

Universidad Internacional de Valencia

Procesamiento de Datos Masivos

03MBID

Tema 1: Computación Ubicua – Cloud Computing

Parte I

Yudith Cardinale

Noviembre 2022

Introducción

Cómputo distribuido – ¿Por qué es necesario?

- ★ La demanda de cómputo ha estado siempre por encima del *status quo* tecnológico.
- ★ **Respuestas obvias:**
 - Diversos elementos de cómputo trabajando en armonía para resolver un único problema
 - Necesidad de tener recursos compartidos entre los diferentes elementos de computación
- ★ **Evolución:**
 - Sistemas centralizados (cómputo centralizado) → Sin SO, SO monotarea, monousuario, SO multitareas, multiusuarios, ...
 - Sistemas distribuidos (cómputo distribuido) → Redes, sistemas paralelos, clusters de cómputo, ...
 - Sistemas ubicuos (**cómputo distribuido ubicuo**) → Computación de alto desempeño, middlewares, grid computacionales, **computación en la nube**, Peer-to-Peer, computación voluntaria, IoT, Big Data, ...

COMPUTACIÓN UBICUA

Cómputación ubicua – Computación de alto desempeño

★ Características

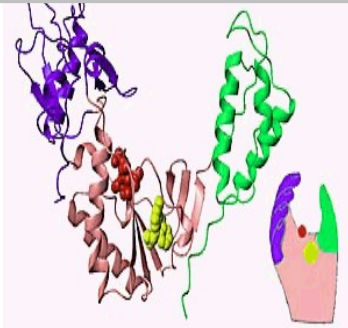
- Implica el uso de ordenadores muy poderosos (super-ordenadores) o **clúster de cómputo** para la resolución de problemas que requieren **mucho cómputo** y **análisis de grandes volúmenes de datos**.
- Está relacionada con programación paralela para usar muchos procesadores simultáneamente.
- **Paralelismo tradicional**: está principalmente asociado con computación para investigación científica.
- **Paralelismo para Big Data**: está principalmente asociado con computación para el negocio, comercio, industria, apoyo a toma de decisiones.

Computación Ubicua

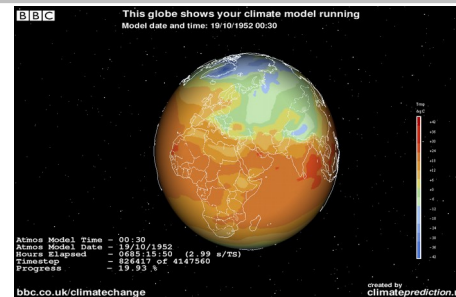
Aplicaciones de alto desempeño - Ejemplos



Teoría de números



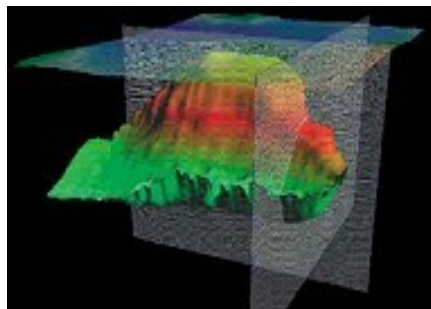
Ciencias de la vida



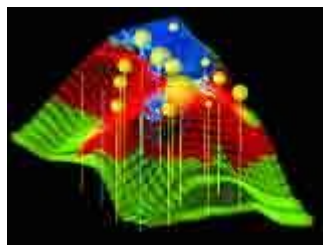
Predicción del tiempo



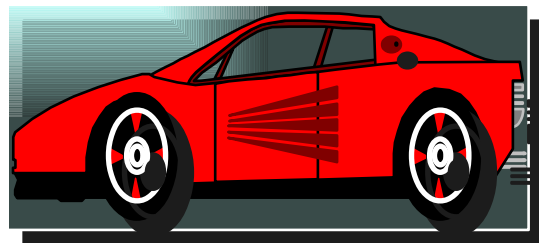
Biología Digital



Petroquímica: dinámica de fluidos, simulación, modelamiento



Internet & e-commerce



CAD/CAM



Aplicaciones Militares



Analítica de Big Data

Arquitecturas de alto desempeño: super-ordenadores

- ★ **Computación Paralela:** Hacer un trabajo **n** veces más rápido usando **n** procesadores
- ★ Requiere arquitecturas de hardware especializadas: clusters, super-ordenadores, top 500 (<https://www.top500.org/>)



SUMMIT, USA; 2,414,592 cores, IBM/NVIDIA



SIERRA, USA; 1,572,480 cores; IBM



Sunway TaihuLight, China; 10,649,600 cores



FRONTERA, USA; 448,448 cores, Dell

Arquitecturas de alto desempeño: super-ordenadores

- ★ **Computación Paralela:** Hacer un trabajo **n** veces más rápido usando **n** procesadores
- ★ Requiere arquitecturas de hardware especializadas: clusters, super-computadores, top 500 (<https://www.top500.org/>)



Fugazu, Japón, 7,630,848 cores (puesto 1)



MareNostrum, España, 153,216 cores, Lenovo (puesto 42)

Arquitecturas de alto desempeño: clusters de cómputo

- ★ **Cluster para balance de carga:** se usan para compartir la carga computacional entre múltiples tareas que pueden o no estar relacionadas, se distribuye la carga de trabajo de manera eficiente entre los procesadores; son usados tanto en ambientes científicos y académicos, como en empresas comerciales;
- ★ **Cluster de alta disponibilidad:** proveen redundancia de datos y servicios (sistemas de respaldo); actúan como granjas de servidores para soportar grandes cantidades de peticiones por parte de millones de clientes (servicios de Google, de Amazon, etc.);
- ★ **Cluster de alto desempeño:** para la ejecución de programas paralelos, usados en ambientes científicos/académicos y para análisis de Big Data.

Computación Ubicua

¡Quiero más!

- ★ Siempre habrá aplicaciones que demandan más potencia que la capacidad de cómputo disponible.
- ★ **Limitaciones de los Clusters:** alto consumo de potencia energética y acondicionamiento del espacio físico limitan la agregación.
- ★ **Solución:** agregar, integrar recursos distribuidos a través de redes de alta velocidad ...

Computación UBICUA

**Uso de la tecnología existente para resolver
problemas más complejos**

Computación Ubicua

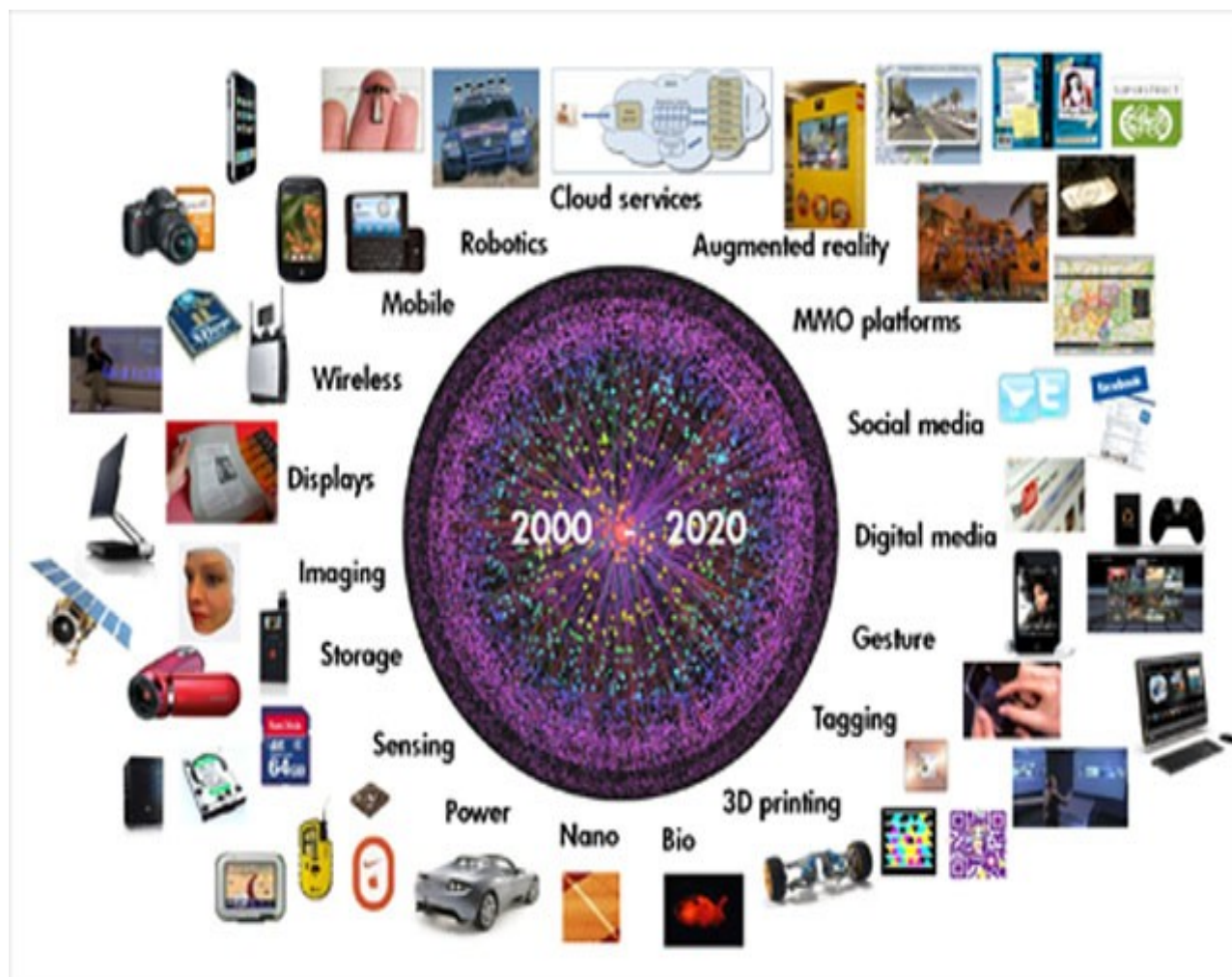
Evolución de la tecnología para computación de alto desempeño:

★ Super-ordenadores

★ Clusters

★ Computación ubicua

- Computación voluntaria,
- *Grid computing*,
- **Cloud computing**,
- IoT,
- **Big Data**
- ...



Computación Ubicua

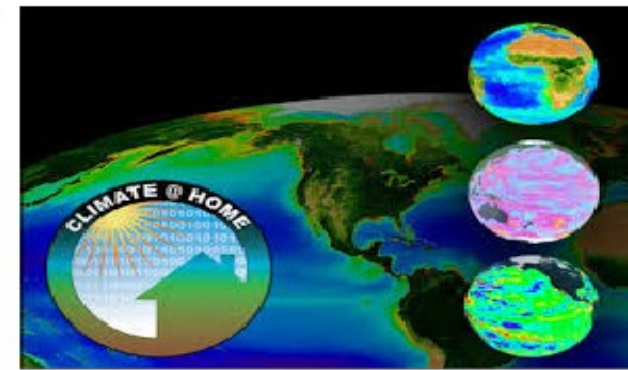
Dos perspectivas

★ Tecnología usada para mejorar la colaboración



Computer Supported Cooperative Work

★ Colaboración usada para mejorar la tecnología y la ciencia

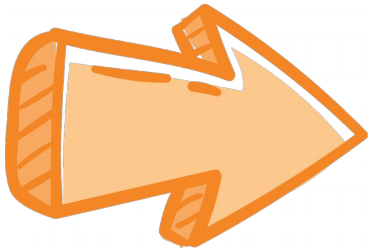


COMPUTACIÓN VOLUNTARIA

Computación Voluntaria

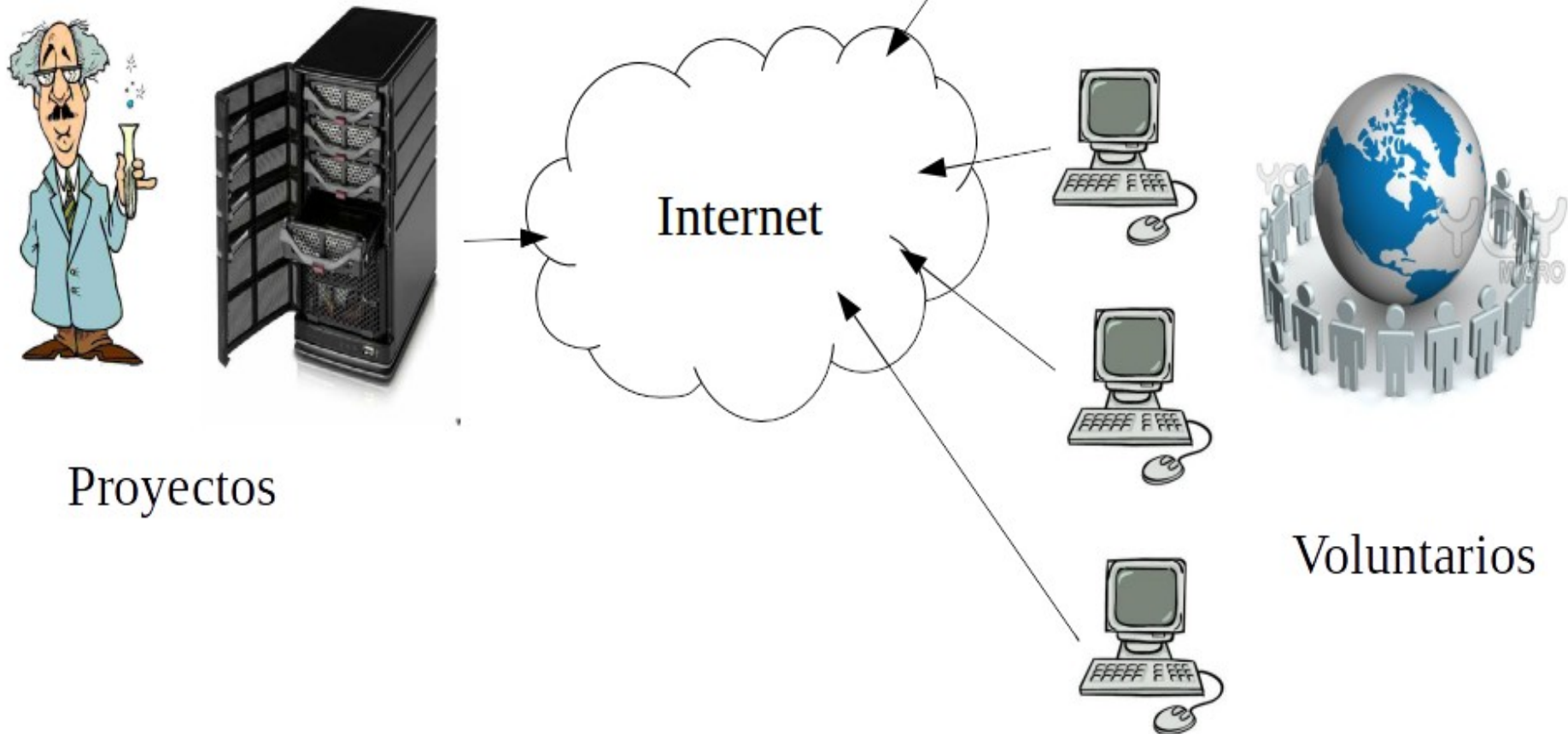
Escenario

- ★ Más de 7 billones de ciudadanos en el mundo
- ★ Muchos tienen acceso a Internet
- ★ Muchos estamos (o podríamos estar) interesados en la ciencia o simplemente colaborar con la ciencia
- ★ Gran capacidad de cómputo y almacenamiento agregado en el mundo
- ★ Alto porcentaje de tiempo ocioso



¿Cómo puede la ciencia aprovechar esta situación?

Computación Voluntaria



Los pioneros

★ GIMPS (Great Internet Mersenne Prime Search) – 1996



Número Primo Mersenne 50th: $2^{77,232,917} - 1$ con más de 23 millones de dígitos (2017)



Número Primo Mersenne 51th: $2^{82,589,933} - 1$ con más de 24 millones de dígitos (2018)

★ Distributed.net – 1997

Área de criptografía:

- En 2002 rompió el RC5 con 64 bits
- Intenta romper RC5 con claves de 72 bits.
- Hasta septiembre 2020, había resuelto el 7% (**más de 100 años para terminar!**)



Computación Voluntaria

- ★ SETI@home (Search for Extraterrestrial Intelligence) – 1999 – se detuvo en marzo 2020 :’(

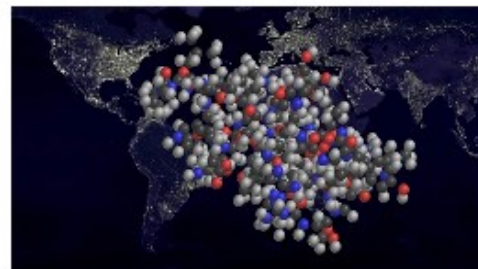
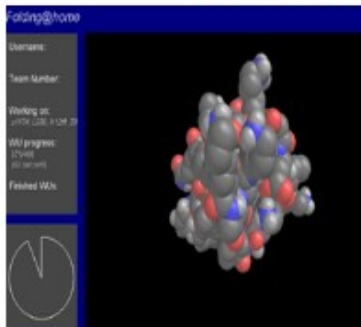


<https://setiathome.berkeley.edu/>

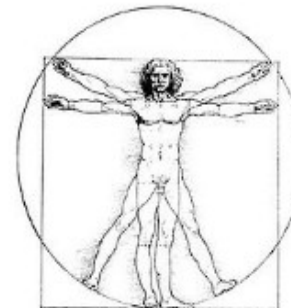


- ★ Folding@home (plegamiento de proteínas) y [genome@home](http://genomeat.home.org/) (proyecto del genoma humano) – 2000

<https://foldingathome.org/>



Proyecto genoma humano
(fin en 2004)



Computación Voluntaria

Más proyectos:

- ★ Climaprediction.net
- ★ Einstein@home (detección de ondas gravitacionales)
- ★ LHC@home (del CERN: Conseil européen pour la recherche nucléaire)
- ★ Rosetta@home (estructuras de proteínas)
- ★ **Boinc** (<https://boinc.berkeley.edu/>)
- ★ **Rendimiento actual:** 500K personas, 1M de computadores, 6.5 PetaFlops. **Potencial:** 1 billón de PCs (hoy — 2018), 2 billones de PCs (2020), GPU, ExaFlops, Exabytes de almacenamiento.



COMPUTACIÓN EN MALLA (grid computing)

Computación en malla



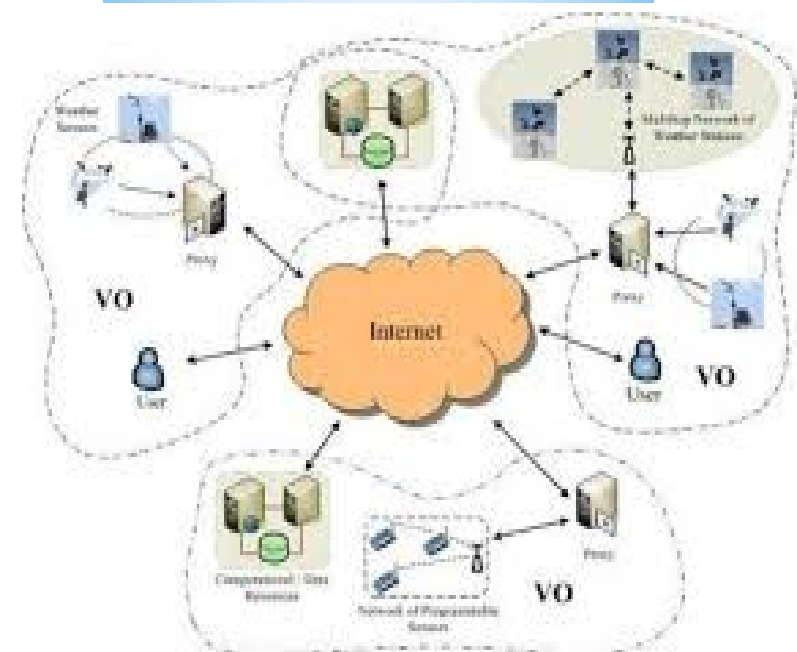
Unificación de recursos distribuidos geográficamente

¿Qué es?

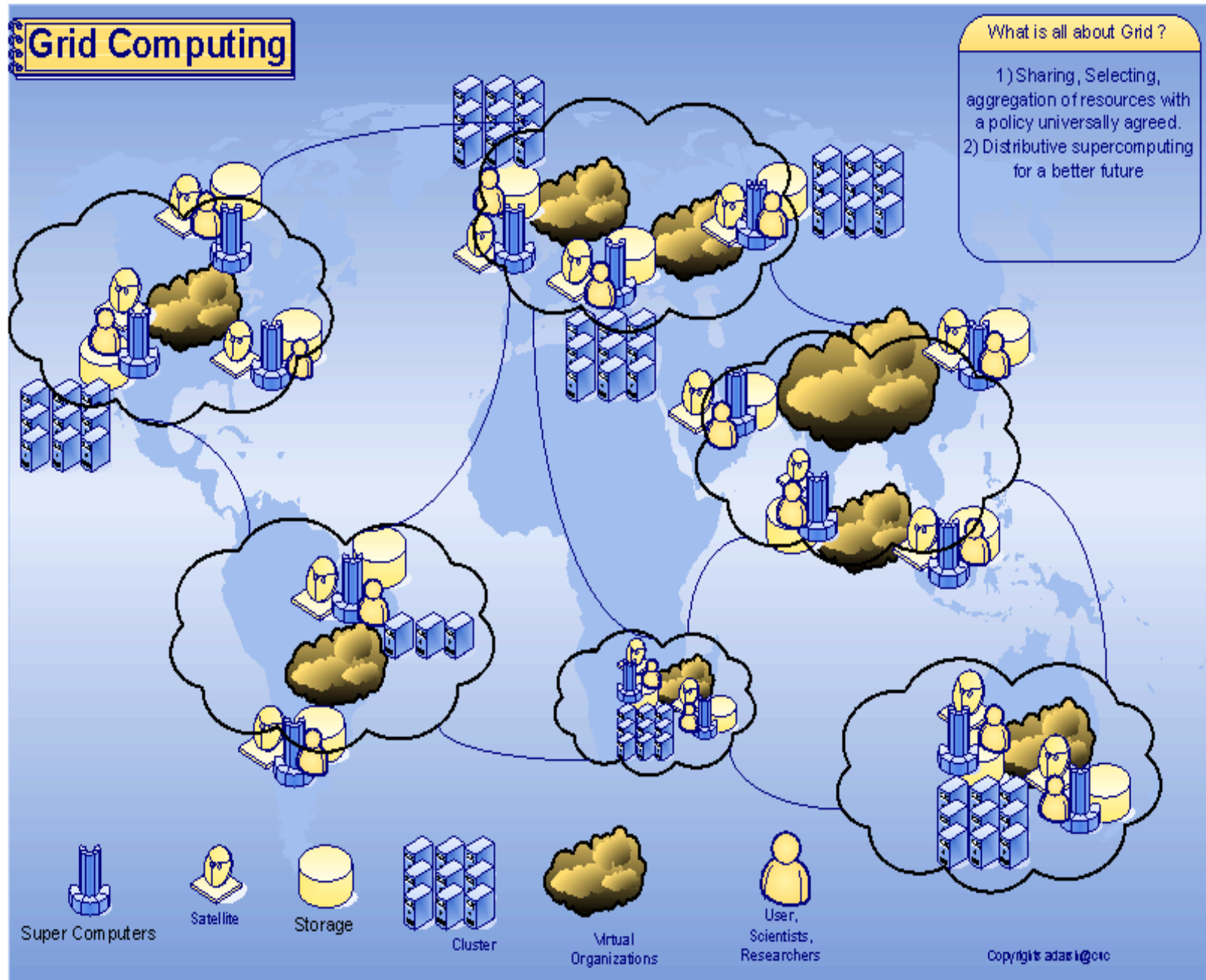
- ★ Infraestructura de hardware y software que provee alto desempeño y alta disponibilidad
- ★ Colección de recursos (personas, computadores, instrumentos y bases de datos) conectados por una red de alta velocidad.
- ★ Mecanismo para que los usuarios puedan usar recursos distribuidos geográficamente de forma transparente, creando la ilusión de un sistema de computación integrado.
- ★ Permite colaboración entre instituciones académicas, de investigación y científicas
- ★ Cada institución ofrece su infraestructura para obtener acceso a otras infraestructuras:
 - Se incrementa la capacidad de cómputo local
 - Mejor uso de los recursos (ejecutar grandes aplicaciones sin tener que invertir en infraestructura)
- ★ Cada institución es responsable de su propia infraestructura (costos, mantenimiento, administración...)

Computación en malla

- ★ Podemos decir que *grid computing* define un modelo de computación social:
 - Todas las instituciones colaboran con objetivos científicos y de investigación comunes
 - Todas las instituciones obtienen beneficios a corto o largo plazo



Computación en malla



COMPUTACIÓN EN LA NUBE

Definiciones

Computación en la nube

Definición de Cloud Computing



Unificación de recursos distribuidos geográficamente PARA LA VENTA

Computación en la nube

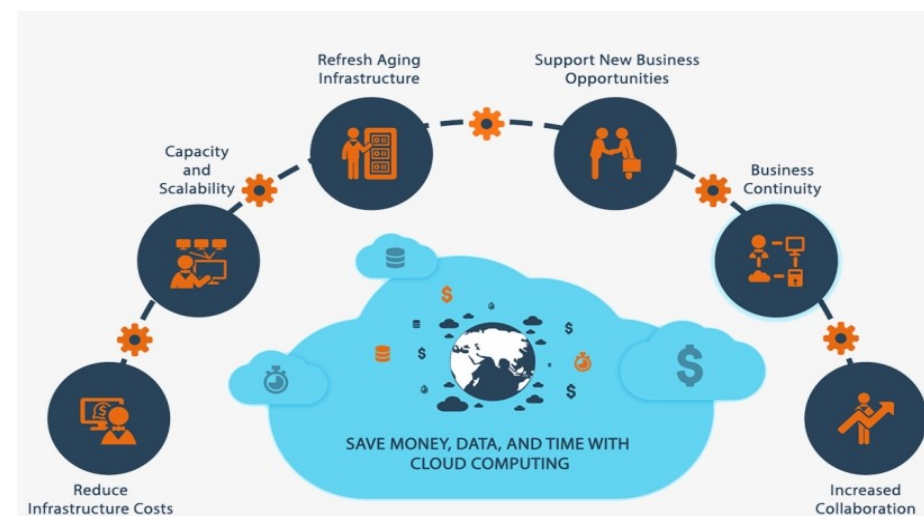
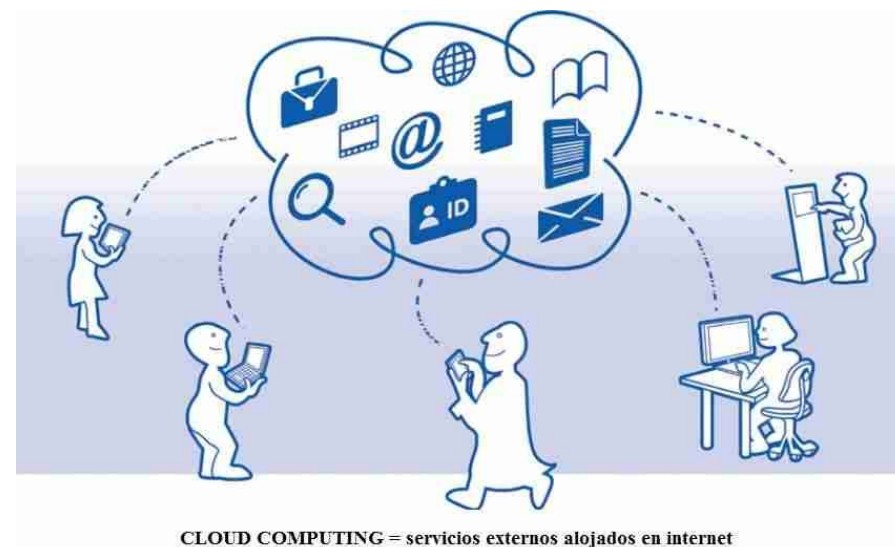
Definición de Cloud Computing

- ★ Permite ofrecer servicios de *hosting* en Internet
- ★ Provee acceso fácil y escalable de recursos de computación y servicios de IT
- ★ Los usuarios no requieren tener conocimientos técnicos, ni experticia, ni control sobre la infraestructura tecnológica de la “nube”
- ★ Los servicios son administrados por los proveedores
- ★ Es un modelo de entrega y consumo de servicios en el que existe **un consumidor** y **un prestador**; el consumidor busca recursos que no posee y el prestador se las proporciona según algunos términos y condiciones.
- ★ Esta relación, entre el que quiere un recurso para hacer su actividad y el que se lo presta, es un acceso directo a las capacidades de consumir, escalabilidad, pago por uso, transparencia en las condiciones de uso, de términos y condiciones y de esta forma poder acceder a ellas de manera muy rápida.

Computación en la nube

Definición de Cloud Computing

- ★ Podemos decir que *cloud computing* define un modelo de computación comercial:
 - Cada participante obtiene beneficios individuales con un mínimo de inversión
 - Los proveedores obtienen beneficios (\$\$\$) de todos los participantes



Computación en la nube

Implicaciones de cloud computing

- ★ Los recursos son propiedad del proveedor
 - Pagas lo que usas
 - No administras hardware ni software
- ★ Los usuarios no requieren tener conocimientos técnicos, ni experticia, ni control sobre la infraestructura tecnológica de la “nube”
- ★ Los servicios son administrados por los proveedores
- ★ Heterogeneidad
 - A todos los niveles (sw, hw, so, lenguajes, arquitecturas, etc.)
- ★ Alta disponibilidad de recursos
- ★ Provee tolerancia a fallas
- ★ Escala fácilmente
- ★ Fácil de usar (clientes)