

## High Scalability with SAP Business Information Warehouse on IBM DB2® Universal Database™ for Linux, UNIX®, and Windows®

Formerly known as mySAP Technology, SAP NetWeaver is an application and integration platform that helps unify and align people, information, and business processes across technologies and organizations. SAP NetWeaver is the technical foundation for all SAP solutions, from the mySAP Business Suite through SAP Industry Solutions to SAP xApps.

### **SAP Business Information**

**Warehouse (SAP BW)** is positioned as a central business intelligence platform and is one of the core components of SAP NetWeaver. The success of SAP BW is proven by its large number of installations in the market.

**IBM DB2 Universal Database Enterprise Server Edition (DB2 ESE) with the Database Partitioning Feature (DPF)** provides a high-performance mechanism to support large databases, and offers greater scalability in Massively Parallel Processors (MPPs) and Symmetric Multiprocessor (SMP) environments. It is ideal for applications requiring parallel processing, particularly data warehousing and data-mining (see sidebar on the next page).

This article, based on a recent case study, refers to the compelling benefits of having IBM DB2 UDB<sup>1</sup> as the underlying database for your SAP BW.

<sup>1</sup> This study used IBM DB2 UDB Enterprise-Extended Edition (DB2 EEE), the predecessor of DB2 Enterprise Server Edition with DPF.

### **Scalability**

*Scalability* is the level of flexibility needed to upscale on a given environment in order to adapt to growth according to business needs.

Scalability is crucial for applications with a growing volume of data like SAP BW. Ideally, application throughput and database performance would scale in a linear manner as more machines were added to the DB2 UDB cluster.

Scalability is not only the key to high performance, it is also very important for running a large SAP BW system at the best price/performance ratio. It is often less costly to distribute a large SAP BW database over multiple medium-sized machines than to run it on one large SMP machine. For that reason, it is important that DB2 UDB scales well in a distributed environment.

For this study, a shared-nothing architecture was utilized based on an IBM Unix Cluster (RS/6000 SP) interconnected by a high-performance, packet-switched network (SP switch). Scalability may have been even better when using a single large SMP machine with a shared-memory architecture. In this case, multiple logical database partitions communicate via shared memory, which is faster than communication via an SP switch. A large SAP BW system can also be distributed over several large SMP machines connected by fast communication channels (e.g., switch technology or Gigabit Ethernet).

Andreas Christian,  
Karl Fleckenstein, and  
Stephan Kammerer,  
SAP IBM DB2 UDB  
Development/Porting Center,  
IBM Germany

The I/O configuration in this test focused on high availability of real customer scenarios.

### **System Setup**

An IBM RS/6000 SP — now IBM eServer pSeries (see sidebar on the next page) — with 4 wide nodes acted as the SAP BW servers.

One SAP BW system was installed on a SSA storage system with 580 GB of disk space. This system was used for SAP BW load tests from Persistent Staging Area (PSA) into InfoCubes and for SAP BW query tests.

A second SAP BW system was installed on an ESS system with 320 GB of disk space. This SAP BW system was used for load tests from file into PSA, and from PSA into Operation Data Store (ODS). All disk subsystems<sup>2</sup> were configured as RAID-5.

### **Results**

The study demonstrates the high scalability of DB2 ESE with DPF for

<sup>2</sup> IBM 7133 Serial Disk System Advanced Model D40 (SSA) and IBM Enterprise Storage System (ESS).

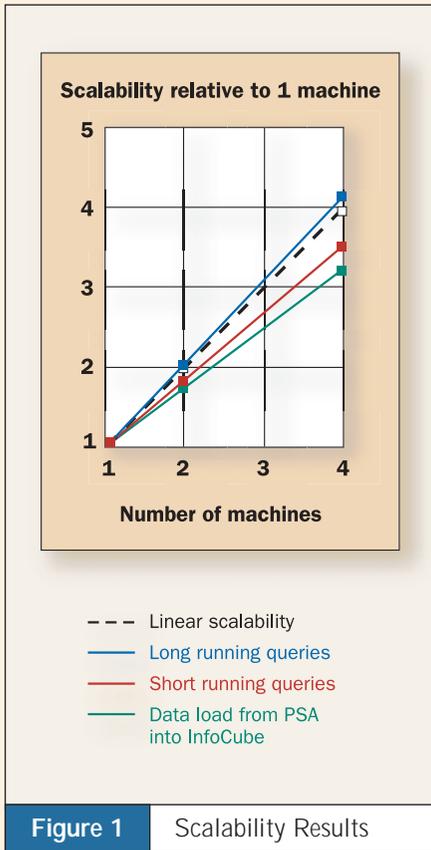


Figure 1 Scalability Results

### IBM DB2 Universal Database

The DB2 UDB provides consistently high levels of transaction processing. IBM DB2 ESE with DPF is the only parallel database on the Unix and Windows market today available for SAP BW. This technology ensures faster response times for complex queries such as those associated with online analytical processing (OLAP) applications handled by SAP BW.

The Database Partitioning Feature (DPF) allows DB2 ESE customers to partition a database within a single system or across a cluster of systems. The DPF capability provides the customer with multiple benefits, including scalability to support very large databases or complex workloads along with increased parallelism for administration tasks, like index creation, backup, and restore.

### IBM pSeries

IBM pSeries — “p” as in performance — is the high-performance and highly scalable Unix line within IBM. In the commercial arena, the pSeries family offers highly scalable SMP systems as well as clustered environments. The pSeries 690 as a high-end system can be used as a large DB server or as a partitioned system for consolidation purposes. The RS/6000 SP “shared nothing” architecture has evolved into the pSeries cluster offerings, allowing for excellent horizontal scalability while maintaining centralized landscape management.

SAP BW. Based on the good results on a single machine, we found an impressive performance increase was achieved when adding additional machines (see Figure 1).

#### Data Load Scalability

- For loading data from the *Persistent Staging Area (PSA) into an InfoCube*, the scalability factor was **3.28** when going from 1 machine to 4 machines.
- For loading data *from a flat file into the PSA*, the scalability factor was **1.69** when going from 1 machine to 2 machines.
- For loading data *from the PSA to the Operation Data Store (ODS)*, the

scalability factor was **1.71** when going from 1 machine to 2 machines.

#### Query Scalability

- Single-running queries:
  - Query runtimes decrease with an increasing number of machines.
  - Long-running queries scale at least linear.
  - Short-running queries remain short running as the number of database partitions and machines increase.
- Multiple concurrent queries:
  - Near-linear scalability for multiple queries executed concurrently.

- For long-running queries, the scalability factor was **4.11** when going from 1 machine to 4 machines.
- For short- to medium-running queries, the scalability was **3.58** when going from 1 machine to 4 machines.

To receive a copy of the full study, please visit [www7b.boulder.ibm.com/dmdd/library/techarticle/0208christian/0208christian.html](http://www7b.boulder.ibm.com/dmdd/library/techarticle/0208christian/0208christian.html) or contact IBM SAP Information Service at [isicc@de.ibm.com](mailto:isicc@de.ibm.com).

For more information on SAP and IBM DB2, please visit [www.ibm.com/software/data/partners/ae1partners/sap](http://www.ibm.com/software/data/partners/ae1partners/sap) or the IBM SAP Alliance Web site at [www.ibm-sap.com](http://www.ibm-sap.com). ■

DB2, DB2 Universal Database, the e-business logo, IBM, the IBM logo, MagStar, pSeries, RS/6000, SP and Tivoli are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both. SAP, mySAP.com and R/3 are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world. Other company, product or service names may be trademarks or service marks of others.