

# DESKTOP GRID Y COMPUTACIÓN VOLUNTARIA: INTRODUCCIÓN A LA TECNOLOGÍA BOINC

José Luis Guisado Lizar

<http://cum.unex.es/profes/profes/jlguisado>

Grupo de Evolución Artificial (GEA)

Universidad de Extremadura



JORNADA Y TALLER SOBRE APLICACIÓN PRÁCTICA DE LA TECNOLOGÍA GRID  
SEMANA DE LA CIENCIA 2006

Centro Universitario de Mérida, Universidad de Extremadura, Noviembre de 2006



# Esquema general:

## 1 - Introducción a BOINC

- Computación voluntaria con BOINC
- Proyectos
- Desktop grid (computación institucional)

## 2 - Estructura de la plataforma BOINC

- Un proyecto BOINC
- Componentes de la infraestructura BOINC
- Aplicaciones adecuadas para BOINC

# 1 - INTRODUCCIÓN A BOINC

# ¿Qué es BOINC?

## BOINC = BERKELEY OPEN INFRASTRUCTURE FOR NETWORK COMPUTING

- Infraestructura de software para realizar computación distribuida mediante el **paradigma "Desktop Grid"**
  - Utilizando un conjunto heterogéneo de computadores interconectados a través de una red del tipo Internet.
- Dos posibles tipos de uso:
  - **Computación voluntaria**, mediante computadores del público
  - **"Desktop grid" o Computación institucional**, mediante computadores de una institución:
    - Universidades: Virtual Campus Supercomputing Center
    - Empresas

# Computación voluntaria con BOINC

- **Uso del tiempo de proceso en el que el computador está ocioso para realizar cálculos de un proyecto científico**
- Uso: **tiempo de proceso**, **memoria** y **espacio en disco** donados por usuarios públicos o instituciones
- **Motivación**: el usuario siente haber realizado una contribución a un importante desarrollo científico. Estímulos:
  - Sitio web divulgativo presentando una descripción interesante de la investigación científica que se persigue
  - Salvapantallas que muestra el estado del cálculo que se está realizando
  - El usuario recibe un nº de "créditos" proporcional al tiempo de cálculo con que ha contribuido su PC.  
Ranking: en la web del proyecto
- **Distintos sistemas operativos** posibles: Windows, Linux, Mac, Solaris.

# ¿Cómo participar en un proyecto de computación voluntaria?

## ■ **Participación:**

- Elegir un proyecto
- Visitar el sitio web del proyecto y crear una cuenta de usuario indicando una dirección de correo electrónico
- Descargar y ejecutar el software cliente de BOINC

## ■ Se puede participar simultáneamente en más de un proyecto:

- El usuario controla el porcentaje de potencia de cómputo que dedicará a cada uno
- Así, su computador realizará cálculos incluso cuando un proyecto no tenga trabajos

# Proyectos de computación voluntaria con BOINC

- Proyecto inicial: **SETI@home**:
- **SETI = Search for Extraterrestrial Intelligence**: Disciplina científica que trata de detectar vida inteligente fuera de la Tierra.
- **Objetivo**: SETI@home analiza señales captadas por el radiotelescopio de Arecibo en busca de **señales con ancho de banda estrecho** procedentes del espacio. Como no se conoce ningún mecanismo natural capaz de producir este tipo de señales, una detección aportaría evidencias de tecnología extraterrestre.
- Proyectos previos usaban supercomputadores para el análisis de datos
- En 1995, David Gedye propuso utilizar un "supercomputador virtual" formado por un gran número de computadores conectados a Internet y organizó el proyecto SETI@home, que se inició en 1999
- **BOINC** es la plataforma de software desarrollada originalmente para SETI@home, que fue liberada para poder ser usada en cualquier otro proyecto



# Algunos proyectos de computación voluntaria con BOINC

- Ciencias de la tierra:
  - [Climateprediction.net](http://Climateprediction.net)
  - [BBC Climate Change Experiment](http://BBC Climate Change Experiment)
  - [Seasonal Attribution Project](http://Seasonal Attribution Project)
- Matemáticas y juegos de estrategia:
  - [SZTAKI Desktop Grid](http://SZTAKI Desktop Grid)
  - [Riesel Sieve](http://Riesel Sieve)
  - [Chess960@home](http://Chess960@home)
  - [Rectilinear Crossing Number](http://Rectilinear Crossing Number)
- Astronomía / física / química:
  - [SETI@home](http://SETI@home)
  - [Quantum Monte Carlo at Home](http://Quantum Monte Carlo at Home)
  - [Einstein@home](http://Einstein@home)
  - [LHC@home](http://LHC@home)
  - [SpinHenge@home](http://SpinHenge@home)
- Biología y medicina:
  - [Predictor@home](http://Predictor@home)
  - [SIMAP](http://SIMAP)
  - [Tanpaku](http://Tanpaku)
  - [World Community Grid](http://World Community Grid)
  - [Malariacontrol.net](http://Malariacontrol.net)
  - [Rosetta@home](http://Rosetta@home)

# Computación voluntaria con BOINC

## ■ **Potencia de cómputo:**

- Limitada en cada equipo
- Enorme si el nº de contribuyentes es muy grande
- Ejemplo: seti@home, datos 9 nov 2006:
  - 1,472,537 computadores
  - Potencia sostenida: 484 TFlops
  - Comparación con el supercomputador nº1 del Top 500: 281 TFlops (eServer Blue Gene Solution)

## ■ **Factor de calidad de los resultados:** No tan alto como en un cluster dedicado, pero resoluble sacrificando algo de potencia de cómputo:

- Se deben prevenir los resultados erróneos (debidos a fallos del computador, interferencias maliciosas, redes lentas, etc).
- Mecanismos:
  - **Computación redundante** ⇒ reducción de la potencia de cómputo pura esperada
  - **Firma del código** para prevenir distribución maliciosa de ejecutables
  - La aplicación y todos los archivos requeridos se sitúan en una "**sandbox**": para poder acceder a ellos, hay que resolver sus nombres físicos usando la API de BOINC

# Desktop grid (computación institucional) con BOINC

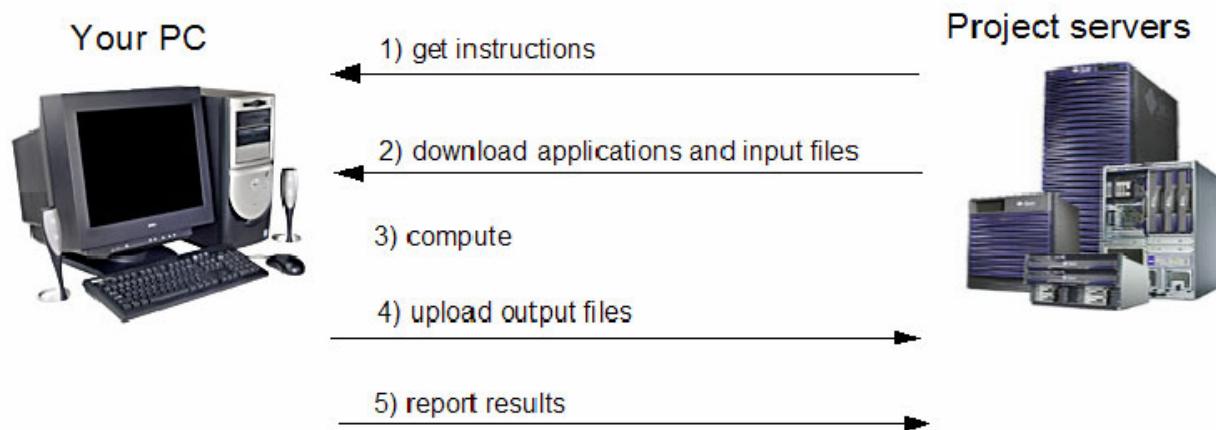
- **Uso de computadores propios de una cierta institución.** Posibilidades:
  - Uso dedicado (100% del tiempo de proceso)
  - Tiempo de proceso en que el computador está ocioso
  - Ejecución con cierta prioridad (en clientes Linux)
- **Universidades: Virtual Campus Supercomputing Center**
  - Puede proporcionar a los investigadores la potencia computacional de un gran cluster por una fracción pequeña de su coste, o con un coste muy pequeño utilizando el tiempo en que los equipos están ociosos
  - VCSC = Proyecto BOINC con aplicaciones de los investigadores, al que se conectan los equipos locales
  - Potencia de computación: suministrada por:
    - PCs del campus
    - PCs del personal
    - Máquinas de los laboratorios de computación
    - PCs personales de los alumnos (en sus casas)
  - Ejemplo: VSCS con 2.000 PCs, corriendo una media del 50% del tiempo:
    - Potencia computacional: equivalente a un cluster de 1.000 nodos.  
Coste inicial de HW  $\approx 800 \text{ €} \times 1000 \text{ nodos} = 800.000 \text{ €}$
    - Coste del VSCS: Servidor  $\approx 2.000 \text{ €}$
- **Empresas**

# **2 - ESTRUCTURA DE LA PLATAFORMA BOINC**

# El proyecto BOINC

- Unidad básica: el **proyecto BOINC**:
  - Grupo de una o más aplicaciones distribuidas que utilizan la plataforma BOINC, manejadas por una cierta organización
  - Cada proyecto BOINC se identifica por una única **URL maestra** (la página de inicio de su sitio web)
  - Computadores cliente: se registran en los proyectos mediante una **cuenta** que se obtiene en su web

## ■ Interacción entre cliente y servidor:



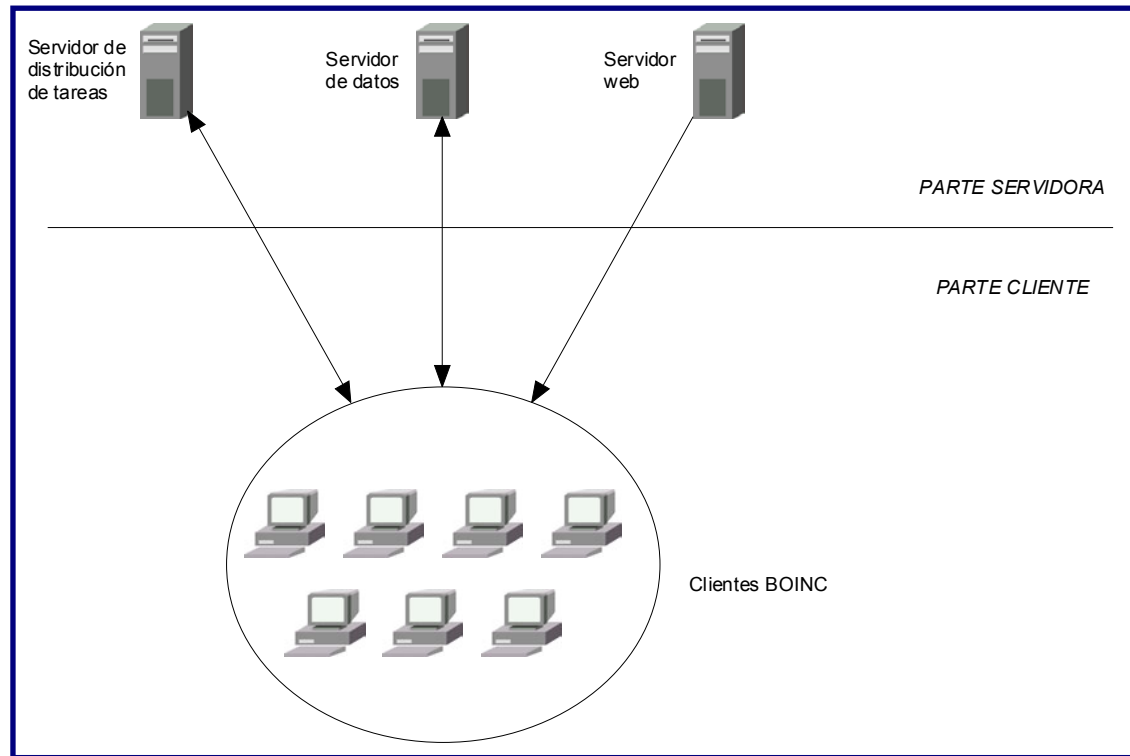
# Infraestructura BOINC

## ■ Parte servidora

- En un único servidor o en varios (parte de la funcionalidad cada uno)
- Consta de una base de datos relacional y un conjunto de servicios web y procesos demonio:
  - Servidor de distribución de tareas
  - Servidor de datos
  - Servidor web

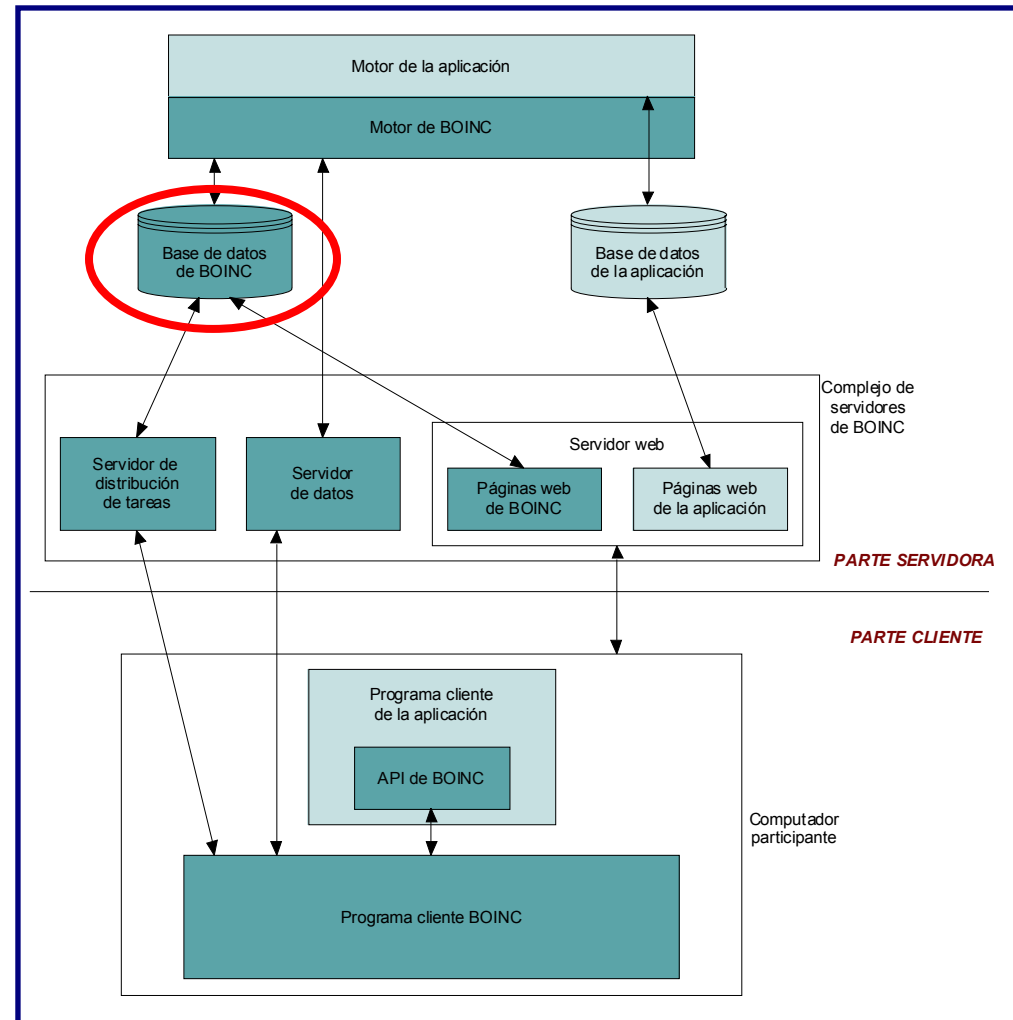
## ■ Parte cliente

- Software a instalar en cada computador cliente



# Componentes de la infraestructura BOINC (1)

- **Componentes con fondo oscuro:** proporcionados como parte del sistema BOINC
- **Componentes con fondo claro:** deben ser desarrollados por los gestores de la aplicación que se pretenda ejecutar
- **Base de datos de BOINC:**
  - Almacena descripciones de todos los componentes del proyecto:
    - Aplicaciones
    - Plataformas
    - Versiones
    - Resultados
    - Cuentas
    - Equipos ...



# Componentes de la infraestructura BOINC (2)

## ■ Servidor/es de distribución de tareas:

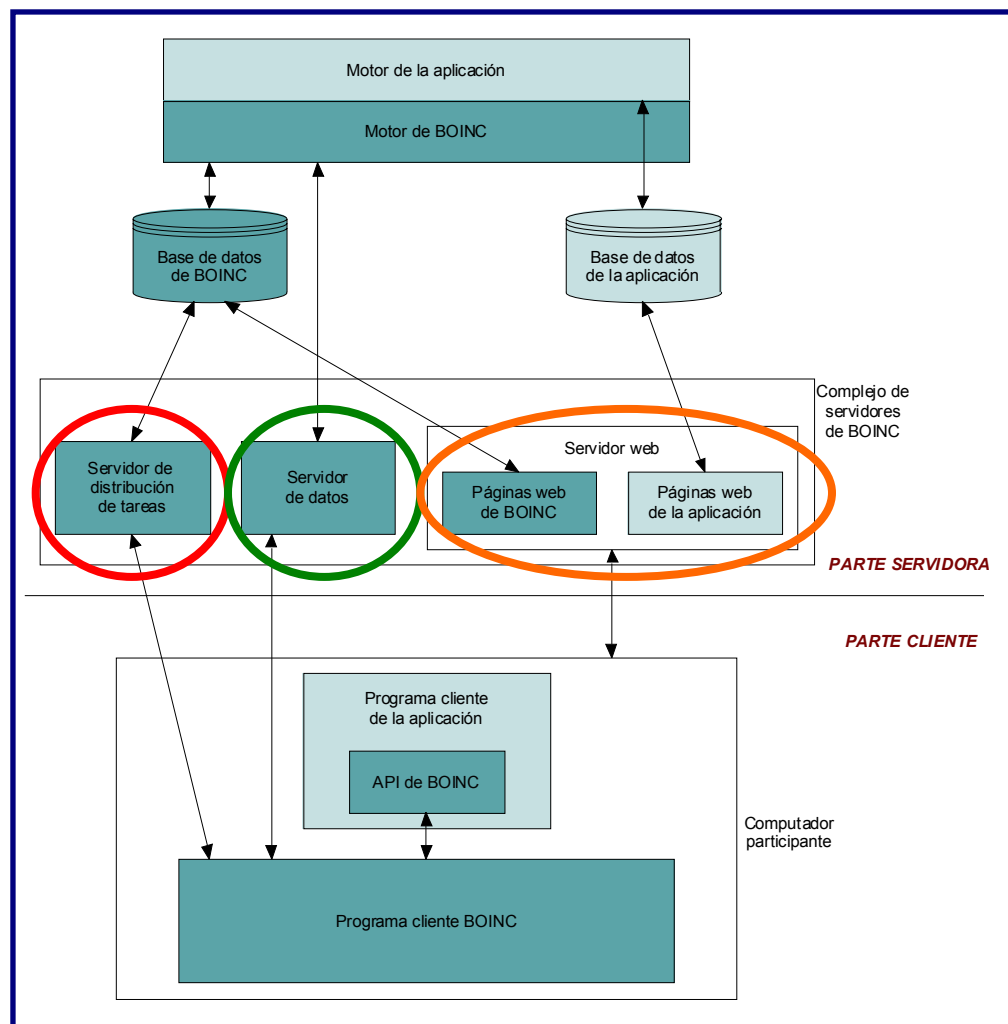
- Envío de trabajos a los clientes
- Recepción de informes acerca de los trabajos ya completados

## ■ Servidor/es de datos:

- Envío de archivos de datos de entrada para la ejecución de los trabajos a los clientes
- Recepción de archivos de datos con resultados de los trabajos completados desde los clientes

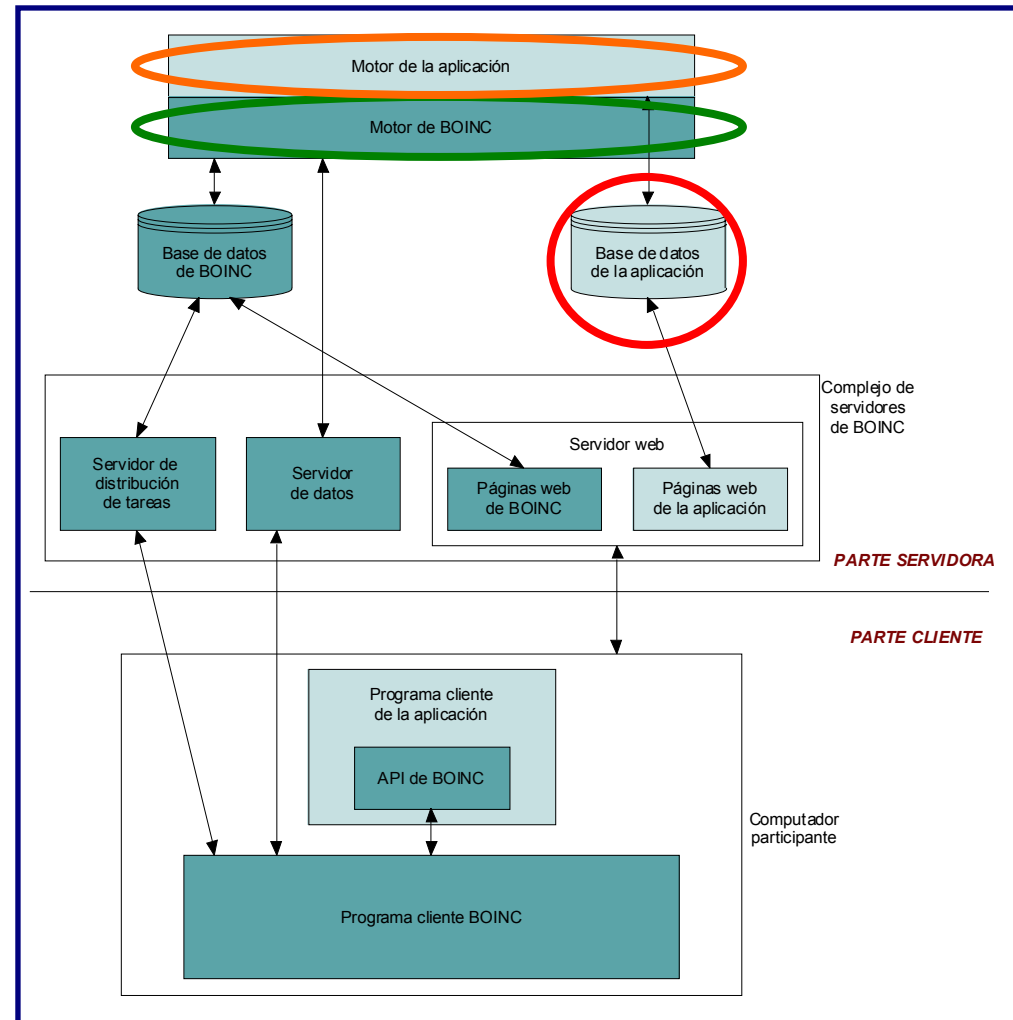
## ■ Servidor web:

- Proporciona interfaz del proyecto hacia los clientes. Incluye:
  - Págs. web estándar del sistema BOINC
  - Págs. web dependientes de la aplicación (opcional)



# Componentes de la infraestructura BOINC (3)

- **Base de datos de la aplicación** (opcional):
  - Almacenar datos propios de las págs. web de la aplicación
- **Software motor de BOINC:**
  - Controla todos los componentes de la parte servidora
- **Software motor de la aplicación** (opcional):
  - Controla funciones específicas de la aplicación
- Todos los servidores pueden estar situados físicamente en una misma máquina



# Componentes de la infraestructura BOINC (4)

## ■ Parte cliente de BOINC:

- Recepción de trabajos y archivos de entrada desde los servidores del proyecto
- Ejecución de los trabajos
- Envío a los servidores de los archivos de datos con los resultados

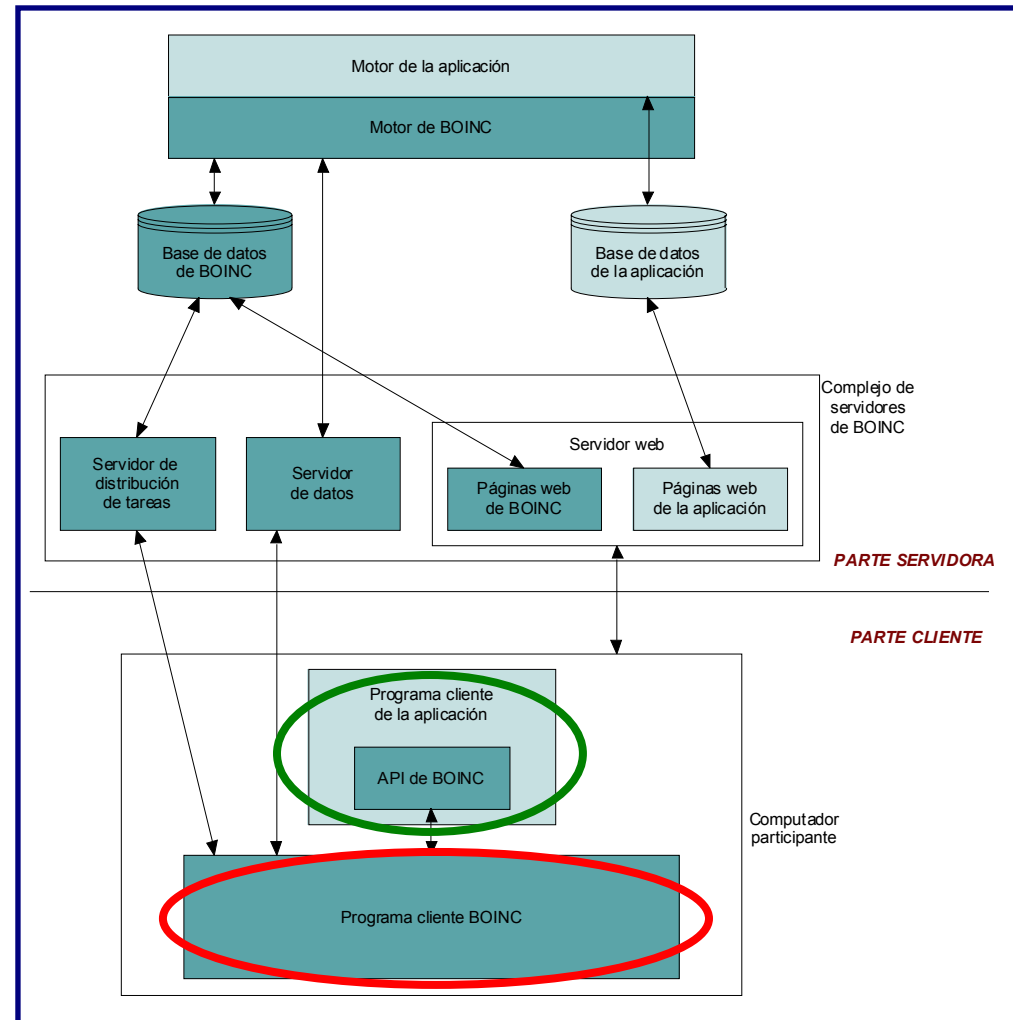
## ■ Programa cliente BOINC:

- Controla la operación de la parte cliente

## ■ Programa cliente de la aplicación:

- Realiza la computación propia de la aplicación en cada computador cliente

- Ambos se comunican usando la **API de BOINC**



# Aplicaciones adecuadas para BOINC

- Las **aplicaciones más adecuadas para la infraestructura BOINC** son las que cumplen las **condiciones**:
  - **Alta ratio computación - comunicación**: los datos de entrada y salida se envía a través de conexiones de Internet, que pueden ser muy lentas. Especialmente para computación voluntaria
  - **Paralelismo independiente**: debe ser divisible en fragmentos paralelos independientes o con muy pocas dependencias de datos. Especialmente para computación voluntaria (los usuarios pueden conectarse esporádicamente y con conexiones muy lentas)
  - **Grandes requerimientos de**:
    - Computación
    - Almacenamiento
    - (O ambos)
  - **Tolerancia a fallos**: No se puede asumir automáticamente que un resultado devuelto por un computador de uso público (o semi-público) es correcto. El uso de computación redundante reduce la probabilidad de error.
  - Las aplicaciones para **computación voluntaria** deben tener **atractivo para el público**.

# Información adicional sobre BOINC

1. Sitio web oficial de BOINC: [boinc.berkeley.edu](http://boinc.berkeley.edu)
2. Wiki no oficial de BOINC: [boinc-wiki.ath.cx](http://boinc-wiki.ath.cx)
3. Información sobre BOINC en la Wikipedia:  
[en.wikipedia.org/wiki/Berkeley\\_Open\\_Infrastructure\\_for\\_Network\\_Computing](http://en.wikipedia.org/wiki/Berkeley_Open_Infrastructure_for_Network_Computing)
4. The 2nd Pan-Galactic BOINC Workshop:  
[boinc.berkeley.edu/workshop\\_06.php](http://boinc.berkeley.edu/workshop_06.php)