CHAPTER 1

Quick Start: Pentaho Examples

COPYRIGHTED MATERIAL

Pentaho is a powerful Business Intelligence Suite offering many features: reporting, OLAP pivot tables, dashboarding and more. In this book you will find a lot of detailed information about Pentaho's components, how they work and interact, the features they deliver, and how to use the Pentaho BI Suite to create solutions for real-world problems. However, it's a good idea to try and grasp the big picture before diving into the details.

This chapter helps you get started by showing you where to get the software and how to install and run it. The Pentaho BI Suite includes many examples demonstrating its features to give new users an idea of what kind of solutions you can build with it. Most of these examples work ''out of the box'' and are thus ideal for an introduction to the product. By reading this chapter, you'll get acquainted with Pentaho by looking at some examples.

Getting Started with Pentaho

In this section, we describe how to obtain the software, install it, and run it. To run the software, you need a regular desktop or laptop computer running any popular operating system, such as Ubuntu Linux, Mac OS X, or Microsoft Windows 7, XP, or Vista. To download the necessary software you will need an Internet connection with sufficient bandwidth to download tens to hundreds of megabytes.

Downloading and Installing the Software

The Pentaho BI Suite is open source software; you are free to use and distribute its programs, and if you like, you can study and even modify its source code. You may do all of this free of charge.

Pentaho is programmed in the Java programming language. Before you can run Java programs, you need to install Java. For Pentaho, you need at least Java version 1.5. You should also be able to use Java 1.6. We assume you already have a recent version of Java installed on your system. You can find more details on downloading and installing Java in Chapter 2.

You can download all of Pentaho's released software from the SourceForge website. The easiest way to find the software is to navigate to `http://sourceforge.net/projects/pentaho/` and click the Download link. You will see a list of products you can download.

For now, you won't need all of the software—all you're interested in at the moment is the Business Intelligence Server. Click the Download link in the far right column. This takes you to a page containing a list of different versions of the software. Here you should take care to find the latest version of the generally available (GA) release, packaged in a way that is appropriate for your platform. For example, Microsoft Windows users should download the `.zip` compressed package, and users of UNIX-based systems should download the `.tar.gz` compressed package.

NOTE **In Pentaho's download pages on SourceForge, you can usually find at least the latest generally available (GA) release as well as a so-called milestone release of the new, upcoming version. If you really want to be on the bleeding edge of development, you can download nightly builds of the software from `http://ci.pentaho.com/`. For this book, we mostly worked with the nightly builds of the Citrus release, which was still being developed at the time of writing, but which should be available as a milestone or GA release by the time of publishing.**

It is always a good idea to try out the milestone releases to keep track of future changes and additions. But beware that milestone releases are still in development; they are not intended for production use, and you may find bugs or experience usability issues. However, this is one of the best reasons why you should run milestone releases—by reporting any issues you experience, you can directly influence the improvement of the software for your own benefit (as well as that of all other users).

After downloading the `.zip` or `.tar.gz` compressed package, you must extract the actual software from the compressed package and copy it to some place you find convenient. Windows users can right-click the `.zip` file and choose Extract Here (in new folder) in the context menu. Alternatively, you can use a third-party program such as Peazip to extract the programs from

the compressed package. Users of UNIX-like systems can open a terminal and extract the package from the command line.

Extraction should result in a single folder containing all of the Pentaho BI Server software. Windows users can place this folder anywhere they like, but it makes most sense to put it in the Program Files directory. For UNIX-like systems, the proper location depends on the exact UNIX flavor, but for checking out the examples, it is best to move the Pentaho Server directory to your home directory. In the rest of this chapter, we refer to the directory containing the Pentaho Server software as the Pentaho home directory or simply Pentaho home.

Running the Software

Now that you have downloaded and installed the software, you can start using it.

Starting the Pentaho BI Server

In the Pentaho home directory, you will find a few scripts that can be used to start the server. Microsoft Windows users can double-click the script named `start-pentaho.bat`.

For UNIX-based systems, the script is called `start-pentaho.sh`. You may first need to allow this script to be executed. Modern Linux desktop environments such as GNOME and KDE will let you do this in the file's Properties dialog, which you can invoke from the file browser. For example, in Ubuntu Linux, you can right-click the file and choose Properties from the context menu to invoke the dialog. In the Permissions tab in the dialog, you can select a checkbox to allow the file to be executed, as illustrated in Figure 1-1.

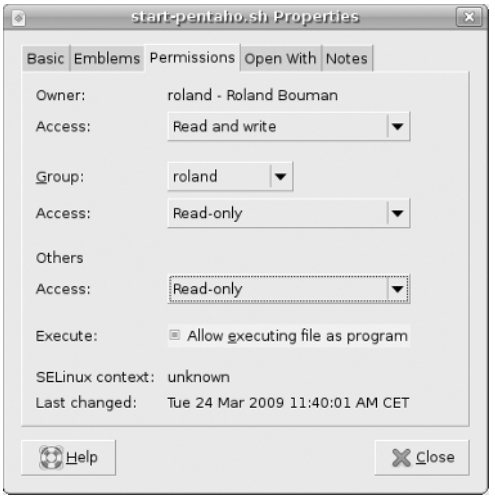


Figure 1-1: Making the start-pentaho.sh script executable

Alternatively, you can open a terminal and change directory (using the `cd` command) to the Pentaho home directory. From there, you can use the following command to make all `.sh` scripts executable:

```
shell> chmod ug+x *.sh
```

Now you can simply start the script by double-clicking it (you may need to confirm in a dialog) or by typing it in the terminal:

```
shell> ./start-pentaho.sh
```

After starting the script, you will see quite some output appearing in the console. You should leave open the terminal window in which you started the script.

NOTE **The `start-pentaho` script does two things.**

First, it starts a HSQLDB database server, which is used by the Pentaho server to store system data, as well as a sample database, which is used by most examples. By default, the HSQLDB database runs on port 9001. You should make sure no other server is running on that port.

Second, it starts a Tomcat server. By default, the Tomcat server listens on port 8080 for web requests. You should make sure no other server is running on that port, or the Pentaho BI Server will not be started successfully.

Logging in

After starting the server you can start your Internet browser to connect to the server. You should be able to use any of the major browsers (such as Mozilla Firefox, Microsoft Internet Explorer, Apple Safari, Opera, or Google Chrome) to do this. Navigate your browser to the following address:

```
http://localhost:8080
```

You are automatically redirected to the following:

```
http://localhost:8080/pentaho/Login
```

Shortly, you should see a Welcome page for the Pentaho user console. From there, you can log in to the server by pressing the large orange Login button. If you press the button, a Login box appears. From there, you can select a username from the drop-down list. For now, log in as the user Joe, as shown in Figure 1-2.

After selecting the username, you can press Login button to actually log in.

Figure 1-2: The Pentaho welcome screen and login dialog

Mantle, the Pentaho User Console

After confirming the login, you should see the Pentaho user console, as shown in Figure 1-3.

In the user console, you'll find a few elements to control the Pentaho BI Server:

- A menu bar, which is located at the top of the page and spans the page horizontally. Here you can find some standard menu items: File, View, Tools and Help.
- A toolbar containing several buttons, located immediately beneath the menu.
- A side pane, located on the left of the page, can be dynamically resized using the gray vertical bar at the far right of the pane. The pane can also be hidden/displayed in its entirety using the Toggle Browser button, which is the rightmost button on the toolbar.

- The tree view that is visible in the upper half of the side pane is called the Repository Browser. In Figure 1-3, this is labelled Browse. You can use this to browse through all BI content available in the Pentaho BI Server.
- A folder contents pane is located in the side pane, right beneath the solution repository browser. In Figure 1-3 this is labelled Files. It shows any contents of the selected folder in the solution repository (such as reports, dashboards and OLAP pivot tables) as a list of items. You can open an item by double-clicking it.
- A workspace. This is the larger pane on the right. When you double-click an item in the folder contents pane, it will be displayed here using a tab interface.

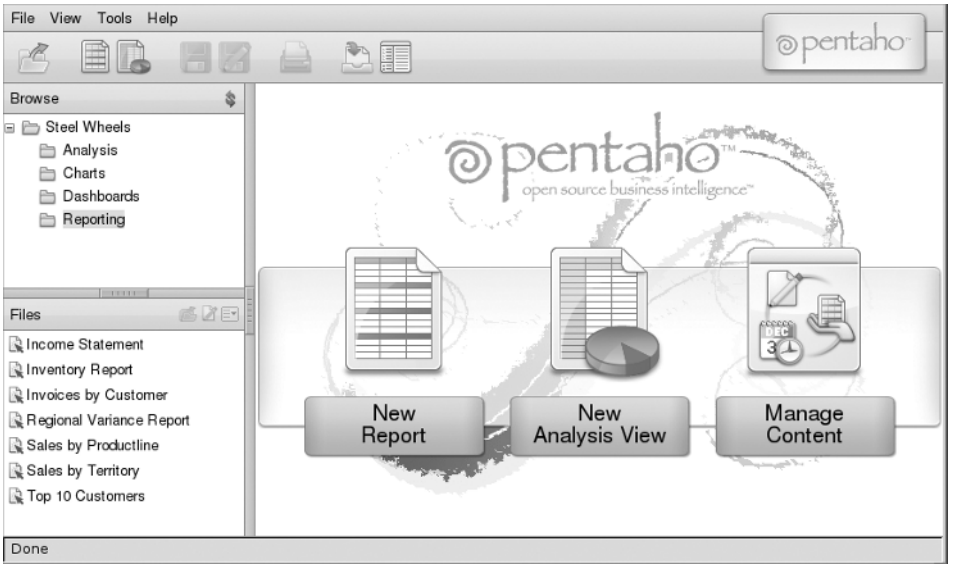


Figure 1-3: The Pentaho user console, also known as Mantle

Working with the Examples

The community edition of the Pentaho BI Server comes with two sets of examples:

- BI Developer Examples
- Steel Wheels

Each set of examples resides in its own Pentaho solution and is visible in the solution repository browser (see Figure 1-4).

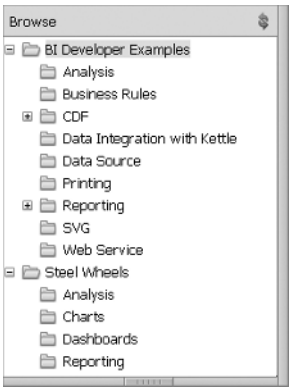


Figure 1-4: Two example solutions included in the Pentaho BI Server

Both of these Pentaho solutions contain good examples to demonstrate the types of reports you can create with Pentaho. Both solutions use the same sample data set. The BI Developer Examples focus more on the technical aspect of accomplishing a particular task, whereas the Steel Wheels examples illustrate how to combine techniques to build an application to support a classic cars business. The Steel Wheels examples also pay more attention to customizing look and feel.

Using the Repository Browser

You can access all of the examples using the repository browser. (This is the top pane of the left side bar in the user console, labelled Browse.) The repository browser offers a tree view that can be used to open and close the folders in the repository. To open a folder and reveal its subfolders, simply click once on the plus icon immediately on the left side of the folder icon. The folder's subfolders will become visible right beneath the parent folder, and the icon left of the folder icon changes to display a minus, indicating the folder is currently expanded. To close a folder and hide its subfolders, click on the minus icon.

To view the contents of a folder, click the folder icon or the folder name that appears directly on the right of the folder icon. The folder title will display a gray highlighting and its contents will become visible in the folder contents pane directly beneath the repository browser (in Figure 1-3, this is labelled Files).

To open an item that appears in the Files pane, double-click it. This will open a new tab page in the workspace, showing the output created by the item.

Understanding the Examples

Although you can learn a lot from the examples by simply running them, you can learn even more if you can see how they were built. Especially if you are a

Business Intelligence developer, you should consider examining the examples more closely using Pentaho Design Studio.

You'll learn the details about Pentaho Design Studio in Chapter 4, but you can follow these steps to get started quickly:

1. Download Pentaho Design Studio from the Pentaho downloads page at SourceForge.net.
2. Unzip the download to some location you find convenient.
3. Start Pentaho Design Studio. Microsoft Windows users can double-click `PentahoDesignStudio.exe`; users of UNIX-based systems can execute the `PentahoDesignStudio` binary file.
4. Use the main menu (File ➢ Switch Workspace) to change the workspace to the directory where you installed the Pentaho BI Server. The program will restart. In the opening splash screen, choose Workbench.
5. Create a new project by choosing File ➢ New ➢ Project. In the dialog, expand the General folder and choose Project to create a plain project. Click Next.
6. In the next dialog, enter `pentaho-solutions` for the project name. Make sure that whatever you type here corresponds exactly to the name of the `pentaho-solutions` directory located in the home directory of the Pentaho BI Server. The Use Default Location checkbox should be selected, and the location should automatically point to the Pentaho BI Server home directory.
7. Confirm the dialog.

In the Navigator tab page in the left side pane in Pentaho Design Studio, you should now see the `pentaho-solutions` project folder (which corresponds exactly with the actual `pentaho-solutions` folder). You can expand this folder and browse through the Pentaho solution repository.

Double-clicking on any items inside the folders will usually load the file in a new tab page in the Pentaho Design Studio Workspace. You can learn a lot, especially from opening the `.xaction` files that are present throughout the repository. Refer to Chapter 4 for more details on these files.

Beware that the items that show up in the repository browser in the user console of the Pentaho BI Server usually have a label that is distinct from the actual file name. This complicates things a bit in case you're looking for the corresponding item in Pentaho Design Studio, as the navigator there only displays file names. To discover the corresponding file name for any item shown in the repository browser, right-click the item and choose Properties in the context menu. This will pop up a dialog with a few tabs. The actual file name is shown in the General tab.

NOTE **The `.xaction` extension indicates an *action sequence*. Action sequences are Pentaho-specific lightweight processes to run or deliver BI content. In this particular case, the action sequence simply calls a Pentaho report. Action sequences are coded in a specific XML-format and typically stored in `.xaction` files. Action sequences are discussed in more detail in Chapter 4.**

Running the Examples

In the remainder of this chapter, we discuss a few items from these examples to give you a feel for what you can do with Pentaho solutions. For each item, we include references to the chapters of this book that relate to the example. We hope this will allow you to quickly get an overview of Pentaho's features and see how this book can help you master them.

Reporting Examples

Reporting is often one of the first requirements of any BI solution. Reporting is covered in detail in Chapter 13. Most of the reports discussed here are invoked from an action sequence; you can find more details on action sequences in Chapter 4.

The following sections examine a few of the reporting examples.

BI Developer Examples: Regional Sales - HTML

The Regional Sales - HTML example is one of the most straightforward reporting examples; as you would assume, it shows the sales figures for an example company broken down by region. You can find it in the Reporting folder in the BI Developer Examples set. The corresponding file name is `JFree_Quad.xaction`.

When you run the example, the report output is immediately shown in the workspace (see Figure 1-5).

In the report output you see an organization detailed by region (Central), department (Executive Management, Finance) and then position title (SVP Partnerships, CEO, and so on). For the position title level, you see the actual data. In this case, the data pertains to sales and shows the actual and projected (budgeted) sales numbers in the first two columns and the variance in the third column. You also see a totals line that sums up the figures for the department level, and if you could scroll down further you would also see the totals for the regional level, followed by the figures for another region. All the way down at the bottom of the report you would see totals for the entire business.

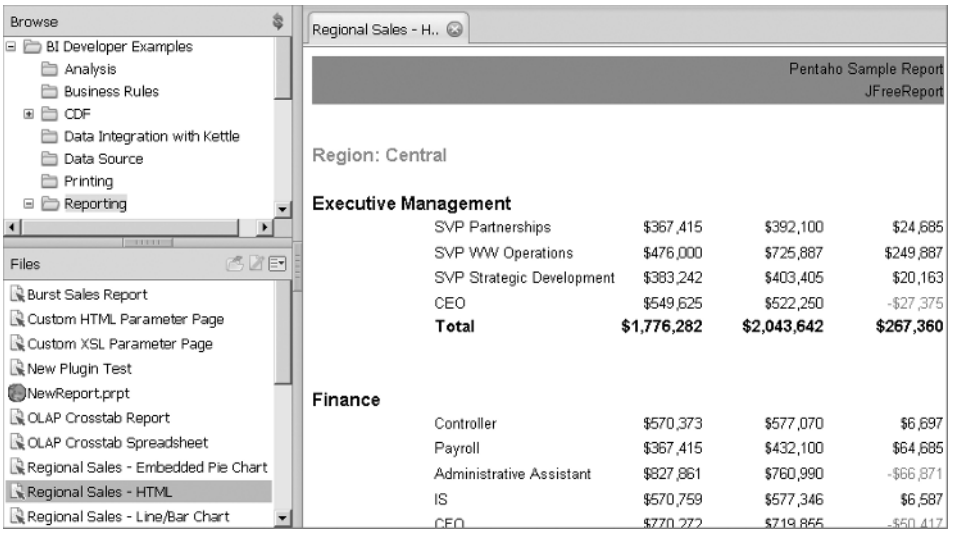


Figure 1-5: The Regional Sales - HTML sample report

Steel Wheels: Income Statement

The Income Statement example report from the Steel Wheels example set is another typical report with a self-explanatory name. You can find it in the Reporting folder beneath the Steel Wheels solution, and the corresponding file name is `Income Statement.xaction`. Figure 1-6 shows the report.

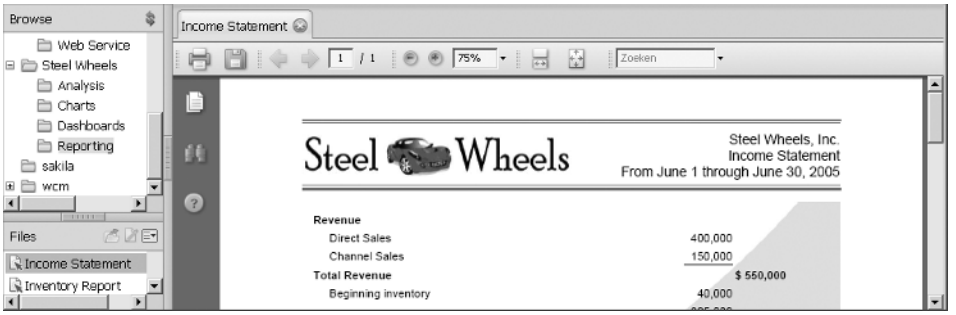


Figure 1-6: The Steel Wheels Income Statement report

A few differences from the Regional Sales report in the previous example are the styling and the output format. Although both reports were created with the Pentaho Report Designer, and both are rendered by the Pentaho reporting engine (which is the component responsible for interpreting reports and generating report output), they look quite different. Whereas the Regional Sales report outputs an HTML page, this report delivers a PDF file as output. In addition, this report shows adornments using a picture for a logo and a page background picture.

Steel Wheels: Top 10 Customers

In the previous section, we mentioned that the Income Statement report delivers output in the form of a PDF file, whereas the Regional Sales example outputs a plain web page. The Top 10 Customers report illustrates two more important features of the report output format. You can find this report also in the reporting folder in the Steel Wheels example set, and its file name is `Top Ten Customer ProductLine Analysis.xaction`. Running this example does not immediately show the report output, but displays the dialog shown in Figure 1-7 instead.

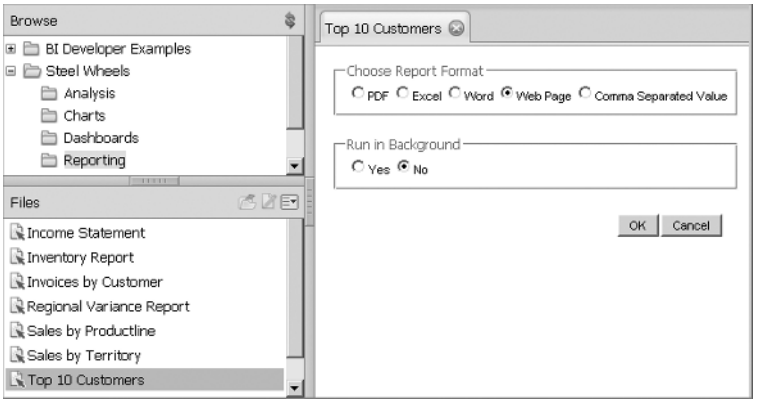


Figure 1-7: The Top 10 Customers report

As indicated by the dialog, you can choose from as many as five different output formats. In the previous reporting examples, the desired output format was stored as part of the report, but there is nothing in the reporting engine that forces this. This allows users to choose whatever format is most appropriate for the purpose at hand.

The dialog shown in Figure 1-7 illustrates another important feature of Pentaho reporting. The user can choose to wait for the report output now, or to have the Pentaho BI Server run the report in the background. The latter option will execute the report, but does not wait for the output to be returned. Rather, the output will be stored in the user's personal storage space on the server. This feature is especially useful for long-running reports.

You can find more on background execution and related features such as scheduling and subscription in Chapter 14.

BI Developer Examples: button-single-parameter.prpt

The previous example reports were all called from action sequences. In the upcoming Citrus release, reports can also be called directly. Examples

using this feature are all located in the Reporting folder in the BI Developer Examples set.

This example takes a closer look at the `button-single-parameter.prpt` example. When you start it, the report loads immediately in the workspace. However, the actual report output won't show until you press one of the Region buttons that appear in the Report Parameters section at the top of the page. Figure 1-8 illustrates what you might see after you press the Central button.

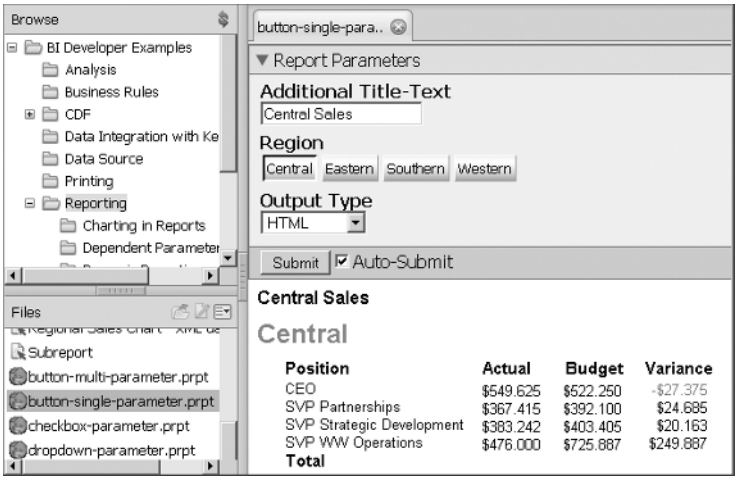


Figure 1-8: The button-single-parameter.prpt example

This example shows yet another feature of Pentaho, namely report parameters. Through parameters, the user can interact with the report and specify values to influence report behavior. Generally, this feature is used to allow the user to select only a portion of all possible report data.

In this example, there are two parameters. The Additional Title-Text parameter allows the user to specify a title that appears above all remaining report output. There is another parameter for Region, which allows the report to render output pertaining to only the specified region.

There are many more things you can do with report parameters, and these examples, as well as Chapter 13 of this book, should offer enough guidance for you to use this feature in a meaningful way.

Charting Examples

Whereas reports are great to communicate detailed information, they are less suitable for obtaining an overview of the data as a whole. For this purpose,

charts and graphs usually work better. Charts are also better suited than reports to display trends over time.

The Pentaho BI Server ships with two different charting solutions:

- JFreeChart—A 100% Java chart library.
- Pentaho Flash Charts—A charting solution based on open flash charts (which requires Adobe Flash).

Pentaho reporting offers full integration with JFreeChart, and you will find detailed information on integrating charts with your reports in Chapter 13. You can find more information about JFreeChart charts and how to integrate them with dashboards in Chapter 17.

Steel Wheels: Chart Pick List

The Chart Pick List example is located in the Charts folder in the Steel Wheels example set. The corresponding file name is `ChartComponent_ChartTypes.xaction`. Executing the item loads a dialog in the workspace that allows you to choose a particular chart type. After picking the chart type, you can press the Run button to actually display the chart. Figure 1-9 shows how this works for a Pie Grid.

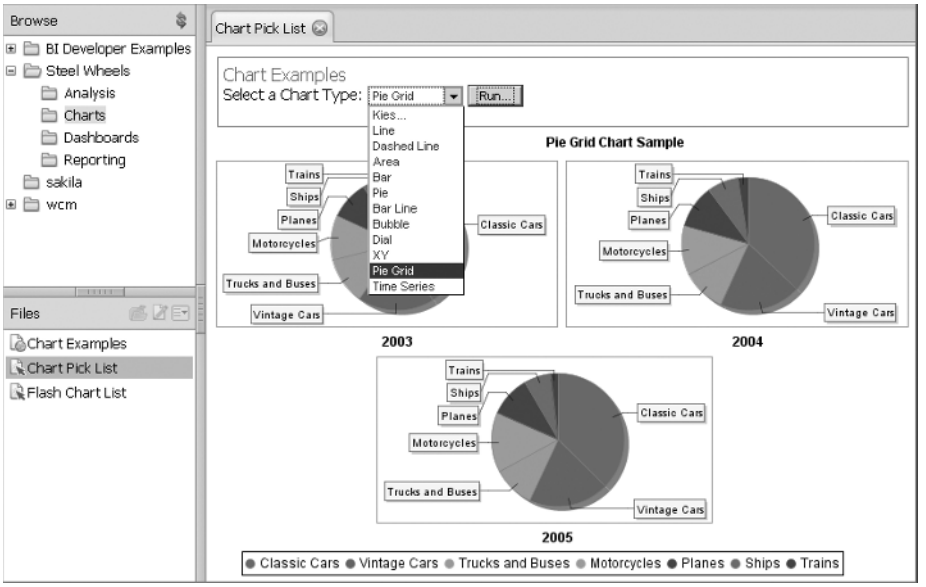


Figure 1-9: Pentaho charting using the JFreeChart Chart Pick List

Steel Wheels: Flash Chart List

Functionally, the Flash Chart List example is similar to the Chart Pick List example (which is based on JFreeChart). The difference is that the JFreeChart

Pick List example is based on the Open Flash Chart project. You can find the Flash Chart List also in the Charts folder within the Steel Wheels example set. The corresponding file name is `pentahoxml_picker.xaction`.

BI Developer Examples: Regional Sales - Line/Bar Chart

The Regional Sales - Line/Bar Chart example is located in the Reporting folder in the BI Developer Examples solution. The corresponding file is `JFree_SQLQuery_ComboChart.xaction`.

This example report displays a chart on the top of the page, and below that, a more detailed report shows the actual figures. In this case the chart is embedded into the report. The example report is shown if Figure 1-10.

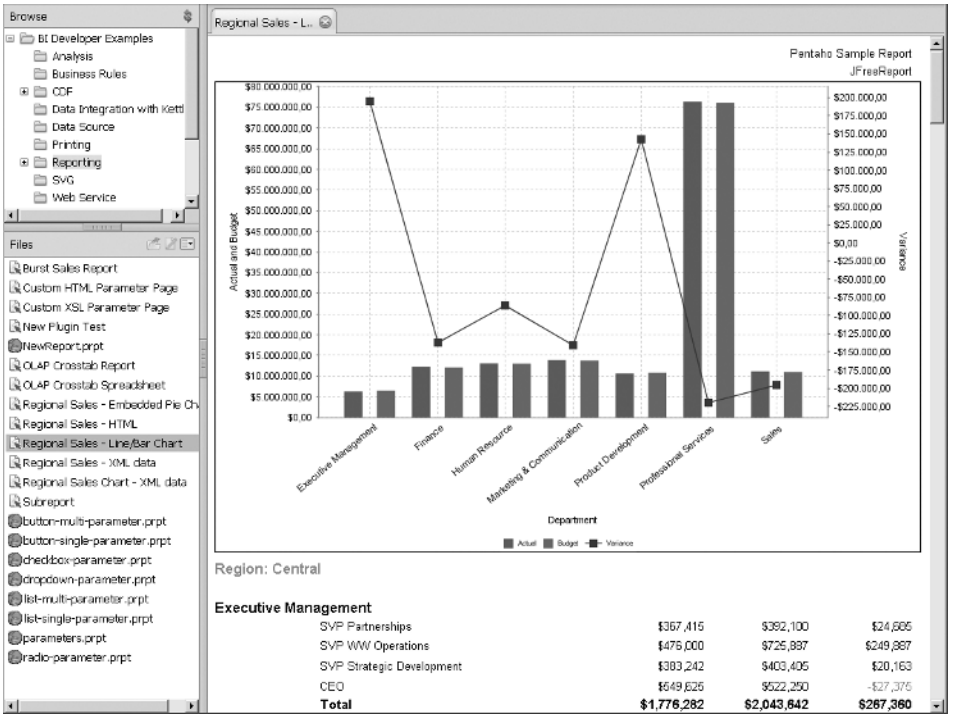


Figure 1-10: Regional Sales - Line/Bar Chart example

Analysis Examples

Like reporting, analysis is another essential feature of all BI solutions. Reports are typically static (save for parameters) and mainly used to support decisions that affect the business at the operational level. Analysis tends to be a lot more dynamic, and is typically used by managers to support decisions at the tactical and strategic level.

One of the typical elements in analytical solutions is that they allow the user to dynamically explore the data in an ad-hoc manner. Typically, the data is first presented at a highly aggregated level, say, total sales per year, and then the user can drill down to a more detailed level, say, sales per month per region. Any interesting differences between regions and/or months can then be used to drill into a new direction until a new insight or understanding of the business is obtained, which could then be used to affect plans for new promotions, next season's product catalog, or development of new products. This, in a nutshell, is what analysis is for.

Closely related to typical analytical questions and solutions is the dimensional model. Ultimately, this is what allows viewing data in aggregated form and features such as drill up/down. You will find detailed information about the dimensional model in Chapters 6, 7, and 8 of this book. In Chapter 15, we discuss the practical implementation of analytical applications using Mondrian and JPivot. All analytical examples presented in this chapter are based on Mondrian/JPivot.

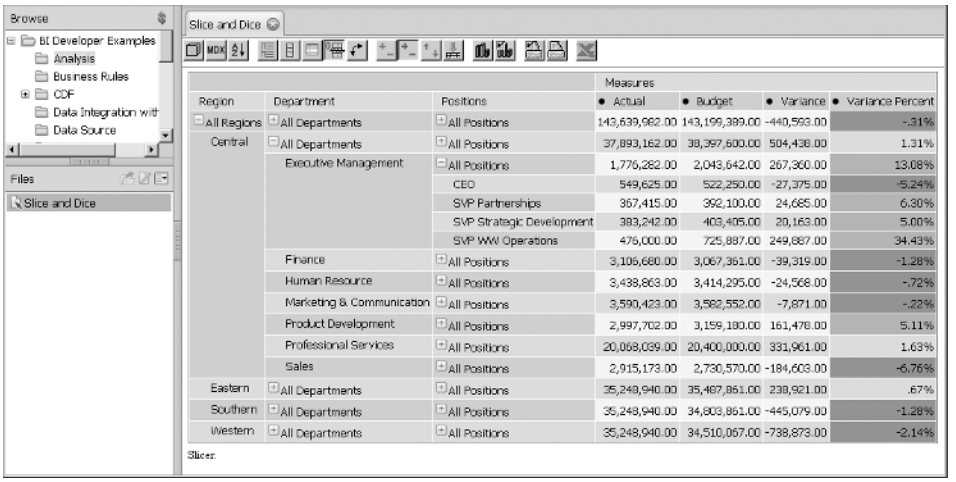
BI Developer Examples: Slice and Dice

The Slice and Dice example is located in the Analysis folder in the BI Developer Examples. Its corresponding file is called `query1.xaction`.

The Slice and Dice example is the most basic analysis example included with the Pentaho BI Server. Running it produces a dynamic crosstab, also known as a *pivot table*. The pivot table shows actual and budgeted sales figures, as well as actual versus budget variance. In the context of Analytics, figures like these are called *measures* or *metrics*. The measures can be split according to Region, Department, and Position. These headings are shown at the left side of the pivot table and represent *dimensions*, which are aspects that describe the context of the metrics.

A typical feature is that the pivot table not only shows the figures themselves but also totals, and that the totals can be computed at several levels of the dimensions (see Figure 1-11).

In Figure 1-11, you can see the columns for Region, Department, and Positions. The first row in the pivot table shows the results for All Regions, Departments, and Positions, and the figures are aggregated or ''rolled up'' along these dimensions. This represents the highest level of aggregation. Below that, you see that the data is split; in the first column, All Regions is split into Central, Eastern, Southern, and Western, forming the second-highest level of aggregation for the Region dimension. In the first row for each individual region, you see the data rolled up only across Department and Positions. For the Central region, the data is again split, this time showing all individual departments. Finally, for the Executive Management department, data is again split according to position.



Region	Department	Positions	Actual	Budget	Variance	Variance Percent
			Measures			
All Regions	All Departments	All Positions	143,639,962.00	143,199,389.00	-440,593.00	-.31%
Central	All Departments	All Positions	37,893,162.00	38,397,600.00	504,438.00	1.31%
	Executive Management	All Positions	1,776,282.00	2,043,642.00	267,360.00	13.08%
		CEO	549,625.00	522,250.00	-27,375.00	-5.24%
		SVP Partnerships	367,415.00	392,100.00	24,685.00	6.30%
		SVP Strategic Development	383,242.00	403,405.00	20,163.00	5.00%
		SVP WW Operations	476,000.00	725,887.00	249,887.00	34.43%
	Finance	All Positions	3,106,680.00	3,067,361.00	-39,319.00	-1.28%
	Human Resource	All Positions	3,438,863.00	3,414,295.00	-24,568.00	-.72%
	Marketing & Communication	All Positions	3,590,423.00	3,582,552.00	-7,871.00	-.22%
	Product Development	All Positions	2,997,702.00	3,159,180.00	161,478.00	5.11%
	Professional Services	All Positions	20,068,039.00	20,400,000.00	331,961.00	1.63%
	Sales	All Positions	2,915,173.00	2,730,570.00	-184,603.00	-6.76%
Eastern	All Departments	All Positions	35,248,940.00	35,487,861.00	238,921.00	.67%
Southern	All Departments	All Positions	35,248,940.00	34,803,861.00	-445,079.00	-1.28%
Western	All Departments	All Positions	35,248,940.00	34,510,067.00	-738,873.00	-2.14%

Figure 1-11: The Slice and Dice pivot table example

The splitting and rolling up is achieved dynamically by clicking on the plus and minus icons that appear next to the labels identifying Region, Department, and Positions. For example, by clicking on the plus icon next to any of the All Departments labels appearing in the second column, you can drill down and see how the rolled-up total value for any of the Sales metrics can be split up. Clicking a minus icon will roll the values back together into the total again, thus drilling up.

Steel Wheels Analysis Examples

In addition to the basic Slice and Dice example, you can find other interesting Analytics examples in the Analysis folder in the Steel Wheels example set. There you will find two examples:

- Market Analysis By Year
- Product Line Analysis

Like the basic Slice and Dice example, these examples display a pivot table, showing aggregated sales figures. In these examples, sales figures can be sliced along Product, Market (region), and Time.

Whereas the Slice and Dice example displayed only the measures on the horizontal axis, these examples show some more variety by placing the market on the horizontal axis. The Product Line Analysis example also places Time on the horizontal axis, beneath the Markets.

If you like, you can use alternative ways to set up the axes using the OLAP Navigator. You can invoke the OLAP Navigator by pressing the button with the cube icon on the toolbar that appears in the very top of the pages showing

the analysis examples. The OLAP Navigator and a part of that toolbar are shown in Figure 1-12.

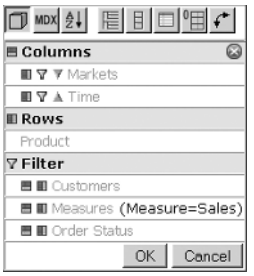


Figure 1-12: The OLAP Navigator

The OLAP Navigator shown in Figure 1-12 was taken from the Product Line Analysis example. In the top of the OLAP Navigator, you can see the caption Columns, and below that are two rows, Markets and Time. This corresponds directly with the Markets and Time shown along the horizontal axis of the pivot table. In the section below that, you see a Rows caption, with one row below it, Product. This corresponds with the products that are listed along the vertical axis of the pivot table. You can move the items in the Columns section to the Rows section and vice versa by clicking the small square in front of it.

There's a third section in the OLAP navigator labelled Filter. In this section, you find Customers, Measures, and Order Status. These items do not currently appear along one of the axes of the pivot table. You can move items from the Rows and Columns sections to the filter by clicking the filter icon. Moving items from the filter to either one of the axes is done by clicking the little square icon that corresponds to the axis to which you want to move the item.

We discuss the OLAP Navigator in detail in Chapter 15.

Dashboarding Examples

Dashboards are discussed in detail in Chapter 17. If you are interested in dashboards, you are strongly encouraged to check out the Community Dashboard Framework (CDF) dashboards examples included in the Pentaho BI Server. You can find them in the CDF folder in the BI Developer Examples solution.

A good way to start with Pentaho Dashboards is by navigating to the Samples subfolder of the CDF folder in the BI Developer Examples solution. Here you will find examples to use Charts, Reports, Analytic Pivot tables, and Maps in a dashboard, and see how you can tie these elements together.

Once you have a taste for what you can do with dashboards, you can read Chapter 17 and follow the detailed steps described there to build your own dashboard. When you are in the process of building your own dashboards, you

will find the documentation included with the CDF examples indispensable. You can find detailed documentation in the Documentation subfolder of the CDF folder. The documentation found in the Component Reference folder will be an especially invaluable companion.

Other Examples

Many more examples are included in the Pentaho BI Server. These include examples to start ETL processes, to call web services, to send report output to a printer or by e-mail, and much more. However, we will not discuss these examples here. Many of these examples require additional setup, and others are not particularly instructive unless you have need for that particular feature. However, readers are encouraged to experiment with the remaining examples.

Summary

This chapter provided an introduction to the Pentaho software and walked you through some of the examples that are shipped with it. After installing the software and exploring the examples, you should have a good idea of what you can do with Pentaho. The rest of this book will teach you how to work with each part of Pentaho to create your own Pentaho solutions.