WINTER CORPORATION

September 10, 2002

Mr. Ian Walsh
Director, Product Marketing
SAP
San Jose, CA 95119

Dear Mr. Walsh:

Winter Corporation has conducted an independent audit of the benchmark test referred to as *The 5 Terabyte Showcase* using the SAP Business Information Warehouse version 3.0B. Our summary of the test conducted, the results obtained, and the audit process appears in this letter and its attachments.

Role of the Auditor

As the auditor acting in a strictly objective capacity, our role was to: witness the benchmark setup and test; determine whether SAP performed the test as described below; determine whether the demonstration functioned correctly; and, determine whether the performance of the system had been properly measured.   The audit fee was paid by SAP under an agreement in which this letter was written solely at the discretion of the auditor; the content and presentation of the letter is strictly under the control of the auditor; and, the letter must be published in its entirety with its appendices. Extracts from the letter and appendices may be published only with the written approval of Winter Corporation.

Background

SAP Business Information Warehouse (SAP BW) is an application component of mySAP Business Intelligence that provides a complete solution for enterprise wide data capture and analysis. SAP BW offers a data warehousing platform and suite of tools to gain insight and knowledge about business performance. It supports an Operational Data Store (ODS) as well as multidimensional models for Online Analytical Processing (OLAP).

SAP conducted a customer survey that revealed that many installations had databases approaching one terabyte of data and some were already in the one to two terabyte range. To address these rapidly growing customer requirements, SAP initiated this project. The project aims to provide both a feasibility study and a demonstration of how SAP BW 3.0B would perform with 20,000 concurrent users accessing a five-terabyte database. It was a joint project between SAP, Sun Microsystems, and Oracle conducted at SAP in Waldorf, Germany and the Sun Garrison Lab in Palo Alto California.

Auditor's Statement

SAP did conduct a benchmark using the SAP Business Information Warehouse version 3.0B. Approximately 3.6 terabytes of raw data was loaded into an Oracle 9i database

running on a Sun Fire 15K server. The database occupied a total 5.5 terabytes of storage, which included 1.1 terabytes of detail data, 0.2 terabytes in aggregates, and 0.6 terabytes as Oracle indexes. The total load time was 169 hours and 27 minutes.

SAP conducted two query workload simulations. The first test consisted of queries that only accessed aggregate views of the data. As suggested by the auditor, a second test was conducted whereby one of the fifteen query types accessed detailed data. The navigation steps and average response times that SAP reported for these tests were measured properly.

A summary of the benchmark process, benchmark results and audit process can be found in the following appendices:

- o *Appendix A - Benchmark Configuration*
- o *Appendix B - Benchmark Process*
- o *Appendix C – Workload Methodology*
- o *Appendix D – Query Workload Simulation*
- o *Appendix E – Benchmark Reports*
- o *Appendix F – Results Obtained*
- o *Appendix G – Audit Process*

A complete description of the benchmark is also available in a technical white paper produced by SAP and Sun Microsystems. This document can be found at (www.sap.com/bi/brochures.asp) or (www.sun.com/sap/collateral/index.html).

Additional Comments

Terabyte scale performance was proven here for an important set of business intelligence requirements, but no single benchmark can address the full range of possibilities. Prospective terabyte scale users of SAP BW should take their requirements into account and evaluate this benchmark test and its results in depth along with other information concerning large-scale performance.

In conclusion, Winter Corporation finds that the tests described in this letter demonstrate that the SAP BW product is capable of loading, managing and querying a database containing 3.6 terabytes of raw data with good response time. In the test, the system handled many typical business intelligence queries simultaneously -- thus simulating the servicing of a large population of concurrent users. Several aspects of scalability that were successfully demonstrated in the benchmark add up to an important step forward for the SAP Business Information Warehouse.

Sincerely,
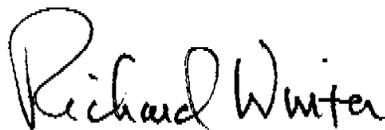
Richard Winter
President
Winter Corporation
richard.winter@wintercorp.com

David J. DiMare
Senior Database Consultant
Winter Corporation
david.dimare@wintercorp.com

Appendix A - Benchmark Configuration

The following hardware and software was used to support the SAP BW benchmark:

Central Server – The benchmark was run on a Sun Fire 15K Server with 72 - 900 megahertz processors and 144 gigabytes of main memory. The Sun Fire 15K uses the Sun Fireplane Interconnect system, which is a shared-memory protocol using UltraSPARC III+ processors. This server functioned both as the database server and SAP BW Server.

Driver Server – A second server was used to drive the transaction workload. It was a Sun Fire 4800 Server with 12 - 900 megahertz processors and 24 gigabytes of main memory. A private network connection was established between the driver server and central server to eliminate interference from network traffic not associated with the benchmark.

Storage Subsystem – Twenty-four Sun StorEdge T3 RAID disk arrays were connected to the server via fibre channel to support data storage requirements. Each T3 contained 9 - 73 gigabyte disks and was configured as single RAID device via the hardware RAID controller. Veritas Volume Manager was also used to manage logical volumes defined across the disk arrays.

The twelve disk arrays used to support the Oracle database were set up as RAID-0 devices. A RAID-0 configuration provides data striping across disks but has no data redundancy (neither parity or mirroring). I/O performance is enhanced because read and writes are done in parallel across devices. This type of storage configuration may be appropriate for temporary or reproducible data but is not appropriate for mission critical applications unless combined with another layer to add reliability.

Software – SAP Business Information Warehouse version 3.0B, Solaris 2.9, Oracle 9i version 9.2.0, and Veritas Volume Manager 3.2

Appendix B - Benchmark Process

The SAP BW benchmark consisted of the following steps:

1) Load master data into the SAP BW Persistent Staging Area (PSA). The master data included data for 100,000 customers, 200,000 products, 100 sales organizations, 20 sales divisions, 42 countries, and 8 distribution channels. This data was later used to populate the dimensions of the SAP BW InfoCube. The size of the master data was approximately 400 megabytes.

2) Load detail sales transaction data stored in 960 ASCII files to the PSA. The 960 files represented monthly sales for 10 years across 8 distribution channels (10 years * 12 months * 8 distribution channels). Each file contained approximately 12 millions sales transactions, which represented 400,000 transactions per day. The data was consolidated into 80 PSA tables, each table representing one year of data for a single distribution channel. This represented a total of 11.7 billion records or approximately 3.6 terabytes of data.

3) Build Oracle indexes for the PSA tables. There was one compound index built for each of the 80 PSA tables. These indexes had been dropped prior to the PSA load in order to speed up load processing but were required for the cube load. This is normal procedure for loading large amounts of data into an Oracle database.

4) Load 80 InfoCubes from the PSA. Similar to the PSA, each cube represented sales data for a single distribution channel for one year or approximately 144 million records (12 million per file * 12 months). The InfoCubes represented the detail data stored in a star schema database design, occupying about 1.1 terabytes of storage in Oracle.

5) Load the Operational Data Store (ODS) from the PSA. The ODS contained a consolidated view of the most recent two years of data stored in flat database tables. There were 4 ODS objects, each representing 6 months of data for all distribution channels; therefore each ODS object contained approximately 580 million records ((144 million records * 8 divisions * 2 years) / 4 ODS objects)

6) Build Oracle bitmap indexes for the InfoCubes. There were 8 Oracle bitmap indexes defined per InfoCube or 640 indexes (8 indexes * 80 InfoCubes). Similar to the PSA load, these indexes were also dropped prior to the cube load to improve load performance.

7) Generate aggregate views of the InfoCubes. There were 10 aggregates defined per InfoCube. Four of the aggregates were built to generate other aggregates and therefore only six of the aggregates were actually used for reporting. The six aggregates ranged in size from approximately 8 rows to 200,000 rows.

8) Simulate a multi-user query workload by running many reports simultaneously. There were 15 different sales reports that were run continuously in a loop using random attributes for report filtering. To simulate multiple users, there were several instances of the loop running in parallel.

Appendix C - Workload Methodology

The methodology SAP used for establishing a multi-user query workload was based on their QuickSizer approach. This is a methodology that SAP uses to determine hardware requirements for new customers. The primary metric used in this methodology is a navigation step (Nav), which is the equivalent of a database request. Based on this methodology it was determined that 20,000 users would generate 158,991 navigation steps in one hour, running simple database requests. The goal of the workload simulation was to meet or exceed the number of navigation steps calculated for the 20,000 users.

The QuickSizer approach classifies users into three categories along with the number of navigation steps each would execute in one hour: normal – 1.11, advanced – 11.11, and power – 33.33. Based on the customer survey the SAP conducted, it was determined that users could be divided into the three categories as follows: 71% normal, 27% advanced, and 2% power. Applying these percentages to the target of 20,000 users is equivalent to 14,200 normal, 5400 advanced and 400 power users.

Queries were also placed into three categories with a relative resource consumption assigned to each: easy – 1 unit, medium – 1.5 units, and heavy - 5 units. Based on SAP tests, it was established that users would execute queries as follows:

User Type	Easy Query - 1	Medium Query - 1.5	Heavy Query – 5
Normal – 1.1 Nav/h	80%	20%	0%
Advanced – 11.11 Nav/h	50%	50%	0%
Power – 33.33 Nav/h	0%	0%	100%

Using the table above, the total number of navigation steps that 20,00 users would generate was calculated as follows:

$$Nav/h = 14{,}200 * 1.1 * (.8 * 1 + .2 * 1.5 + 0 * 5)$$
$$+ 5400 * 11.11 * (.5 * 1 + .5 * 1.5 + 0 * 5)$$
$$+ 400 \ * 33.33 * (0 * 1 + 0 * 1.5 + 1 * 5)$$

$$= 158{,} 991 \text{ Navigation Steps Per Hour}$$

The goal of the workload simulation was to achieve the 158,991 navigation steps per hour by submitting a fixed number of requests, measuring the response time, and extrapolating the navigation steps for one hour of activity.

Appendix D – Query Workload Simulation

To simulate the query workload, 15 sales reports were submitted from within a loop running on the driver processor. The reports were submitted as HTTP requests from the driver processor to the central server. Multiple parallel processes were established to submit many reports simultaneously.

The report parameters, including the InfoCube used to satisfy the request, were chosen at random. Although the parameter changed when a new loop was initiated, within a single loop the parameters remained the same. This was intended to simulate a user accessing multiple reports for a single unit of data (e.g. sales for a customer, followed by credits issued to the same customer).

A single report could only access data from one cube and therefore reports did not span more that one year or more than one distribution channel. To support this type of access, a SAP BW multicube would have been required. Multicubes were not used for this benchmark.

Two simulation tests were measured. In the first test, the queries were tuned to access aggregates that had been generated from the InfoCubes. None of the reports ran against data that was stored in the fact tables, which resulted in very little I/O activity. In the second simulation test, one of the fifteen queries was modified to obtain results by going directly to the InfoCube.

Simulation Test #1

SAP BW was configured with 1 central instance and 10 application servers, which provided 11 instances of the application. Each instance had 60 dialog processes, which provided 660 parallel processes (11 instances * 60 dialog processes each), each submitting requests.

Each of the 660 connections ran through 5 loops of all 15 sales reports. Therefore, a total of 49,500 reports were run during the simulation (660 connections * 5 loops * 15 reports). All reports ran to completion and the results were extrapolated to determine the navigation steps executed per hour.

Simulation Test #2

During this test, SAP BW was also configured with 1 central instance and 10 application servers. Two of the instances were dedicated to processing the query that accessed the InfoCubes. It was expected that these queries would generate high I/O. These instances had 5 dialog processes whereas the other 9 had 60 dialog processes. This provided a total of 550 parallel processes ((9 instances * 60 processes) + (2 instances * 5 processes)).

The 2 instances dedicated to the heavy I/O query processed 2 loops of 15 reports whereas the other 9 instances processed 5 loops. Therefore, a total of 40,800 reports were run during this simulation ((9 instances * 60 dialog processes * 5 loops * 15 reports) + (2 instances * 5 dialog processes * 2 loops * 15 reports)). All reports ran to completion and the results were extrapolated to determine the navigation steps executed per hour.

Appendix E - Benchmark Reports

The following reports were used in the query workload simulation:

Report Title	Input Parameters	Results	Used InfoCube or Aggregate	InfoCube or Aggregate Row Count	Joins	Columns Returned
Product Analysis	Product Number	Orders by product	Aggregate	200,000	3	11
Sales Volume	Customer Number	Sales by customer	Aggregate	100,000	3	8
Credit Memos	Customer Number	Credits by customer	Aggregate	100,000	3	8
Incoming Orders	Customer Number	Orders by customer	Aggregate	100,000	3	9
Customer Returns	Customer Number	Sales & returns by customer	Aggregate	100,000	3	11
Customer Statistics	Customer Number	Sales, orders, returns, & credits by customer	Aggregate	100,000	3	12
Customer Returns	Month / Year	Orders & returns by calendar month	Aggregate	12	3	7
Customer Incoming Orders	Sales Organization & Division	Sales and credit amounts by sales organization & division	Aggregate	2,000	4	11
Order Entry Analysis	Sales Organization, & Month / Year	Orders by calendar month for a sales organization	*(1)* Aggregate	12	3	6
			(2) InfoCube	145,000,000	6	7
Product Profita-bility Analysis	Product Number	Sales volume by product	Aggregate	200,000	3	6
Top Customers	Customer Number	Sales, orders, & returns by customer	Aggregate	100,000	3	7
Incoming Order Analysis	None	Orders by calendar month for 1 year	Aggregate	12	3	6
Sales Volume/ Cost Analysis	None	Sales by calendar month for 1 year	Aggregate	12	3	6
Distribution Channel Analysis	None	Sales by distribution channel	Aggregate	8	3	6
Sales Data Overview	None	Sales & orders by division	Aggregate	20	3	11

(1) Aggregate used for Simulation Test #1
(2) InfoCube used for Simulation Test #2

For those reports that required input parameters, a range of values was provided. The low end of the range was chosen at random and the high end was calculated by adding an offset to the low-end value:

Input Parameter	Offset
Customer	10
Division	5
Materials	50

| Sales Organizations | 10 |
| Calendar Month | 3 |

Therefore, the offset determined the number of items that appeared on the report (e.g. sales by customer would have a report showing 10 customers).

During the first simulation, it was established that all 15 reports were accessing aggregate views of the data and none of the reports were accessing the InfoCubes. At the auditor's request a second simulation was added to the benchmark. In the second simulation, a modification was made to allow the *Order Entry Analysis* report go directly against the InfoCube instead of the aggregate.

Appendix F - Results Obtained

The following results were measured for each of the SAP BW benchmark test phases:

Test Phase	Results
Load/Build Phases	
- Load ASCII data into PSA	44:50:25 hours
- Create PSA indexes & database statistics	35:00:00 hours
- Load data from PSA into ODS & InfoCubes	65:17:19 hours
- Create InfoCube bitmap indexes & database statistics	09:12:23 hours
- Build aggregates	14:14:45 hours
Total Load/Build Time	169:27:04 hours
Query Workload Simulation	
- Simulation Test #1	356,000 navigation steps Average Response Time: 2.5 seconds (aggregates)

- Simulation Test #2	208,000 navigation steps
	
	Average Response Time:
	  5 seconds (aggregates)
	  335 seconds (cubes)

Appendix G - Audit Process

The following steps were conducted to audit the SAP BW benchmark:

1) Reviewed the SAP project plan for the entire project.

2) Reviewed the hardware and software configuration for the driver and benchmark servers to ensure the specifications matched the configuration supplied by SAP:

   - Checked the Solaris configuration and setup by running various UNIX diagnostic commands.

   - Reviewed the Veritas configuration supplied by SAP and checked that devices either had been defined as file systems or were referenced in Oracle as raw partitions.

   - Checked the Oracle installation by reviewing the *init.ora* file, the database creation script, and by running various PL/SQL commands.

3) Examined the scripts that were used to generate the 3.6 terabytes of transaction data and 400 megabytes of master data. Checked that the data had been generated with a skewed distribution to reflect a realistic operating environment.

4) Verified the ASCII input files had been generated and for multiple files, checked that file counts matched the numbers provided by SAP.

5) To verify the load of the ASCII data to the PSA, a random check of multiple PSA objects was performed:

   - Checked the load processes and times by reviewing random InfoSources in the SAP BW Administrator Workbench

   - Checked Oracle row counts on several of the underlying PSA tables using PL/SQL

6) To verify the movement of data to the InfoCubes and to verify that aggregates had been generated, a random check of multiple InfoCubes and aggregates was performed:

   - Checked the load processes and times via the SAP BW Administrator Workbench

   - Checked Oracle row counts on several aggregates and InfoCubes using PL/SQL.

7) Reviewed the perl scripts that were used to simulate the query workload.

8) Examined the 15 sales reports used in the workload simulation by running each report manually with random input.

9) Reviewed the Oracle query plans for the 15 sales reports.

10) Witnessed the execution of the query workload simulation with the SAP benchmark team at their facilities in Waldorf, Germany. While the simulations were running, the following system metrics were captured by the auditor:

   - Server utilization and I/O performance statistics (benchmark server only)

   - Oracle performance report using Oracle Statspack

   - SQL Trace using the SAP BW Database Performance Monitor