

Scalable Clusters with IBM DB2 Universal Database, Intel Servers and Network Appliance Storage

Technical
Bulletin

intel®

September 2001

Revision 1.0

IBM®


NetworkAppliance®

Table of Contents

Executive Summary	2
DB2, Intel Servers and Network Appliance Filers	3
About DB2	3
Intel AD450NX Server	3
NetApp F840 Filer	3
System Design	4
Experimental Design	5
DB2 Configuration	5
Network Appliance Filer Configuration	6
Physical Disk Layout	6
Volumes and Qtrees for Data Storage	6
CIFS Shares	7
Connecting the Filer to DB2	7
Networking Issues	8
Configuration Parameters	8
Server Configuration	9
Results Overview	10
Scaling from One Node to Two Nodes	10
Scaling from Two Nodes to Four Nodes	12
Scaling from Two Servers/Two Filers to Four Servers/Two Filers	13
Summary	15
Appendix 1. Performance Tuning: Experiences, Lessons and Recommendations	16
PREFETCHSIZE	16
Bufferpools and Sortheap	16
Tablespaces	16
Filer Configuration Parameters	16
Network Isolation	16
Filer Gigabit Ethernet	17
Iometer	17
Monitoring	17
Appendix 2. DB2 Scripts	18
Appendix 3. DB2 Configuration Parameters and Environment Variables	19
Appendix 4. References	23

Executive Summary

Today, businesses are collecting more data than ever before and putting that data to work for a broader range of purposes—from enterprise resource planning and decision support to business intelligence and customer relationship management. This rapid and continuous growth means scalability must be a primary consideration for a company's database solution, including the database software, server engines and storage.

At the same time, an increasingly competitive and profit-oriented business climate is making companies more focused on ensuring that their information technology purchases represent a good value. As a result, a growing number of companies are considering newer and more affordable approaches—including clustered computing and network-attached storage—that offer outstanding price/performance and scalability.

This technical bulletin reports on collaborative efforts by engineers at IBM, Intel, and Network Appliance to evaluate the scalability of a solution that combines three technologies known for their scalability and cost-effectiveness:

- IBM DB2* Universal Database Enterprise-Extended Edition (DB2)
- Intel® AD450NX High Performance Servers
- Network Appliance NetApp* F840 Filers

DB2, with its ability to support up to 1,000 data partitions, is renowned for both scalability and ease of use. With respect to hardware, Intel® Architecture-based servers and NetApp Filers offer an incremental, “pay as you go” growth model that enhances price performance and scalability. All three product families emphasize ease of use and manageability, and amplify the price/performance benefits by providing savings in total cost of ownership.

The study reported on in this bulletin evaluated the results of scaling from one to four Intel servers and quadrupling the database size from 50GB to 200GB. Our results demonstrated excellent scalability. We also found that the database layout was relatively simple to set up and adjust, further contributing to ease of use and low cost of ownership.

The focus of this study was to validate and understand the scalability of the system, not to achieve peak performance figures. Only a moderate amount of performance tuning was implemented. This bulletin also documents findings from the evaluation and offers some best practice recommendations.

The collaborative efforts of these three industry leaders offers companies with growing databases the assurance that they can deploy DB2 on Intel-based servers with NetApp filers with confidence, knowing that this solution will support their scalability requirements.

DB2, Intel Servers and Network Appliance Filers

About DB2

DB2 is IBM's object-relational database for UNIX*, Linux*, OS/2, and Microsoft Windows* operating environments. DB2 offers easy installation, integrated functionality, a rich bundle of development tools, full Web enablement, Online Analytical Processing (OLAP) capabilities, and flexibility to scale and change platforms.

DB2 provides high performance support for large databases and offers outstanding scalability. DB2's shared-nothing architecture offers a superior foundation for overcoming scalability bottlenecks. These bottlenecks stem from the competition for limited computing resources such as CPUs and memory and, in database applications, from database concurrency-guarding mechanisms.

The performance tests used DB2 UDB EEE V 7.2 for Windows.

Intel AD450NX Server

The Intel AD450NX is a high performance server that supports symmetrical multiprocessing and a variety of operating systems. The system architecture is optimized for scalability, permitting users to add additional processors, memory, I/O boards and peripheral devices. The openness of the Intel Architecture gives businesses the freedom to select from an exceptionally broad array of hardware and software products and services to create innovative, best-of-breed solutions, as typified by the use of DB2 UDB EEE and Network Appliance Filers showcased in this bulletin.

Each AD450NX server used in this study had four Intel® Pentium® III Xeon™ processors at 550MHz with 2GB of RAM. The operating system was Microsoft Windows 2000 Advanced Server with SP1.

NetApp F840 Filer

NetApp F800 series enterprise filers provide a simple yet powerful data management solution for improving performance, ensuring continuous availability and minimizing infrastructure costs. The flexible, multiprotocol Network Appliance architecture enables simultaneous file sharing of UNIX, Windows, and Web data by users and application servers across the enterprise. The F840 filer offers built-in RAID protection, automatic load balancing, expanded memory and PCI slots to deliver maximum scalability and performance for large enterprises. In addition, NetApp filers feature dynamic online disk expansion, providing the ability to dynamically add one or more disks to a volume with no application server downtime or performance degradation. Superior throughput and response times enable the F840 to support thousands of users with up to 6TB of raw data.

Two NetApp F840 filers were used for this project, each configured with:

- 733MHz Intel Pentium III processor
- 3GB ECC memory
- 128MB Non-volatile memory
- 2 Gigabit Ethernet adapters (referred to as 'GbE' in this document)
- 1 Fiber Channel Arbitrated Loop (FC-AL) adapter
- Dual redundant fans
- Hot-pluggable redundant power supplies

The operating system on the filers was Network Appliance DataONTAP* version 6.1 with Microsoft CIFS* used as the file service protocol with the Intel servers. The Microsoft Common Internet File System (CIFS) protocol is natively integrated into DataONTAP.

System Design

The primary variable in the scalability study is the number of servers: the amount of data per server remained fixed at 50GB for one, two and four servers. DB2 was installed on each server, and either one or two filers were attached. Table 1 summarizes the configurations that were evaluated.

System	Servers	Filers	Database Size (GB)		
			Total	Per Server	Per Filer
A	1	1	50	50	50
B	2	2	100	50	50
C	2	1	100	50	100
D	4	2	200	50	100

Table 1: System Configurations Tested

Figure 1 provides an overall view of the system. Its topology is straightforward. The AD450NX servers are connected via an Emulex cLAN Virtual Interface (VI) Gigaset switch, and are connected to the filers via two Cisco Gigabit Ethernet (GbE) switches (Catalyst 3500 and Catalyst 6000). Communication among the servers is separated from that between the servers and filers to relieve network traffic over the Gigabit Ethernet.

This configuration was fully utilized in the four-node test and partially used in the other three tests. In the single-node test, only AD450NX-1 and filer 1 were in use. For the two-node test, two configurations were used—a one filer configuration and a two filer configuration.

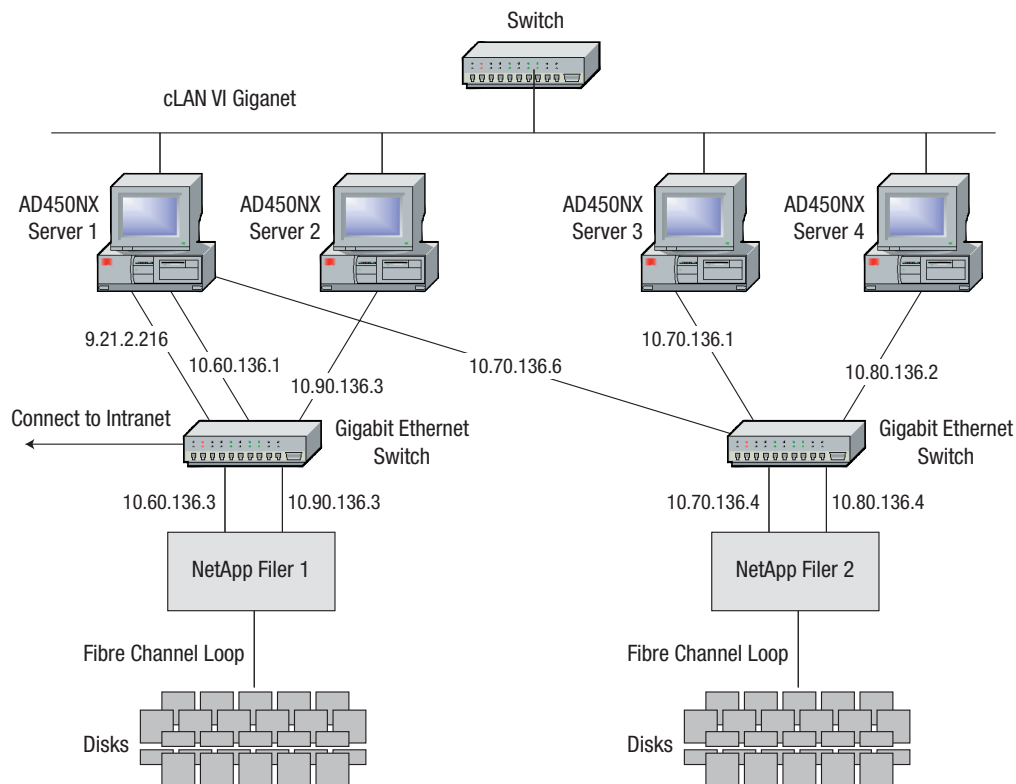


Figure 1. The Four-Server Configuration System (System D)

Experimental Design

A pre-determined set of decision support queries was used, and the same procedure was followed for each test—build the database and take the measurements.

The data was partitioned so that each node was populated with 50GB of data. There were two very small dimension tables that were not partitioned and were loaded only on the catalog node. All other tables were partitioned evenly among the nodes involved in each specific configuration. The other six tables contained approximately 300, 75, 40, 10, 7.5 and 0.5 million rows respectively on each node.

A set of complex queries modeled on a typical data warehouse were selected. They included all the operations that are fundamental to decision support applications—sequential table scans, index scans, multi-table joins, large aggregations, extensive sorting, etc. These were used to assess the scalability of the system as a whole. We also included a number of simple atomic queries to highlight the scalability of some critical database operations (e.g., pure sequential table scan, table scan with index, index creation, load, etc.).

Each set of queries was run several times to ensure consistency. In all cases, variability was minimal—less than 5%. The results reported correspond to the best runs for each system.

DB2 Configuration

The first DB2 configuration issue to be considered was the definition of tablespaces. DB2 supports two kinds of tablespaces: System Managed Storage (SMS) and Database Managed Storage (DMS). A DB2 tablespace is defined over containers. There are three types of containers: directory, file and device. An SMS tablespace is managed by the operating system and can only use directory container(s). A DMS tablespace is built on a pre-allocated portion of storage, either file or raw device, and managed by the database manager. (For detail, please reference the DB2 Administration Guide and relevant materials.) Because NetApp filers support the native file service protocols for both Windows 2000 and UNIX—Common Internet File System (CIFS) for Windows and the Network File System (NFS) for UNIX—either System Managed Space (SMS) or Database Managed Space (DMS) file tablespace could be used. For simplicity, SMS was chosen.

A 4K pagesize was used for the DB2 tablespaces in order to be consistent with the filer's 4K block size. It is strongly recommended that the DB2 database pagesize be set to a multiple of the filer block size (which is 4K). A single 4K bufferpool was used consisting of 200,000 4K pages (where 200,000 represents the value of the BUFFPAGE parameter as indicated in Table 4). Table 2 shows tablespace characteristics. Tablespace performance characteristics were set as follows: `overhead=6.5` and `transferrate=0.25`.

Tablespace	Type	Page Size (KB)	Extent Size (Page)	Prefetch Size (Page)
Main	SMS	4	24	288
Others	SMS	4	24	288
Temp	SMS	4	24	288
Small Tables	SMS	4	16	32

Table 2. Tablespace Characteristics

In DB2, one or more tables are stored in a tablespace. Tablespaces are typically symmetrical across nodes and it is common practice to partition the storage system to match the storage requirements of each node. For simplicity, each of these tablespaces was associated with storage from a specific filer (i.e., the tablespaces do not span the filers) and consists of a single container. Table 3 shows the tablespace arrangement adopted for each server. Note that the total size per node exceeds 50GB (it is actually 100GB) because additional space is required for the indexes and temporary tablespace.

Tablespace	Size (GB)
Main	55
Others	25
Temp	20
Small Tables	Very Small

Table 3. Server Tablespaces

A reasonable but not heroic amount of tuning was undertaken for the single-node system. Following standard tuning practice, the amount of memory allocated to the sortheap and bufferpool was maximized (total user memory was approximately 1.2GB). The tuning was efficient rather than exhaustive—ultimate performance was not our primary objective. Values of some of the most critical tuning parameters are listed in Table 4. Appendix 3 provides a full list of DB2 configuration parameters (database, database manager, and environment variables).

Bufferpool Size (Pages)	(BUFFPAGE) = 200,000
Sort List Heap (4KB)	(SORTHEAP) = 20,000
Sortheap Threshold (4KB)	(SHEAPTHRES) = 240,000
Degree of Parallelism	(DFT_DEGREE) = 12

Table 4. Key DB2 Tuning Parameters

Some additional configuration was required for the multi-node runs. DB2 enumerates the nodes in the db2nodes.cfg file. While there are also various communication parameters that can be altered, these parameters were left at their default values. The selection of SQL plans was left entirely to DB2’s industry-leading, cost-based optimizer. In all other respects, the multi-node configurations were essentially identical to the single-node configuration.

Network Appliance Filer Configuration

Physical Disk Layout

The storage subsystem employed by this project consisted of two F840 filers and eight NetApp FC-9 disk shelves. Each disk shelf was configured with seven 36GB disks.

The eight disk shelves were divided evenly between the two filers. Two disks were reserved for the root volume (and the operating system of each filer), leaving a total of 2 x 26 disks for data. Each of the two sets of 26 disks was then allocated to a single logical volume with two 12+p RAID groups (where “+p” represents a parity disk).

Although slightly better performance might have been obtained with a larger RAID group, our choice represents a reasonable compromise between performance and security. Tests conducted by NetApp determined that a 13+p configuration is optimal in configurations using 36GB disks (as in the course of this project.)

File system performance tends to increase as the number of disks in a RAID group/volume increases. However, the performance increases tend to level off above a 14 disk RAID group size. To maximize both performance and reliability, NetApp recommends using a RAID group size of 14 disks (13+p) in configurations that utilize 36GB disks. The choice of EXTENTSIZE follows from the definition of the RAID groups. Since 24 data disks were used during the testing, setting EXTENTSIZE equal to 24 ensured that data was evenly spread across the 24 data drives.

Volumes and Qtrees for Data Storage

NetApp Filers store data in file systems called volumes. Each volume is an independent file system with its own RAID groups. Each filer has a special volume known as a root volume. The root volume contains a file known as the */etc/rc* file, which contains startup commands. In addition to volumes, there are Qtrees—special subdirectories in the root directory of a volume—that act as virtual sub-volumes with special attributes, primarily quotas and permissions. One may store data either directly in volumes or in Qtrees within the volumes. The use of Qtrees is not required, but they may be defined to help organize the data. For the purposes of this project, the following Qtrees were defined on each volume:

On filer1:

```
qtrees for AD450NX-1 data:  db2_main_11
                             db2_others_11
                             db2_temp_11
```

```
qtrees for AD450NX-2 data:  db2_main_12
                             db2_others_12
                             db2_temp_12
```

```
db2_rawdata_11
```

On filer2:

```
qtrees for AD450NX-3 data:  db2_main_21
                             db2_others_21
                             db2_temp_21
```

```
qtrees for AD450NX-4 data:  db2_main_22
                             db2_others_22
                             db2_temp_22
```

```
db2_rawdata_21
```

CIFS Shares

CIFS Shares, commonly known as “shares,” are directories or directory structures that have been made available to network users and can be mapped to a drive letter on a CIFS client.

To permit access from the servers, each Qtree described on the previous page was mapped to a Windows 2000 networked drive using the autoEXNT service, which is triggered every time the system reboots. The batch file below is the autoexnt.bat file on AD450NX-1:

```
net use L: /d
net use L: \\10.60.136.3\db2_main_11
net use M: /d
net use M: \\10.60.136.3\db2_others_11
net use N: /d
net use N: \\10.60.136.3\db2_temp_11
net use V: /d
net use V: \\10.60.136.3\db2_rawdata_11
```

The autoexnt.bat files on AD450NX-2, AD450NX-3, and AD450NX-4 were virtually identical to the one shown on the left, except for the IP addresses used for the CIFS shares. Each AD450NX server used its own specific network path to the appropriate GbE interface on its respective filer to map the drive letters. For example, on AD450NX-2 the autoexnt.bat file contained:

```
net use L: /d
net use L: \\10.90.136.3\db2_main_12
net use M: /d
net use M: \\10.90.136.3\db2_others_12
net use N: /d
net use N: \\10.90.136.3\db2_temp_12
```

Note: When two or more network adapters are used, using the numeric IP addresses is recommended.

Connecting the Filer to DB2

Making the NetApp filer storage available to DB2 is straightforward. Simply define the tablespaces appropriately. Once the tablespaces are defined, the filers are transparent to DB2. This procedure is particularly simple and elegant using SMS tablespaces. The tablespaces can be defined by referring to the networked drives. No additional information is needed.

As described earlier, with respect to the decision support database, there is a set of tablespaces (and Qtrees) associated with each server. For simplicity, each tablespace can be assigned to a particular drive letter. This allows the same tablespace definitions to be used for all the nodes (see Appendix 2). The assignments take the form shown in Table 5.

Tablespace Name	Location	Qtrees on Filers			
		AD450NX-1	AD450NX-2	AD450NX-3	AD450NX-4
Main	L:	db2_main_11	db2_main_12	db2_main_21	db2_main_22
Others	M:	db2_others_11	db2_others_12	db2_others_21	db2_others_22
Temp	L:	db2_temp_11	db2_temp_12	db2_temp_21	db2_temp_22

Table 5. Tablespace Assignments

Networking Issues

Defining the networked drives is conceptually quite simple. Network Appliance filers running Data ONTAP* 6.1 include support for the Microsoft Active Directory and Kerberos authentication. Therefore, filers may be installed into Windows NT* domains, or Windows 2000 “mixed-mode” or “native-mode” Active Directory (AD) domains. (For further information, please refer to the technical report *Windows 2000 and NetApp Filers: An Overview* on the Network Appliance Tech Library Web site: http://www.netapp.com/tech_library/3113.html.)

As a result, CIFS shares can be referred to directly in defining networked drives. However, one outstanding issue remains—the paths by which the shares are to be accessed. Since there are two Gigabit Ethernet adapters per filer (and multiple GbEs inside each server), there are a number of ways to do this.

To minimize network bottlenecks and maximize performance, each filer GbE was assigned to a specific node. In the present case, this was done by associating a distinct network with each filer-server pair: These were 10.60.136, 10.70.136, 10.80.136, and 10.90.136, corresponding to AD450NX-1, AD450NX-3, AD450NX-4, and AD450NX-2 respectively. The networked drives were then defined by explicitly referring to the appropriate filer GbE. For AD450NX-1 (IP address 10.60.136.1 and matching GbE at 10.60.136.3), these definitions took the form:

```
net use L: \\10.60.136.3\db2_main_11
net use M: \\10.60.136.3\db2_others_11
net use N: \\10.60.136.3\db2_temp_11
net use V: \\10.60.136.3\db2_rawdata_11
```

In this way, each server was assured of a unique path to the filer. What’s more, this made access to the data particularly straightforward (e.g., on each node the same drive letter could be used to access the same tablespace). No further configuration was required to connect the filer to DB2; indeed, one can switch relatively quickly from one network topology to another—the system does not even need to be rebooted.

It is possible to have multiple paths connecting each server-filer pair (each filer supports up to a total of six GbEs on three PCI buses, and each AD450NX server has five 64-bit and six 32-bit PCI slots). This is a valuable option for users who demand maximum performance. In the present context, however, it was believed that it would be best to stay with the simpler (and more representative) one-path arrangement. As will be discussed in the *Results Overview* sections, there is excellent scalability between one and four nodes, and the number of paths should not be a critical issue.

An additional GbE card was installed in AD450NX-1 server, the domain controller, to permit external access outside the 10.x.136 networks. The system was completely private, with the exception of this one link to the outside world. To minimize the possibility of broadcast storms, the broadcast transmission window was lowered to 0.01%.

Configuration Parameters

The following filer configuration parameters were changed:

```
options cifs.tcp_window_size 64240
options cifs.max_mpx 253
options cifs.oplocks.enable off

vol options nosnap=yes
vol options minra on
```

Configuration Parameters (cont.)

For the first group of options to take effect, the corresponding Windows 2000 registry variables had to be modified accordingly:

```

\\HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\LanmanServer\
    Parameters\MaxMpxCt
    Datatype: DWORD
    Value: 253 (0xFD)

\\HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\LanmanServer\
    Parameters\MaxWorkItems
    Datatype: DWORD
    Value: 4096 (0x1000)

\\HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\
    Parameters\TcpWindowSize
    Datatype: DWORD
    Value: 64240 (0xFAF0)
    
```

Microsoft Hotfix Q271148_W2K_SP2_x86 was also applied to enable values of MaxMpxCt greater than 50. This Hotfix is included in Windows 2000 Service Pack 2.

Server Configuration

Server	Server Model	CPU and Frequency	L2 Cache	System Memory	BIOS Version
AD450NX-1	AD450NX	4-Way 550MHz Intel® Pentium® III Xeon™ Processor	1MB	2GB	ASPNO.86B.0009.P13
AD450NX-2	AD450NX	4-Way 550MHz Intel® Pentium® III Xeon™ Processor	1MB	2GB	ASPNO.86B.0009.P13
AD450NX-3	AD450NX	4-Way 550MHz Intel® Pentium® III Xeon™ Processor	1MB	2GB	ASPNO.86B.0009.P13
AD450NX-4	AD450NX	4-Way 550MHz Intel® Pentium® III Xeon™ Processor	1MB	2GB	ASPNO.86B.0009.P13

Results Overview

The measure of scalability is meaningful only when the underlying resources are held fixed as the number of servers is varied. As a result, it is helpful to organize the results according to the number of filers per server. The test began with one filer per server and moved on to one filer for every two servers. There were not enough filers available for a proper comparison of one and four servers; however, in the final subsection there is an examination of what happens when the number of filers per node is reduced.

As is standard practice, the scalability is expressed as a percentage of the baseline results, i.e.,

$$\text{Scalability} = 100 + 100 \cdot (t_{\text{ref}} - t_{\text{new}}) / t_{\text{ref}}$$

where t_{ref} is the query time for the baseline and t_{new} is the query time for the case under consideration. In other words, the scalability equals 100% if the query times are unchanged, greater than 100% if the query takes less time, and less than 100% if there is a degradation. The second term on the right-hand side may be thought of as the percentage improvement over the baseline.

Scaling from One Node to Two Nodes

The system displays excellent scalability between one and two nodes (1 filer per server in each case). This is most easily seen using the complex queries. Table 6 and Figure 2 show query times for the 16 complex queries.

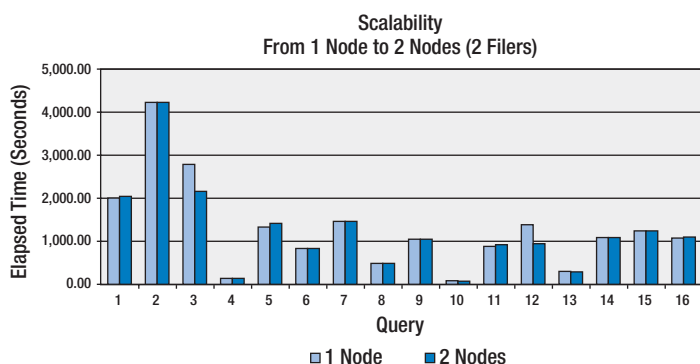


Figure 2. Complex Query Elapsed Times for System A (1 server + 1 filer) and B (2 servers + 2 filers).

Query	1 Node 1 Filer 24 Data Spindles 50GB Data	2 Nodes 2 Filers 48 Data Spindles 100GB Data	Scalability (%)
1	2,002.30	2,040.90	98.07
2	4,237.30	4,213.50	100.56
3	2,801.70	2,164.80	122.73
4	145.50	145.40	100.07
5	1,337.60	1,397.40	95.53
6	817.00	825.60	98.95
7	1,451.50	1,469.40	98.77
8	493.50	488.80	100.95
9	1,052.40	1,037.10	101.45
10	89.80	97.70	91.20
11	871.20	910.30	95.51
12	1,391.70	927.00	133.39
13	311.40	306.60	101.54
14	1,075.00	1,073.30	100.16
15	1,237.80	1,232.90	100.40
16	1,063.70	1,090.90	97.44
Total	20,379.40	19,421.60	104.70

Table 6. Complex Query Times (in seconds) for Systems A (1 server + 1 filer) and B (2 servers + 2 filers).

A key item of interest in Table 6 is that the total time required to complete the complex queries is no greater for two nodes with 100GB than for one node with 50GB. In fact, the total time is slightly less for two nodes—there is a scalability of approximately 104%. In fact, scalability is very good across the board.

This excellent scalability can be seen in performance metrics other than the query times. The performance of the servers (Table 7) and filers (Table 8) scales almost perfectly between one and two nodes.

Scaling from One Node to Two Nodes (cont.)

	1 Node	2 Nodes
Average CPU Utilization (%)	69	64
Average Bytes Transferred (MB/s)	33	35
Average User Memory (GB)	1.27	1.19

Table 7. Server Statistics for Systems A (1 node) and B (2 nodes)

	1 Node	2 Nodes	
	Filer 1	Filer 1	Filer 2
Average CPU Utilization (%)	40.2	42.3	42.4
Max CPU (%)	70	58	61
Average OPS	1,183	1,110	1,107
Max OPS	4,666	3,503	3,516
Average Disk Read (KB/s)	33,158	33,753	33,669
Maximum Disk Read (KB/s)	48,513	48,394	48,582

Table 8. Filer Statistics for Systems A (1 node) and B (2 nodes)

These numbers reinforce our conclusion that every element of the Intel/NetApp/DB2 system shows excellent scalability. By adding an extra server and filer, the performance has literally been doubled. The efficiency of the system is exemplary.

The max disk read numbers are particularly noteworthy. They are very close to the maximum throughput we have been able to obtain with a single filer and a single server. Using the Intel iometer utility, we obtained a max read throughput of 49MB/s. Higher throughput can be obtained with more servers.

As one might expect, the scalability of the simple atomic tests is equally good (Table 9).

Query	1 Node 1 Filer 24 Data Spindles	2 Nodes 2 Filers 48 Data Spindles	Scalability (%)
1. Pure table scan without index	1,076.25	1,013.35	105.8
2. Table scan with data sorted without index	1,122.85	1,116.45	100.6
3. Index only access	8.75	9.10	96.0
4. Index and table access	12.65	18.90	50.6
5. Aggregation, count distinct, one table, sorted. No Index	294.80	293.10	100.6
6. Partition key search criteria, indexed	0.25	0.35	60.0
7. Co-located join, two tables indexed	6.70	7.25	91.8
8. Broadcast join, multi-table join	2.55	2.60	98.0
9. Create non-unique index	538.84	554.82	97.0
10. Create unique index	588.65	572.15	102.8
11. Insert records	777.09	722.80	107.0
Total:	4,429.38	4,310.87	102.7

Table 9. Atomic Tests (in seconds) for Systems A and B

The overall scalability is more than 100%. There is some degradation in queries 4 and 6, but the query times here are so short that a subtle system overhead could be significant for them.

Scaling from Two Nodes to Four Nodes

The scalability between two and four nodes (with one filer for every two servers) is nearly as good as scaling from one node to two nodes. Table 10 demonstrates that the scalability is very good overall, with average scalability of 97% and some configurations producing better than linear scalability.

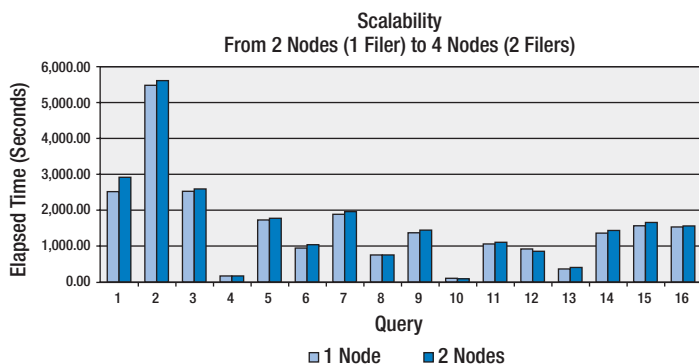


Figure 3. Complex Query Elapsed Times for System C (2 servers + 1 filer) and D (4 servers + 2 filers).

Query	2 Servers 1 Filer 100GB Data 24 Data Spindles	4 Servers 2 Filers 200GB Data 48 Data Spindles	Scalability (%)
1	2,569.30	2,913.40	86.61
2	5,448.30	5,553.30	98.07
3	2,546.20	2,561.40	99.40
4	190.40	198.90	95.54
5	1,806.20	1,830.00	98.68
6	944.60	1,013.80	92.67
7	1,816.80	1,926.40	93.97
8	657.20	658.50	99.80
9	1,439.00	1,481.50	97.05
10	138.80	136.00	102.02
11	1,031.30	1,067.60	96.48
12	935.80	859.70	108.13
13	409.10	431.70	94.48
14	1,435.80	1,476.30	97.18
15	1,626.30	1,660.30	97.91
16	1,489.20	1,498.20	99.40
Total	24,484.30	25,267.00	96.80

Table 10. Complex Query Elapsed Times (in seconds) for Systems C and D

Not surprisingly, the performance of the servers (Table 11) and the filers (Table 12) effectively doubles from two nodes to four nodes. Again, the server and filer statistics are essentially independent of the number of nodes—one of the necessary conditions for scalability is satisfied.

	2 Nodes 24 Data Spindles	4 Nodes 48 Data Spindles
Average CPU Utilization (%)	48.20	49.40
Average Bytes Transferred (MB/s)	28.30	27.70
Average User Memory (GB)	1.18	1.18

Table 11. Server Statistics for Systems C (2 servers + 1 filer) and D (4 servers + 2 filers).

Data for both system C and D came from AD450NX-2, the coordinator node.

	2 Nodes Filer 1	Filer 1	Filer 2
Average CPU Utilization (%)	72	69.4	71.2
Max CPU (%)	98	98	98
Average OPS	1,777	1,712	1,705
Max OPS	5,707	5,382	5,777
Average Disk Read (KB/s)	53,713	52,081	52,071
Maximum Disk Read (KB/s)	73,333	73,665	69,698

Table 12. Filer Statistics for Systems C (2 servers + 1 filer) and D (4 servers + 2 filers).

Finally, as Table 13 indicates, the atomic queries show excellent scalability as well.

Scaling from Two Nodes to Four Nodes (cont.)

Query	2 Nodes 1 Filer 24 Data Spindles	4 Nodes 2 Filers 48 Data Spindles	Scalability (%)
1. Pure table scan without index	1,392.75	1,400.95	99.4
2. Table scan with data sorted without index	1,443.65	1,417.55	101.8
3. Index only access	9.10	9.20	98.9
4. Index and table access	20.00	19.45	102.8
5. Aggregation, count distinct, one table, sorted. No Index	342.55	334.95	102.2
6. Partition key search criteria, indexed	0.35	0.25	128.6
7. Co-located join, two tables indexed	7.80	7.40	105.1
8. Broadcast join, multi-table join	3.35	3.35	100.0
9. Create non-unique index	604.17	607.74	99.4
10. Create unique index	657.00	647.60	101.4
11. Insert records	714.52	777.57	91.2
Total:	5,195.24	5,226.01	99.4

Table 13. Atomic Tests for Systems C (2 servers + 1 filer) and D (4 servers + 2 filers).

Scaling from Two Servers/Two Filers to Four Servers/Two Filers

In the preceding subsections, the number of filers doubled—going from one to two servers, and from two to four servers. In other words, the resources were not held fixed.

In practice, however, it is difficult to ensure that the resources available to each server remain constant. Typically one might wish to add more servers before adding more I/O and storage capacity. It is therefore important to examine the scalability of the Intel/NetApp/DB2 system under less-than-ideal conditions. This section addresses a natural extension of the preceding analysis—two servers with two filers, and four servers with two filers.

Doubling the number of servers but keeping the number of filers fixed is a simple and straightforward way of dealing with twice as much data.

We begin with the complex queries (Table 14). This time we do not have perfect scalability. Nevertheless, given the imposed limitations—four times as much data but only twice as much disk I/O—the scalability is quite impressive. By merely adding two more servers, 70% of the performance theoretically obtainable from a four-node system is achieved.

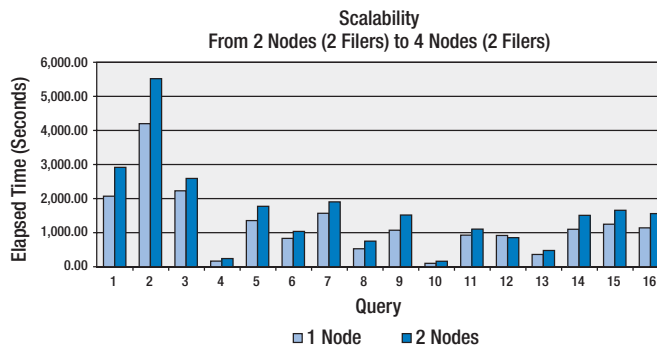


Figure 4. Complex Query Elapsed Times for Systems B (2 servers and 2 filers) and D (4 servers and 2 filers).

Query	2 Servers 2 Filers 100GB Data 48 Data Spindles	4 Servers 4 Filers 200GB Data 48 Data Spindles	Scalability (%)
1	2,040.90	2,913.40	57.25
2	4,213.50	5,553.30	68.20
3	2,164.80	2,561.40	81.68
4	145.40	198.90	63.20
5	1,397.40	1,830.00	69.04
6	825.60	1,013.80	77.20
7	1,469.40	1,926.40	68.90
8	488.80	658.50	65.28
9	1,037.10	1,481.50	57.15
10	97.70	136.00	60.80
11	910.30	1,067.60	82.72
12	927.00	859.70	107.26
13	306.60	431.70	59.20
14	1,073.30	1,476.30	62.45
15	1,232.90	1,660.30	65.33
16	1,090.90	1,498.20	62.66
Total	19,421.60	25,267.00	69.90

Table 14. Complex Queries (in seconds) for Systems B (2 servers + 2 filers) and D (4 servers + 2 filers)

Scaling from Two Servers/Two Filers to Four Servers/Two Filers (cont.)

It is instructive to examine the performance of the servers (Table 15) and filers (Table 16).

	2 Nodes 48 Data Spindles	4 Nodes 48 Data Spindles
Average CPU Utilization (%)	64.10	49.40
Average Bytes Transferred (MB/s)	35.30	27.70
Average User Memory (GB)	1.19	1.18

Table 15. Server Statistics for Systems B (2 servers + 2 filers) and D (4 servers + 2 filers).

Data for system B came from AD450NX-1 and that for system C came from AD450NX-4.

	2 Nodes 48 Data Spindles	4 Nodes 48 Data Spindles
Average CPU Utilization (%)	42.4	70.3
Max CPU (%)	61	98
Average OPS	1,109	1,709
Max OPS	3,516	5,777
Average Disk Read (KB/s)	33,711	52,076
Maximum Disk Read (KB/s)	48,394	73,665

Table 16. Filer Statistics for Systems B (2 servers + 2 filers) and D (4 servers + 2 filers).

While the I/O throughput of the filer jumps up markedly when two servers are attached to each filer, it does not quite double (as it would if there were perfect scalability). Roughly speaking, each filer is producing about 77% of the throughput that would be needed for perfect scalability. Likewise, the average number of operations per second also increases to about 77% of the value required for perfect scalability. Since there are effectively half as many spindles assigned to each server, this performance is quite good.

The atomic queries (Table 17) show similar behavior, although the scalability is slightly better than for the complex queries.

Query	2 Nodes 2 Filers 48 Data Spindles	4 Nodes 2 Filers 48 Data Spindles	Scalability (%)
1. Pure table scan without index	1,013.35	1,400.95	61.8
2. Table scan with data sorted without index	1,116.45	1,417.55	73.0
3. Index only access	9.10	9.20	98.9
4. Index and table access	18.90	19.45	97.1
5. Aggregation, count distinct, one table, sorted. No Index	293.10	334.95	85.7
6. Partition key search criteria, indexed	0.35	0.25	128.6
7. Co-located join, two tables indexed	7.25	7.40	97.9
8. Broadcast join, multi-table join	2.60	3.35	71.2
9. Create non-unique index	554.82	607.74	90.5
10. Create unique index	572.15	647.60	86.8
11. Insert records	722.80	777.57	92.4
Total:	4,310.87	5,226.01	78.8

Table 17. Atomic Tests for Systems B (2 servers + 2 filers) and D (4 servers + 2 filers)

This result is consistent with earlier observations regarding the IO throughput of the filers. While there is naturally a price to be paid for doubling the data but not the resources (i.e., the number of data disk spindles), the scalability remains quite good.

Summary

To address the business need for scalable database solutions that can handle the accelerating growth of data, IBM, Intel, and Network Appliance examined the scalability of a robust, high-value database solution implemented on three industry-leading technologies—IBM DB2, Intel-based servers and NetApp filers. The collaborative work by these three companies demonstrates that even with only moderate tuning, the solution delivers linear scalability when the storage resources are kept constant for each server (see Figures 5, 6, and 7). It also scales very well even when storage resources are not added to—although clearly, the preferred path is to scale the storage devices as the amount of data and the number of servers scale. Along with their excellent scalability, the technologies used in this study have outstanding track records for manageability, reliability and ease of use, making them an outstanding choice for the most demanding—and rapidly growing—database deployments.

Figure 5. Scalability from 1 Node to 2 Nodes (1 node per filer)

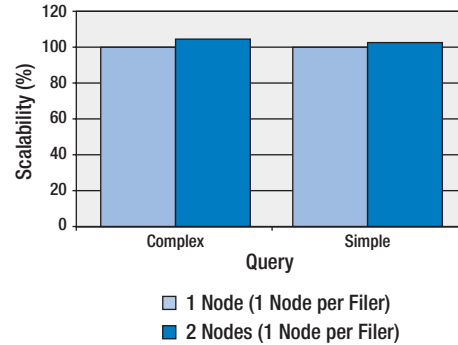


Figure 6. Scalability from 2 Nodes to 4 Nodes (2 nodes per filer)

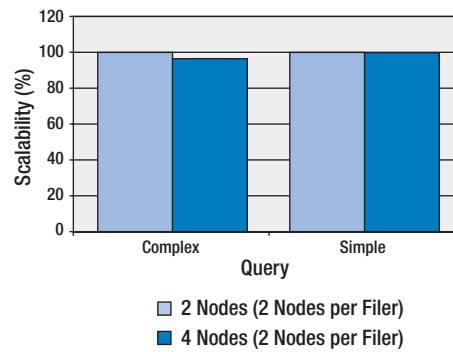
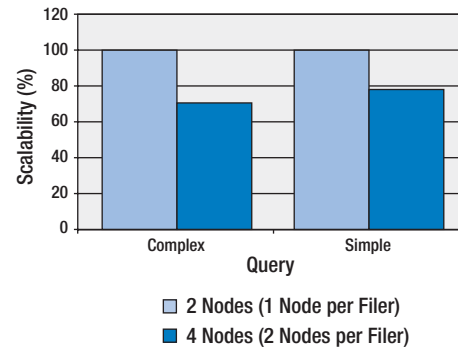


Figure 7: Scalability from 2 Nodes (1 node per filer) to 4 Nodes (2 nodes per filer)



Appendix 1. Performance Tuning: Experiences, Lessons and Recommendations

Although this technical bulletin focuses on scalability, performance issues were not ignored. Clearly the findings would be of limited value if they didn't represent results obtained by typical users. To obtain this level of performance, a certain amount of time and experimentation was required to overcome unexpected quirks and configure the system quasi-optimally. This section describes some of our more useful performance tuning experiences and offers recommendations. Table 19 summarizes these recommendations.

PREFETCHSIZE

Increasing the PREFETCHSIZE increases the amount of data that the DB2 prefetches attempt to retrieve at a single time. It is generally recommended that the PREFETCHSIZE be a small multiple of the EXTENTSIZE.

We found that sequential prefetching is optimized for **PREFETCHSIZE=288**:

PREFETCHSIZE	Time(s)	Net in (MBs)	Read (MB/s)	CPU (%)
96	361	46	43	50
192	241	66	61	71
288	221	72	70	82
394	216	75	72	82

Table 18. Query Times and Server Statistics vs. PREFETCHSIZE for a Simple Sequential Table Scan That Returns No Rows

Setting the appropriate PREFETCHSIZE coordinates the DB2 I/O patterns to match the RAID group size on the filer. Using the formula $PREFETCHSIZE = EXTENTSIZE \times \# \text{ disks in the RAID group}$, given the values outlined elsewhere in this document where $EXTENTSIZE = 24$ and the number of data disks per RAID group = 12, calculates to a PREFETCHSIZE of 288, which was the value indicated above and that worked well during the testing.

Bufferpools and Sortheap

Following standard tuning practice, we divided the overwhelming majority of the total available memory nearly evenly between the bufferpool and the sortheap. We found that doubling **buffpage** and **sheapthres** to their final values improved overall performance by about 20%. Queries requiring sorts show an improvement of roughly 50%.

We used **SHEAPTHRES=SORTHEAP*DFT_DEGREE**, where **DFT_DEGREE** is the degree of parallelism. We found that **DFT_DEGREE=12** is optimal (for a single node). While tuning the bufferpool size, it was necessary to increase **dbheap** in conjunction with **buffpage** (it should be roughly 3.5% of the total bufferpool size).

Tablespaces

When using 4K tablespaces, be sure to explicitly drop the default 4K **TEMPSPACE1** tablespace; otherwise, DB2 will alternate between **temp_tbsp**, with its large BP4K bufferpool, and **TEMPSPACE1** and its comparatively tiny bufferpool. Use “one temporary tablespace for a given page size” (*DB2 UDB V7.1 Performance Tuning Guide*, p.57.)

Filer Configuration Parameters

The only parameter we examined in detail is **minra**, which controls the amount of read ahead that the filer will perform. In general, for random workloads, **minra** should be turned on, and for more sequential workloads **minra** should be turned off. In this case, **minra** has a minimal effect on filer performance. We recommend turning on **minra** when perfecting is required.

Network Isolation

We used a private network for our final results. In obtaining the preliminary results, we occasionally used a public network. The difference between the public and private networks was insignificant.

Filer Gigabit Ethernet

To minimize network bottlenecks and maximize performance, each filer's gigabit Ethernet should be assigned to a specific node. If the DB2 workload across the nodes is relatively balanced, assigning a dedicated filer GbE to each node should help maintain that workload balance by eliminating the possibility of overburdening any single GbE with the workload of multiple nodes.

iometer

Intel's iometer utility is a performance analysis tool that lets you quantify a server's I/O throughput. It works equally well with the local and networked drives. Following the recommendations in the documentation, we used 64KB sequential reads to determine the maximum throughput.

iometer is available from:

<http://developer.intel.com/design/servers/devtools/iometer/>

Monitoring

Monitoring of the filers and servers is easily done. The servers can be monitored using the Windows 2000 Performance Monitor, or more efficiently, with `typeperf`, which is a command-line version included in the Windows 2000 Resources Kit. We used the following invocation:

```
typeperf $statsint $cpu $gig $mem > typeperf.op
```

where

```
$statsint=10
$cpu="\\Processor(_Total)\\% Processor Time"
$gig="\\Network Interface(Gigabit Ethernet
  Adapter from IBM)\\Bytes Total/sec"
$mem="\\Memory\\Committed Bytes"
```

The filers can be monitored using the `sysstat` command. To save the output to disk, you can invoke `sysstat` via `rsh`, e.g.:

```
rsh filer1 -l guojian sysstat >>filermon.op
```

From this data, summary statistics can be easily computed.

Table 19. Summary of Performance Tuning Recommendations

Characteristic	Recommendation
PREFETCHSIZE	<ul style="list-style-type: none"> ■ PREFETCHSIZE = EXTENTSIZE x # data disks in the RAID group
Bufferpools and sorheap	<ul style="list-style-type: none"> ■ Set DFT_DEGREE=12 for a single node ■ Divide the majority of the total available memory nearly evenly between the bufferpool and the sorheap, and increase dbheap in conjunction with buffpage to keep it at roughly 3.5% of the total bufferpool size.
Tablespaces	<ul style="list-style-type: none"> ■ Explicitly drop the default 4K TEMPSPACE1 tablespace when using 4K tablespaces. ■ Use one temporary tablespace for a given page size.
Fiber gigabit Ethernet	<ul style="list-style-type: none"> ■ Assign each filer's gigabit Ethernet to a specific node.
iometer	<ul style="list-style-type: none"> ■ Use the Intel iometer tool to evaluate server throughput, with 64KB sequential reads to determine maximum throughput.
Monitoring	<ul style="list-style-type: none"> ■ Use Windows 2000 Performance Monitor or typeperf to monitor servers. ■ Use the sysstat command to monitor filers.

Appendix 2. DB2 Scripts

i) Create Bufferpools

```
alter bufferpool IBMDEFAULTBP size 1000;
create bufferpool BP4K size -1 pagesize 4K
commit;
terminate;
db2stop;
db2start;
```

ii) Create Nodegroups

```
create nodegroup catalog_node on node (0);
create nodegroup all_nodes;
commit;
```

iii) Create Tablespaces

```
create regular tablespace small_tables
  in nodegroup catalog_node
  pagesize 4K
  managed by system
  using ('C:\small_tables')
  on node (0)
  bufferpool BP4K
;
```

```
create regular tablespace main_tbsp
  in nodegroup All_NODES
  pagesize 4K
  managed by system
  using ('L:\main_tbsp')
  prefetchsize 48 but you document the
  prefetch size should be 288
  extentsize 24
  bufferpool BP4K
  overhead      6.5
  transferrate  0.25
;
```

```
create regular tablespace others_tbsp
  in nodegroup All_NODES
  pagesize 4K
  managed by system
  using ('M:\others_tbsp')
  prefetchsize 48
  extentsize 24
  bufferpool BP4K
  overhead      6.5
  transferrate  0.25
;
```

```
create temporary tablespace temp_tbsp
  in nodegroup IBMTEMPGROUP
  pagesize 4K
  managed by system
  using ('N:\temp_tables')
  prefetchsize 48
  extentsize 24
  bufferpool BP4K
  overhead      6.5
  transferrate  0.25;
commit work;

drop tablespace temp_space1;

commit work;
```

Appendix 3. DB2 Configuration Parameters and Environment Variables

i) Database Configuration

Database configuration release level		= 0x0900
Database release level		= 0x0900
Database territory		= US
Database code page		= 1252
Database code set		= IBM-1252
Database country code		= 1
Dynamic SQL Query management	(DYN_QUERY_MGMT)	= DISABLE
Directory object name	(DIR_OBJ_NAME)	=
Discovery support for this database	(DISCOVER_DB)	= ENABLE
Default query optimization class	(DFT_QUERYOPT)	= 5
Degree of parallelism	(DFT_DEGREE)	= 12
Continue upon arithmetic exceptions	(DFT_SQLMATHWARN)	= NO
Default refresh age	(DFT_REFRESH_AGE)	= 0
Number of frequent values retained	(NUM_FREQVALUES)	= 10
Number of quantiles retained	(NUM_QUANTILES)	= 300
Backup pending		= NO
Database is consistent		= YES
Rollforward pending		= NO
Restore pending		= NO
Multi-page file allocation enabled		= NO
Log retain for recovery status		= NO
User exit for logging status		= NO
Data Links Token Expiry Interval (sec)	(DL_EXPINT)	= 60
Data Links Number of Copies	(DL_NUM_COPIES)	= 1
Data Links Time after Drop (days)	(DL_TIME_DROP)	= 1
Data Links Token in Uppercase	(DL_UPPER)	= NO
Data Links Token Algorithm	(DL_TOKEN)	= MACO
Database heap (4KB)	(DBHEAP)	= 10000
Catalog cache size (4KB)	(CATALOGCACHE_SZ)	= 386
Log buffer size (4KB)	(LOGBUFSZ)	= 512
Utilities heap size (4KB)	(UTIL_HEAP_SZ)	= 4000
Extended storage segments size (4KB)	(ESTORE_SEG_SZ)	= 16000
Number of extended storage segments	(NUM_ESTORE_SEGS)	= 0
Max storage for lock list (4KB)	(LOCKLIST)	= 8000
Max appl. control heap size (4KB)	(APP_CTL_HEAP_SZ)	= 8000
Sort list heap (4KB)	(SORTHEAP)	= 20000
SQL statement heap (4KB)	(STMTHEAP)	= 8000
Default application heap (4KB)	(APPLHEAPSZ)	= 4000
Package cache size (4KB)	(PCKCACHESZ)	= 320

Appendix 3. DB2 Configuration Parameters and Environment Variables (cont.)

```

Statistics heap size (4KB)                (STAT_HEAP_SZ) = 4384

Interval for checking deadlock (ms)       (DLCHKTIME) = 5000
Percent. of lock lists per application    (MAXLOCKS) = 100
Lock timeout (sec)                        (LOCKTIMEOUT) = 200

Changed pages threshold                   (CHNGPGS_THRESH) = 80
Number of asynchronous page cleaners      (NUM_IOCLEANERS) = 4
Number of I/O servers                     (NUM_IOSERVERS) = 49
Index sort flag                           (INDEXSORT) = YES
Sequential detect flag                    (SEQDETECT) = YES
Default prefetch size (pages)             (DFT_PREFETCH_SZ) = 16

Track modified pages                      (TRACKMOD) = OFF

Default number of containers              = 1
Default tablespace extentsize (pages)    (DFT_EXTENT_SZ) = 32

Max number of active applications         (MAXAPPLS) = 12
Average number of active applications     (AVG_APPLS) = 1
Max DB files open per application         (MAXFILOP) = 64

Log file size (4KB)                      (LOGFILSIZ) = 8192
Number of primary log files               (LOGPRIMARY) = 50
Number of secondary log files             (LOGSECOND) = 20
Changed path to log files                 (NEWLOGPATH) =
Path to log files                         = D:\DB2MPP\NODE0000\SQL00001\SQLOGDIR\
First active log file                     =

Group commit count                       (MINCOMMIT) = 3
Percent log file reclaimed before soft chckpt (SOFTMAX) = 500
Log retain for recovery enabled           (LOGRETAIN) = OFF
User exit for logging enabled             (USEREXIT) = OFF

Auto restart enabled                     (AUTORESTART) = ON
Index re-creation time                    (INDEXREC) = SYSTEM (ACCESS)
Default number of loadrec sessions        (DFT_LOADREC_SES) = 1
Number of database backups to retain      (NUM_DB_BACKUPS) = 12
Recovery history retention (days)        (REC_HIS_RETENTN) = 366

TSM management class                     (TSM_MGMTCLASS) =
TSM node name                            (TSM_NODENAME) =
TSM owner                                 (TSM_OWNER) =
TSM password                             (TSM_PASSWORD) =
    
```

Appendix 3. DB2 Configuration Parameters and Environment Variables (cont.)

ii) Database Manager Configuration

Node type = Partitioned Database Server with local and remote clients

```

Database manager configuration release level          = 0x0900

Maximum total of files open                        (MAXTOTFILOP) = 16000
CPU speed (millisec/instruction)                  (CPUSPEED)   = 1.039157e-006
Communications bandwidth (MB/sec)                  (COMM_BANDWIDTH) = 1.000000e+000

Max number of concurrently active databases        (NUMDB)      = 1
Data Links support                                (DATA LINKS) = NO
Federated Database System Support                  (FEDERATED)  = NO
Transaction processor monitor name                 (TP_MON_NAME) =

Default charge-back account                        (DFT_ACCOUNT_STR) =

Java Development Kit 1.1 installation path         (JDK11_PATH) =

Diagnostic error capture level                     (DIAGLEVEL)  = 3
Notify Level                                       (NOTIFYLEVEL) = 2
Diagnostic data directory path                     (DIAGPATH)   =

Default database monitor switches
  Bufferpool                                       (DFT_MON_BUFPOOL) = OFF
  Lock                                             (DFT_MON_LOCK)   = OFF
  Sort                                             (DFT_MON_SORT)   = OFF
  Statement                                       (DFT_MON_STMT)   = OFF
  Table                                           (DFT_MON_TABLE)  = OFF
  Unit of work                                    (DFT_MON_UOW)    = OFF

SYSADM group name                                (SYSADM_GROUP) =
SYSCTRL group name                               (SYSCTRL_GROUP) =
SYSMAINT group name                              (SYSMAINT_GROUP) =

Database manager authentication                   (AUTHENTICATION) = SERVER
Cataloging allowed without authority              (CATALOG_NOAUTH) = NO
Trust all clients                                (TRUST_ALLCLNTS) = YES
Trusted client authentication                     (TRUST_CLNTAUTH) = CLIENT

Default database path                             (DFTDBPATH)    = D:

Database monitor heap size (4KB)                  (MON_HEAP_SZ)  = 12
UDF shared memory set size (4KB)                  (UDF_MEM_SZ)   = 256
Java Virtual Machine heap size (4KB)              (JAVA_HEAP_SZ) = 512
Audit buffer size (4KB)                           (AUDIT_BUF_SZ) = 0

Backup buffer default size (4KB)                   (BACKBUFSZ)   = 1024
Restore buffer default size (4KB)                   (RESTBUFSZ)   = 1024

Agent stack size                                  (AGENT_STACK_SZ) = 16
Minimum committed private memory (4KB)            (MIN_PRIV_MEM) = 32
Private memory threshold (4KB)                    (PRIV_MEM_THRESH) = 1296

Sortheap threshold (4KB)                          (SHEAPTHRES)  = 240000
    
```

Appendix 3. DB2 Configuration Parameters and Environment Variables (cont.)

```

Directory cache support                (DIR_CACHE) = YES

Application support layer heap size (4KB) (ASLHEAPSZ) = 15
Max requester I/O block size (bytes)    (RQRIOBLK) = 32767
DOS requester I/O block size (bytes)    (DOS_RQRIOBLK) = 4096
Query heap size (4KB)                   (QUERY_HEAP_SZ) = 1000
DRDA services heap size (4KB)           (DRDA_HEAP_SZ) = 128

Priority of agents                      (AGENTPRI) = SYSTEM
Max number of existing agents           (MAXAGENTS) = 400
Agent pool size                         (NUM_POOLAGENTS) = 40
Initial number of agents in pool        (NUM_INITAGENTS) = 3
Max number of coordinating agents        (MAX_COORDAGENTS) = (MAXAGENTS - NUM_INITAGENTS)
Max no. of concurrent coordinating agents (MAXCAGENTS) = MAX_COORDAGENTS
Max number of logical agents            (MAX_LOGICAGENTS) = MAX_COORDAGENTS

Keep DARI process                      (KEEPDARI) = YES
Max number of DARI processes             (MAXDARI) = MAX_COORDAGENTS
Initialize DARI process with JVM        (INITDARI_JVM) = NO
Initial number of fenced DARI process   (NUM_INITDARIS) = 0

Index re-creation time                  (INDEXREC) = ACCESS

Transaction manager database name        (TM_DATABASE) = 1ST_CONN
Transaction resync interval (sec)       (RESYNC_INTERVAL) = 180

SPM name                                (SPM_NAME) =
SPM log size                            (SPM_LOG_FILE_SZ) = 256
SPM resync agent limit                   (SPM_MAX_RESYNC) = 20
SPM log path                             (SPM_LOG_PATH) =

NetBIOS Workstation name                 (NNAME) =

TCP/IP Service name                      (SVCENAME) =
APPC Transaction program name            (TPNAME) =
IPX/SPX File server name                 (FILESERVER) =
IPX/SPX DB2 server object name           (OBJECTNAME) =
IPX/SPX Socket number                    (IPX_SOCKET) = 879E

Discovery mode                           (DISCOVER) = SEARCH
Discovery communication protocols        (DISCOVER_COMM) =
Discover server instance                  (DISCOVER_INST) = ENABLE

Directory services type                   (DIR_TYPE) = NONE
Directory path name                       (DIR_PATH_NAME) = ../../subsys/database/
Directory object name                     (DIR_OBJ_NAME) =
Routing information object name           (ROUTE_OBJ_NAME) =
Default client comm. protocols           (DFT_CLIENT_COMM) =
Default client adapter number            (DFT_CLIENT_ADPT) = 0

Maximum query degree of parallelism      (MAX_QUERYDEGREE) = ANY
Enable intra-partition parallelism       (INTRA_PARALLEL) = YES

```

Appendix 3. DB2 Configuration Parameters and Environment Variables (cont.)

No. of int. communication buffers(4KB) (FCM_NUM_BUFFERS) = 8000
Number of FCM request blocks (FCM_NUM_RQB) = 8000
Number of FCM connection entries (FCM_NUM_CONNECT) = (FCM_NUM_RQB * 0.75)
Number of FCM message anchors (FCM_NUM_ANCHORS) = (FCM_NUM_RQB * 0.75)

Node connection elapse time (sec) (CONN_ELAPSE) = 10
Max number of node connection retries (MAX_CONNRETRIES) = 1
Max time difference between nodes (min) (MAX_TIME_DIFF) = 60

db2start/db2stop timeout (min) (START_STOP_TIME) = 10

iii) environment variables:

DB2_ANTIJOIN=Y
DB2_NEW_CORR_SQ_FF=Y
DB2_STRIPED_CONTAINERS=ON
DB2_VI_DEVICE=nic0
DB2_VI_VIPL=VIPL.DLL
DB2_VI_ENABLE=ON
DB2_LIKE_VARCHAR=2.5
DB2_CORRELATED_PREDICATES=Y
DB2_HASH_JOIN=Y
DB2ACCOUNTNAME=AD450NX-s\guojian
DB2INSTOWNER=AD450NX-2
DB2PORTRANGE=50002:50005
DB2OPTIONS= -t -v +c
DB2NTNOCACHE=ON
DB2INSTPROF=\\AD450NX-2\DB2-DB2MPP
DB2_PARALLEL_IO=*

Appendix 4: References

Further information about the products used in this study is available at:

- **DB2:** <http://www-4.ibm.com/software/data/>
- **Intel Servers:** <http://www.intel.com/pentiumiii/xeon/home.htm> and <http://www.intel.com/titanium/index.htm>
- **NetApp Filers:** <http://www.netapp.com/products/filer>



Intel Corporation
2200 Mission College Blvd.
P.O. Box 58119
Santa Clara, CA 95052-8119
Tel. (408) 765-8080
www.intel.com

(c) International Business Machines Corporation, Intel Corporation, and Network Appliance Inc. 2001. All Rights Reserved. US Government User Restricted Rights—Use, Duplication or disclosure restricted by GSA ADB Schedule Contract with IBM Corporation. The information included in this Technical Bulletin is provided AS IS without warranty of any kind. IBM, Intel, and Network Appliance expressly disclaim any warranties, express or implied, including the implied warranties of merchantability or fitness for a particular purpose. IBM, Intel and Network Appliance do not warrant any results or performance measurements obtained. The information included in this Technical Bulletin concerning IBM products was written for IBM products and services offered in the U.S.A. IBM may not offer the products, services or features discussed in this document in other countries, and the information is subject to change without notice. The information in this document relating to Intel products is provided as a convenience to Intel's general customer base. IBM, Intel and Network Appliance each individually and collectively make no warranty or representation, express or implied with respect to the accuracy and completeness of the information, and assume no responsibility for any errors, which may appear in the document. Any performance data contained herein was determined in a controlled test environment. Therefore, the results obtained in other operating environments may vary significantly. Actual performance and environment cost will vary depending on individual customer configurations and conditions. No license, express or implied, to any intellectual property is granted by this document. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

DB2®, DB2 Universal Database, IBM and the IBM logo are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both. Microsoft, Windows, and Windows 2000 are trademarks or registered trademarks of Microsoft Corporation in the United States, other countries or both. Intel and Xeon are trademarks or registered trademarks of Intel Corporation in the United States, other countries or both. Network Appliance, NetApp and DataOntap are trademarks or registered trademarks of Network Appliance Inc. in the United States, other countries or both.

*Third-party brands and names are the property of their respective owners

Note: This technical bulletin IS NOT a certified benchmark by IBM, Intel, or Network Appliance.

This is a preliminary data sheet and solution review featuring IBM DB2 Universal Database v7.2, Intel Pentium III Xeon processors, and Network Appliance NetApp F840 Filers.