# **SAP BW on HANA Sizing Report**

Version 1.8 – July 2013





Sizing recommendations apply for <u>certified hardware only</u>. Please contact your hardware vendor for suitable hardware configuration.

Note that HANA is constantly being optimized. This might have impact on sizing recommendations, which will be reflected in this document. Therefore, check for the latest version of this document, the sizing software, and related notes.

The sizing guideline in this document refers to SAP BW on HANA only. Additional applications running on top of HANA are not covered in this document or in the report that is described. Please refer to application specific sizing guidelines for details in that case.

Actual memory requirements in HANA strongly depend on operational conditions which cannot be taken into account without analyzing the HANA system itself. The sizing results derived from this guideline are therefore to be understood as estimates based on best practices and experiences.

### **Purpose and Pre-Requisites**

### **Mission statement:**

Provide an easy to use, source database independent tool to predict resource requirements for a BW system on a SAP HANA database.

The sizing report /SDF/HANA\_BW\_SIZING is a convenient replacement of the database dependent sizing scripts provided in SAP note 1637145. Major advantages of the ABAP report as opposed to the sizing scripts:

- Easy to deploy and use no DB administrator required
- Versatile parameterization control resource consumption and speed
- Independent of source database compression
- Considers user defined future growth

The report is available with ST-PI 2008\_1\_7xx SP7. A preliminary version based on ST-PI 2008\_1\_7xx SP6 as well as an updated version for SP7 can be obtained from note 1736976. The report requires SAP NetWeaver BW 7.0 SP 1 or higher.

Important note for DB2 on iSeries (AS/400): Please implement note 1677958!

### /SDF/HANA\_BW\_SIZING: How it works

### Main characteristics of /SDF/HANA\_BW\_SIZING:

- Report obtains list of tables from ABAP dictionary (table DD02L)
- For each table, a sample data set is selected, and data field contents are analyzed.
   NOTE: MaxDB does not support data sampling. Database statistics data are used for sizing in this case. Expect longer report execution times or less accurate results for this platform.
- Calculates table sizes based on sampled data in ABAP memory and total record count → source database compression is taken into account automatically
- Automatic uplift for non-unicode source systems

© 2013 SAP AG. All rights reserved.

- Separation of tables in row store and column store
- Computation of overall SAP HANA memory requirements (including runtime memory) with table type specific compression factors.
- Parallel processing of subsets of tables to speed up processing time (degree of parallelism can be adapted to availability of work processes)
- Computation of future resource requirements based on relative or absolute yearly growth (on demand)
- Subset sizing (either for a complete BW system containing specified subsystem, or for the set of specified objects as isolated add-on for an already existing BW)
- Sizing of data that is stored in a Near-line Storage solution (NLS) is not in scope of this report

### /SDF/HANA\_BW\_SIZING: Optimized Sizing Procedure

### **Table Type Specific Compression Factors**

- Analysis of various PoC customer and internal test systems has revealed different average compression factors for InfoCubes, DSOs, PSAs, ChangeLogs, etc.
- Overall sizing result reflects data distribution across different table types much better now.

### **General Sizing Remarks**

- The report determines for which tables an uplift needs to be considered to provide memory for dynamic runtime objects, delta indexes, and merge results. For these tables, the table size is multiplied by a factor 2.
- Processing of row store also requires additional runtime memory. The runtime uplift is as large as the row store itself, however, it is capped by 100 GB.
- On each server node a fixed amount of memory has to be taken into account for
  - HANA statistics service, HANA name service (10 GB)
  - HANA system caches (40 GB)

## /SDF/HANA\_BW\_SIZING: Recommendations and Remarks

# Before you run the report /SDF/HANA\_BW\_SIZING, please consider the following:

Selection of parallel degree and availability of DIA work processes
 In order to run the report in parallel, make sure that for the specified dergree of parallelism a corresponding number of dialog work process is available.

#### Increase dialog runtime limitations

To prevent dialog processes from terminating due to time out conditions, we recommend temporarily setting the profile parameter rdisp/max\_wprun\_time to 0 (transaction RZ11), or check option "One table per RFC", which may slightly increase the report runtime.

#### Result accuracy

Due to the nature of data sampling, two consecutive executions of the report might deliver results which slightly differ from each other. However, the difference will not be significant enough to have an impact on sizing and can therefore be ignored.

We strongly recommend to run the report on up-to-date database statistics!

#### Precision Settings

The default precision (medium precision) of the report results in a sampling rate which is large enough to calculate reliable table sizes. However, tests have shown that even the low precision setting delivers results which are very close to medium precision, but returns in significantly less time.

#### Report Execution Time

Typical execution times for the report on a customer source database with ~1 TB:

- 9 minutes (low precision, 18 parallel processes)
- 45 minutes (medium precision, 18 parallel processes)

### /SDF/HANA\_BW\_SIZING: Selection Screen

Frogram Edit Goto System Help ②	Store output to file directly. If not set, a result screen will be displayed (wich still can be saved as a file at a later point of time).
Store output in file     X     File name     Extent.txt       Number of parallel procs     50	File name for output (relative to SAP work directoy, or absolute path name)
Suppress tables < TMB     A       One table per RFC     X       Precision	Number of parallel work processes used for sampling. Note: free dialog porcesses required for each process!
High O Medium O Low ©	In detailed list of tables, suppress those with sizes lower than 1 MB.
Scenario / Subset Selection	Use separate RFC per table if you cannot set max_wprun_time to 0 (may increase runtime)
Use system subset only List of top level InfoProv. Subset for other BW system BW system using this subset C Exclude listed objects O	Precision setting: decreases / increases data sample rate. Note: lower precision means less run time!
Future Growth Simulation	Specify list of top-level InfoProviders (InfoCubes, DSOs, Multi-Providers, Semantically Partitioned Objects, etc.) for subset / scenario sizing
Number of years     3       Relative growth (in %)     Image: Control of the second s	Future growth: calculate future HANA requirements for next few years, based on relative or absolute yearly growth. Enter number of years and amount of growth (absolute in GB or relative in %)
Non-active Data       Consider non-active data       X       WARM write-optimized DSOs       YTD*       to	Non-active data: switch on default handling for PSA, enter names of write-optimized DSOs that should be treated as warm' (see slide 7 for details)

### Subset / Scenario Sizing



Switch on by checking flag "Use system subset only".

You can determine the HANA size for parts of your source system by identifying all top level InfoProviders that need to be part of the scenario, or that you would like to exclude, respectively.

For inclusion scenarios, the report automatically considers all sizing relevant objects that depend on the provided list (e.g. all InfoObjects of an InfoCube, or all other InfoProviders in the downward data flow.

#### Object types that can be specified:

- InfoCubes
- DataStore Objects
- MultiProviders
- Semantically Partitioned Objects

- InfoObjects
- Hybrid Providers
- Open Hub Destinations
- InfoSets

## Subset / Scenario Sizing (cont.)

### Select type of subset:

Scenario / Subset Selection			
Use system subset only		ta	
Subset for other BW system	0	10	
BW system using this subset	۲		
Exclude listed objects	0		

#### Subset for other BW System:

- Report includes objects specified in list AND all dependent objects, but no other BW related system tables.
- Select this option if you plan to move the scenario to an already existing BW system and you want to determine the additional amount of memory that the scenario will require in the target system

#### BW System using this subset:

- Report includes objects specified in list, all dependent objects and all BW related system tables
- Select this option if you plan for a new BW system that only contains the specified scenario.

#### **Exclude listed objects:**

© 2013 SAP AG. All rights reserved.

- Report includes all tables except those belonging to the objects specified in list. Dependent objects will NOT be excluded.
- Select this option if you plan to move InfoCubes / DSO objects to Near Line Storage (NLS).

### Subset / Scenario Sizing: Subset Types

### **Source BW System**



- Business Data (Cubes, DSO, PSA, ...)
- BW System Tables
- Top Level Objects Specified in List
- Derived Dependent Objects (InfoObjects, etc.)

Subset for other BW System:

BW System using this subset:



Exclude listed objects:



### **Non-Active Data**

hot	Data is read/written frequently In Memory, additional memory required for dynamic objects (merge, intermediate results, etc.)						
warm	Infrequent access – no need to keep in memory all the time Relevant for non-active data concered on disk, loaded to memory only on demand, good candidate for displacement if memory runs short						
cold	Sporadic access – not stored in HANA DB Restricted to NLS capabilities						

### **Objects in SAP HANA can be marked as "non-active"**

Tables / partitions belonging to these objects are ...

- displaced from RAM (per column) with higher priority in case of RAM shortage (but only then) or when a cleanup is triggered
- loaded into RAM only when accessed (read access, merges) as usual

SAP BW automatically marks all PSA tables and write-optimized DataStore Objects for early displacement, so no extra maintenance or tuning is necessary. All other BW objects are treated as usual.

In addition, you can override the default handling for PSA tables and write-optimized DataStore Objects (e.g. DSOs used for frequent reporting). Note that this requires customization in BW metadata!

For more detail please refer to SAP note 1767880.

© 2013 SAP AG. All rights reserved.

### Requirements: SAP BW 7.30 SP08 and SAP HANA 1.0 SPS05

### **Non-Active Data: Recommendations and Remarks**



Enter list of write-optimized DataStore Objects that are considered to be "warm". Wildcards are allowed.

- NOTE: please enter technical object names, NOT the names of the database tables!
- Please refer to SAP note 1767880 for further information on the non-active data concept

### Do avoid too aggressive downsizing by marking objects as warm which are accessed frequently!

### /SDF/HANA\_BW\_SIZING: Output Screen / Report (1)



Provide additional disk space for exports / dumps / backups etc.!

### /SDF/HANA\_BW\_SIZING: Output Screen / Report (2)

SIZING DETAILS 	data [GB]	total [GB] incl. dyn.	total [GB] (non-act.)	Split of resources for row store and column store on master and slave servers (data only, data including dynamic runtime memory withou
Row Store Master Column Store Caches / Services	45 5 50	89 10 50	89 10 50 <mark>\</mark>	and with non-active data consideration)
TOTAL (MASTER)	100	149	149	Split of resources on master server. Note that master server contains both row and column
Slave Column Store Caches / Services TOTAL (SLAVES)	357 100 457	700 100 800	693 100 793	store data. Total (including memory for services / caches) needs to fit on one node
TOTAL (All Servers)	557	949	943	

Split of resources on slave servers. Total memory determines the number of slaves to distribute data on.

### /SDF/HANA\_BW\_SIZING: Output Screen / Report (3)

Time considered for growth: 3 years Assumed yearly growth rate: 25.0 %			future growth
Phys. memory per node:	512 GB	1024 GB	requirements
Memory Requirement (after 1 year ): Disk Space Requirement - data (after 1 year ): Disk Space Requirement - logs (after 1 year ): Number of Nodes incl. master (after 1 year ):	1041 GB 1041 GB 1041 GB 3	941 GB 941 GB 941 GB 1	future grov
Memory Requirement (after 2 years): Disk Space Requirement - data (after 2 years): Disk Space Requirement - logs (after 2 years): Number of Nodes incl. master (after 2 years):	1215 GB 1215 GB 1215 GB 4	1165 GB 1165 GB 1165 GB 3	
Memory Requirement (after 3 years): Disk Space Requirement - data (after 3 years): Disk Space Requirement - logs (after 3 years):	1369 GB 1369 GB 1369 GB	1319 GB 1319 GB 1319 GB	
SYSTEM INFORMATION	7		System inform run time, paran and DB syster
Report version :1.7Execution date and time:21.05.201314:03:51Runtime:00:03:29Parallel degree:50 (s)Precision settings:LConsider non-active data:YESUse system subset:NORDBMS:ORACLEOperating System:Linuxhostname:1d9168SID:PWS			

Sizing recommendations considering future growth with detailed sizing requirements for each year (only if future growth was selected)

System information, including report run time, parameter settings, operating and DB system information and SAP BW release

### /SDF/HANA\_BW\_SIZING: Output Screen / Report (4)

#### NON-ACTIVE DATA

All PSA tables are considered 'warm' by default

MANUALLY CLASSIFIED OBJECTS

WARM write-optimized DataStore Objects:

YTDWD YTDWD001 YTDWD002

TABLE DETAILS

#### MASTER NODE

Type Row	Table Name RSBATCHDATA	ABAP Size [MB] 33981.0	HANA SIZE [MB] 22654.0	rec count [million] 12.094	
Row Row	RSDDSTATEVDATA REPOLOAD	4480.5 4429.9	2987.0 2953.3	88.500 0.142	
Row more	RSDRV_TC	1.0	0.6	0.000	
Туре	Table Name	ABAP Size	HANA SIZE	rec count	
		[MB]	[MB]	[million]	
Col	RSDDSTATDTP	[MB] 1751.6	[MB] 1030.3	[million] 12.003	
Co1 Co1	RSDDSTATDTP BALDAT	[MB] 1751.6 1068.0	[MB] 1030.3 628.2	[million] 12.003 2.032	
Co1 Co1 Co1	RSDDSTATDTP BALDAT WBCROSSGT	[MB] 1751.6 1068.0 713.7	[MB] 1030.3 628.2 419.8	[million] 12.003 2.032 14.393	

#### SLAVE NODES

Туре	Table Name		ABAP Size	HANA SIZE	nonact.	na.	rec count	
			[MB]	[MB]			[million]	
Co1	/BIC/B0001824000	(C)	79821.4	9977.7	9977.7		215 .02	[
Co1	/B49/ADS0_PCL00		79069.0	26356.3	26356.3		215.429	
Co1	/B49/ADS0_SCL00		78782.5	26260.8	26260.8		214.572	
Co1	/BIC/EPHSDBLC03		58260.8	8323.0	8323.0		434.389	
Co1	/BIC/B0001345000	(P)	57897.9	5263.4	1226.3	(w)	121.267	
Co1	/B49/ADS0_XXXM00		57350.7	19116.9	19116 9		174 770	
Co1	/BIC/B0001764000	(C)	49528.2	0191.0	0191.0		100.702	
Col	/BIC/FPHSDDLC02		46132.7	6590.4	6590.4		439 292	

List of DataStore Objects that have been specified to be considered as non-active

Detailed size information for each master node table (row store, column store, sorted by size descending). Large tables may be good candidates for data cleansing to reduce memory requirements.

Report shows ABAP size and estimated size in HANA memory

Detailed size information for each column store table (sorted by size descending). Helps to identify good candidates for data cleansing to reduce memory requirements (e.g. very large PSA tables, etc.)

Marks tables which are handled specially as part of "non-active" data:

(w): Warm DataSources (by default)(M): Warm objects (as specified in object list)

Distinguishes between PSA (P), Change Log (C) and Error stack (E) tables

### **BW on HANA Sizing: Scale Out**

## If a single HANA node cannot accommodate data due to limited memory, data has to be distributed across multiple nodes (scale-out).

- Symmetric solution: 1 master node, n slave nodes, all on identical hardware
- Master node will handle system load and transactional load: ABAP system tables and general
  operational data of the BW are stored on the master node. Note that this includes both column store
  and row store data. DDL statements are executed on this node, global locks are acquired here.
- Slave nodes will handle OLAP queries as well as loading/staging/activation/merging. BW data (master data + cubes/DSOs/PSAs – all tables that have been generated by BW) is distributed across the column stores of all slaves. This ensures a balanced utilization of the available CPU and memory resources. Note that no column store data (except system tables) may be stored on the master node!
- Easy extensibility: When a new slave is added, BW data can quickly be reorganized to fit the new system. More information here: <u>http://help.sap.com/hana/hana\_db\_part\_en.pdf</u>
- Optional: stand-by node(s) can take over in case of node failure
- For more detailed information please refer to document "HANA\_BW\_Sale\_Out" attached to note 1736976.

### **BW on HANA Sizing: How to Size HANA Systems**

#### Sizing Report delivers necessary information at a glance:

MINIMUM SIZING RECOMMEN	NDATION - CURF	RENT					
	Phys.	memory per node:	512 GB	1024 GB			
Memory Requirement (Minimum Total):943 GB843 GBDisk Space Requirement - data (Minimum Total):943 GB843 GBDisk Space Requirement - logs (Minimum Total):943 GB843 GBNumber of Nodes incl. master (Minimum Total):31							
<ul> <li>NOTE:</li> <li>Please carefully read documentation attached to SAP NOTE 1736976 for a detailed description of the sizing procedure and its results!</li> <li>Disk space calculation for data does not include space for backups, dumps, etc. anymore. This space has to be provided on additional disk volumes.</li> </ul>							
SIZING DETAILS 	data [GB]	total [GB] incl. dyn.	total [GB] (non-act.)				
Row Store Master Column Store Caches / Services TOTAL (MASTER)	45 5 50 100	89 10 50 149	89 10 50 149				
SLAVES: Slave Column Store Caches / Services TOTAL (SLAVES)	357 100 457	700 100 800	693 100 793				
TOTAL (All Servers)	557	949	943				

Total memory including data, runtime, services (on each node)

#### Number of nodes (slaves + 1 master for scale-out configurations)

Detailed split of memory requirements for data, runtime objects and services with / without impact of non-active data. If scale-out landscape is required, master and slave memory requirements are shown separately

### **BW on HANA Sizing: How to Size HANA Systems**

#### **Remarks:**

HANA can allocate at most 90% of the physical memory of each node. In case of a 512GB node the available memory is 460 GB, for a 1TB node 921 GB will be available.

In both cases, HANA baseline consumption is 50GB on each node for additional services, caches, etc., **leaving 410 GB or 871 GB free**, respectively.

In case of scale-out configurations, row store and non-generated column store tables must fit in one node.

If total row store is significantly larger than 200 GB, **consider house keeping measures** (see SAP notes 706478, 1729988, and 1829728 for details)

Disk sizing shows only the **disk space that is required for the persistence of the HANA data files**. Additional disk space must be provided for:

- Executables, instance directories, etc.
- Dumps
- Exports, backups

We recommend reserving 2x - 3x of the total memory as additional disk space for these data.

### BW on HANA Sizing: How to Size a Single Node System



### BW on HANA Sizing: How to Size a Scale Out System



### Grand Total:

149 GB + 793 GB = 942 GB → Scale-out: 3 nodes with 512 GB each

### **Frequently Asked Questions**

 Do I have to apply any further calculation on the report results to get the final memory requirements?

No, the sizing report does all the math for you. You can find the overall memory and disk space requirements for different hardware configurations in the section "SIZING RECOMMENDATION – CURRENT" of the report. The information in section "SIZING DETAILS" helps understanding how the overall memory requirements are mapped to master and slave servers.

Why is the size of a table calculated by the report different to the size reported by the database (e.g. in transaction DB02)?

The sizing report reads a sample set of records from the DB, analyzes the sample size in its ABAP representation, and extrapolates the total size of the table if it was entirely loaded into ABAP memory, using the total row count information of the DB. By looking at the ABAP representation, the impact of DB specific features like intrinsic compression or special data formats on sizing is automatically eliminated, and the report can estimate the size of the table in ABAP by applying a suitable compression factor. In the output of the report you can find the ABAP size of tables along with the estimated HANA size. Due to the different representation of data in ABAP and the underlying DB system, the DB table sizes may significantly differ from the ABAP sizes.

 In section "sizing recommendation - CURRENT" the report prints "n/a", but section "sizing DETAILS" shows proper results

The values in section "SIZING RECOMMENDATION – CURRENT" are calculated for specific hardware configurations (currently for server nodes with 512GB and 1024GB RAM). While the overall result is valid, a mapping to the existing server configurations was not possible. One reason may be that the master data cannot be stored in a single server node due to its size.

## **Thank You!**

**Thomas Becker** 

TIP HANA BW Performance Services mailto:Th.Becker@sap.com



THE BEST-RUN BUSINESSES RUN SAP™