



# Best Practices

## **SAP Applications with the DB2 pureScale Feature on SUSE Linux Enterprise Server and IBM System x: A best-of-breed Solution Stack**

Introduction .....	4
Overview of the DB2 pureScale Feature.....	5
Hardware requirements.....	8
Supported IBM System x servers.....	8
Supported IBM storage systems .....	10
Supported high-speed RDMA networks.....	11
Hardware configurations.....	11
SUSE Linux Enterprise Server as an optimized operating system for SAP Solutions .....	19
SUSE Linux Enterprise Server for SAP Applications .....	19
Installation wizard for SAP solutions .....	20
Optimized paging support for large in-memory workloads.....	20
SAP NetWeaver compatible Java Virtual Machine (JVM).....	21
SAP specific update channel .....	21
Extended service pack overlap support .....	21
Priority support option .....	22
Premium Support Entitlement.....	22
Linux support for the DB2 pureScale Feature .....	22
Virtualization.....	23
DB2 pureScale Feature Use Cases and Benefits for SAP Applications .....	25
A complete HA solution for SAP Applications.....	25
Non-disruptive maintenance .....	26
Workload isolation within an SAP solution.....	27
Use of the most cost-efficient computing hardware .....	27
Efficient management of capacity growth: Grow DB2 database servers as your business grows .....	28
Temporary capacity increases to cover workload peaks .....	28
Performance considerations .....	30
Workload characteristics and the Caching Facility (CF).....	30
Avoidance of page reclaims .....	31

GPFS and database I/O.....	35
Integration and operation of the DB2 pureScale Feature with SAP applications .....	37
Setting up SAP Central Services components for HA .....	40
Monitoring the DB2 pureScale Feature Using the DBA Cockpit .....	42
Application-transparent maintenance of a DB2 member or CF.....	43
Conclusion .....	45
Glossary .....	46
Further reading.....	48
Authors.....	50
Contributors.....	51

## Introduction

Today's rapidly changing business generates increasing demands for flexibility and continuous availability of business applications. This paper describes why a solution stack of IBM System x servers, SUSE Linux Enterprise Server optimized for SAP applications, the IBM DB2 pureScale Feature, and SAP applications provides an excellent response to these demands. This paper is targeted at both a technical audience and decision makers who want to understand the benefits that are provided through this best-of-breed solution stack.

SAP solutions on the DB2 pureScale Feature are horizontally scalable on both the SAP NetWeaver Application Server component and the DB2 database server tier. Optimal scaling is supported through specific technologies such as Remote Direct Memory Access (RDMA) capabilities over a high-speed cluster interconnect. The combination of IBM System x servers with SUSE Linux Enterprise Server for SAP applications is an excellent platform, offering the price and performance advantages of modern x86-based systems.

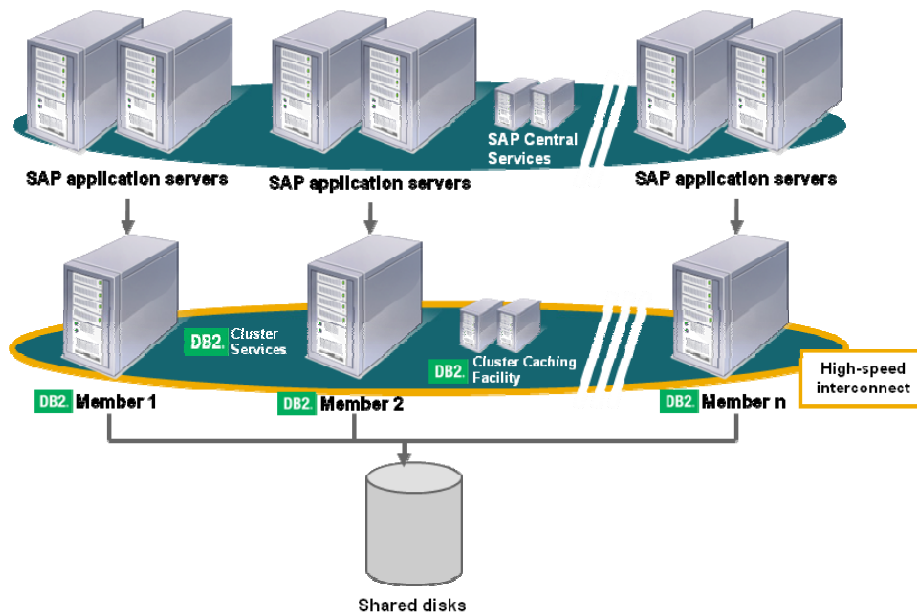


Figure 1. Overview of SAP applications on DB2 pureScale

The horizontal scaling of the solution adds redundancy to the system, which on its own increases availability. Integrated DB2 cluster services take care of all servers and ensure continuous availability in the event of failures. During maintenance, you can seamlessly take individual components out of the cluster, ensuring that operations are not affected by the maintenance. The section *DB2 pureScale Feature Use cases and Benefits for SAP Applications* covers the availability, maintenance, and scalability aspects of this solution in more detail and explains the business value of these aspects based on use cases that are specific to SAP solutions. The performance characteristics are described in the section *Performance considerations*.

IBM, Novell, and SAP have closely worked together to tightly integrate the components. Specific optimizations have been added to the components to ease the initial setup and operations of the system. These optimizations and the resulting benefits are described in the section *Integration and operation of the DB2 pureScale Feature with SAP Applications*.

## Overview of the DB2 pureScale Feature

The DB2 pureScale Feature [1] delivers a high level of availability and scalability, giving you the ability to scale out your database over a number of servers in an “active-active” configuration. The DB2 pureScale Feature offers virtually unlimited capacity for growth because you can dynamically add servers to a DB2 pureScale instance with no disruption to cluster operation. The DB2 pureScale Feature also provides continuous availability by being able to tolerate multiple component failures while continuing to provide full access to data that does not have to be recovered.

The DB2 pureScale Feature offers true application transparency. Applications require no knowledge of the underlying database system topology and do not require any modifications to run on the DB2 pureScale Feature. Applications running need not be “cluster aware” to scale, as can be the case with other scale-out architectures that rely on a heavy network-based messaging infrastructure for sharing data in a cluster. SAP applications running on the DB2 pureScale Feature can take immediate advantage of additional servers by using either the Round-Robin Connectivity Setup or reloading the Client Affinity File. As described later, the DB2 pureScale Feature also supports transparent application scaling by exploiting the latest network and hardware architectures and providing a centralized server for locking and caching.

### Main components of a DB2 pureScale instance

A DB2 pureScale instance consists of one or more DB2 servers, referred to as *members*. Each member runs a full DB2 engine with its own facilities, for example buffer pools, local lock lists, agent threads, and log files. The DB2 pureScale Feature is a shared data distributed database architecture, meaning each member has access to the same data for read and write operations. Synchronizing access to this data is required for database consistency, and being able to do so efficiently is essential for good database performance. The solution for ensuring database consistency while maintaining high levels of performance in the DB2 pureScale environment is based on DB2 for z/OS Parallel Sysplex, a shared-data DBMS that is known for its exceptional availability and scalability.

DB2 for z/OS Parallel Sysplex achieves its recognition largely because of its use of the coupling facility. The coupling facility provides a centralized location for managing locks and a global shared cache for data. Part of every DB2 pureScale instance is the pureScale cluster caching facility (CF) server, which provides the same centralized locking and global shared cache architecture. Whenever SQL processing on a member requires a lock, such as for updating a row on a page, it requests the lock from the CF. The CF determines whether the lock is available and if so, the lock is granted, and execution can continue on the member. Requests to the CF are also made when a member wants to read a page into its local buffer pool (LBP) or commit changes to a page. During a read request, if the CF has the page cached, it sends the page to the member, thus saving the member from having to read the page from disk. When the CF receives a page

write request, the modified page is sent to the CF as part of the request and is cached in the group buffer pool (GBP). At this point, the CF has the most recent version of the page and proceeds to “invalidate” (mark as “stale”) any other copies of the page that exist in the LBPs of other members.

To eliminate the CF as a single point of failure, the DB2 pureScale instance is generally configured to use two CFs. In a two-CF configuration, all requests are sent to the CF that is designated as the primary CF and are also synchronously duplexed to the secondary CF. The secondary CF can then quickly take over as the primary CF if the primary CF fails. Communication to and from both CFs is optimized with extremely low latency communication performed over an RDMA-capable network fabric.

InfiniBand and 10 Gigabit Ethernet are supported for communication in a DB2 pureScale environment. InfiniBand and 10 Gigabit Ethernet support interrupt-free RDMA, which plays a critical role in the scalability of the DB2 pureScale Feature. With RDMA, each member can directly access memory on the CF, and the CF is able to directly access memory on each member, without using any CPU time at the server being remotely accessed. For example, during commit processing, all the pages that were modified during the transaction are sent to the CF. The CF then caches these pages and determines which (if any) of the other members have (now stale) copies of these pages in their LBPs. After obtaining this information, the CF uses RDMA to directly access the pages in the members' LBPs and marks the pages as invalid. The ability to mark these pages as invalid with no interruption to the execution taking place on the member is critical for the scalability of write-intensive applications.

All servers in a DB2 pureScale instance are connected to the same disk storage through a storage area network (SAN). IBM General Parallel File System (GPFS) is used to provide access to the data with high scalability, high availability, and seamless capacity expansion. GPFS is shipped as part of the DB2 pureScale Feature and is installed and configured seamlessly during the DB2 pureScale Feature installation. The **db2cluster** command creates and manages GPFS file systems and automatically configures GPFS with known best practices for settings such as file system block size and file system cache size.

The DB2 pureScale Feature uses IBM Tivoli Systems Automation (TSA) and IBM Reliable Scalable Cluster Technology (RSCT) to automate failure detection and recovery. By offering a way to monitor the state of each of the components in a DB2 pureScale environment, TSA and RSCT help to provide continuous high availability. Upon failure detection, TSA and RSCT take the appropriate action to bring the failed component back to an operational state. Examples of

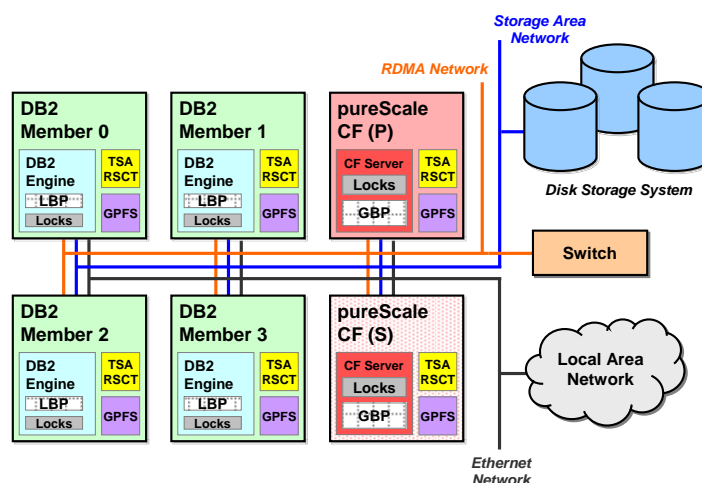


Figure 2. DB2 pureScale architecture

actions include remounting file systems, restarting a DB2 member, and rebooting a server. Like GPFS, TSA and RSCT are shipped as part of the DB2 pureScale Feature and are installed and configured seamlessly during the DB2 pureScale Feature installation. You can easily manage TSA and RSCT after installation through the **db2cluster** command.

Figure 2 illustrates the topology of a DB2 pureScale instance with four DB2 members and two DB2 pureScale CF servers. One CF server is designated as primary (P) and the other as secondary (S).

## Hardware requirements

The following hardware is required for a system on which you plan to install the IBM DB2 pureScale Feature for Enterprise Server Edition [2]:

- IBM System x servers (model x3850 X5, x3690 X5, or x3650 M3), as database server hosts for the DB2 pureScale database members and CFs.
- Any storage system that is supported by GPFS. The storage must be connected to the database server hosts through a SAN.
- An Ethernet network (LAN) to provide the communication infrastructure between the database server hosts, SAP NetWeaver Application Server hosts, and application clients.
- A high-speed RDMA network to connect the CFs and the database members. This network can be built on either InfiniBand or 10 Gigabit Ethernet architecture. The DB2 pureScale Feature currently supports only one high-speed RDMA network adapter card per server host.

Virtualized environments are currently not supported for the DB2 pureScale Feature on Linux.

### *Supported IBM System x servers*

IBM System x servers help organizations make the most of their mission-critical applications by delivering leading, mainframe-inspired reliability, availability, and serviceability (RAS):

- Redundant and fast-access components enhance availability and simplify serviceability.
- Using Chipkill, Memory ProteXion, memory mirroring, and DIMM rack sparing, System x servers help achieve maximum memory integrity and recovery from memory errors.
- An Integrated Management Module allows for remote control access.
- Advanced Light Path Diagnostics and Predictive Failure Analysis support proactive problem resolution.
- IBM Systems Director provides easy centralized management of physical and virtual resources. Active Energy Manager is an IBM Systems Director extension that helps to lower wattage and costs.



Table 1 shows an overview of the IBM System x servers that are supported for the DB2 pureScale Feature. The chart includes processor, memory, and network capabilities. The maximum RAM values are based on 16 GB or 32 GB memory modules.

IBM server model	Form factor	Intel Xeon processor series	Chassis nodes / Total processor sockets / Max. cores / Max. threads	Max. freq. in GHz	Max. DIMM Num.	Max. RAM in GB	Max. PCIe adapt. slots
X3650 M3	2U	5600	1 / 2 / 12 / 24	3.46	18 12	144 192	4
X3690 X5	2U	E7-2800	1 / 2 / 20 / 40	2.40	32	1024	5
X3690 X5 + MAX5	3U	E7-2800	1 / 2 / 20 / 40	2.40	64	2048	5
X3850 X5 (1 node)	4U	E7-4800	1 / 4 / 40 / 80	2.40	64	2048	7
x3850 X5 (1 node) + MAX5	5U	E7-4800	1 / 4 / 40 / 80	2.40	96	3072	7
x3850 X5 (2 node)	8U	E7-8800	2 / 8 / 80 / 160	2.40	128	4096	14
x3850 X5 (2 node) + 2 MAX5	10U	E7-8800	2 / 8 / 80 / 160	2.40	192	6144	14

*Table 1. IBM System x servers supported for the DB2 pureScale Feature*

All the supported IBM System x server models have an integrated dual-port Gigabit Ethernet (GbE) controller and at least four high-speed x8 (“by 8”) PCIe Gen 2 adapter slots to support high-performance adapters, such as 10Gb Ethernet, Fibre Channel and InfiniBand cards. The integrated GbE controller is primarily used for the administration of the servers and the heartbeat communication of the server cluster.

To improve throughput and resilience in the event of Ethernet network adapter failures, use Ethernet bonding. Ethernet bonding, which is also known as channel bonding, is a setup in which two or more network interfaces are combined. Similarly, use redundant SAN interface cards.

The IBM System x3650 M3 server [3] provides a good entry-level configuration. This two-socket server is built on the Intel Xeon processor 5600 series and currently supports up to 192 GB. The industry-leading, energy-efficient design of IBM System x3650 M3 servers offers up to 60% performance improvement per watt over servers built on the Intel Xeon processor 5500 series.

The x3850 X5 and x3690 X5 servers are members of the IBM eX5 high-end server family [4]. The eX5 servers are equipped with processors from the new Intel Xeon E7-X800 processor series or from the previous Intel Xeon 7500 processor series. All these processors combine exceptional raw computing power with increased memory bandwidth and support for significantly greater memory capacity than previous server generations. The x3850 X5 server has four sockets in a single chassis and memory can be scaled up to 2 TB. The two-socket server x3690 X5 is scalable to 1 TB RAM. These new IBM eX5 servers provide about four times better performance than previous System x servers that were based on the Intel Xeon 7400 processor series.

You can scale up a single-node x3850 X5 server by adding a second chassis node through Intel Quick Path Interconnect (QPI) and IBM eXA scaling, a unique IBM eX5 technology. The dual-

node x3850 X5 server allows for an 8-socket single-image system, providing up to 4 TB of memory capacity.

A further unique feature of IBM eX5 technology is a memory expansion unit, named MAX5, that you can attach to the x3850 X5 and x3690 X5 server models. MAX5 adds support for up to 1 TB of additional memory, without investments in larger servers. If you attach two MAX5 memory extensions to the dual-node x3850 X5 server, the memory capacity can reach up to 6 TB of RAM.

The maximum memory values of the IBM eX5 servers are based on 32 GB memory modules, which are supported by the new Intel Xeon E7-X800 processor series.

IBM eX5 servers achieved many leadership results at 2-tier benchmarks across the board - with two, four, and eight sockets – and often left the competition far behind them. The performance capacity of the dual-node x3850 X5 server scales up beyond 100,000 SAPS [5].

IBM eX5 servers can utilize eXFlash, which is a local storage option based on solid state drives (SSDs). Each IBM eXFlash unit supports eight 1.8-inch high IOPS SSDs, each of which provides 50 Gigabytes or 200 Gigabytes storage capacity. Organizations can add up to three such eXFlash modules per x3690 X5 system, and up to two modules in one x3850 X5 server. The local storage of an x3650 X5 server is scalable up to 4.8 Terabytes. One eXFlash unit can service approximately 48,000 read and 16,000 write I/O instructions per second (IOPS), which would require about 125 parallel mechanical disks spinning at 15000 rpm. eXFlash offers superior uptime with three times the reliability of mechanical disk drives. SSDs have no moving parts that could fail, and they use Enterprise Wear-Leveling to even further extend their lifespan. eXFlash storage also takes up much less space and requires less power and cooling than HDD storage. The eXFlash units are connected to the same types of disk controllers as internal SAS/SATA disks. For higher IOPS performance, you can configure two disk controllers per eXFlash option.

When you plan to use two-socket servers for the DB2 pureScale Feature and need more or full height PCIe adapter slots, alternatively select the four-socket x3850 X5 servers with just two processors.

## ***Supported IBM storage systems***

IBM System Storage [6] ranges from high-performance, high-capacity storage platforms such as the IBM System Storage DS8000 series and the IBM XIV Storage System, through midrange systems such as the IBM Storwize V7000 and IBM System Storage DS5000 series, to entry-level configurations such as the IBM System Storage DS3000 series.

When one or more database members fail, the functional sub-cluster requires operational quorum for Reliable Scalable Cluster Technology (RSCT). If exactly half of the database members remain in communication, the sub-cluster can exclusively reserve a tiebreaker to obtain the operational quorum.

To rapidly fence and recover failed database members, use disk storage that is certified for SCSI-3 Persistent Reserve (PR) and GPFS software. Disk storage devices without SCSI-3 PR support

work with the DB2 pureScale Feature, but the fencing time is longer, which increases DB2 member failover time.

Shared storage devices and multipath I/O driver combinations that are supported by the DB2 pureScale Feature [7] can be divided into three categories: Category 1 combinations successfully support both the DB2 cluster services tiebreaker and fast I/O fencing, resulting in the highest resiliency and fastest recovery times. Category 2 devices only support the DB2 cluster services tiebreaker. Category 3 combinations have not been validated with the DB2 pureScale Feature, though they are supported.

Improve data center productivity by using storage virtualization. The IBM System Storage SAN Volume Controller (SVC) creates a single point of control, simplifying administration and enhancing security by combining storage systems from multiple vendors into a single pool of capacity. The Storwize V7000, DS5000, and DS3000 storage systems are ideal building blocks for SVC virtualized environments.

Solid-state drive (SSD) options for DS8000, XIV Storage System, Storwize V7000, and SVC products accelerate the performance of the SAP solution. The IBM System Storage Easy Tier feature optimizes SSD deployments simply and automatically. For example, the feature can transparently move the data that are more frequently accessed to faster SSDs while less frequently used data is stored on slower, less-expensive devices.

### ***Supported high-speed RDMA networks***

The DB2 pureScale Feature supports two options for the high-speed RDMA network: the InfiniBand (IB) network or the 10 Gigabit Ethernet (10 GbE) network.

For an InfiniBand (IB) network, the following components are required:

- Mellanox IB (40 Gb) dual-port network cards (MT26428),
- Mellanox InfiniScale IV QDR IB switches (36 QSFP ports), and
- Mellanox FabricIT IB Fabric Management.

For a 10 Gigabit Ethernet (10 GbE) network, the following components are required:

- Mellanox 10 Gigabit Ethernet (10 GbE) dual-port network cards (MT26448) and
- 10 GbE network switches supporting priority-based flow control.

The 10 GbE option is currently supported only on SUSE Linux Enterprise Server 10.

### ***Hardware configurations***

The DB2 pureScale Feature gives organizations the ability to scale out a database system to multiple physical servers. While system requirements are growing, new server nodes can be easily added. The existing database server hosts do not have to be replaced; they are still being used. The horizontal scaling further increases database availability by redundancy. Both aspects (scalability and redundancy) imply that organizations select a set of preferably identical physical servers, which can easily be replaced or extended.

Table 2 outlines T-shirt sizing of supported IBM System x cluster configurations. Select a suitable configuration depending on current and future system requirements. Each cluster configuration uses one of the supported System x server models as a building block: The x3650 M3 servers are ideal for small "S" size configurations at small and midsize enterprises (SME) or test setups. For mid-range system requirements, the medium "M" size server model x3690 X5 is the best choice. Even though both 2-socket servers have the same 2U rack size, the x3690 X5 model offers higher computing performance and more memory and storage capacities than the x3650 M3 server. High-end configurations for large enterprises build on the large "L" size servers x3850 X5 having 4 sockets, or on the very large "XL" 8-socket versions. The figure after the T-shirt size refers to the number of physical servers in the cluster configuration.

In a two-node configuration, for example, "S2", "M2", and "L2", one cluster caching facility (CF) and one database member reside on each of the two separate hosts. The CF and the database member share computing power, memory, and I/O resources of the same host.

In the "plus" configurations, for example, "S 2 plus" and "M 2 plus", the DB2 pureScale processes only consume a small portion of the total hardware resources. The physical servers are capable of hosting SAP NetWeaver Application Server components, preferably in a two-node high availability cluster.

When the requirements of your database are growing, move the SAP NetWeaver Application Server components to separate physical hosts, which can be IBM BladeServer models. Only the DB2 pureScale Feature processes continue running on the cluster nodes. Further growth of the database requires that the CFs and database members run on separate hosts. The cluster size can easily increase to four nodes with two members, six nodes with four members, or eight nodes having six database members. In the large "L" size category, the corresponding cluster configurations are called "L 4", "L 6", and "L 8".

Find more detailed sizing recommendations for memory and the CPUs of the CFs and database members in the section *Performance considerations*. While all database member hosts can have the same CPU and memory configurations, the resource requirements of the CFs depend on the number of members. In the case of four or even six database members, the memory sizing of the CF host could come up with a MAX5 memory attachment to provide additional memory modules, for example, 8 GB DIMMs. The alternative would be a partial or complete replacement with double-size memory modules, for example, 16 GB DIMMs. The sizing of an x3650 M3 CF host in an "S 6" cluster could result in more than 192 Gigabytes of memory. In this case, an X5 server would be a better choice.

In the last column of Table 2, the character "N" refers to an open number of application server hosts, which are separate from the database cluster hosts.

T-shirt size	IBM server models	Number of cluster nodes hosting		
		DB2 pureScale CFs	DB2 pureScale Members	SAP NetWeaver Application Servers
S 2 plus	x3650 M3 (2S)	2		
S 2	x3650 M3 (2S)	2		N
S 4	x3650 M3 (2S)	2	2	N
S 6	x3650 M3 (2S)	2 (X5)	4	N
M 2 plus	x3690 X5 (2S)	2		
M 2	x3690 X5 (2S)	2		N
M 4	x3690 X5 (2S)	2	2	N
M 6	x3690 X5 (2S)	2 (MAX5)	4	N
L 2 / XL 2 plus	X3850 X5 (4S/8S)	2		
L 2 / XL 2	X3850 X5 (4S/8S)	2		N
L 4 / XL 4	X3850 X5 (4S/8S)	2	2	N
L 6 / XL 6	X3850 X5 (4S/8S)	2 (MAX5)	4	N
L 8 / XL 8	X3850 X5 (4S/8S)	2 (MAX5)	6	N

Table 2. T-shirt sizing of supported IBM System x cluster configurations

The following figures illustrate five sample IBM System x server and IBM System Storage configurations for using the DB2 pureScale Feature, ranging from a high-end configuration to an entry-level configuration. The hardware running SAP NetWeaver Application Server is not shown in the figures.

## High-end system configuration

Figure 3 shows a high-end configuration based on eight four-socket IBM System x3850 X5 servers.

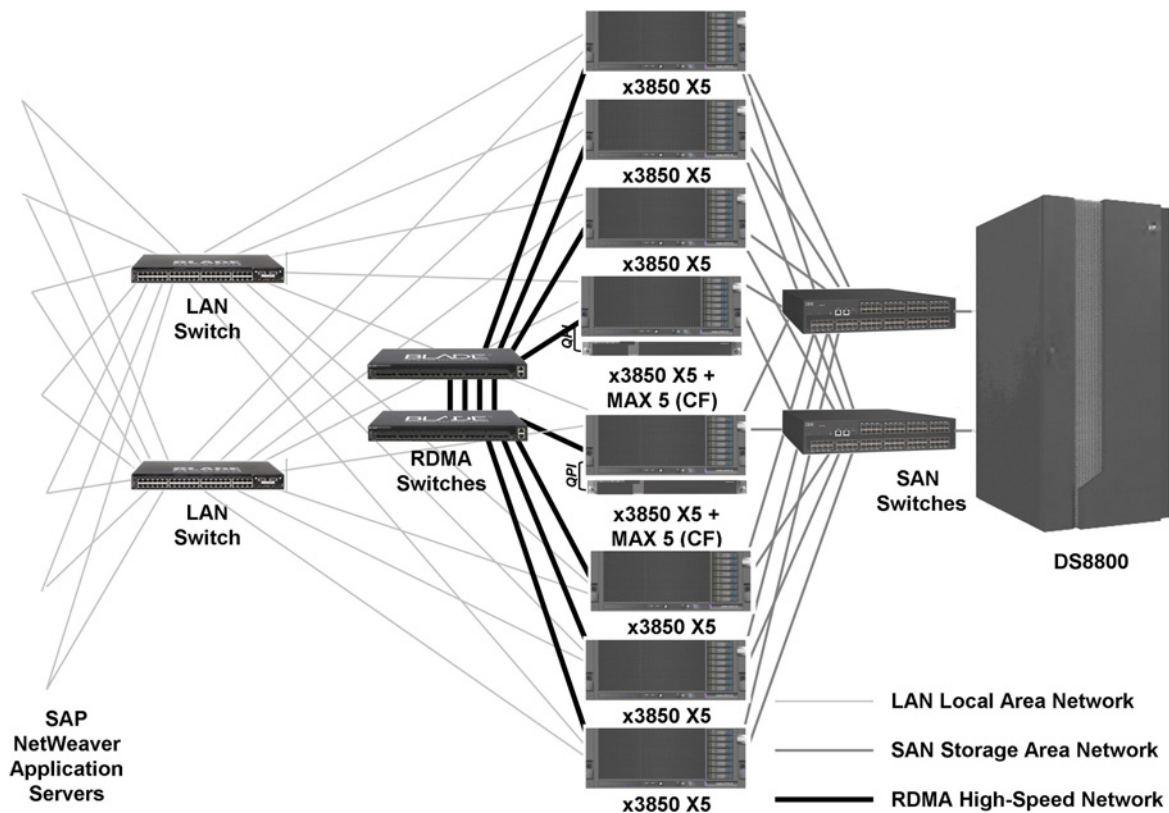


Figure 3: High-end configuration of System x for the DB2 pureScale Feature

The two servers in the middle, marked with (CF), host the primary and secondary CFs. The remaining six servers host the database members. The figure illustrates how a MAX5 memory extension unit can be attached to each of the CF servers to meet increasing memory demands.

IBM currently supports DB2 pureScale clusters with up to two high-speed RDMA network switches. A database server host can currently have only one direct connection to one such network switch. In the figure, the upper database server hosts are directly connected to the upper high-speed RDMA network switch, and the lower database server hosts are directly connected to the lower high-speed RDMA network switch. If the lower network switch fails, the upper network switch and the database server hosts that are attached to the upper switch are still functional. In high-end configurations, you can select either InfiniBand or 10 Gb Ethernet technology for the high-speed RDMA network.

The figure also shows the LAN and SAN infrastructure with redundant switch devices.

IBM System Storage DS8000 is an example of a high-end storage system that is supported by DB2 pureScale Feature in category 1 [7].



## Midrange system configurations

Figure 4 depicts an upper midrange configuration comprising six four-socket IBM System x3850 X5 servers, hosting database members and CFs.

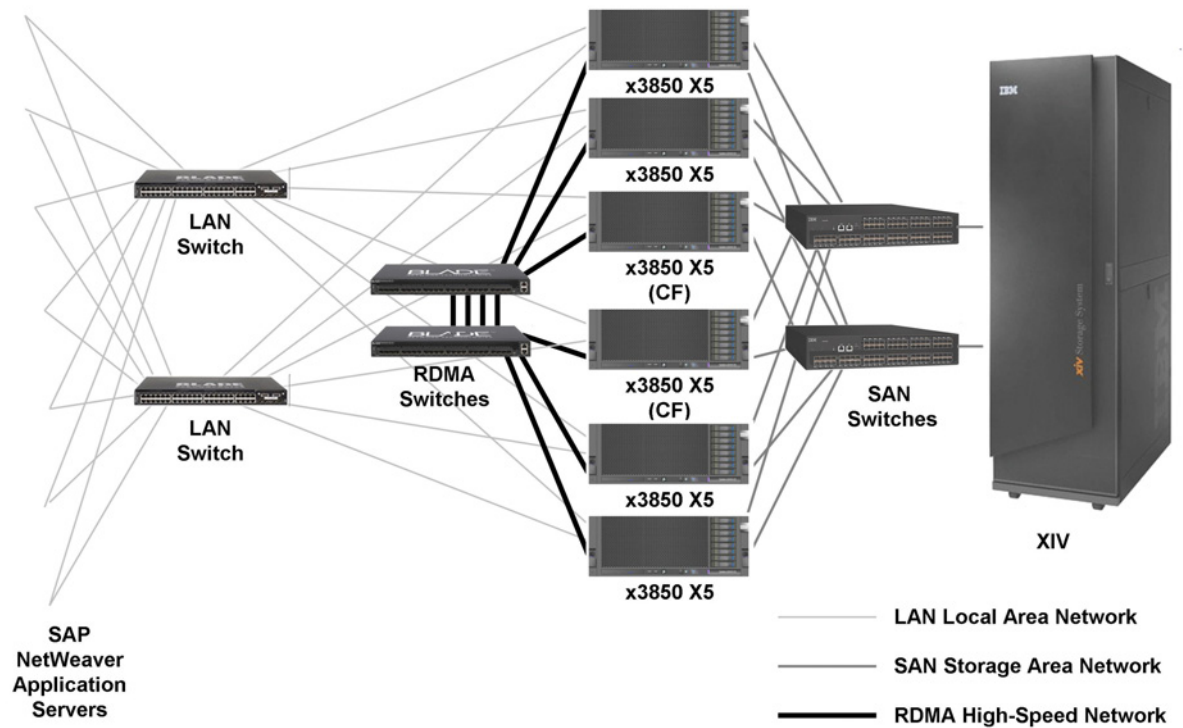


Figure 4. Upper midrange configuration of System x for the DB2 pureScale Feature

In the figure, all the database server hosts are coupled to an IBM XIV Storage System. Alternatively, select a set of IBM Storwize V7000 systems. Regarding DB2 cluster services tiebreaker and fast I/O fencing support, check out the validations of storage devices and multipath I/O drivers [7].

Figure 5 illustrates a midrange configuration using six two-socket IBM System x3690 X5 servers.

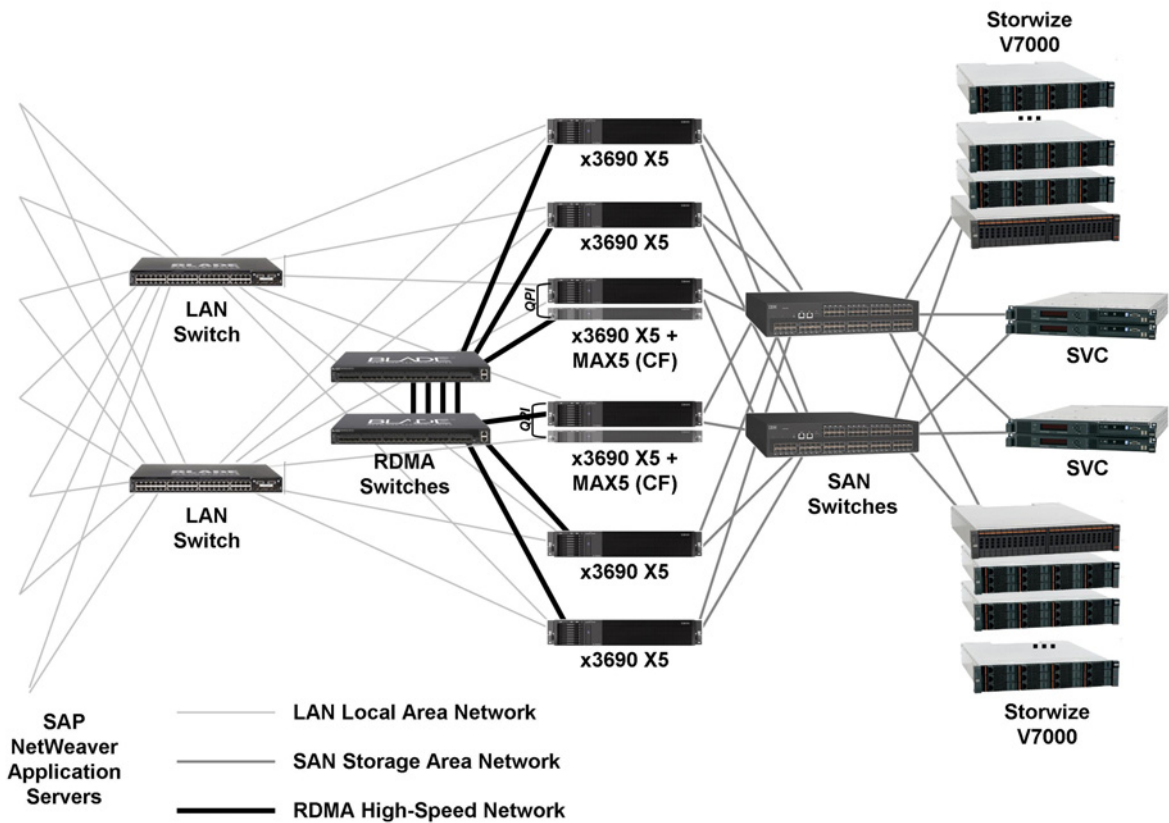


Figure 5. Midrange configuration of System x for the DB2 pureScale Feature

The database members and CFs run on the two-socket IBM System x3690 X5 servers. When the number of database members increases, scale up the CF hosts by replacing memory modules or attaching MAX5 memory expansion units. Optionally, use four-socket x3850 X5 servers only for the CFs.

In the sample configuration, all the database server hosts are coupled to a cluster of SAN Volume Controllers (SVCs) and Storwize V7000 systems.. For fast I/O fencing support, see the list of validated device combinations [7].



Figure 6 illustrates a lower midrange configuration comprising four two-socket IBM System x3690 X5 servers, hosting the database members and CFs. If the CF hosts have free resources, they can be shared with database members.

The LAN and SAN infrastructures use non-redundant network switches. If the Ethernet switch for the LAN has a sufficient number of free 10 Gb ports, they can be used for the RDMA network.

All database member and CF hosts are coupled to an IBM Storwize V7000 system through a SAN.

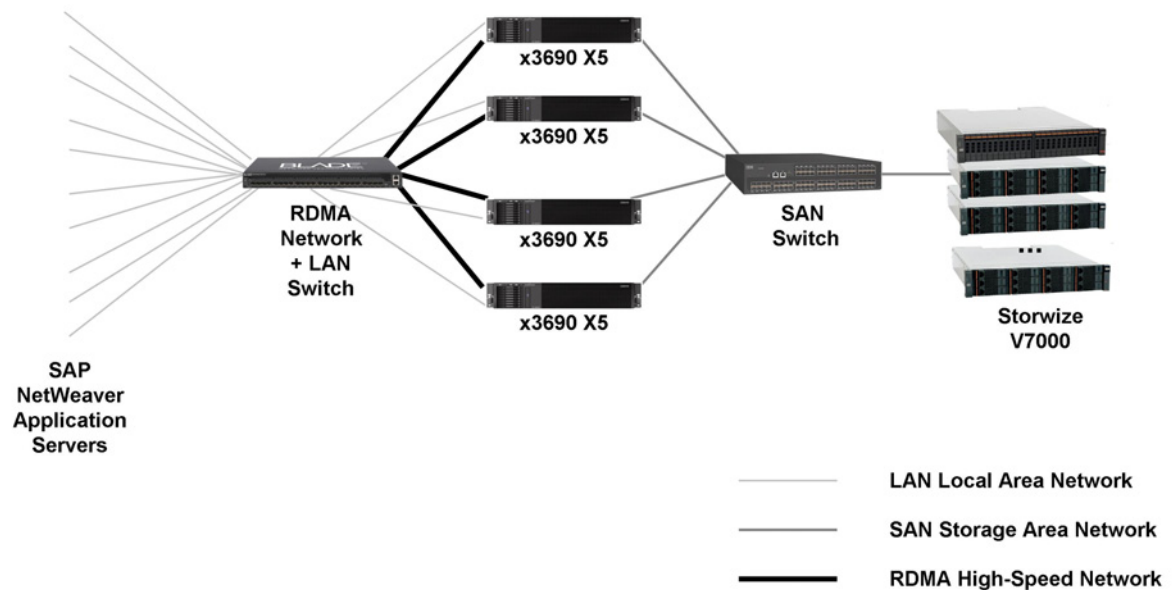


Figure 6. Lower midrange configuration of System x for the DB2 pureScale Feature

## Entry-level system configuration

An entry-level or test configuration is shown in Figure 7, which has a reduced acquisition cost but still allows for future capacity expansion.

Two IBM System x3650 M3 servers host the DB2 pureScale Feature software. Each x3650 M3 server has one CF and one database member. The high-speed RDMA network is preferably built on 10 Gb Ethernet, which can share one switch with the LAN infrastructure. At least one SAN

fabric switch is needed to couple both x3650 M3 servers to a storage system, which can be an entry-level storage system of the DS3000 series.

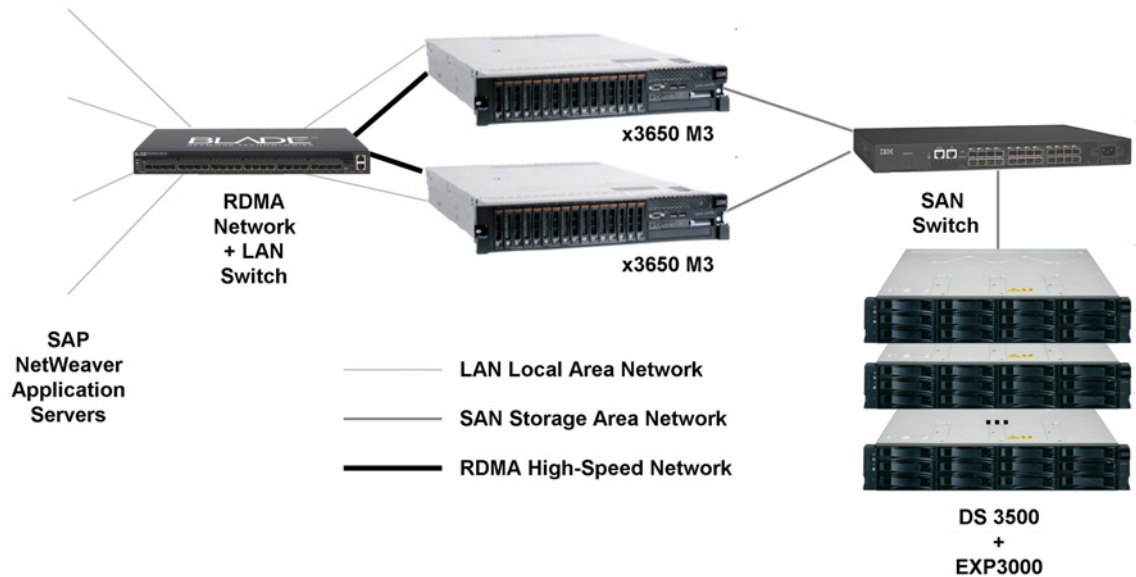


Figure 7. Entry-level or test configuration of System x for the DB2 pureScale Feature

## SUSE Linux Enterprise Server as an optimized Operating System for SAP Solutions

Through collaboration with Novell, SAP selected SUSE Linux Enterprise Server as the only operating system for solutions such as SAP NetWeaver Business Warehouse Accelerator software, the SAP NetWeaver Enterprise Search application, the SAP Discovery System, and the SAP StreamWork™ application. Also the new SAP HANA™ application software is based on SUSE Linux Enterprise Server 11. SUSE Linux Enterprise Server is one of two reference platforms for SAP software development.

SUSE Linux Enterprise Server already includes the “sap-conf” and “sap-locale” packages, which prepare SUSE Linux Enterprise Server for the installation of SAP software.

### *SUSE Linux Enterprise Server for SAP Applications*

Novell provides SUSE Linux Enterprise Server for SAP applications, which is optimized for mission-critical SAP software solutions. SUSE Linux Enterprise Server for SAP applications comprises software, maintenance, and support offerings.

SUSE Linux Enterprise Server for SAP applications version 11 service pack 1 differs from the regular SUSE Linux Enterprise Server (SLES) version 11 SP1 in that the SAP-specific version contains several additional product features, which are listed in Table 3.

Product feature description included in	SUSE Linux Enterprise Server 11	SUSE Linux Enterprise Server for SAP applications 11
Installation wizard for SAP solutions		x
Optimized paging support for large in-memory workloads		x
SAP NetWeaver-compatible JVM		x
Dedicated SAP-specific update channel		x
Extended service pack overlap support		x
High-availability extension		x
Priority support option	x	x
Premium support entitlement		x

Table 3. SUSE Linux Enterprise Server for SAP applications 11 product features

SUSE Linux Enterprise Server for SAP applications version 11 inherits all SAP hardware and software certifications from SUSE Linux Enterprise Server 11. These SAP certifications result from a quality assurance process of these products released for your mission-critical SAP applications. Hardware certifications (for example of IBM System x), and software certifications (for example of IBM DB2) are valid for the combination of the certified product and each of the Linux versions: SUSE Linux Enterprise Server and SUSE Linux Enterprise Server for SAP applications.

These SUSE Linux Enterprise Server versions support the latest multicore processor architectures from Intel and AMD, which are built in IBM System x and BladeCenter servers. SUSE Linux Enterprise Server for SAP applications is available only for the x86\_64 architecture.

The following sections explain SUSE Linux Enterprise Server for SAP applications features that are relevant to a DB2 pureScale Feature deployment. For a complete explanation of all features and functions, see the white paper *IBM System x powered by SUSE Linux Enterprise Server for SAP applications* [8].

## **Installation wizard for SAP solutions**

An SAP software installation normally requires a skilled and SAP-certified consultant, and often takes several days. The SAP software installation may comprise manual installation steps, which are tedious and error-prone. Although IT organizations often develop best practices to set up an SAP solution more efficiently, the manual installation steps are repeated again and again.

Over the last years, the Novell SUSE engineering staff automated and simplified this SAP software installation process. Based on SUSE Autoyast technology, the SUSE engineers developed an Installation wizard for SAP solutions that not only installs the Linux operating system but also automates the complete SAP software installation process.

Previous versions of the installation wizard, based on SUSE Linux Enterprise Server 10, required SAP software installation images that were adapted to the installation on this specific operating system. However, SUSE Linux Enterprise Server for SAP applications 11 now provides a new Installation wizard for SAP solutions that supports automated installations from generic SAP software installation images, which are independent of the operating system. The new Installation wizard for SAP solutions can help to automate the setup of SAP NetWeaver Application Server components in a DB2 pureScale environment.

## **Optimized paging support for large in-memory workloads**

SUSE Linux Enterprise Server for SAP applications supports optimized paging behavior for large workloads from SAP applications, which keep large amounts of business application data in memory. A regular Linux kernel uses remaining free memory for other purposes. For example, the page cache accelerates I/O operations by storing data read from or written to storage devices. When the business applications need more memory, the regular Linux kernel can dynamically reduce the page cache size.

To free up further memory, the kernel can copy data pages from memory to swap devices, which is called page-out. The kernel decides which memory areas are paged out, for example, based on a least recently used algorithm. When backup jobs cause many I/O operations, the kernel will

preferably page out business application data, which has not been accessed for some time. This paging may have a negative impact on the performance of the business application. When the application tries to access paged-out data, the corresponding pages have to be copied back to memory. This page-in process dramatically reduces response times of the business application.

SUSE engineers developed a new Linux kernel feature that allows system administrators to limit the size of the page cache. When business application data and page cache data are competing for free memory, the kernel now favors to keep business application data in memory over data cached for I/O operations. This feature is in particular useful for large in-memory database and SAP applications running on the Java stack, where a Java Virtual Machine performs garbage collections accessing all business application memory at one time. Thus, the page cache limit feature can help to minimize page-in activities and improve the performance of the business applications.

Reasonable performance improvements have been observed with large SAP ERP and SAP Customer Relationship Management applications, which use large amounts of memory when the page cache limit was enabled.

The Page cache limit feature is only supported with SUSE Linux Enterprise Server for SAP applications 11 [9].

### **SAP NetWeaver compatible Java Virtual Machine (JVM)**

SAP supports only the specific IBM Java Virtual Machine (JVM) version 1.4.2 for SAP NetWeaver Java stacks. However, the regular SUSE Linux Enterprise Server provides a newer JVM version 1.6. To ensure compatibility with the SAP NetWeaver technology platform, you must remove the newer JVM version and replace it with the SAP supported version at each SUSE Linux Enterprise Server installation.

SUSE Linux Enterprise Server for SAP applications not only automatically provides the JVM version that is required by SAP, but it also provides support for this JVM version. You do not have to downgrade the JVM version, and you receive a fully supported software stack for the SAP applications, Java and the Linux operating system.

### **SAP specific update channel**

SUSE Linux Enterprise Server for SAP applications has an additional update channel for patches, fixes and updates of the SAP specific software packages.

### **Extended service pack overlap support**

About once a year, Novell delivers a new service pack (SP) for the version 11 of SUSE Linux Enterprise Server for SAP applications. A service pack is an assembly of all patches and fixes provided since the previous service pack. In most cases, it also contains new or updated drivers to support new hardware models. After the release of the new service pack, patches and bug fixes will be provided for the new service pack (# n), and, over a period of 6 months, for the previous service pack (# n-1). After this period, users of the previous service pack (# n-1) need to upgrade to the new service pack (# n) if they want to receive further patches.

Because a period of 6 months is not convenient for many SAP customers, SUSE Linux Enterprise Server for SAP applications has an extended support period of 18 months for all preceding service packs (# n-1, # n-2, # n-3, and so on) This extended service pack overlap support is included in each subscription of SUSE Linux Enterprise Server for SAP applications.

### **Priority support option**

Customers of SUSE Linux Enterprise Server for SAP applications can subscribe to either basic or priority support. Both types of subscriptions are respectively available for either a one-year or a three-year period. The basic support subscription entitles customers to receive fixes, patches, updates and service packs for SUSE Linux Enterprise Server for SAP applications. After registration of a system at the Novell Customer Center, the system gets access to the update channels for the registered Novell products.

A basic support subscription does not include support services. In most cases, customers of basic support subscriptions have a support contract, for example with IBM.

In addition to basic support, the priority support option for SUSE Linux Enterprise Server for SAP applications includes the following services:

- 24 x 7 support via e-mail or phone
- Integration in SAP support

Priority support customers are entitled to receive SUSE Linux Enterprise Server operating system support through SAP Solution Manager and other established SAP support channels. SAP support has a specific support component for problems related to SUSE Linux Enterprise Server. A dedicated Novell Technical Support (NTS) team is integrated in SAP's support workflow. Thereby, SAP support provides a single point of contact for all problems that are related to SAP or SUSE Linux software.

### **Premium Support Entitlement**

In addition to basic and priority support, Novell offers add-on premium support to customers who would like to have a dedicated support engineer and/or support account manager. A condition for a premium support contract with Novell is that the customer has priority support subscriptions for the business-critical systems. Because SUSE Linux Enterprise Server for SAP applications already includes additional maintenance and support services in the basic support offering, customers of SUSE Linux Enterprise Server for SAP applications basic support qualify for a premium support contract with Novell. The price of premium support is not included in the SUSE Linux Enterprise Server for SAP applications subscription.

### ***Linux support for the DB2 pureScale Feature***

SAP, IBM, and Novell have a close technology and support partnership. SAP founded the SAP LinuxLab [10] not only to support SAP business solutions on Linux operating systems but also to ensure enterprise-class support for Linux systems that are used as platforms for SAP solutions. As SAP global technology partners, IBM and Novell have technical subject matter experts who

are continuously present in SAP's LinuxLab in St. Leon-Rot, Germany. Furthermore, Novell has technical engineers who are dedicated to working with IBM specialists at IBM labs.

Novell and SAP have extended their support partnership with SUSE Linux Enterprise Priority Support for SAP applications. This unique Enterprise Linux support subscription integrates with SAP Solution Manager and entitles SAP users to receive Linux support through their established SAP support channel. Furthermore, this support offering leverages the SAP LinuxLab.

Figure 8 describes, at a high level, how the support process leverages the SAP Customer Support Network (CSN) to deliver support for the DB2 pureScale Feature on Linux operating systems. A support incident that is issued by a customer support organization

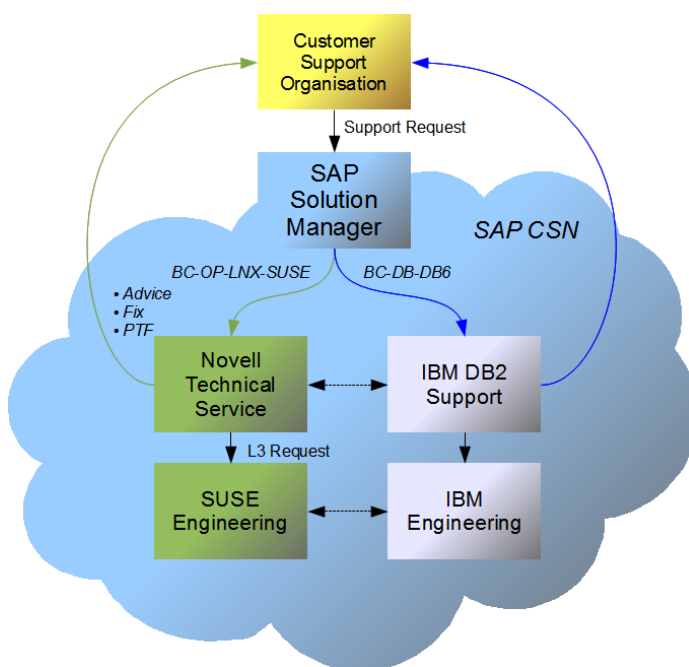


Figure 8. SAP support for DB2

(L1) through SAP Solution Manager or other SAP support channels is rerouted to Novell Technical Service (NTS, Level 2) when designated as a Linux issue (SAP support queue BC-OP-LNX-SUSE). NTS then works with the SUSE engineering staff (L3) and IBM DB2 support to provide a problem solution. This can be advice on how to solve the problem, a problem fix, or a program temporary fix (PTF), which is usually provided by the SUSE engineering staff. PTFs can be made available to other customers through the SUSE maintenance channels and are consolidated in SUSE Linux Enterprise Server Service Packs.

If the incident is designated as a DB2 issue, the incident is routed to IBM DB2 support (SAP support queue BC-DB-DB6). Again, the support organizations from IBM, SAP, and Novell jointly work on a problem solution. From your perspective, there is only one support interface.

## Virtualization

You can use SUSE Linux Enterprise Server for physical installations and for virtual machine guests on VMware vSphere, XEN hypervisor, and KVM. Although Novell supports Microsoft Hyper-V virtualization, SAP does not support Microsoft Hyper-V virtualization on Linux operating systems [11].

As mentioned earlier, the DB2 pureScale Feature, which requires RDMA, is currently not yet supported in virtualized environments. However, virtualization may be applied to the SAP NetWeaver Application Server components. Even when the database instances of the DB2 pureScale Feature can be deployed in virtualized environments, administrators can decide to



keep them on physical hardware to avoid potential I/O bottlenecks caused by the hypervisor.

SUSE Linux Enterprise Server for SAP applications licensing is based on physical servers. One subscription is valid for a system with up to 32 CPU sockets. A system with 64 CPU sockets requires two subscriptions. This licensing model is also valid for virtualized servers: the number of subscriptions is independent of the number of virtual machines. For example, running four virtual machines on one eight-socket server requires only one subscription. The licensing is also independent of the virtualization technology you use.



## DB2 pureScale Feature Use Cases and Benefits for SAP Applications

The DB2 pureScale Feature offers unlimited capacity and continuous availability for transaction processing, enabling you to adapt your SAP solution to changing workloads without the need for system downtime. This section describes several use cases and system configuration options that benefit from the scalability and high-availability (HA) capabilities of the DB2 pureScale Feature.

### *A complete HA solution for SAP Applications*

The DB2 pureScale Feature has included the cluster service components TSA and RSCT, which provide built-in HA for the DB2 database. HA for SAP applications, however, is not a database-only task. An HA solution must cover the database server and the SAP NetWeaver Application Server component to be effective. Consequently, the cluster service components of the DB2 pureScale Feature provide HA for the complete SAP application (for both the DB2 pureScale database server and the SAP NetWeaver Application Server tier).

On the SAP NetWeaver Application Server component, the TSA component covers HA for SAP central services and SAP NetWeaver Application Server components. The `sapscssetup` tool, which is provided by SAP, sets up HA for SAP central services such as the SAP enqueue server and enqueue replication service (ERS), the message server, the SAP Web dispatcher, the SAP router, the SAP gateway, and the central system message log. TSA policies set up and manage HA for the SAP NetWeaver Application Server components. If an SAP component fails, TSA automatically orchestrates failover and restart of the failed component, ensuring overall system availability.

If a DB2 member fails, the DB2 pureScale Feature automatically detects the failure. The SAP NetWeaver Application Server components that are connected to the failed member are automatically redirected to other members of the cluster. The DB2 pureScale Feature automatically recovers the failed member and brings it back online. Recovery can take place on the home host of the failed member if it is still available or on a different host, in the case of a hardware failure. During the recovery of the failed member, only the portion of data that was inflight at the failed member is locked. All other data and all other members remain fully online and unaffected. After the DB2 member is back online, any SAP NetWeaver Application Server component that was redirected to other DB2 members is automatically and transparently redirected back to the original member at the next transaction boundary.

If the primary CF fails, DB2 also automatically detects and corrects the failure. The secondary CF takes over and becomes the primary CF. Because of the permanent duplexing of data between the primary and secondary CFs, there is only a minimal delay in takeover while the secondary CF completes the global lock list with page read locks that it retrieved from the DB2 members. The takeover is fully transparent to the applications, and no inflight data is lost.

If the secondary CF fails, there is no impact. After the restart, it is integrated as the secondary CF again. When the secondary CF is restarted, it first enters a CATCHUP state to synchronize the

data with the primary CF. When this is done, the state switches to PEER state, and the secondary CF is available for failover again.

If the failure of a DB2 member or CF was due to an unrecoverable hardware failure, an alert is displayed in the DBA Cockpit to make you aware that you must take action and bring the server back online. In all other cases, the DB2 pureScale Feature corrects the problem automatically, without your intervention.

To make best use of DB2 pureScale HA, you should reserve some spare capacity in your DB2 pureScale database server so that the cluster can handle your workload if a DB2 member fails. There are two ways to accommodate this. One option is to size particular DB2 members such that these members can carry the additional workload of one other DB2 member, if it fails. In this case, the connected SAP NetWeaver Application Server components fail over to the surviving members and remain connected to them until the failed member comes back online. Another option is to add one or more spare members to the cluster, which will take over the work if a member or members fail. This second option has two advantages. Firstly, redirected work does not affect work on other DB2 members. Secondly, the spare members provide additional capacity, which you can use to cover temporary workload spikes or to take over work during maintenance.

### *Non-disruptive maintenance*

Non-disruptive maintenance is a very high priority for businesses today. With the DB2 pureScale Feature, you can perform stealth maintenance for operating system and server hardware. To make the maintenance transparent, you should allocate enough resources to your DB2 pureScale cluster to cover your workload demands in the absence of a DB2 member. To do so, you can size the DB2 members big enough to handle the workload in the absence of a DB2 member, or you can have one or more dedicated spare DB2 members that will compensate for the temporary absence of a DB2 member.

To perform maintenance on a DB2 member, drain the workload off the DB2 member by using the DB2 **QUIESCE** command. Workload from the SAP solution is redirected transparently to other DB2 members at transaction boundaries without disrupting the business. The workload can be directed either to DB2 members that were already processing work or to spare DB2 members that you dedicated to the cluster for HA and maintenance tasks.

After you drain all work off the DB2 member, you can put the server into maintenance mode and start the maintenance work. After you have completed the maintenance task, you can reintegrate the server back into the DB2 pureScale cluster. The SAP NetWeaver Application Server components detect that the DB2 member is back online and automatically route their work back to the DB2 member at the next transaction boundary, transparently to users.

To perform maintenance on a CF, you can temporarily stop and take the CF out of the DB2 pureScale cluster. In the case of maintenance on the primary CF, the secondary CF automatically takes over and become the primary CF. After you have completed the maintenance work on the CF and you have re-started the CF, then the newly reintegrated CF starts automatically as the secondary CF. Again, the maintenance is transparent to users. The application is unaffected.

## ***Workload isolation within an SAP solution***

You might want to isolate individual workloads in your SAP solution. Workload isolation helps ensure that dedicated resources are assigned to specific workloads and that service level agreements that are related to criteria such as response times or maximum processing times are met. Although you have been able to use DB2 workload management to set up and manage workload isolation since DB2 V9.5, the DB2 pureScale Feature provides additional options to drive workload assignment and isolation.

A DB2 pureScale cluster consists of multiple DB2 members, which process the workload from the SAP NetWeaver Application Server components. You can use this implicit division of work for workload isolation. For example, if you want to isolate dialog workload from batch workload, you can set up SAP NetWeaver Application Server components that are dedicated to batch workload and connect these batch application servers to dedicated DB2 members.

Similarly, you can assign dedicated SAP NetWeaver Application Server components that are connected dedicated DB2 members to high-priority workload or to particular users or applications. Users can be directed to their dedicated SAP NetWeaver Application Server components at logon time through SAP logon groups. For example, if you run both the sales and distribution (SD) and HR functionalities in SAP ERP, dedicated SAP NetWeaver Application Server components for SD and HCM that are connected to dedicated DB2 members ensure that these workloads have minimal interference, dedicated resources, and optimal data locality.

Last but not least, using dedicated servers for specific workloads can greatly help you understand the impact of new processes or businesses on your SAP solution and thus reduce the risk of change. Suppose you plan to add a new business entity or process to your SAP solution. By setting up dedicated SAP NetWeaver Application Server components and DB2 members for this new workload, you create an isolated environment where the new workload has minimal impact on the existing business. You can monitor the resource consumption of the new business and see whether and how it interferes with the existing workload. After you have a firm understanding of the impact of the new workload, you can then decide whether it should stay separate or be handled within the existing infrastructure.

## ***Use of the most cost-efficient computing hardware***

The transactional workload in your SAP application determines the computing power needed to process your business. You can provide the needed computing power by using a few database servers with more computing power or a higher number of database servers with smaller computing power. Although using fewer, bigger database servers helps reduce the number of operating system images and database servers that you must maintain, using a higher number of smaller database servers is typically more cost efficient from a hardware perspective.

SAP applications have always provided an efficient scale-out capability on the application server tier through its three-tier client/server architecture. You can scale computing capacity through the use of multiple smaller SAP NetWeaver Application Server components instead of fewer, larger ones.

With the DB2 pureScale Feature, you can now follow the same paradigm for the database server tier of SAP applications. The total computing capacity of a DB2 pureScale database server comprises the capacity of the multiple computers that form the DB2 pureScale cluster.

The design of the DB2 pureScale Feature integrates several software and hardware technologies that result in excellent scaling. For example, the central component CF helps keep the communication cost low as the number of members grows. Also, data pages in the global cache are shared between the members and the CF through RDMA without requiring any processor time or I/O cycles on members. All operations are performed over the high-speed interconnect, resulting in very fast round-trip communications.

Using the DB2 pureScale Feature as the database server for SAP applications can help reduce hardware cost noticeably. Also, because the DB2 pureScale Feature provides continuous availability to SAP applications, the overall solution reduces cost of ownership and operations.

### ***Efficient management of capacity growth: Grow DB2 database servers as your business grows***

As your business grows, your current database server might not be able to keep up with your increased workload. If you are not using a DB2 pureScale database server, you must purchase a larger database server to accommodate the growing demand. With larger hardware typically being much more costly than smaller hardware, this purchase can result in a significant additional investment. In exceptional cases, you might not even be able to use larger hardware, because you already have the largest available machine.

The DB2 pureScale Feature helps manage additional growth more efficiently. Using the scale-out capabilities of the DB2 pureScale Feature, you can add DB2 members to your database server to increase capacity without replacing your existing hardware. You can grow your DB2 database server by adding newer and more powerful hardware to the DB2 pureScale cluster while still using your existing hardware and protecting your existing investments.

### ***Temporary capacity increases to cover workload peaks***

In business processing, a temporary demand for increased capacity often arises. Such situations occur, for example, during quarter-end or year-end closing. Other situations that might require a temporary capacity increase are seasonal changes such as year-end business in retail or the profitability calculation process in banking.

To temporarily increase the database server capacity of your SAP solution, you can add DB2 members to your DB2 pureScale cluster. You can configure your DB2 pureScale cluster with additional spare members up front allowing you to add capacity later without further administrative action. These spare members can be initially inactive if you want to use the hardware for other purposes. Alternatively, you can have the spare members up and running so that they also serve as additional capacity for HA and maintenance purposes.

When additional capacity is required, the spare DB2 members are activated if they were not already active. The SAP NetWeaver Application Server components transparently redirect their work to the additional DB2 members at the next transaction boundary.

After the need for additional capacity has ended, the SAP NetWeaver Application Server components can be notified to transparently route their work away from the spare member back to the original DB2 members as configured.

## Performance considerations

### *Workload characteristics and the cluster caching facility (CF)*

The CF plays a critical role in a DB2 pureScale instance because it manages the synchronization of data and locks between all DB2 pureScale members to maintain data consistency. To help achieve optimal performance in a DB2 pureScale environment, the CF must be able to efficiently handle incoming requests from members so that statement execution can continue on the members. The volume of incoming CF requests depends on various characteristics of the workload running against the database. For example, workloads with a high percentage of read activity likely make fewer CF requests for locking and reading pages. Fewer read requests are made because it is more likely that the pages were previously read and are still available in the member's local buffer pool (LBP). Fewer lock requests are required because members keep a shared lock on pages even after the completion of the transaction, meaning that future reads that require shared-lock requests will not require communication with the CF. Conversely, write-intensive workloads can result in more CF read requests. One reason is that when a member writes a page to the CF, other members with copies of the same page in their LBPs must each reread that page from the CF the next time they reference the page, because their local copies are no longer current.

The last example introduces the concept of data sharing, which is a fundamental characteristic of shared-data database systems such as the DB2 pureScale Feature. The level of data sharing in a workload directly affects the amount of communication required between members and the CF. Partitioned workloads in which the level of data sharing between members is small do not generate as much CF communication. The additional read requests in the previous example involving a write-intensive workload, which occurred as a result of one member modifying a page, are less likely to occur in a workload with very little data sharing. A write-intensive workload with high levels of data sharing between members can generate page contention between those members. When two or more members contend for a page, a *page reclaim* process occurs. This process causes the page to be taken from one member and given to another, even in the middle of a transaction. The requesting member can continue working without waiting for the holding member to commit its transaction. Although this process is heavily optimized through the use of interrupt-free RDMA, it is always a good practice to minimize communications where reasonable. Page reclaims are described further in later sections.

### **Configuration of CPU resources for the CF**

For the majority of workloads, one CF core is required for every five DB2 pureScale member cores. The CF is a multithreaded application that uses polling to minimize the overall CF response time. Requests made to the CF are handled by CF worker threads. It is highly recommended that you provide the CF with dedicated cores to avoid CPU resource contention between CF worker threads and other processes on the system that are consuming CPU resources. As a rule of thumb, on System x machines, the number of CF worker threads should be equal to the number of cores minus one.



## Allocation of memory resources for the CF

A reasonable initial CF memory allocation is 35% - 40% of the total size of all LBPs from all DB2 pureScale members. Assign this value to the `cf_db_mem_sz` database configuration parameter. Of this amount, you should assign approximately 80% to the global buffer pool (GBP), 15% to the global lock manager (GLM), and 5% to the shared communications area (SCA).

Recall that one of the benefits of the CF is that the GBP acts as a cache for modified database pages. When a page required by a member is not found in the member's LBP, the member requests the page from the CF. If the page is in the GBP, the CF will send the page to the member and a costly disk access will be avoided. In general, increasing the size of the GBP results in a higher GBP hit ratio and improved performance. To determine your GBP hit ratio, use the `MON_GET_BUFFERPOOL` table function. You can get the same information by viewing the buffer pool screen of the DBA Cockpit, which has been enhanced to include monitoring specific to the DB2 pureScale Feature, including the GBP hit ratio.

## Avoidance of page reclaims

A page reclaim occurs when one member in the DB2 pureScale instance is forced to release a page because another member requested access to the same page in a conflicting mode. Although this is a very powerful algorithm for improving concurrency across a cluster, it incurs a cost. Processing a page reclaim involves transferring the contended-for page between the members that are involved and the CF. First, the member holding the page sends the modified page to the CF and releases the lock that it held on that page. Upon receiving the updated page, the CF invalidates any copies of that page in the LBPs of other members. At this point, the member requesting the page is granted access to the page, after which it retrieves the updated page from the CF.

Although the RDMA-based communication with the CF is very fast, minimizing page contention (and thus page reclaims) is good practice for any database system, and the DB2 pureScale Feature is no exception.

## Use of client affinity with SAP solutions

You should direct all database connections from one SAP NetWeaver Application Server to one specific DB2 member. Having connections to multiple DB2 members from one SAP NetWeaver Application Server can result in page reclaims because connections from the same SAP NetWeaver Application Server are likely to access a similar set of tables. A configuration in which all connections of an SAP NetWeaver Application Server go to the same DB2 member uses *client affinity*.

Client affinity is configured through the central connectivity configuration file, named `db2dsdriver.cfg`. This file is located in the `/sapmnt/SAPSID/global/db6` global directory, which is shared by all SAP NetWeaver Application Server components. The `db2dsdriver.cfg` file contains a list of all DB2 members and all SAP NetWeaver Application Server components. For every SAP NetWeaver Application Server, DB2 automatically determines (through a round-robin algorithm) which DB2 member the SAP NetWeaver Application Server should connect to

and which other DB2 members should be selected as failover targets if the primary DB2 member fails.

The following example displays the client affinity configuration for an SAP application with six application servers and three DB2 members. The first application server in the list, `sap_as_1` on host `sapas1.wdf.sap.corp`, uses the first member in the list, `member_0` on host `db2dsf0.wdf.sap.corp`, as its primary member. Member `member_1` serves as the first failover target for `sap_as_1`, and `member_2` serves as the secondary failover target. The second application server, `sap_as_2`, uses `member_1` as its primary member, `member_2` as its first failover target, and `member_0` as its secondary failover target, and so on. Because there are more application servers in the list than DB2 members, assignment continues with the first DB2 member again. That is, SAP NetWeaver Application Server `sap_as_4` has the same affinity settings as `sap_as_1`.

```
<configuration>
  <dsncollection>
    <dsn alias="TD1" name="TD1" host="db2dsf0.wdf.sap.corp" port="5912" />
  </dsncollection>
  <databases>
    <database name="TD1" host="db2dsf0.wdf.sap.corp" port="5912">
      <acr>
        <parameter name="enableAcr" value="true"/>
        <parameter name="enableSeamlessAcr" value="true"/>
        <parameter name="affinityFailbackInterval" value="60"/>
        <alternateserverlist>
          <server name="member_0" hostname="db2dsf0.wdf.sap.corp" port="5912" />
          <server name="member_1" hostname="db2dsf1.wdf.sap.corp" port="5912" />
          <server name="member_2" hostname="db2dsf2.wdf.sap.corp" port="5912" />
        </alternateserverlist>
        <clientaffinityroundrobin>
          <client name="sap_as_1" hostname="sapas1.wdf.sap.corp"/>
          <client name="sap_as_2" hostname="sapas2.wdf.sap.corp"/>
          <client name="sap_as_3" hostname="sapas3.wdf.sap.corp"/>
          <client name="sap_as_4" hostname="sapas4.wdf.sap.corp"/>
          <client name="sap_as_5" hostname="sapas5.wdf.sap.corp"/>
          <client name="sap_as_6" hostname="sapas6.wdf.sap.corp"/>
        </clientaffinityroundrobin>
      </acr>
    </database>
  </databases>
</configuration>
```

The `db2dsdriver.cfg` file is automatically created and maintained by the `SAPinst.`

## Advanced client affinity configurations with SAP applications

In environments where you isolate specific workloads and direct them to specific application servers, you can further reduce page reclaims by assigning all SAP NetWeaver Application



Server components that execute the same workload to the same DB2 members. You can define this affinity configuration by altering the values in the `db2dsdriver.cfg` file.

The following example displays a user-defined client affinity configuration for an SAP solution with six application servers and four DB2 members. One of the DB2 members, `spare_0`, is dedicated as a spare member, providing additional computing power in case of a member failure or member maintenance.

There are three affinity groups defined, HR, SD, and finance (FI), each with distinct affinity settings. The HR group, for example, is configured to work with `member_0` as the primary member and `spare_0`, `member_1`, and `member_2` as the failover targets. That means every SAP NetWeaver Application Server belonging to the HR group normally works with `member_0`. In the absence of `member_0`, they fail over to `spare_0`. If both `member_0` and `spare_0` are down, the application servers work with `member_1`. Lastly, if `member_1` is also down, `member_2` is used. Similar definitions are provided for the SD and FI groups. Application servers `sap_as_1` and `sap_as_2` are defined as part of the HR group and share the HR group affinity settings. The other application servers are assigned to the SD and FI groups.

```
<configuration>
  <dsncollection>
    <dsn alias="TD1" name="TD1" host="db2dsf0.wdf.sap.corp" port="5912" />
  </dsncollection>
  <databases>
    <database name="TD1" host="db2dsf0.wdf.sap.corp" port="5912">
      <acr>
        <parameter name="enableAcr" value="true"/>
        <parameter name="enableSeamlessAcr" value="true"/>
        <parameter name="affinityFailbackInterval" value="60"/>
        <alternateserverlist>
          <server name="member_0" hostname="db2dsf0.wdf.sap.corp" port="5912" />
          <server name="member_1" hostname="db2dsf1.wdf.sap.corp" port="5912" />
          <server name="member_2" hostname="db2dsf2.wdf.sap.corp" port="5912" />
          <server name="spare_0" hostname="db2dsf3.wdf.sap.corp" port="5912" />
        </alternateserverlist>
        <affinitylist>
          <list name="HR" serverorder="member_0,spare_0,member_1,member_2" />
          <list name="SD" serverorder="member_1,spare_0,member_2,member_0" />
          <list name="FI" serverorder="member_2,spare_0,member_0,member_1" />
        </affinitylist>
        <clientaffinitydefined>
          <client name="sap_as_1" hostname="sapas1.wdf.sap.corp" listname="HR" />
          <client name="sap_as_2" hostname="sapas2.wdf.sap.corp" listname="HR" />
          <client name="sap_as_3" hostname="sapas3.wdf.sap.corp" listname="SD" />
          <client name="sap_as_4" hostname="sapas4.wdf.sap.corp" listname="SD" />
          <client name="sap_as_5" hostname="sapas5.wdf.sap.corp" listname="FI" />
          <client name="sap_as_6" hostname="sapas6.wdf.sap.corp" listname="FI" />
        </clientaffinitydefined>
      </acr>
    </database>
  </databases>
```

```
</configuration>
```

## Multiplexing of SAP update tables

The update tables of an SAP application (VBDATA, VBMOD, and VBHDR) have some of the characteristics described previously that make them a likely source of page reclaims. Application servers for which you configure update work processes read from and write to these tables, potentially causing contention for the rows in the tables.

Because the data written into these tables by one application server need not be visible to another application server, SAP provides, through SAP Note 52310, the ability to create a set of update tables for every application server that runs update work processes. This ability is known as *update table multiplexing* and is a highly recommended optimization that you can apply to completely eliminate page reclaims for the update tables if you configure the SAP NetWeaver Application Servers to use client affinity.

## Tables with frequent inserts

Tables that involve frequent inserts and are in table spaces with a small extent size can be a source of page reclaims. When tables are extended to accommodate more pages, the number of additional pages to grow the table by is a multiple of the table space extent size. A small extent size results in having to extend the table more frequently. This can result in page reclaims because multiple members are likely to be attempting to grow the table at the same time.

As a rule of thumb, place tables that you expect to grow quickly into a table space with a larger extent size. For example, SAP applications use table spaces with an extent size of two pages by default, and moving tables with a high insert rate into a table space with an extent size of eight pages can improve the insert performance for these tables.

In addition, to further optimize insert performance in the DB2 pureScale Feature, you can place a table into APPEND mode, just as you can without the DB2 pureScale Feature. This avoids the maintenance of free space control records (FSCRs), which can, in turn, also minimize page reclaims. The same considerations for using APPEND mode without the DB2 pureScale Feature also apply with the DB2 pureScale Feature (for example, tables where inserts are significantly more frequent than deletes).

## Monitoring of page reclaims

To monitor page reclaims, use the MON\_GET\_PAGE\_ACCESS\_INFO table function. Using this table function, you can quickly identify which tables and indexes contribute the most pages to the reclaim count and reclaim wait time in the database.

You can take advantage of the monitoring enhancements to the DBA Cockpit for the DB2 pureScale Feature to view page reclaim statistics. Specifically, the page contention screen of the DBA Cockpit provides a summary of the details available through the MON\_GET\_PAGE\_ACCESS\_INFO table function.

In a low-end system, if you observe 50 or more page reclaims per second for a table or index, consider investigating and perhaps taking further action. The threshold is larger in a mid- or high-end system. After identifying the table or index, examine the characteristics of the object to identify what is causing page reclaims to occur.

The `MON_GET_PKG_CACHE_STMT` table function can provide a statement-level gauge of the affect of page reclaims through the `reclaim_wait_time` monitor element. This monitor element represents the amount of time during statement execution that was spent waiting on page locks, where the lock request caused a page to be reclaimed. If the value of the `reclaim_wait_time` monitor element represents 50% or more of the total statement execution time, consider taking action to reduce page reclaims on the tables or indexes that are accessed during statement execution.

## *GPFS and database I/O*

Although avoiding the cost of disk accesses is a common goal for many database systems, it is a cost that must eventually be incurred. When determining how many disks you need to support your DB2 pureScale environment, it is important to keep in mind that all members might access the same disks when reading and writing data. However, log records and data should not be stored on the same set of disks. Using separate disks for data and log records is common practice and important for good performance in many applications. In a DB2 pureScale environment, an additional good practice is to give each member a dedicated set of disks for logging such that no two members will write log records to the same disks. These dedicated disks must be accessible by all members through the SAN. Based on this recommendation, create one GPFS file system per set of log disks for each member in the cluster, plus at least one GPFS file system for the database itself.

During host failure scenarios, GPFS suspends access to file systems until the failed host has been evicted from the cluster. While waiting for this eviction process to be completed, transactions requiring disk access cannot run. Therefore, minimizing the time that is required for host eviction is critical for achieving fast failure recovery times. Using the SCSI-3 Persistent Reserve (SCSI-3 PR) technology, host eviction is completed almost immediately upon failure detection. Assuming that your storage system and disks support SCSI-3 PR, you can enable it in GPFS by using the `mmchconfig` utility with the `usePersistentReserve=yes` option. There are several other best practices that you should follow when using GPFS:

- When creating GPFS file systems, use a block size of 1 MB because larger block sizes tend to use the GPFS file system cache (referred to as the *pagepool* in GPFS documentation) more efficiently. If you use the `db2cluster` command to create the file system, a block size of 1 MB is used by default.
- The default size of the pagepool is 64 MB. You should increase this to at least 256 MB. Note that the memory used for the pagepool is pinned meaning that this memory cannot be swapped out to disk if the amount of free memory on the system becomes low. Ensure that there is enough free memory on the system to support the size that you want.
- The maximum number of concurrent operations on a GPFS file system is controlled by the `worker1Threads` parameter, which is set to 40 by default. A value of 256 is recommended as a reasonable starting point for most DB2 pureScale environments. You

should tune this parameter based on your application and the number of threads (such as agents, page cleaners, and prefetchers) that you expect to perform file system operations.

## Integration and operation of the DB2 pureScale Feature with SAP applications

The DB2 pureScale Feature is tightly integrated into SAP applications. The integration covers all aspects of database usage with SAP applications: installation, simple database operation, system management, monitoring, and integration of HA for SAP central services.

The tight integration of the DB2 pureScale Feature into SAP applications allows you to perform a wide spectrum of database related tasks from SAP tools rather than with DB2 commands.

The DB2 database management system provides two very comprehensive scaling technologies: The DB2 pureScale Feature and the Database Partitioning Feature (DPF). DPF is a shared-nothing scale-out technology that has been supported for OLAP applications such as SAP NetWeaver Business Warehouse components for many years.

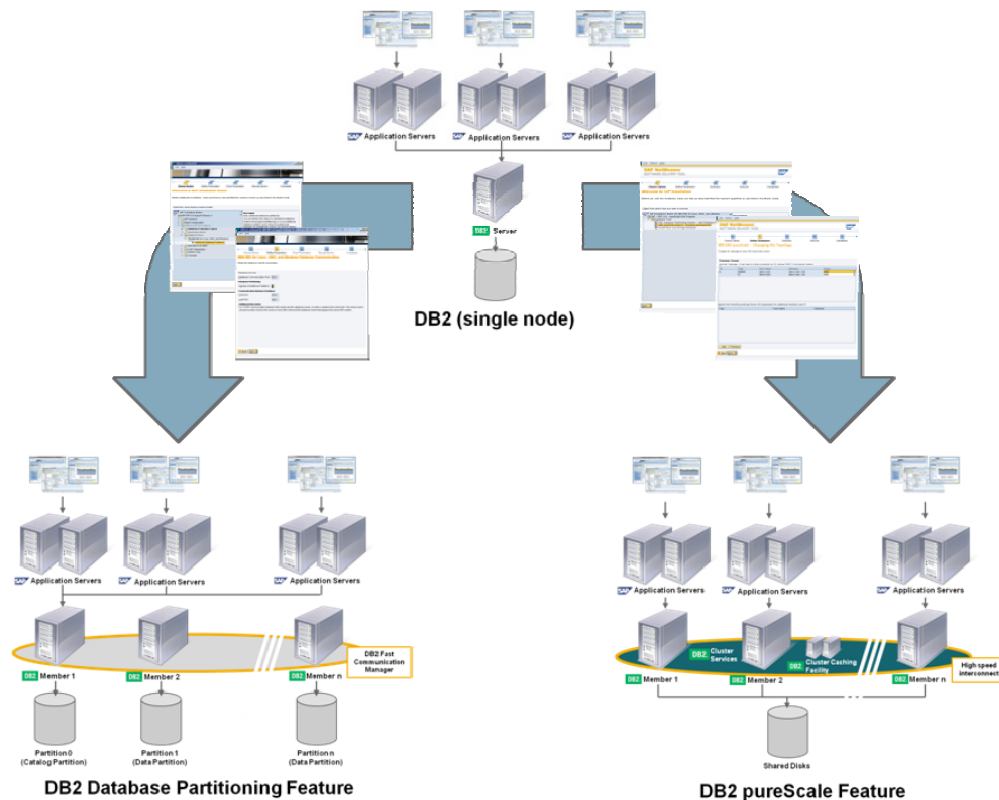


Figure 9. Managing initial setup and topology changes

## Installing an SAP application with the DB2 pureScale Feature

Most SAP customers start with a single-node database and then later decide to extend the topology of their database servers to multiple computing nodes. The philosophy for managing DB2 topology follows this paradigm. The initial SAP software installation on DB2 consists of a single-node database. After the initial installation step, you can extend the DB2 topology to a multiple node setup using SAPInst installation tool.

The DB2 pureScale Feature requires certain data, such as the table space data, the DB2 logs, and the database directory, to be in a GPFS file system. For manageability reasons, you should use DB2 commands for creating and maintaining a GPFS file system.

After creating the GPFS file systems, install the SAP application on a single-node DB2 V9.7 database by using the regular SAP software installation master. After the SAP software installation is finished on a DB2 V9.7 database, you can upgrade the database to DB2 pureScale (= DB2 V9.8), as described in the following sections.

## Upgrading an SAP application on DB2 V9.7 to DB2 V9.8

If you have an SAP application that does not use GPFS file systems, you must move the existing data to GPFS before upgrading the DB2 database to DB2 V9.8. For manageability reasons, you should always use DB2 commands for creating and maintaining a GPFS file system. This ensures that your GPFS file system is owned and maintained by DB2. To create and set up GPFS, install the DB2 V9.8 software on your database server. After completing the DB2 installation, use the **db2cluster** command to create the file systems. You must create file systems for the sapdata, log, db2dump, and database directory. After creating the new DB2 file systems, move the existing data to GPFS. You can use the **rebalance** command to move table space data without downtime. For a detailed description of the procedure, see the *SAP Upgrade Guide for DB2 pureScale*.

After you have moved all table space data to GPFS, stop the SAP NetWeaver Application Server components and the DB2 instance, move the remaining files, logs, db2dump and database directory files to GPFS, upgrade the database to DB2 V9.8, and then start the SAP software installation program SAPInst to change the topology of your DB2 instance. For this task SAPInst has a new installation service, "DB2 pureScale Feature – Topology Management," that upgrades your current database to a DB2 pureScale database and extends it to your target topology.

## Using SAPInst for the DB2 pureScale Feature topology changes

You can use SAPInst to change the topology of the DB2 pureScale cluster, for example, to add or remove DB2 members or CFs. The SAPInst software installation service prompts you for information about your SAP NetWeaver Application Server components and then displays two screens that you can use to define and modify the topology of your DB2 pureScale cluster. Figure 10 illustrates the SAPInst screen for DB2 pureScale topology changes.

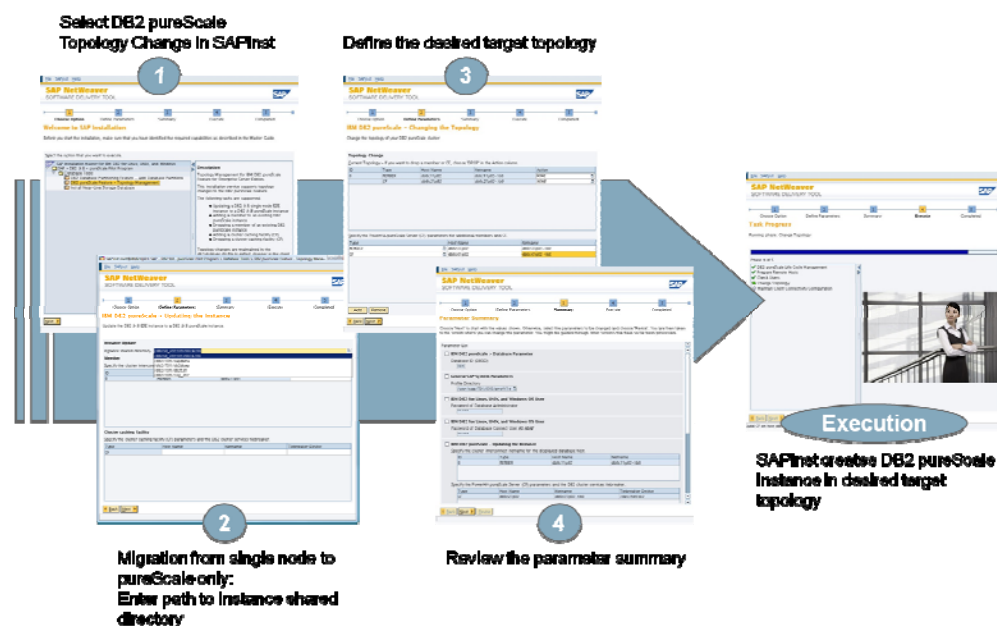


Figure 10. Making topology changes

During the execution phase of SAPinst, all required execution steps on remote hosts are done automatically. The SAP software installation master is copied to all new hosts that will join the cluster, and the installation runs silently on the new hosts. You do not have to execute SAPinst on any host other than the current database host.

When you use SAPinst to upgrade your DB2 single node instance to a DB2 pureScale instance, SAPinst retrieves connectivity information from the `db2cli.ini` file and replaces the `db2cli.ini` with a `db2dsdriver.cfg` file containing the round-robin client affinity information. All members are added to the alternative server list, and all SAP NetWeaver Application Server components are added to the round-robin client affinity list.

If you already have a DB2 pureScale cluster, you can use SAPinst to add or remove members or CFs as needed to adjust your current topology to your target topology. You can make multiple changes in one SAPinst run if you keep at least one member and one CF. If you want to replace all database hosts, you must run SAPinst twice and split the topology changes such that at least one member and CF are kept in each run.

SAPinst maintains the alternative server list section of the `db2dsdriver.cfg` file during all topology changes. If you use the default round-robin connectivity, you do not have to make any further changes: all new members are available to the application servers. If you use user-defined client affinity, you can edit the `db2dsdriver.cfg` file after the SAPinst run and manually define the affinity that you want.

After SAPinst has finished, your SAP NetWeaver Application Server components are ready to use the DB2 pureScale cluster.



## Changing the SAP NetWeaver Application Server topology

Changes to the topology of the SAP NetWeaver Application Server components also require updates to the central `db2dsdriver.cfg` client affinity configuration file. Consequently, `SAPinst` has been extended to maintain the central `db2dsdriver.cfg` file during installation of additional SAP NetWeaver Application Server components. The installation of a new SAP NetWeaver Application Server adds the new server to the client round-robin section of the `db2dsdriver.cfg` file. The new SAP NetWeaver Application Server can then take full advantage of the high availability of your DB2 pureScale instance.

If you want to use user-defined affinity, you can modify the configuration manually by removing the new SAP NetWeaver Application Server from the round-robin client list and adding it to the user-defined affinity list after the application server installation.

## Setting up SAP Central Services components for HA

HA is a system-wide strategy that is not limited to the database server. HA for other components of your SAP application, such as the SAP central services, is an important requirement. SAP central services are the key services of an SAP application. They are essential for the SAP application to run correctly and without HA they can be a single point of failure.

The license that SAP provides for the DB2 pureScale Feature includes a full license of TSA. This provides an HA setup for your complete SAP application, including the central services and the

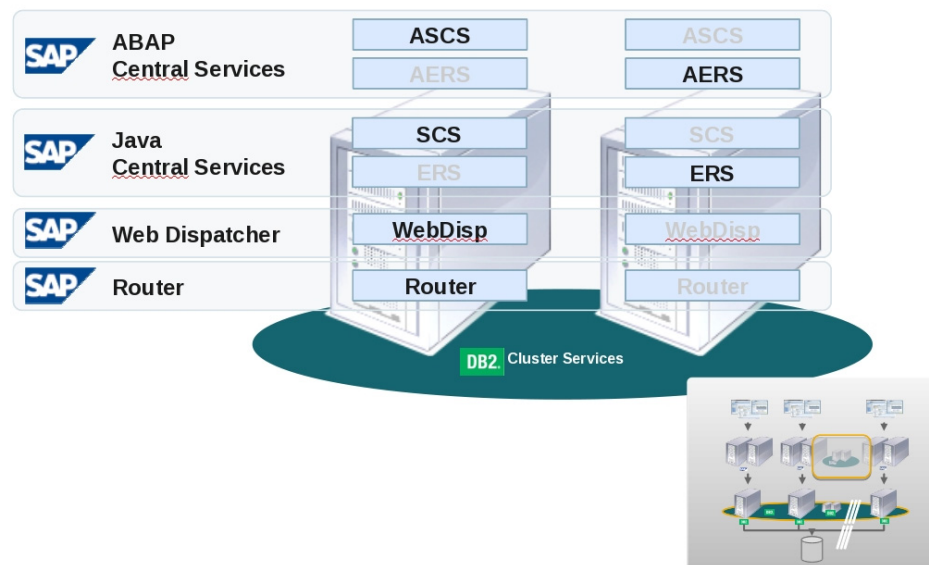


Figure 11. SAP Central Services HA



database, eliminating any single point of failure for your SAP application.

The ABAP™ central services and Java central services consist of the following components:

- **Enqueue server and enqueue replication server**  
The SAP application has its own locking mechanism at the business transaction level to synchronize data access. If the enqueue server fails, the SAP application can no longer acquire locks and therefore, the SAP application hangs. SAP provides a mechanism for replicating the lock table to another host. This mechanism is the enqueue replication server (ERS).
- **Message server**  
Only one message server can run in an SAP application. The message server handles, among other things, central communication between SAP NetWeaver Application Server components and load distribution of logons using logon groups for the SAPGUI or RFC connections. A failure of the message server results in unavailability of the SAP application.
- **Gateway**  
The SAP gateway carries out RFC services to allow SAP applications and external programs to communicate with each other. If the gateway fails, no RFC communication or execution of registered programs is possible.

You can use the `sapscscluster.sh` tool, which is provided by SAP, to help set up HA for the SAP central services.

You must separate the SAP central services from the SAP dialog instances in an HA setup. For a new installation, you should install SAP central services separately from other SAP services by using SAPinst High-Availability System installation services. For an existing system, you can use the `sapscscluster.sh` tool to separate the Central Services of your existing SAP application to avoid a system copy. After the Central Services are separated, the `sapscscluster.sh` tool automatically creates all TSA resources and the cluster environment. This tool substantially simplifies the cluster setup process and provides you with SAP HA for the SAP application.

After you make the central services highly available, you can also make the SAP NetWeaver Application Server components highly available. But in general there is more than one SAP NetWeaver Application Server connected to the DB2 pureScale cluster. For that reason SAP NetWeaver Application Server components are not a single point of failure in SAP applications and HA solutions for single SAP NetWeaver Application Server are not needed in the majority of cases.

In the SAP solution landscape, you could also have the following optional components. If you use these, you must protect them in an HA setup as well.

- **SAP Web Dispatcher**  
The SAP Web Dispatcher is the entry point for HTTP requests into your SAP application. If the Web Dispatcher fails, no HTTP requests can be sent to your SAP application.
- **SAP Router**  
SAP Router acts as an intermediate station (proxy) for connections between SAP

applications and external networks. If SAP Router fails, connections cannot be established.

- **Central NFS Share**

The `/sapmnt/SID` shared directory contains the binary files and configuration data for the application servers. If this directory is not available, SAP NetWeaver Application Servers cannot be started. You can protect the NFS shared directory by using TSA or an NAS storage server.

## *Monitoring the DB2 pureScale Feature Using the DBA Cockpit*

The DBA Cockpit is the central point for administering and monitoring DB2, regardless of whether it uses the DB2 pureScale Feature. The DBA Cockpit has been enhanced to support new monitoring elements for the DB2 pureScale Feature. When deploying an SAP application on the DB2 pureScale Feature, you should import the DBA Cockpit enhancements, as described in SAP Note 1409540.

The new monitoring elements are shown on the following screens:

- **Cluster Status**

The Cluster Status screen presents the status of all DB2 members and CFs, including current states and alerts. The current and home hosts for members are also displayed. You can use this information to help understand whether a member is performing a restart light on a different host because of a host failure.

- **Cluster Performance**

The Cluster Performance screen provides an overview of performance metrics for all members. You can drill down for more information about a member by double-clicking the member in the overview table. The screen shows wait time for CFs, global lock wait time, total reclaim wait time, and GBP and LBP hit ratios.

- **Buffer Pools**

The Buffer Pools screen displays hit ratios for the GBP and LBPs. You can see more detailed information for a member by double-clicking the entry of the member.

- **Package Cache**

The Package Cache screen provides additional DB2 pureScale metrics at the statement level.

- **CF**

The CF screen shows detailed information about the size and consumption of the startup memory, the GBP, the global lock memory, and the SCA memory. For each host, there is also information about resources at the operating-system level.

- **Page Contention**

The Page Contention screen shows information about page reclaims in the DB2 pureScale cluster. You can drill down to get more detailed information about each table.

- **Client Affinity**

The Client Affinity screen shows which application servers are connected to which member. You can also display the contents of the `db2dsdriver.cfg` file to see the definition of the client affinity for your SAP application and reload the configuration file

for all DB2 clients on the SAP NetWeaver Application Server components to make topology changes effective without a restart.

Figure 12 shows some sample screen captures of the Cockpit enhancements:

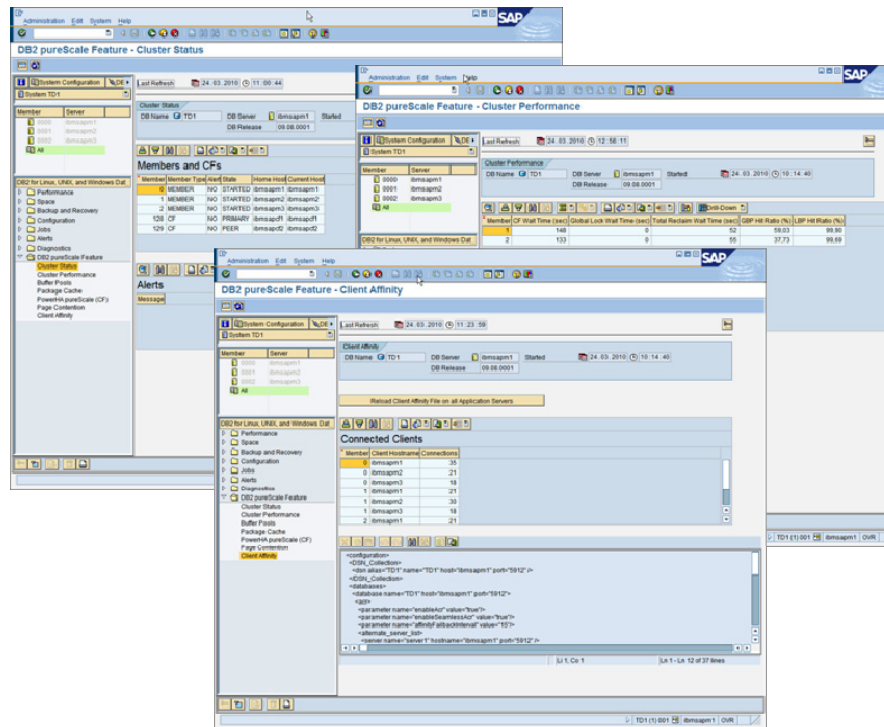


Figure 12. Screen captures of Cockpit enhancements

## *Application-transparent maintenance of a DB2 member or CF*

The DB2 pureScale Feature helps minimize the outages for planned hardware or operating system maintenance.

To perform the maintenance on a host belonging to the DB2 pureScale cluster, first move the work off that host. For a DB2 member, use the DB2 **QUIESCE** command to force work off. Current work continues until the transaction ends and then transparently fails over to the configured failover member. To move the work to a different member other than the configured failover member, update the client affinity file before issuing the DB2 **QUIESCE** command, and trigger a configuration reload in the DBA Cockpit. For example, you might want to temporarily define a spare member as a failover target.

When there are no active connections on the member, the member is taken offline. Then you can take the member host out of the cluster if you want and perform maintenance. After the

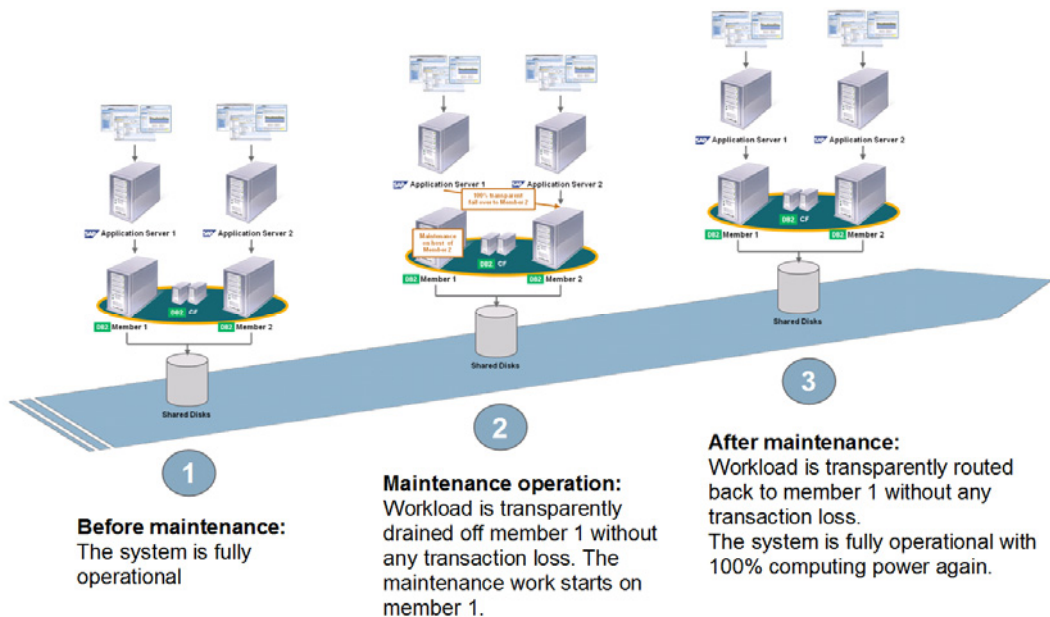


Figure 13. Maintenance of a member in a DB2 pureScale cluster

maintenance is complete, the host and the DB2 member can be reintegrated into the cluster. At the next transaction boundary, the SAP NetWeaver Application Server components automatically route work back to their primary member according to the client affinity configuration.

Maintenance on the CF works in a very similar way. First, stop the CF using the command `db2stop <cf_id>`. If you stop the secondary CF, the primary CF continues its work unaffected. If you stop the primary CF, the secondary CF takes over the primary role, transparent to the application, provided that the CF is in peer state. Depending on the planned maintenance, you can then put the host into maintenance mode. After you complete the maintenance, you can bring up the CF, which reintegrates transparently as the secondary CF.

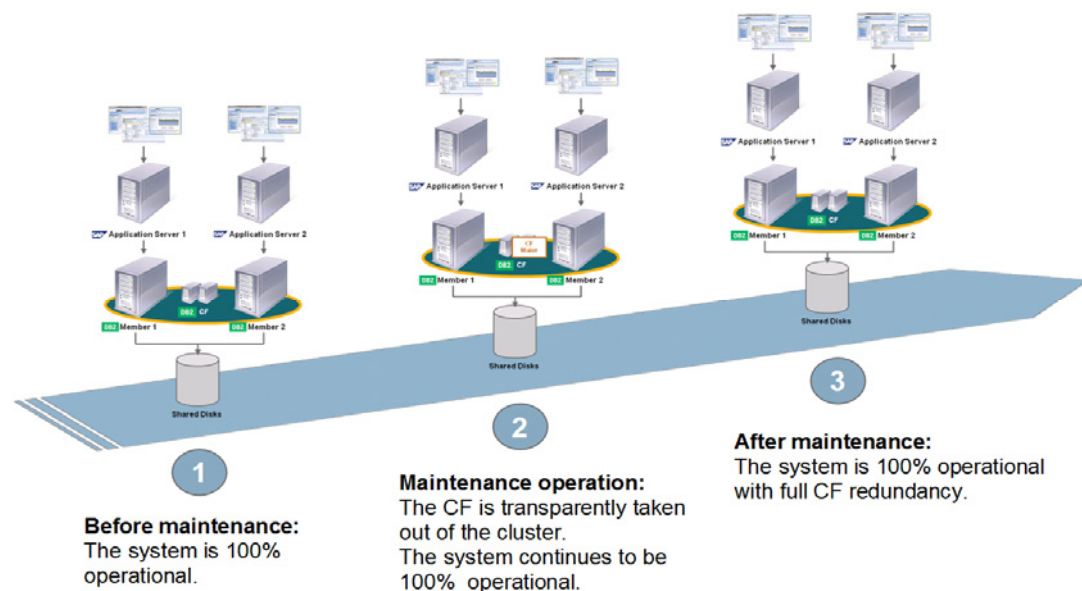


Figure 14. Maintenance of the CF

## Conclusion

Scale out or scale up? This is one of the most important questions that enterprise architects have had to answer since client/server architectures have been implemented in enterprise data centers. Scale out means providing more capacity by adding more servers, storage, and network components. Scale up usually means using bigger hardware with more CPUs, memory, network, and storage components. Both approaches have their advantages and disadvantages. Scale-out architectures work very well with modern x86 servers, in particular, with blade systems. The obvious disadvantage is that there are more physical servers to manage. Another challenge with scale-out architectures is concurrent access to data, in particular, concurrent write access. As a result, mixed two-tier environments are common, in which application servers are implemented on top of a scale-out infrastructure and the backend system, which is a database, is deployed on a large, scaled-up system.

This paper describes in depth how the DB2 pureScale Feature solves the problem of concurrent data access while providing the advantages of a scale-out architecture. The most obvious advantage is scalability itself. You can easily expand database capacity without a service interruption. In a competitive economy, this is an important requirement.

Tight integration with SAP applications minimizes the efforts of administering a DB2 pureScale cluster. You can perform many administrative tasks with SAP tools, without knowledge of DB2 commands. The integration covers all aspects of database usage with SAP applications: installation, simple database operation, system management, monitoring, and integration of HA for SAP central services.

With its scale-out architecture, the DB2 pureScale Feature provides redundancy and HA, including a complete HA solution for SAP applications. Except in a single-server solution, the failure of one component or member does not cause a service outage.

The DB2 pureScale Feature supports SUSE Linux Enterprise Server from Novell, a common platform for SAP software development. SAP trusts SUSE Linux Enterprise Server as an operating system for many SAP solutions, some of them exclusively available on SUSE Linux Enterprise Server. In addition, this Linux operating system leverages the latest x86 server technology from IBM. As described in this paper, you can implement small, midsize and very large SAP solutions with the DB2 pureScale Feature on the Linux operating system and x86 hardware and benefit from the price/performance advantage of that infrastructure.

The combination of SAP business solutions and the DB2 pureScale database product on SUSE Linux Enterprise Server is a best-of-breed solution that provides highest reliability and scalability at an very attractive price/performance advantage. Also, through the SAP Customer Support Network, IBM, Novell, and SAP jointly provide first-class enterprise support for the combined solution of SAP applications and the DB2 pureScale Feature on SUSE Linux Enterprise Server.

## Glossary

CF	Cluster caching facility
DPF	Database Partitioning Feature
DR	Disaster recovery
ERS	Enqueue replication server
FC	Fibre Channel
GbE	Gigabyte Ethernet
GBP	Group buffer pool
GPFS	IBM General Parallel File System
HA	High Availability
LAN	Local area network
LBP	Local buffer pool
MAX5	IBM System x memory expansion technology
NTS	Novell Technical Service
OFED	Open Fabrics Enterprise Distribution
PTF	Program Temporary Fix
QPI	QuickPath Interconnect
RAS	Reliability, availability, scalability
RDAC	Redundant Disk Array Controller
RDMA	Remote Direct Memory Access
RSCT	IBM Reliable Scalable Cluster Technology
SAN	Storage area network
SCS	SAP Central Services
SLES	SUSE Linux Enterprise Server
TSA	IBM Tivoli System Automation





## Further reading

[1] DB2 Best Practices are available at

<http://www.ibm.com/developerworks/db2/bestpractices/>

[2] Installation prerequisites for DB2 pureScale Feature (Linux) can be found in the *IBM DB2 pureScale Feature Information Center* at

<http://publib.boulder.ibm.com/infocenter/db2luw/v9r8/topic/com.ibm.db2.luw.s.d.doc/doc/r0057441.html>

[3] Details about the IBM System x server x3650 M3 can be found at

<http://www-03.ibm.com/systems/x/hardware/rack/x3650m3/>

or in the Solution Brief *SAP on System x3650 M3* at

<http://www.ibm.com/common/ssi/fcgi-bin/ssialias?htmlfid=XSS03018USEN>.

[4] For details on IBM eX5 technology, see

<http://www.ibm.com/systems/info/x86servers/ex5>

or read the Solution Brief *Virtualize, consolidate and scale SAP with IBM System x and eX5*, which can be downloaded from

<http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=XSS03017USEN>.

More technical details on eX5 servers can be found in the IBM Redpaper *IBM eX5 Portfolio Overview: IBM System x3850 X5, x3950 X5, x3690 X5, and BladeCenter HX5*, REDP-4650-03, October 2010, at

<http://www.redbooks.ibm.com/redpapers/pdfs/redp4650.pdf>,

in the *IBM System x3850 X5 / x3950 X5 - Product Guide*, April 2011, at

<http://public.dhe.ibm.com/common/ssi/ecm/en/xso03099usen/XSO03099USEN.PDF>,

or in the *IBM System x3690 X5 - Product Guide*, April 2011, at

<http://public.dhe.ibm.com/common/ssi/ecm/en/xso03120usen/XSO03120USEN.PDF>.

[5] SAPS is a hardware-independent unit of measurement describing performance of system configurations in SAP environments. SAPS requirements result from SAP's sizing tool *QuickSizer* and are starting points for the configuration and sizing of server systems. Details on *SAP standard application benchmarks* are available at

<http://www.sap.com/solutions/benchmark/index.epx>

and through the home page of *SAP Quicksizer* at

<http://service.sap.com/quicksizer>.

- [6] More information on IBM System Storage is available in the Internet at

<http://www.ibm.com/systems/storage/>

- [7] "Shared storage considerations for DB2 pureScale environments" are available in the "IBM DB2 pureScale Feature Information Center" at

<http://publib.boulder.ibm.com/infocenter/db2luw/v9r8/topic/com.ibm.db2.luw.sd.doc/doc/c0059360.html>

That web page gives an overview of validated shared storage device and multipath I/O driver combinations in the categories 1 and 2. (Serge: please replace "... " quotes with italic font here).

- [8] The whitepaper *IBM System x powered by SUSE Linux Enterprise Server for SAP applications* can be downloaded from Novell and IBM web sites. For example, see *IBM System x and BladeCenter solutions for SAP solution environments* at

<http://www-03.ibm.com/systems/x/solutions/infrastructure/erpcrm/sap.html>

or *Novell SUSE Linux Enterprise Applications* at

<http://www.novell.com/products/sles-for-sap.html>,

- [9] See SAP Note 1557506 *Linux paging improvements* at

<https://service.sap.com/sap/support/notes/1557506>

- [10] See *SAP on Linux in General FAQ* at

<http://www.sdn.sap.com/irj/sdn/index?rid=/webcontent/uuid/509a0be3-b280-2a10-74bb-b5a6f0484417>.

- [11] See SAP Note 1492000 *General Support Statement for Virtual Environments* at

<https://service.sap.com/sap/support/notes/1492000>.

## *Authors*

Torsten Ziegler

*Development Manager, SAP NetWeaver on IBM DB2  
for Linux, UNIX, and Windows*

*SAP*

Peter Schinagl

*Technical ISV Relationship Manager  
SUSE Linux Products, Novell*

Andreas Nemeth

*Technical Director SAP  
Novell SAP Global Alliance*

Matthew Huras

*Distinguished Engineer*

*DB2 for Linux, UNIX, and Windows Development  
IBM Canada*

Steve Rees

*DB2 for Linux, UNIX, and Windows Performance  
IBM Canada*

Kostas Rakopoulos

*DB2 Performance Quality Assurance  
IBM Canada*

Frank M. Neumann

*IT Specialist System x  
IBM Infrastructure SAP Solutions  
IBM Germany*

Karl Fleckenstein  
*Senior Technical Staff Member*  
*Integration IBM Data Server solutions into SAP*  
*IBM Germany*

Thomas Drescher  
*SAP DB2 Installation Tools - SAP DB2 Development*  
*IBM Germany*

## ***Contributors***

Serge Boivin  
*Senior Writer, DB2 Information Development*  
*IBM Canada*

Paul Dermody  
*Writer, DB2 Information Development*  
*IBM Ireland*

Maik Gasterstaedt  
*SAP Technical Sales Enablement for IBM Systems*  
*Storage - Global Alliance Solutions*  
*IBM Germany*

Holger Hellmuth  
*Technical Marketing and Sales Enablement: IBM DB2*  
*optimized for SAP software*  
*IBM Germany*

Liwen Yeow  
*Software Sales Canada SWG IM*  
*IBM Canada*

## Notices

This information was developed for 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. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

**The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:** INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

Without limiting the above disclaimers, IBM provides no representations or warranties regarding the accuracy, reliability or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The use of this information or the implementation of any recommendations or techniques herein is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Anyone attempting to adapt these techniques to their own environment do so at their own risk.

This document and the information contained herein may be used solely in connection with the IBM products discussed in this document.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. 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.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE: © Copyright IBM Corporation 2011. All Rights Reserved.

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

## **Trademarks**

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)

Windows is a trademark of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.