



## Pentaho Aggregation Designer 1.1 User Guide



This document is copyright © 2009 Pentaho Corporation. No part may be reprinted without written permission from Pentaho Corporation. All trademarks are the property of their respective owners.

## About This Document

If you have questions that are not covered in this guide, or if you find errors in the instructions or language, please contact the Pentaho Technical Publications team at [documentation@pentaho.com](mailto:documentation@pentaho.com). The Publications team cannot help you resolve technical issues with products.

Support-related questions should be submitted through the Pentaho Customer Support Portal at <http://support.pentaho.com>.

For information about how to purchase support or enable an additional named support contact, please contact your sales representative, or send an email to [sales@pentaho.com](mailto:sales@pentaho.com).

For information about instructor-led training on the topics covered in this guide, visit <http://www.pentaho.com/training>.

## Limits of Liability and Disclaimer of Warranty

The author(s) of this document have used their best efforts in preparing the content and the programs contained in it. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, express or implied, with regard to these programs or the documentation contained in this book.

The author(s) and Pentaho shall not be liable in the event of incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of the programs, associated instructions, and/or claims.

## Trademarks

Pentaho (TM) and the Pentaho logo are registered trademarks of Pentaho Corporation. All other trademarks are the property of their respective owners. Trademarked names may appear throughout this document. Rather than list the names and entities that own the trademarks or insert a trademark symbol with each mention of the trademarked name, Pentaho states that it is using the names for editorial purposes only and to the benefit of the trademark owner, with no intention of infringing upon that trademark.

## Company Information

Pentaho Corporation  
Citadel International, Suite 340  
5950 Hazeltine National Drive  
Orlando, FL 32822  
Phone: +1 407 812-OPEN (6736)  
Fax: +1 407 517-4575  
<http://www.pentaho.com>

E-mail: [communityconnection@pentaho.com](mailto:communityconnection@pentaho.com)  
Sales Inquiries: [sales@pentaho.com](mailto:sales@pentaho.com)  
Documentation Suggestions: [documentation@pentaho.com](mailto:documentation@pentaho.com)

Sign-up for our newsletter:  
<http://community.pentaho.com/newsletter/>

# Contents

Introduction .....	2
Pentaho BI Suite Enterprise Edition Documentation .....	3
Installing the Pentaho Aggregation Designer .....	4
Defining the Data Source .....	5
Selecting a Model .....	6
Getting Recommendations Using Aggregate Advisor .....	7
Customizing Aggregates .....	8
Customizing an aggregate .....	8
Adding Aggregates .....	8
Deleting Aggregates .....	9
Exporting Aggregates .....	10
Glossary of Terms .....	11
Aggregate .....	11
Aggregate Tables .....	11
Aggregation .....	11
Data Definition Language (DDL) .....	11
Mondrian Schema .....	11
Relational Online Analytic Processing (ROLAP) .....	11
Snowflake Schema .....	11
Star Schema .....	12

# Introduction

This user's guide provides you with instructions for recommending and/or designing aggregate tables. The use of aggregate tables can dramatically improve the query performance of ROLAP solutions.

## **Assumptions**

This document is written for Database Administrators and consultants who design specific aggregate tables or get recommendations for aggregate tables based on an intelligent adviser algorithm. It is assumed that you, the reader, have a strong understanding of database design and concepts (such as database modeling, SQL security, and performance), and are familiar with aggregate table concepts.

# Pentaho BI Suite Enterprise Edition Documentation

As an Enterprise Edition customer, you have access to an array of professional documentation in the Pentaho Knowledge Base , including a collection of tech tips and articles, and the following comprehensive guides for users and administrators:

Document Name	Purpose
Customizing the Pentaho User Console	A short but informative developer's guide on how to customize the theme and graphics in the Pentaho User Console
Pentaho BI Suite 3.5 Standard Installation Guide	An installer's step-by-step guide to deploying the BI Suite with the preconfigured Tomcat application server and a MySQL solution database
Pentaho BI Platform 3.5 Manual Deployment Guide	An installer's step-by-step guide to building and deploying the BI Platform and Pentaho Enterprise Console for all supported solution repository databases and application servers
Pentaho BI Suite 3.5 Mac Installation Guide	An installer's step-by-step guide to installing the BI Suite on a Mac
Pentaho BI Suite 3.5 Archive-Based Installation Guide	An installer's step-by-step guide to installing the BI Suite along with a preconfigured Tomcat application server on Windows, Solaris, or Linux by using small, pre-configured ZIP or TAR archives
Pentaho User Console 3.5 Guide	A comprehensive guide to the Pentaho User Console, including the Dashboard Designer plugin
Pentaho BI Suite 3.5 Administrator's Guide	A sysadmin and database administrator's guide to advanced BI Platform configuration, including security integration, product maintenance, and performance tuning
Pentaho Report Designer 3.5 User Guide	A task-based user guide for Report Designer
Upgrading to the Pentaho BI Suite Enterprise Edition 3.5	An administrator's step-by-step guide to upgrading a production system from the BI Suite version 2.x to 3.5

# Installing the Pentaho Aggregation Designer

This section provides you with information and instructions for installing and configuring the Pentaho Aggregation Designer.

## System Requirements

The Pentaho Aggregation Designer requires Java SE runtime version 1.5 or later. Java versions earlier than Java 1.5 are not supported. Pentaho recommends that you have at least 1 Gigabyte of free physical memory. The Pentaho Aggregation Designer 1.0 supports Pentaho Open BI Suite versions 1.7 and higher.

Category	Minimum	Recommended
Disk	CPU	RAM
1 GB	1 GHz (single core)	1GB
10 GB	1.5 GHz + (multi-core)	2 GB

Refer to the table below for operating system support:

Windows	Linux
Windows 2003	Windows XP
Red Hat Enterprise Linux 5	SUSE Linux 10.1

**Note:** Other versions of the above-listed Linux distributions, and other distributions entirely, will probably work with the Pentaho Aggregation Designer 1.0; however, only the distributions listed here are tested by Pentaho.

## Installation Instructions

The **pad-open-1.0-xx.zip** file contains all the libraries and script files necessary to run Pentaho Aggregation Designer. To install the Pentaho Aggregation Designer, unzip this file into a directory of your choice.

### To launch the Aggregation Designer on Windows...

Run the **startaggregationdesigner.bat** script located in the root of your installation directory.

### To launch the Aggregation Designer on Linux...

Run the **startaggregationdesigner.sh** script located in the root of your installation directory.

**CAUTION:** Place your JDBC driver JARs in the Drivers directory. Once in this directory, the drivers are added to the classpath automatically when the Pentaho Aggregation Designer starts.

## Defining the Data Source

To design an aggregate, you must first establish a connection with your target relational database, then select the OLAP model to optimize. You can connect to any relational database that is supported by Mondrian. In some instances, you may need to define additional parameter-related values for your JDBC driver.

To define a data source connection...

1. In the Pentaho Aggregation Designer tool bar, click **Connection** to open the **Connect to Data Source** dialog box.
2. Click **Configure**. The **Connection Properties** dialog box appears.
3. In the **Connection Name** field, enter a name for your connection; this is a free-text field. A connection name uniquely defines a connection.
4. In the **Connection Type** list, select a database.
5. In the **Access** list, keep the default choice, which should be **Native (JDBC)**.
6. In the **Settings** section, type the host name of the database server into the **Host Name** field. In the **Database Name** field, type the name of the database you're connecting to. In the **Port Number** field, enter the TCP port number. Optionally, in the **User Name** and **Password** fields, type the user name and password used to connect to the database.
7. Click **Test**.

If the settings you typed in are correct, a success message appears.

8. Click **OK**.

## Selecting a Model

After defining your data source, you must select the cube you want to use for defining and building aggregate tables.

To select a model...

1. In the **Connect to Data Source** dialog box, under **OLAP Model**, select **Mondrian Schema File**.
2. Click the ellipsis (...) to display a file dialog box.
3. Browse to locate and select your Mondrian schema file, then click **OK**.
4. Click **Apply**. The Cube list is populated with a list of cubes defined in your schema.
5. Select the Mondrian cube you want to optimize, then click **Connect**.

When the Pentaho Aggregation Designer establishes a connection, it runs several validation tests to ensure that your database structure is ready to support aggregate tables. A validation summary dialog box appears with a list of test results. If you see an error message, contact your database administrator.



## Getting Recommendations Using Aggregate Advisor

If you are unfamiliar with aggregate table design and need help creating aggregates to optimize a cube, you can rely on the Aggregate Advisor to provide you with a list of recommendations. The Pentaho Aggregation Designer uses your schema file and the data in your database to create aggregate definitions.

To display recommended aggregates...

1. In the Pentaho Aggregation Designer toolbar, click **Advisor**.
2. Specify your **Advisor Input Parameters**.

Setting	Description
Max Tables to Create	Allows you to specify the maximum number of aggregates you want the Advisor to recommend.
Max Time to Run	Allows you to specify the maximum amount of time (in seconds) you want the Advisor to run before making recommendations. <i>Note : Allowing the Advisor to run for longer periods of time allows for more potential recommendations to be evaluated and results in more accurate recommendations.</i>

3. Click **Recommend**.

The Advisor runs for a few seconds before it displays an initial list of recommended aggregates. The Advisor is designed to keep running until it finds an optimal solution. If you stop the Advisor prematurely, the Advisor returns the best set of recommendations it has found up to the point when it was stopped.

# Customizing Aggregates

When you select an aggregate, the Pentaho Aggregation Designer pulls information from the schema file to display its dimensions and levels. You can modify any aggregate the Pentaho Aggregation Designer recommends, customizing it for your needs. You can also create an aggregate from scratch and delete aggregates you do not want.

## Impact summary

The impact summary in the lower right pane provides you with information on the estimated impact for creating all of the currently selected aggregates. This summary includes the number of aggregate tables that will be created, the estimated number of rows contained in those tables, and the estimated amount of space it will occupy on the hard drive. The impact summary is automatically updated as you select and deselect aggregates from the list of proposed aggregates.

## Cost/benefit chart

The Cost/Benefit chart provides a high-level comparison of the benefit of all currently selected aggregates relative to their estimated cost. The benefit scale represents the relative number of queries that can be fulfilled by an aggregate table versus having to be retrieved from the base fact table. The cost scale is an indicator of the impact in terms of number of tables and disk space needed to create the selected aggregate recommendations.

## Saving your design

The Pentaho Aggregation Designer allows you to save all aggregate-related data (custom- or advisor-created) in your workspace at any time. Saving ensures that all of the data (your designs) in the workspace is retained; you are saving the state of your workspace as an XML file in a location you specify. To save, go to the **File** menu and click **Save As**. To open a saved file, go to the **File** menu and click **Open**, then navigate to the design you previously saved.

# Customizing an aggregate

To customize an aggregate...

1. In the **Pentaho Aggregation Designer**, click on an aggregate in the proposed aggregate list to select it.

**Note:** When you modify an aggregate created using the Advisor, the aggregate becomes a Custom aggregate as indicated by the Type column in the proposed aggregate list.

2. In the left pane, you can (optionally) modify the **Name** and **Description** for your custom aggregate.
3. In the **Aggregation Levels** tab, click the down arrows to make changes to the hierarchy and levels associated with the aggregate definition you are customizing.
4. Click **Apply**.

The Pentaho Aggregation Designer updates the proposed aggregate list, cost/benefit chart, and impact summary.

# Adding Aggregates

To add an aggregate...

1. In the right pane of the Pentaho Aggregation Designer, click **Add**.
2. In the left pane, type a **Name** and **Description** for your new aggregate.
3. Under **Level**, click the down arrows to define the hierarchy and levels associated with the aggregate you are creating.
4. Click **Apply**.

Your aggregate is added to the aggregate list.

## Deleting Aggregates

Select an aggregate from the proposed aggregate list and click Remove.

## Exporting Aggregates

The Pentaho Aggregation Designer allows you to preview the DDL, DML, and schema (for relational databases) outputs before you build aggregate tables. You can also save the outputs and edit them later. If you are using OLAP, DML is the only available output.

To preview the DDL and DML outputs...

1. Select the aggregates that have DML/DDDL output you want to preview.
2. In the Pentaho Aggregation Designer toolbar, click **Export**.
3. In the Execute and Publish dialog box, click **Preview**.
4. Click **Copy to Clipboard** or **Save** to retain the output.
5. If you examine the DDL/DML outputs and are satisfied with the results, you can allow the Pentaho Aggregation Designer to build (Execute/Publish) the aggregate tables. Follow the instructions for publishing and exporting included in the **Execute and Publish** dialog box.

# Glossary of Terms

Below is a list of terms used in this document.

## Aggregate

Definitions for aggregate tables that help optimize a cube; also, summarized data.

## Aggregate Tables

Coexists with the base fact table, and contains pre-aggregated measures built from the fact table. It is registered in Mondrian's schema, so that Mondrian can choose whether to use the aggregate table rather than the fact table, if applicable for a particular query.

## Aggregation

The process of merging multiple data values into one value. For example, sales data collected daily can then be aggregated to the week level, the week data could be aggregated to the month level, and so on. The data can then be referred to as aggregate data. Aggregation and summarization are synonyms, as are aggregate data and summary data.

## Data Definition Language (DDL)

Originally a subset of SQL, this language defines data structures, including rows, columns, tables, indexes, and database specifics such as file locations. DDL SQL statements are more a part of the database management system, and have large differences between SQL implementations.

## Mondrian Schema

Defines a multi-dimensional database. A Mondrian schema contains a logical model, consisting of cubes, hierarchies, and members, and a mapping of this model onto a physical model. The logical model consists of the constructs used to write queries in the MDX language: cubes, dimensions, hierarchies, levels, and members. The physical model is the source of the data presented through the logical model. It is typically a star schema, which is a set of tables in a relational databases.

## Relational Online Analytic Processing (ROLAP)

An alternative to MOLAP (Multidimensional OLAP) technology. While both ROLAP and MOLAP analytic tools are designed to allow analysis of data through the use of a multidimensional data model, ROLAP differs significantly in that it does not require the pre-computation and storage of information. Instead, ROLAP tools access the data in a relational database and generate SQL queries to calculate information at the appropriate level when an end user requests it. With ROLAP, it is possible to create additional database tables (summary tables or aggregations) which summarize the data at any desired combination of dimensions.

## Snowflake Schema

A way of arranging tables in a relational database such that the entity relationship diagram resembles a snowflake in shape. At the center of the schema are fact tables which are connected to multiple dimension tables. Thus a snowflake simplifies to a star schema when relatively few dimensions are used. The star and snowflake schemas are most commonly found in data warehouses where the speed of data retrieval is more important than the speed of insertion. As such, these schemas are not normalized much, and are frequently left in third normal form or second normal form.

## Star Schema

The simplest style of data warehouse schema that consists of a few fact tables (possibly only one) referencing any number of dimension tables (sometimes referenced as star join schemas) .