archivos.
 - Transferencias server-to-server
- GSI
 - Comunicación segura entre elementos del grid.
 - Soporte de "single sign on", incluyendo delegación de credenciales para realizar cómputo que involucre múltiples recursos y sitios.



Computación Grid

Arquitectura OGSA

- Establecimiento de identidad y negociación de autenticaciones
- Descubrimiento, monitorización y gestión de servicios
- Negociación y monitorización de niveles de servicio
- Comunicación y Gestión de la Virtualización de miembros de la Grid.
- Uso jerárquico de servicios Grid
- Integración de recursos de datos en
 - los procesos de computación
- Gestión de recursos a través de plataformas heterogéneas
- Generación de Calidad de Servicio (QoS) adecuada
- Definición de una base común para gestión autónoma (capacidad de los sistemas de auto verificarse y autogestionarse)
- Interfaces públicos y abiertos.
- Utilización de tecnología estándar: SOAP; XML, ...
- Integración con los recursos de TI existentes.



Computación Grid

- Modelo de programación OGSA: algunas de las definiciones del modelo de programación son obligatorias y otras optativas.
- Éstas definiciones son:
 - *Factory* (factoria de clases)
 - *Registry* (registro)
 - *Discovery* (localización de recursos)
 - *Life cycle* (Ciclo de Vida)
 - *Query service data* (petición de servicio)
 - Notificación
 - Invocación fiable



Computación en Grid

- Grids científicas y Grids comerciales
 - Mientras que a la comunidad científica le interesa hacer descubrimientos, a la comunidad comercial le interesa hacer dinero -> percepciones sobre el grid son distintas
 - Muchos proveedores han ofrecido soluciones grid basadas mas que nada en el modelo “Enterprise”
 - Bajan costos y mejoran soluciones, pero el nivel “sharing” es bajo
 - Intereses científicos implican poder poner a colaborar estaciones de trabajo con hardware y software de distintos proveedores, y con posibilidades de acceso bastante mas universalizados



Computación en Grid

- La comunidad comercial esta interesada en implementar Grids dentro de una misma organización
- La comunidad científica esta implementando y elaborando testbeds, que involucran gran cantidad de países y organizaciones



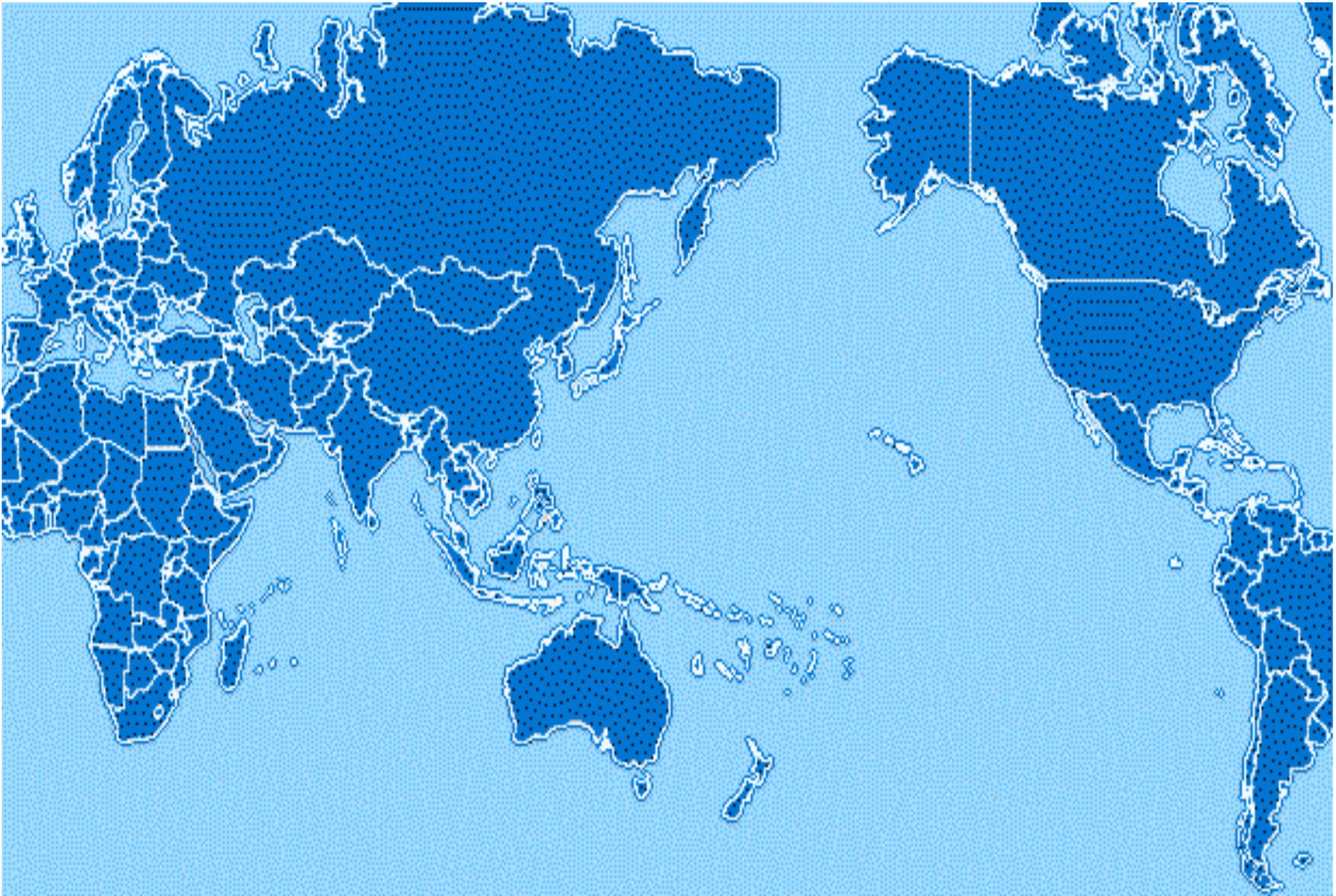
Computación en Grid

- Las cinco ideas principales
 - Compartir recursos
 - Mecanismos para establecer confianza y responsabilidad
 - Seguridad
 - (pólíticas de acceso, autenticación, autorización)
 - Balance de tareas
 - Balance óptimo de tareas
 - Eliminación del factor distancia (ya es realidad en telefonía)
 - Estándares abiertos
 - OGSA Globus Toolkit



Computación en Grid

- Testbeds
 - Actualmente se implementan en redes de alto rendimiento
 - Red GEANT Europea (10Gbps)
 - UK Super Janet Network (10Gbps)
 - Nec Earth simulator 10GFlops/seg 640 procesadores con 10Terabytes memoria y 700Terabytes en disco

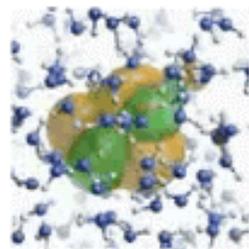


30 Clusters from 27 institutions in 14 countries

Source Cindy Zheng

BioPfuga: *Biosimulation Platform United* *on Grid Architecture*

A platform where individual application programs at the different levels are united to execute a hybrid computation. In particular, *BioPfuga* is a platform for biosimulation.



[About](#)
[Libwrapper](#)

<http://www.biogrid.jp/>

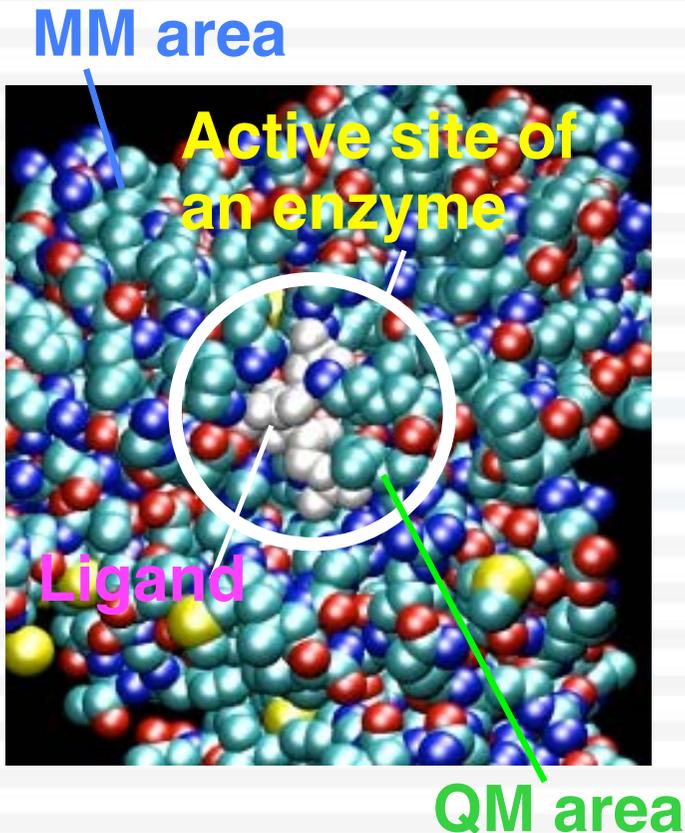
BioPfuga

Development of *BioPfuga* (Biosimulation Platform United on Grid Architecture)

The usual usage of the grid architecture is running one computation on many distributed CPUs through a rapid network. However, in order to analyze much more complicated biological systems, composed of simulations at different levels, along the new paradigm for biological science, more integrated computational approaches are required. The individual programs should be driven on their own corresponding machines on the grid system. For this purpose, we have designed and developed a new platform, *BioPfuga* (**Biosimulation Platform United on Grid Architecture**) where individual applications, corresponding to the different levels of bio-simulations, are united and executed as a hybrid application.

Quantum mechanics (QM) and Molecular mechanics (MM) simulation coupled on *BioPfuga*

Coupled simulation on Grids



Hamiltonian of total system

$$H_{total} = H_{QM}(x,x) + H_{QM/MM}(x,y) + H_{MM}(y,y)$$

Calculate in QM area

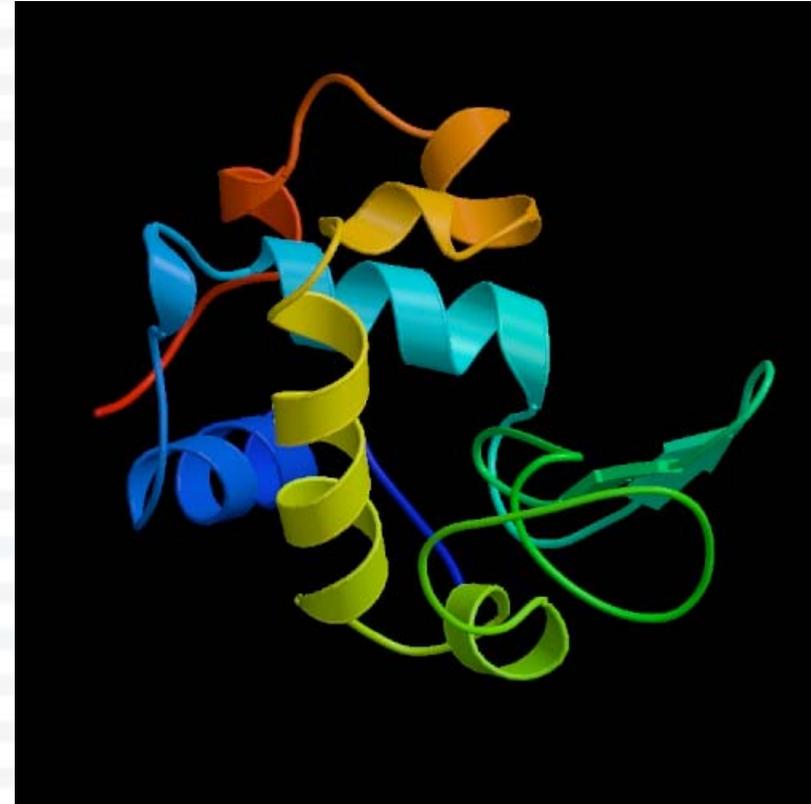
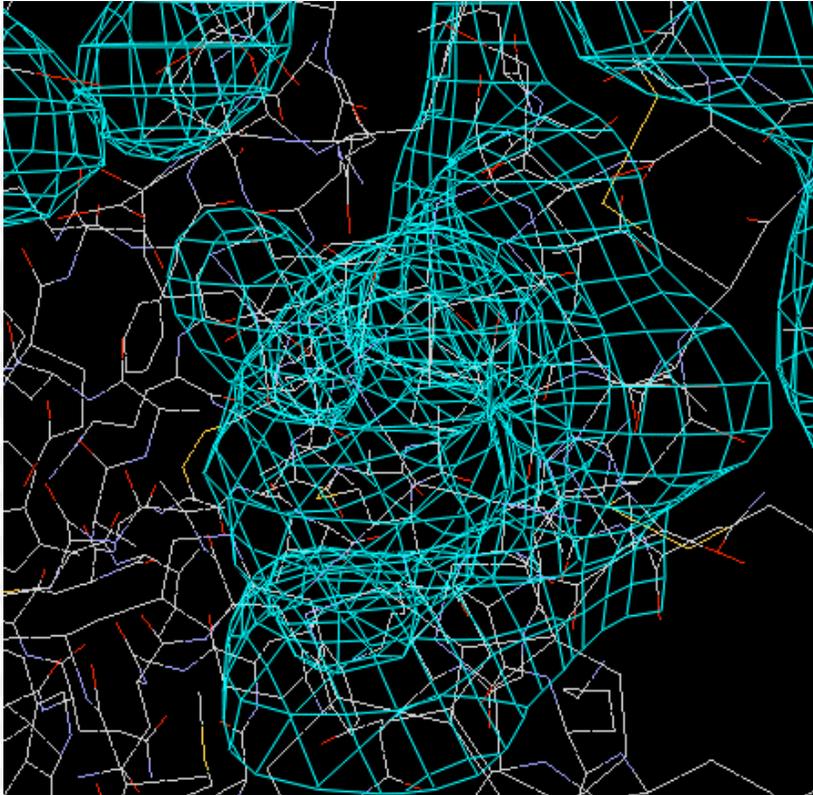
Calculate in MM area



PDB (Protein Data Bank): Protein Tertiary Structure Database

Instituto Tecnológico de Costa Rica

Atom species and coordinates, amino acid residues, any other experimental results with raw data, experimental methods, and the conditions.



X-ray crystallography, NMR, and Electron microscope experiments.

Protein Tertiary Structure

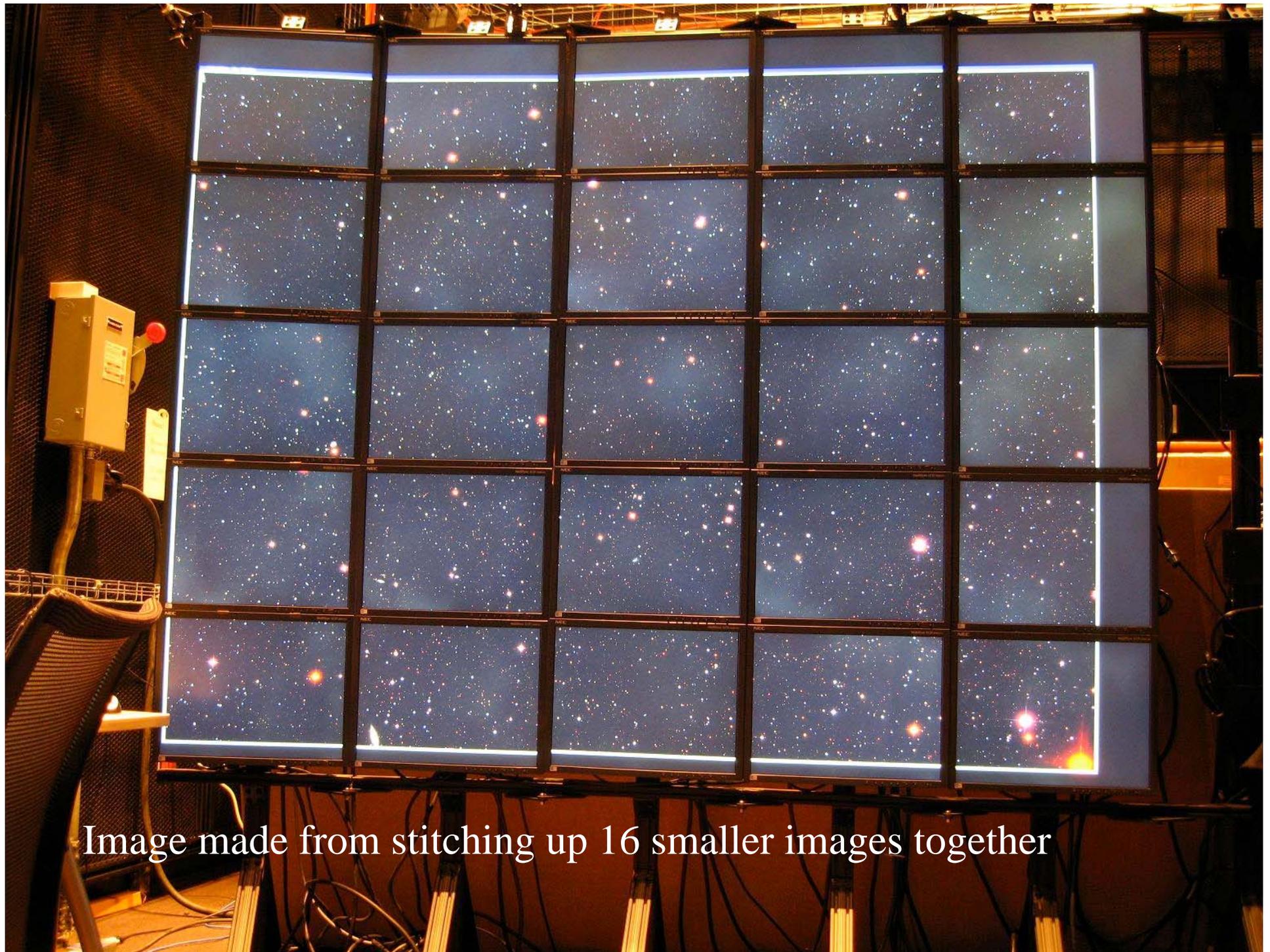


Image made from stitching up 16 smaller images together



Computación Grid

- Y el Grid en Costa Rica?



Computación Grid

- Grupos colaboradores
 - PRAGMA
 - NCSA
 - UCHRI
 - NUCRI
 - CENAT



Computación Grid

- Motivación para la discusión



Computación Grid

- Preguntas?