%let pgm=sys\_benccwyam;

This is somewhat dated.

Benchmarks before and after upgrade (not the latest upgrade – which may be worse)

Before the latest upgrade I could run 32 simultaneous jobs without the grid now

I can only run 4)

Performance seems to be the same for table scans and sorts, but

index creation and use seem more than twice as fast.

Benchmarks before and after EG Upgrade

compared with same benchmarks on my 2008 Workstation

all results in seconds seconds

$600

May 18 June 1 2008 Dell T7400

EG Before EG After EG After Workstation

======== ======== ======== ===========

Create data (22gb) 265 351 355 129

Sort data 305 333 331 181

Index data 227 152 51 111

sql query (with index) 7.35 2.56 2.12 3.78

sql query (without index) 32.34 27.13 10.74 10.59

SASFILE

SQL query (no index) N/A\*\* N/A\*\* N/A\*\* 4.30

SPDE(free with base sas)

SDS not available in EG?

SQL query (with SPDE index) N/A\* N/A\* N/A\* 0.45 \*\* best parallel index

GRID

Job (4 tasks)\* 231 98 99 114

Workstation 8 simultaneous systasks (could run 32 on EG before upgrade – unix server had 32 cores)

options noxwait noxsync;

%let tym=%sysfunc(time());

systask kill sys1 sys2 sys3 sys4 sys5 sys6 sys7 sys8;

systask command "&\_s -termstmt %nrstr(%sort(1);) -log d:\log\a1.log" taskname=sys1;

systask command "&\_s -termstmt %nrstr(%sort(2);) -log d:\log\a2.log" taskname=sys2;

systask command "&\_s -termstmt %nrstr(%sort(3);) -log d:\log\a3.log" taskname=sys3;

systask command "&\_s -termstmt %nrstr(%sort(4);) -log d:\log\a4.log" taskname=sys4;

systask command "&\_s -termstmt %nrstr(%sort(5);) -log d:\log\a5.log" taskname=sys5;

systask command "&\_s -termstmt %nrstr(%sort(6);) -log d:\log\a6.log" taskname=sys6;

systask command "&\_s -termstmt %nrstr(%sort(7);) -log d:\log\a7.log" taskname=sys7;

systask command "&\_s -termstmt %nrstr(%sort(0);) -log d:\log\a8.log" taskname=sys8;

waitfor sys1 sys2 sys3 sys4 sys5 sys6 sys7 sys8;

%put %sysevalf( %sysfunc(time()) - &tym);?

\* - parallel indexing and partioned data not supported in EG

SPDE is free in workstation SAS

Note only 10 grided jobs are allowed by my EG grid manager? A Workstation Dell dual XEON

or I7 should support eight. Newer workstaions easily support 16.

Note Dell I7 Laptops support 32gb ram and 4 sata3 ports, DVD, C drive, esata on laptop and

esata on docking station and RAID 0.

\*My workstation is an T7400(circa 2008) with 64gb ram(DDR2) and two raid 0 SSDs(sata II) arrays

of 4 disks each. You can get these off lease for about $500.

SPDE is so fast because it uses all three indexes, without SPDE SAS chooses just one index

the one with the highest cardinality. Note indexes can be slower than a table scan.

Don't use an index if pulling 15% or more from the slow EG drives.

Workstation gridding of the sums of squares for regression analysis can be extremely fast.

This is why ordinary least squares is so fast in Teradate.

Might be interesting to compare regression on workstation and Teradata.

EG is a good data farm for extracting data for the workstation to analyze.

Generally datasets <1TB can easily and inexpensively be handled on the desktop.

EG is also good as a learning tool for non-programmers.

\* \* \* \* \* \*\*\*\*\* \* \* \*

\* \* \* \* \*\* \* \* \* \* \* \*

\* \* \* \* \* \* \* \* \* \* \* \*

\* \* \*\*\*\*\* \* \*\* \* \* \* \*\*\*\*\*

\* \* \* \* \* \* \* \* \* \* \*

\* \* \* \* \* \* \* \* \* \* \*

\* \* \* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \*;

options fullstimer;

\* 22gb dataset;

\* create main dataset for use with spde, sasfile, ssd and spinning disks;

data gig23ful\_main;

array num[50] n1-n50;

do seq=1 to 50000000;

bene\_id=int(uniform(-1)\*1e12);

do idx=31 to 80;

num[idx-30]=uniform(-1);

end;

idx1=int(50\*uniform(-1));

idx2=int(25\*uniform(-1));

idx3=int(12\*uniform(-1));

output;

end;

run;quit;

proc sort data=gig23ful\_main noequals;

by idx1 idx2 idx3;

run;quit;

proc datasets lib=work;

modify gig23ful\_main;

index create idx1 idx2 idx3;

quit;

proc sql;

create

table modbene\_idx as

select

count(\*) as cnt

,sum(n1) as sum\_n1

,sum(n2) as sum\_n2

,sum(n3) as sum\_n3

,sum(n4) as sum\_n4

,sum(n5) as sum\_n5

,sum(n6) as sum\_n6

from

gig23ful\_main

where

idx1<25 and idx2<12 and idx3=3

;quit;

/\*

1273! quit;

NOTE: PROCEDURE SQL used (Total process time):

real time 3.78 seconds

user cpu time 1.15 seconds

system cpu time 2.65 seconds

memory 1187.70k

OS Memory 18628.00k

Timestamp 01/04/2015 07:17:48 PM

Up to 40 obs WORK.MODBENE\_IDX total obs=1

Obs CNT SUM\_N1 SUM\_N2 SUM\_N3 SUM\_N4 SUM\_N5 SUM\_N6

1 999272 499184.84 499667.98 499315.52 499536.51 499974.98 500281.85

\*/

\*\*\* \*\*\*\* \*\*\*\* \*\*\*\*\*

\* \* \* \* \* \* \*

\* \* \* \* \* \*

\* \*\*\*\* \* \* \*\*\*\*

\* \* \* \* \*

\* \* \* \* \* \*

\*\*\* \* \*\*\*\* \*\*\*\*\*;

proc datasets library=spde;

delete gig23ful\_spde;

run;quit;

libname spde spde 'd:/tmp'

datapath=("f:/wrk/spde\_f" "e:/wrk/spde\_e" "g:/wrk/spde\_g")

partsize=128;

data spde.gig23ful\_spde(compress=no);

set gig23ful\_main(

sortedby=idx1 idx2 idx3);

run;quit;

proc datasets lib=spde;

modify gig23ful\_spde(asyncindex=yes);

index create idx1 idx2 idx3;

quit;

options fullstimer;

proc sql;

create

table modbene as

select

count(\*) as cnt

,sum(n1) as sum\_n1

,sum(n2) as sum\_n2

,sum(n3) as sum\_n3

,sum(n4) as sum\_n4

,sum(n5) as sum\_n5

,sum(n6) as sum\_n6

from

spde.gig23ful\_spde

where

idx1<25 and idx2<12 and idx3=3

;quit;

/\*

OTE: PROCEDURE SQL used (Total process time):

real time 0.45 seconds

user cpu time 1.01 seconds

system cpu time 0.79 seconds

memory 1407.76k

OS Memory 18624.00k

Timestamp 01/04/2015 07:20:17 PM

Up to 40 obs WORK.MODBENE total obs=1

Obs CNT SUM\_N1 SUM\_N2 SUM\_N3 SUM\_N4 SUM\_N5 SUM\_N6

1 999272 499184.84 499667.98 499315.52 499536.51 499974.98 500281.85

\*/

\* \* \*\*\* \*\*\*\*\* \*\*\*\* \* \*

\*\* \* \* \* \* \* \* \* \*

\* \* \* \* \* \* \* \* \* \*

\* \*\* \* \* \* \* \* \*

\* \* \* \* \* \* \* \* \*

\* \* \* \* \* \* \* \* \*

\* \* \*\*\* \*\*\*\*\* \*\*\*\* \* \*;

\* NO INDEX;

proc sql;

drop index idx1, idx2, idx3 from gig23ful\_main;

quit;

proc sql;

create

table modbene as

select

count(\*) as cnt

,sum(n1) as sum\_n1

,sum(n2) as sum\_n2

,sum(n3) as sum\_n3

,sum(n4) as sum\_n4

,sum(n5) as sum\_n5

,sum(n6) as sum\_n6

from

gig23ful\_main

where

idx1<25 and idx2<12 and idx3=3

;quit;

/\*

1399! quit;

NOTE: PROCEDURE SQL used (Total process time):

real time 10.59 seconds

user cpu time 1.71 seconds

system cpu time 8.89 seconds

memory 795.20k

OS Memory 18368.00k

Timestamp 01/04/2015 07:24:40 PM

Up to 40 obs WORK.MODBENE total obs=1

Obs CNT SUM\_N1 SUM\_N2 SUM\_N3 SUM\_N4 SUM\_N5 SUM\_N6

1 999272 499184.84 499667.98 499315.52 499536.51 499974.98 500281.85

\*/

sasfile gig23ful\_main load;

proc sql;

create

table modbene as

select

count(\*) as cnt

,sum(n1) as sum\_n1

,sum(n2) as sum\_n2

,sum(n3) as sum\_n3

,sum(n4) as sum\_n4

,sum(n5) as sum\_n5

,sum(n6) as sum\_n6

from

gig23ful\_main

where

idx1<25 and idx2<12 and idx3=3

;quit;

/\*

NOTE: Table WORK.MODBENE created, with 1 rows and 7 columns.

1418! quit;

NOTE: PROCEDURE SQL used (Total process time):

real time 4.30 seconds

user cpu time 4.19 seconds

system cpu time 0.09 seconds

memory 313.31k

OS Memory 29937600.00k

Timestamp 01/04/2015 07:28:48 PM

\*/

\*\*\*\*\* \*\*\*

\* \* \*

\* \*

\*\*\*\* \* \*\*\*

\* \* \*

\* \* \*

\*\*\*\*\* \*\*\*;

options fullstimer;

\* 22gb dataset;

\* create main dataset for use with spde, sasfile, ssd and spinning disks;

data gig23ful\_main(compress=no);

array num[50] n1-n50;

do seq=1 to 50000000;

bene\_id=int(uniform(-1)\*1e12);

do idx=31 to 80;

num[idx-30]=uniform(-1);

end;

idx1=int(50\*uniform(-1));

idx2=int(25\*uniform(-1));

idx3=int(12\*uniform(-1));

output;

end;

run;quit;

/\*

NOTE: The data set WORK.GIG23FUL\_MAIN has 50000000 observations and 56 variables.

NOTE: DATA statement used (Total process time):

real time 4:24.62 (2:09)

user cpu time 2:52.91

system cpu time 1:31.09

memory 553.00k

OS Memory 15672.00k

Timestamp 01/04/2015 10:51:19 PM

Page Faults 0

Page Reclaims 0

Page Swaps 0

Voluntary Context Switches 314

Involuntary Context Switches 24240

Block Input Operations 13

Block Output Operations 1

\*/

proc sort data=gig23ful\_main noequals;

by idx1 idx2 idx3;

run;quit;

/\*

NOTE: There were 50000000 observations read from the data set WORK.GIG23FUL\_MAIN.

NOTE: The data set WORK.GIG23FUL\_MAIN has 50000000 observations and 56 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 5:05.22 (3:00)

user cpu time 5:20.14

system cpu time 5:16.28

memory 266630.87k

OS Memory 281056.00k

Timestamp 01/04/2015 11:00:12 PM

Page Faults 2

Page Reclaims 0

Page Swaps 0

Voluntary Context Switches 1487597

Involuntary Context Switches 26111

Block Input Operations 2

Block Output Operations 2

\*/

proc datasets lib=work;

modify gig23ful\_main;

index create idx1 idx2 idx3;

quit;

/\*

NOTE: MODIFY was successful for WORK.GIG23FUL\_MAIN.DATA.

NOTE: PROCEDURE DATASETS used (Total process time):

real time 3:47.12

user cpu time 2:07.29

system cpu time 3:10.19

memory 264131.17k

OS Memory 278088.00k

Timestamp 01/04/2015 11:05:14 PM

2

Page Faults 1

Page Reclaims 0

Page Swaps 0

Voluntary Context Switches 149929

Involuntary Context Switches 7667

Block Input Operations 0

Block Output Operations 3

\*/

proc sql;

create

table modbene\_idx as

select

count(\*) as cnt

,sum(n1) as sum\_n1

,sum(n2) as sum\_n2

,sum(n3) as sum\_n3

,sum(n4) as sum\_n4

,sum(n5) as sum\_n5

,sum(n6) as sum\_n6

from

gig23ful\_main

where

idx1<25 and idx2<12 and idx3=3

;quit;

/\*

NOTE: PROCEDURE SQL used (Total process time):

real time 7.35 seconds

user cpu time 3.39 seconds

system cpu time 3.96 seconds

memory 715.79k

OS Memory 15672.00k

Timestamp 01/04/2015 11:06:14 PM

Page Faults 0

Page Reclaims 0

Page Swaps 0

Voluntary Context Switches 31

Involuntary Context Switches 54

Block Input Operations 0

Block Output Operations 0

\*/

\*\*\* \*\*\*\* \*\*\*\*\* \*\*\*\*

\* \* \* \* \* \* \*

\* \* \* \* \* \*

\* \*\*\* \*\*\*\* \* \* \*

\* \* \* \* \* \* \*

\* \* \* \* \* \* \*

\*\*\* \* \* \*\*\*\*\* \*\*\*\*;

data \_null\_;file "&\_r/oto/sys\_100cutdat.sas" lrecl=512;input;put \_infile\_;putlog \_infile\_;

cards4;

%macro sys\_100cutdat(beg=,end=);

%utlopts;

data gig23ful&beg (compress=yes);

array num[50] n1-n50;

array chr[50] $40 c1-c50;

do seq=&beg to &end;

bene\_id=int(uniform(-1)\*1e12);

do idx=31 to 80;

num[idx-30]=uniform(-1);

chr[idx-30]=repeat(byte(idx),40);

end;

output;

end;

run;quit;

%mend sys\_100cutdat;

;;;;

run;quit;

/\*

options obs=1000;

%sys\_100cutdat(beg=1,end=10000000);

options obs=max;

\*/

%let \_s=/sas/sashome/compute/SASFoundation/9.4/sasexe/sas -sysin /dev/null -log /dev/null -autoexec &\_r/oto/batch\_autoexec.sas;

dat(beg=1,end=10000000);) -altlog &\_r/log/sys101.log" taskname=sys101;

systask command "&\_s -termstmt %nrstr(%sys\_100cutdat(beg=10000001,end=20000000);) -altlog &\_r/log/sys102.log" taskname=sys102 ;

systask command "&\_s -termstmt %nrstr(%sys\_100cutdat(beg=20000001,end=30000000);) -altlog &\_r/log/sys103.log" taskname=sys103 ;

systask command "&\_s -termstmt %nrstr(%sys\_100cutdat(beg=30000001,end=40000000);) -altlog &\_r/log/sys104.log" taskname=sys104 ;

waitfor \_all\_ sys101 sys102 sys103 sys104;

systask list;

%put %sysevalf( %sysfunc(time()) - &tym);

%unx(ps -ef | grep sysin); \* watch them run;