Configure and tune Epic ODB on IBM Power

Best practices guide for successfully implementing Epic applications on IBM AIX

Table of contents

Introduction 3
A general description of the Epic product 3
Configuring IBM Power and AIX 5
Additional recommendations 18
Summary 26
Get more information 26
About the author 26

Introduction

Epic is a Healthcare Information System (HIS) provider that develops and delivers a comprehensive electronic medical record (EMR) system covering all aspects of the medical healthcare profession. The Epic solution includes a variety of applications that cover areas, such as medical billing, emergency room, radiology, outpatient, inpatient, and ambulatory care. Epic is a privately held company founded in 1979 and based in Verona, Wisconsin.

The Epic production uses an electronic database management system from InterSystems Corporation called IRIS (replacing and extending the previous application, Caché). IRIS is a multi-modal, post-relational database with unique characteristics, and these are described in this paper. The Caché ObjectScript language is a modern implementation of M (formerly MUMPS), a language originally created for healthcare applications, and is used to program applications and code to manage and manipulate the IRIS database.

Epic's Operational Database (ODB) is known as Chronicles and interfaces with IRIS as the database engine for all their core applications. Epic's analytics DB runs on a relational database, either MS-SQL or Oracle. The analytics DB has the highest bandwidth requirements, but the IRIS ODB is by far the most critical to end user performance, and consequently is where most attention needs to be focused. This Best Practices guide is therefore centered on the Epic ODB.

A general description of the Epic product

Epic uses the following two fundamental architecture models:

- Single-server symmetric multiprocessing (SMP)
- Multi-server Enterprise Cache Protocol (ECP)

Most Epic customers are using the SMP architecture. Each architecture has a production database server that is clustered in an active-passive configuration to a failover server.

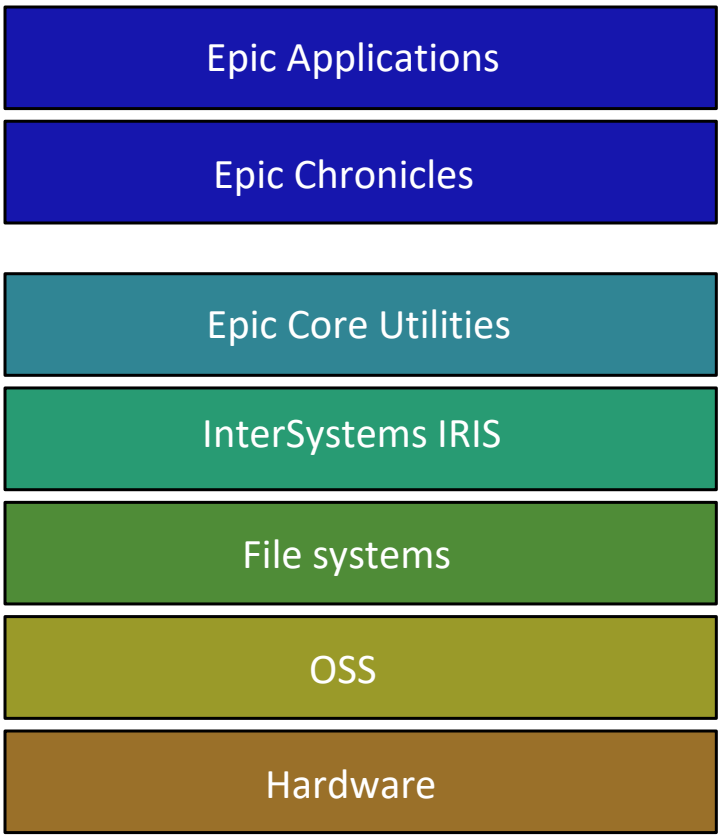

Figure 1: Functional layers of the Epic architecture

SMP database server

The single database server architecture provides a greater ease of administration. The SMP model today scales well up to 28.9 million global references (GREFs) at 90 IBM® Power® E1080 cores, and up to 29 million GREFs at 96 Power 1050 processor cores, or up to 23 million GREFs at 120 Power E980 processor cores. For this workload the metric to define scalability is Global references per second (GREFs) representing the IRIS metric for database references per second. Beyond this point, the ECP model is required.

Tiered database architecture utilizing IRIS ECP technology

The tiered architecture retains a central database server with a single data storage repository. Unlike the SMP architecture, some processing needs are offloaded to application servers, as well as utility servers. The application servers contain no permanent data. This architecture offers increased scaling over the SMP architecture.

Configuring IBM Power and AIX

For IBM Power and AIX to provide optimal performance when running Epic software, a specific set of changes to the default system tunables is required. These system parameters have been tested with core and non-core Epic software components for many different scenarios.

Mounting with concurrent I/O

The primary change to a default AIX system is invoking the use of Concurrent I/O (CIO). By default, AIX uses Enhanced Journal File System (JFS2). CIO bypasses the caching features which are enabled within JFS2. The principal reason for disabling JFS2 cache is that the IRIS DB application is already caching the required data blocks. IRIS determines what data needs to be written to permanent storage and what data should remain in the IRIS global buffers. Having the JFS2 cache also making this determination will typically cause unnecessary extra work to be performed by the system. In addition, the JFS2 cache requires real memory which could otherwise be used by the IRIS global buffer.

CIO is invoked using the `–o cio mount` option. This option should be used on the database file systems - typically /epic/prd01 – /epic/prd08, the primary and alternate journal file systems, and the IRIS Write Image Journal file location (default directory for the WIJ file is <install_directory>/mgr. This can be changed in the Journal Settings screen in the management GUI).

Creation of volume groups, logical volumes, and file systems for use by IRIS

Following are the steps necessary to create and mount the volumes which will host the Epic data volumes. It is assumed that the storage logical unit numbers (LUNs) which correspond to the volumes have already been created either through the storage system or the IBM SAN Volume Controller (SVC) if available.

Step 1: Make a top-level root directory for the Epic/IRIS.

Example:

```
mkdir /epic
```

```
mkdir /epic/prd01
```

Step 2: Create the volume groups.

Example:

```
mkvg -S -y epicprvg -s 16 hdisk1 hdisk2 hdisk3 .....
```

Step 3: Create the logical volumes.

Example:

```
mklv -a e -b n –e x –t jfs2 -y prdlv01    epicprvg  10G
hdisk1 hdisk2 hdisk3 .....
mklv -a e -b n –e x –t jfs2 -y prdlv02    epicprvg  10G
hdisk2 hdisk3 hdisk4 .....
```

Step 4: Create the file systems.

Example:

```
crfs -v jfs2 -d prdlv01 -m /epic/prd01 -A yes -a
logname=INLINE –a options=cio
```

Step 5: Mount the file systems.

Example:

```
mount /epic/prd
mount /epic/prd01
```

Step 6: Check that the appropriate entries and options are added to /etc/filesystems.

These steps should be repeated for the eight production volumes, and the WIJ and journal files. The WIJ file should share the same LUNs as the production volumes. The journal files should utilize a separate set of LUNs under a separate volume group. When the volumes are mounted, the results from the mount command, the df command, and the path command should resemble the following output:

```
# df /epic/prd0*

Filesystem    1024-blocks      Free %Used    Iused %Iused Mounted on
/dev/prdlv01     78577664   1995268   98%       12     1% /epic/prd01
/dev/prdlv02     78577664   1995268   98%       12     1% /epic/prd02
/dev/prdlv03     78577664   1995260   98%       12     1% /epic/prd03
/dev/prdlv04     78577664   1995264   98%       12     1% /epic/prd04
/dev/prdlv05     78577664   1995292   98%       12     1% /epic/prd05
```

```
/dev/prdlv06      78577664   1995280   98%        12     1% /epic/prd06
/dev/prdlv07      78577664   1995376   98%        12     1% /epic/prd07
/dev/prdlv08      78577664   1985984   98%        12     1% /epic/prd08
```

Additional system settings

Following are the recommended changes to a subset of the AIX system tunable parameters as of November 1, 2022. These settings have been tested with the following recommended minimum versions for Epic applications:

- AIX 7.2 TL4 SP3
- AIX 7.3 TL0 SP2

Kernel parameters for all AIX 7 versions

Parameter	Description and validation command	Recommended setting
AIX license limit	Maximum AIX license usage. Check with: `lslicense`	32767 (maximum possible)
ATTnum of pty0	Maximum number of pseudo terminals. Check with: `lsattr -El pty0 | grep ATTnum | awk '{print $1 " = " $2}'`	256, or number of expected concurrent OS login operations, whichever is larger. Set with: `chdev -l pty0 -a ATTnum=256`

Core dump location	Path for process cores dumps.
	Check with:
	`lscore -d`
	This should display:
	`compression: off`
	`path specification: on`
	`corefile location: /epic_core`
	`naming specification: on`
		Core dump settings can be changed with the following command, which will take effect on reboot:
		chcore -p on -l /epic_core/ -n on -d
		/epic_core/ should be a dedicated rootvg file system with 777 permissions.

Parameter	Description and validation command	Recommended setting

`$CORE_NOSHM`	Controls whether shared memory is included in a core dump.
	`grep CORE_NOSHM /etc/environment`
	`echo $CORE_NOSHM`
		True

Disable CPU folding for micro-partitioned LPARs	The use of micro-partitioning is not recommended due to the importance of the production logical partitions (LPARs) and its workload being a poor fit for micro-partitioning. If you choose to use micro-partitioning against this recommendation, it is better to disable CPU folding to avoid performance issues. Check with: `schedo -o vpm_fold_policy`	`schedo -p -o vpm_fold_policy=4`
Disable TCP traffic regulation (tcptr)	`tcptr` should be disabled. It puts restrictive limits on the number of TCP connections for common port ranges, which limits the number of Hyperspace sessions that can be connected. Note that some elevated security profiles enable `tcptr`. The test team recommends against using elevated security profiles. Check if enabled with: `no -o tcptr_enable`	`tcptr_enable` should be set to 0 with: `no -po tcptr_enable=0`
Dynamic tracking and fast fail	Enable dynamic tracking and fast fail to reduce the impact of path failure. It should be made on all fscsi devices.	Set `fc_err_recov` to `fast_fail` and set `dyntrk` to `yes`. This can be set with:

Parameter	Description and validation command	Recommended setting
	Check with: `for DEV in $(lsdev -l fscsi* | awk '{ print $1 }'); do echo $DEV; lsattr -El $DEV | grep -E "fc_err_recov|dyntrk"; done`	`for DEV in $(lsdev -l fscsi* | awk '{ print $1 }'); do chdev -P -l $DEV -a 'dyntrk=yes fc_err_recov=fast_fail'; done`
`lgpg_regions` and `lgpg_size`	These settings define the number and size of large pages on AIX. Only 16 MB large pages should be used for IRIS. Check with: `vmo -o lgpg_regions -o lgpg_size`	Login to Epic UserWeb and see Large Pages and Shared Memory Sizing for IRIS for sizing your server's number of large pages. Set with `vmo -r -o lgpg_regions=<number of large pages> -o lgpg_size=16777216` This requires `bosboot` and a reboot to set. If you need to make large page adjustments on an HA cluster, remember to make the same adjustments on both sides of the cluster. Do not set `lgpg_regions` to a value larger than the amount of memory allocated to the server, doing so will cause the server to refuse to boot. Log in to Epic UserWeb and see Updating Your Large Page Settings for more information.
`maxuproc`	Maximum number of processes allowed per user.	This should be set to `100000`. Set with:

Parameter	Description and validation command	Recommended setting
	`lsattr -El sys0 -a maxuproc`	`chdev -l sys0 -a maxuproc=100000`
Microcode and firmware level	For IBM Power10 processor-based servers, a certain microcode level is required to benefit from the AIX improvements to the Epic job-off process. Check with: `lsmcode -c`	Firmware should be FW1010.20 or higher version.
minperm percentage	Target minimum percentage of memory used for non-computational pages. Check with: `vmstat -v`	1% Set with: `vmo -p -o minperm%=1`
`num_locks_per_semid`	Value of `num_locks_per_semid`. Check with: `vmo -o num_locks_per_semid`	The CUR and BOOT values should be set to 64. Set with: `vmo -p -o num_locks_per_semid=64`
Paging space	Disk space reserved for memory in case free physical memory is low. `lsps -a`	Greater than or equal to 8 GB for a system with less than 128 GB total memory. Greater than or equal to 16 GB for a system with 128 GB or more total memory. Greater than or equal to 32 GB for a system with 1024 GB or more total memory. Set with: `chps -s <number of physical partitions to add to paging`

Parameter	Description and validation command	Recommended setting
		`space> <paging space logical volume>`
`pgz_lpgrow`, `pgz_mode`, and `pgz_num_workers`	These values are related to the use of large pages for shared memory segment. Setting the recommended values speeds up allocation and deallocation of large pages. Check with: `vmo -o pgz_lpgrow -o pgz_mode -o pgz_num_workers`	`pgz_lpgrow` and `pgz_mode` should both be 3. `pgz_num_workers` should be 12. `Set with: vmo -p -o pgz_lpgrow=3 -o pgz_mode=3 -o pgz_num_workers=12`
Power saving mode	All IBM Power8® or later processor-based servers except for the IBM Power E850 should use Fixed Maximum Frequency, introduced with HMC Version 8.820. For servers that are older than Power8, it is recommended to set the power saver mode to *disabled*. Related IBM documentation: • Power8 EnergyScale Whitepaper • Power9 EnergyScale Introduction wiki • Power9 EnergyScale - Configuration and Management wiki	In HMC power management: • Set to enable Fixed Max Frequency (can show as "Maximum Performance Mode" for IBM Power9™ processor-based systems) • In ASM power management: • Set to enable fixed maximum frequency (can show as "Maximum Performance Mode" for Power9 processor-based systems) • Disable idle power saver or list as "Feature is not supported on this system"
Restrict trace command	Prevents non-root users from running the trace command. Unintentional use of the trace command can cause outages. Check with `trcctl -l` and confirm that "Restrict non-privileged	Set with: `trcctl -R enable`

Parameter	Description and validation command	Recommended setting
	users from using trace Channel-0" is set to `enabled`. This setting defaults to enabled on AIX 7.3.	

`rfc1323`	Enable TCP window scaling. This needs to be set at both the system and interface level.
	Check with:
	`no -o rfc1323`
	or
	`ifconfig -a | grep rfc1323`
		`no -p -o rfc1323=1`
		`chdev -l <interface> -a rfc1323=1`

`sack`	Enable TCP selective acknowledgement. Check with: `no -o sack`	`no -p -o sack=1`
`sb_max`	Increase the maximum TCP buffer size; particularly useful for mirroring journal file transfer with latency. Check with: `no -o sb_max`	`no -p -o sb_max=33554432`
Set SMT to the appropriate value for your platform	For optimal performance, Power9 processor-based servers should use SMT8, except for Power E980 servers with more than 80 cores, which should use SMT4. Power10 processor-based servers should use SMT4, except for Power E1080 servers, which should use SMT8. Check with: `smtctl`	Set to SMT8 with: `smtctl -t 8 -w boot; bosboot -a` (must then reboot server) Set to SMT4 with: `smtctl -t 4 -w boot; bosboot -a` (must then reboot server)

Parameter	Description and validation command	Recommended setting

System environment variable TZ	Time zone setting.
	Check with:
	`grep -i tz /etc/environment`
	or
	`echo $TZ`
		An Olson format, such as America/New_York, should be used.
		Set with:
		Edit /etc/environment with correct timezone (TZ=<*timezone*>)

TCP idle timeout and TCP keep alive interval	Controls how long inactive TCP sessions are kept alive. Reducing this value can alleviate issues caused by network connections keeping records locked. It is recommended to set this for each ODB server that has Hyperspace sessions connecting to it. Value is in 0.5 second units.
	Check with:
	`no -o tcp_keepidle -o tcp_keepintvl`
		Set with:
		`no -p -o tcp_keepidle=1200`
		`no -p -o tcp_keepintvl=120`

`tcp_nodelayack`	Send immediate acknowledgment packets, which can help with high volume interfaces. Check with: `no -o tcp_nodelayack`	Use `tcp_nodelayack=1` if you need it for particular third-party systems. Set with: `no -p -o tcp_nodelayack=1`

`tcp_recvspace` and `tcp_sendspace`	Default TCP send and receive sizes.
	Check with:
	`no -o tcp_recvspace -o tcp_sendspace`
		Recommended setting commands are:
		`no -p -o tcp_recvspace=262144`
		and
		`no -p -o tcp_sendspace=262144`

Parameter	Description and validation command	Recommended setting
`ulimit` settings	Critical `ulimit` settings for the `epicadm`, `epicdmn`, and root users	`epicadm`, `epicdmn`, and `root` should all have unlimited file size and unlimited core size. `root` should have unlimited data size. `epicadm` should have its data `ulimit` set to `4194304`. (Because `/etc/security/limits` uses 512-byte blocks, this corresponds to a value of `2097152` KB.) `epicdmn` should have its data `ulimit` set to `409600`. (Since `/etc/security/limits` uses 512-byte blocks, this corresponds to a value of `204800` KB.) `epicdmn` should also have unlimited user processes set. See: ulimit Command - IBM Documentation.
`vmm_mpsize_support`	Allow the use of 64 KB pages. Required for use of `ldedit` (see "64k Page Size" below) to enforce use of 64 KB pages for `irisdb` executable. `vmm_mpsize_support = -1` is the AIX default which enables AIX to determine the optimal page size to use on boot time.	Set to default (-1) to support 64 KB page size with: `vmo -r -d vmm_mpsize_support` Afterwards, run `bosboot` followed by a reboot.

Parameter	Description and validation command	Recommended setting
	Check with: `vmo -L vmm_mpsize_support` This command will show the AIX determined value (expected to be 3) as the CUR value and -1 should be the BOOT value.	
XL C++ runtime version	Check with: `lslpp -l xlC\.*`	The following libraries must be version 13.1.0.0 or later: • xlC.rte • xlC.aix61.rte • xlC.msg.en_US.rte On IRIS, the xlC.rte library must be version 16.1 or later.

64k Page Size	Applicable to both Power9 and Power10 but **required** for Power10 processor-based systems. This setting requires approximately a 30% increase in per-process memory, but Epic's Power10 hardware sizing has taken this increase into account.
	The command documented here must be performed once per IRIS installation or upgrade.
	This setting can be validated by running the following command and checking that DPageSize is 64k.
	`dump -ov -X64 <path to irisdb file>`
		To configure, run one of the following commands as appropriate:
		`ldedit -bdatapsize=64k /epic/prd/irissys/bin/irisdb`
		`ldedit -bdatapsize=64k /epic/prd/cachesys/bin/irisdb`

Parameter	Description and validation command	Recommended setting
File system bufstructs	Specifies the minimum number of file system bufstructs for Enhanced JFS. Will improve performance of JFS2 file systems. Check with: `ioo -o j2_nBufferPerPagerDevice`	Set with: ioo -p -o j2_nBufferPerPagerDevice=2048
LVM working element slots	Set to prevent running out of LVM working element slots. Check with: `lvmo -o workQ_size`	Set with: `lvmo -o workQ_size=1024`
Physical I/O buffers per physical volume	Specifies the minimum number of physical I/O buffers per physical volume. Check with: `ioo -o pv_min_pbuf`	Set with: `ioo -p -o pv_min_pbuf=4096` (this is a restricted tunable, so you must confirm the change by answering “yes”)
`hdisk queue depth`	Sets the hdisk queue depth to 64 (default is 20). Follow your storage provider’s recommendation. Check with: `lsattr -El <hdisk#> | grep queue_depth`	The following command sets the `queue_depth` value for a single Epic hdisk. Be sure to change all the Epic hdisks to the same setting. Set with: `chdev -l <hdisk#> -P -a queue_depth=64`
`vmm_vmap_policy`	Use the copy-on-reference policy for shared library mappings. Check with: vmo -o vmm_vmap_policy	This should be equal to the default (0). Set with: vmo -r -d vmm_vmap_policy

Kernel parameters specific to AIX 7.2

Parameter	Description and validation command	Recommended setting

`vmm_default_pspa`	Require a memory range to have 100% real memory occupancy to be promoted.
	Check with:
	`vmo -o vmm_default_pspa`
		`vmm_default_pspa=0`
		Set with:
		`vmo -r -d vmm_default_pspa`

`waitpid_direction`	This changes how certain wait states are handled by the OS. Check with: `schedo -o waitpid_direction`	This should be equal to 1. Set with: `schedo -p -o waitpid_direction=1`

Kernel parameters specific to AIX 7.3

Parameter	Description and validation command	Recommended setting
`lff` file system flag	Allows for files larger than 16 TB and file systems larger than 32 TB, if needed. Enable `lff` on file systems if your IRIS.DAT files are larger than 16 TB or your file systems are larger than 32 TB. Check with: `Mount | grep lff`	Set with options: `-o lff=yes` to the `mkfs` command or, `-a lff=yes` to the `chfs/crfs` commands
`hugeseg_shm_mode`	Allows for 1 TB segment size. Check with: `vmo -o hugeseg_shm_mode`	This should be equal to 2. Set for AIX 7.3 (7.2 recommendation pending) Set with: `vmo -p -o hugeseg_shm_mode=2`
NTP v4	AIX 7.3 removes support for NTP version 3. Ensure that you have NTP version 4 configuration to disallow time steps.	Should have a line with `tinker step 0` in `/etc/ntp.conf`.

	Check with: `grep tinker /etc/ntp.conf`	
`vmm_default_pspa`	Require a memory range to have 100% real memory occupancy to be promoted. Check with: `vmo -o vmm_default_pspa`	This should be equal to 0. Set with: `vmo -p -o vmm_default_pspa=0`

Additional recommendations

The following are the additional recommendations for running Epic applications on IBM systems and AIX.

Boot from SAN

Boot from SAN is **not** recommended when running the Epic environment. Both IRIS and PowerHA depend on the OS running correctly. If a SAN failure occurs such that the OS can no longer communicate with the rootvg volume, even for a brief interval, the condition of the OS is unsure. The system may appear to be operating correctly. However, if any OS specific data was lost during transfer between RAM and disk, the OS is no longer viable. Because all software running on the system depends entirely on the OS, user products or supporting middleware may no longer function correctly.

Epic recommends that customers do not boot from SAN so that Epic can log into the system following a failure to troubleshoot. However, PowerHA recommends that customers boot from SAN partly because of the Live Partition Mobility (LPM) feature. The decision to boot from SAN should be discussed with your Epic representative.

PowerHA

There are multiple resources for information regarding the best method of configuring a PowerHA failover cluster. Epic provides its customers with PowerHA callable scripts containing the necessary instructions to effectively shut down and start up the Epic and IRIS environment.

Most IT system administrators view PowerHA as being capable of recovering from all events that could occur to an Epic environment. As much as we would like to imagine such a safety mechanism, it does not exist.

What PowerHA will do:

PowerHA can recover from real hardware failure that includes servers, switches, disk systems, and any other type of device which could experience a physical failure due to power loss, electronic component failure, or a catastrophic event.

What PowerHA will not do:

PowerHA cannot recover from user errors, either intentional or accidental. Because PowerHA depends on the operating system, it is assumed that if the operating system started running the Epic environment without a problem, it should continue to support the environment without a problem. There are two conditions where the OS could fail (a) A hardware failure, or (b) A change made to the OS environment by a user. In case of (a), PowerHA will recognize the hardware failure and initiate a failover. PowerHA, however, will not support case (b).

PowerHA requires diligent administration and monitoring. PowerHA cannot be installed and left alone to run by itself. Taking this approach certainly results in eventual failure of the correct operation of PowerHA.

PowerHA documentation provides two major recommendations:

1. Whenever a change is made to the cluster that is being managed by PowerHA, no matter how trivial it might seem, PowerHA must always be re-tested to ensure that nothing was modified in such a way that PowerHA can no longer function properly.

2. Regardless of whether the system was modified or not, a manual PowerHA failover should be conducted at regular intervals (for example, every three months).

3. Item (a) provides two benefits: It gives confirmation that a PowerHA failover will work when an unexpected failure occurs. By executing a planned failover, any problems can quickly be identified and resolved.

4. PowerHA depends heavily on the environment that it is assigned to manage. Due to its flexibility, there are many ways to misconfigure a PowerHA environment. There is only one way to be certain that PowerHA has been configured to run successfully: Test fail-over.

PowerHA and single point of failure (SPOF)

For PowerHA to work, it must not be limited by single points of failures (SPOFs). For example, for PowerHA to maintain inter-nodal communication within the HA cluster there must exist more than a single communication path. This requires the availability of completely redundant switches, cables, and adapters from one end to the other. Having eight communication adapters on each node does no good if the two nodes are connected through a single data path (Ethernet cable). Having multiple redundant zones on a switch would not help if the switch lost power.

Therefore, building in redundancy is a must. This requires that half of the equipment may be sitting idle, until a failure occurs, which unfortunately, is a cost of maintaining a High Availability environment.

PowerHA and enhanced concurrent volume group

Customers who are using Epic are required to provide a failover system which will take over in the event of a primary online transaction processing (OLTP) system failure. IBM offers this facility on IBM Power processor-based systems with PowerHA.

If the active compute system that is running Epic encounters a failure, PowerHA will recognize the loss of the active system. The failover process causes the resources being used by the primary system (primarily the attached storage system) to be acquired by the takeover system. The backup system will then attempt to start the same Epic environment. Although the takeover is not instantaneous, it does provide an automated method to recover from a catastrophic hardware failure.

In more recent versions of PowerHA, IBM has introduced the use of enhanced concurrent volume groups (ECVG). The primary advantage of ECVG is that the Epic database volumes are already varied-on to both PowerHA nodes (active and standby nodes). In the event of a failure, the time required for the takeover node to acquire the Epic volumes is greatly reduced. Therefore, IBM has encouraged its PowerHA customers to take advantage of ECVG mounted volumes that are associated with a PowerHA cluster.

In the unlikely event that PowerHA itself fails, ECVG can potentially cause a ‘split brain’ event. When both the nodes in the cluster can no longer communicate, or especially if the takeover node believes that the primary node has failed, it is possible for both

nodes to become active. Therefore, it is possible that the Epic software could start running on the takeover node while the primary node is still in play. Recent versions of PowerHA (versions 6.1 and 7.1) have significantly reduced the possibility of a 'split brain' event occurring. In PowerHA version 6.1, ECVG can safely be used in the Epic environment. In PowerHA version 7.1 and later, ECVG is mandatory. Therefore, PowerHA ECVG can safely be used in an Epic environment.

When logical volumes are mounted concurrently, it allows access from more than one compute node simultaneously. Therefore, when a volume group is mounted concurrently, data on the volumes can be updated by both nodes.

Micro-Partitioning

Micro-Partitioning® or shared processing LPAR (SPLPAR) is currently not supported within an Epic production environment. Dynamic LPAR (DLPAR), however, is supported.

There are several reasons why Epic does not recommend the use of SPLPAR.

- If both CPU and memory resources were to be shared between Epic and other applications, there is always a possibility that a non-Epic application could choke resources away from Epic during a critical time.
- When Epic provides the sizing information, the assumption is that the Epic products are the only ones actively running on the system. Therefore, at a minimum, the Epic partition would need to be fully configured with the Epic required resources. It is assumed that those resources are always available. Thus, in effect, the Epic LPAR would really be regarded as a fixed resource LPAR, or DLPAR.
- Epic sizes the DB server so that, under normal load, the customer is not running above 75% CPU utilization while set to SMT4 and above 65% CPU utilization while set to SMT8. The test team does not know how quickly a shared partition can obtain resources from another shared partition before those resources can begin to provide some relief during a sudden and unplanned increase in resource demand originating from the Epic partition. In any case, the priority for *spare capacity* to the Epic partition would require top priority over all other partitions, thereby, once again making the Epic partition an effectively independent DLPAR.

- If a performance-related problem occurs, Epic must be in a position to be able to reproduce the problem. If performance was degraded due to shared resources being unavailable, it would be more difficult for Epic (or IBM), to identify whether the cause was due to something that happened within the Epic partition, or whether an external load-driven event was the cause.
- At this time, the team has not adequately tested the interaction between SPLPAR and PowerHA. As an example, what would happen, or what would the test team expects to happen, is the system to experience a physical CPU failure. What should PowerHA do if Epic happened to be using one tenth or more of the physical CPU at the time? Normally, loss of a resource would trigger a failover. However, this CPU is now a *virtual* resource.
- Epic, however, has no objection to the use of SPLPAR in a non-production environment, if performance is not being evaluated within that environment.

Virtual I/O

Virtual I/O (VIO) may be used in the Epic environment. Although Virtual I/O may provide better use of existing hardware resources, the performance impacts must be considered in the production environment. The number of the adapters that are being included in a VIO environment must continuously provide the same level of performance as in a non-VIO environment.

N_Port ID Virtualization (NPIV) virtualizes a physical Fibre Channel adapter, thereby allowing the assignment of multiple worldwide names (WWNs). Again, the total load of multiple LPARs being supported by a physical adapter must be considered.

Epic prefers the use of physical adapters over VIO servers for the production OLTP system. If VIO servers are preferred for enterprise virtualization or consolidation practices, the following considerations apply when using VIO with the production OLTP LPAR and its failover LPAR:

- Follow IBM’s best practices to efficiently set up redundancy at the VIO layer to avoid single points of failure.
- Follow IBM’s recommendation to properly size the VIO servers for the overall activities on the server frame.

  - When using Oracle on an IBM Power server as the Clarity RDBMS, make sure that the Clarity RDBMS Oracle server is on separate VIO servers from the production OLTP LPAR and its failover LPAR.
  - You should employ redundant VIO servers. Each VIO server must have sufficient CPU and memory resources to support the full load expected. If they are in a shared processor pool, the VIO servers should have the highest weight within the pool to avoid being starved by activities from other application LPARs.
  - Each VIO server must have a total of at least four ports from at least two physical host bus adapters (HBAs). The total I/O bandwidth provided by the HBAs must accommodate the total input/output operations per second (IOPS) projection from all LPARs, with sufficient redundancy. The IOPS projections from the main Epic components can be found in the previous I/O projection and requirements section.
  - The total network bandwidth provided by the Ethernet adapters must accommodate the network traffic expected from all LPARs, with sufficient redundancy. 25 Gb interfaces are generally more appropriate for large scale systems. If using 10 Gb interfaces, multiple interfaces may have to be aggregated to provide adequate bandwidth and acceptable latency. The Ethernet network must provide enough bandwidth for all the Epic functional requirements (for example, mirror, backup, and so on). You may still find it beneficial to use separate network interface controllers (NICs) for traffic that may have unbounded bandwidth usage patterns.

- There are two technologies available to provide I/O access using VIO: virtual SCSI and NPIV. Discuss with the IBM support team and identify the technology that best suits your needs.
  - Be aware that queue_depth needs to be properly tuned at both VIO server layer and the production LPAR layer when using virtual SCSI.
  - Epic has conducted performance tests with NPIV and found the results acceptable.

There could be different VIO considerations for SAN boot. If you want to use SAN boot, follow IBM’s best practices for SAN boot.

Live Partition Mobility

Live Partition Mobility (LPM) provides the ability to move an existing running Epic instance from one Power frame to another. During a migration, impact on performance may be observed depending on the size of the Epic environment being migrated. The database activity may be momentarily suspended. This may result in user clients being disconnected temporarily. The alternative for migrating an Epic production instance from one Power frame to another is to initiate a manual PowerHA failover. Using PowerHA would result in anywhere from at least a 5 to 15-minute outage, versus a brief user client disconnect of less than a minute when using LPM.

LPM requires VIO servers on both the source and the target Power frames. Use of NPIV is strongly recommended to support LPM.

An LPM migration must be done only during low-use hours (whenever there is minimal use of the Epic production database).

Information to collect when a problem occurs

A performance issue can be caused by any part of either the server system or the storage system. Because each stage of the computational process depends on all others, it can often be difficult to identify the true culprit causing a problem. For example, although it seems that obtaining data from storage appears slow, it may in fact be the case that the server is running out of I/O buffers or disk queues to handle the incoming data from the storage system. Therefore, each stage of the process must be analyzed and diagnosed. The primary task is to determine whether a stage in the process is waiting for something, (starving), or whether the stage is overloaded.

For example, the disk I/O throughput may seem reasonable for the given configuration. However, users are noting a substandard response time. Upon further investigation, it is determined that the Logical Volume Manager (LVM) has run out of resources on the server. This may not be immediately evident because you do not see large amounts of CPU being consumed. However, lack of certain JFS buffers could result in a bottleneck.

Following is a partial list of information which should be collected when reporting a problem either to IBM support or to anyone involved in technical support of Epic.

- Have they filed a problem management record (PMR) with IBM? If so, provide the PMR number.

- Has Epic been made aware of the problem? Who is the primary Epic contact that the customer is dealing with?
- What is the type of System P server, Model, number of CPUs, total memory, DLPARs, SPLPARs, and so on.
- What is the type of storage, number of HDisks, storage configuration, (for example, RAID 5, RAID 10, stripe size, number of ranks, LUNs, and so on)? Is the customer using SVC?
- Is the storage or SVC being shared with other non-Epic applications?
- What has been changed prior to experiencing the performance problem? For example, increased users, change in storage configuration, additional workloads, application upgrades, and so on.
- Did the performance degrade suddenly or was it a slow degradation over time?
- Is there a particular hour of day or night that the performance degrades? Is it constant?
- Can the customer provide results from the Epic RanRead facility?
- Does the performance degradation occur during an IBM FlashCopy® operation or other back-end copy procedures?

Also, provide a topology diagram showing the OLTP, mirror, failover servers, the storage switches, and the associated interconnects to each component which supports the entire Epic environment. In most cases, the IBM Support team provides instructions on collecting performance data and other important data for troubleshooting issues and identifying the root cause. Often, customers will be asked to collect perfpmr information while an issue occurs. This can be done by running the script perfpmr.sh found at: https://www.ibm.com/support/pages/perfpmr-tool.

It is recommended to run perfpmr.sh -T -s -N prior to any system upgrades or changes that may impact performance and during any performance-related issues. The -s option omits svmon data that may take a very long time on Epic servers and the-T option omits iptrace/tcpdump. Epic recommends the following changes to perfpmr.cfg prior to running perfpmr.sh:

```
get_inode_table = false
lock_trace_level = 9
```

Summary

IBM Power and AIX is a powerful combination for running Epic electronic medical records applications. This paper has provided installation, configuration, and troubleshooting best practices to achieve optimal results for your organization. Epic software requires a server platform that has been tested and approved by their technical team and that is continuously monitored by them while deployed at customer sites. IBM Power and AIX meet these requirements and are a target platform for Epic implementation.

Get more information

To learn more about running Epic EMR software on AIX, contact your IBM representative or IBM Business Partner.

About the author

Chip Elmblad is a Technical Engagement Leader with IBM Systems ISV Engineering Organization. He has more than 10 years of experience working with AIX. As part of the IBM Systems ISV Engineering Organization, he is responsible for assisting Epic customers to achieve optimal performance, availability, and scalability from IBM products.

IBM Corporation New Orchard Road Armonk, NY 10504

Produced in the
United States of America December 2022