

Tuning Tips for Sybase IQ on IBM Power Systems



This document can be found on the web, www.ibm.com/support/techdocs
Search for document number WP##### under the category of “White Papers.”

Version 01
November 21, 2008

Jointly prepared by IBM and Sybase

Table of Contents

- 1. INTRODUCTION 3
- 2. PURPOSE OF THIS TUNING GUIDE..... 3
 - 2.1. WHAT THIS PAPER IS 3
 - 2.2. WHAT THIS PAPER IS NOT..... 3
- 3. BEST PERFORMANCE PRACTICES..... 3
 - 3.1. AIX BEST PERFORMANCE PRACTICES 4
 - 3.2. SYBASE IQ BEST PERFORMANCE PRACTICES 5
 - 3.2.1. Database Tuning Example for 300GB DB 6
 - 3.3. ADDITIONAL TUNING PARAMETERS TO CONSIDER 6
 - 3.3.1. AIX tunables: 6
 - 3.3.2. Sybase IQ tunables in <servername>.cfg: 7
 - 3.3.3. Sybase IQ database options: 7
 - 3.4. ADDITIONAL REFERENCE MATERIAL 7
- 4. PROOF OF CONCEPT (POC) VS. PRODUCTION ENVIRONMENTS 8
- 5. PROOF-OF-CONCEPT RECOMMENDATIONS 9
- 6. SUMMARY 10
- 7. TRADEMARKS 10

1. Introduction

Sybase® IQ is a highly optimized analytics server designed specifically to deliver fast results for mission-critical business intelligence, data warehouse and reporting solutions on any standard hardware and operating systems.

IBM® Power Systems™ launched POWER6™ processors in 2007, the fastest microprocessor ever built and an ultra-powerful new computer server that leverages the chip's many breakthroughs in energy conservation and virtualization technology.

This paper is the result of a joint effort between Sybase and IBM and provides recommendations and best practice guidelines for optimal performance when deploying Sybase IQ on IBM Power Systems.

2. Purpose of this Tuning Guide

2.1. What this paper IS

The primary purpose of this paper is to provide general guidance and education to technical field personnel for tuning and optimizing the performance of Sybase IQ on IBM Power Systems. It is a collection of tuning tips that have shown to be successful to some degree for a variety of configurations and environments. These best practices and guidelines are intended to provide general guidance to performance tuning Sybase IQ on IBM Power Systems by identifying tunables that have shown to be beneficial in the past.

The best recommendation is to try these variables independently in your specific environment and configuration and measure what performance benefits (if any) are observed.

2.2. What this paper IS NOT

This paper is not a definitive and absolute guide to performance tuning of Sybase IQ on IBM Power Systems. The countless variations of environments and configurations make it nearly impossible to state with complete certainty that any particular tuning tip will always provide a performance improvement in each and every instance.

3. Best Performance Practices

In this section, we will provide an overview of tuning tips which are applicable in enough Sybase IQ and IBM Power Systems environments that they can be recommended without a lot of specific knowledge of the environment. Section 3.1 will cover the parameters available in AIX while section 3.2 will cover Sybase IQ parameters.

Note: There are additional tuning parameters discussed in section 3.3. However, those parameters might benefit in some cases and might not benefit in other cases. They require further experimentation and observation to determine if they should be exploited and what if any benefit they provide.

3.1. AIX Best Performance Practices

Shown below in Table 1 are the VMM page replacement defaults for AIX 5.2/5.3 and AIX 6.1. For most customer configurations running AIX 5.3, we recommend applying the AIX 6.1 defaults (except for **page_steal_method** which is an optional recommendation and it is probably not needed in a Sybase environment).

Table 1 Default values for AIX Virtual Memory Management parameters

AIX 5.2/5.3 defaults	AIX 6.1 defaults
minperm% = 20	minperm% = 3
maxperm% = 80	maxperm% = 90
maxclient% = 80	maxclient% = 90
strict_maxperm = 0	strict_maxperm = 0
strict_maxclient = 1	strict_maxclient = 1
lru_file_repage = 1	lru_file_repage = 0
page_steal_method = 0	page_steal_method = 1

We no longer recommend using low values for **maxclient** and **maxperm** since we introduced the **lru_file_repage** tunable. Turning off the LRU file repage check (**lru_file_repage=0**) allows us to tune VMM page replacement in a way that the system can use up to 90% of its memory for file caching but favors computational pages over file pages. The system will not page to paging space unless the amount of virtual memory exceeds 97% of the size of the real memory (100% - **minperm%**). IBM defined the default values for AIX 6.1 to be the same as the recommended values for AIX 5.3 and most of the tunables are restricted, which means that a customer shouldn't need to change them.

Note: Legacy VMM page replacement tuning recommendations included increasing **minfree** and **maxfree** using a formula to calculate their values based on the number of CPUs and the value for page read ahead. We don't recommend using the legacy formula.

Current recommendation for tuning **minfree** and **maxfree** is to increase both by the same value, for example increase both by 1000, in the case the free list drops to 0.

For multi-threaded applications, the **AIXTHREAD_SCOPE** environment variable should be set to "S" to get one kernel thread per application thread. This environment variable has no impact on single threaded applications.

Although we are presently unaware of any Sybase customers currently experiencing problems with the following two issues, they could potentially be of concern to POWER5 and POWER6 users and should be monitored and addressed as recommended below:

HW	Problem	Work around
POWER6	Power saving mode can cause an impact on response time sensitive applications running on mainly idle systems	Turn off smt_snooze_delay on dedicated partition.
POWER5/ POWER6	Possible performance impact due to IERAT misses when changing from 4KB page size to 64KB page size and back. Example: Application running with 4KB page size is calling shared library subroutines with 64KB page size	Preferred: Run the application using 64KB page size Alternative: Disable multiple page size support

3.2. Sybase IQ Best Performance Practices

This section talks about some of the Sybase IQ tuning parameters which help achieve the best performance in most of the cases. It is important and assumed that the initial physical database design and storage configuration is completed for the optimal performance.

- **Main Buffer Cache:** This memory segment stores all user and static data. This is the location where data that resides in tables is cached for queries and loads. This is set via **-iqmc <value>** in the config file. The value is in Megabytes. For any active database, the default main buffer cache size of 16MB is too low. For optimal performance, allocate as much memory as possible to the IQ main buffer cache. Depending on the availability of the total physical memory available, this configuration needs to be supplied. General guideline is to allocate 40% of total memory dedicated to IQ.
- **Temp Buffer Cache:** This is the memory segment that will cache all volatile and temporary data during loads and queries. This is also a workspace for the IQ engine when it needs memory for sorting or joins algorithms. This is set via **-iqtc <value>** in the config file. The value is in Megabytes. For any active database, the default temp buffer cache size of 8MB is too low. For optimal performance, allocate as much memory as possible to the IQ temp buffer cache. General guideline is to allocate 60% of the total memory dedicated to IQ.
- **Number of Threads for Sort:** This specifies the number of threads to be used in a sort. Use this option for performance analysis and tuning. If you change this option, experiment to find the best value to increase performance, as choosing the wrong value might decrease performance. For large IQ temporary buffer cache size, Sybase recommends setting the **SORT_PHASE1_HELPERS** option between 5 and 10.
- **Number of CPUs:** While the Sybase IQ engine automatically decides the number of CPUs available on the system, it is preferred to provide that input using **-iqnumbercpus** parameter. This is very helpful for cases where other applications may share the same system or where actual numbers of “physical CPUs” are not same as the number of logical CPUs visible to IQ due to hyper threading, SMT or CMT technology.

3.2.1. Database Tuning Example for 300GB DB

The following chart in Table 2 shows the Sybase IQ server configuration parameters and database options used for a set of ad hoc, decision support tests run on IBM Power Systems with a 300GB database. Initially **-iqnumbercpus** was set to equal the same number of cores in the system but we found that in the SMT environment the default value (2*number of cores) had advantages. The database option **'Sort_Phase1_helpers'** (number of threads to be used in a sort) had notable performance improvement when set to a higher value than the default value of 3.

Table 2 Sybase IQ Server Configuration Parameters Used for Ad Hoc Tests

Configuration parameters	Database options
-c 32M	SET OPTION "PUBLIC".Allow_Nulls_By_Default='Off';
-gd all	SET OPTION "PUBLIC".Append_Load='On';
-gm 25	SET OPTION "PUBLIC".Flatten_Subqueries = 'On';
-gc 5000	SET OPTION "PUBLIC".Force_No_Scroll_Cursors='On';
-gr 5000	SET OPTION "PUBLIC".Garray_Fill_Factor_Percent=2;
-gp 4096	SET OPTION "PUBLIC".Load_Memory_Mb=0;
-tl 0	SET OPTION "PUBLIC".Max_IQ_Threads_Per_Connection=100;
-iqmt 1600	SET OPTION "PUBLIC".Minimize_Storage='On';
-iqmc 9000	SET OPTION "PUBLIC".Notify_Modulus=10000000;
-iqtc 5000	SET OPTION "PUBLIC".Query_Temp_Space_Limit=0;
-iqpartition 4	SET OPTION "PUBLIC".Row_Counts='On';
-iqgovern 10	SET OPTION "PUBLIC".Sort_Phase1_Helpers=5;
	SET OPTION "PUBLIC".Sweeper_Threads_Percent=8;
	SET OPTION "PUBLIC".Wash_Area_Buffers_Percent = '20';
	SET OPTION "PUBLIC".Prefetch_Threads_Percent = 15;
	SET OPTION "PUBLIC".Max_Hash_Rows=10000000;
	SET OPTION "PUBLIC".Default_Having_Selectivity=1;
	SET OPTION "PUBLIC".Hash_Thrashing_Percent=100;

3.3. Additional Tuning Parameters to Consider

The OS and Sybase IQ tuning parameters outlined above conform to IBM and Sybase best practices for Sybase IQ on AIX. Limited testing and experimentation by Sybase and IBM field specialists have identified some variants which may help to improve performance on a case-by-case basis. The most promising ones are as follows:

3.3.1. AIX tunables:

- Consider creating or mounting a dedicated Sybase IQ bulk load file system with the mount option **rbr** (release behind read) to avoid caching pages only read once.
- Consider increasing the **ioo** parameter **j2_maxPageReadAhead**
- In shared pool mode, consider setting schedo parameter **vpm_xvcpus=-1** to disable virtual processor folding.

- In environments utilizing POWER processors with 8-cores or more, consider setting the following tunables that have shown improvements in other similar environments.

```

YIELDLOOPTIME=0
SPINLOOPTIME=500
MALLOCOPTIONS=watson,multiheap:32,considersize
AIXTHREAD_SCOPE="S"
AIXTHREAD_MNRATIO="1:1"
AIXTHREAD_MUTEX_FAST=ON

```

In addition to the environment variables above, you can turn off krlock with **schedo -o krlock_enable=0**

- While the general recommendation is setting **AIXTHREAD_SCOPE="S"** (which sets **AIXTHREAD_MNRATIO=1:1**), an experiment which might be worth trying is to export **AIXTHREAD_MNRATIO=2:1**. Modify the default value in `$ASDIR/bin/start_asiq` of **AIXTHREAD_MNRATIO=4:1** and comment out **AIXTHREAD_SCOPE="S"**.

3.3.2. Sybase IQ tunables in <servername>.cfg:

- Consider setting **-iqnumbercpus=<number of physical processors>**. This parameter defaults to <number of logical processors> but restricting it to <number of physical processors> has often been beneficial in the field.
- **-iqmt** should be defaulted unless high values are used for database options **Max_IQ_Threads_Per_Team** and **Max_IQ_Threads_Per_Connection**.
- **-iqtss** governs the amount of stack memory per thread. Sybase IQ documentation recommends a value of 1024 for 64 bit system

3.3.3. Sybase IQ database options:

- **Max_IQ_Threads_Per_Team** and **Max_IQ_Threads_Per_Connection** are often tuned specifically for a customer application. This benchmark used a value of **Max_IQ_Threads_Per_Connection=100**. Consider a value of 75 for each.

Sort_Phase1_Helpers defaults to 3. Sybase documentation recommends increasing the value to 5 or more for systems with large memory. A value of 5 was beneficial to this benchmark.

3.4. Additional Reference Material

- "[Hardware Sizing Guide for Sybase IQ 12.6 and 12.7](#)" by Mark Mumy, Sybase. This contains valuable recommendations on memory sizing, main store to temp store ratio, LUNs per CPU ratio and much more.

- “Sybase BI for System P: IQ and the Risk Analytics Platform” by Peter Barnett – presented at STG 2008, contains a how-to for setting up, loading, tuning and running Sybase IQ on AIX and Linux for Power.
- “Sybase IQ version 12.7 Query Engine Internals” by David Walrath, Sybase – a deep dive into the IQ Optimizer, index selection and SQL tuning.

4. Proof of Concept (POC) vs. Production Environments

Across the industry, designers optimize computer systems via micro architecture decisions, trading off area and power for performance-enhancing features. As a result, for a POC to accurately represent the performance of the customer’s true production environment, it is imperative that the POCs be structured to stress the computer system in a similar manner to the customer’s production environment – benefiting from the strengths of a given micro architecture and suffering from its weaknesses. With respect to POWER5 systems and POWER6 systems, there are several areas where the IBM systems provide significant performance leverage often not available on the competitive systems. For Sybase users, two of areas of interest are SMT (Simultaneous Multi-threading) and large off-chip caches (32-36MB). Each of these features translates to POC and production environment characteristics which must be leveraged to enable the POC to accurately predict the performance that the customer will experience in their production environment.

To fully leverage the benefit of large off-chip caches, the primary high-level concern is that the POC must include data similar to the customer’s typical production environment with respect to 1) size of the database, 2) size of the tables, 3) number of tables, and 4) data accessing patterns for the tables.

SMT on POWER5 and POWER6 processors is implemented with additional hardware (duplicate register files, added control logic and queues) that allows a single physical core to concurrently execute instructions from more than one thread – and therefore appear as multiple “logical” processors. Therefore, simulating an accurate thread count is critical to evaluating performance in an SMT-capable processor. The primary areas of interest here are the complexity of the queries and the number of concurrent queries. It is worthwhile to note that “single query” does not imply “single-thread”. Sybase IQ may create multiple threads of work for a single query depending on the number of tables, the complexity of the query, and the predicates used. The average number of active queries further increases the number of concurrent threads. The number of concurrent threads will further increase in environments where a single OS has multiple applications running. Finally, for production environments which include virtualization supporting multiple partitions, hardware thread exploitation is further increased. While some POCs are structured to run several queries sequentially, in general, true production environments rarely purchase a 4-way to 16-way system to run one query in isolation.

Prior to the invention of SMT, a first-order approximation of the performance of an N-way system might be to assume roughly N times the performance of a single-threaded test. Therefore, historically, single-threaded tests may have been a reasonable, simplified, and quick performance indicator. As shown in Figure 1 below, for a single-core system without SMT, there is little performance difference between the time required to run two single-threaded queries sequentially (upper diagram) and the time to run the two queries concurrently, allowing the operating system to “time slice” between the two queries (middle diagram). However, with the advent of SMT which allows both queries to run concurrently on a single core, running each test sequentially (upper diagram) dramatically understates the system performance one

would observe if instead multiple threads of work per core were allowed as in a real production environment. Therefore, for POC tests to accurately predict production-environment performance, concurrent thread count must be approximated (including number of concurrent queries, complexity of queries, multiple users, and number of virtual partitions). This is becoming increasingly important as the number of cores per chip increases, since there is often adequate bandwidth to support a single core but there may not be sufficient bandwidth to support all the cores on a chip. In these cases, even single-core tests are not good indicators of N-way performance and accurate system performance projections require POC tests which reflect the stress of the expected concurrency level and bandwidth requirements.

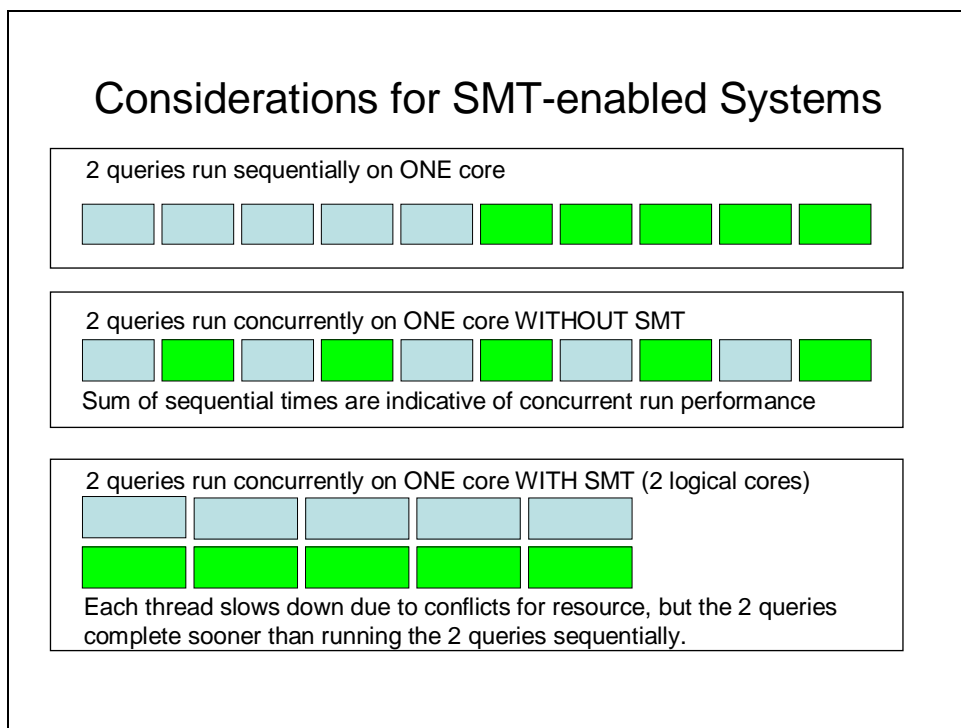


Figure 1 Why sequentially run tests may not accurately predict the actual production environment

5. Proof-of-Concept Recommendations

Our recommendations for Sybase IQ and IBM Power Systems POC engagements are:

- Before the POC is initiated, work with the customer to ensure that the POC tests are representative of the expected production environment with respect to size of the database, types of queries, number of concurrent queries, number of users, and types of virtualization. Locating the POWER6 system at the customer site will often facilitate exploitation of real customer data in the POC as typically much of the customer's data is sensitive.
- Since the leverage of the enhanced SMT capability of POWER6 increases with processor utilization, POCs which are structured such that the utilization is moderate to high will benefit a POWER6 system more than it benefits the competitive hardware. If possible, the utilization level of the existing system (and the differences between the existing system and the new system) should be used to predict the utilization of the new system.

- Simple single-attribute characterizations of computer systems, such as clock rate, are not sufficient to accurately predict performance. It is recommended to use the appropriate sizing metrics (such as the IBM rPerf metric) when setting performance expectations.
- While the POC is ongoing, start with the parameters specified in section 3.1, experiment with the tuning parameters in section 5.0, and then experiment with parameters in Appendix A, starting with the parameters which have shown benefits in similar POCs.
- Consult the reference material in Appendix B.
- As the POC progresses, if performance expectations appear to be in jeopardy, it is better to get help earlier while there is time to get others involved and more analyses and experiments can be performed.
- Sybase IQ has a strong preference for raw devices and the current recommendation is to use raw LUNs. These can be raw disk devices (hdisks in AIX or /dev/raw/rawxx in Linux) or raw logical volumes. Sybase recommends that disk striping size at the SAN level be in multiples of the Sybase IQ database page size. Sybase also recommends dedicated partitions for the main and temporary dbspaces and not to define multiple partitions for the same physical disks.

6. Summary

Further optimization tests and experiments continue and are expected to yield subsequent improvements. These results will be provided in future versions of this report or an alternatively suitable whitepaper.

7. Trademarks

Sybase is a registered trademark of Sybase Inc. in the United States, other countries or both.

IBM, the IBM logo, AIX, Power Architecture, Power Systems, POWER and System Storage 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

Sun, and Sun Fire are trademarks of Sun Microsystems, Inc.

AMD Opteron is a trademark of Advanced Micro Devices

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

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

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