

# Informatica Optimization

## Cheat Sheet

**datAvail**

BI/Analytics • Applications • Databases

Reference this quick cheat sheet for information on optimizing Informatica performance with health checks and performance tuning.

### Health Checks

Datavail has developed a five-pronged strategy to assess the performance of Informatica and your databases, known as the **"5S Approach."**

# 1

#### Source Code

Identify long-running processes and how long the popular processes take to run.

# 2

#### Statistics

Are the database statistics being updated frequently enough?

# 3

#### Space/Indexing

How much total space, temporary space, and storage space do you have?

# 4

#### Sessions

Are people consuming too many resources on your system sessions?

# 5

#### Scheduled Processes

Are your processes overlapping and creating performance issues?

## Common Questions to Ask About the Health of Your Data Integration Environment and Repository

- ✓ What is the origination of the data and data structures (e.g. [lift and shift](#))?
- ✓ Is the database optimized for its current platform: IBM Netezza, Microsoft SQL Server, Oracle, etc.?
- ✓ What are the strategies for indexing and partitioning?
- ✓ Is there proper session connection management at the database and application layers?
- ✓ If parallelism is possible, does it match the capability and architecture of your choice of platform?
- ✓ Are the ETL jobs optimized for best performances, are there any performance bottlenecks?

## Common Issues with Health Checks

- ✓ Consider the possibility that DML commands in SQL (i.e. SELECT, INSERT, UPDATE, and DELETE) are causing problems with fragmentations and statistics.
- ✓ The lack of a proper TEMP tablespace will cause failures that you can diagnose by reading watch logs or using dedicated tools such as Oracle Automatic Storage Management (ASM).
- ✓ File system lookups during this stage can be slow if there is a lack of indexing, and file systems can become fragmented depending on the operating system.
- ✓ Consider potential issues such as OLAP write speeds and I/O, as well as indexing and statistics concerns.
- ✓ Make sure that you confirm your asynchronous I/O settings in the operating system and the database.
- ✓ For your SQL code, INSERTS can benefit from adding HINT "APPEND" or "APPEND\_VALUES." In addition, INSERTs and UPDATEs can benefit from PARALLEL hinting.

## Performance Tuning

Identify and address the below issues:

- ☑ Heavy data loads
- ☑ Long-running, inefficient workflows and queries
- ☑ Poorly coded mappings
- ☑ Incorrect design of source, lookups & target systems

In addition to the above items contributing to performance issues, these additional concerns need to be addressed.

## Network Testing

Testing the connection speeds between servers is critical in order to determine if your database is slow due to its configuration or simply the network input/output. Recommended tools: [ping](#), [netcat](#), and [lperf](#).

## Memory Footprints

The [jmap](#) tool is used to print memory maps or heap memory details of a given Java process. Passing a Java VM process ID to jmap will give you the memory footprint for this point in time. Using jmap provides information such as the heap size, Eden space, Young Generation, etc.

In order to see what's running in the JVM, the [jstack](#) tool will give you the stack trace of a given Java process or core file. Executing jstack with the -l flag (lowercase L) will reveal any locked threads that could be causing performance issues.

When it comes to Informatica health checks and performance tuning, the possibilities extend far beyond the database itself. Taking a comprehensive, bottom-up, end-to-end view will give you the best results.

**Need support?** Contact [Datavail's data integration Informatica experts](#). In addition to Informatica, we can support SSIS, Talend, Data Stage, and other leading data integration platforms. You can also download our essential guide "[Informatica Health Checks & Performance Tuning](#)" for a more in depth discussion on this topic.

## Informatica Grid Architecture

In a power center domain with multiple nodes, we can configure the workflows and sessions to run on a grid. The grid option delivers load balancing, dynamic partitioning, parallel processing and high availability to ensure optimal scalability, performance and reliability. When we run a workflow on a grid, the Integration Service runs a service process on each available node of the grid to increase performance and scalability. When you run a session on a grid, the Integration Service distributes session threads to multiple DTM processes on nodes in the grid to increase performance and scalability.

## Indexing Strategies

Sometimes certain indexes in your database are skipped for reasons that may seem inexplicable. If you dive further down, however, you can find that this often has to do with the index's [clustering factor](#), which is a measure of how ordered an index is in relation to the table on which it's based. Large clustering factors are an indication that the index is not well-ordered, and the database may therefore be performing expensive [skip scans](#).

