

Gridificación de LCAOMIN

Joint CHAIN/GISELA/EPIKH Workshop

Daniel Bellomo¹ Antonio Juan Rubio Montero²
Manuel Aurelio Rodriguez Pascual²

¹Universidad Nacional de Río Cuarto (UNRC)
Unidad de Tecnología de la Información (UTI)

²Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
(CIEMAT)

10 de Diciembre de 2010, Valparaíso (Chile)



Facultades

- Agronomía y Veterinaria
- Ciencias Económicas
- Ciencias Exactas, Físico-Químicas y Naturales
- Ciencias Humanas
- Ingeniería

Río Cuarto, Córdoba, Argentina, www.unrc.edu.ar



Director general Dr. Oscar Taurian



Director general Dr. Oscar Taurian

Antecedentes en Grid computing

- Inicio de actividades en el 2007
- Alumnos y tutores en actividades de EELA-2
- Escuela de grid de EELA-2: aplicación AEROVANT
- Aplicaciones: Física, Matemática, Ingeniería
- Colaboración en la implementación de sitios grid en instituciones de la región



Director general Dr. Oscar Taurian

Antecedentes en Grid computing

- Inicio de actividades en el 2007
- Alumnos y tutores en actividades de EELA-2
- Escuela de grid de EELA-2: aplicación AEROVANT
- Aplicaciones: Física, Matemática, Ingeniería
- Colaboración en la implementación de sitios grid en instituciones de la región

Recursos técnicos y humanos

- Cluster
- Sitio Grid (en proceso de integración)
- Cluster y Grid staff



Proyectos Involucrados en la UNRC

- Implementación de modelos matemáticos
- Parámetros magnéticos y estructura molecular. Aplicaciones al estudio de compuestos orgánicos de interés alimentario y ambiental

mas de 10 personas involucradas

mas 10 años de investigación y desarrollo



Implementación de modelos matemáticos

Director

- Dr. Juan Cesco

Integrantes

- Mg. Claudia Denner
- Ing. Guillermo Frascetti
- Mg. Ana Rosso
- Mg. Carlos Bagetta

Alumna

- Carmina Alturria

Asesor

- Dr. Jorge Pérez

Secretaría de Ciencia y Técnica, Dpto. Matemática, Facultad de Ciencias Exactas, Físico-Químicas y Naturales, UNRC



Parámetros magnéticos y estructura molecular. Aplicaciones al estudio de compuestos orgánicos de interés alimentario y ambiental

Director

- Dr. Oscar Taurian

Integrantes

- Dr. Félix Ortiz
- Ing. Mario Romero
- Dr. Jorge Pérez

Secretaría de Ciencia y Técnica Dpto. Física, Facultad de Ciencias Exactas, Físico-Químicas y Naturales, UNRC



Descripción del problema

En Física Molecular es muy utilizado el método de Hartree-Fock ^a para obtener funciones de onda electrónicas aproximadas. El programa LCAOMIN es la versión piloto (no publicada hasta la fecha) para chequear el grado de aproximación (Hartree-Fock) que se puede llegar a obtener utilizando solamente orbitales atómicos 1s. El programa está permitiendo utilizar distintos modelos de bases atómicas para predecir, en una primera etapa, la geometría molecular.

Este problema tiene, inherentemente, alto costo computacional debido a la gran cantidad de átomos que puede llegar a tener la molécula en estudio.

^aA. Szabo-N. Ostlund: Modern quantum chemistry: Introduction to Advanced Electronic Structure Theory (Macmillan Publishing Co, New York)



- 1 Debido que el programa calcula la geometría molecular óptima se debe especificar la molécula a considerar



Algoritmo de cálculo

- 1 Debido que el programa calcula la geometría molecular óptima se debe especificar la molécula a considerar
- 2 Se calculan cuatro matrices: $S_{\mu\nu}$, $H_{\mu\nu}$, $(\mu\nu|\sigma\lambda)$, $F_{\mu\nu}$; donde $F_{\mu\nu}$ depende de $H_{\mu\nu}$, $(\mu\nu|\sigma\lambda)$. (Los índices van de 1 a n, por lo cual el cálculo escala aproximadamente como n^4)



- 1 Debido que el programa calcula la geometría molecular óptima se debe especificar la molécula a considerar
- 2 Se calculan cuatro matrices: $S_{\mu\nu}$, $H_{\mu\nu}$, $(\mu\nu|\sigma\lambda)$, $F_{\mu\nu}$; donde $F_{\mu\nu}$ depende de $H_{\mu\nu}$, $(\mu\nu|\sigma\lambda)$. (Los índices van de 1 a n , por lo cual el cálculo escala aproximadamente como n^4)
- 3 El programa resuelve el siguiente problema de autovalores generalizado $FC_i = \varepsilon_i SC_i$. Debido que F depende de C_i , que es la solución, el procedimiento es iterativo

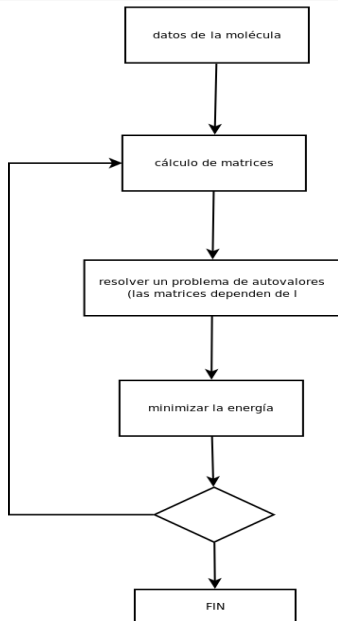
- 1 Debido que el programa calcula la geometría molecular óptima se debe especificar la molécula a considerar
- 2 Se calculan cuatro matrices: $S_{\mu\nu}$, $H_{\mu\nu}$, $(\mu\nu|\sigma\lambda)$, $F_{\mu\nu}$; donde $F_{\mu\nu}$ depende de $H_{\mu\nu}$, $(\mu\nu|\sigma\lambda)$. (Los índices van de 1 a n , por lo cual el cálculo escala aproximadamente como n^4)
- 3 El programa resuelve el siguiente problema de autovalores generalizado $FC_i = \varepsilon_i SC_i$. Debido que F depende de C_i , que es la solución, el procedimiento es iterativo
- 4 Cuando se comprueba la convergencia del proceso, el proceso finaliza, caso contrario vuelve a 2

Algoritmo de cálculo

- 1 Debido que el programa calcula la geometría molecular óptima se debe especificar la molécula a considerar
- 2 Se calculan cuatro matrices: $S_{\mu\nu}$, $H_{\mu\nu}$, $(\mu\nu|\sigma\lambda)$, $F_{\mu\nu}$; donde $F_{\mu\nu}$ depende de $H_{\mu\nu}$, $(\mu\nu|\sigma\lambda)$. (Los índices van de 1 a n , por lo cual el cálculo escala aproximadamente como n^4)
- 3 El programa resuelve el siguiente problema de autovalores generalizado $FC_i = \varepsilon_i SC_i$. Debido que F depende de C_i , que es la solución, el procedimiento es iterativo
- 4 Cuando se comprueba la convergencia del proceso, el proceso finaliza, caso contrario vuelve a 2
- 5 La obtención de la geometría molecular óptima, se realiza minimizando la energía. Se realiza el cálculo anterior para diferentes posiciones relativas de los átomos



Diagrama de flujo



geometría <- vectores en R3
orbitales atómicos

S, H, matriz de 4 índices, F

$F C_i = E_i S C_i$

La aplicación

- Desarrollada en MatLab y Windows
- No requiere software adicional para su ejecución
- Se ejecuta en PC
- Corridas con experimentos simples
- Los altos tiempos de ejecución son un problema



Aumentar

- Complejidad de los experimentos
- Cantidad de experimentos simultáneos
- Disponibilidad del recurso de cómputo
- Acceso a recursos de cómputo



Expectativas de los usuarios al gridificarlo

Aumentar

- Complejidad de los experimentos
- Cantidad de experimentos simultáneos
- Disponibilidad del recurso de cómputo
- Acceso a recursos de cómputo

Restricciones

- No modificar el código
- Mantener el código en privado (hasta publicar la aplicación)



Pasar de

- Windows a Linux (c:\lcaomin -> /home/\$HOME/lcaomin)
- Matlab a Octave (para usar software libre)
 - Pruebas en equipo local (SL)
 - Verificación de los datos obtenidos



¿Existen los recursos necesarios para ejecutar la aplicación?

- Se buscan sitios grid con Octave para conocer su método de instalación
 - en GILDA está instalado en pocos nodos
 - en EELA solo está instalado en UTFSM
 - en SEE se encontraron multitud de nodos, pero con versiones anteriores (2.x)
 - en todos los casos la instalación era directa con RPM



Desde la aplicación

- Parametrizar para permitir realizar varios experimentos simultáneamente
 - Se identifican los parámetros de entrada
- Estimar el tiempo de ejecución
 - Se ejecutan jobs de pruebas con una molécula simple
 - Se realizan estimación de como escalaría en la grid
 - Se evalúa utilizar el servicio de MyProxy e implementar checkpoints
- Realización de un script master para el manejo del job



- Se realizan ejecuciones en la UI
- Se comienza a trabajar con la infraestructura de testing (GILDA)
- Se identifican los CEs que tienen octave instalado y se ejecutan job de pruebas
- Se realizan varios experimentos en paralelo (jobs paramétricos)



Dos maneras de ejecutarlo

- Instalación en el sitio grid
 - Se realiza la compilación estática (sin dependencias) de la última versión de Octave y se lo comprime (tgz) para distribuirlo a los admins de la VO o de cada sitio
 - Se realiza un procedimiento detallando la instalación standalone
- Ejecución en un sitio donde no esté instalado
 - Se almacena el archivo tgz (95 MB) en un SE para descargarlo y ejecutarlo localmente
 - La compilación fue realizada en SL5
 - Aún existen sitios que tienen nodos con SL4

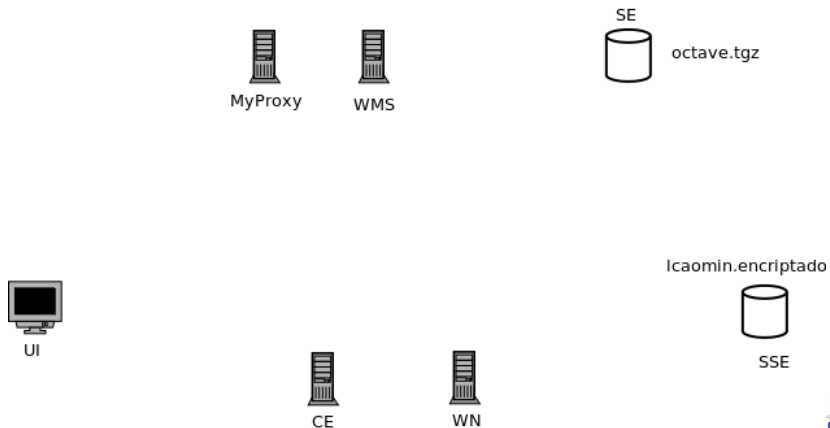


environment

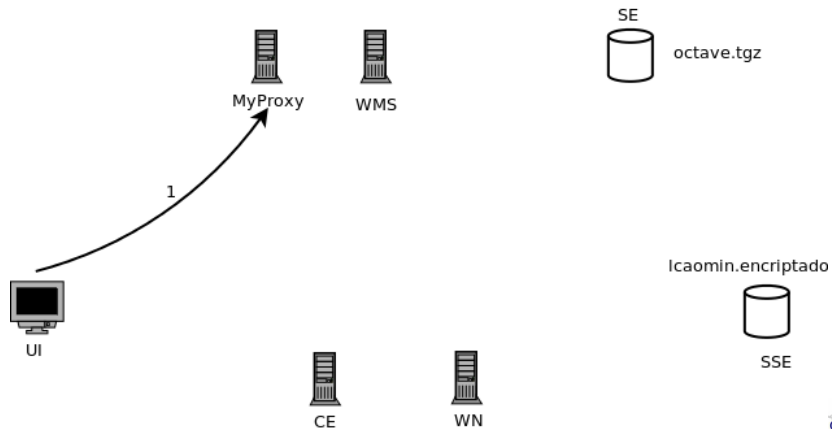
```
OCTAVE_HOME=$VO_PROD_VO_EU_EELA_EU_SW_DIR/  
octave-3.2.4-x86_64-v.1.0  
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$OCTAVE_HOME/  
lib/octave-3.2.4
```



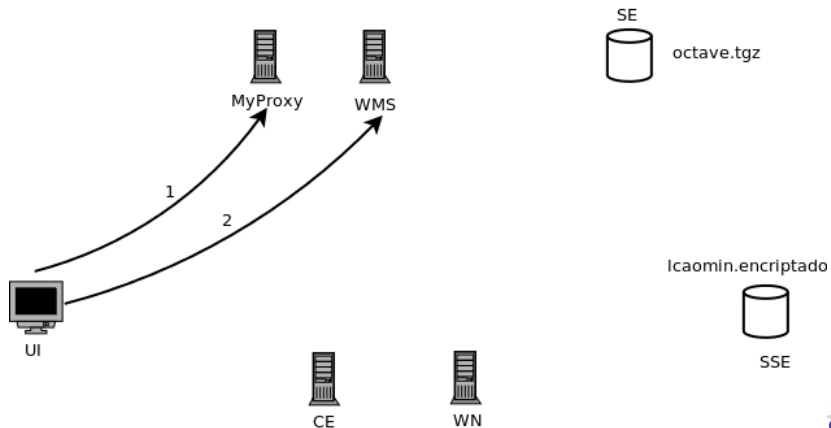
Esquema (0)



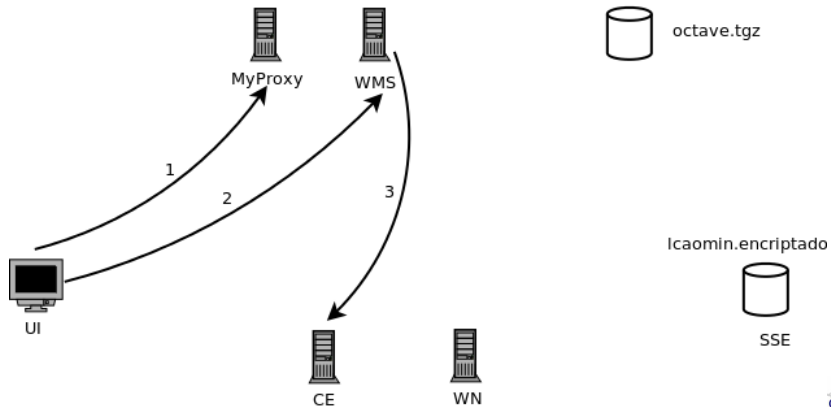
Esquema (1)



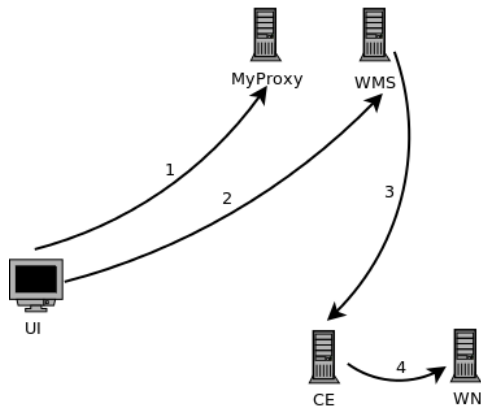
Esquema (2)



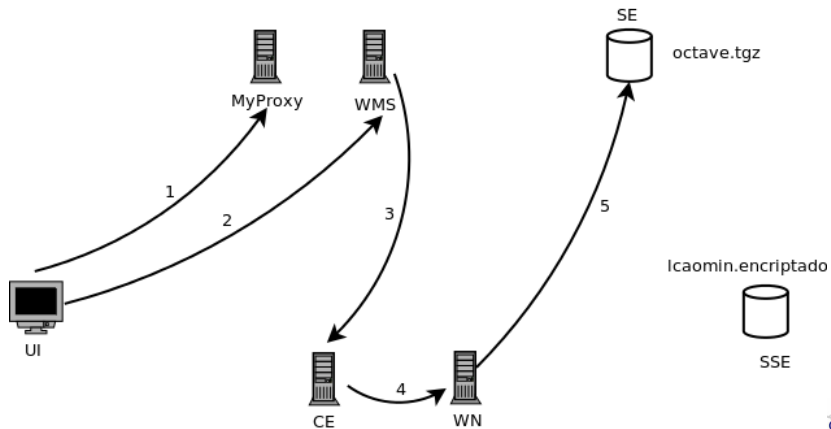
Esquema (3)



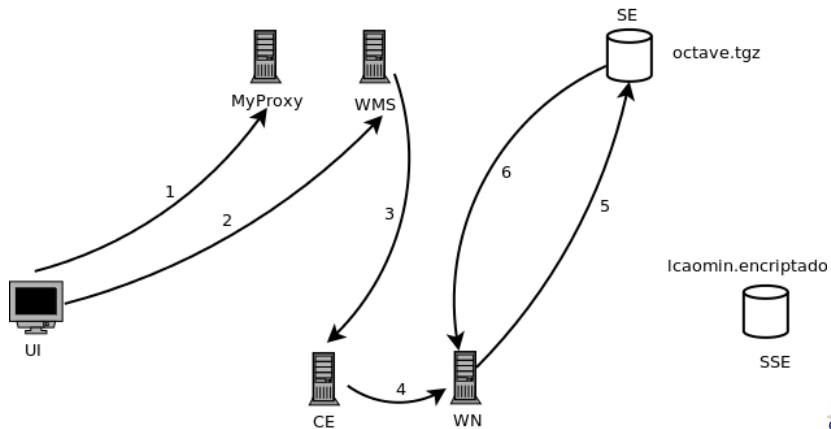
Esquema (4)



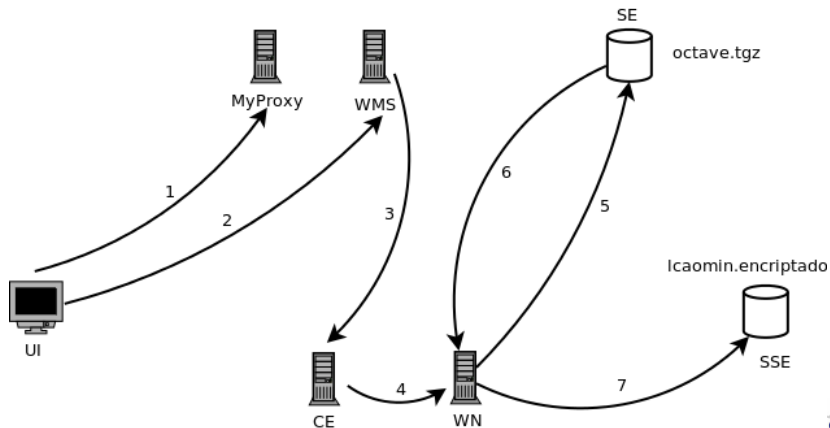
Esquema (5)



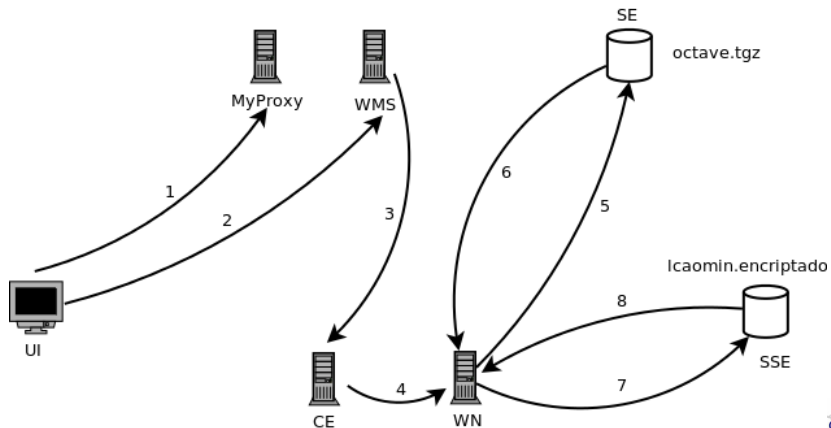
Esquema (6)



Esquema (7)



Esquema (8)



```
[  
  JobType = "Parametric";  
  MyProxyServer = "px.eela.ufrj.br";  
  Executable= "lcaomin.sh";  
  Arguments= "_PARAM_";  
  Stdoutput= "output-_PARAM_.log";  
  StdError= "error-_PARAM_.log";  
  Parameters = {agua, carbono};  
  InputSandbox= {"lcaomin.sh", "lcaomin.tgz", "_PARAM_.tgz"};  
  OutputSandbox= {"resultados.tgz"};  
]
```



```
[dbellomo@uis grid_school]$ ll parametric_job_1/
total 12
-rw-rw-r-- 1 dbellomo  438 Dec  9 11:10 ids_nodes.map
drwxr-xr-x 2 dbellomo 4096 Dec  9 11:10 Node_agua/
drwxr-xr-x 2 dbellomo 4096 Dec  9 11:10 Node_carbono/
```



Servicios

Myproxy

Checkpoints

SE

SSE

Servicios

Myproxy

Checkpoints

SE

SSE

Aplicación

Se satisfacen los requerimientos

Status S3

Octave en GISELA

Servicios

Myproxy

Checkpoints

SE

SSE

Aplicación

Se satisfacen los requerimientos

Status S3

Octave en GISELA

To do

Implementar checkpoints (se necesita modificar el código)

Dar respuesta a nuevos requerimientos de los usuarios

Documentar en el wiki de GISELA “Octave howto”

¡Muchas Gracias!

- UNRC / UTI
- GISELA
- CIEMAT
- UTFSM / Organización local
- Tutores
- Compañeros





Creer... Crear... Crecer...