Gaussian, ADF and NWChem on SHARCNET

Jemmy Hu

SHARCNET HPC Analyst

Email: jemmyhu@sharcnet.ca

April 7, 2008

AG Seminar

Contents

1. Gaussian license issues

2. G03 input issues
   (%mem, %nproc, %rwf, %chk)

3. G03 run scripts
   (g03, g03-local, g03-SN)

4. ADF and NWChem

Gaussian License Issue

- Campus site license

  - cheap but runs only on campus machines

Site	**Version**	**Clusters**
McMaster	C.02	requin, wobbie
Guelph	C.02	narwhal
UW	C.02	whale
UWO	C.02	bull, goblin, greatwhite
Windsor	D.01	tiger
York	D.01	dolphin

Gaussian License Issue

- **SHARCNET g03 license, g03-D.01**

  - expensive commercial license
  - installed on clusters in one physical/campus site, can be used by any Sharcnet users
  - license cost: $1000/yr per group

  To join Sharcnet G03 group, send an request to help@sharcnet.ca

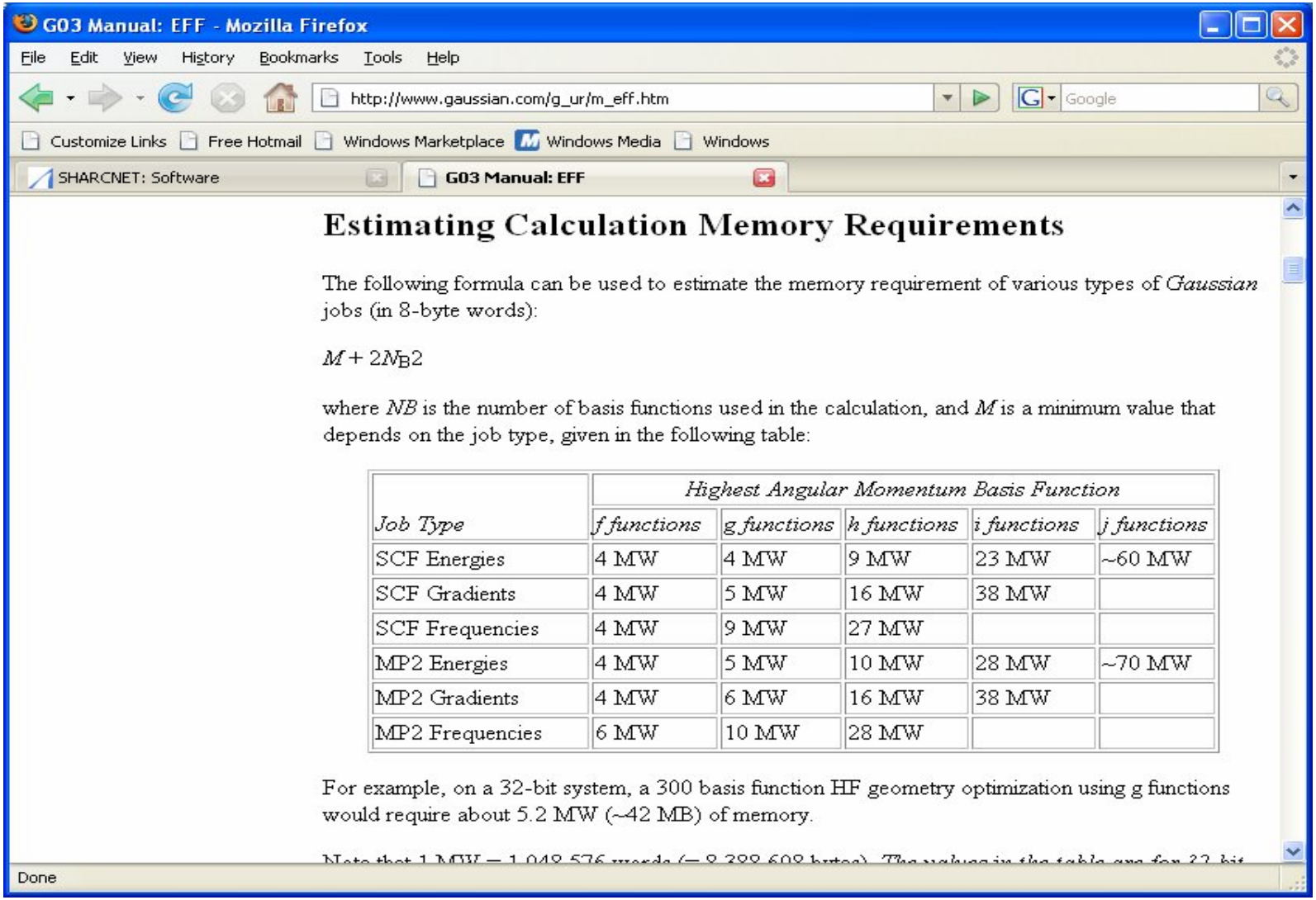
- **Clusters**

  bull: 4-cpu, 32GB RAM, dedicated gaussian queue

  goblin: 4-cpu or 2-cpu nodes

  greatwhite: 1-cpu serial g03 job only

G03 input issues: %mem



Estimating Calculation Memory Requirements

The following formula can be used to estimate the memory requirement of various types of *Gaussian* jobs (in 8-byte words):

$M + 2N_B2$

where *NB* is the number of basis functions used in the calculation, and *M* is a minimum value that depends on the job type, given in the following table:

	Highest Angular Momentum Basis Function				
Job Type	*f functions*	*g functions*	*h functions*	*i functions*	*j functions*
SCF Energies	4 MW	4 MW	9 MW	23 MW	~60 MW
SCF Gradients	4 MW	5 MW	16 MW	38 MW	
SCF Frequencies	4 MW	9 MW	27 MW		
MP2 Energies	4 MW	5 MW	10 MW	28 MW	~70 MW
MP2 Gradients	4 MW	6 MW	16 MW	38 MW	
MP2 Frequencies	6 MW	10 MW	28 MW		

For example, on a 32-bit system, a 300 basis function HF geometry optimization using g functions would require about 5.2 MW (~42 MB) of memory.

Note that 1 MW = 1,048,576 words (= 8,388,608 bytes). *The values in the table are for 32-bit*

http://www.gaussian.com/g_ur/m_eff.htm

G03 input issues

- %nproc

  Gaussian software uses OpenMP parallel solution, which performs very good on the SMP node for up to 4 threads. %nproc depends on cluster's node infrastructure

  - %nproc = 4

    bull, narwhal, whale, dolphin, tiger, wobbie, goblin
  - %nproc = 2

    requin, goblin, wobbie
  - %nproc=1 **on greatwhite (serial)**

you can also specify %nproc=1 and submit the job to the serial queue on any clusters, but it's not optimal in most cases.

G03 input issues

- **%rwf** (.rwf, .inp, .scr, .d2e, .int)

  By default, Gaussian will generate 5 scratch files with random name. Those files will be removed once the job is finished successfully.

  .inp is a binary input file

  .scr occasionally used in one MP2 routine

  .d2e and .int are useless

  All those files are written into /scratch/$USER/

- Name the .rwf file

  %rwf=/scratch/yourid/name.rwf

  %NoSave

Remove those scratch files once a job was failed !

G03 input issues

- %chk

  - why checkpoint?

    restart a geometry optimization job

    used for further property study such as freq, etc.

    on a diff. basis sets, or to use another method

- Name %chk

  %chk=jobname

  jobname.chk file in current directory, i.e., your /work

  %chk=/scratch/yourID/jobname

  jobname.chk file in your /scratch directory

input file structure

Traditional script, still used by some users in Sharcnet

Stop this kind of script/input on Sharcnet clusters!

Do not need to include g03 settings any more !!

```
#!/bin/bash
#
# use current directory for output and runtime scratch files
g03root="/opt/sharcnet/gaussian"
GAUS_SCRDIR=`pwd`
export g03root GAUS_SCRDIR
. $g03root/g03/bsd/g03.profile
#
prog=$g03root/g03/g03
# setup the output and checkpoint file names
#
$prog << ! >& CO-6311-4.log
```

```
%rwf=/scratch/$USER/CO-6311-4
%nosave
%mem=3200MB
%chk=/scratch/$USER/CO-6311-4
%nproc=4
# mp2/6-311++g(2df,p) opt freq

<< CO-6311++g(2df,p)—MP2-Opt -freq>>

0 1
c
o 1 co

co=1.1281
```

Run G03 on SHARCNET

- **g03** – common run script on all clusters

  - do a ‘dos2unix’ on the .com file

  - rewrite **%rwf, %mem, %nproc, %chk** for .com file

Clusters	**%mem**	**%nproc**
whale	3GB	4
bull	29GB	4
narwhal, dolphin, tiger	7GB	4
requin	7GB	2
greatwhite	1~2GB	1

  %rwf, %chk in /scratch/yourID/jobname

  - create and run a temp .com file

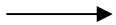
  - remove the temp .com file when job terminated

g03 run script

- **On bull:**

CO-6311-4.com

```
%rwf=/scratch/$USER/CO-6311-4
%nosave
%mem=3200MB
%chk=CO-6311-4
%nproc=4
# mp2/6-311++g(2df,p) opt freq

<< CO-6311++g(2df,p)—MP2-Opt >>

0 1
c
o 1 co

co=1.1281
```



temp .com

```
%rwf=/scratch/$USER/CO-6311-4
%nosave
%mem=29GB
%chk=/scratch/$USER/CO-6311-4
%nproc=4
# mp2/6-311++g(2df,p) opt freq

<< CO-6311++g(2df,p)—MP2-Opt >>

0 1
c
o 1 co

co=1.1281
```

g03 run script

Attention 1: no space lines before # keyword line

```
%rwf=/scratch/$USER/CO-6311-4
%nosave
%mem=3200MB
%chk=CO-6311-4
%nproc=4

# mp2/6-311++g(2df,p) opt freq

<< CO-6311++g(2df,p)—MP2-Opt-freq >>

0 1
c
o 1 co

co=1.1281
```



```
%rwf=/scratch/$USER/CO-6311-4
%nosave
%mem=29GB
%chk=/scratch/$USER/CO-6311-4
%nproc=4
#
# mp2/6-311++g(2df,p) opt freq

<< CO-6311++g(2df,p)—MP2-Opt-freq >>

0 1
c
o 1 co

co=1.1281
```

wrong .com file

g03 run script

Attention 2:

For multiple g03 jobs linked in the .com file, make sure the .chk file is in your /scratch directory

g03 only take care of the lines before the first # keyword line

Use a checkpoint file for another named job, you have to change the name of .chk to match the new jobname in your /scratch directory

For safe, use .com for input file, not .gjs or other type of files

```
%rwf=/scratch/$USER/CO-6311-4
%nosave
%mem=3200MB
%chk=CO-6311-4                          <--
%nproc=4
# mp2/6-311++g(2df,p) opt

<< CO-6311++g(2df,p)—MP2-Opt >>

0 1
c
o 1 co

co=1.1281

--Link 1--
%rwf=/scratch/$USER/CO-6311-4
%nosave
%mem=3200MB
%chk=/scratch/yourID/CO-6311-4          <--
%nproc=4
# mp2/6-311++g(2df,p) freq
……
```

G03 run script

- **Submit command**

  sqsub –q threaded –n 4 g03 jobname.com (-n 2 for requin)

  bsub –q gaussian –n 4 g03 jobname.com (bull SN license users)

  sqsub g03 jobname.com (serial job on greatwhite)

  by default, output file is jobname.log

Use –o jobname.log

sqsub –q threaded –n 4 –o jobname.log g03 jobname.com

bsub –q gaussian –n 4 –o jobname.log g03 jobname.com

Append LSF message to the end of the g03 output, good for debugging for failed jobs, encourage you to do so.

Run your own .com file

- **g03-local**, for site license on any local clusters
  sqsub –q threaded –n 4 g03-local jobname.com (-n 2 for requin)
- **g03-SN**, for SN license on bull, (-q threaded for goblin)
  bsub –q gaussian –n 4 g03-SN jobname.com
- **g03-SN**, for SN license on greatwhite
  sqsub g03-SN jobname.com

Have to specify %mem, %nproc, %chk, %rwf in your .com file

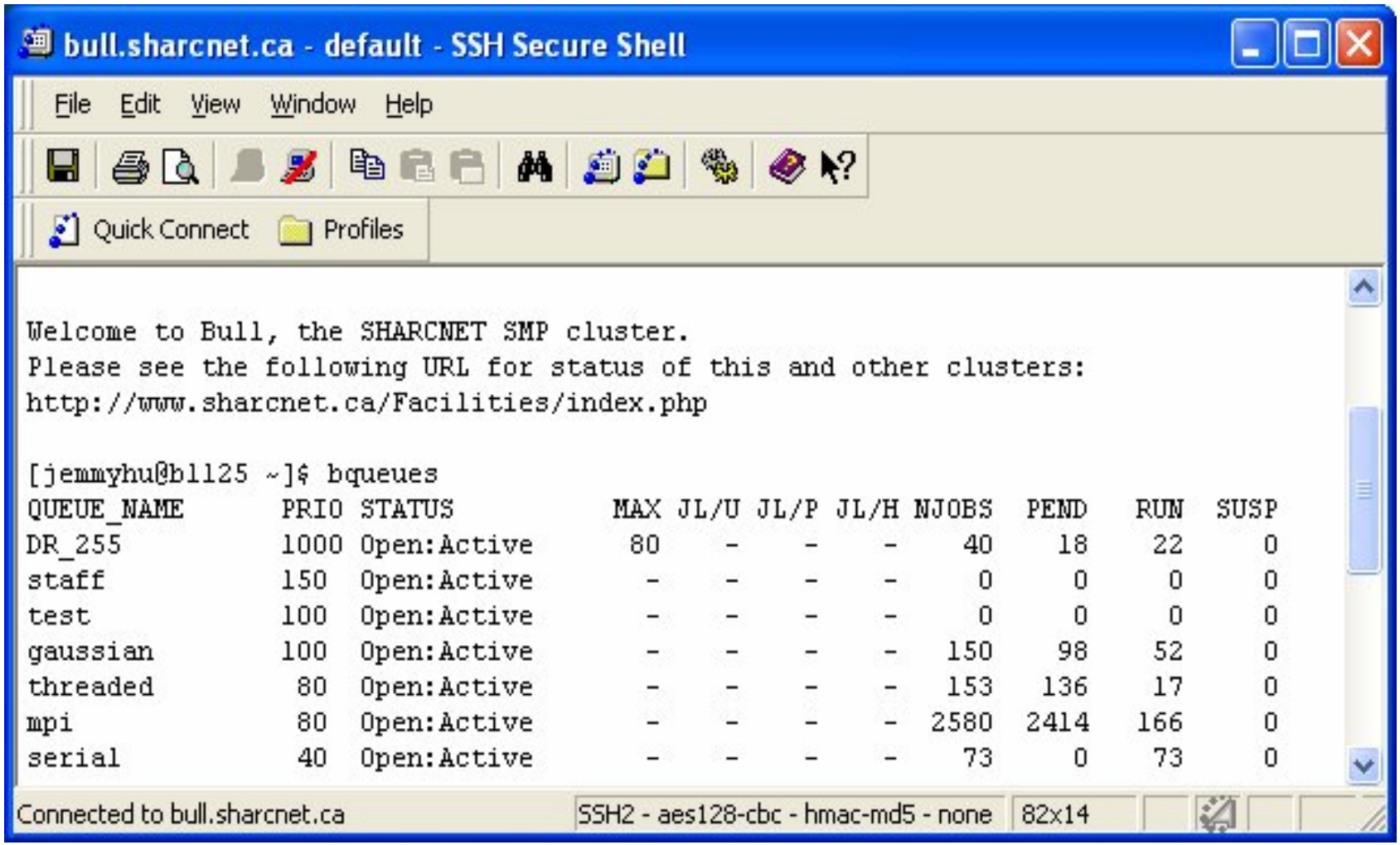
Case 1: for some MP2 job, where large %mem may not work correctly

Case 2: serial %nproc=1 job on the clusters (except for greatwhite) as g03 script does not set %nproc=1 automatically.

```
%rwf=/scratch/$USER/CO-6311-4
%nosave
%mem=3200MB
%chk=CO-6311-4
%nproc=4
# mp2/6-311++g(2df,p) opt freq

<< CO-6311++g(2df,p)—MP2-Opt-freq >>

0 1
c
o 1 co

co=1.1281
```

Queue system:

bull: bqueues



*

ADF / BAND, NWChem

- NWChem

  parallel solutions, free license

http://www.sharcnet.ca/Facilities/software/soft_detail.php?id=12

- ADF, BAND

  DFT based

  method for heavy atoms where relativistic effect is important.

  license cost: $1000/yr per group

http://www.sharcnet.ca/Facilities/software/soft_detail.php?id=56

Problem reporting online

https://www.sharcnet.ca/my/problems

or email jemmyhu@sharcnet.ca

Sharcnet Gaussian software webpage:

http://www.sharcnet.ca/Facilities/software/soft_detail.php?id=21

- Questions, Concerns ?