

World Community Grid

 almost **620,000**
volunteers

in **80**
countries 

using
over **2.3** 
million PCs

By pooling resources, IBM
World Community Grid PCs
have performed computations
in the equivalent of over



730,000
years

to help
process **21** 
research projects

Statistics reflect data as of May 15, 2013.





IIC2523 - Sistemas Distribuidos
Escuela de Ingeniería
Departamento de Ciencias de la Computación
2017-2

Grid/Volunteer Computing

Sebastián Amenábar
Raimundo Herrera

¿Qué es?

- Básicamente compartir poder de cómputo
- > 1 billón de computadores en el mundo
- Volunteer: Anonimidad - Plug&Play
- Grid: Responsabilidad - Configuración

World Community Grid



Statistics reflect data as of May 15, 2013.



2. Select projects

Select all projects

- | | |
|--|---|
| <input type="checkbox"/> Microbiome Immunity Project |  |
| <input type="checkbox"/> Smash Childhood Cancer |   |
| <input type="checkbox"/> OpenZika |   |
| <input type="checkbox"/> Help Stop TB |  |
| <input type="checkbox"/> FightAIDS@Home - Phase 2 |  |
| <input type="checkbox"/> Outsmart Ebola Together |   |
| <input type="checkbox"/> Mapping Cancer Markers |  |
| <input type="checkbox"/> FightAIDS@Home |   |

You must select at least one project before you can continue

Next >

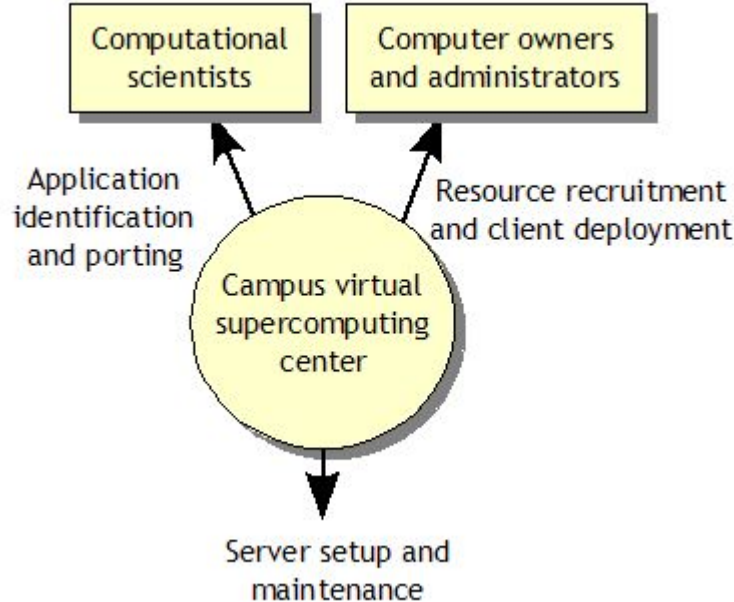
Usar BOINC

- Registrarse en algún sitio como:
<https://www.worldcommunitygrid.org/>
- Descargar un cliente de BOINC (multiplataforma)
- Elegir la causa a la que se quiere ayudar.

Programar BOINC

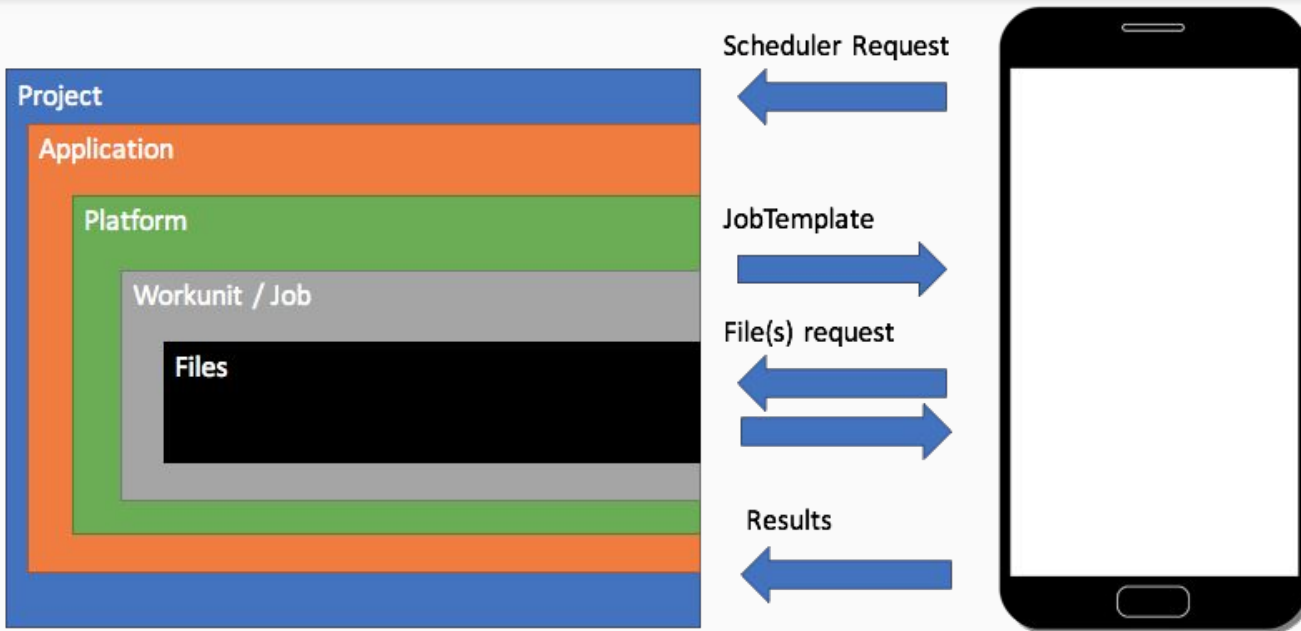
- Descargar el código fuente de una “example_app” desde el github oficial:
<https://github.com/BOINC/boinc>
- Administración mediante Remote Procedure Call (RPC)

VCSC



“The initial hardware cost is roughly \$5 million, with ongoing yearly energy and maintenance costs of at least \$1 million. The VSCS, in contrast, has hardware costs of about \$10K.”

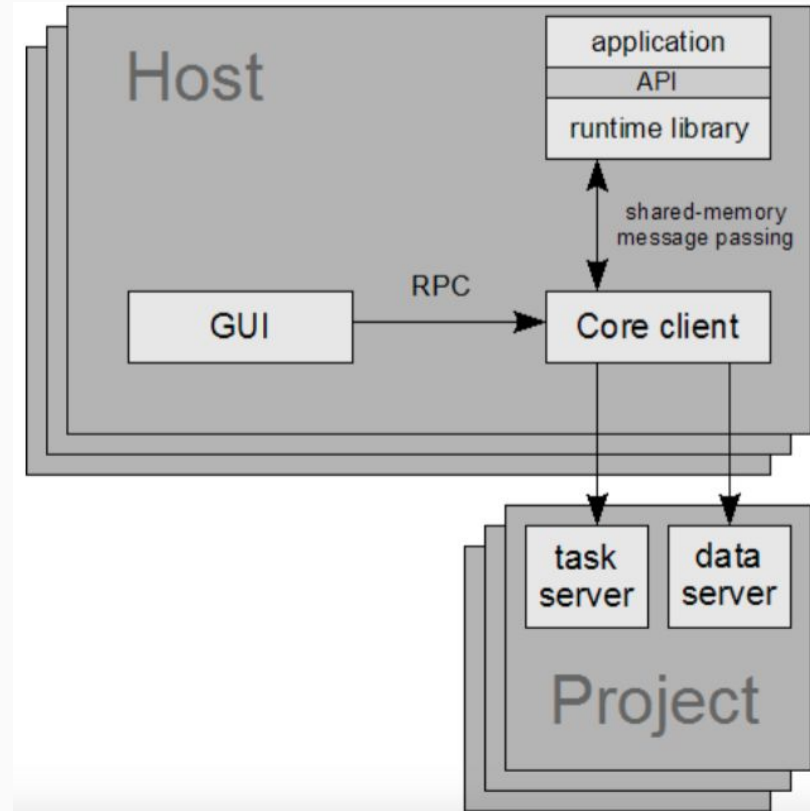
Berkeley Open Infrastructure for Network Computing



Lado del Servidor

- Manejo mediante RPC y XMLs
- Directory tree:
 - Applications:
 - Versions
 - Platforms
 - Files
 - Inputs
 - Outputs
- MySQL database:
 - platform
 - app
 - app_version
 - user
 - host
 - workunit
 - result

Lado del Cliente



Scheduling

Primeras versiones:

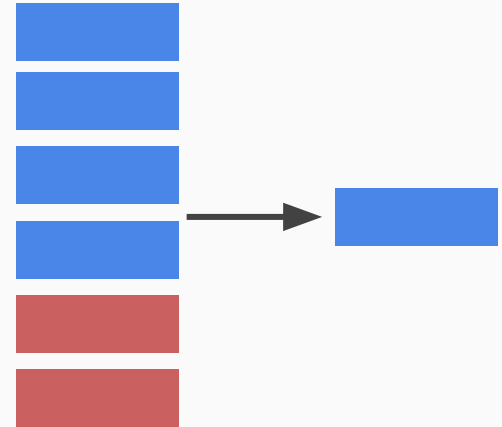
- Weighted Round Robin entre proyectos: recursos a utilizar.
- Mantener suficiente trabajo para utilizar el *ConnectionInterval*, al menos un job por proyecto.

Situación Actual

- Uso de *batches* para minimizar la transferencia, principio de localidad: Diferentes jobs pueden usar el mismo archivo, diferentes archivos pueden ser usados por un mismo job.
- Completar *batches* rápidamente.
- Usuarios con *throughput* similar (en base a tiempo, memoria, FLOPS) son asignados a un mismo *batch* en lo que se llama un *puntero*.

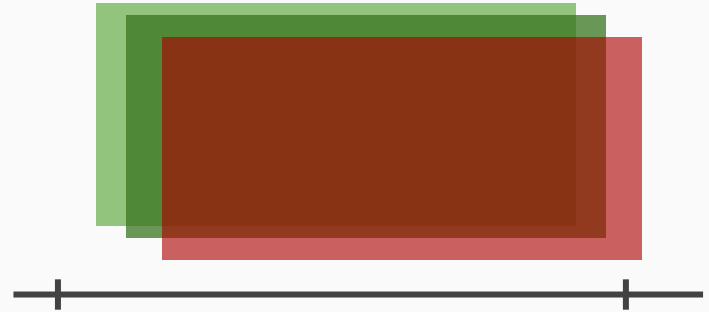
Validación: replicación

- Sin replicación:
 - Cada trabajo solamente se hace una vez
- *N*-Replicación:
 - *N* hosts realizan el trabajo y un resultado es válido si es generado por la mayoría absoluta.
- Replicación adaptativa:
 - Hay hosts más confiables, por lo que se necesitan menos validaciones de sus resultados. Puede usar los métodos siguientes para validar.



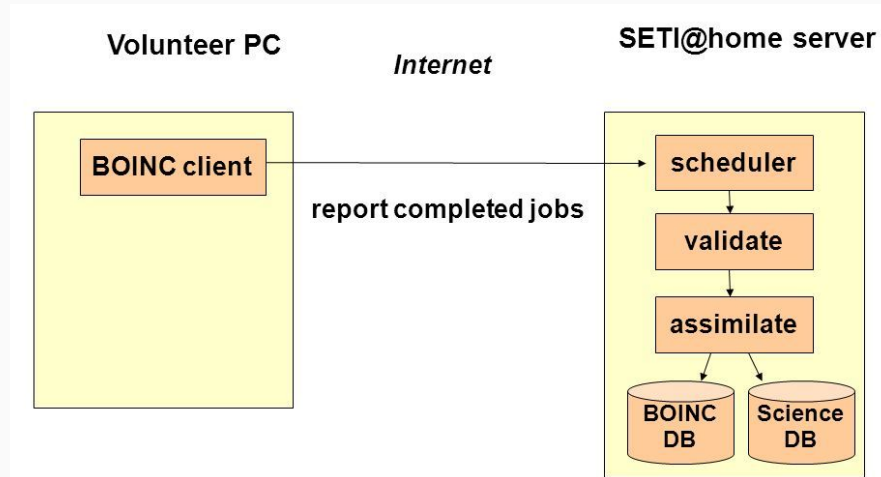
Validación: precisión

- Problemas de precisión
 - Usar librerías y opciones del compilador para asegurar precisión absoluta.
 - Fuzzy-comparison: función comparadora que tolera cierto nivel de diferencia.
 - Replicación homogénea: Jobs son realizados por hosts “numéricamente equivalentes”



Asimilación

- Programa específico para cada aplicación que sube los resultados a un directorio permanente o los sintetiza en una base de datos.

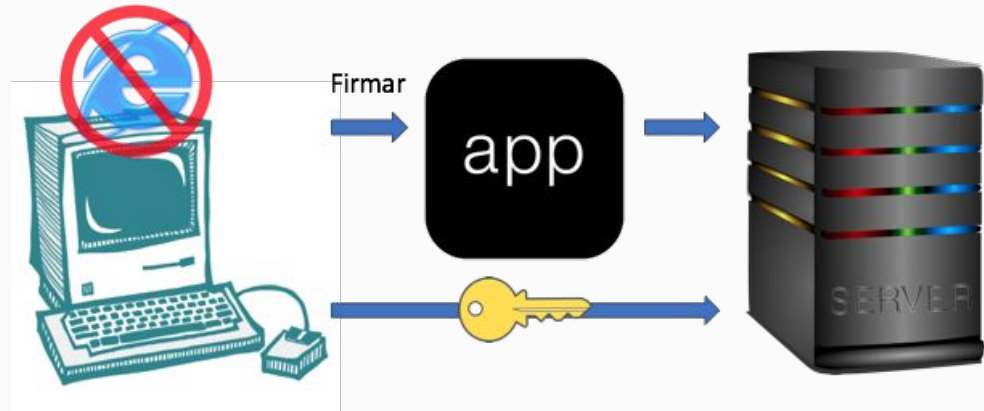


Seguridad

- Result and credit falsification
- Malicious executable distribution
- Denial of server attacks on data servers: Tamaño de output máximo y firma de subida.
- Robo de datos desde el servidor del proyecto: Responsabilidad de cada proyecto.

Prevención de Malware

- Mediante llaves RSA
- Computador Off-the-Grid firma las aplicaciones.
- Servidor contiene la llave pública y la comparte con los *hosts*.



Seguridad

- Sniffing de datos del usuario: Sin protección
- Robo de archivos de proyectos
- Abuso intencional sobre los clientes:
 - Impedido por *sand-boxing* mediante usuario especial.
- Abuso accidental sobre los clientes:
 - BOINC tiene medidas preventivas; aplicaciones que usan mucha memoria o CPU pueden ser abortadas.

¿Dudas?

Principal: <https://boinc.berkeley.edu/trac/wiki/ProjectMain>
Scheduling: https://boinc.berkeley.edu/boinc_papers/sched/paper.pdf
Batch Schedule: <https://boinc.berkeley.edu/trac/wiki/LocalityNew>
Validación: <https://boinc.berkeley.edu/trac/wiki/ValidationSummary>
Seguridad: <https://boinc.berkeley.edu/trac/wiki/SecurityIssues>
Firmado: <https://boinc.berkeley.edu/trac/wiki/CodeSigning>
Volunteer: <https://boinc.berkeley.edu/trac/wiki/VolunteerComputing>
VCSC: <https://boinc.berkeley.edu/trac/wiki/VirtualCampusSupercomputerCenter>