



Informatica™

Informatica® Multidomain MDM  
10.4 HotFix 1

# Performance Tuning Guide

Informatica Multidomain MDM Performance Tuning Guide  
10.4 HotFix 1  
September 2020

© Copyright Informatica LLC 2016, 2021

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at [infa\\_documentation@informatica.com](mailto:infa_documentation@informatica.com).

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2021-10-28

# Table of Contents

<b>Preface</b> .....	<b>4</b>
Informatica Resources. ....	4
Informatica Network. ....	4
Informatica Knowledge Base. ....	4
Informatica Documentation. ....	4
Informatica Product Availability Matrices. ....	5
Informatica Velocity. ....	5
Informatica Marketplace. ....	5
Informatica Global Customer Support. ....	5
<b>Chapter 1: Introduction to Performance Tuning</b> .....	<b>6</b>
Overview. ....	6
Factors that Influence the Performance of the MDM Hub. ....	6
Acronyms. ....	7
<b>Chapter 2: Recommendations</b> .....	<b>9</b>
Overview. ....	9
Recommendations for Java. ....	9
Recommendations for Virtual Image Environment. ....	12
Recommendations for Database. ....	13
Recommendations for Oracle Database. ....	15
INIT.ORA Recommendations for Oracle. ....	17
Recommendations for Microsoft SQL Server. ....	20
Recommendations for IBM Db2. ....	21
Recommendations for the MDM Servers. ....	25
Recommendations for the MDM Hub. ....	26
Recommendations for Batch Job Optimization. ....	41
Multithreaded Batch Job – Process Flow. ....	49
Recommendations for the Hub Console Optimization. ....	51
Recommendations for Data Director and SIF or BES API Optimization. ....	52
Recommendations for Elasticsearch. ....	57
Recommendations for Customer 360. ....	58
Recommendations for Environment Validation Tools and Utilities. ....	59
<b>Appendix A: Glossary</b> .....	<b>60</b>
<b>Index</b> .....	<b>62</b>

# Preface

See the Informatica® *Multidomain MDM Performance Tuning Guide* to learn how to optimize the overall performance of Multidomain MDM within the database and the application server environments.

## Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

### Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <https://network.informatica.com>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

### Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at [KB\\_Feedback@informatica.com](mailto:KB_Feedback@informatica.com).

### Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at [infa\\_documentation@informatica.com](mailto:infa_documentation@informatica.com).

## Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

## Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <http://velocity.informatica.com>. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at [ips@informatica.com](mailto:ips@informatica.com).

## Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

## Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

To find online support resources on the Informatica Network, visit <https://network.informatica.com> and select the eSupport option.

# CHAPTER 1

## Introduction to Performance Tuning

This chapter includes the following topics:

- [Overview, 6](#)
- [Factors that Influence the Performance of the MDM Hub, 6](#)
- [Acronyms, 7](#)

### Overview

You can use the performance tuning recommendations to configure specific parameters that optimize the performance of the MDM Hub. Experiment with these parameters to arrive at optimum values. You can also achieve a baseline performance by using these recommendations.

### Factors that Influence the Performance of the MDM Hub

The following table lists the different components that you can fine-tune to optimize the performance of the MDM Hub:

Components	Description
Java	Parameters that you can fine-tune in the Java layer. These parameters are applicable for the MDM Hub and for the Process Servers, including clustered environments.
General Database Recommendations	Parameters that you can fine-tune in the database layer. These parameters are related to sizing and storage.
Number of CPUs	Calculate this number as a number of simultaneously promoted ORSs +1.
Oracle	Parameters that you can fine-tune in the Oracle database, including the RAC environment. These parameters are applicable for both the MDM Hub Master Database schema and any Operational Reference Store (ORS) schema.

Components	Description
Microsoft SQL Server	Parameters that you can fine-tune in the Microsoft SQL Server. These parameters are applicable for both the MDM Hub Master Database schema and any ORS schema.
IBM DB2	Parameters that you can fine-tune in the IBM Db2 database. These parameters are applicable for both the MDM Hub Master Database schema and any ORS schema.
General MDM Server Recommendations	Recommendations related to the MDM Hub, the Process Server, and the Elasticsearch server environment.
The MDM Hub Specific Configuration	Parameters that you can configure in the MDM Hub. These parameters are applicable for both the MDM Hub Master Database settings and any ORS specific settings.
Batch Job Optimization	Parameters that you can configure for the better performance of the batch jobs.
Hub Console Optimization	Parameters that you can configure for the better performance of the Hub Console.
Informatica Data Director and Services Integration Framework or Business Entity Services Optimization	Parameters that you can configure for the better performance of Informatica Data Director with business entities or subject areas and Services Integration Framework (SIF) or Business Entity Services.
Environment Validation Tools and Utilities	List of tools and utilities that you can use to verify the current environment to identify the area of improvement.

## Acronyms

The following table lists the different acronyms used in this guide:

Term	Definition
BES	Business Entity Services
BO	Base Object
EJB	Enterprise Java Beans
GC	Garbage Collection
HM	Hierarchy Manager
JVM	Java Virtual Machine
MDM	Informatica Master Data Management
PAM	Product Availability Matrix
PGA	Program Global Area (Oracle database)
RAC	Real Application Cluster

<b>Term</b>	<b>Definition</b>
RAID	Redundant Array of Independent Disks
REST	Representational state transfer
RMAN	Recovery Manager
SAN	Storage Area Network
SGA	System Global Area
SIF	Services Integration Framework
SOAP	Simple Access Object Protocol
SSD	Solid State Drive



## CHAPTER 2

# Recommendations

This chapter includes the following topics:

- [Overview, 9](#)
- [Recommendations for Java, 9](#)
- [Recommendations for Virtual Image Environment, 12](#)
- [Recommendations for Database, 13](#)
- [Recommendations for Oracle Database, 15](#)
- [Recommendations for Microsoft SQL Server, 20](#)
- [Recommendations for IBM Db2, 21](#)
- [Recommendations for the MDM Servers, 25](#)
- [Recommendations for the MDM Hub, 26](#)
- [Recommendations for Batch Job Optimization, 41](#)
- [Recommendations for the Hub Console Optimization, 51](#)
- [Recommendations for Data Director and SIF or BES API Optimization, 52](#)
- [Recommendations for Elasticsearch, 57](#)
- [Recommendations for Customer 360, 58](#)
- [Recommendations for Environment Validation Tools and Utilities, 59](#)

## Overview

The recommendations are based on the regular volumes of data with standard hardware. For larger systems, you can adjust the settings accordingly. You can configure parameters related to Java, Oracle database, Microsoft SQL server, IBM Db2, and the MDM Hub to optimize the performance of the MDM Hub.

## Recommendations for Java

You can configure the parameters related to the Java Virtual Machine (JVM) settings and database connection pool. You can change or fine-tune the Java parameters to improve the MDM Hub performance.

Definition of these parameters might vary for each release. You must check the appropriate ergonomics document, Java guidelines, and recommendations.

The following table lists the recommendations for the JVM settings:

Parameter	Recommended Setting	Description
JVM random number generation	<code>-Djava.security.egd=file:/dev/./urandom</code>	Applicable only for the UNIX-based environments. When you first log in to Data Director and perform a search, it might run slower without the property. See the Informatica Knowledge Base article number 523036 for further details.
Elasticsearch server to client connection	<code>-Dmdm.es.index.client.ttl=86400</code> Default is 7200 seconds.	Number of seconds for the application server to wait before it closes the Elasticsearch server stale connection. Informatica recommends that you use 86400 seconds.
Heap size	<code>-Xms2048m</code> <code>-Xmx8192m</code>	In an environment that uses more SIF or business entity services calls, maintain the maximum heap size equal to or greater than 8 GB. If the number of Operational Reference Stores is higher irrespective of active or obsolete, you must allocate more heap memory. If you require memory, add another JVM with a new node.
Metaspace size	<code>-XX:MaxMetaspaceSize=3G</code>	Maximum amount of native memory used for class metadata of JVM.
Code cache	On Oracle or Azul Zulu JRE: <code>-XX:ReservedCodeCacheSize=512m</code> On IBM JRE: <code>-XX:codecachetotal=512m</code>	Maximum amount of memory for the JIT code cache size. See the Informatica Knowledge Base article number 352050 for further details.
Native initial stack size in IBM JRE	<code>-Xmso4096k</code>	Applicable for an AIX and WebSphere environment that uses IBM JRE.
Stack Size	<code>Xss2000k</code> or <code>-XX:ThreadStackSize=2000</code>	Use the platform as default. If required, you can adjust the setting. Stack Size is the memory used by Java for each thread it spans. This memory is outside the heap. If the memory is too low, the server might fail with the 'Stack Over Flow' error. If the memory is too high, the server might fail with the 'Out of memory' error. If you need to fine-tune the value, analyze the heap dump to arrive at an optimal value. See the Informatica Knowledge Base article number 144477 for further details.

Parameter	Recommended Setting	Description
Garbage Collection Policy	Use the default policy, as chosen automatically by the JVM.	<p>If you decide to modify the setting, you must do a complete analysis to determine the garbage collection (GC) policy. Garbage collection policy, includes (Garbage First Garbage Collector) G1GC, parallel GC, and concurrent sweep.</p> <p>For example, the parallel GC in heavy real time usage can give longer pauses, but the concurrent sweep do not give long pauses. You can use a combination of the policies. See the Java Guide to set the exact parameters.</p> <p>See the Informatica Knowledge Base article number 144944 for further details.</p>
Application Server Sizing	Optimal application server sizing.	<p>Proper analysis is required to determine the application server sizing. Different factors including average number of concurrent users, peak number of concurrent users, current data volume, rate of data growth, future data volume, Hierarchy Manager configuration, Hierarchy Manager relationship cardinality, and number of Data Director applications must be considered for sizing.</p> <p>For assistance on server sizing, contact the Informatica Professional Services team.</p>

The following table lists the recommendations for the database connection pool:

Parameter	Recommended Setting	Description
CMX_SYSTEM Database Connection Pool	Ensure that the default values of the application server are accurate. For example, if you use WebLogic, the value must be 10.	<p>For batch or real-time operations, do not use <code>cmx_system</code> connections, so the default values are adequate.</p> <p>See the Informatica Knowledge Base article number 121471 for further details.</p>
ORS Database Connection Pool	<p>Max Connection: <math>(N+T) \times 2.5</math>, where:</p> <ul style="list-style-type: none"> <li>- N is the number of concurrent IDD and SIF or BES API users.</li> <li>- T is the number of concurrent threads in Batch.</li> </ul> <p>Min Connection: 0</p> <p>Test Connection on Lease: Disable</p> <p>Statement Caching: 10</p>	<p>On average, each thread takes 2.5 connections. Therefore, multiply the number of concurrent threads by 2.5.</p> <p>Test Connection on Lease: Disable this property to avoid additional database cost incurred.</p> <p>Min Connection: In some instances, application servers might have some connection leaks with respect to rollback transaction where a connection caught in a rollback is not released to the pool. If such instances are found, you can set Min Connection to 0.</p> <p>Statement Caching (WebLogic): Initially, set 10 as the minimal number.</p> <p>See the Informatica Knowledge Base article number 121471 for further details.</p>

# Recommendations for Virtual Image Environment

The following table lists the recommendations for the virtual image environment:

Parameter	Recommended Setting	Description
Data file storage	Use a physical drive instead of a virtualized drive.	Use a physical drive to store the data files instead of a virtualized drive within the image. The virtualized drive causes input or output (I/O) contention and latency through another layer with no actual benefit.
Hardware and software specification	Equivalent physical or standalone instance.	The hardware and software that you use must be equivalent to the physical machine. Ensure that the specifications meet all the PAM or sizing requirements.
CPU Cores	100% allocated to virtual image.	The physical machine, where the virtual image is hosted, must allocate 100% of the CPU cores to the virtual image. Informatica does not recommend sharing the cores.

# Recommendations for Database

You can configure the parameters related to the database environment, virtual image environment, and database tables and data.

The following table lists the recommendations related to the database environment:

Parameter	Recommended Setting	Description
Environment sharing	<p>Always use non-shared environment:</p> <p>The production ORS must have exclusive use of the database server instance.</p> <p>The production database server instance must have exclusive use of the host machine.</p>	<p>You must not share the MDM Hub database server instances, such as MDM Hub Master Database, ORS, and ActiveVOS database with other MDM Hub installations and applications.</p> <p><b>Note:</b> Each additional level of sharing affects the performance.</p>
Connectivity: Application server to the Hub Store databases, such as ping or tracert (Trace Route)	<p>Enable fastest connectivity to the Hub Store by using fiber-optic connections.</p> <p>Run <code>ping</code> and <code>tracert</code> commands from the application server to the database to verify the basic network performance.</p> <p>The <code>tracert</code> command must preferably return with one hop away.</p> <p>Response latency must be less than 10 milliseconds.</p>	<p>Connection latency might have a major performance impact. You must use the fastest connectivity possible.</p>
Connectivity: Database server to data file storage	<p>Enable faster connectivity to the data by using fiber-optic connections.</p> <p>You must have a dedicated point-to-point connection to avoid network contention.</p>	<p>Connection latency might have a major performance impact.</p> <p>You must use the fastest connectivity possible.</p>
Database server sizing	<p>To determine the database server sizing, you need proper analysis.</p>	<p>You must use optimal database server sizing. You must consider the different factors, such as the current data volume, rate of data growth, future data volume, SIF or BES calls, and batch volume.</p> <p>For assistance on server sizing, contact the Informatica Professional Services team.</p>

Parameter	Recommended Setting	Description
MDM Hub Master Database (cmx_system) host name in cluster environments	Absolute host name or IP address	To avoid caching issues in multi-node or cluster environments, use the absolute database host name or IP address in place of the default localhost. The host name is configured during the MDM Hub installation. You can update the host name in the DATABASE_HOST column of the C_REPOS_DATABASE table.
Test specific disk partition IO	Reasonable MB/s	Follow the instructions in the Informatica Knowledge Base article number 139805 to test all disk partitions involved in the MDM Hub, including partitions where tablespaces reside and where the database debug log is written. If it is greater than 10 MB/s, you must fix that specific partition.

The following table lists the recommendations related to the database tables and data:

Parameter	Recommended Setting	Description
Custom and Backup tables created with names starting with C_REPOS%	Do not name the custom tables or backup tables with names starting with C_REPOS%.	During HUB server restart and Met Migration, the performance of the Hub Console degrades if the schema has large volume tables with names that start with C_REPOS%. Ensure that backup tables are not created with names starting with 'C_REPOS%'.
Fragmentation	Minimize likelihood of fragmentation.	Maintain the Oracle schema to ensure that fragmentation is kept to a minimum. Monitor and de-fragment whenever the degree of fragmentation has an impact on the MDM Hub performance.
High volume of data in C_REPOS_TABLES with historical data	Perform regular maintenance on the following METADATA tables: - C_REPOS_AUDIT - C_REPOS_MQ_DATA_CHANGE - C_REPOS_JOB_CONTROL - C_REPOS_JOB_METRIC - C_REPOS_MET_VALID_RESULT - C_REPOS_MET_VALID_MSG - C_REPOS_TASK_ASSIGNMENT_HIST	If the number of records in the metadata tables is too high, it might cause issues such as slow startup, out of memory errors, and performance issues. It is recommended to back up and truncate or reduce the data volume on the metadata tables. If you enable raw retention on any base object then you can purge C_REPOS_JOB_* tables for any date beyond the maximum raw retention period. For more information, search the Informatica Knowledge Base for article number 141201.

# Recommendations for Oracle Database

You can configure the parameters related to the Oracle database environment, tablespace, Oracle table statistics, RAC recommendations, and Oracle flashback.

The following table lists the recommendations related to the Oracle database environment:

Parameter	Recommended Setting	Description
C_REPOS_APPLIED_LOCK	Enable caching for this table, C_REPOS_APPLIED_LOCK	The application uses this table frequently so you can cache this table to improve performance.
RMAN backups	Suppress RMAN backups from running during batch processing.	RMAN backups are good for a fast backup and restore. However, performance is decidedly lower when the RMAN backup is performed.
Archive Logging	Turn off archive logging during the initial data load. For steady-state operation, you can turn on archive logging after the initial data load ends.	<p>Archive logging is unnecessary during the initial data load and adds an overhead. If there is a failure during the initial data load, Oracle rolls back the entire transaction (the current batch cycle of a batch job). The process can be re-run with no data loss.</p> <p>During the initial data load, back up the ORS schema that loads at regular checkpoints with no jobs running. For example, back up the ORS schema after major long running jobs have completed. At an absolute minimum, take backups after completion of each phase: Stage, Load, Match, and Merge. You require backups to safeguard the work already done before you proceed.</p> <p>You might enable archive logging for all steady-state operations (post initial data load).</p> <p>If you use a Standby database (database mirroring), disable the standby before doing the initial data load. When the initial data load is complete, copy the database to the standby site, and then enable the standby.</p>

The following table lists the recommendations for the tablespace:

Parameter	Recommended Setting	Description
Tablespace	<p>Use the following recommended settings for tablespace:</p> <ul style="list-style-type: none"> <li>- Locally Managed.</li> <li>- Uniform Extent.</li> <li>- Auto Segment Management.</li> <li>- Default Tablespace Block Size of 8 KB or 16 KB.</li> </ul>	<p>Applicable for all tablespaces involved in the MDM Hub including CMX_DATA, CMX_INDX, and CMX_TEMP.</p> <p>Default Tablespace Block Size of 8 KB is good for high API workload implementations. Use 16 KB as a balanced block size to support a mix of API and batch processing (small transactions and bulk read and write transactions).</p> <p>If you use a Default Tablespace Block Size of 16 KB and the database was created with default block size of 8 KB, then configure an appropriate DB_CACHE for it.</p>
Tablespace Storage	<p>Use the following Oracle recommendations for storage of tablespace data files:</p> <ul style="list-style-type: none"> <li>- If you use RAID: use either RAID 1+0 or RAID 0.</li> <li>- If you do not use a RAID controller: each tablespace must comprise multiple data files spread across different disks.</li> </ul>	<p>Many small disks perform better than a few large disks if everything else remains equal.</p> <p>Use the following steps to improve the performance:</p> <ul style="list-style-type: none"> <li>- RAID1+0 (RAID 10) has a high degree of fault tolerance with mirroring. Use as many disks as possible. Disks must be as fast as possible.</li> <li>- Avoid RAID 5 due to a writing overhead and poor performance if there is a disk failure.</li> <li>- If there is no RAID controller: use multiple single disks to split several data files over more disks. Do not dedicate data files for one tablespace to a single disk. Use the fastest disks possible and stripe the disks. In such cases, each tablespace uses part of each disk instead of a tablespace being dependent on a single disk. Similarly, it is better to keep redo and undo logs in a different physical disk.</li> </ul>

The following table lists the recommendations for the Oracle table statistics:

Parameter	Recommended Setting	Description
Table Statistics	<p>Analyze the ORS schema on a regular (frequent) basis.</p> <p>Use the following options:</p> <ul style="list-style-type: none"> <li>- Analyze the full schema (perform outside business hours).</li> <li>- Analyze individual tables whenever 10% of the data is changed.</li> <li>- Perform unplanned table analysis.</li> <li>- Set the DBMS_STATS .SET_GLOBAL_PREFS (see Environment Sharing in general database recommendations).</li> <li>- With no sharing:  <code>DBMS_STATS.SET_GLOBAL_PREFS( 'DEGREE', DBMS_STATS.AUTO_DEGREE);</code> </li> <li>- If forced to share:  <code>DBMS_STATS.SET_GLOBAL_PREFS( 'DEGREE', &lt;(number of CPUs on DB Server) minus 1 &gt; );</code> </li> </ul>	<p>Analyze the entire ORS schema on a regular basis as a best practice.</p> <p>Analyze individual tables whenever 10% of the data is changed by using data sampling. This is the Oracle recommendation. See Oracle database documentation for details.</p> <p>With sampling there is a trade-off. The best execution plans are always chosen when the statistics represent the entire table. When you use sampling, the execution plan is as appropriate to the tables as the data sample. For example, you can use 10% of the table as representative of the table. If you use a sample, be sure to use a large enough sample. A sample of 1 to 2% is too small unless the table is large. A sample of 10% is more representative on smaller tables. You can switch off sampling and perform a full analyze, which provides the most appropriate execution plans. However, a full analyze is achieved at the cost of far higher overhead in terms of the time taken when running the analyze. It is best to run outside of business hours to mitigate impact.</p>



The following table lists the recommendations for RAC:

Parameter	Recommended Setting	Description
Environment	Each node must satisfy the same recommendations as set for the standalone environment.	-
_PKQ Sequence	Use NOORDER . Set sequence cache to 20000.	Informatica recommends to increase initial data load performance. These sequences are used to populate the ROWID_OBJECT in the base object tables.  For more information, search the Informatica Knowledge Base for article number 115788.

The following table lists the recommendations for the Oracle flashback:

Parameter	Recommended Setting	Description
Flash recovery area	Fast file system.	Use a fast file system for your flash recovery area, preferably without operating system file caching.
Disk spindles	As needed.	Configure enough disk spindles for the file system to hold the flash recovery area.
Striped storage volume	Smaller stripe size.	If flash recovery area does not have non-volatile RAM, opt for striped storage volume with smaller stripe size such as 128k.  This will allow each write to the flashback logs to be spread across multiple spindles, improving performance.

For more information, search the Informatica Knowledge Base for article number 333718.

## INIT.ORA Recommendations for Oracle

The INIT.ORA recommendations are based on the standard hardware of 24 GB RAM with 8-core CPU Oracle server.

The following table lists the basic initialization parameters for the Oracle database:

INIT.ORA Parameter	Applicable to version	Value	Description
cursor_sharing		EXACT	Only allows statements with identical text to share the same cursor.
db_block_checking		FALSE	To avoid additional overhead.
db_block_size		8192	This parameter affects the maximum value of the FREELISTS storage parameter for tables and indexes. Oracle uses one database block for each freelist group.

INIT.ORA Parameter	Applicable to version	Value	Description
db_cache_size		2000M	Reduces additional overhead on dynamic allocation.
db_file_multiblock_read_count		0	To be auto-determined by Oracle database.
db_writer_processes (Single Core)		1	Oracle guideline.
db_writer_processes (Multi Core)		CPU / 8	Oracle guideline.
disk_async_io		TRUE	Oracle guideline.
filesystemio_options		SETALL	This bypasses file system buffer cache, especially on Linux.
java_pool_size		0	To be auto-determined by Oracle database.
large_pool_size		400M	If undefined, RMAN would use the SHARED POOL.
log_buffer		10M	A value of 10 MB is a reasonable initial size. Increase based on Oracle AWR report taken under typical heavy load. <b>Section: Cache Sizes &gt; Log Buffer</b>
memory_target		Two thirds of available memory.	Two thirds of the available physical memory for Oracle. One third is reserved for system and other processes.
memory_max_target		Two thirds of available memory.	Two thirds of the available physical memory for Oracle. One third is reserved for system and other processes. Based on your requirement, you can increase this value.
open_cursors		1000	The MDM Hub uses parallel processing requires opening up multiple cursors.
parallel_adaptive_multi_user		TRUE	The MDM Hub uses multi-sessions for the same Oracle instance.
Processes		1000	Sufficient Oracle processes are allocated to support connection, parallel thread, internal process, and other usage. If set too small, some processes might fail to run.

INIT.ORA Parameter	Applicable to version	Value	Description
Recyclebin		OFF	Set the value of the recyclebin parameter to OFF. The MDM Hub has many temporary tables that the recycle bin spends time trying to maintain when dropped.
shared_pool_size		400M	Use 400 MB initially. Increase this according to the value seen in the Oracle AWR report for this instance taken under typical heavy load. See the Section: <b>Cache Sizes &gt; Shared PoolSize</b> .
streams_pool_size		0	The MDM Hub does not use Oracle streams functionality. Disable the functionality to prevent Oracle from reserving memory for it.
utl_file_dir		Do not set this	Use an Oracle Directory object instead of this parameter. For more information, see the Informatica Knowledge Base article 90456.
workareas_size_policy		AUTO	To be auto-determined by Oracle.
pga_aggregate_target	11g	0 in case of Automatic Memory Management (AMM) or One third of memory allocated to Oracle in case of Manual Memory Management (MMM).	Set PGA explicitly for MMM. For AMM, you need not set PGA.
sga_target	11g	Two thirds of Max_Memory in AMM or two thirds of allocated memory to Oracle in MMM.	For SGA, Informatica recommends that you allocate two thirds of memory available for Oracle regardless of AMM or MMM.  In AMM configuration, the setting of SGA explicitly always reserves two thirds of memory for SGA. This makes Buffer Cache ready for large data loaded into it during the batch. If you do not SGA, AMM might allocate memory for something else. The remaining memory might also not be enough for a big batch and it might run slower.

INIT.ORA Parameter	Applicable to version	Value	Description
optimizer_capture_sql_plan_baselines		False	Setting this parameter to false makes SQL plan management to not recalculate the execution plan for each repeatable SQL statement.
optimizer_index_caching		0	Setting this value to 0 defaults the behavior of cost-based optimization to favor nested loop joins and IN-list iterators.
optimizer_index_cost_adj		100%	To make use of default optimization based on table indexes, setting to default value evaluates index access paths at the regular cost. Default value is 100%.
optimizer_adaptive_features	12c	FALSE (Scope = BOTH).	Disables all the adaptive optimizer features, including adaptive plan, for better performance.
optimizer_use_sql_plan_baselines		TRUE	Enables or disables the use of SQL plan baselines.

For more information about the INIT.ORA parameters, see Informatica knowledge base article 90408.

## Recommendations for Microsoft SQL Server

You can configure parameters related to the Microsoft SQL Server to optimize the performance of the MDM Hub.

The following table lists the recommendations related to the Microsoft SQL Server environment:

Database Parameters	Recommended Setting	Description
AUTO_UPDATE_STATISTICS_ASYNC	ON	To enforce statistics update in asynchronous mode.
PARAMETERIZATION	FORCED	To parameterize SQL statements. Parameterized statements reduce the frequency of query compilations and recompilations to improve performance.
READ_COMMITTED_SNAPSHOT	ON	To allow other sessions to see data that is not yet a committed transaction.

# Recommendations for IBM Db2

You can configure the parameters related to the IBM Db2 environment, registry variables, and database file configuration parameters.

The following table lists the recommendations related to the IBM Db2 environment:

Parameter	Recommended Setting	Description
Physical Disk Drives (Tablespace)	Different physical drives for different tablespaces.	To reduce the amount of blocked input/output, you can increase input/output parallelism. Achieve input/output parallelism by storing user data tablespaces, temporary tablespaces, and transaction logs on different physical disk drives. Batch operations can access all the paths in parallel, which increases the throughput by reducing the input wait times and output wait times.
Physical Disk Drives (Container)	Different physical drives for different containers.	If more physical disk drives are available, you can increase input/output parallelism by extending parallelism to the container level. To increase input/output parallelism, place all containers for a tablespace on different physical disks. The IBM Db2 prefetchers and input/output cleaners access these containers in parallel without blocking each other, thereby increasing the throughput.
Processing Large Dataset		<p>When you process a large data set, use the following command to rebind packages:</p> <pre>db2 bind @db2cli.lst blocking all grant public sqlerror continue CLIPKG 10</pre> <p>Perform the step every time you process a large data set. It is better to do so after the initial data load.</p>
Reorganize Match Tables		<p>Routinely determine and re-organize the match key tables (C_&lt;Base Object&gt;_STRP).</p> <p>To improve the performance of the SearchMatch API, you need to reorganize match key tables based on their primary key column, <code>SSA_KEY</code>. To determine if a match key table needs reorganization, perform a reorganization check and analyze the results. The cluster ratio of the primary key index appears in the <code>CLUSTERRATIO</code> column, F4, of the reorganization check result. The cluster ratio must be close to 100% for optimal performance.</p> <p>Determine when to reorganize the match key table by noting the cluster ratio at which you observe degradation in the SearchMatch API performance. Use the IBM DB2 <code>REORGCHK</code> and <code>REORG</code> commands to reorganize tables.</p> <p>Update the match key table statistics so that the IBM Db2 optimizer can use the table layout that the reorganization generates.</p>

The following table lists the recommendations for the registry variables:

Database Parameters	Recommended Setting	Description
DB2_INLIST_TO_NLJN	NO	Configures the optimizer to prefer or not prefer nested loop joins. The Db2 SQL compiler might rewrite an IN list predicate as a join. The rewrite might provide better performance if you define an index on the joined columns. When the optimizer does not have accurate statistics, it might not be able determine the best join for the rewritten join. Set the variable to NO to prevent the optimizer from favoring nested loop joins in such cases.
DB2_ANTIJOIN	YES	Configures the optimizer to transform subqueries into anti-joins. The MDM Hub has queries that use NOT EXISTS subqueries. Set the registry variable to YES to look for possibilities to transform NOT EXISTS subqueries into anti-joins that IBM Db2 processes more efficiently.
DB2_REDUCED_OPTIMIZATION	REDUCE_LOCKING	Configures reduced optimization features or rigid use of optimization features at the specified optimization level. Set the registry variable to REDUCE_LOCKING to favor NLJOIN over MSJOIN whenever possible to reduce the amount of locking on the outer table.
DB2_EXTENDED_OPTIMIZATION	ON, ENHANCED_MULTIPLE_DISTINCT, IXOR, SNHD	Configures whether or not the query optimizer uses optimization extensions to improve query performance. The ON, ENHANCED_MULTIPLE_DISTINCT, and SNHD values specify different optimization extensions.
DB2_HAS_JOIN	YES	Configures hash join as a possible join method when compiling an access plan. Tune hash join to get the best performance. For best performance of a hash join, avoid hash loops and overflows to disk. To tune hash join performance, perform the following tasks: <ul style="list-style-type: none"> <li>- Estimate the maximum amount of memory available for the sheapthres configuration parameter.</li> <li>- Tune the sortheap configuration parameter.</li> </ul>

The following table lists the recommendations for the database file configuration parameters:

Database Parameters	Recommended Setting	Description
LOCKLIST	AUTOMATIC	<p>Allocates the amount of storage to the lock list of a database.</p> <p>Multiple MDM Hub processes use "locks at the row-level" to complete tasks and to support concurrency. The number of locks that IBM Db2 needs to acquire depends on the number of rows to process. If the incoming volume differs greatly in size, set the parameter to AUTOMATIC to allow the database manager to determine the appropriate value. If you conservatively tune the LOCKLIST parameter value, lock escalations can occur, and some of the MDM Hub operations can fail due to lock timeouts.</p>
MAXLOCKS	AUTOMATIC	<p>Configures the percentage of the lock list that one application can use.</p> <p>Most MDM Hub processes run under the scope of a single application. Such single applications can acquire many row-level locks, consuming most of the available lock list. If the incoming volume differs greatly in size, predicting the MAXLOCKS parameter value is difficult. Set the parameter to AUTOMATIC to allow the database manager to determine the appropriate value. If you conservatively tune the MAXLOCKS parameter value, lock escalations can occur, and some of the MDM Hub operations can fail due to lock timeouts.</p>
CATALOGCACHE_SZ	25000 or higher	<p>Configures the maximum memory that the catalog cache can use from the shared memory of the database.</p> <p>IBM Db2 stores system catalog information in the catalog cache. The MDM Hub comprises many dynamic SQL queries that reference multiple metadata objects. If the catalog cache is large, IBM Db2 can retain information for some of the metadata objects from the system catalogs in the memory. If subsequent dynamic SQL queries require the same metadata objects, the compilation is quick. The MDM Hub comprises many frequently accessed metadata objects. Therefore, you must set the CATALOGCACHE_SZ parameter value to 25000 or higher.</p>
LOGBUFSZ	4096 or higher	<p>Configures the amount of the database heap to use as a buffer for log records before writing the records to disk.</p> <p>The MDM Hub creates logs for most of its operations in the IBM Db2 transaction logs. IBM Db2 buffers the log records in the log buffer before writing them to the disk.</p> <p>If the log buffer is large, IBM Db2 writes the log records to the disk less frequently. This makes disk input/output for log records more efficient. The default value for the database configuration parameter is not sufficient for an average MDM Hub environment. Set the parameter to 4096 pages or higher.</p>

Database Parameters	Recommended Setting	Description
LOGFILSIZ	128000 or higher	<p>Configures the number of log records written to the log files.</p> <p>A single MDM Hub transaction can contain many DML queries resulting in many log records that might span across many log files. A large log file size avoids the need to create new log files frequently. If IBM Db2 creates new log files frequently, it adversely influences the performance of input/output bound systems.</p> <p>The total log space for a database is equal to the total number of primary and secondary log files multiplied by the log file size. The database must have adequate log space to ensure that the MDM Hub transactions do not run out of log space and fail. If the MDM Hub transactions fail, the database needs more time to roll back the transactions. Set LOGFILSIZ to 128000 or higher to ensure that the MDM Hub transactions do not fail and need a roll back. Also, you must consider the number of primary and secondary logs for a complete equation of log space.</p>
LOGPRIMARY	100	<p>Configures the number of primary log files to be pre-allocated. IBM Db2 creates primary log files when you activate the database. If the uncommitted transaction exhausts the primary log space, IBM Db2 creates secondary log files as needed. Set the LOGPRIMARY parameter value to 100 to efficiently handle the MDM Hub processes. Secondary log files act as a backup in cases where long transactions can exhaust the entire primary log space.</p>
LOGSECOND	100	<p>Configures the number of secondary log files that IBM Db2 can create and use for recovery log files.</p> <p>Log file creation can adversely impact the performance based on the size of the log file. If you allocate sufficient primary log space, transaction performance increases because the database does not create secondary log files frequently. Set the LOGSECOND parameter value to 100 to cover unexpected long transactions due to large incoming volumes, especially during batch processes. The sum of the values of LOGPRIMARY and LOGSECOND must be 200.</p>
PCKCACHESZ	128000	<p>Configures the package cache size which is allocated out of the database shared memory.</p> <p>The MDM Hub has many dynamic SQL statements. Each dynamic SQL statement has a compiled package associated with it. IBM Db2 caches these packages in the package cache memory. You must configure an appropriate package cache size to avoid package cache overflows, which adversely influence performance.</p> <p>Experiment with the values for the package cache size. Initially, set the parameter value to 50000 and monitor the different phases of the MDM Hub processes. If you observe frequent package overflows, tune the parameter again.</p>
STMHEAP	AUTOMATIC	<p>Configures the limit of the statement heap, which is used during the compilation of an SQL statement.</p> <p>If the statement heap is not sufficient, it might prevent the optimizer from evaluating all possible access plans for an SQL query. This might result in a suboptimal plan and adversely influence performance. Set the STMHEAP parameter to automatic to allow the optimizer to weigh all possible access plans for the compilation an SQL query.</p>



Database Parameters	Recommended Setting	Description
SORTHEAP	AUTOMATIC	Configures the sort heap size. The MDM Hub processes perform many sorts. If the sort heap size is not sufficient, large sorts can spill from the memory to disk. Disk input/output is slower compared to memory, and such sort spills can cause queries to run longer. Sort spills to disk can adversely influence performance and is more evident with larger spills. Set the parameter value to AUTOMATIC for the memory tuner to dynamically size the memory area as the sort requirements change.
SHEAPTHRES_SHR	AUTOMATIC	Configures the limit on the total amount of database shared memory that the sort memory consumers can use at a time. Set the SHEAPTHRES_SHR parameter to AUTOMATIC if you set the SORTHEAP parameter to AUTOMATIC.
UTIL_HEAP_SZ	50000 or higher	Configures the maximum amount of memory that the BACKUP, RESTORE, and LOAD utilities can use simultaneously. During some batch operations, the MDM Hub uses the IBM Db2 LOAD utility to move data between tables. The LOAD utility uses the utility heap to complete the data movement process. The size of the utility heap has an impact on the performance of the LOAD operation. Set the UTIL_HEAP_SZ parameter to an appropriate value to provide better throughput for the MDM Hub processes.

## Recommendations for the MDM Servers

The following table lists the recommendations for the Hub Server, the Process Server and the Elasticsearch servers:

Parameter	Recommended Setting	Description
Ping or tracert (Trace Route)	Response latency must be less than 10 milliseconds.	Run a <code>ping</code> or <code>tracert</code> to the MDM Hub Server or the Process Server from the client box or the Hub Server box. If applicable, repeat this step for the following boxes: <ul style="list-style-type: none"> <li>- All nodes in a cluster</li> <li>- All Elasticsearch server nodes</li> <li>- Load balancer</li> <li>- Web server</li> </ul>
Environment sharing	The application server and the Elasticsearch server must run on their own dedicated server machines without having any additional products, such as Informatica platform, on the same machine.	Each additional level of sharing affects the performance.

# Recommendations for the MDM Hub

You might be able to improve performance by changing how the MDM Hub operates within the environment.

## MDM Hub Environment

The following table lists the recommendations related to the MDM Hub environment:

Parameter	Recommended Setting	Description
Obsolete Data Director applications and Operational Reference Stores databases	Remove obsolete items.	Obsolete Data Director applications and Operational Reference Stores schemas impact the performance of server startups, run time memory, and Security Access Manager profile caching.
Order of authentication provider in Security Providers	Configure the security providers in order with the first provider being the provider that authenticates the heaviest user load.	<b>Configuration &gt; Security Providers &gt; Authentication Providers.</b> The MDM Hub authenticates the user based on the order of the security providers configured. If most of the users are authenticated by using the custom security provider (if applicable), it is recommended to move it to the first position. <b>Note:</b> Each authentication request has a cost of few milliseconds associated with it. The number of authentication requests is reduced significantly by using the User Profile Cache.

## Application Server

The following table lists the recommendations for the application server configuration:

Parameter	Recommended Setting	Description
enable-http2	Disable	Applicable for JBoss. Disable http2 on the undertow listener in the <code>standalone-full.xml</code> file of JBoss. Default is true.  <pre>&lt;http-listener name="default" socket-binding="http" redirect-socket="https" enable-http2="false"/&gt;</pre> See the Informatica Knowledge Base article number 618247 for further details.
Maximum thread count for the thread pool	300 or higher	For example, in the JBoss application set the following property in the <code>standalone-full.xml</code> file:  <pre>&lt;thread-pools&gt; &lt;thread-pool name="default"&gt; &lt;max-threads count="300"/&gt; &lt;/thread-pool&gt;&lt;/thread-pools&gt;</pre>

Parameter	Recommended Setting	Description
Maximum connections in HTTP connection pool	300 or higher	<p>For example, in the JBoss application set the following property in the <code>standalone-full.xml</code> file:</p> <pre>&lt;connector name="http" protocol="HTTP/1.1" scheme="http" socket-binding="http" max-connections="300"/&gt;</pre>
JDBC logging level	OFF	<p>For example, in the JBoss application, set the following log level property in the <code>standalone-full.xml</code> file:</p> <pre>&lt;subsystem xmlns="urn:jboss:domain:logging:1.2"&gt;   &lt;logger category="com.microsoft.sqlserver.jdbc" &gt; &lt;level name="OFF"/&gt;&lt;/logger&gt;</pre>
Transaction timeout	Greater than 3600 seconds.	<p>Set the transaction timeout to at least 3600 seconds (1 hour).</p> <p>For example, in the Boss application set the following property in the <code>standalone-full.xml</code> file:</p> <pre>&lt;coordinator-environment default-timeout="3600"/&gt;</pre>
Maximum beans instance pool size in JBoss	20 or higher. The maximum beans pool size for the data source connection is based on the off-peak resource demand.	<p>Stateless session bean pool. You must increase it to improve concurrency and performance of the Cleanse and CleansePut APIs.</p> <p>In JBoss, set the following property in the <code>standalone-full.xml</code> file:</p> <pre>&lt;bean-instance-pools&gt; &lt;strict-max-pool name="slsb-strict-max-pool" max-pool-size="20" instance-acquisition-timeout="5" instance-acquisition-timeout-unit="MINUTES"/&gt; &lt;strict-max-pool name="mdb-strict-max-pool" max-pool-size="20" instance-acquisition-timeout="5" instance-acquisition-timeout-unit="MINUTES"/&gt; &lt;/bean-instance-pools&gt;</pre>
Load jobs and user exits in WebSphere	Update the JAR files.	<p>Due to conflict in the IBM Java transaction JAR files with the MDM cleanse libraries, the Infinispan is not initialized for some components.</p> <p>See the Informatica Knowledge Base article number 567591 for further details.</p>

## Operational Reference Store

The following table lists the recommendations for the ORS configuration:

Parameter	Recommended Setting	Description
Production Mode	Enable this property in Production.	<b>[Configuration &gt; Database &gt; Database Properties]</b> . Enable this property to remove additional overhead of pre-scheduled daemons that refresh the metadata cache.
Batch API Interoperability	Enable if both real time and batches are used. You can also enable the setting for concurrent real time API performance improvement.	<b>[Configuration &gt; Database &gt; Database Properties]</b> . Enabling this configuration has an impact on performance. Enable the configuration if batches and real time API calls are used or if Data Director is used. If your application uses neither real time API updates nor Data Director, do not enable API Batch Interoperability. <b>Tip:</b> During Initial Data Load, disable this property for faster loading of data.
Auditing	Disable the auditing completely.	Auditing introduces additional overhead. You must disable auditing completely.
Write lock monitor Interval	<code>cmx.server.writelock.monitor.interval=10</code>	When more than one Hub Console uses the same ORS, a write lock on a Hub Server does not disable caching on the other Hub Servers. The unit is in seconds.  For more information, see the <i>Multidomain MDM Configuration Guide</i> .

## Schema Design

The following table lists the recommendations for the schema design:

Parameter	Recommended Setting	Description
Child Base Objects	Avoid too many child base objects for a particular parent base object.	The performance of load, tokenize, and automerger batch jobs decreases as the number of child base objects for a base object increases.
Business Entities	Avoid creating cyclic or recursive relationship.	Ensure that you do not create cyclic relationship, such as A has a referenceOne link to B and B has a referenceOne link to A, to avoid recursive calls. Recursive calls affect the performance when these objects are referenced.

Parameter	Recommended Setting	Description
Match columns	Avoid too many match columns.	The performance of tokenize and match jobs decreases with the increase in the number of match columns.
Lookup Indicator	Enable Lookup Indicator only for 'Lookup' tables and not for any other base objects unrelated to lookup.	<b>Schema &gt; [base object] &gt; Advanced &gt; Lookup indicator.</b> Enabling lookup indicator for non-lookup base objects unnecessarily caches the base object data in the memory. Doing so results in out of memory errors, slow Data Director performance, and slower rate of lookup cache refresh.
Lookup Display Name	Configure Lookup Display Name to be the same as the lookup column.	For high volume lookup tables: If you set the lookup display name to any column other than the column on which the relationship is built, SIF PUT calls must send the lookup display name values in the SIF call. When inserting data into the base object, the lookup value is validated by querying the lookup table. The order of lookup is predefined: the lookup display column value comes first followed by the actual column value second. In high volume lookup tables this becomes an overhead.
History	Enable History if you want to retain historical data for the specific base object. Otherwise, disable it.	If you enable History for a base object, the MDM Hub additionally maintains history tables for base objects and for cross-reference tables. The MDM Hub already maintains some system history tables to provide detailed change-tracking options, including merge and unmerge history. The system history tables are always maintained. Over a period, history in the database keeps growing. Consider keeping months or at most a few years of history in the system and to preserve database access performance.
History	To avoid very large history tables that cause performance issues, you can partition the tables.	For more information, search the Informatica Knowledge Base for article number 306525.
Cross Reference Promotion History	Enable Cross Reference Promotion History if you want to retain historical data for the specific base object.	<b>Schema &gt; [base object] &gt; Advanced &gt; Enable History of Cross Reference Promotion.</b> Enabling history incurs performance cost both to real time and to batch operations. Use the history option cautiously and if required.

Parameter	Recommended Setting	Description
Trust	Configure trust only for required columns.	<p>A higher number of trust columns and validation rules on a single base object incur higher overhead during the Load process and the Merge process.</p> <p>If the more trusted and validated columns are implemented on a particular base object:            Longer SQL statements (in terms of lines of code) are generated to update the <code>_CTL</code> control table and the <code>_VCT</code> validation control table.</p> <p>Minimize the number of trust and validation columns to conserve good performance.</p>
Case-insensitive Search	<p>Enable case-insensitive search only for characters. For example, VARCHAR2 columns.</p> <p>Ensure that you do not include any column with a data type other than character, such as VARCHAR in the search query.</p>	<p>Enabling case-insensitive search for non-character columns hinders performance. In the Microsoft SQLServer, if the collate is case-insensitive, do not enable this property.</p>
Message Trigger Setup	Avoid configuring multiple message triggers for different event types.	<p>Do an in-depth analysis before configuring message triggers. There is a performance cost associated with them during the execution of load jobs.</p>
	Tune Message Trigger Query.	<p>The best approach to tuning the query used in the Package Views is to use Explain Plan. Add custom indexes wherever required to avoid full table scans, and analyze tables/schema on a regular basis. When you use Explain Plan, retrieve the plan by wrapping the query around an outer query that contains a "where" clause for a <code>rowid_object</code> equal to.</p> <p>For more information about message triggers, search the Informatica Knowledge Base for article number 142115.</p>
	Throughput can be greatly improved if you increase the Receive Batch Size and reduce the Message Check Interval.	<p>The Message Queue Monitoring settings have a major impact related to the throughput message posting time. Configure these settings from the Hub Console in the Master Reference Manager (MRM) Master Database (<code>CMX_SYSTEM</code>) in the Configuration section.</p>
	Avoid unnecessary column selection in message trigger.	<p>Do not select "Trigger message if change on any column" if you do not need to monitor all the columns. Also, try to minimize the selection of columns.</p>

Parameter	Recommended Setting	Description
'Read Database' Cleanse Function	Use the following recommended settings: <ul style="list-style-type: none"> <li>- Use with caution.</li> <li>- Enable 'cache' if used.</li> </ul>	<p>The 'Read Database' cleanse function incurs a performance overhead compared to using a similar MDM Cleanse Function to perform the same function. The performance overhead is more pronounced on a high volume table. The overhead is caused by the creation of a new database connection and the corresponding transmit to, processing by, and receipt of the results from the database. These would otherwise be managed within the Process Server application layer.</p> <p>If use of this function cannot be avoided, if applicable, enable caching behavior of the Read Database function. Pass a Boolean 'false' value to the 'clear cache' input field of the Read Database function'. Doing so reduces performance lag by enabling future operations to use the cached value rather than creating a new database connection on each access of the function.</p>
Cleanse Functions	Do not make it very complex.	The performance of batch jobs increases with reduced number and reduced complexity of cleanse functions.
Timeline (Versioning)	<p>If it is strictly required, you must enable dynamic timeline or versioning only on an entity base object. The base object must be a regular base object and not a base object with Hierarchy Manager Relationship.</p> <p>Versioning has a performance impact on the base object associated with it.</p> <p>For Hierarchy Manager Relationship base objects, you must enable versioning without an option to disable it.</p>	<p>Versioning must only be enabled on those Entity base objects (regular MDM Hub base object) which strictly need it to maintain the fastest performance possible. With the Versioning functionality, the additional associated metadata and processing carry a significant amount of complex processing when running any process on a version-enabled base object. Enabling versioning on a base object brings an additional performance cost to all processing performed on that base object.</p> <p>For more information, search the Informatica Knowledge Base for article numbers 138458 and 140206.</p>
State Management	Disable state management if you do not require it.	<p>State Management carries an associated performance overhead.</p> <p>If you use Data Director with workflows, you must enable State Management.</p> <p>However, enabling History for State Management Promotion at the cross-reference level is optional.</p>

Parameter	Recommended Setting	Description
Delta Detection	Enable it only on the minimum number of columns that strictly need it.	<p>Delta Detection carries a sizable associated overhead on performance.</p> <p>If a Landing Table has only new and updated records in every staging job, you can disable delta detection. If you want to enable Delta Detection the least impactful approach is to use the <code>last_update_date</code>. If you need additional columns, for each additional column you enable, analyze if the involvement of this additional column is worth the associated performance overhead. Avoid blindly enabling Delta Detection on all the columns.</p>
Cleanse Mappings	Minimize the complexity of Mappings.	<p>Minimized the complexity in the mapping to have better performance.</p> <p>If you use a Cleanse List in a Cleanse mapping, use static data in the Cleanse List. Consider using lookup tables only for dynamic data.</p>
Validation Rules	Optimize Validation Rule SQL code.	<p>The detection piece of each Validation Rule SQL runs against every record during the Load process to determine if it applies.</p> <p>Poorly performing Validation Rule SQL influences performance on every Loading record.</p>
User Exits and External Calls	Optimize user exit or external call code for performance.	<p>User exit or external call code influences performance if it is not optimized.</p> <p>Applicable to Data Director, batch user exits, and external calls. See the best practices when you implement user exits and external calls.</p>
Packages	Optimize the SQL code written in each MDM Queries which is called from an MDM Hub Package.	<p>These MDM Hub Packages are used in Data Director, SIF API calls, Data Manager, Merge Manager, and in search operations.</p> <p>If you do not tune the MDM Hub for performance, it results in an expensive operation whenever it is called.</p>



Parameter	Recommended Setting	Description
Custom Indexes	Use caution when adding custom indexes. Each index added has an associated cost. Ensure that gain received outweighs the cost of each additional custom index.	<p>Index management has a performance cost associated.</p> <p>Perform the following steps to improve the performance cost:</p> <ol style="list-style-type: none"> <li>1. Get a log of the real queries run on the base object (content data) and base object shadow tables (content metadata) on a typical day. Ignore temporary T<math>\\$</math>% tables and system C_REPOS_% tables;</li> <li>2. Identify indexes which exist on these tables to avoid unnecessary overlap.</li> <li>3. Before adding any indexes, review a regular day of logs and take an inventory of: <ol style="list-style-type: none"> <li>a. SIF API or BES call duration.</li> <li>b. Data Director process durations.</li> <li>c. Batch jobs: <ol style="list-style-type: none"> <li>a. Duration of each batch job.</li> <li>b. Duration of each cycle within that batch job.</li> <li>c. Duration of longest running statements within a batch job.</li> </ol> </li> </ol> </li> <li>4. Consider the longest running process for potential benefit from a custom index.</li> <li>5. Consider adding indexes so the longest running SQL query or queries hit the new index in their execution plan. Avoid indexing fields which have many updates or inserts.</li> </ol> <p>After each new custom index added return to Step 3 and assess if there is still potential to improve performance through adding more custom indexes.</p>
Parallel Degree on Base Object	Between one and number of CPU cores on database machine.	<p>Parallel degree is an advanced base object property. For optimum performance of batch jobs, set a value between one and the number of processor cores on the database server machine.</p> <p>For more information, search the Informatica Knowledge Base for article number 181313.</p>

## Match and Merge

The following table lists the recommendations for the match and merge configuration:

Parameter	Recommended Setting	Description
Match Path Filter	Filter on match path instead of at the match rule level.	<p>If you need to exclude records from the match process, filter on the appropriate match path instead of at the match rule level.</p> <p>When you filter at the match path level, it excludes the records from tokenization and they, therefore, do not participate in the match.</p>
Check for missing children	Use it with caution.	<p>This match patch property indicates if parent records must be considered for matching based on the existence of child records.</p> <p>If you need a fuzzy match on a base object, tokenization of a parent base object record must occur. Tokenization of a parent base object record occurs if all child base objects that have the option to check for missing children disabled have a related child base object record. If a parent base object record has a child, where the option to check for missing children is disabled yet contains no record, the parent record is not tokenized.</p> <p>The MDM Hub performs an outer join between the parent and the child tables when the option to check for missing children is enabled. This option has an impact on the performance on each match path component on which the option is enabled. Therefore, when not needed, it is more efficient to disable this option.</p>
Match Key	The tighter the key the better the performance.	<p>The width of the match key determines the number of rows in the tokenization table (number of tokenized records which are used to match each record to be matched) and the number of records to be considered for each match candidates. Usually the standard key width is enough.</p> <p><b>Search Level:</b> Use the narrowest possible search level to generate acceptable matches. Usually the typical search level is enough.</p> <p><b>Match Rules:</b> For each match rule, add one or more exact match columns to act as a filter to improve the performance for each rule.</p>

Parameter	Recommended Setting	Description
Match Rule	Search Level: Narrow Match or Search Strategy: Filtered	<p>Search Level. To generate acceptable matches, you must use the narrowest possible search level. You can run a sample test with the required level to avoid under matching. Based on your requirement, you can fine-tune the settings.</p> <p>For each match rule, add one or more exact match columns that acts as a filter. The process might improve the performance.</p> <p>Filtered. Use filtered match rule instead of exact match rule to make use of fuzzy filtering on exact columns. The filtered rule enables bulk processing in the applications server rather than the database server. For implementing match rule, see the Informatica Knowledge Base article number 121044.</p>
Dynamic Match Analysis Threshold (DMAT)	Change if required. Default is 0.	<p>To set the DMAT, identify any large ranges that cause bottlenecks. After you identify the ranges, use the "Comparison Max Range" count from the logs.</p> <p>Proper analysis is required to change this value. See the Informatica Knowledge Base article number 90740</p>
STRIP_CTAS_DELETE_RATIO	Change if required. Default is 10%.	<p><b>C_REPOS_TABLE [base object] &gt; STRIP_CTAS_DELETE_RATIO.</b></p> <p>Proper analysis is required if you decide to change the default 10% value. See the Informatica Knowledge Base article number 145312.</p> <p>If the volume of data change in the <code>_STRP</code> table is more than this percentage, tokenization would instead use 'Create Table ... As Select ...' code to recreate the <code>_STRP</code> table with the needed changes rather than delete and/or insert operations to arrive at the same result in less time. The optimal value might vary for each implementation and it depends on the size of the table and the percentage of records that must be updated.</p>

Parameter	Recommended Setting	Description
STRIP_CTAS_DELETE_UPPER_LIMIT	10% of the BO records.	You must set the value above the peak daily incremental delta, such as 1 million records, or more.
COMPLETE_STRIP_RATIO	Change if required. Default is 60%.	<b>[Model &gt; Schema &gt; [base object] Advanced &gt; Complete Tokenize Ratio].</b> Proper analysis is required if you decide to change the default 60% value. If the volume of data change in the _STRP table is more than this percentage, the tokenization process would drop and re-create the entire _STRP table rather than (re)tokenizing only the updated records.

For more information about tuning match and merge, search the Informatica Knowledge Base for article number 357214.

### Services Integration Framework (SIF) APIs

The following table lists the recommendations for the SIF API:

Parameter	Recommended Setting	Description
Protocol	Use EJB protocol over HTTP or SOAP.	EJB Protocol is faster and reliable. See the Informatica Knowledge Base article number 138526 for further details.
Disable Paging	Set to True if not required.	SIF Request parameter: disablePaging If paging is not required (if results are not going to return many records), it is better to set this flag to true. If set to false (default value), this would incur two database calls for each SIF call.
Return Total	Do not set any value.	SIF request parameter: returnTotal If you do not require total count, it is better to not set this flag. If set to true, it would incur two database calls for each SIF call.

## Hub Server Properties

The following table lists the recommendations for the Hub Server properties:

Parameter	Recommended Setting	Description
User Profile Cache	True Default is true.	<code>cmx.server.provider.userprofile.cacheable</code> This property is found in the <code>cmxserver.properties</code> file. When you set this flag to true, once a user profile is authenticated, it is cached. Set the flag to true to suppress the need for explicit user authentication requests for every SIF or BES call.
User Profile Life Span	60000 Default is 60000.	<code>cmx.server.provider.userprofile.lifespan</code> This property is found in the <code>cmxserver.properties</code> file. Time to retain the cached user (milliseconds) before refreshing. A few minutes is adequate-avoid setting this to longer durations.
Security Access Manager (SAM) cache refresh interval	5 clock ticks Default is 5 clock ticks at a rate of 60,000 milliseconds for 1 clock tick, which is equivalent to 5 minutes.	<code>cmx.server.sam.cache.resources.refresh_interval</code> This property is found in the <code>cmxserver.properties</code> file. Refreshes the SAM cache after the specified clock ticks. To specify the number of milliseconds for 1 clock tick, use the <code>cmx.server.clock.tick_interval</code> property. For more information, see the <i>Multidomain MDM Configuration Guide</i> .
Cleanse Poller	30 (seconds) Default is 30.	<code>cmx.server.poller.monitor.interval</code> This property is found in the <code>cmxserver.properties</code> file. For every number of seconds configured, the MDM Hub Server would poll the availability of the Process Server and accordingly flag the status of the Process Server as valid or invalid. For more information, search the Informatica Knowledge Base for article number 151925.

## Infinispan Metadata Caching

The following table lists the recommendations for Infinispan metadata caching parameters, which are available in the following file:

```
<MDM Hub installation directory>/hub</cleanse>/server/resources directory/  
infinispanConfig.xml
```

:

Parameter	Recommended Setting	Description
expiration lifespan	86400000 (milliseconds)	<p>Maximum lifespan of a cache entry in milliseconds. When a cache entry exceeds its lifespan, the entry expires within the cluster.</p> <p>You can increase the lifespan for the following caches: DISABLE_WHEN_LOCK, DATA_OBJECTS, and REPOS_OBJECTS. For example, you can increase a lifespan from one hour (3600000) to one day (86400000).</p> <p>Each cache has its own default value for this parameter. To find the default values, open the <code>infinispanConfig.xml</code> file.</p>
expiration interval	300000 (milliseconds)	<p>Maximum interval for checking the lifespan.</p> <p>For example, you can increase an interval from five seconds (5000) to five minutes (300000).</p>

For more information about Infinispan parameters, search the Informatica Knowledge Base for article number 509572.

## Logging

The following table lists the recommendations for the logging:

Parameter	Recommended Setting	Description
Hub Server Logging	Set to ERROR mode.	<p>Change the <code>log4j.xml</code> file to use the ERROR mode. If clustered, update the <code>log4j.xml</code> file in all nodes under</p> <pre>&lt;INFAHOME&gt;/hub/server/conf/log4j.xml</pre> <p>After the <code>log4j</code> configuration file is updated, you can see the changes reflected in the log within a few minutes.</p>
Process Server Logging	Set to ERROR mode.	<p>Change the <code>log4j.xml</code> file to use the ERROR mode. If clustered, update the <code>log4j.xml</code> file in all nodes under</p> <pre>&lt;INFAHOME&gt;/hub/cleanse/conf/log4j.xml</pre> <p>After the <code>log4j</code> configuration file is updated, you can see the changes reflected in the log within a few minutes.</p>

For more information about logging, search the Informatica Knowledge Base for article number 120879.

## Search

The following table lists the recommendations for search:

Parameter	Recommended Setting	Description
Limit Searchable Fields	Do as needed.	Do not index unnecessary searchable fields. Multiple searchable fields increase the indexing and searching time, so configure only the required fields as searchable fields. Also keep only the required fields and facets. Facets should only be on the fields with low entropy. Also limit the number of fuzzy fields.

## Task Assignment

The following table lists the recommendations for task assignments:

Parameter	Recommended Setting	Description
<code>task.creation.batch.size</code>	Default is 1000.	In MDM 10.0 and earlier, the default value is 50. Available in <code>cmxserver.properties</code> . Sets the maximum number of records to process for each match table. If more tasks need to be assigned on the run, you can increase this value.

## Operational Reference Store and SIF APIs

The following table lists the recommendations for ORS-specific SIF API generation:

Parameter	Recommended Setting	Description
Required objects	As needed.	ORS specific API generation depends on the number of objects selected. It is preferable to add only the required objects to gain performance during the SIF API generation.
SIF API (Java Doc Generation) Heap Size	Default is 256m.	<code>sif.jvm.heap.size</code> Available in <code>cmxserver.properties</code> . Sets the heap size used during the creation of Java Doc. As Java Doc creation takes a lot of heap memory, you can increase this to a higher value if required. Note that this heap size setting is not connected to the heap size of the MDM application, which is set during the server startup.

## Informatica Data Quality

The following table lists the recommendations for Informatica Data Quality cleansing:

Parameter	Recommended Setting	Description
Batch Size	Default is 50.	<p><code>cmx.server.cleanser.number_of_recs_batch</code></p> <p>Available in <code>cmxcleanse.properties</code>.</p> <p>If the workflow supports minibatch, then you can set this value to any desirable value depending on the number of records to be cleansed at a time. If this attribute is set, then MDM automatically groups the records for cleansing.</p> <p><b>Note:</b> This property can be used in other cleansing engines if they support minibatch.</p>

For more information, search the Informatica Knowledge Base for article number 153419.

## Initial Data Load

The following table lists the recommendations for the initial data load for database or environment settings:

Parameter	Recommended Setting	Description
Distributed Matching	1	Set the <code>cmx.server.match.distributed_match=1</code> parameter in the <code>cmxcleanse.properties</code> file. This often improves performance by spreading the match load across multiple servers. Ensure that you configure the Cleanse Match servers to perform the match process in the batch mode on servers that you expect to spread the match.
UNBO Tablespace	Sufficient UNDO tablespace must be available.	The Match process utilizes a lot of UNDO tablespace. Ensure that sufficient UNDO tablespace is available in the database. You must adjust your batch match size accordingly.
Database Archive Log	Off	During IDL, switch off or disable to improve the performance.
Batch API Interoperability	Disable	During IDL, you must disable this setting.
Application Server Performance		See the sections, such as JVM , thread counts, and block size, in this guide.
Log4j.xml in siperian.performance	Off	Switch OFF the <code>siperian.performance</code> parameter in the <code>log4j.xml</code> file. For better performance, you must also ensure that the default async appenders are used in all the levels.
Match Setting		<p>Match Only Previous Rowid Objects. Enable it during the initial load match process for additional performance gain. Ensure that you disable this property for the incremental match process.</p> <p>Match Only Once. Enable it during the initial match job. You must assess the trade-off between performance and match quantity based on the characteristics of your data and your particular match requirements.</p>
Match Rule Settings	Default settings	You must use the default settings of the match configuration. If you want to modify the setting, you must analyze the requirement and implement an extensive testing.



Parameter	Recommended Setting	Description
Connection Pool Size		Ensure that sufficient connection pool is available for the sizing. See the connection pool recommendations in this guide.
Custom Indexes	Disable	Disable the custom indexes.
Constraints	Disable	To disable the constraints on NI indexes on the base object and all the indexes on XREF, you must choose the option <code>Allow constraints to be disabled</code> .
History	Disable	Disable history.
Analyze Schema		Prior to the initial data load, analyze the database schema.
Production Mode	Enable	Enable the production mode flag.
Elasticsearch Server	Disable	During the initial data load with high volume of data, you must disable the Elasticsearch server. During incremental load, you must enable the Elasticsearch server.

For further information, see the Informatica Knowledge Base article numbers 158622 and 158822.

## Recommendations for Batch Job Optimization

A batch job is a program in the MDM Hub that you can run to complete a discrete unit of work. You can launch batch jobs individually or as a group from the Hub Console or with the SIF or BES APIs. You can configure settings to optimize the performance of batch jobs.

The following table lists the different batch job parameters and their recommended settings to achieve a base-level performance:

Parameter	Recommended Setting	Description
<b>Minimum records to enable multithreading</b> Used in all the batch jobs.	Default is 1000.	<code>com.informatica.mdm.batchcontroller.BatchJob.min_rec_for_multithreading</code> This property is available in the <code>cmxcleanse.properties</code> file. When the number of records reaches the specified value, multithreading starts on all the batch jobs. To disable multithreading, you must set the property value to 1.
<b>Cleanse Thread Count</b> Used in the following batch jobs: <ul style="list-style-type: none"> <li>- Match Job</li> <li>- Generate Match Tokens process on Load job</li> <li>- Stage job</li> </ul>	Start with the number of cores available. Based on CPU utilization, number of threads can be increased. Default is 1.	Available in <b>"Process Server &gt; Threads for Cleanse Operations"</b> . Total number of threads used by the Master or Slave Process Server when executing. Generate Match Tokens after Load, Match, and Stage jobs.

Parameter	Recommended Setting	Description
<p><b>Threads for Batch Processing</b></p> <p>Used in the following batch jobs:</p> <ul style="list-style-type: none"> <li>- Automerge Job</li> <li>- Load Job</li> <li>- Batch Delete</li> <li>- Batch Unmerge</li> <li>- Batch Revalidate</li> <li>- Initially Index Smart Search</li> </ul>	<p>Specify a value that is equivalent to four times the number of CPU cores on the system on which the Process Server is deployed to start with.</p> <p>Default is 20.</p>	<p>Available in <b>"Process Server &gt; Threads for Batch Processing"</b>.</p> <p>Maximum number of threads to use for a batch process.</p> <p>For example, if the host machine has 16 CPU cores, set the Threads for Batch Processing in the Process Server registration to 64. Based on the usage and requirements, you can set a higher value. Applicable only if the Process Server is marked for batch processing.</p> <p><b>Note:</b> From the total number of threads available on the Process Server, dedicate n threads for Batch jobs by setting a value for the property number of threads for Batch processing.</p>
<p><b>Controller Thread Time Out</b></p> <p>Used in the following batch jobs:</p> <ul style="list-style-type: none"> <li>- Automerge Job</li> <li>- Load Job</li> <li>- Batch Delete</li> <li>- Batch Unmerge</li> <li>- Batch Recalculate</li> </ul>	<p>300000 (5 minutes).</p> <p>Default is 300000.</p>	<p><code>com.informatica.mdm.loadbalance.ControllerThread.timeout</code></p> <p>This property is found in the <code>cmxcleanse.properties</code> file.</p> <p>When distributing the load to different slave Process Servers, after the last block is sent to a slave Process Server, all slave Process Servers which are processing the blocks <b>MUST</b> complete the job within the timeout period.</p> <p><b>Note:</b> If not completed, such blocks are marked with 'No Action' in the batch result. Note that, the batch is not marked as failed because the remaining blocks are successfully loaded.</p>
<p><b>Load analyze threshold rate</b></p> <p>Used in the following batch jobs:</p> <ul style="list-style-type: none"> <li>- Automerge Job</li> <li>- Load Job</li> <li>- Batch Delete</li> <li>- Batch Unmerge</li> <li>- Batch Recalculate</li> </ul>	<p>Default is 10.</p>	<p><code>cmx.server.batch.load.analyze_threshold_rate</code></p> <p>Available in <code>cmxserver.properties</code></p> <p><b>For ORACLE only. Available from MDM 10.0 HotFix 1.</b></p> <p>Specifies the frequency that the MDM Hub gathers analytical statistics for tables affected by a batch Load job. Set to 0 to disable statistic collection. Set to 1 to collect statistics only at the end of a Load job for base object and cross-reference tables.</p> <p>For example, if the threshold is 10, then statistics would be gathered at every 10<sup>n</sup> records. For example, new statistics would be gathered whenever the insert record count reaches 100, 1000, 10000, and so on.</p>

Parameter	Recommended Setting	Description
<p><b>Recycler Thread Max Idling</b></p> <p>Used in the following batch jobs:</p> <ul style="list-style-type: none"> <li>- Automerge Job</li> <li>- Load Job</li> <li>- Batch Delete</li> <li>- Batch Unmerge</li> <li>- Batch Recalculate</li> </ul>	<p>300000 (5 minutes).</p> <p>Default is 300000 (5 minutes).</p>	<p>com.informatica.mdm.batchserver.RecyclerThread.max_idling</p> <p>This property is found in the <code>cmxcleanse.properties</code> file.</p> <p>If a slave Process Server is processing a block of batch job and is idle for a duration specified in this attribute then the specific thread is marked as 'dead.'</p> <p><b>Note:</b> If a slave Process Server is timed out as noted earlier, the corresponding block is marked with 'No Action' in the batch result. Note that the batch is not marked as failed as the remaining blocks are successfully loaded.</p>
<p><b>Automerge:</b></p> <p>Automerge Threads Per Job</p>	<p>Default is 1.</p>	<p>cmx.server.automerge.threads_per_job</p> <p>This property is found in the <code>cmxserver.properties</code> file.</p> <p>Maximum number of threads distributed across different Process Servers to process the automerge job.</p> <p>For example, if this value is 20, automerge would be distributed across two Process Servers each with 10. The distribution depends on factors such as CPU weightage of the Process Server and other jobs running on the Process Server.</p> <p>This value must be less than the value in 'Threads for Batch' attribute specified for the Process Server.</p> <p><b>Note:</b> The optimum value for a database server with a 16 core processor and a solid-state drive (SSD) set up in a RAID is 20. Based on CPU utilization on different Process Servers, you can increase the threads.</p>
<p><b>Automerge:</b></p> <p>Automerge Block Size</p>	<p>Default is 250.</p>	<p>cmx.server.automerge.block_size</p> <p>This property is found in the <code>cmxserver.properties</code> file.</p> <p>Maximum number of records to be sent for merges to each Process Server in one block.</p> <p>For example, consider the scenario of two Process Servers with 1000 records to be merged. If this value is 250, each Process Server gets 250 records first followed by another 250 records next.</p> <p>Increasing this value can provide performance improvement based on how powerful the application servers and database servers are.</p>

Parameter	Recommended Setting	Description
<b>Load:</b> Batch Threads Per Job	Default is 10.	<code>cmx.server.batch.threads_per_job</code> This property is found in the <code>cmxserver.properties</code> file. Maximum number of threads distributed across different Process Servers to process the load job. For example, if this value is 20 then load process would be distributed across two Process Servers each with 10. The distribution depends on factors such as CPU weightage of the Process Server and other jobs running on the Process Server. This value must be less than the value in 'Threads for Batch' attribute specified for the Process Server. <b>Note:</b> The optimum value for a database server with a 16 core processor and a solid-state drive (SSD) set up in a redundant array of independent disks (RAID) is 20. Based on CPU utilization on different Process Servers, you can increase the threads.
<b>Load:</b> Batch Block Size	1000. Default is 250.	<code>cmx.server.batch.load.block_size</code> This property is found in the <code>cmxserver.properties</code> file. Maximum number of records in one block for each Process Server. For example, consider the scenario of two Process Servers with 1000 records to load. If the specified value is 250, each Process Server gets 250 records first followed by another 250 records next. Increasing this value can provide performance improvement based on how powerful the application servers and database servers are.
<b>Load:</b> Threads per job for generate tokens, if 'Generate Match Tokens on Load' attribute is enabled on the base object	Same as "Threads for cleanse processing".	See 'Threads for Cleanse Processing' attribute described earlier. Note that, this thread attribute is different from the core threads per job attribute of the load job described earlier. If 'Generate Match Tokens on Load' is not selected, this attribute does not have any impact on the performance of the Load job.
<b>Load:</b> Validate Lookup for the timeline enabled BO.	Default is false.	<code>cmx.server.validate.lookup.timeline.onload</code> This property is found in the <code>cmxserver.properties</code> file. For timeline enabled BO, the XREF is validated for the correct timeline among multiple timeline values, before selecting a valid lookup row. If the load job is taking more time and heap memory, when you have the timeline enabled BO and inserts large lookup stream cluster, then you must enable this property in the <code>cmxserver.properties</code> file.

Parameter	Recommended Setting	Description
<b>Batch Recalculate (SIF API Request):</b> Recalculate Threads Per Job	Same property, re-used from LOAD Job. See LOAD Job section for more details.	<code>cmx.server.batch.threads_per_job</code> This property is found in the <code>cmxserver.properties</code> file. Same property, re-used from LOAD Job. See LOAD Job section for more details.
<b>Batch Recalculate (SIF API Request):</b> Recalculate Block Size	Default is 250.	<code>cmx.server.batch.recalculate.block_size</code> This property is found in the <code>cmxserver.properties</code> file. Maximum number of records to be sent, to recalculate BVT, to each Process Server in one block. For example, consider the scenario of two Process Servers with 1000 records to be recalculated. If this value is 250, each Process Server gets 250 records first followed by another 250 records next. Increasing this value can provide performance improvement based on how powerful the application servers and database servers are.
<b>Batch Recalculate (SIF API Request):</b> Threads Per Job	Same property, re-used from LOAD Job. Refer to LOAD Job section for more details.	<code>cmx.server.batch.threads_per_job</code> Available in <code>cmxserver.properties</code> Same property, re-used from LOAD Job. Refer to LOAD Job section for more details.
<b>Batch Unmerge (SIF API Request):</b> Unmerge Block Size	Default is 250.	<code>cmx.server.batch.batchunmerge.block_size</code> This property is found in the <code>cmxserver.properties</code> file. Maximum number of records to be sent for unmerges, to each Process Server in one block. For example, consider the scenario of two Process Servers with 1000 records to be unmerged. If this value is 250, each Process Server gets 250 records first followed by another 250 records next. Increasing this value can provide performance improvement based on how powerful the application servers and database servers are.
<b>Batch Delete (SIF API Request):</b> Threads per job	Same property, re-used from LOAD Job. See LOAD Job section for more details.	<code>cmx.server.batch.threads_per_job</code> This property is found in the <code>cmxserver.properties</code> file. Same property, re-used from LOAD Job. See LOAD Job section for more details.

Parameter	Recommended Setting	Description
<b>Batch Delete (SIF API Request):</b> Delete Batch Block Size	Default is 250.	<code>cmx.server.batch.delete.block_size</code> This property is found in the <code>cmxserver.properties</code> file. Maximum number of records to be sent for deletion, to each Process Server in one block. For example, consider the scenario of two Process Servers with 1000 records to be deleted. If this value is 250, each Process Server first gets 250 records and then another 250 records. Increasing this value can provide performance improvement. This performance improvement depends on how powerful the application servers and database servers are.
<b>Tokenize:</b> Tokenization File Loader Option	Default is true.	<code>cmx.server.tokenize.file_load</code> This property is found in the <code>cmxcleanse.properties</code> file. Applicable for Oracle and Db2. The MS SQL Server always uses Java. If true, Db2 file loader or Oracle SQL Loader is used to load the records during the tokenization job. If file writing is causing performance issue, this can be changed to false, thereby, data is directly written to the database every time instead of file loader option. Generally, file loader is faster than the direct database writes. You might choose the option according to your environment.
<b>Stage:</b> Threads per job	See 'Cleanse Thread Count' attribute described earlier.	See 'Cleanse Thread Count' attribute described earlier.
<b>Stage:</b> Cleanse Minimum Distribution	Default is 1000.	<code>cmx.server.cleansse.min_size_for_distribution</code> This property is found in the <code>cmxcleanse.properties</code> file. The MDM Hub distributes the cleanse job across different cleanse server only if the number of records is higher than this minimum size. When distributing the load, each slave Process Server would use the Cleanse Thread Count for the number of worker threads.

Parameter	Recommended Setting	Description
<b>Stage:</b> Stage JDBC Loader	Default is false. Usually, file writing must be faster than the direct database writing.	<code>cmx.server.java_jdbc_loader</code> Applicable for Oracle and Db2. Default is false. This property is found in the <code>cmxcleanse.properties</code> file. If true, Db2 and Oracle use direct database connections during the stage job instead of Db2 file loader or Oracle SQL loader options If file writing is causing performance issue, this can be changed to true. On doing so, data gets directly written to the database every time instead of file loader option. Note that, generally, file loader is faster than the direct database write. You might choose the option according to your environment.
<b>Match:</b> Threads per job	See 'Cleanse Thread Count' attribute described earlier.	See 'Cleanse Thread Count' attribute described earlier.
<b>Match:</b> Match Distribution Flag	Enable this flag to 1, if the MDM Hub has to distribute the match job load across different cleanse servers.	<code>cmx.server.match.distributed_match</code> This property is found in the <code>cmxcleanse.properties</code> file. The MDM Hub distributes the match job across different cleanse server only if this value is set to 1. When distributing the load, each slave Process Server would use the Cleanse Thread Count for the number of worker threads.
<b>Match:</b> Match File Loader Option	Default is true. Usually, file writing must be faster than the direct database writing.	<code>cmx.server.match.file_load</code> Applicable for Oracle and Db2. This property is found in the <code>cmxcleanse.properties</code> file. If true, Db2 file loader or Oracle SQL Loader is used to load the records during the tokenization job. Microsoft SQL Server always uses the Java Database Connectivity (JDBC) loader. If file writing causes performance issue, change it to false to directly write data to the database every time instead of file loader option. Generally, file loader is faster than the direct database writes. You might choose the option according to your environment.

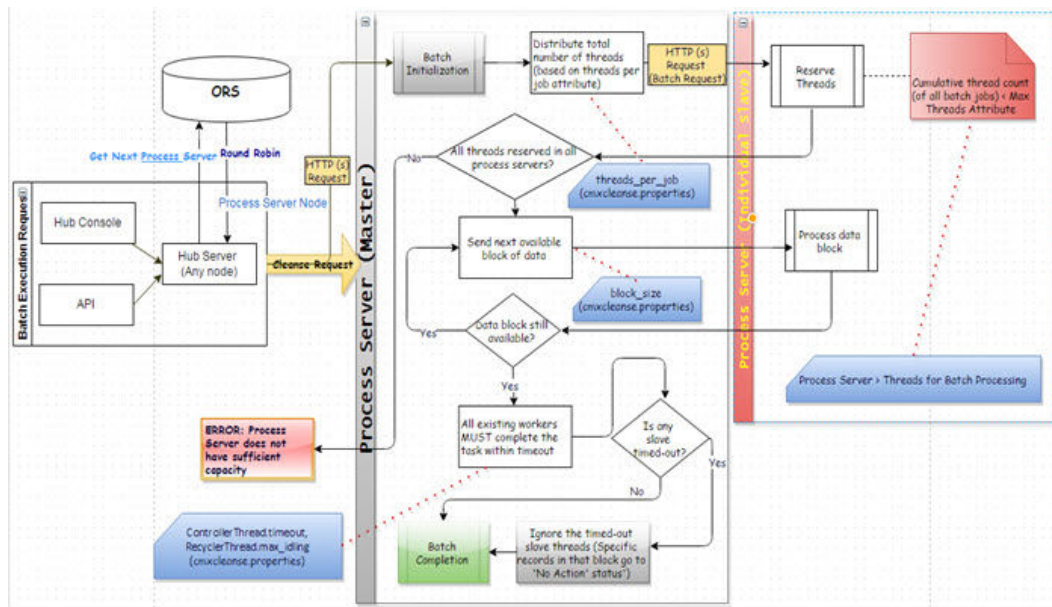
Parameter	Recommended Setting	Description
<b>Match:</b> Match Loader Batch Size	Default is 250.	<code>cmx.server.match.loader_batch_size</code> This property is found in the <code>cmxcleanse.properties</code> file. Applicable if JDBC load is used in match processing instead of file loader option. Maximum number of records to be sent for match in each worker thread. Increasing this value can provide performance improvement based on how powerful the application servers and database servers are.
<b>Match:</b> Match Elapsed Time	Default is 20 (minutes).	<b>Hub Console &gt; Base Object &gt; Max Elapsed Match Minutes.</b> The execution timeout in minutes when executing a match rule. If this time is reached, the match process will exit. This must be increased only if the match rule and the data are very complex. Generally rules must be able to complete within 20 minutes.
<b>Match:</b> Match Batch Size	Default is 20000000.	<b>Hub Console &gt; Base Object &gt; Match/Merge Setup &gt; Number of rows per match job batch cycle.</b> Maximum number of records to be processed by the MDM Hub for matching. This number would affect the duration of match process. Also, lower the match batch size, you have to run the match process more times. <b>Note:</b> When running large Match jobs with large match batch sizes, if there is a failure of the application server or the database, you must re-run the entire batch.
<b>Match:</b> Maximum records per ranger node	Default is 5000.	<code>max_records_per_ranger_node</code> This property is found in the <code>cmxcleanse.properties</code> file. Number of records per match ranger node (limits memory use). Ranger is an internal component used within the match process where sorting and merging operations are performed based on this maximum records attribute. You can optimize this value to get better performance based on the memory available in your application server.
<b>Initially index smart search data:</b> Block Size	5000. Default is 250.	<code>cmx.server.batch.smartsearch.initial_block_size</code> This property is available in the <code>cmxserver.properties</code> file. Maximum number of records that the initially index smart search data batch job can process in each block. This property is not applicable through regular indexing outside this specific batch job. Based on the power of your application server, you can increase this value to improve the performance.



Parameter	Recommended Setting	Description
<b>Initially index smart search data:</b> Smart search threads	Default is 10. Same property, re-used from LOAD Job.	<code>cmx.server.batch.threads_per_job</code> This property is available in the <code>cmxserver.properties</code> file. Maximum number of threads distributed across different Process Servers to process the batch job initially index smart search data. You can increase this value to achieve more performance during this batch job. This property is not applicable for regular indexing outside this specific batch job.
<b>Load job with indexing smart search optimization</b>	True. Default is false.	<code>cmx.ss.index.build.optimization=true</code> This property is available in the <code>cmxserver.properties</code> and <code>cmxcleanse.properties</code> files. If the smart search indexing from a load job takes more time in a specific model, set this property for a better optimization on the indexing.

## Multithreaded Batch Job – Process Flow

The following image shows the process flow of Automerge, Load, Batch Delete, Batch Unmerge, and Batch Recalculate jobs:



### Master Process Server

The following list describes the properties to configure for the multithreaded batch jobs:

- Threads allocated in Slave Process Server #1 + Threads allocated in Slave Process Server #2 + ... [All Slave Process Servers] must be equal to the specific "Threads Per Job" parameter in `cmxcleanse.properties` file.

The following properties are related to threads:

- `cmx.server.automerge.threads_per_job`
- `cmx.server.batch.threads_per_job`

- Each Slave Process Server would get number of records as specified in the block size.

The following properties are related to block\_size:

- `cmx.server.automerge.block_size`
- `cmx.server.batch.block_size`
- `cmx.server.batch.recalculate.block_size`
- `cmx.server.batch.batchunmerge.block_size`
- `cmx.server.batch.delete.block_size`

- After the last block is sent to the next available Slave Process Server, all Slave Process Servers that process the blocks MUST complete the job within the timeout period.

`com.informatica.mdm.loadbalance.ControllerThread.timeout`

## Slave Process Server

The following list describes the properties to configure for multi-threaded batch jobs:

- Threads allocated for Batch Job 1 + Threads allocated for Batch Job 2 + ... [All parallel batch jobs] must not exceed "Threads for Batch Processing" of the specific Process Server.
- The `com.informatica.mdm.batchserver.RecyclerThread.max_idling` property specifies the idle time for a Process Server thread. The Process Server recycles the thread when it is idle for more than the configured value.

# Recommendations for the Hub Console Optimization

The Hub Console parameters can be optimized for better performance.

The following table lists the recommendations for the Hub Console optimization:

Parameter	Recommended Setting	Description
Client Java Version (JRE)	As specified in the Product Availability Matrix (PAM).	<p>Ensure that the client Java version is as specified in the PAM document for the product.</p> <p>It is not required for the Hub Console to pick the latest JRE that the client box has. The JRE version selected depends on the PATH variable. To ensure that the Hub Console uses the correct JRE, temporarily enable the console log as listed in the Informatica Knowledge Base article.</p> <p>When enabled, the next launch of Hub Console opens a Java console window where you can see the used JRE version in the top of the console.</p>
Console log	Disable (Default).	<p>Disable the logging by launching the javaws -viewer option on the client box in the run command.</p> <p>In the javaws window, go to Java Control Panel in the Advanced tab and perform the following steps:</p> <ul style="list-style-type: none"> <li>- Clear 'Enable tracing' and 'Enable logging' options in the Debugging section.</li> <li>- Check 'Hide Console' option in the Java Console section.</li> </ul>
High Data Volume	See the parameter "High Volume of Data in C_REPOS_TABLES with historical data" in Database - General Recommendation Section.	<p>Having huge number of records in the tables listed in the referenced section might cause multiple issues on Hub console.</p> <p>The following lists are the issues that might be include:</p> <ul style="list-style-type: none"> <li>- batch groups screen loads too slowly.</li> <li>- base object record is saved too slowly.</li> <li>- cleanse function is saved too slowly.</li> </ul>
Network latency	Good network connection.	<p>Hub Console communicates with the Hub Server (application server) frequently. Therefore, the Hub Console must process a considerable amount of data to and from the Hub Server. Have good network connection between the client box and the application server.</p>
Console mode	<code>cmx.appserver.console.mode=https (/ http)</code>	<p>MDM 10.3 and earlier versions use http(s) instead of the default EJB.</p> <p><b>Note:</b> The MDM 10.4 is configured only with http(s).</p>

For more information, search the Informatica Knowledge Base for article numbers 310913 and 139923.

# Recommendations for Data Director and SIF or BES API Optimization

The following table lists the recommendations for Data Director and Services Integration Framework (SIF) or Business Entity Services (BES) API calls optimization:

Parameter	Recommended Setting	Description
<b>Read BES call:</b> Parallel read call	Default is 30. Parallel threads.	<code>cmx.server.enrichcopager.thread_pool</code>  This property is available in the <code>cmxserver.properties</code> file. Sets the number of threads the EnrichCoPager property uses from the thread pool to perform parallel ReadCO operations. If you want to disable multithreading, set this value to 1. Based on your requirement, increase the value for better performance.
<b>Read BES call:</b> Minimum record to enable multithreading.	Default is 2. Number of records returned.	<code>cmx.server.enrichcopager.min_rec_for_multithreading</code>  This property is available in the <code>cmxserver.properties</code> file.  Sets the minimum number of records to return before the EnrichCoPager property uses multithreading.
<b>Hierarchy REST API:</b> Maximum depth for hierarchy search.	Default is 100.	<code>cmx.server.hierarchy.max-search-depth</code>  This property is available in the <code>cmxserver.propertiesHub</code> file.  Maximum depth searched when you use hierarchy REST APIs to find a hierarchy path. If you set to a higher value, it causes performance impact on hierarchy search.
<b>Hierarchy REST API:</b> Maximum width for exporting the hierarchy.	Default is 1000000.	<code>cmx.server.hierarchy.max-search-width</code>  This property is available in the <code>cmxserver.propertiesHub</code> file. Maximum width of the searched hierarchy to include when you use the hierarchy REST APIs to export. If you set to a higher value, it causes performance impact on hierarchy export.
<b>BE internal jobs:</b> Minimum record to enable multithreading.	Default is 30.	<code>com.informatica.mdm.batchcontroller.BEBatchJob.min_rec_for_multithreading</code>  This property is available in the <code>cmxcleanse.propertiesProcess server</code> file. The least number of records trigger a multi-threaded job. If the number of records is below the threshold, the job runs on a single thread. It is specific for jobs that are indirectly triggered from user actions, such as find and replace.  In an instance where you consistently see that the users are waiting for a long time to implement UI operations for a smaller number of records, it might be worthwhile to decrease this number. But note that, if you have multiple users performing the same operation with a lower threshold value, you require a higher number of concurrent threads to handle the work, even if it is for a shorter duration.

Parameter	Recommended Setting	Description
<b>Cache:</b> Lookup Cache Update Period	Default is 300000 (5 minutes).	lookupCacheUpdatePeriod in CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL  The amount of time the lookup data is cached in the server. Change it to a higher value if your lookup data is not going to change in real time frequently. Keep a higher value as it would improve the performance as the cache refresh is reduced.
<b>Cache:</b> SAM Cache Update Period	Default is 600000 (10 minutes).	samCacheUpdatePeriod in CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL  The amount of time the data about Security Access Manager (SAM) roles (no users) is cached in the server. Change it to a higher value if your role data is not going to change in real time very frequently. Keeping a higher value improves the performance, as the cache refresh is reduced.
<b>Cache:</b> Cache Refresh Daemon Thread Monitor (both Lookup and SAM Cache)	Default is 5000 (5 seconds).	threadSchedulerIdleTime in CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL  Time interval of the daemon thread that checks and triggers the Lookup cache and the Security Access Manager cache operations based on their update period. That is, every 5 seconds, the daemon thread would check and trigger the lookup cache and Security Access Manager cache if their update period is reached.  This period can be increased to improve performance, if it is not necessary to check the SAM and Lookup Cache time periods every 5 seconds.
<b>Search Operation:</b> Maximum number of records retrieved at a time	Default is 100.	serverPageSize in CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL  Number of records to be retrieved from the Hub Server when a search operation is performed. Note that even though Hub Server shows 10 records per page, internally it loads these many numbers of records (given by this parameter). And the Hub Server caches the remaining records that are not shown in the view. <b>Note:</b> This parameter is used in Search Results and in functional areas such as potential matches, hierarchy relationship records, wherever a list of records is retrieved for the user.  A higher value could make the search results performance better, provided the Hub Server can handle the load and cache.
<b>Search Operation:</b> Maximum number of records during search results export	Default is 5000.	maxSearchResultsExportedRows in CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL  Number of rows that are exported when using the export option in the search results page. Do not increase the value unless there is a valid use case. The greater the value, the lower will be the performance of the export operation.

Parameter	Recommended Setting	Description
<b>Search Operation:</b> Case Insensitive Search	Default is false. Adds performance overhead when enabled.	<code>case.insensitive.search</code> This property is found in the <code>cmxserver.properties</code> file. If set to true, users can enable the case insensitiveness attribute for individual columns in the base object, to enable case insensitive query search in Data Director. Enabling this flag would incur a new index created for such columns. As index management has its own performance overhead, use this flag with caution. <b>Note:</b> If there is any system column that you want to include with case insensitive search, create a functional index manually for such columns. For more information, search the Informatica Knowledge Base for article numbers 154132 and 154139.
<b>Search Operation:</b> Default effective date	Default is true.	<code>isEffectiveDateIncluded</code> in <code>CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL</code> By default, effective date is automatically populated with current date. This may impact performance for normal queries that do not need effective date. In such scenarios, you can set this flag to false to remove the default value. For more information, search the Informatica Knowledge Base for article number 347953.
<b>Search Operation:</b> Remove Duplicates Flag	As applicable. Default is true.	<code>cmx.server.remove_duplicates_in_search_query_results</code> Default is true. This property is found in the <code>cmxserver.properties</code> file. By default, Data Director filters the duplicate results from the search query that joins child records. However, this flag if enabled, can affect the performance depending on the time it takes to de-duplicate the search results. If duplicates are not an issue in the search result, you can disable this flag. For more information, search the Informatica Knowledge Base for article number 122829.
<b>Data View:</b> Asynchronous Child Load	Default is false. Adds performance overhead when enabled.	<code>asyncChildLoading</code> in <code>CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL</code> If true, all child records are opened (expanded) automatically when you open a parent subject area.
<b>Hierarchy:</b> Maximum records in Relationship Flyover	Default is 10000.	<code>sif.api.hm.flyover.max.record.count</code> This property is found in the <code>cmxserver.properties</code> file. This number denotes the maximum number of relationships that are to be shown in Relationship Flyover in Data Director. Internally, the best version of the truth (BVT) calculation is done for all such relationships. Therefore, having a high number here would impact the performance of relationship flyover. For more information, search the Informatica Knowledge Base for article number 157628.

Parameter	Recommended Setting	Description
<b>Hierarchy:</b> Inactive relationship records	Default is false.	<code>hmlInactiveRelationshipsAvailable</code> in <code>CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL</code> If true, Hierarchy Manager in the Data Director would show all the inactive relationship records also. Do not keep this value as true unless there is a valid use case to do so.
<b>Task Context:</b> Parallel Promote Threads	Same as the number of Hub Server Cores.	<code>maxParallelPromoteThreads</code> in <code>CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL</code> Default is 1. Determines the maximum number of parallel threads when a task is approved (Promote API). This is helpful when you promote parent, child, and grandchild subject areas together.
<b>Task Context:</b> Parallel BVT Threads	Same as the number of Hub Server Cores.	<code>maxParallelBvtThreads</code> in <code>CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL</code> Default is 1. Determines the maximum number of parallel threads when a task is viewed (PreviewBVT API).
<b>Task Context:</b> Maximum tasks assigned to a user	Default is 25.	<code>sip.task.maximum.assignment</code> This property is found in the <code>cmxserver.properties</code> file. Maximum number of tasks assigned to each user when automatic task assignment is enabled. Keeping this a high number would impact the performance of task context screens in Data Director. For more information, search the Informatica Knowledge Base for article number 134833.
<b>ActiveVOS, Identity Service</b>	Enable the identity caching from the ActiveVOS Console.	Identity service uses cache for subsequent calls within a specified time or maximum entries specified in the cache. The cache boosts the ActiveVOS performance.
<b>Bulk Import:</b> Bulk Import Threads	Same as the number of Hub Server Cores.	<code>maxImportThreads</code> in <code>CMX_SYSTEM.C_REPOS_DS_PREF_DETAIL</code> Default is 5. Maximum number of threads to be used while using the Bulk Import module.
<b>Bulk Import:</b> User Exit	As applicable.	PUT user exits are called for each and every record individually. Having a complex logic in the PUT user exit would deteriorate the performance. Use user exit with caution.
<b>Bulk Import:</b> Children, Foreign Keys, and Lookups	As applicable.	More number of children, foreign key relationships, and lookup columns in the data import would impact the performance.

Parameter	Recommended Setting	Description
<b>Search Operation and HM Hierarchy:</b> BVT Optimization Limit	<code>temptableTimeToLive</code> default is 20. <code>MaxRowCount</code> default is 5000.	<p>This property is found in the <code>cmxserver.properties</code> file. Applicable for SearchQuery, SearchHMQuery, GetOneHop, and GetEntityGraph SIF APIs (also the relevant Data Director modules) when effective dates are specified.</p> <p><code>searchQuery.buildBvtTemp.MaxRowCount</code>: Specifies the maximum number of records to be used in the BVT calculation when processing a search operation with effective date filters.</p> <p><code>sif.search.result.query.temptableTimeToLive.seconds</code>: Specifies the number of seconds, the temporary table of the specific API must reside in the server. That is, the earlier mentioned APIs would store the full results in its own temporary tables which are cleaned-up automatically based on this timeout. Pagination would work based on the data from this temporary table instead of re-querying the API repeatedly from the core tables. For search operations involving larger number of records, it is recommended to increase this timeout value.</p> <p>Example:            If a hierarchy entity can have more than 10,000 related records, set the values as shown:</p> <pre>sif.search.result.query.temptableTimeToLive.seconds=3600 searchQuery.buildBvtTemp.MaxRowCount =100000</pre> <p>For more information, see the <i>Multidomain MDM Configuration Guide</i> and the <i>Multidomain MDM Services Integration Framework Guide</i>.</p>
Traffic compression	Default is TRUE.	<code>cmx.bdd.server.traffic.compression_enabled</code> This property is found in the <code>cmxserver.properties</code> file. Specifies if Data Director server traffic compression is enabled. For more information, see the <i>Multidomain MDM Configuration Guide</i> .
<b>SearchMatch API:</b> Thread count	Default is 1.	<code>cmx.server.match.searcher_thread_count</code> Available in <code>cmxcleanse.properties</code> file. Configures the thread count for the SearchMatch API. Optimal value can be decided based on your environment. You may need to test with an incremental value of thread count until you see an optimal performance.
<b>SearchMatch API:</b> JDBC direct connection	false	<code>cmx.server.batch.use_direct_connection=false</code> This property is found in the <code>cmxserver.properties</code> file. This property helps you to improve the performance of realtime SearchMatch API. <b>Note:</b> Batch jobs and APIs use this property. Modifying the property might impact your batch jobs. See the Informatica Knowledge Base article number 301843 for further details.



# Recommendations for Elasticsearch

The Elasticsearch parameters can be optimized for better performance.

The following table lists the recommendations for Elasticsearch:

Parameter	Recommended Setting	Description
Heap Size	-Xms8192m -Xmx8192m	Based on your requirement, contact the system administrator to increase the values as per the available memory on the system. However, do not exceed 31 GB of heap. Instead, you can add additional clustered nodes. <b>Note:</b> Informatica recommends that you have the same value for Xms and Xmx.
Elasticsearch indexing commit interval during the initial index job.	Default is 30 seconds.	<code>es.index.refresh.interval=30</code> This property is found in the <code>cmxserver.properties</code> file. Sets the interval in seconds for Elasticsearch to commit the changes to the data after an Initially Index Smart Search Data batch job is run. The data is available for search after this time interval. Modifying this value has a performance impact. Before you modify this property, you must contact Informatica Global Customer Support.
Elastic search server number of shards	Default is 2.	<code>search.provisioning.numshards=2</code> Number of shards to create on your Elasticsearch environment. The value depends on the maximum number of shards and the total number of nodes. For example, if the maximum number of shards is 1 and the number of nodes is 3, you can create 3 shards. Based on the requirement and available nodes, you can increase the value. If you want to modify this value, you require re-indexing.
Elasticsearch server number of replicas.	1 or more. Default is 0 The number of total shards, including the primary and replicas are evenly divided by the number of nodes in the elastic cluster.	<code>search.provisioning.numreplicas=1</code> Number of copies of the Elasticsearch engine documents that you want to create on different nodes. Use the replication factor to create multiple copies of the documents in the shards of different nodes. You require multiple copies of the documents to achieve high availability if one or more nodes shut down unexpectedly. For example, if the replication factor is 2, you get two copies of the documents in two nodes.
Elastic search maximum connection per host.	Default is 100.	<code>ex.max.conn.per.host=100</code> Sets the maximum number of Elasticsearch nodes that you want to connect to the host. The maximum number of total connection value for the Elasticsearch server is the number of nodes x maximum connection. For example, if you use 3 nodes for Elasticsearch server, then 3 x 100 = 300 maximum number of total connection value is available for the Elasticsearch server.

Parameter	Recommended Setting	Description
Number of Elasticsearch nodes.	Odd number. For example: 1, 3, 5.	To avoid split-brain issue, you must set the number of nodes in odd numbers in the Elasticsearch server cluster.
Elasticsearch server maximum worker threads.	Default 1.	<code>ex.max.threads=1</code> Sets the maximum number of threads that you want the Apache asynchronous non-blocking receiver to use for each node in the Elasticsearch cluster. You can set the maximum of number of processors available for the node. Contact Informatica Global Customer Support before modifying this property.
Elasticsearch server facets.	Limit the use of facets and choose this fields with less unique values.	Search performance is good with limited facets.
Elasticsearch server searchable fields.	Mark only the required fields as searchable.	If you have more searchable fields, it slows down the performance of jobs that needs to delete the indexes.
Elasticsearch server disk space monitoring.	You must have a monitoring tool to keep an eye on the disk space of the Elasticsearch server.	If you have less than 15% of free disk space, Elasticsearch server marks all indexes as read.
Elasticsearch server maximum virtual memory areas.	<b>OS command:</b> <code>sysctl -w vm.max_map_count=262144</code>	The MMap FS type stores the shared index on the file system by mapping a file into memory. Ensure that you set the values based on your requirement. See the Informatica Knowledge Base article number 589503 for further details.

## Recommendations for Customer 360

The following table lists the recommendations for Customer 360:

Parameters	Recommended Setting	Description
Searchable fields configuration	Remove the searchable fields that are not required.	By default, Customer 360 has many fields configured as searchable in the Provisioning tool. The file cause indexing performance issues. You must review all the fields and remove the fields that are not required.

# Recommendations for Environment Validation Tools and Utilities

You can verify the current environment and identify the area of improvement by using different tools and utilities.

The following table lists the recommendations related to support tools :

Parameter	Recommended Setting	Description
Test IO (utility) applicable for Oracle	Available in the Resource Kit Utility.	Test the Input or Output (I/O) capability of the data files of the database tablespaces. See the Informatica Knowledge Base article number 506035. The test result must meet or exceed the standards documented for 'Good Performance' seen in the Informatica Knowledge Base article number 506035.
Match Analysis Tool	Available in the Hub Console. To access the tool from the Hub Console, click <b>Help &gt; MDM Support Tools</b> .	For match relates issues, analyze the match process. See the Informatica Knowledge Base article number 239270 for further details.
User Migration Tool	Available in the Hub Console. To access the tool from the Hub Console, click <b>Help &gt; MDM Support Tools</b> .	Use the tool to migrate users across different environments and to bulk insert or update the existing users. See the Informatica Knowledge Base article number 158095 for further details

# APPENDIX A

## Glossary

### **\_PKQ sequence**

Sequence that is used to populate the ROWID\_OBJECT of base object record. For example, the C\_PARTY BO uses C\_PARTY\_PKQ sequence is used to populate the party records.

### **<INFAHOME>**

Physical location where the Hub Server and the Process Server are installed.

### **heap size**

Amount of memory allocated to Java processes which are created on the same JVM.

### **Hub Server**

The server that manages core and common services for the MDM Hub.

### **master database**

Database instance that stores metadata to manage individual domain schemas called ORS schemas. The database instance is unique to each the MDM Hub environment.

### **ORS**

Operational Reference Store. A database instance where you store domain data.

### **Process Server**

The server that cleanses and matches data and performs batch jobs such as load, recalculate BVT, and revalidate.

### **response latency**

The time duration between request and response.

### **Tracert**

Network diagnostic tool that displays the route to a destination with the transit delay information.

### **user profile**

An internal object of the MDM Hub that stores the user details including authentication and associated roles.

### **Xms**

A JVM parameter that specifies the initial Java heap size.

**Xmx**

A JVM parameter that specifies the maximum Java heap size.

**XREF**

Cross reference. Data that relates the base object data with the relevant source information.

**Xss**

A JVM parameter that specifies the memory assigned for stacking the threads created within the Java process.

# INDEX

## A

acronyms [7](#)

## G

glossary [60](#)

## I

IBM Db2 recommendations  
database file configuration parameters [21](#)

## J

Java recommendations  
database connection pool [10](#)  
JVM settings [10](#)

## M

MDM Hub recommendations  
Hub Server properties [26](#)  
Infinispan [26](#)

MDM Hub recommendations (*continued*)  
JBoss Cache *See Infinispan*  
metadata cache *See Infinispan*  
ORS configuration [26](#)  
ORS specific SIF API generation [26](#)

## O

Oracle database recommendations  
INIT.ORA recommendations [17](#)  
RAC recommendations [15](#)

## R

recommendations  
batch job optimization [41](#)  
Data Director and SIF optimization [52](#)  
database [13](#)  
environment validation tools and utilities [59](#)  
Hub Console optimization [51](#)  
IBM Db2 [21](#)  
Java [10](#)  
MDM Hub [26](#)  
Microsoft SQL Server [20](#)  
Oracle database [15](#)