



Quick Facts about the SAS Scalable Performance Data Engine

Scott Bass
Senior Information Analyst
Bupa Australia

Purpose of Presentation

- Introduce you to the SPDE engine
- Whet your appetite to learn more about SPDE and try it yourself
- This presentation is not meant to be a complete discussion of the SPDE engine

Issues and Workarounds – File Locking

- The default file caching can cause file locking problems
- Merely viewing library members in DMS Explorer can lock all data sets
- Workaround: specify SAS invocation option

```
/* turn off file level caching for SPDE libraries */  
-spdedebug 'cache_disable'
```

- Recommend setting this in the root level SASV9.CFG file
- See <http://support.sas.com/kb/18/467.html>
- Note: if you have the OLAP Server, you should enable file level caching for that server

Segmented File System – File Naming Convention

SPDE embeds some metadata in the long filenames it uses.

Here's the format:

Dsname.FtypeSuffix.Fuid.P#.V#.SPDS9

where

Dsname	is the data set name
Ftype	is MDF, DPF, IDX, or HBX
Suffix	is empty for MDF or DPF files is the index name for IDX and HBX files
Fuid	is the name of the primary path
P#	is the partition number
V#	is the version number
SPDS9	is the file extension

Segmented File System - Example

```
4,473,480 claimfactmedical.mdf.0.0.0.spds9

25,616,740 claimfactmedical.dpf.19541ab6.0.3.spds9
25,654,705 claimfactmedical.dpf.19541ab6.1.3.spds9
26,894,839 claimfactmedical.dpf.19541ab6.2.3.spds9
...
34,185,216 claimfactmedical.hbxaddresskey.19541ab6.0.3.spds9
    24,576 claimfactmedical.hbxagentbranchkey.19541ab6.0.3.spds9
    81,920 claimfactmedical.hbxclmdt.19541ab6.0.3.spds9
37,044,224 claimfactmedical.hbxcoverkey.19541ab6.0.3.spds9
200,425,472 claimfactmedical.hbxepisodeid.19541ab6.0.3.spds9
45,449,216 claimfactmedical.hbxmemberkey.19541ab6.0.3.spds9
...
280,944,640 claimfactmedical.idxaddresskey.19541ab6.0.3.spds9
12,886,016 claimfactmedical.idxagentbranchkey.19541ab6.0.3.spds9
8,110,080 claimfactmedical.idxclmdt.19541ab6.0.3.spds9
282,279,936 claimfactmedical.idxcoverkey.19541ab6.0.3.spds9
292,626,432 claimfactmedical.idxepisodeid.19541ab6.0.3.spds9
286,130,176 claimfactmedical.idxmemberkey.19541ab6.0.3.spds9
```

Multi-Threaded I/O

Large Fact Table Load (76.3M records)

BASE Engine:

NOTE: The SAS System used:

real time	1:48:05.53	
user cpu time	1:09:47.29	
system cpu time	6:23.59	
Memory		206838k
OS Memory		208408k
Timestamp	07-Nov-11	14:27:37

SPDE Engine:

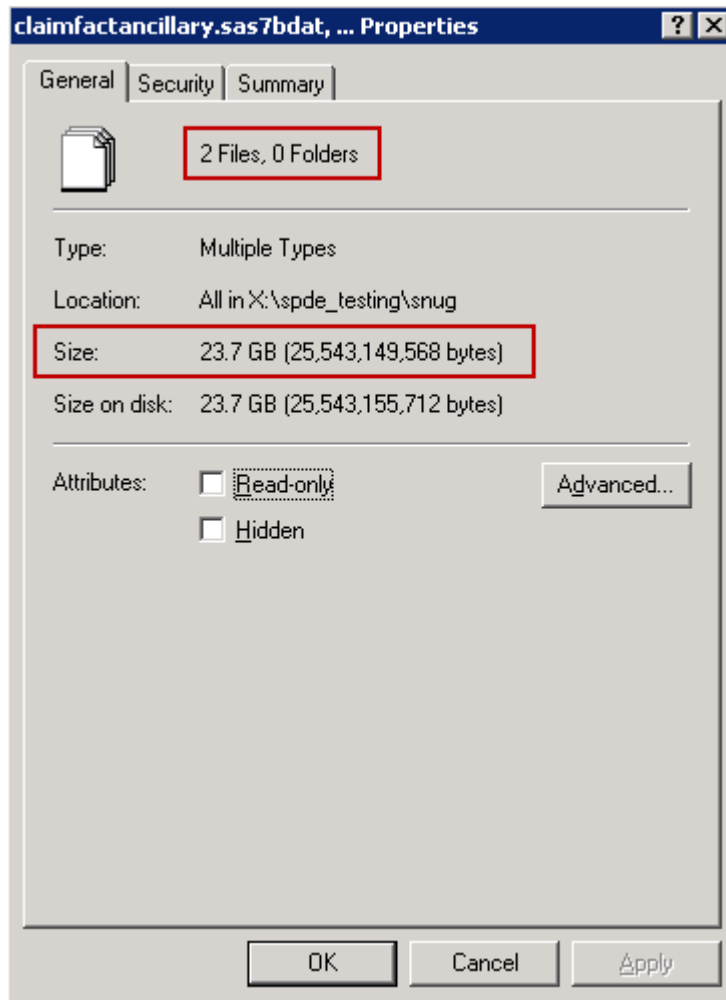
NOTE: The SAS System used:

real time	36:41.88	<<<<<
user cpu time	40:37.64	<<<<<
system cpu time	3:54.76	
Memory		1342014k
OS Memory		1527508k
Timestamp	07-Nov-11	22:58:55

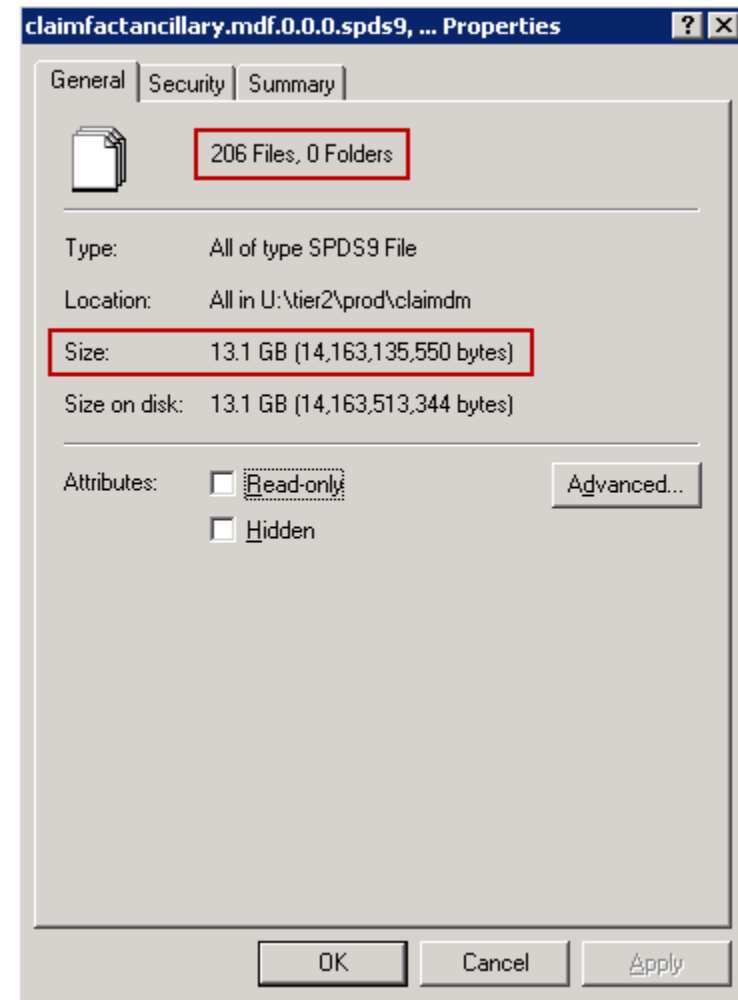
Higher levels of data set compression

COMPRESS=BINARY

BASE Engine:



SPDE Engine:



Very Fast Index Creation – asyncindex=yes data set option

Single pass over the data set for all indexes (15 indexes)

BASE Engine:

NOTE: MODIFY was successful for BASE_S._CFA_TOGGLE.DATA.

NOTE: PROCEDURE DATASETS used (Total process time):

real time	50:16.75	
user cpu time	42:55.31	
system cpu time	2:38.23	
Memory		67441k
OS Memory		78040k
Timestamp	07-Nov-11	14:24:06

SPDE Engine:

NOTE: MODIFY was successful for SPDE_S._CFA_TOGGLE.DATA.

NOTE: Compressing data set SPDE_S._CFA_TOGGLE decreased size by 77.65 percent.

NOTE: PROCEDURE DATASETS used (Total process time):

real time	9:15.43	
user cpu time	17:33.73	
system cpu time	56.57 seconds	
Memory		1334067k
OS Memory		1527508k
Timestamp	07-Nov-11	22:57:06

Large Data Mart Table Index Creation (3 Indexes)

BASE Engine:

NOTE: MODIFY was successful for BASE_T.CLMA004T.DATA.

NOTE: PROCEDURE DATASETS used (Total process time):

real time	1:07:34.70	
user cpu time	14:28.73	
system cpu time	2:09.62	
Memory		68407k
OS Memory		78920k
Timestamp	07-Nov-11	17:41:39

SPDE Engine:

NOTE: MODIFY was successful for SPDE_T.CLMA004T.DATA.

NOTE: Compressing data set SPDE_T.CLMA004T decreased size by 82.57 percent.

NOTE: PROCEDURE DATASETS used (Total process time):

real time	3:37.64	
user cpu time	5:13.71	
system cpu time	29.73 seconds	
Memory		104585k
OS Memory		126812k
Timestamp	08-Nov-11	00:03:49

Summary of SPDE Features and Benefits

- Partitioned data files
- Hybrid index (very cool – optimizes both high and low cardinality queries)
- Parallel index creation (also very cool – creates all indexes in a single pass)
- Parallel WHERE evaluation
- WHERE processor uses multiple indexes
- Parallel block interface to procedures
- Implicit sort for BY
- Parallel load (with the APPEND procedure)
- More than 2GB (2,147,483,647) rows (whether 32 or 64 bit O/S)
- Data set compatibility between SPDE and SPD Server

References

SAS 9.2 Scalable Performance Data Engine: Reference

<http://support.sas.com/documentation/cdl/en/engspde/61887/PDF/default/engspde.pdf>

Scalable Access to SAS® Data, Billy Clifford, SAS® Institute Inc., Austin, TX

<http://www.scsug.org/SCSUGProceedings/2004/Clifford%20-%20Scalable%20Access%20to%20SAS%20Data.pdf>

Very Large Dataset Performance as Measured Across Multiple SAS 9.1.3 Data Storage Options, Don Kros, Humana, Inc. Louisville, KY

<http://support.sas.com/resources/papers/proceedings09/263-2009.pdf>

An Annotated Guide: The New 9.1, Free & Fast SPDE Data Engine,

Russ Lavery, Ardmore PA, Independent Contractor
Ian Whitlock, Kennett Square PA

<http://www.nesug.org/proceedings/nesug06/io/io15.pdf>

Up and Out: Where We're Going with Scalability in SAS Version 9,

Cheryl Doninger, SAS Institute Inc.

<http://www2.sas.com/proceedings/sugi27/p279-27.pdf>



Questions?

