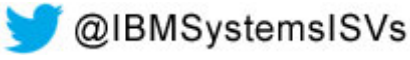
# IBM Power Systems solution for EnterpriseDB Postgres Plus Advanced Server

*A white paper on EnterpriseDB Postgres Plus Advanced Server running on Linux on IBM POWER8 processor-based servers*

*Deepak Narayana*
*IBM Systems and Technology Group ISV Enablement*

*August 2014*



## Table of contents

# Abstract

*This paper demonstrates the benefit of EnterpriseDB Postgres Plus Advanced Server database on IBM Power Systems running Linux servers when compared to a similar configuration using the Intel Xeon processor-based system.*


# Introduction

The primary focus of this paper is on the throughput of the EnterpriseDB Postgres Plus Advanced Server running different load scenarios on the IBM® Power Systems™ servers featuring the new IBM POWER8™ processor technology.

**Note:** The Red Hat Enterprise Linux (RHEL) 6.5 operating system was used for both the systems tested in this paper. The scope of this white paper is to demonstrate the use and scaling of EnterpriseDB open source based database, Postgres Plus Advanced Server on IBM Power Systems running Linux. It is based on the open source database, PostgreSQL, and is capable of handling a wide variety of high-transaction and heavy-reporting workloads.

This paper covers the setup and configuration of the Postgres Plus Advanced Server on IBM Power Systems running Linux for best throughput when loaded. In this exercise, the test team used a benchmark tool called *pgbench* which is loosely based on TPC-B. The *Select* query option will be used in this tool to run the tests.

The results show how Power Systems can sustain and surpass other hardware throughputs with lesser number of cores while still leaving room for more scaling potential.

## Logical partitions and dynamic logical partitions

All IBM POWER5™, IBM POWER6®, POWER7® and POWER8 processor-based UNIX® and Linux® servers are logical partition (LPAR)-capable. That is, they support LPARs, which are multiple OS images of varying sizes and capacities that exist on a single physical server frame. Even if there is only one image on an applicable hardware platform, it is still an LPAR.

The IBM POWER® processor-based servers support the dynamic reallocation of memory and processors, referred to as dynamic LPARs (DLPARs). The system administrator can add or remove processors and memory in an LPAR without restarting. This assumes that the chosen values stay within the minimum and maximum memory and processor values specified in the LPAR profile. The same applies to virtual resources. IBM PowerVM®, (earlier known as Advanced Power Virtualization) is a combination of hardware, firmware, and software that provides processor, network, and disk virtualization, and hence provides micro-partitions. For this exercise, the test team chose the full system as a single LPAR, with dedicated resources.

# Environment

This section describes the hardware and software environment used for this exercise. The configuration might vary depending on each user's requirement.

## Test setup

The test environment consists of the following hardware and software.

**Hardware:**

- IBM Power® System S822L server
  - Two sockets with 10 cores/ 1 socket active
  - IBM POWER8 processors at 3.42 GHz
  - 128 GB memory
  - Local disk for OS (RHEL 6.5) and EnterpriseDB Postgres Plus Advanced Server 9.3 binary files

- IBM Flex System® x240 compute node
  - Two sockets with 16 cores total
  - Intel® Xeon® processors E5-2670v at 2.60GHz
  - 128 GB memory
  - Local disk for OS (RHEL 6.5) and EnterpriseDB Postgres Plus Advanced Server 9.3 binary files

**Software:**

- RHEL 6.5 (64-bit)
- EnterpriseDB Postgres Plus Advanced Server 9.3
- pgbench (a benchmarking tool that is part of EnterpriseDB Postgres Plus Advanced Server 9.3)

**Note:** For this exercise, *pgbench*, the benchmarking tool, ran on the same system where EnterpriseDB Postgres Plus Advanced Server resided.

## General recommendation

It is generally recommended to have a root and data disk on separate physical disk or Redundant Array of Independent Disks (RAID). In order to eliminate the impact of performance differences between disparate storage technologies, the database was instead kept in memory on a *tmpfs* virtual file system.

## EnterpriseDB Postgres Plus Advanced Server on IBM Power systems running Linux

Postgres Plus Advanced Server is built upon PostgreSQL, one of the most advanced open source databases and has additional functionality and capabilities in the following areas:

- Performance
- Compatibility
- Security
- Tooling

With Postgres Plus Advanced Server's Oracle compatibility features, you can use existing Oracle-based applications on a low-cost, high-performance PostgreSQL-based platform, or run adjacent applications that integrate seamlessly with your mission-critical databases without additional Oracle licenses.

Always ensure that EnterpriseDB is installed with the latest service pack available from the EnterpriseDB website at: http://www.enterprisedb.com/downloads/postgres-postgresql-downloads.

## Prerequisites

Refer to EnterpriseDB manuals for software prerequisites.

- Make sure that there is adequate paging space and free space on the /tmp and / (root) file systems.
- Verify the ulimits setting for each of the products using the `ulimit –a` command
- Refer to the EnterpriseDB tuning guide and best practices guide.
- EnterpriseDB setup requires Java™. You can download the IBM Java from IBM developerWorks website. Follow the instruction to install Java at: **ibm.com**/developerworks/java/jdk/linux/download.html

# Setting up EnterpriseDB Postgres Plus Advanced Server

EnterpriseDB can be installed using the **rpm** command line or the yum installer. It is recommended to perform the installation using the yum tool.

First, you need to make sure whether a yum repository for RHEL is defined. Also, you can define a yum repository for EnterpriseDB rpms for easy one-step installations. Defining a yum repository for the base OS and software installation packages enables the installation of software to automatically fetch the rpm packages (which are prerequisites).

## Configuring the yum repository

A RHEL yum repository can be defined by pointing to a CD or a DVD media, a shared network location, or a local directory that contains the RHEL installation media. The test team defined it using a local directory. The yum configuration files are in the /etc/yum.repos.d/ directory.

You can find more information about yum repositories at:

https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Deployment_Guide/sec-Managing_Yum_Repositories.html

```
[root@plinux ~]# cat /etc/yum.repos.d/rhel-dvd.repo
[rhel64-dvd]
name="Red Hat Enterprise Linux Installation DVD"
baseurl=file:///home/RHEL6.4-dvd
enabled=1
gpgcheck=0
```

*Figure 1: Sample yum configuration fro RHEL 6.5*

The same need to be defined for EnterpriseDB packages. Refer to Figure 2.

```
[root@plinux ~]# cat /etc/yum.repos.d/edb.repo
[edb-ppa93]
name=EnterpriseDB PPA9.3
baseurl=file:///home/downloads/ppas-9.3-rhel-6-ppc.rpms
enabled=1
gpgcheck=0
```

*Figure 2: Sample yum configuration for EnterpriseDB*

## Installing EnterpriseDB Postgres Plus Advanced Server

You can find details about installing EnterpriseDB Postgres Plus Advanced Server at:

http://enterprisedb.com/docs/

After the yum repositories are defined, you can perform the installation of EnterpriseDB using a simple `install` command:

```
yum install ppas93-server
```

This command automatically installs all the required prerequisites, and also creates a database administrator user, *enterprisedb*.

Before proceeding with creating and configuring the DB instance, it is best to apply tunings. In this exercise, the test team performed various system tunings, including kernel tuning.

## Power Systems configurations

The following settings were used for this exercise.

- Single partition mode (non-virtualized LPAR) – The whole system was assigned to a single partition with all cores *dedicated*. And Hardware Management Console (HMC) access is needed to create this LPAR. This would be the only LPAR on the system.
- Disable the following power saving options for the POWER8 processor-based system.
  - Turn off the **Idle Power Saver** option.
  - Set **Dynamic Power Saver** mode to **enabled, favor performance**.

This setting can only be done in the Advanced System Management (ASM) interface. For this test, HMC was used to access the ASM.

Launch ASM from HMC for the particular system.

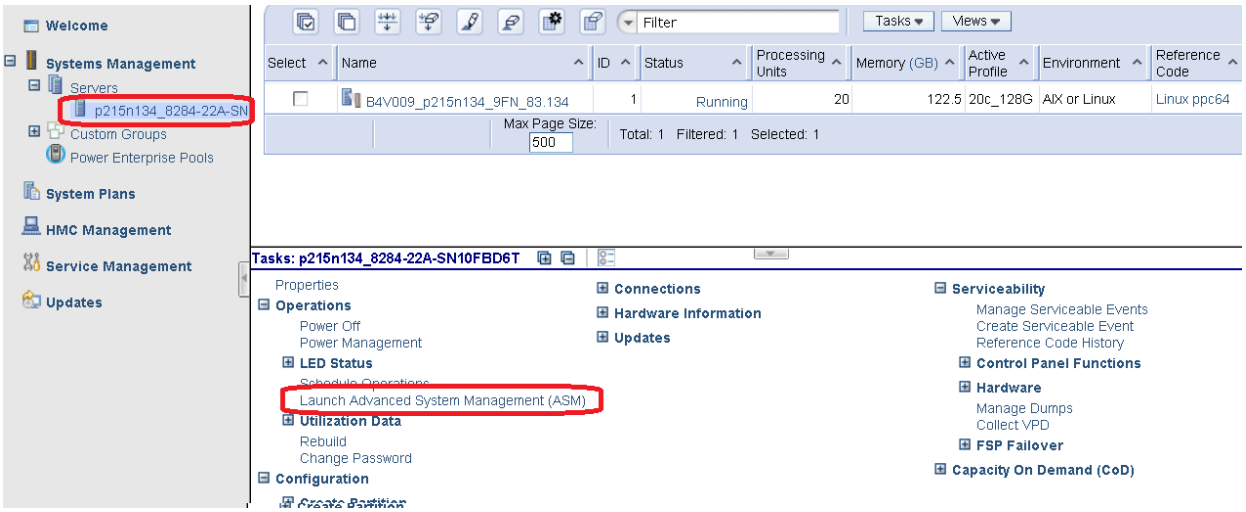



*Figure 3: Accessing ASM interface from HMC*

To turn off **the Idle Power Saver** option, log in and go to **System Configuration → Power Management → Idle Power Saver**.

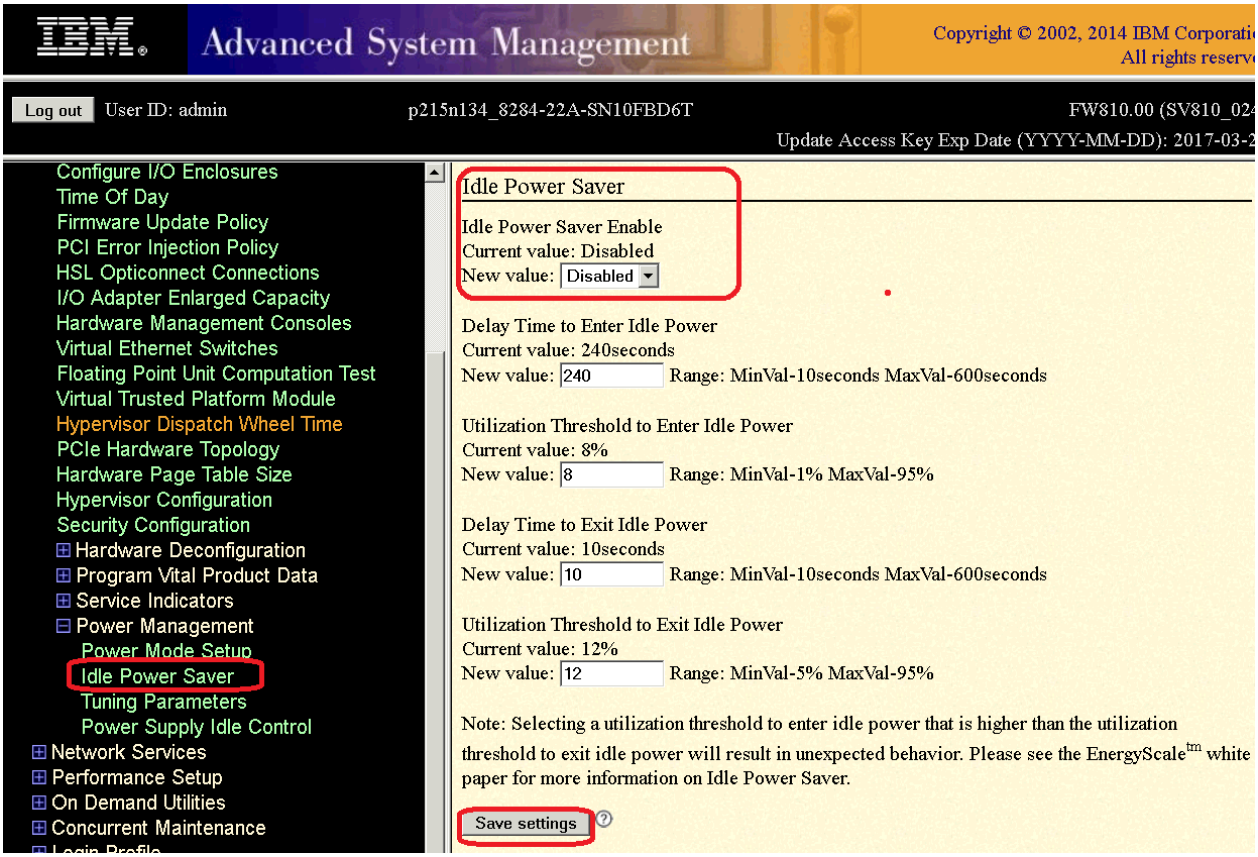



*Figure 4: ASM interface - Idle power saver*

To enable the dynamic power saver mode, perform the following steps.

1. Click **Power Mode Setup** in the left pane, and select the **Enable Dynamic Power Saver (favor performance) mode** option in the right pane.
2. Click **Continue**.

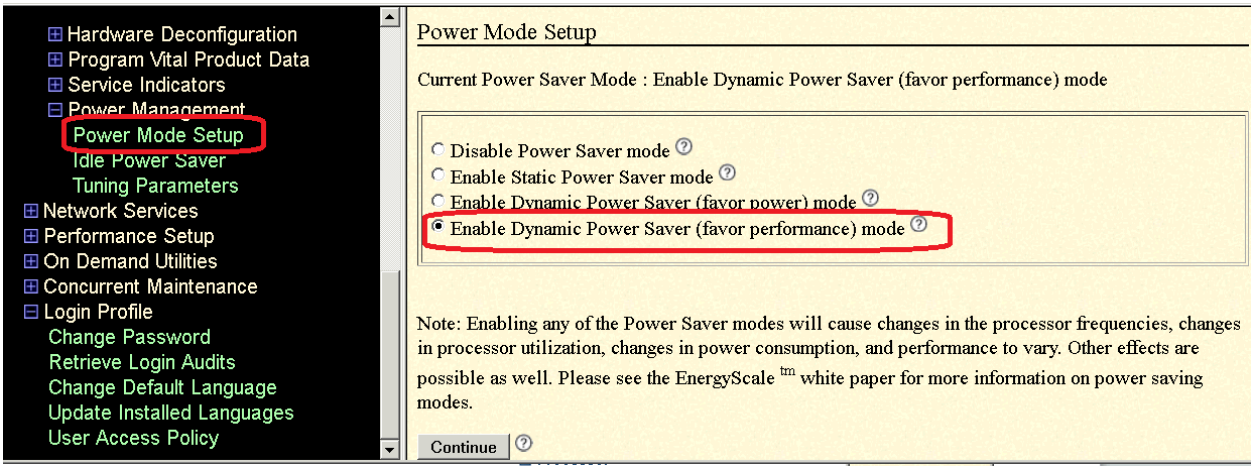



*Figure 5: ASM interface – Power Mode Setup*

These settings help in increasing the processor speed. This can be verified at the shell by using the `ppc64_cpu –freq` command. This operation does not require the system to be shut down or restarted. It takes effect immediately.

## Kernel tunings and OS tunings

The following kernel parameter values were applied as a root user for this exercise. Perform the activity with caution and set the values according to your hardware specifications and operating system implementation.

- sysctl -w fs.file-max=65535
- sysctl -w vm.dirty_background_bytes=37108864
- sysctl -w vm.dirty_ratio=40
- sysctl -w vm.dirty_background_ratio=40
- sysctl -w vm.hugetlb_shm_group=`id -g enterprisedb`
- sysctl -w vm.dirty_bytes=296870912
- sysctl -w vm.nr_hugepages=2000
- sysctl -w vm.swappiness=0
- sysctl -w vm.hugepages_treat_as_movable=0
- sysctl -w vm.nr_overcommit_hugepages=512
- sysctl -w vm.zone_reclaim_mode=0
- sysctl -w vm.drop_caches=3
- sysctl -w kernel.sched_migration_cost=500000
- sysctl -w  kernel.sched_autogroup_enabled=1

**OS tunings**

- Set the processor's simultaneous multithreading (SMT) snooze delay to be higher using the following command.

  ```
  ppc64_cpu --smt-snooze-delay=16777215
  ```

  The default value is 100. This is a tunable parameter to delay the entry to nap state.

- Use the following command to turn off I/O preemption.

  ```
  mount -t debugfs debugfs /sys/kernel/debug
  ```

  ```
  echo NO_WAKEUP_PREEMPT > /sys/kernel/debug/sched_features
  ```

  The preempt scheduler relates to the releasing of the processor time when a process of a higher priority wants to use the processor. So, if you have a lot of processes with similar priorities, the processor time can be consumed with this swapping in and out.

- Run the following command on your Linux system to disable hardware data prefetch.

  ```
  ppc64_cpu -dscr=1
  ```

- Run the following command to restore the default value.

  ```
  ppc64_cpu -dscr=0
  ```

- Add the following entries to the limits file for the enterprise user - /etc/security/limits.conf

```
enterprisedb     soft    memlock        68719476736

enterprisedb     hard    memlock        68719476736
```

This limits the maximum amount of locked-in-memory address space.

- Optionally, enable large pages for EnterpriseDB by adding the following lines to the EnterpriseDB user shell profile.

```
HUGETLB_SHM=yes

LD_PRELOAD='/usr/lib64/libhugetlbfs.so'

export HUGETLB_SHM

export LD_PRELOAD
```

In case, the Postgres Plus Advanced Server complains of usage of large pages, it could be resolved by running this command as root.

```
hugeadm --pool-pages-min DEFAULT:16M
```

**File system for database to reside in memory**

The name of the file system used for this exercise is *tmpfs*. A memory of 20GB was allocated for EnterpriseDB purposes. The total space used during the run was noted to be in the range of 15 GB to 19 GB for *pgbench* scale of 1000.

Run the following command to create the *tmfs* file system -

```
mkdir -p /media/tmp.

mount -t tmpfs -o size=20G tmpfs /media/tmp
```

**Note:** A system reboot deletes this file system and its information. You need to back it up as needed.

Also, stop or disable any services or process that is not needed for this run or is critical to the system. For example, IPV6 can be disabled, the mail server can be stopped, and so on.

# EnterpriseDB database setup

This section explains how to initialize a database cluster instance, apply the DB cluster parameter tunings, and create a database.

To begin with, log in as an *enterprise* user and make sure the EnterpriseDB Postgres Plus Advanced Server utilities are defined in the path. You can find the binary files at the default installation location at /usr/ppas-9.3/bin.

## a) Initialize the DB cluster instance

The *initdb* utility is used to create a DB cluster instance. The location for the DB files needs to be specified as a parameter in the command. The *tmpfs* file system created in the earlier step is used here as the database instance folder.

```
initdb -D /media/tmp/data
```

This creates a database repository along with a database parameter file, postgresql.conf.

### b) Tune the EnterpriseDB Postgres Plus Advanced Server parameters

You can find the default DB cluster parameter in the postgresql.conf file. For this exercise, the following parameters were changed to the recommended values.

- shared_buffers = 20GB (you need to tune this based on the memory allocated for the LPAR. The recommendation is that this cannot exceed 1/4th of the LPAR memory.
- maintenance_work_mem = 512MB
- checkpoint_completion_target = 0.9
- effective_cache_size = 64GB
- work_mem = 512MB
- wal_buffers = 16MB
- checkpoint_segments = 300
- synchronous_commit = off
- Disable Dynatune by commenting out the edb_dynatune and edb_dynatune_profiles

### c) Start the DB cluster and create a DB

The DB cluster is started using the *pg_ctl* utility that comes with the EnterpriseDB Postgres Plus Advanced Server. Output is redirected to a file using the **–l** option.

- `pg_ctl -D /media/tmp/data -l logfile start`
  It can also be started using the `edb-postgres -D /media/tmp/data` command.
- Database can be created using the *createdb* command-line utility. Here, the *pgbench* is chosen as the database name.

  ```
  createdb pgbench
  ```

## EnterpriseDB Postgres Plus Advanced Server database benchmarking

This section discusses the benchmark test that ran on the database after initializing the workload.

Reinitializing the DB is not needed for every run, but recommended as it refreshes the DB.

### Benchmarking tool *pgbench*

*pgbench* is a simple utility to run the benchmark tests on EnterpriseDB Postgres Plus Advanced Server or PostgreSQL. You can find more information about this utility at:
http://www.postgresql.org/docs/devel/static/pgbench.html

pgbench offers various read-only (*select only* query) and read-write (*select, update* and *insert* queries) modes. For this exercise, the *select only* option of the pgbench was used. This utility can be run on the same system as EnterpriseDB Postgres Plus Advanced Server. This utility can also be run on a separate computer over the network as well.

**Initialize the database**

First, the database must be initialized. The *pgbench* utility is invoked using the **-I** option and a scaling factor is specified. In this exercise, the test team used a scaling factor of 1000.

```
pgbench -i -s 1000 pgbench
```

Initialization takes some time as it populates the DB. Scaling of 1000 typically consumes around 16 GB of memory.

**Run the benchmarking tool**

The parameter passed is **-T**, which specifies the duration for which the tool runs, **-S** enables *select only* loads, **-c** is the number of clients, and **-j** is the number of worker threads. This run is performed for various client and thread counts. For this test, the client and thread resided on the same core. The *pgbench* tool is run for 5 minutes (300 seconds). For stable results, running *pgbench* for 5 minutes or longer is recommended

```
pgbench -n -S -T 300 -c 100 -j 100 pgbench
```

At the end of the run, this utility displays the results as transactions per second (TPS) including and excluding connections.

```
transaction type: SELECT only
scaling factor: 1000
query mode: simple
number of clients: 100
number of threads: 100
duration: 300 s
number of transactions actually processed: 68026668
tps = 226750.289567 (including connections establishing)
tps = 226981.752582 (excluding connections establishing)
```

*Figure 6: Sample result output from pgbench tool*

## Results

The following results were measured using the *pgbench* tool for various client threads. Data points were gathered at key intervals. Servers based on both Intel and POWER8 processors were evaluated. The systems were maintained in similar states of tunings.

| IBM Power S822L server with POWER8 processor and 10 cores | | | | |
|---|---|---|---|---|
| **Number of clients** | **Scaling factor** | **Thread** | **TPS** | **Per core TPS** |
| 10 | 1000 | 4 | 98462 | 9846 |
| 20 | 1000 | 4 | 172131 | 17213 |
| 40 | 1000 | 4 | 217059 | 21706 |
| 80 | 1000 | 4 | 261358 | 26136 |
| 100 | 1000 | 4 | 268918 | 26892 |
| 120 | 1000 | 4 | 283884 | 28388 |

*Table 1: pgbench results from a POWER8 system*

| IBM Flex System x240 compute node with Xeon E5-2670 processor and 16 cores | | | | |
|---|---|---|---|---|
| **Number of clients** | **Scaling factor** | **Thread** | **TPS** | **Per core TPS** |
| 10 | 1000 | 2 | 68575 | 4286 |
| 20 | 1000 | 2 | 148692 | 9293 |
| 40 | 1000 | 2 | 241847 | 15115 |
| 80 | 1000 | 2 | 225007 | 14063 |
| 100 | 1000 | 2 | 226750 | 14172 |
| 120 | 1000 | 2 | 226978 | 14186 |

*Table 2: pgbench results from an x86 system*

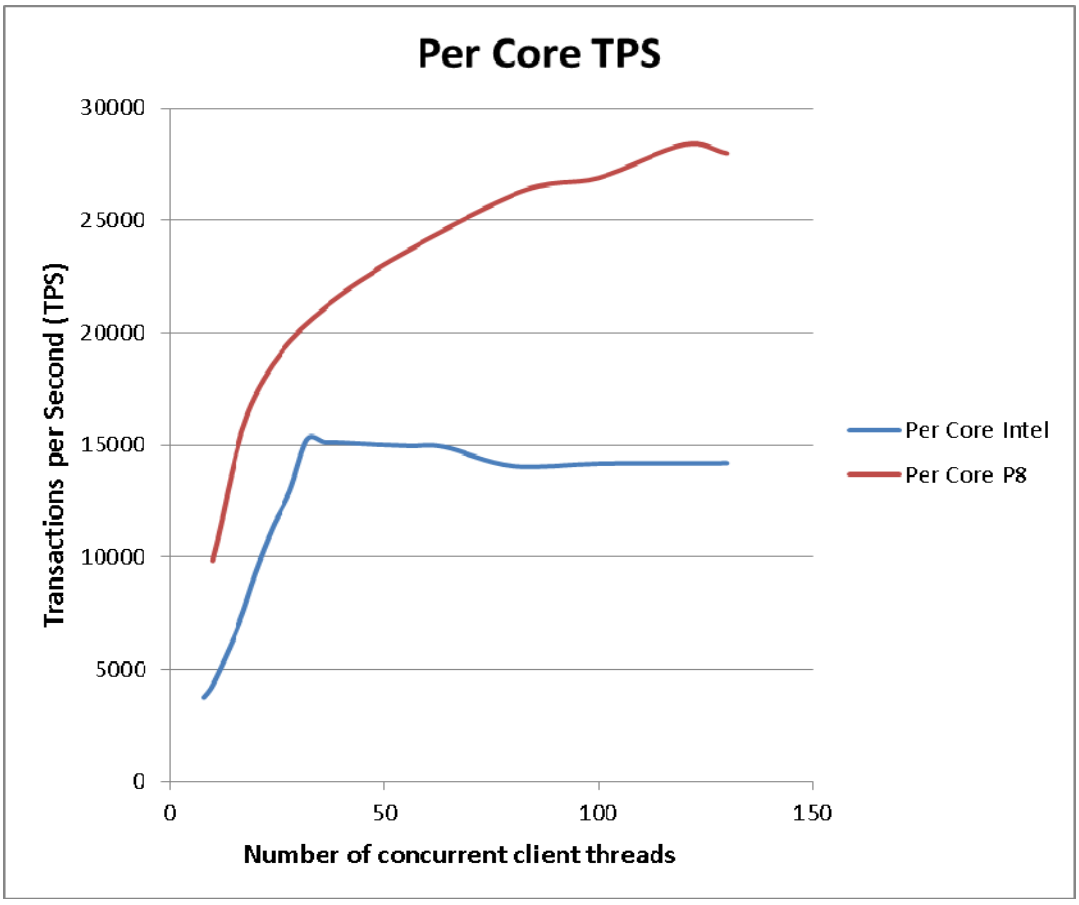



*Figure 7: Graph representing EnterpriseDB pgbench results on a per core basis*

**Note:** *pgbench* workload varies at lower client threads; the results can vary for each run.

# Summary

The IBM Power S822L server with 10 cores on one -socket system provides a throughput of 283,884 transactions per second and the scalability is stable up to 120 client threads. The x240 server with 16 cores of Xeon E5 processors on a two-socket system, running the same software components, provides a throughput of 241,847 transactions per second and the scalability drops after 40 client threads.

**Note:** The POWER8 processor-based system runs in SMT4 mode (four hardware threads per physical core), whereas, the Intel processor-based system supports only two hardware threads per core on the Xeon E5 processor.

Additional tuning and optimizations are feasible. RHEL7 supports optimizations and tuning specifically for the new POWER8 processor including SMT8. RHEL7 testing is planned after supported by EnterpriseDB.

Scaling the software environment after 40 hardware threads showed limitations. Because the tested configurations already sustained very large transaction rates, it was not explored in this paper. If sizing larger systems, the hardware thread scalability might need to be taken into account and further researched.

# Appendix A: Sample postgres.conf parameter file

This section shows a sample postgres.configuration which were modified for this exercise.

```
# -----------------------------
# PostgreSQL configuration file
# -----------------------------
#
# This file consists of lines of the form:
#
#   name = value
#
# (The "=" is optional.)  Whitespace may be used.  Comments are introduced with
# "#" anywhere on a line.  The complete list of parameter names and allowed
# values can be found in the PostgreSQL documentation.
#
# The commented-out settings shown in this file represent the default values.
# Re-commenting a setting is NOT sufficient to revert it to the default value;
# you need to reload the server.
#
# This file is read on server startup and when the server receives a SIGHUP
# signal.  If you edit the file on a running system, you have to SIGHUP the
# server for the changes to take effect, or use "pg_ctl reload".  Some
# parameters, which are marked below, require a server shutdown and restart to
# take effect.
#
# Any parameter can also be given as a command-line option to the server, e.g.,
# "postgres -c log_connections=on".  Some parameters can be changed at run time
# with the "SET" SQL command.
#
# Memory units:  kB = kilobytes        Time units:  ms  = milliseconds
#                MB = megabytes                     s   = seconds
#                GB = gigabytes                     min = minutes
#                                                   h   = hours
#                                                   d   = days


#------------------------------------------------------------------------------
# FILE LOCATIONS
#------------------------------------------------------------------------------

# The default values of these variables are driven from the -D command-line
# option or PGDATA environment variable, represented here as ConfigDir.

#data_directory = 'ConfigDir'           # use data in another directory
                                        # (change requires restart)
#hba_file = 'ConfigDir/pg_hba.conf'     # host-based authentication file
                                        # (change requires restart)
#ident_file = 'ConfigDir/pg_ident.conf'      # ident configuration file
                                        # (change requires restart)

# If external_pid_file is not explicitly set, no extra PID file is written.
#external_pid_file = ''                      # write an extra PID file
                                        # (change requires restart)


#------------------------------------------------------------------------------
# CONNECTIONS AND AUTHENTICATION
#------------------------------------------------------------------------------

# - Connection Settings -

#listen_addresses = '*'                 # what IP address(es) to listen on;
listen_addresses = '127.0.0.1'          # what IP address(es) to listen on;
                                        # comma-separated list of addresses;
                                        # defaults to 'localhost'; use '*' for all
```

```
                                        # (change requires restart)
#port = 5444                            # (change requires restart)
#max_connections = 100                  # (change requires restart)
max_connections = 150                   # (change requires restart)
# Note:  Increasing max_connections costs ~400 bytes of shared memory per
# connection slot, plus lock space (see max_locks_per_transaction).
#superuser_reserved_connections = 3     # (change requires restart)
#unix_socket_directory = ''             # (change requires restart)
#unix_socket_group = ''                         # (change requires restart)
#unix_socket_permissions = 0777                 # begin with 0 to use octal notation
                                        # (change requires restart)
#bonjour = off                          # advertise server via Bonjour
                                        # (change requires restart)
#bonjour_name = ''                      # defaults to the computer name
                                        # (change requires restart)

# - Security and Authentication -

#authentication_timeout = 1min                  # 1s-600s
#ssl = off                              # (change requires restart)
#ssl_ciphers = 'ALL:!ADH:!LOW:!EXP:!MD5:@STRENGTH'   # allowed SSL ciphers
                                        # (change requires restart)
#ssl_renegotiation_limit = 512MB        # amount of data between renegotiations
#ssl_cert_file = 'server.crt'           # (change requires restart)
#ssl_key_file = 'server.key'            # (change requires restart)
#ssl_ca_file = ''                       # (change requires restart)
#ssl_crl_file = ''                      # (change requires restart)
#password_encryption = on
#db_user_namespace = off

# Kerberos and GSSAPI
#krb_server_keyfile = ''
#krb_srvname = 'postgres'               # (Kerberos only)
#krb_caseins_users = off

# - TCP Keepalives -
# see "man 7 tcp" for details

#tcp_keepalives_idle = 0                # TCP_KEEPIDLE, in seconds;
                                        # 0 selects the system default
#tcp_keepalives_interval = 0            # TCP_KEEPINTVL, in seconds;
                                        # 0 selects the system default
#tcp_keepalives_count = 0               # TCP_KEEPCNT;
                                        # 0 selects the system default


#------------------------------------------------------------------------------
# RESOURCE USAGE (except WAL)
#------------------------------------------------------------------------------

# - Memory -

#shared_buffers = 32MB                  # min 128kB
shared_buffers = 20000MB
                                        # (change requires restart)
#temp_buffers = 8MB                     # min 800kB
#max_prepared_transactions = 0                  # zero disables the feature
                                        # (change requires restart)
# Note:  Increasing max_prepared_transactions costs ~600 bytes of shared memory
# per transaction slot, plus lock space (see max_locks_per_transaction).
# It is not advisable to set max_prepared_transactions nonzero unless you
# actively intend to use prepared transactions.
#work_mem = 1MB                                 # min 64kB
work_mem = 512MB                                # min 64kB
#maintenance_work_mem = 16MB            # min 1MB
maintenance_work_mem = 512MB            # min 1MB
#max_stack_depth = 2MB                  # min 100kB

# - Disk -
```

```
#temp_file_limit = -1                  # limits per-session temp file space
                                       # in kB, or -1 for no limit

# - Kernel Resource Usage -

#max_files_per_process = 1000          # min 25
                                       # (change requires restart)
shared_preload_libraries = '$libdir/dbms_pipe,$libdir/edb_gen'
                                       # (change requires restart)

# - Cost-Based Vacuum Delay -

#vacuum_cost_delay = 0ms               # 0-100 milliseconds
#vacuum_cost_page_hit = 1              # 0-10000 credits
#vacuum_cost_page_miss = 10            # 0-10000 credits
#vacuum_cost_page_dirty = 20           # 0-10000 credits
#vacuum_cost_limit = 200               # 1-10000 credits

# - Background Writer -

#bgwriter_delay = 200ms                       # 10-10000ms between rounds
#bgwriter_lru_maxpages = 100           # 0-1000 max buffers written/round
#bgwriter_lru_multiplier = 2.0               # 0-10.0 multipler on buffers scanned/round

# - Asynchronous Behavior -

#effective_io_concurrency = 1          # 1-1000; 0 disables prefetching

# - InfiniteCache
#edb_enable_icache = off
#edb_icache_servers = ''  #'host1:port1,host2,ip3:port3,ip4'
#edb_icache_compression_level = 6


#------------------------------------------------------------------------------
# WRITE AHEAD LOG
#------------------------------------------------------------------------------

# - Settings -

#wal_level = minimal                   # minimal, archive, or hot_standby
                                       # (change requires restart)
#fsync = on                            # turns forced synchronization on or off
#synchronous_commit = on               # synchronization level;
synchronous_commit = off               # synchronization level;
                                       # off, local, remote_write, or on
#wal_sync_method = fsync               # the default is the first option
                                       # supported by the operating system:
                                       #   open_datasync
                                       #   fdatasync (default on Linux)
                                       #   fsync
                                       #   fsync_writethrough
                                       #   open_sync
#full_page_writes = on                 # recover from partial page writes
#wal_buffers = -1                      # min 32kB, -1 sets based on shared_buffers
wal_buffers = 16MB                     # min 32kB, -1 sets based on shared_buffers
                                       # (change requires restart)
#wal_writer_delay = 200ms              # 1-10000 milliseconds

#commit_delay = 0                      # range 0-100000, in microseconds
#commit_siblings = 5                   # range 1-1000

# - Checkpoints -

#checkpoint_segments = 3               # in logfile segments, min 1, 16MB each
checkpoint_segments = 300              # in logfile segments, min 1, 16MB each
#checkpoint_timeout = 5min             # range 30s-1h
#checkpoint_completion_target = 0.5    # checkpoint target duration, 0.0 - 1.0
```

```
checkpoint_completion_target = 0.9    # checkpoint target duration, 0.0 - 1.0
#checkpoint_warning = 30s             # 0 disables

# - Archiving -

#archive_mode = off           # allows archiving to be done
                              # (change requires restart)
#archive_command = ''         # command to use to archive a logfile segment
                              # placeholders: %p = path of file to archive
                              #               %f = file name only
                              # e.g. 'test ! -f /mnt/server/archivedir/%f && cp %p
/mnt/server/archivedir/%f'
#archive_timeout = 0          # force a logfile segment switch after this
                              # number of seconds; 0 disables


#------------------------------------------------------------------------------
# REPLICATION
#------------------------------------------------------------------------------

# - Sending Server(s) -

# Set these on the master and on any standby that will send replication data.

#max_wal_senders = 0          # max number of walsender processes
                              # (change requires restart)
#wal_keep_segments = 0        # in logfile segments, 16MB each; 0 disables
#replication_timeout = 60s    # in milliseconds; 0 disables

# - Master Server -

# These settings are ignored on a standby server.

#synchronous_standby_names = ''      # standby servers that provide sync rep
                              # comma-separated list of application_name
                              # from standby(s); '*' = all
#vacuum_defer_cleanup_age = 0 # number of xacts by which cleanup is delayed

# - Standby Servers -

# These settings are ignored on a master server.

#hot_standby = off                   # "on" allows queries during recovery
                                     # (change requires restart)
#max_standby_archive_delay = 30s     # max delay before canceling queries
                                     # when reading WAL from archive;
                                     # -1 allows indefinite delay
#max_standby_streaming_delay = 30s   # max delay before canceling queries
                                     # when reading streaming WAL;
                                     # -1 allows indefinite delay
#wal_receiver_status_interval = 10s  # send replies at least this often
                                     # 0 disables
#hot_standby_feedback = off          # send info from standby to prevent
                                     # query conflicts


#------------------------------------------------------------------------------
# QUERY TUNING
#------------------------------------------------------------------------------

# - Planner Method Configuration -

#enable_bitmapscan = on
#enable_hashagg = on
#enable_hashjoin = on
#enable_indexscan = on
#enable_indexonlyscan = on
#enable_material = on
#enable_mergejoin = on
```

```
#enable_nestloop = on
#enable_seqscan = on
#enable_sort = on
#enable_tidscan = on
#enable_hints = on                     # enable optimizer hints in SQL statements.

# - Planner Cost Constants -

#seq_page_cost = 1.0                   # measured on an arbitrary scale
#random_page_cost = 4.0                      # same scale as above
#cpu_tuple_cost = 0.01                 # same scale as above
#cpu_index_tuple_cost = 0.005          # same scale as above
#cpu_operator_cost = 0.0025            # same scale as above
#effective_cache_size = 128MB
effective_cache_size = 64GB

# - Genetic Query Optimizer -

#geqo = on
#geqo_threshold = 12
#geqo_effort = 5                       # range 1-10
#geqo_pool_size = 0                    # selects default based on effort
#geqo_generations = 0                  # selects default based on effort
#geqo_selection_bias = 2.0             # range 1.5-2.0
#geqo_seed = 0.0                       # range 0.0-1.0

# - Other Planner Options -

#default_statistics_target = 100      # range 1-10000
#constraint_exclusion = partition     # on, off, or partition
#cursor_tuple_fraction = 0.1          # range 0.0-1.0
#from_collapse_limit = 8
#join_collapse_limit = 8              # 1 disables collapsing of explicit
                                      # JOIN clauses


#------------------------------------------------------------------------------
# ERROR REPORTING AND LOGGING
#------------------------------------------------------------------------------

# - Where to Log -

log_destination = 'stderr'            # Valid values are combinations of
                                      # stderr, csvlog, syslog, and eventlog,
                                      # depending on platform.  csvlog
                                      # requires logging_collector to be on.

# This is used when logging to stderr:
#logging_collector = off                     # Enable capturing of stderr and csvlog
                                     # into log files. Required to be on for
                                     # csvlogs.
                                     # (change requires restart)

# These are only used if logging_collector is on:
#log_directory = 'pg_log'             # directory where log files are written,
                                     # can be absolute or relative to PGDATA
#log_filename = 'enterprisedb-%a.log' # log file name pattern,
                                     # can include strftime() escapes
#log_file_mode = 0600                 # creation mode for log files,
                                     # begin with 0 to use octal notation
#log_truncate_on_rotation = on               # If on, an existing log file with the
                                     # same name as the new log file will be
                                     # truncated rather than appended to.
                                     # But such truncation only occurs on
                                     # time-driven rotation, not on restarts
                                     # or size-driven rotation.  Default is
                                     # off, meaning append to existing files
                                     # in all cases.
#log_rotation_age = 1d                # Automatic rotation of logfiles will
```

```
                                        # happen after that time.  0 disables.
#log_rotation_size = 0                  # Automatic rotation of logfiles will
                                        # happen after that much log output.
                                        # 0 disables.

# These are relevant when logging to syslog:
#syslog_facility = 'LOCAL0'
#syslog_ident = 'postgres'

# This is only relevant when logging to eventlog (win32):
#event_source = 'PostgreSQL'

# - When to Log -

#client_min_messages = notice           # values in order of decreasing detail:
                                        #   debug5
                                        #   debug4
                                        #   debug3
                                        #   debug2
                                        #   debug1
                                        #   log
                                        #   notice
                                        #   warning
                                        #   error

#log_min_messages = warning             # values in order of decreasing detail:
                                        #   debug5
                                        #   debug4
                                        #   debug3
                                        #   debug2
                                        #   debug1
                                        #   info
                                        #   notice
                                        #   warning
                                        #   error
                                        #   log
                                        #   fatal
                                        #   panic

#log_min_error_statement = error        # values in order of decreasing detail:
                                        #   debug5
                                        #   debug4
                                        #   debug3
                                        #   debug2
                                        #   debug1
                                        #   info
                                        #   notice
                                        #   warning
                                        #   error
                                        #   log
                                        #   fatal
                                        #   panic (effectively off)

#log_min_duration_statement = -1        # -1 is disabled, 0 logs all statements
                                        # and their durations, > 0 logs only
                                        # statements running at least this number
                                        # of milliseconds


# - What to Log -

#debug_print_parse = off
#debug_print_rewritten = off
#debug_print_plan = off
#debug_pretty_print = on
#log_checkpoints = off
#log_connections = off
#log_disconnections = off
#log_duration = off
```

```
#log_error_verbosity = default              # terse, default, or verbose messages
#log_hostname = off
log_line_prefix = '%t '                     # Use '%t ' to enable log-reading
                                    # features in PEM and pgAdmin
                                    # special values:
                                    #   %a = application name
                                    #   %u = user name
                                    #   %d = database name
                                    #   %r = remote host and port
                                    #   %h = remote host
                                    #   %p = process ID
                                    #   %t = timestamp without milliseconds
                                    #   %m = timestamp with milliseconds
                                    #   %i = command tag
                                    #   %e = SQL state
                                    #   %c = session ID
                                    #   %l = session line number
                                    #   %s = session start timestamp
                                    #   %v = virtual transaction ID
                                    #   %x = transaction ID (0 if none)
                                    #   %q = stop here in non-session
                                    #        processes
                                    #   %% = '%'
                                    # e.g. '<%u%%%d> '
#log_lock_waits = off                # log lock waits >= deadlock_timeout
#log_statement = 'none'                      # none, ddl, mod, all
#log_temp_files = -1                 # log temporary files equal or larger
                                    # than the specified size in kilobytes;
                                    # -1 disables, 0 logs all temp files
log_timezone = 'US/Eastern'


#------------------------------------------------------------------------------
# EDB AUDIT
#------------------------------------------------------------------------------

#edb_audit = 'none'                  # none, csv or xml

# These are only used if edb_audit is not none:
#edb_audit_directory = 'edb_audit'   # Directory where log files are written
                                    # Can be absolute or relative to PGDATA

#edb_audit_filename = 'audit-%Y-%m-%d_%H%M%S' # Audit file name pattern.
                                    # Can include strftime() escapes

#edb_audit_rotation_day = 'every'    # Automatic rotation of logfiles based
                                    # on day of week. none, every, sun,
                                    # mon, tue, wed, thu, fri, sat

#edb_audit_rotation_size = 0         # Automatic rotation of logfiles will
                                    # happen after this many megabytes (MB)
                                    # of log output.  0 to disable.

#edb_audit_rotation_seconds = 0              # Automatic log file rotation will
                                    # happen after this many seconds.

#edb_audit_connect = 'failed'        # none, failed, all
#edb_audit_disconnect = 'none'               # none, all
#edb_audit_statement = 'ddl, error'  # none, dml, ddl, select, error, all


#------------------------------------------------------------------------------
# RUNTIME STATISTICS
#------------------------------------------------------------------------------

# - Query/Index Statistics Collector -

#track_activities = on
#track_counts = on
```

```
#track_io_timing = off
#track_functions = none                        # none, pl, all
#track_activity_query_size = 1024     # (change requires restart)
#update_process_title = on
#stats_temp_directory = 'pg_stat_tmp'


# - Statistics Monitoring -

#log_parser_stats = off
#log_planner_stats = off
#log_executor_stats = off
#log_statement_stats = off


#------------------------------------------------------------------------------
# AUTOVACUUM PARAMETERS
#------------------------------------------------------------------------------

#autovacuum = on                       # Enable autovacuum subprocess?  'on'
                                       # requires track_counts to also be on.
#log_autovacuum_min_duration = -1      # -1 disables, 0 logs all actions and
                                       # their durations, > 0 logs only
                                       # actions running at least this number
                                       # of milliseconds.
#autovacuum_max_workers = 3            # max number of autovacuum subprocesses
                                       # (change requires restart)
#autovacuum_naptime = 1min             # time between autovacuum runs
#autovacuum_vacuum_threshold = 50      # min number of row updates before
                                       # vacuum
#autovacuum_analyze_threshold = 50     # min number of row updates before
                                       # analyze
#autovacuum_vacuum_scale_factor = 0.2# fraction of table size before vacuum
#autovacuum_analyze_scale_factor = 0.1       # fraction of table size before analyze
#autovacuum_freeze_max_age = 200000000       # maximum XID age before forced vacuum
                                       # (change requires restart)
#autovacuum_vacuum_cost_delay = 20ms  # default vacuum cost delay for
                                       # autovacuum, in milliseconds;
                                       # -1 means use vacuum_cost_delay
#autovacuum_vacuum_cost_limit = -1     # default vacuum cost limit for
                                       # autovacuum, -1 means use
                                       # vacuum_cost_limit


#------------------------------------------------------------------------------
# CLIENT CONNECTION DEFAULTS
#------------------------------------------------------------------------------

# - Statement Behavior -

#search_path = '$user,public'          # schema names
#default_tablespace = ''               # a tablespace name, '' uses the default
#temp_tablespaces = ''                 # a list of tablespace names, '' uses
                                       # only default tablespace
#check_function_bodies = on
#default_transaction_isolation = 'read committed'
#default_transaction_read_only = off
#default_transaction_deferrable = off
#session_replication_role = 'origin'
#statement_timeout = 0                 # in milliseconds, 0 is disabled
#vacuum_freeze_min_age = 50000000
#vacuum_freeze_table_age = 150000000
#bytea_output = 'hex'                  # hex, escape
#xmlbinary = 'base64'
#xmloption = 'content'

# - Locale and Formatting -

#datestyle = 'iso, mdy'                       # PostgreSQL default for your locale
```

```
datestyle = 'redwood,show_time'
#intervalstyle = 'postgres'
timezone = 'US/Eastern'
#timezone_abbreviations = 'Default'     # Select the set of available time zone
                                       # abbreviations.  Currently, there are
                                       #   Default
                                       #   Australia
                                       #   India
                                       # You can create your own file in
                                       # share/timezonesets/.
#extra_float_digits = 0                        # min -15, max 3
#client_encoding = sql_ascii           # actually, defaults to database
                                       # encoding

# These settings are initialized by initdb, but they can be changed.
lc_messages = 'en_US.UTF-8'                    # locale for system error message
                                       # strings
lc_monetary = 'en_US.UTF-8'                    # locale for monetary formatting
lc_numeric = 'en_US.UTF-8'                     # locale for number formatting
lc_time = 'en_US.UTF-8'                                # locale for time formatting

# default configuration for text search
default_text_search_config = 'pg_catalog.english'

# - Other Defaults -

#dynamic_library_path = '$libdir'
#local_preload_libraries = ''

#oracle_home =''      # path to the Oracle home directory;
                                       # only used by OCI Dblink; defaults
                                       # to ORACLE_HOME environment variable.

#------------------------------------------------------------------------------
# LOCK MANAGEMENT
#------------------------------------------------------------------------------

#deadlock_timeout = 1s
#max_locks_per_transaction = 64              # min 10
                                       # (change requires restart)
# Note:  Each lock table slot uses ~270 bytes of shared memory, and there are
# max_locks_per_transaction * (max_connections + max_prepared_transactions)
# lock table slots.
#max_pred_locks_per_transaction = 64  # min 10
                                       # (change requires restart)


#------------------------------------------------------------------------------
# VERSION/PLATFORM COMPATIBILITY
#------------------------------------------------------------------------------

# - Previous PostgreSQL Versions -

#array_nulls = on
#backslash_quote = safe_encoding      # on, off, or safe_encoding
#default_with_oids = off
#default_with_rowids = off
#escape_string_warning = on
#lo_compat_privileges = off
#quote_all_identifiers = off
#sql_inheritance = on
#standard_conforming_strings = on
#synchronize_seqscans = on

# - Other Platforms and Clients -

#transform_null_equals = off

# - Oracle compatibility -
```

```
edb_redwood_date = on                # translate DATE to TIMESTAMP(0)
edb_redwood_strings = on             # treat NULL as an empty string in
                                     # string concatenation
#edb_stmt_level_tx = off            # allow continuing on errors instead
                                     # rolling back
db_dialect = 'redwood'               # Sets the precedence of built-in
                                     # namespaces.
                                     # 'redwood' means sys, dbo, pg_catalog
                                     # 'postgres' means pg_catalog, sys, dbo
#optimizer_mode = choose             # Oracle-style optimizer hints.
                                     # choose, all_rows, first_rows,
                                     # first_rows_10, first_rows_100,
                                     # first_rows_1000

#------------------------------------------------------------------------------
# ERROR HANDLING
#------------------------------------------------------------------------------

#exit_on_error = off                 # terminate session on any error?
#restart_after_crash = on            # reinitialize after backend crash?


#------------------------------------------------------------------------------
# CUSTOMIZED OPTIONS
#------------------------------------------------------------------------------

#dbms_pipe.total_message_buffer = 30kB  # default: 30KB, max: 256MB, min: 30KB
#dbms_alert.max_alerts = 100         # default 100, max: 500, min: 0

#------------------------------------------------------------------------
# DYNA-TUNE
#------------------------------------------------------------------------

#edb_dynatune = 66                   # percentage of server resources
                                     # dedicated to database server,
                                     # defaults to 0
#edb_dynatune_profile = mixed        # workload profile for tuning.
                                     # 'oltp', 'reporting' or 'mixed',

#------------------------------------------------------------------------
# QREPLACE
#------------------------------------------------------------------------

#qreplace_function = ''                     # function used by Query Replace.

#------------------------------------------------------------------------
# RUNTIME INSTRUMENTATION AND TRACING
#------------------------------------------------------------------------

timed_statistics = off               # record wait timings, defaults to on

# Add settings for extensions here
```

# Appendix B: Resources

The following websites provide useful references to supplement the information contained in this paper:

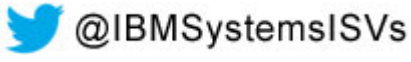
- IBM Linux on Power (all topics)
  **ibm.com**/systems/power/software/linux/resources.html
- IBM Linux Technology Centers
  **ibm.com**/linux/ltc/index.html
- IBM PowerLinux Community
  **ibm.com**/developerworks/community/groups/service/html/communityview?communityUuid=fe313521-2e95-46f2-817d-44a4f27eba32
- IBM Publications Center
  www.elink.ibmlink.**ibm.com**/public/applications/publications/cgibin/pbi.cgi?CTY=US
- IBM developerWorks – A Technical Community for PowerLinux
  **ibm.com**/developerworks/group/tpl
- IBM PowerVM Virtualization Introduction and Configuration
  **ibm.com**/redbooks/abstracts/sg247940.html
- IBM PowerVM Virtualization Managing and Monitoring
  **ibm.com**/redbooks/abstracts/sg247590.html
- IBM developerWorks®
  **ibm.com**/developerworks/linux/
- IBM Java at DeveloperWorks
  **ibm.com**/developerworks/java/jdk/linux
- IBM Techdocs – technical white papers
  **ibm.com**/support/techdocs/atsmastr.nsf/Web/TechDocs
- EnterpriseDB PPAS general information
  www.enterprisedb.com/products-services-training/products/postgres-plus-advanced-server
- EnterpriseDB PPAS manuals address general installations and configuration
  enterprisedb.com/resources-community/tutorials-quickstarts/all-platforms

  Red Hat Enterprise manuals address the system requirements and physical hardware setups
  https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Installation_Guide/pt-install-info-ppc.html

## Acknowledgements

Deepak Narayana (IBM) authored this paper. It was built on research and guidance from Mark Nellen, Paul Clarke, Steven Pratt and Mala Anand.

The authors welcome readers to share their experiences and to provide feedback at: narayana@us.ibm.com (Deepak Narayana).



# Trademarks and special notices



References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

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.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Intel, Intel Inside (logos), MMX, and Pentium are trademarks of Intel 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 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.

Information is provided "AS IS" without warranty of any kind.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local IBM office or IBM authorized reseller for the full text of the specific Statement of Direction.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the

storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.

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