Improving the Performance of the PowerCenter Repository
Abstract

The PowerCenter repository is a relational database instance managed by the PowerCenter Repository Service. The database stores tables that hold metadata about repository objects, such as mappings and transformations. Over time, PowerCenter repository performance can slow for many reasons. This article describes methods to improve PowerCenter repository performance.

Supported Versions

- PowerCenter 9.x

Table of Contents

- Overview.................................................................. 2
- Truncate Workflow and Session Logs. .......................... 3
- Rules and Guidelines for Truncating Logs. ....................... 3
- Truncating Logs with the PowerCenter Repository Manager. 4
- Truncating Logs with the pmrep TruncateLog Command. . 4
- Manage Logs. .......................................................... 5
  - Configuring Log Management Properties. ..................... 5
- Update Repository Statistics. ........................................ 6
  - Updating Repository Statistics Using pmrep UpdateStatistics. 6
- Skip Information During Repository Operations. ............. 6
- Index Repository Tables. ............................................ 6

Overview

The PowerCenter repository is a relational database instance managed by the PowerCenter Repository Service. The database stores tables that hold metadata about repository objects, such as mappings and transformations. Over time, PowerCenter repository performance can slow for many reasons. This article describes methods to improve PowerCenter repository performance.

You can perform the following tasks to improve the performance of the PowerCenter repository:

**Truncate workflow and session logs.**

Use the pmrep TruncateLog command or the Repository Manager to remove workflow and session logs from the repository. The TruncateLog command increases performance by reducing the time that it takes the PowerCenter repository to access and write to the logs.

**Manage logs.**

Configure log management properties to restrict the size and the retention time of PowerCenter workflow and session logs.

**Update repository statistics.**

Update repository statistics to allow the repository database to use its internal query optimization techniques more effectively.
Skip information during repository operations.

When you back up, restore, or copy a repository, you can choose to skip types of information, and reduce the time it takes to perform the operation.

**Truncate Workflow and Session Logs**

You can remove workflow and session logs from the repository to reduce the time that it takes the PowerCenter repository to access and write to the logs.

When you configure a session or workflow to archive session logs or workflow logs, the PowerCenter Integration Service saves those logs in local directories. The repository also creates an entry each time that a workflow or a session runs. If you move or delete a session log or workflow log from the workflow log directory or session log directory, you can remove the entries from the repository.

Use the Repository Manager or the `pmrep TruncateLog` command to truncate the workflow logs for workflows that have completed.

You can truncate all logs for a workflow, or logs that were created before a specified date. You cannot truncate a workflow log for a workflow that is still running. The PowerCenter Repository Service truncates the workflow log list and the session log list at the same time.

When the PowerCenter Repository Service truncates logs for sessions and workflows, it also deletes the following runtime information for the sessions and workflows:

- Workflow details
- Session statistics
- Task details
- Source and target statistics
- Partition details
- Performance details

**Note:** When you truncate log entries from a Microsoft SQL Server repository, verify that no workflow is running. If you truncate log entries when a workflow is running, the workflow fails.

**Rules and Guidelines for Truncating Logs**

Use the following rules and guidelines when you truncate PowerCenter repository logs:

- **Schedule regular retention of logs.**
  
  Truncate logs on a regular schedule to minimize log sizes. Informatica recommends that you truncate logs weekly or every two weeks.

- **Determine length of log entry retention.**
  
  When you truncate logs, you can choose to remove log entries older than a certain date or older than a certain number of days. PowerCenter repository administrators commonly retain 10 days of log entries.

- **Regularly back up the PowerCenter Repository.**
  
  Follow a standard practice of regular PowerCenter repository backup.
  
  Optionally, you can include logs in the backup before you truncate them. Backing up the repository immediately before or after you truncate repository logs is not required.

- **Update statistics before and after truncating logs.**
  
  To ensure accurate repository statistics, run the `pmrep UpdateStatistics` command immediately before and after truncating PowerCenter repository logs.
Only use the TruncateLog command or the Repository Manager to truncate logs.

The pmrep truncateLog command and the Repository Manager truncate logs according to internal rules and algorithms. Do not manually remove or truncate log tables without guidance from Informatica Global Support.

**Truncating Logs with the PowerCenter Repository Manager**

You can truncate workflow and session logs in the PowerCenter Repository Manager:

1. In the Repository Manager, select the workflow in the Navigator window or in the Main window.
2. Choose **Edit > Truncate Log**.
   
The **Truncate Workflow Log** dialog box appears.
3. Choose to delete all workflow and session log entries, or to delete all workflow and session log entries with an end time before a particular date.
4. If you want to delete all entries older than a certain date, enter the date and time.
5. Click **OK**.
   
The PowerCenter Repository Service deletes the workflow and session log entries from the repository.

**Truncating Logs with the pmrep TruncateLog Command**

You can use the pmrep TruncateLog command to truncate workflow and session logs in the PowerCenter repository.

The command returns "truncatelog completed successfully" or returns "Failed to execute truncatelog." The truncate operation might fail for the following reasons:

- The folder name is not valid.
- The workflow does not exist in the given folder.
- You specified a workflow, but no folder name.

The TruncateLog command uses the following syntax:

```
truncatelog
-t <logs_truncated (all or up to end time in MM/DD/YYYY HH24:MI:SS format or as number of days before current date)>
[-f <folder_name>]
[-w <workflow_name>]
```
The following table describes `pmrep TruncateLog` options and arguments:

<table>
<thead>
<tr>
<th>Option</th>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-t</td>
<td>logs_truncated</td>
<td>Required. Use “all” to delete all logs, or enter an end time. <code>pmrep</code> deletes all logs older than the end time. You can enter the end time with the format MM/DD/YYYY HH24:MI:SS, or you can specify the number of days before the current date. If you specify the number of days, the end time must be an integer greater than 0.</td>
</tr>
<tr>
<td>-f</td>
<td>folder_name</td>
<td>Optional. Deletes logs associated with the folder. If you do not give both the folder name and the workflow name, then <code>pmrep</code> deletes all logs from the repository.</td>
</tr>
<tr>
<td>-w</td>
<td>workflow_name</td>
<td>Optional. Deletes logs associated with the workflow. The Repository Service deletes all logs from the repository if you do not give both the folder name and the workflow name. If you give both the folder name and workflow name, the Repository Service deletes logs associated with the workflow. If you enter the workflow name, you must also provide the folder name.</td>
</tr>
</tbody>
</table>

**Manage Logs**

You can configure log management properties to restrict the size and the retention time of PowerCenter workflow and session logs.

You can perform the following tasks:

**Limit the time that the repository retains logs.**

When you configure log management properties to preserve log events for a limited number of days, you reduce the total size of logs.

Many administrators find that 10 days is sufficient to keep logs. Base your setting on the load in your environment.

**Limit the size of repository logs.**

Configure the maximum disk size for the directory that contains the log event files. The standard guideline is 1 GB. Base your setting on the load in your environment.

**Configuring Log Management Properties**

Configure the log management properties in the **Log Management** dialog box in Informatica Administrator.

1. In the Administrator console, click the **Logs** tab.
2. Select **Log Actions > Log Management**.
3. Enter the number of days for the Log Manager to preserve log events.
4. Enter the maximum disk size for the directory that contains the log event files.
5. Enter the time zone in the following format:
   ```
   GMT(+|-)<hours>:<minutes>
   ```
   For example: GMT+08:00
6. Click **OK**.
Update Repository Statistics

You can update repository statistics with the pmrep UpdateStatistics command to improve PowerCenter repository performance.

Almost all PowerCenter repository tables use at least one index to speed up queries. Most databases keep and use column distribution statistics to determine which index to use to execute SQL queries optimally. Database servers do not update these statistics continuously.

In frequently used repositories, these statistics can quickly become outdated, and SQL query optimizers might not choose the best query plan. In large repositories, choosing a sub-optimal query plan can have a negative impact on performance. Over time, repository operations gradually become slower.

When you use the pmrep UpdateStatistics command at regular intervals, you prevent this effect.

If you have a repository with a high number of objects, perform this task weekly. You can perform this task during periodic maintenance downtime.

You should run the pmrep UpdateStatistics command immediately before and after you truncate workflow and session logs.

Updating Repository Statistics Using pmrep UpdateStatistics

You can update statistics for PowerCenter repository tables and indexes with the pmrep UpdateStatistics command. The command returns "updatestatistics completed successfully" or returns "updatestatistics failed."

The UpdateStatistics command uses the following syntax:

```
updatestatistics
```

Skip Information During Repository Operations

When you back up, restore, or copy a repository, you can choose to skip types of information, and reduce the time it takes to perform the operation.

The pmrep Backup and pmrep Restore commands include options to skip the following types of information:

- Workflow and session logs
- Deployment group history
- Metadata Exchange (MX) data

Index Repository Tables

Creating indexes on large, stable repository tables is a standard database administration technique to increase performance. A database administrator performs this task on the source and target databases that you use with PowerCenter.

When you install PowerCenter, the installer creates a repository database and creates indexes on certain database tables. Ordinarily, you do not have to take any steps to improve table indexes. Sometimes, Informatica Global Customer Support makes a specific recommendation based on an understanding of your environment, as part of a support case.

**Warning:** PowerCenter indexes repository tables according to internal rules and algorithms. Although you can view the repository tables, Informatica strongly advises against altering the tables or data within the tables. Informatica is not responsible for corrupted data that is caused by customer alteration of the repository tables or data within those tables. Therefore, do not make changes to PowerCenter Repository table indexes without guidance from Informatica Global Customer Support.
Author

Mark Pritchard
Principal Technical Writer