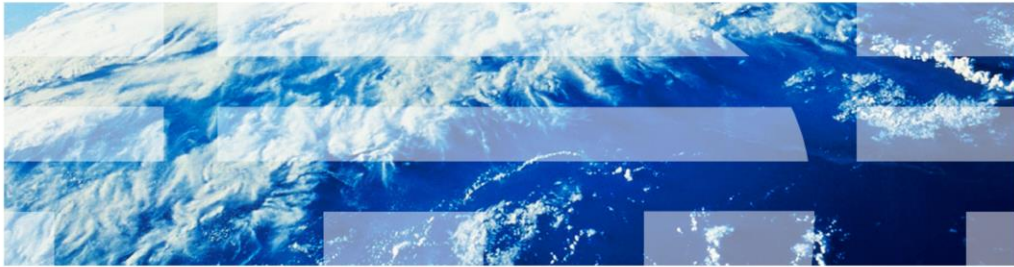


IBM InfoSphere Master Data Management Collaboration Server 10.0.0

Maintenance tasks to ensure optimum performance



© 2012 IBM Corporation

The purpose of this presentation is to review all the essential tasks expected of the IBM InfoSphere® Master Data Management Collaboration Server administrator to ensure optimum performance, good user experience and provide continuity of operation. This topic is important because it will help you maintain the product more effectively and with greater efficiency.

Background

- Presentation objective
 - Overview of expected maintenance tasks
- Terminology
 - TOP
- Not discussed
 - Comprehensive guide
 - Troubleshooting issues
- Assumption
 - Availability of database and network administrator
 - Understanding of product architecture
- Architecture
 - http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.ovr.doc/pim_con_archfo.html

This presentation is meant to provide an overview of the essential maintenance tasks expected of the product administrator.

TOP is a term unique to the product and is used to reference an environment variable which points at the installation directory of the application.

This presentation is not meant to be used as a comprehensive guide or manual, see the Information Center for those details. This is also not a presentation that teaches you how to troubleshoot performance issues. Rather, it should serve as a maintenance checklist to prevent performance related problems.

Note that some of the tasks in this presentation may require collaboration with the database and network administrator. This presentation was prepared assuming that you have an understanding of the product architecture. Follow the link displayed on this slide for details on the product architecture.

Environment setup (1 of 3)

- Database setup
 - DB2®

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.ins.doc/pim_con_ibm_db2_udb_database_options.html

- Oracle

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.ins.doc/pim_con_setting_up_oracle.html

- Database setup checklist
 - \$TOP/bin/db_checklist/*checklist.sql
- Use optimal database server configuration
 - Configure key variables
- Batch processing parameters

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.cof.doc/properties/pim_con_cp_batchprocessing.html

In order for the application to run smoothly, it is paramount that the environment has been set up properly, especially the database. Review the Information Center links displayed on this slide with your database administrator to ensure that all the necessary tasks have been completed. You can use the scripts in the db_checklist directory to verify whether the setup is correct.

To ensure better response time, output, and availability of the server, configure key variables such as your database lock wait, memory, and buffer pool parameters so as to optimize your hardware capabilities. Here, lock wait specifies the amount of time that an application will wait before returning without a lock and the buffer pools should be set high enough to avoid unnecessary data reads from the table spaces.

In addition, there are batch processing parameters that can cause a performance bottleneck. Go to the link displayed on this slide for batch processing parameters.

Environment setup (2 of 3)

- Application server and database connection
 - Latency
 - Gigabit Ethernet connection
- Cache settings
 - Maintain high cache hit ratio, run
 - \$TOP/bin/compat.sh
 - \$JAVA_RT com.ibm.ccd.common.wpcsupport.util.SupportUtil –
cmd=getRunTimeCacheDetails
- Cache property files

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.cof.doc/properties/pim_con_cp_cachepropertyfile.html

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.cof.doc/properties/pim_con_cp_memorycache.html

The connection between the application server and the database is integral to the product. The connection latency should be under point two five milliseconds and use a gigabit Ethernet connection capable of transferring large files at approximately 25 megabytes per second by way of secure file transfer protocol.

The physical input output operations are very resource intensive and so keeping frequently accessed objects in the buffer cache is a primary technique to improve performance. The effectiveness of cache is determined by how frequently the cached objects are being accessed. Low cache hit ratio may result in performance degradation. To monitor the cache hit ratio, use the command displayed on this slide. Follow the Information Center links on this slide to learn more about caching and cache property files. Optimizing these parameters is significant to achieve good performance.

Environment setup (3 of 3)

- Central processing unit
 - At least eight GB RAM
 - Use multi-processor systems
- Monitor processor usage
 - Clustering

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.adm.doc/sys_admin/pim_con_clusteringadmin.html

- Monitoring tools: vmstat, sar, mpstat, ps/top
- Nmon
 - http://www.ibm.com/developerworks/aix/library/au-analyze_aix/

The application requires a lot of processing power. Use at least eight gigabytes of memory with a dual-processor dual-core system for peak load handling. Multiprocessor systems will improve performance significantly as well.

The processor usage should be regularly monitored during peak usage hours. If the usage is close to system capacity, consider moving to a server with more processors or set up horizontal clustering. Refer to the Information Center link displayed on this slide for an overview of clustering.

This slide also displays some of the available tools to help you monitor the processor usage. You can use any of them or choose a different one. The Nmon tool is designed for AIX and Linux performance specialists to use for monitoring and analyzing performance data.

Maintenance tasks (1 of 2)

- Database statistics
 - Oracle

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.adm.doc/data_admin/pim_tsk_reorgstatsoracle.html

- DB2

<http://www.ibm.com/developerworks/data/library/techarticle/dm-0706tang>

- \$STOP/bin/db/analyze_schema.sh

- Database defragmentation
 - DB2 (two-step process)
 - Re-organization: db2 reorg table table_name
 - db2 reorg indexes all for table table_name allow write access
 - Oracle
 - Reorganization: alter table table_name move
 - alter index index_name rebuild

The database optimizer requires statistical metadata information about its user tables to choose the most efficient way to access data. The required information includes number of rows and cardinality. The database administrator must manage and organize tables and indexes of your database schema to find and avoid potential problems or issues. You should set up automatic statistic gathering but you can use `analyze_schema.sh` to do so as well. Use the Information Center link displayed on this slide for details on Oracle statistics gathering and a DeveloperWorks article on setting up automatic statistics gathering in DB2.

Reorganization is used to store data into un-fragmented, physically contiguous spaces. This helps to save space and retrieve data efficiently. In DB2, run the `REORGCHK` command to verify the need for reorganization. Reorganization is recommended for the tables and indexes that contain the asterisk character in the `REORG` column in the output of the `REORGCHK` command. The command to reorganize a table or index is displayed on this slide.

Maintenance tasks (2 of 2)

- Data maintenance

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.adm.doc/sys_admin/pim_con_datamaintenance.html

- Old data version

- \$TOP/src/maintenance/delete_old_versions.sh
- Partitioning

- DeveloperWorks

<http://www.ibm.com/developerworks/data/library/techarticle/dm-0605ahuja2/>

- Technote(v-6)

<http://www-01.ibm.com/support/docview.wss?uid=swg27018196>

- Temporary tables and indexes

- \$TOP/src/db/schema/dbscripts/db2/drop_temp_agg_tables.pl
- \$TOP/src/db/schema/dbscripts/db2/drop_log_tables.pl

The product contains some report jobs to estimate, delete, or archive obsolete or unused data. Visit the Information Center link displayed on this slide to learn more about these reports.

Versioning is a backup feature for all objects so that you can roll back your objects to a specific object version from within the user interface if you encounter problems. The object version handling process creates duplicate entries in your respective database tables for every new version of your objects. Hence, an item may have multiple older versions which are not used by the product. But SQL queries will still loop through these versions creating a performance bottleneck. You should delete these older versions regularly using `delete_old_versions.sh`. If you do not want to delete these older versions, you may instead set up a separate partition for these older versions. Refer to the technote or DB2 DeveloperWorks article link displayed on this slide to learn more about setting up partitioning.

Daily activities like import and update jobs or custom scripts may create temporary tables in the database. So, ensure these tables are removed and one way of doing so is the `drop_temp_agg_tables.pl` perl script. Similarly, all temporary user log tables should be dropped as well, starting with `TCTG_ULE_USER_LOG_ENTRY` and associated temporary sequences using the `drop_log_tables_seq.pl` perl script. There is a separate directory for Oracle's scripts within the `dbscripts` directory.

Monitoring(1 of 3)

- Database
 - DB2 logs: DB2_instance/sqllib/db2dump
 - Oracle logs: ORACLE_BASE/admin/Oracle_SID/directory
 - DB2 table space: List table spaces show detail
- Server resources
http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.trb.doc/pim_con_ts_performanceslowness.html
- Performance profiling data (if enabled)
http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.adm.doc/sys_admin/pim_con_performanceprofile.html

The database should be monitored regularly (especially during peak load) to ensure better performance and availability. The ongoing tasks include monitoring the database for errors, monitoring usage patterns, whether there is enough free disk and table space, and so on. See the links displayed on this slide for the location of database log files and the command for checking the table space in DB2.

You should also monitor the server resources like processor usage, memory swapping, disk input output, network input output, and so on. Use the Information Center link displayed on this slide for troubleshooting system wide slowness. Slowness may be caused by hitting the physical resource limitation that needs to be fixed by adding processor resources or it may be due to some implementation bottleneck.

If performance profiling was used for performance analysis, then the data should be periodically purged upon completion of the analysis. Go to the performance profiling line displayed on this slide for more information.

Monitoring (2 of 3)

- Slow queries
 - profiling_query_delay_threshold
 - \$TOP/logs/service_Name/db.log
 - “Explain plan”, checking statistics, detecting rollback, and deadlocks

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.trb.doc/pim_con_ts_performance.html

 - Open Service Request

<https://www-946.ibm.com/support/servicerequest/>
- Periodic database health check
 - DB2: Use “Health Center” to set-up warnings and alerts
 - Oracle: Alerts, Automatic Workload Repository (AWR) reports, and Oracle Enterprise Manager

In the \$TOP/etc/default/common.properties, there is a data profiling parameter named profiling_query_delay_threshold. This variable defines the maximum time in milliseconds that a query can take without getting logged as a delayed query in the \$TOP/logs/service_Name/db.log file of the particular service where this query originated. The application administrators should regularly check the db.log file for each service to detect delayed queries. The administrator should be vigilant to identify queries which are taking an extremely long time or if there are a lot of similar queries which are significantly delayed. After finding them, work with your database administrator to troubleshoot the probable cause. You should check whether statistics are up to date, monitor for rollback of a large transaction, run “Explain Plan” utility, and look for deadlocks. Refer to the Information Center for details on how to do this. If you still cannot determine the cause of a very slow query, open a Service Request ticket with IBM Support. Note that individual entries are not significant; you should look for a pattern of queries which are delayed. IBM and Oracle offer useful tools for database maintenance and you should use these tools. IBM offers DB2 Health Center to monitor the state of a set of health indicators by defining the warning and threshold values for the indicators. Similarly, Oracle provides features like Oracle Alerts, Automatic Workload Repository reports, referred to as AWR, and Oracle Enterprise Manager for Database.

Monitoring (3 of 3)

- System memory settings
 - \$TOP/bin/conf/service_mem_settings.ini

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.adm.doc/sys_admin/pim_tsk_configuringmemoryflags.html
- Tracking memory usage
 - \$TOP/bin/pimSupport.sh (health check option)

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.ssc.doc/scripts/pim_ref_pimsupport_script.html

 - Enable verbose garbage collection and use IBM monitoring and diagnostic tool for Java

<http://www.ibm.com/developerworks/java/jdk/tools/gcmv/>

 - Tuning WebSphere Application Server process

http://www.ibm.com/developerworks/websphere/techjournal/0909_blythe/0909_blythe.html
- Adequate number of open descriptors

IBM InfoSphere Master Data Management Collaboration Server runs using six key services, each of which run on their own Java Virtual Machine, or JVM. The memory setting for these depend on your hardware and are not optimized by default. Review the settings in the `service_mem_settings.ini` and make necessary changes according to your environment. A very high value may lead to wasted hardware resources while a low value may hinder performance.

To track memory usage for each service over time, there are several tools that you may use. First, there is the health check option of `pimSupport.sh`. For a more dynamic analysis, you may enable verbose garbage collection and use “IBM Monitoring and Diagnostic Tools for Java - Garbage Collection and Memory Visualizer” to see average memory usage of the services JVM. Review the link displayed on this slide for details on general WebSphere Application Server tuning processes.

Unexpected problems can be avoided by checking the number of open descriptors and verifying that they are set according to WebSphere Application Server guidelines to 8000. The number of open descriptors can be checked using the `ulimit -a` command on most machines.

Application continuity

- File system
 - Regular full backup of \$TOP
- Database schema
 - Exporting DB dump
 - <http://www-01.ibm.com/support/docview.wss?uid=swg21303356>
 - DB2
 - \$TOP/src/db/scripts/backup/db2_export.sh
 - \$TOP/src/db/scripts/backup/db2_fast_import.sh
- Configure load balancer
 - Allow for fail-over

Data in the IBM InfoSphere Master Data Management for Collaboration Server can be rendered unusable because of hardware or supporting software failure. To prevent the potential loss of data, set up a backup and recovery plan, which includes scheduling regular full backups and a daily incremental back up of the \$TOP directory.

Regular back up of the database schema is extremely important. This can be done using several techniques like exports or mirroring. Online backups allowing point in time recovery are recommended. You can also follow the guidelines in the technote displayed on this slide to periodically export your database dump and keep it as a backup copy in case of emergency.

Certain DB2 scripts are provided for exporting the database; db2_export.sh and db2_fast_import.sh. These scripts are bundled with the product. DB2_export.sh will create a .tar file containing the database schema, while db2_fast_import.sh will import the subsequent .tar file. There are not bundled scripts for Oracle. Follow the guidelines in the technote for more information.

An efficient way to address potential overloading of the application server is to use clustered application servers and then balance the load using a load balancer. Consult your network administrator or IBM Services to deploy the load balancer. Note that if you use clustering for the scheduler, they will load balance themselves automatically.

Additional resources

- Information Center

<http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.help.ic.doc/out/ic-homepage.html>

- Direct link to performance section

http://publib.boulder.ibm.com/infocenter/mdm/v10r0m0/topic/com.ibm.pim.per.doc/pim_con_performanceplanning.html

For additional resources, see the links displayed on this slide.

Trademarks, disclaimer, and copyright information

IBM, the IBM logo, ibm.com, AIX, DB2, DB, InfoSphere, and WebSphere are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "[Copyright and trademark information](http://www.ibm.com/legal/copytrade.shtml)" at <http://www.ibm.com/legal/copytrade.shtml>

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

© Copyright International Business Machines Corporation 2012. All rights reserved.