Pentaho Operations Mart

HITACHI
Inspire the Next

Change log (if you want to use it):

Date	Version	Author	Changes

Contents

This page intentionally left blank.

Overview

This document covers some best practices on the Pentaho Business Analytics (BA) Operations Mart and the Data Integration (DI) Operations Mart. The Operations Mart (Ops Mart) is used to monitor system performance for information such as how long and how often a report runs, and aggregates data from the Pentaho server log files into a set of prebuilt audit reports.

You will also learn how to modify the Ops Mart for your needs, if the prebuilt reports do not give you the exact information you are looking for.

Our intended audience is network or systems administrators, or anyone with an interest in performance tuning, logging, and monitoring.

Software	Version(s)
Pentaho	7.x, 8.x

The Components Reference in Pentaho Documentation has a complete list of supported software and hardware.

> *In Pentaho 7.x, Ops Mart works for BA and DI, but only on Postgres databases. The Ops Mart in Pentaho 8.x works for all repository database types.*

Before You Begin

This document assumes that you have knowledge about computer networking concepts, and that you have already installed and configured Pentaho. In the Pentaho documentation, both Business Analytics Operations Mart and Data Integration Operations Mart have information about installing and configuring the Operations Marts.

More information about related topics outside of this document can be found in Monitoring System Performance in Pentaho's documentation.

Terms You Should Know

- **Network Logging**: The process of collecting data and then storing it over varying periods of time so that network trends can be analyzed.
- **Network Monitoring**: A network's effort to detect slow or failing components. These could include things such as failing routers, switches, or servers.
- **Network Auditing**: A process in which a network is analyzed for performance, security, availability, and management.
- **Data Mart**: A collection of tables, schemas, transformations, and reports, that explores log data for auditing, reporting, and analysis.

Pentaho Operations Mart

Before you dig into the use cases in this document, here are some basics about both versions of the Pentaho Operations Mart and its applicability to real-world situations.

You can find details on these topics in the following sections:

What is the Pentaho BA Operations Mart?

The BA Operations Mart is organized as a data warehouse using a star schema. The BA Ops Mart is used to monitor system performance for information such as how long and how often a report runs, then aggregates data from the Pentaho server log files into a set of prebuilt audit reports.
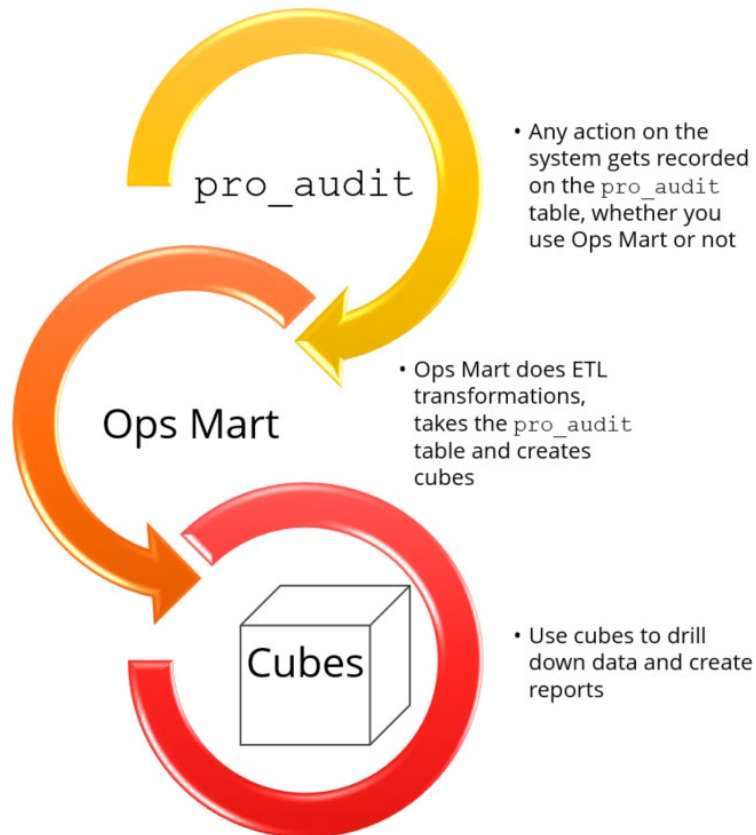

Figure 1: BA Ops Mart Data Flow

- **BA Ops Mart Data Collection**: For the BA Ops Mart, data is collected in the `Hibernate` database, in the `pro_audit` table.
- **BA Ops Mart Data Transformation**: Data is transformed using the Update Audit Data `xaction` that is run from the Pentaho User Console (PUC).
- **BA Ops Mart Reports**: To find these, look in **BA Audit Reports** in Pentaho User Console (PUC):

Figure 2: BA Audit Reports Link

The following diagram shows you the process of using the BA Ops Mart. The actions listed proceed in order from left to right:
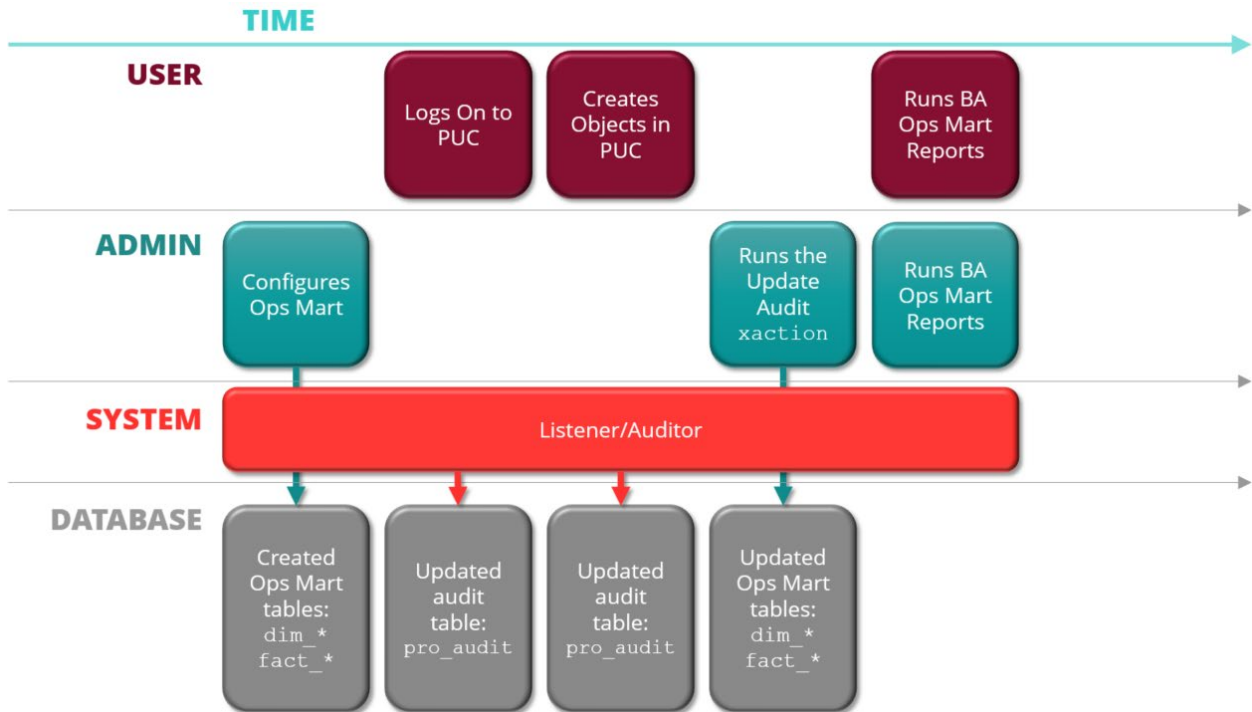

Figure 3: BA Ops Mart Process Overview

How is the Data Integration (DI) Ops Mart Different?

The DI Ops Mart uses data integration logging data, instead of the audit data used by the BA Ops Mart. It is a centralized data mart, a collection of tables, schemas, transformations, and reports that stores job or transformation log data for auditing, reporting, and analysis.

The DI Ops Mart enables you to collect and query DI log data and then use the Pentaho server tools to examine the log data in reports, charts, and dashboards. The data mart is a collection of tables organized as a data warehouse using a star schema. Together, the dimension tables and a fact table represent the logging data.
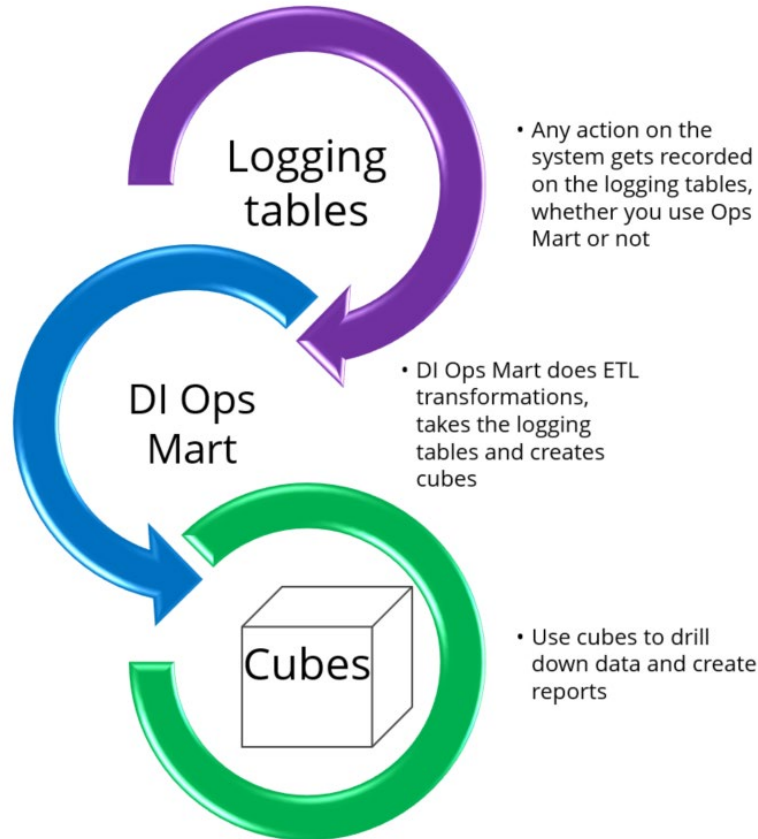

Figure 4: DI Ops Mart Data Flow

- **DI Ops Mart Data Collection**: Data is collected in the `pentaho_dilogs` database, in tables including: `trans_log`, `job_log`, `jobentry_log`, `step_log`, `transperf_log`, and so forth.
- **DI Ops Mart Data Transformation**: Data is transformed with the `Update_Dimensions_then_logging_Data mart` job, run from PDI or the PUC.

- **DI Ops Mart Reports**: To find these, look in **DI Audit Reports** in the PUC:

Figure 5: DI Audit Reports Link

The following diagram shows you process of using the DI Ops Mart. The actions listed proceed in order from left to right:
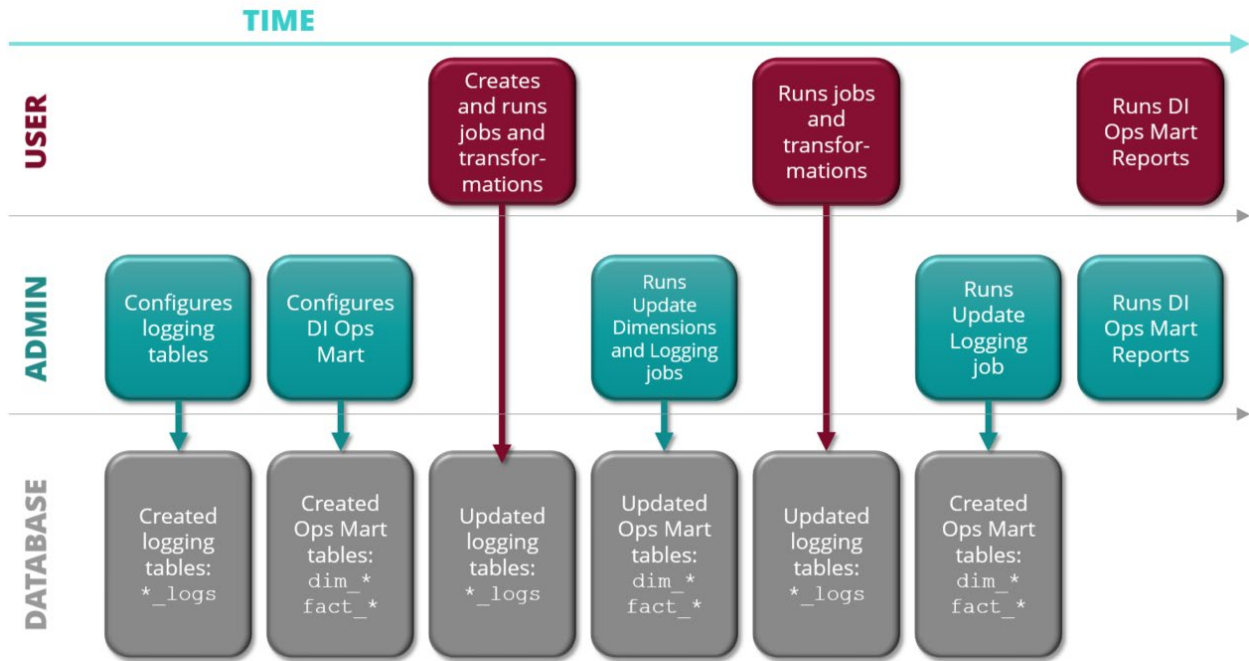

Figure 6: DI Ops Mart Process Overview

What Can I Do with the Ops Mart?

Because both the BA and DI Ops Marts follow the same basic flow of collecting, transforming, and reporting data, you can use Ops Mart for purposes like auditing, resource planning, and performance root cause analysis:

Auditing

Auditing with the Ops Mart can tell you, among other things:

- Who is logging on?
- How many times are they logging on, and for how long?
- What reports is each user running?
- What reports have errored?
- What are the most used reports?

Resource Planning

Resource planning with the Ops Mart will help you find out:

- How many jobs and reports are scheduled at specific times or on specific days?
- What is the best time to schedule new reports or extract/transform/load (ETL) operations?
- How do the numbers of users or volume of usage through time compare?
- How will the system scale?

Performance Root Cause Analysis

Doing performance root cause analysis with the Ops Mart will help you identify what caused performance issues in the environment:

- What was running at the time of the performance issue?
- How long was this job or report running at the time of the problem?

For optimal performance, be sure to clean the BA Ops Mart and DI Ops Mart periodically.

Use Case 1: Why Would I Use Either Pentaho Operations Mart?

Fabiola must audit systems resource usage for Pentaho reports. This will help give her a bigger picture of how many resources are being used and the effect on the systems. She will begin by using the prebuilt Pentaho Operations Mart reports to conduct audits, examine the data in the reports, and then make some decisions about how best to use the findings. In this case, she will need to use the BA Ops Mart to find out session duration counts and the DI Ops Mart to find out what jobs and transformations might be her biggest bottlenecks.

BA Operations Mart

In this case, Fabiola will use a **Session Duration Counts** report to find out how long users are logged in, and how many times they login during specified periods of time.

The BA Ops Mart contains this report and many others. This table lists the reports, what information each shows, and gives an example use for each:

Table 1: Prebuilt BA Operations Mart Reports

Report Name	Information Shown	Example Use
Content Duration	Amount of time it takes a report to run	Learn how long a report is taking to run
Content Failures	Filenames of all content failures within a defined length of time	Find what content has been failing recently
Content Request Day of Month	Compare login metrics for days within a specified month	See how login metrics differ during different time periods
Day of Week Request and Login Metrics	Compare login metrics for days in a week within a specified year	
Hours in Day Request and Login Metrics	Compare login metrics by hour for a specified day	
Session Duration Counts	Length of time for logins and the number of logins per user for specified length of time	
Content Type Usage	List of content sorted by type, used within a defined length of time	See what content, or type of content, is being used over different time periods
Content Usage	List of content usage within a defined length of time	

Using a BA Ops Mart Prebuilt Report

To set up a prebuilt report for use, Fabiola must:

1. Identify the report she wants to open from the Prebuilt BA Operations Mart Reports table.
2. Launch **PUC** and double-click on **Browse Files**.
3. Click on **Public** > **Pentaho Operations Mart** > **BA Audit Reports**.
4. Select the appropriate file and click on **Open**. The monitoring and logging data appears in the prebuilt report.
5. If Fabiola needs to edit this report, she can do that from within the opening tool.

See Pentaho Analyzer, Pentaho Interactive Reports, and Pentaho Dashboard Designer for more information about editing a report in Analyzer or Interactive Reporting. Pentaho Report Designer contains more information about editing a report using the designer.

DI Operations Mart

In Fabiola's case, she has also chosen to use a **PDI Longest Running** report from the DI Ops Mart to find out where she can make improvements for resource utilization.

The DI Ops Mart contains this report and many others. This table lists the reports, what information each report shows, and gives an example use for each:

Table 2: Prebuilt DI Operations Mart Reports

Report Name	Information Shown	Example Use
Currently Running Jobs and Transformations	See what is actively being executed in the server so you can debug performance issues	If you notice a problem going on right now and want to track it down
Failure Rate	See failure rate of jobs and transformations	Researching failed jobs and transformations
Top 10 Failed	See which jobs/transformations fail most frequently or recently	
Most Bottlenecked Steps	Find details on where you can make improvements in steps and transformations to keep performance high	Improving steps and transformations for resource utilization
Most Bottlenecked Transformations		
Most Bottlenecked Transformations with Steps		
PDI Longest Running		
Top 10 Bottlenecked Steps		
Top 10 Bottlenecked Transformations		

Using a DI Ops Mart Prebuilt Report

To set up a prebuilt report for use, Fabiola must:

1. Identify the report she wants to open from the Prebuilt DI Operations Mart Reports table.
2. Launch **PUC** and double-click on **Browse Files**.
3. Click on **Public** > **Pentaho Operations Mart** > **DI Audit Reports**.
4. Select the appropriate file and click on **Open**. The monitoring and logging data appears in the prebuilt report.
5. If Fabiola needs to edit this report, she can do that from within the opening tool.

See Pentaho Analyzer, Pentaho Interactive Reports, and Pentaho Dashboard Designer for more information about editing a report in Analyzer or Interactive Reporting. Pentaho Report Designer contains more information about editing a report using the designer.

Use Case 2: Why Would I Extend the Prebuilt Reports?

Janice has been using the prebuilt Ops Mart reports to do her analysis. She's noticed that the prebuilt reports, while useful to get started, aren't as detailed or refined as she would need to capture all the information within her networking environment. She wants to add more details on which user is executing each content on a monthly basis.

*Janice decides to extend one of the prebuilt reports, **Content Duration (Last N)**, to meet these unique needs.*

It is possible to change and add to the prebuilt reports to make them more useful for your needs.

In this case, Janice must add details to **Content Duration (Last N)** to sort users and aggregate them to see who is using the server, and how much they do so. She will follow this process:
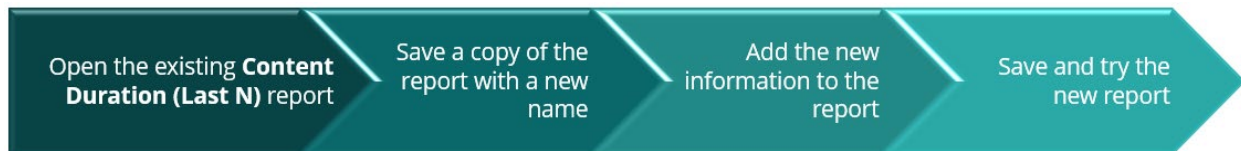

Figure 7: Extending Prebuilt Reports Example

Remember, when you are extending prebuilt reports:

1. Use existing reports as guides for how you want to set up your report.
2. View the existing ETL by looking at the zip files under the default content. It will be located in `pentaho-operations-mart-etl-8.0.0.0-28.zip`, or a filename similar to that based on your Pentaho installation version.
3. Create new ETL to add new fields or external data to your report.

> ⚠️ *Avoid losing your work. Always make* copies *of the reports you plan to extend. Do not save over existing jobs, schemas, or reports.*

See Pentaho Analyzer, Pentaho Interactive Reports, and Pentaho Dashboard Designer for more information about editing a report in Analyzer or Interactive Reporting. Pentaho Report Designer contains more information about editing a report using the designer.

Use Case 3: Why Should I Clean My Ops Mart?

Janice has been using the Ops Mart with great success for several months, but she has noticed that operations are getting much slower than they used to be. The tables created from using the Ops Mart have grown out of control and Janice needs to trim them. She has decided that the best way to do this is to use the Ops Mart cleaning process. Here is a use case about a time when cleaning the Ops Mart would be useful.

Both the DI and BA Ops Marts need periodic cleaning to prevent tables from growing out of control.

> 💡 *The Ops Mart will* not *clean itself unless you set it up. When you do set it up, it will then* default *to cleaning itself every 365 days. Change the setting if you need to clean the Ops Mart more frequently.*

BA Ops Mart Cleaning

To change that schedule so that the operational data is deleted using a different time frame, you will need to add a variable, `ba.cleanup.max.age`, in the `kettle.properties` file located in the `{user.home}/.kettle` directory.

1. Stop the Pentaho server.
2. Open the `kettle.properties` file in the administrative user's `{user.home}/.kettle` directory.
3. Add the `ba.cleanup.max.age` variable to indicate the age above which entries are deleted. For example, if you want to delete entries that are older than 30 days, add:

```
ba.cleanup.max.age=30
```

4. Save and close the `kettle.properties` file.
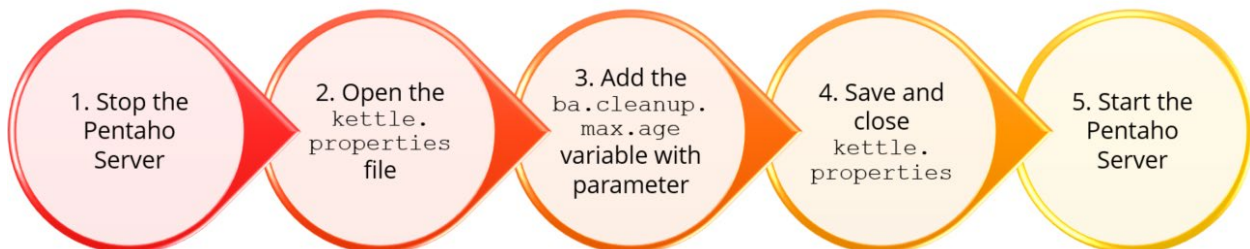5. Restart your Pentaho server for the setting to take effect.

Figure 8: Cleaning the BA Ops Mart

Data Integration (DI) Ops Mart Cleaning

Cleaning the DI Operations Mart consists of running either a job or transformation that deletes data older than a specified maximum age.

Figure 9: Cleaning the DI Ops Mart

Perform the following steps to clean the PDI Operations Mart:

1. Using the **PDI Client**, open the **etl** folder, and choose either
   `Clean_up_PDI_Operations_Mart.kjb` for jobs or the
   `Clean_up_PDI_Operations_Mart_fact_table.ktr` for transformations.
2. Set the following parameters:
   a. `max.age` (required). This is the maximum age, in days, of the data you want to keep.
   b. `schema.prefix` (optional). For PostgreSQL databases, enter the schema name followed by a period (`.`); this will be applied to the SQL statements. For other databases, leave the value blank.
3. Run the job or transformation. This will delete job and transformation data older than the maximum age (`max.age`) from the data mart.

To schedule regular cleanup of the PDI Operations Mart, see Schedule Perspective in the PDI Client.

Related Information

Here are some links to information that you may find helpful while using this best practices document:

- [Business Analytics Operations Mart](#)
- [Cleaning the BA Operations Mart](#)
- [Cleaning the DI Operations Mart (Clean Up Operations Mart Tables)](#)
- [Components Reference](#)
- [Data Integration (DI) Operations Mart](#)
- [Monitoring System Performance](#)
- [Pentaho Analyzer](#)
- [Pentaho Dashboard Designer](#)
- [Pentaho Installation](#)
- [Pentaho Interactive Reports](#)
- [Pentaho Report Designer](#)
- [Schedule Perspective in the PDI Client](#)