

Gaussian98

- Actualizado (16.12.2005)

Execución de Gaussian 98 nos ordenadores do CESGA

Para executar Gaussian 98 (<http://www.gaussian.com/>) en calquera dos ordenadores en donde se instalara no CESGA, é suficiente utilizar o comando

```
g98 ficheiro_de_entrada
```

Este comando definirá automaticamente as variables de entorno necesarias para a execución. Ademais, non permite a execución en interactivo. Si necesita esa funcionalidade, contacte con .

Axuste da memoria

Os ficheiros de entrada de Gaussian permiten a selección da memoria a utilizar para a resolución do problema a través da clave %mem situada antes de cada un dos links. Por exemplo,

```
%mem=6mw
```

```
#P HP/ST0-3G FORCE
```

A sintaxe deste comando é:

```
%mem=<numero><unidade>
```

onde número é a cantidade de memoria a utilizar (no exemplo 6) e unidade pode ser KB, MB, GB (cando a cantidade se expresa en Bytes) ó KW, MW, GW (cando se expresa en palabras. Unha palabra equivale a 8 bytes no HPC320 e no SUPERDOME)

En función do problema e da memoria seleccionada, Gaussian selecciona o modo de execución, que pode ser:

Directo. Todas as integrais necesarias para a resolución son calculadas cando se necesitan. É o modo por defecto.

Convencional. As integrais calcúlanse unha vez e acumúlanse en disco. Este sistema é máis lento que o directo, polo cal está desaconsellado.

InCore. As integrais necesarias acumúlanse en memoria (coidado, si se selecciona este método de cálculo, se non hai suficiente memoria o cálculo non se executará).

Semi-directo. Fundamentalmente utilizado en cálculos MP. É unha combinación do modo directo co convencional.

Nas probas realizadas viuse que o rendemento do software é fortemente dependente da selección de memoria que se realice, incrementándose notablemente si se reserva memoria innecesaria. As cantidades mínimas de memoria necesarias veñen dadas pola expresión:

$M+2 NB^2$

Donde NB é número de bases do problema e M e unha cantidade en MW dada pola seguinte táboa (véxase o manual de Gaussian 98 para máis información):

Job Type

Highest Angular Momentum Basis Function

f functions

g functions

h functions

i functions

j functions

SCF Energies

8 MW

8 MW

18 MW

46 MW

~120 MW

SCF Gradients

8 MW

10 MW

32 MW

76 MW

SCF Frequencies

8 MW

18 MW

54 MW

MP2 Energies

8 MW

10 MW

20 MW

56 MW

~140 MW

MP2 Gradients

8 MW

12 MW

32 MW

76 MW

MP2 Frequencies

12 MW

20 MW

56 MW

Estas cantidades hanse de dividir por 2 cando se execute no SVG.

Tamén é posible utilizar o comando freqmem cuxa sintaxe é:

```
freqmem number-of-atoms number-of-basis-functions  
rhf/uhf conventional/direct sp/spd/spdf
```

Por exemplo,

```
SC1> freqmem 100 300 rhf direct spd
```

RHF direct frequencies with spd functions:

One pass requires 27.59 megawords.

Que indica que para calcular as frecuencias dun problema destas características é necesario un mínimo de 27.59 MW.

Almacenar integrais en memoria

Cando sexa posible, é preferible que as integrais se almacenen en memoria. Actualmente, na versión de Gaussian instalada no SUPERDOME e no HPC320 existe un límite de 16GB, co cal non será posible almacenar as integrais que necesiten máis memoria. Para obter o valor real necesario pódese executar un cálculo con pouca memoria (por exemplo, 6MW) solicitando que se realice InCore. Se as integrais necesarias para o cálculo non se poden almacenar en memoria, o cálculo párase aos poucos segundos cunha mensaxe como esta:

```
Could not store integrals in canonical form, short by 289038185 words
```

A cantidade expresada haberá que sumala á que fixamos e indicala como valor na clave %mem (si é inferior a 16GB, límite actual do Gaussian 98).

Si a cantidade de memoria necesaria é moderada (2-4GB) é preferible realizalo desta maneira, xa que o tempo de execución pode reducirse entre un tercio (véxase o exemplo) e un oitavo do necesario se se fai de forma directa.

Para grandes cálculos de enerxías e de gradientes utilizando os métodos AM1, PM3, NMDO e DFT existe a posibilidade de seleccionar a opción de sparse para almacenar as matrices. Esta clave aplica un corte para anular os valores moi pequenos (ver a descrición no manual de Gaussian 98 en <http://doc.cesga.es/gaussian98/sparsek.htm>)

Execución en paralelo

Gaussian 98 permite a execución en paralelo utilizando memoria compartida (único caso dispoñible no CESGA) ou memoria distribuída (por medio do paquete LINDA). Gaussian 98 ao executarse vai chamando a través de fork() aos diferentes links (ou overlays) que son necesarios para a execución do problema. Destes links, só están paralelizados algúns, entre os que se atopan os seguintes:

Link

Descrición

L502

Solución da ecuación SCF iterativa

L703

Derivadas primeira e segunda (spdf)

L914

Estados excitados de CI-singles, RPA y Zindo. Estabilidade SCF

L1002

Solución das ecuacións CPHF; cálculo de propiedades (como NMR)

L1110

Contribución das integrais de 2e a Fx

En xeral, tendo en conta isto, todos os cálculos de enerxía, de gradientes e de frecuencia a través de SCF (isto é, HF, DFT e CIS) xunto con TD-DFT están aceptablemente paralelizados, permitindo unha execución cunha alta eficiencia cando o número de procesadores está entre 2 e 4. Outros métodos, como Moller-Plesset, non melloran suficientemente cando se executa con máis dun procesador, polo que se desaconsella o seu uso neses casos. Para executar en Gaussian 98 en paralelo hase de utilizar a clave %nproc no ficheiro de entrada ó - P - no ficheiro Default.Route. Por exemplo, para o caso anterior

```
%nproc=4
```

```
%mem=6mw
```

```
#P HF/STO-3G FORCE
```

O problema é que a memoria que require para executarse aumenta co número de procesadores. En xeral, a memoria mínima necesaria pódese calcular multiplicando o número de procesadores solicitados pola memoria necesaria para un só procesador (excluíndo o espacio reservado a integrais no caso de que o cálculo sexa InCore). Unha aproximación experimental ven dada por C.P.Sosa en donde a memoria se pode axustar pola fórmula:

$$MN = M1 + 0.75 \cdot (N-1) \cdot M1$$

Donde M1 é a memoria necesaria para un procesador e MN a memoria para a execución con N procesadores.

Se ten memoria suficiente para acumular as integrais en memoria (InCore) é posible que a gañancia por incluír máis procesadores sexa mínima. Por exemplo, para o problema seguinte con HF

```
#P HF/aug-cc-pVDZ Opt=CalcAll Test
```

```
Test
```

```
0 1
```

```
AI
```

```
-0.9123310000
```

```
-0.9123310000
```

```
-0.9123310000
```

```
AI
```

```
0.9123310000
```

```
-0.9123310000
```

```
0.9123310000
```

```
AI
```

```
0.9123310000
```

```
0.9123310000
```

```
-0.9123310000
```

```
AI
```

```
-0.9123310000
```

```
0.9123310000
```

```
0.9123310000
```

```
O
```

```
0.9071410000
```

0.9071410000
0.9071410000

O
-0.9071410000
-0.9071410000
0.9071410000

O
-0.9071410000
0.9071410000
-0.9071410000

O
0.9071410000
-0.9071410000
-0.9071410000

F
-1.8524860000
-1.8524860000
-1.8524860000

F
1.8524860000
-1.8524860000
1.8524860000

F
1.8524860000
1.8524860000
-1.8524860000

F
-1.8524860000
1.8524860000
1.8524860000

Os resultados obtidos no Superdome son:

Método

Tempo (Segundos)

Speedup

Speedup

Directo, 1 CPU

5938

1

InCore, 1 CPU

2380

2,49

1

InCore, 2 CPUs

1798

3.30

1.33

$SpeedUp = \frac{Tempo1}{Tempo2}$

Se ben con respecto ao cálculo directo, a inclusión de 2CPUs mellora novamente, esta mellora é só do 33% (é dicir, só un tercio da esperada).

Recomendacións

1. Axuste a memoria a utilizar o mellor que poida. Máis memoria non sempre significa mellor rendemento.
2. Para cálculos con HF, DFT, TD-DFT e CIS, se a memoria necesaria para o cálculo en InCore está dispoñible (xeralmente menos de 4GB), realice o cálculo desa forma. En caso contrario, faga o cálculo de forma directa (por defecto en Gaussian 98) con varios procesadores (entre 2 e 4. En SVG non solicite máis de 2, xa que o traballo fallará. En HPC320 non poden pedirse máis de 4.).
3. Non utilice varios procesadores para métodos combinados (como G3) ou para Moller-Plesset, xa que a gañancia é pouca ou casi nula.
4. Ao enviar o traballo a cola, asegúrese de solicitar correctamente o número de CPUs e a cantidade de memoria e de disco necesaria.

5. Ante cualquier problema que teña, non dubide en poñerse en contacto cos técnicos do CESGA a través da conta de correo ou chamando directamente por teléfono.

Referencias:

-
- C.P. Sosa et.al. "Gaussian Performance Using AIX Large Pages"

 - C.P. Sosa et. al. "Some Practical Suggestions for Performing Gaussian Benchmarks on a pSeries 690 System"

 - C.P.Sosa et. al. "Gaussian Tutorial: Estimating Resource Requirements"

 - Notur Advance User Group: "Gaussian In-Depth Performance Analysis"