Scalability Study
SAP Business Information Warehouse on
## DB2® Universal Database™ EEE for Linux, UNIX® and Windows®

Overview

SAP Business Information Warehouse (SAP BW) is the component of mySAP Business Intelligence that delivers enterprise-wide data warehousing, a business intelligence platform and a suite of BI tools. SAP BW is a successful product, proven by a large number of installations in the market.

DB2 Universal Database Enterprise-Extended Edition (DB2 UDB EEE) provides a high performance mechanism to support large databases and offers greater scalability in Massively Parallel Processors (MPP's) and Symmetric Multiprocessor (SMP) environments. Ideal for applications requiring parallel processing, particularly data warehousing and data mining.

Scalability studies already exist for "classical" data warehouses. This particular study has been carried out because there are crucial differences between SAP BW and other data warehouses:

- SAP BW dynamically generates detailed SQL queries; sometimes 80 database tables or more are included in one SQL statement.

- SAP BW uses dynamically generated tables to store intermediate results. These tables are accessed from other SQL queries in additional SAP BW query processing steps.

SAP Business Information Warehouse

SAP BW Application Server
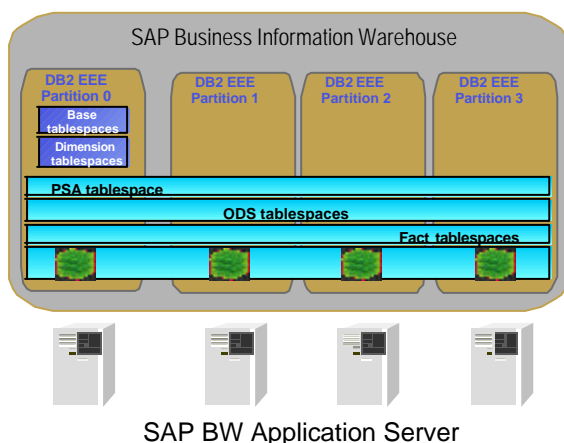
Results

The study demonstrates the **superior scalability of DB2 UDB EEE for SAP BW.**
Based on the good results on a single ᴧ RS/6000 SP node, an impressive performance increase was achieved when adding additional SP nodes.
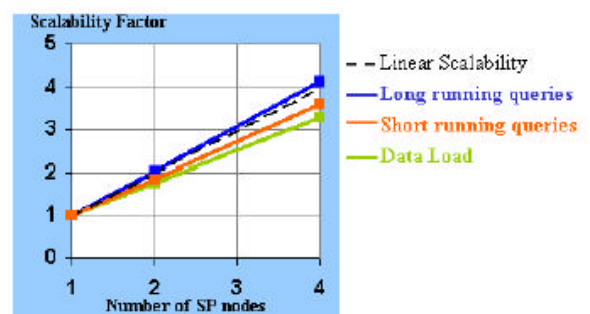
Data Load Scalability:
- For loading data from the Persistent Staging Area **(PSA) into an InfoCube** (SAP BW 2.1C), the scalability factor is **3.28** when going from a one-node to a four-node configuration**.**
- For loading data from **a flat file into the PSA** (SAP BW 3.0A), the scalability factor is **1.69** when going from a one-node to a two-node configuration.
- For loading data from the **PSA to the Operation Data Store (ODS)** (SAP BW 3.0A)**,** the scalability factor is **1.71** when going from a one-node to a two-node configuration.

Query Scalability:
- Single-running queries (SAP BW 2.1C):
  - ✓ The query runtimes decrease with increasing number of SP nodes.
  - ✓ Long-running queries scale at least linear.
  - ✓ Short-running queries (< 1sec) remain short running as the number of database partitions and SP nodes increase.
- Multiple concurrent queries (SAP BW 2.1C):
  - ✓ Near-linear scalability for multiple queries executed concurrently.
  - ✓ For **long running queries**, the scalability factor is **4.11** when going from a one-node to a four-node configuration**.**
  - ✓ For **short- to medium-running queries**, the scalability is **3.58** when going from a one-node to a two-node configuration.
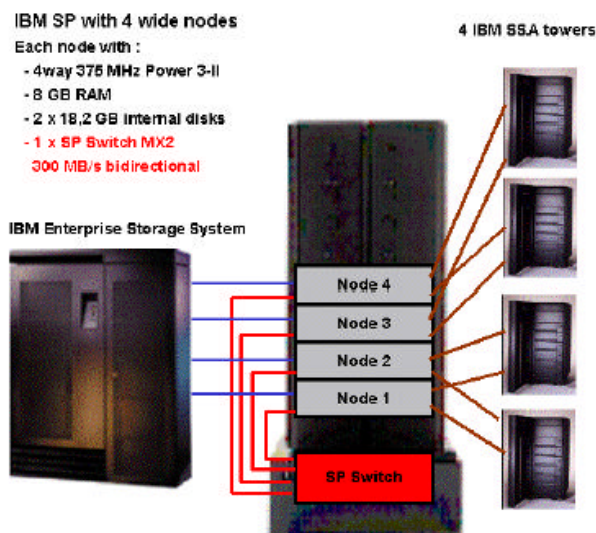
System setup

Hardware:
- ✓ IBM RS/6000 SP (now IBM ⌃ pSeries)
- ✓ IBM 7133 Serial Disk System Advanced Model D40 (SSA)
- ✓ IBM Enterprise Storage System (ESS)

Software:
- ✓ IBM AIX 4.3.3
- ✓ IBM DB2 EEE for Unix 7.2
- ✓ SAP BW 2.1C, SAP BW 3.0A

An IBM RS/6000 SP with 4 wide nodes based on POWER3 technology acted as the SAP BW servers. All disk sub-systems (SSA, ESS) had been configured as RAID-5.

The SAP BW 2.1C was installed on a SSA storage system with 580 GB space distributed on 65 SSA disks. The SAP BW 3.0A was stored on an ESS system with 320 GB of space.

IBM SP with 4 wide nodes
Each node with :
- 4way 375 MHz Power 3-II
- 8 GB RAM
- 2 x 18,2 GB internal disks
- 1 x SP Switch MX2 300 MB/s bidirectional

4 IBM SSA towers

IBM Enterprise Storage System

Node 4 / Node 3 / Node 2 / Node 1 / SP Switch

IBM ⌃ pSeries

"p" as in performance - is the high-performance and highly scalable Unix line within IBM.

In the commercial (OLTP) arena the RS/6000 SP is typically found as a consolidation system. Whereas in a SAP BW implementation users can fully exploit it's massive parallel capabilities.

Serial Storage Architecture (SSA)

A storage solution with high capacity and versatility for bridging greater distances at high performance levels.

Enterprise Storage Server (ESS)

ESS provides the basis for data management, high availability and storage sharing in heterogeneous environments with advanced management functionality's.

Scalability

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

For example with a scalable solution the response time of long running queries can be reduced by adding additional CPUs to the existing machines. Without a scalable solution the only way to reduce response time is to exchange the exisitng CPUs by high performance processors or where this is not applicable the actual hardware needs to be substituted with a more efficient one.

For scalable solutions like SAP BW on DB2 UDB EEE, performance can be improved by adding additional CPUs and/or machines in order to spread the total workload over more CPUs. Adding machines is usually less expensive than switching processors or exchanging the existing hardware environment.

The same applies to I/O bottlenecks. It is often more cost-effective to improve I/O performance by extending a MPP cluster (adding machines with dedicated I/O adapter) rather than upgrading the existing I/O sub-system.

This study indicates that SAP BW on DB2 UDB EEE is a highly scalable solution allowing customers to keep their systems at the best price/performance ratio. HW bottlenecks can be resolved by adding additional CPUs or machines to a MPP cluster instead of upgrading or replacing a large SMP.

Get a free copy of the whole study:
www7b.boulder.ibm.com/dmdd/library/techarticle/0208christian/0208christian.html

More Information:
SAP and IBM DB2:
www.ibm.com/software/data/partners/ae1partners/sap
IBM SAP Alliance:
Www.ibm-sap.com

Contacts:
IBM SAP Information Service
E-Mail: isicc@de.ibm.com
FREE! Subscriptions:
DB2 Magazin: www.db2mag.com