

BIRT Integration and Report Tutorial

V1.1

December 12, 2013

Introduction

Summary

This tutorial will demonstrate how to use BIRT (**B**usiness **I**ntelligence and **R**eporting **T**ools) to develop reports to access the Hortonworks Sandbox and its integrated Hive Server. BIRT is an open source Eclipse-based reporting system that integrates with your Java/Java EE application to produce compelling reports. BIRT provides core reporting features such as report layout, data access and scripting.

Introduction

This tutorial will demonstrate how to integrate Open Source BIRT with the Hortonworks Sandbox by creating a data source connecting to the Sandbox, and a couple of data sets that retrieve data from the Sandbox's Hive 2 Server.

The data we will be retrieving is a snapshot in time of occupational information that includes an occupational code, description, total number of employees in that field, and the average salary for each of these occupations.

In this fictitious tutorial scenario; our goal is to identify occupations that have the best potential for the highest salary. Since we're curious, we'll be creating a report to visualize which occupations have the highest salaries; but more importantly, we also want to identify opportunities where we have the fewest number of employees, in a particular occupation, that also pays the highest average salary.

Prerequisites:

To complete this tutorial you will need to have successfully completed the "BIRT Connectivity" tutorial which demonstrates how to download and install the required dependencies as well as how to create a BIRT Project, BIRT Report, Data Source, and Data Set.

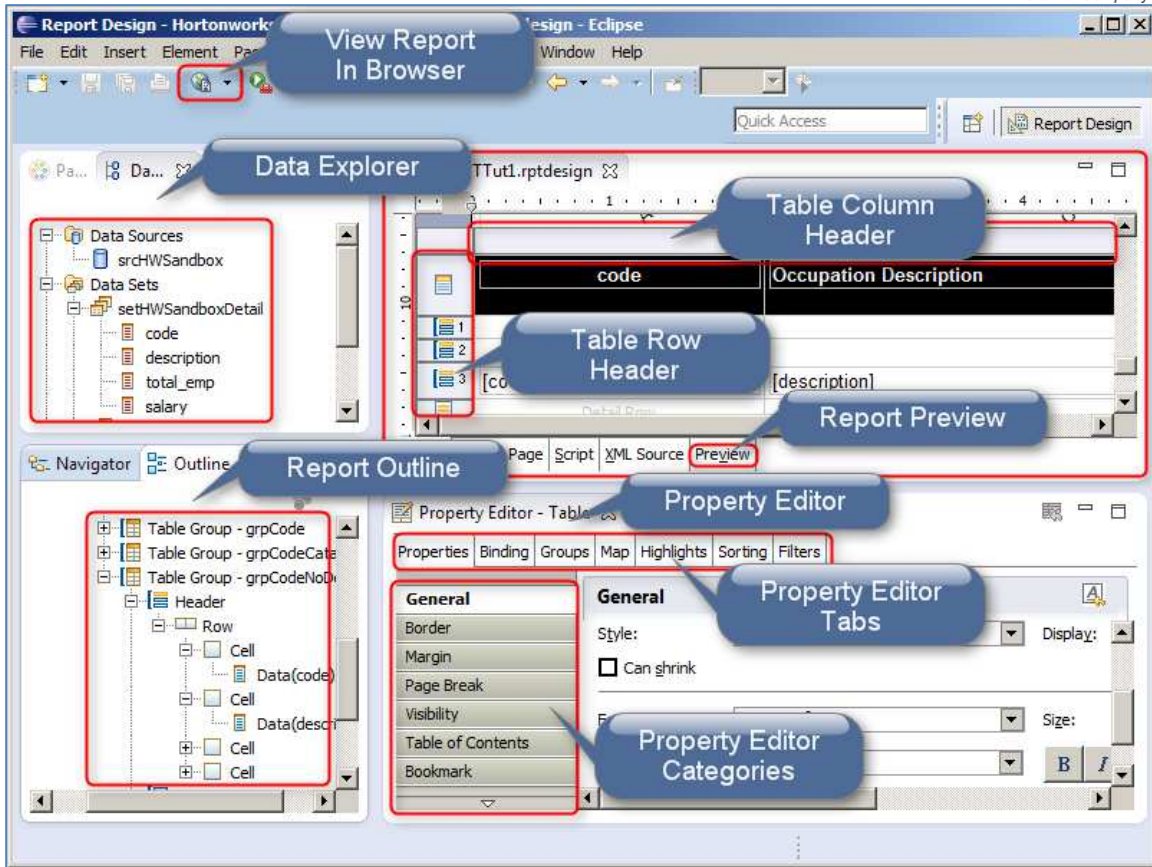
Overview

The main steps to complete this tutorial include:

1. Creating a BIRT Project, BIRT Report, Data Source, and Data Sets by connecting to the Hortonworks Sandbox Hive Server.
2. Creating a BIRT report utilizing charting and conditional table formatting features.
3. Viewing the BIRT in the browser report utilizing Paging and Exporting to formats such as PDF and Excel functionality.

Overview - Eclipse Open Source BIRT IDE

The following image is an overview of the Eclipse BIRT IDE with callouts identifying some of the major IDE components mentioned in this tutorial.



Overview - Tutorial Sample Report

The following image displays part of the BIRT report we will be building with data from the Hortonworks Sandbox.

LOOKING TO MAKE MORE MONEY ?

FIND THE BEST OPPORTUNITY

Although we often choose a job based on a number of factors, such as an aptitude for working with numbers, a natural attention to detail or an interest in serving people, salary is a major factor in selecting any career path.

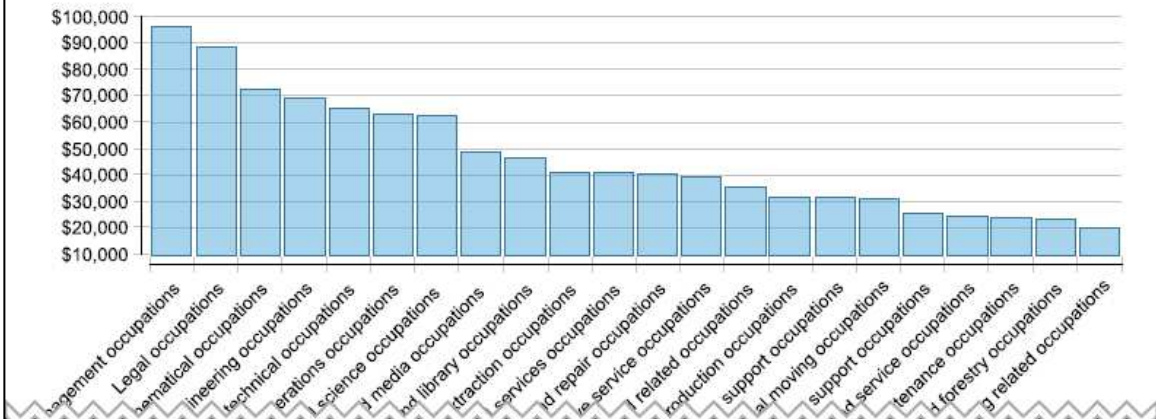
Satisfaction in your career is definitely not based solely on how much money you make. However, to help facilitate the salary portion of your job search, we are providing this report to help you determine target occupations that may provide the best opportunity for the highest salary.

The basic premise of this report is to help you identify job classifications that pay well, combined with how many existing employees in each job classification, to determine target job classification opportunities that pay well and have low competition.



Top 5 Salary Occupation Classifications	Average Salary
Management occupations	\$96,150
Legal occupations	\$88,450
Computer and mathematical occupations	\$72,190
Architecture and engineering occupations	\$68,880
Healthcare practitioners and technical occupations	\$65,020

Average Salary per Occupation Classification



Occupation Description	Employees in Job	Average Salary
Management occupations	6,003,930	\$96,150
Chief executives	299,160	\$151,370
General and operations managers	1,655,410	\$103,780
Legislators	61,110	\$33,880
Advertising and promotions managers	36,300	\$91,100
Marketing managers	165,240	\$113,400
Sales managers	322,170	\$106,790
Public relations managers	47,210	\$97,170
Administrative services managers	239,360	\$76,370
Computer and information systems managers	264,990	\$113,880
Financial managers	484,390	\$106,200
Compensation and benefits managers	41,780	\$88,400
Training and development managers	28,170	\$90,300
Human resources managers, all other	58,100	\$99,810
Industrial production managers	152,870	\$87,550

Step 1 - Create BIRT Project, BIRT Report, Data Source and Data Sets

Watch the [Video](#)

As outlined in the "BIRT Connectivity" tutorial create the following:

- A new BIRT Project named "HortonworksBIRTTutorial"
- A new BIRT Report named "HWBIRTTut1.rptdesign"
- A new Hive Data Source named "srcHWSandbox"
- 2 new Data Sets named 1) "setHWSandboxDetail" and 2) "setHWSandboxSummary" both using the Select Query of "SELECT * FROM sample_07" (we will be modifying the data sets below)

Now that we've done this setup, follow these steps to configure our data sets to retrieve the desired data.

First we'll modify the detail data set to eliminate an aggregate row existing in the data.

1. Double click the "setHWSandboxDetail" Data Set to bring up the Data Set Editor dialog.
2. Click **Filters > New...**
3. Enter or select the following values
 - a. (1st dropdown) select "code"
 - b. (2nd dropdown) select "Not Match"
 - c. (3rd dropdown) enter "00-0000" - include the quotes
4. Click **OK**
5. Click **OK**

Next we'll modify the summary data set to only include the aggregate summary rows existing in the data.

1. Double click the "setHWSandboxSummary" Data Set to bring up the Data Set Editor dialog.
2. Click **Filters > New...**
3. Enter or select the following values
 - a. (1st dropdown) select "code"
 - b. (2nd dropdown) select "Not Match"
 - c. (3rd dropdown) enter "00-0000" - include the quotes
4. Click **OK**
5. Click **New...** again to add a second filter
6. Enter or select the following values
 - a. (1st dropdown) select "code" then append ".substr(3,4)" resulting in "row["code"].substr(3,4)"
 - b. (2nd dropdown) select "Match"
 - c. (3rd dropdown) enter "0000" - include the quotes
7. Click **OK**
8. Click **OK**

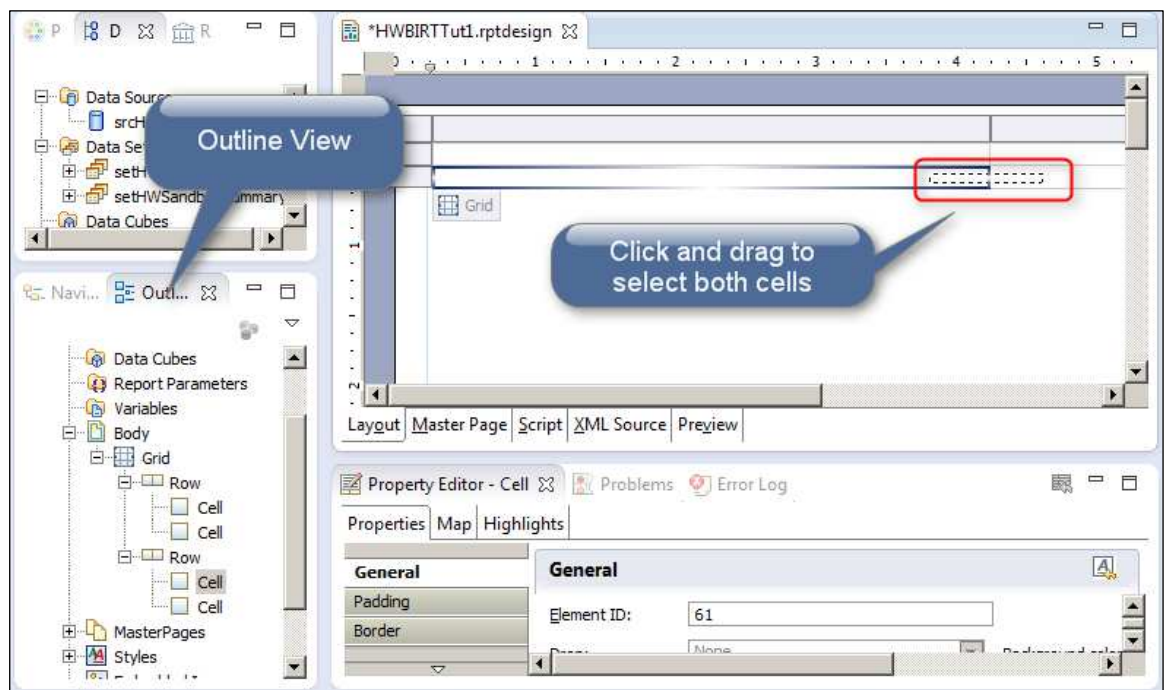
Now that we are retrieving the desired data we'll start designing our report.

Step 2 – Designing Page 1 of Report – Narrative

Watch the [video](#).

In this section, we will design the narrative for the first page of our report. To help with our design open the "Outline" view by click it's tab as shown in the first image of this tutorial "Eclipse BIRT IDE ". Note you can preview the report at anytime by clicking the "Preview" tab, or the "View in Browser" icon in the toolbar as shown in the first image of this tutorial "Eclipse BIRT IDE ".

1. We're not going to have a header or footer in this report so click the **Master Page** tab and set the header and footer heights to 0
2. We also want to use the entire width of the report so in **Margin** set the left and right margin to 0
3. Create a grid to hold our content. Right click on blank report background and select **Insert > Grid**, set number of columns to 2 and number of rows to 2, click **OK**.
4. In the bottom row of the grid you've just create click and drag the mouse across both cells to select them both as show in the image below. Then right click and select **Merge Cells**.



5. Select the top left cell from the outline view, right click and select **Insert Element** and choose **Label**. Repeat this to add a second and third label. Then repeat again but this time add a **Text** element which will bring up the **Edit Text Item** dialog (you should have 3 labels and one text element).
6. In the **Edit Text Item** dialog set the type to be **HTML** then copy and paste the text below into the HTML text area (without the quotes).
 - a. "Although we often choose a job based on a number of factors, such as an aptitude for working with numbers, a natural attention to detail or an interest in serving people, salary is a major factor in selecting any career path.

Satisfaction in your career is definitely not based solely on how much money you make. However, to help facilitate the salary portion of your job search, we are providing this

report to help you determine target occupations that may provide the best opportunity for the highest salary.

The basic premise of this report is to help you identify job classifications that pay well, combined with how many existing employees in each job classification, to determine target job classification opportunities that pay well and have low competition."

7. Double click the first label you created and enter the text "LOOKING TO MAKE MORE MONEY ?" (without the quotes)
8. Double click the second label you created and enter the text "FIND THE BEST" (without the quotes)
9. Double click the second label you created and enter the text "OPPORTUNITY" (without the quotes)
10. Style these labels and text item as you see fit using the Property Editor. The following is how it was done as shown in this tutorial.
 - a. Label 1
 - i. Font: Times New Roman
 - ii. Size: 16
 - iii. Color: RGB(128,128,64)
 - iv. Bold
 - b. Labels 2 and 3
 - i. Font: Times New Roman
 - ii. Size: 36
 - iii. Color: RGB(0,64,128)
 - iv. Bold
 - c. Text Item
 - i. Font: Sans Serif
 - ii. Size: 10
 - iii. Color: RGB(0,64,128)
 - iv. Padding: Left 50, Right 50

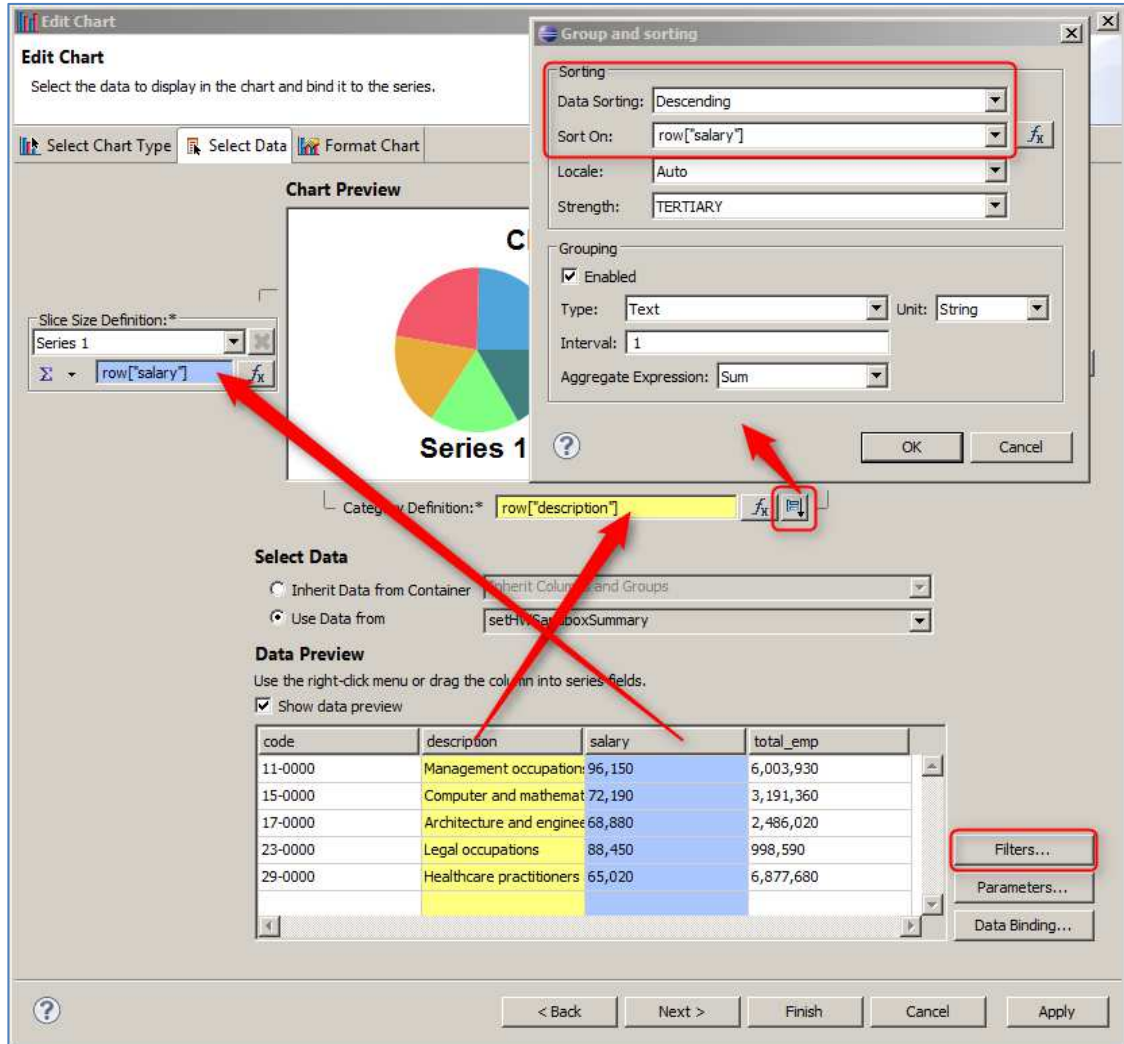
Step 3 – Designing Page 1 of Report - Pie Chart

Watch the [video](#).

In this section, we will design the pie chart for the first page of our report.

1. Select the top right, cell from the outline view, right click and select **Insert Element** and choose **Chart**.
2. In the New Chart dialog, select **Pie** as the chart type and click **Next >**
3. Click **Use Data From** and select "setHWSandboxSummary" from the drop down
4. Click the "salary" column header and drag to **Slice Size Definition**
5. Click the "description" column header and drag to **Category Definition**
6. Click **Filters... > Add...**
7. Enter or select the following values
 - a. (1st dropdown) select "salary"
 - b. (2nd dropdown) select "Top n"
 - c. (3rd dropdown) enter "5" - do not include the quotes
8. Click **OK**
9. Click **OK**
10. Click the sorting icon to the far right of **Category Definition**
11. In Data Sorting, select Descending

12. In Sort On, select "row[\"salary\"]"
13. Click **OK**



Edit Chart

Select the data to display in the chart and bind it to the series.

Select Chart Type | Select Data | Format Chart

Chart Preview

Slice Size Definition:*

Series 1: Σ row[\"salary\"]

Category Definition: row[\"description\"]

Group and sorting

Sorting

Data Sorting: Descending

Sort On: row[\"salary\"]

Locale: Auto

Strength: TERTIARY

Grouping

Enabled

Type: Text Unit: String

Interval: 1

Aggregate Expression: Sum

OK Cancel

Select Data

Inherit Data from Container Inherit Columns and Groups

Use Data from setHortonSandboxSummary

Data Preview

Use the right-click menu or drag the column into series fields.

Show data preview

code	description	salary	total_emp
11-0000	Management occupation	96,150	6,003,930
15-0000	Computer and mathemat	72,190	3,191,360
17-0000	Architecture and engine	68,880	2,486,020
23-0000	Legal occupations	88,450	998,590
29-0000	Healthcare practitioners	65,020	6,877,680

Filters... Parameters... Data Binding...

< Back Next > Finish Cancel Apply

14. Click **Next >**
15. Style this chart as you see fit. The following is how it was done as shown in this tutorial.
 - a. Series
 - i. Delete the **Title** in **Value Series**
 - ii. Check **Translucent**
 - b. Series | Value Series
 - i. Slice | By Distance: 2
 - ii. Check **Show Series Labels**
 - iii. Click **Labels** > set **Position**: Inside, **Prefix**: \$
 - c. Chart Area | Title
 - i. Uncheck **Visible**
 - d. Chart Area | Legend
 - i. Uncheck **Visible**
16. Click **Finish**
17. In the Property Editor - set this chart's height and width to 3 inches

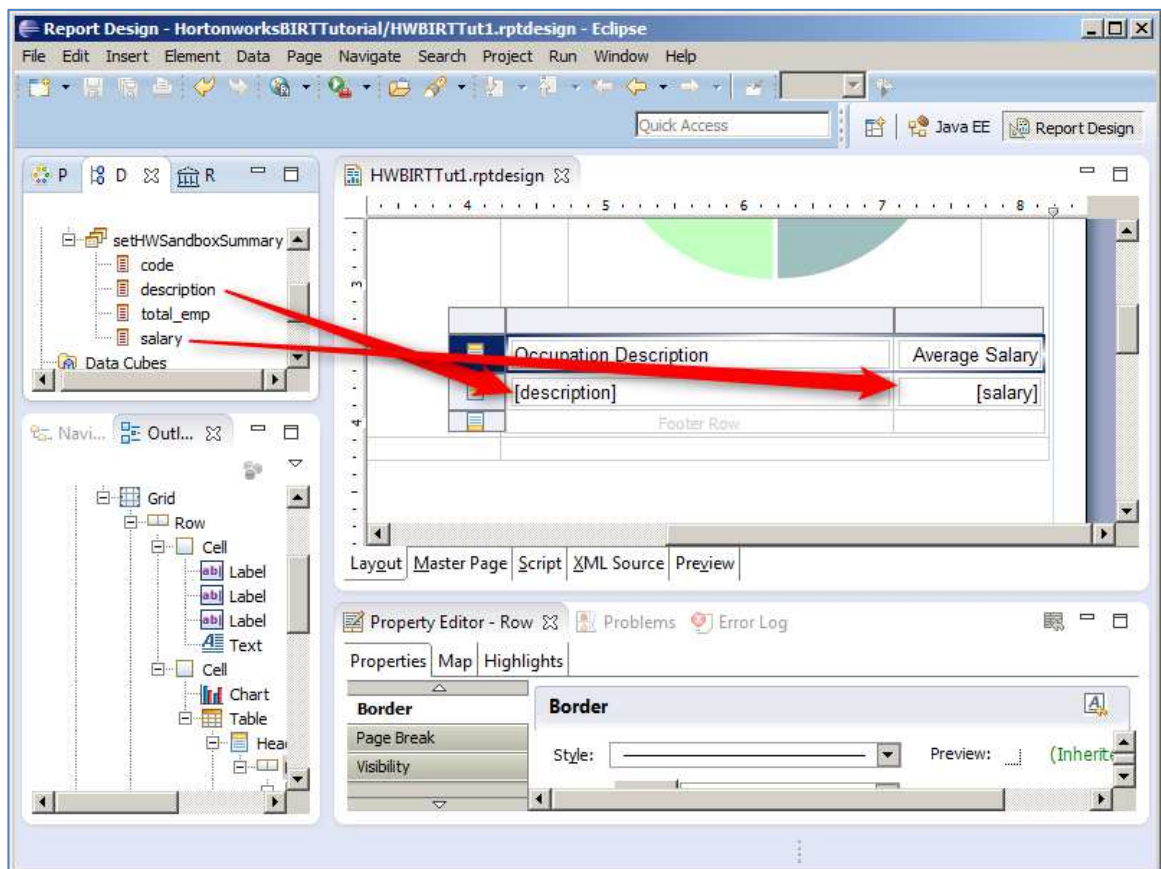
- Click the grid's cell containing the pie chart, and set its horizontal alignment to center in the Property Editor.

Step 4 – Designing Page 1 of Report - Summary Table

Watch the [video](#).

In this section, we will design the summary table located under the pie chart for the first page of our report.

- Select the top right, cell from the outline view, right click and select **Insert Element** and choose **Table**. In the Insert Table dialog set the Number of columns to 2, and choose the Data Set: "setHWSandboxSummary". Click **OK**.
- In the **Property Editor** click the **Sorting** tab. Click **Add...**, specify a descending sort on salary as you did in the pie chart.
- In the **Property Editor** click the **Filters** tab. Click **Add...**, specify a top 5 filter on salary as you did in the pie chart.
- In the **Data Explorer**, expand the Data Set: "setHWSandboxSummary", and drag the column "description" to this table's first column in the detail row and drag the column "salary" to this table's second column in the detail row.



- Select the "salary" data field you've just dragged into this table and in **Property Editor** select **Format Number** and set as **Currency**, 0 decimal places, check the **Use 1000s separator**, and select a "\$" as the **Symbol**.

6. Style this table as you see fit. The following is how it was done as shown in this tutorial.
 - a. Left align the "description" column
 - b. Right align the "salary" column
 - c. Change the "description" header to "Top 5 Salary Occupation Classifications"
 - d. Change the "salary" header to "Average Salary"
 - e. Reduce the width of the "salary" column by dragging the column header divider to the right
 - f. Select the header row's header, and specify a 1px solid bottom border in the **Property Editor**

Step 5 – Designing Page 1 of Report - Bar Chart

Watch the [video](#).

In this section, we will design the bar chart for the first page of our report.

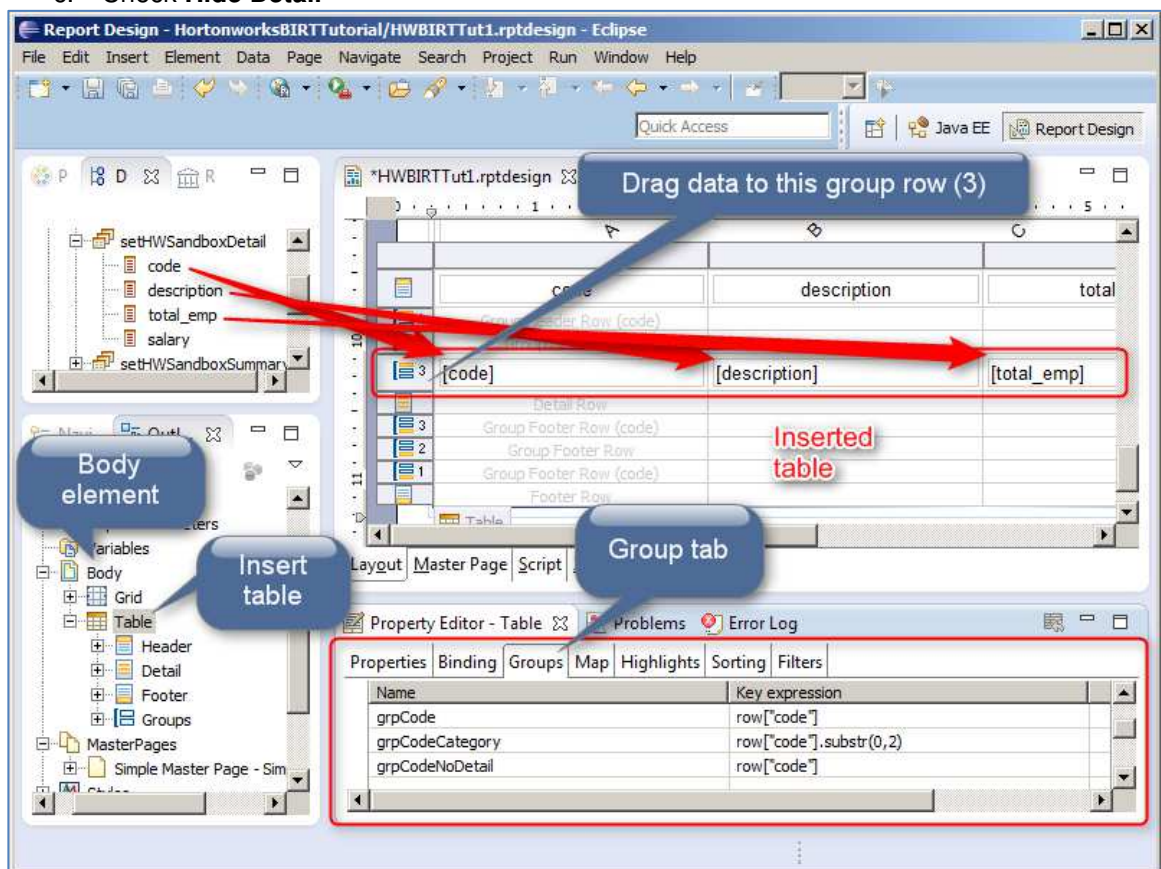
1. Select the bottom cell (this is the merged cell we did earlier) of the grid from the outline view, right click and select **Insert Element** and choose **Chart**.
2. In the New Chart dialog, select **Bar** as the chart type and click **Next >**
3. Click **Use Data From** and select "setHWSandboxSummary" from the drop down
4. Click the "salary" column header and drag to **Slice Size Definition**
5. Click the "description" column header and drag to **Category Definition**
6. Click the sorting icon to the far right of **Category Definition**
7. In Data Sorting, select Descending
8. In Sort On, select "row["salary"]"
9. Click **OK**
10. Click **Next >**
11. Style this chart as you see fit. The following is how it was done as shown in this tutorial.
 - a. Series
 - i. Check **Translucent**
 - b. Chart Area
 - i. Click **General Properties**, set **Unit Spacing** to 10
 - c. Chart Area | Axis | X-Axis
 - i. Click **A** to the right of **Labels**
 - ii. Set the **Degree** to 45
 - d. Chart Area | Axis | Y-Axis
 - i. Click **Format**
 - ii. Select Standard, Prefix: "\$", Fraction Digits: 0
 - iii. Click **Gridlines**
 - iv. Check **Major Grid | Visible**
 - e. Chart Area | Title
 - i. Set Chart Title to "Average Salary per Occupation Classification"
 - ii. Click **A** to the right of **Font**
 - iii. Set Font size to 10
 - f. Chart Area | Legend
 - i. Uncheck **Visible**
12. Click **Finish**
13. In the Property Editor - set this chart's width to 8 inches and height to 5.25 inches

Step 6 – Designing Remaining Pages of Report - Detail Table

Watch the [video](#).

In this section, we will design the details table for the remainder of our report.

1. Select the **Body** from the outline view, right click and select **Insert Element** and choose **Table**. In the Insert Table dialog set the Number of columns to 4, and choose the Data Set: "setHWSandboxDetail". Click **OK**.
2. In the **Property Editor** select **Page Break** and set **Before** to "Always" and **Page Break Interval** to 50
3. We're going to group this table with three individual groupings to get the results we desire. Click the Table you just added in the outline, then click the **Groups** tab in the **Property Editor**.
4. Click **Add...**
 - a. Name: grpCode
 - b. Group On: code
5. Click **Add...**
 - a. Name: grpCodeCategory
 - b. Group On: code
 - c. Click function **Fx** | append ".substr(0,2)" to the existing value - resulting in "row["code"].substr(0,2)"
6. Click **Add...**
 - a. Name: grpCodeNoDetail
 - b. Group On: code
 - c. Check **Hide Detail**



7. Delete all data fields that were added to the table during the groupings

8. In the **Data Explorer**, expand the Data Set: "setHWSandboxDetail", and drag column "code" to this table's first column in the group (3) row, drag the column "description" to this table's second column in the group (3) row, drag the column "total_emp" to this table's third column in the group (3) row, drag the column "salary" to this table's fourth column in the group (3) row.
9. Click the column header for the column "code", and check **Hide Element** in **Visibility** in the **Property Editor** - we won't be displaying this column
10. Click each of the 5 row headers for the groups (headers and footers) that do not contain the data fields and set their **Visibility** to **Hide Element**
11. In the outline view click the cell that contains the data field "salary" (this is under the "grpCodeNoDetail" Table Group), set its background color to yellow in the Property Editor, and give it a 1px solid white right side border and add the following conditional formatting by clicking the **Highlights** tab | **Add...**
 - a. We want to show high salary as green so set...
 - i. (1st dropdown) select "salary"
 - ii. (2nd dropdown) select "Top percent"
 - iii. (3rd dropdown) enter 25
 - iv. Background Color: Green
 - b. We want to show low salary as red so click **Add..** again and set...
 - i. (1st dropdown) select "salary"
 - ii. (2nd dropdown) select "Bottom percent"
 - iii. (3rd dropdown) enter 25
 - iv. Background Color: Red
 - c. We don't want to color code the summary occupation rows so set this to be a consistent looking header by clicking **Add...** again and set...
 - i. (1st dropdown) select "code" and append ".substr(3,4)" resulting in "`row["code"].substr(3,4)`"
 - ii. (2nd dropdown) select "Match"
 - iii. (3rd dropdown) enter "0000" - including the quotes
 - iv. Color: White
 - v. Bold
 - vi. Background Color: Gray
12. In the outline view click the cell that contains the data field "total_emp" (this is under the "grpCodeNoDetail" Table Group), set its background color to yellow in the Property Editor, and give it a 1px solid white right side border and add the following conditional formatting by clicking the **Highlights** tab | **Add...**
 - a. We want to show high number of employees as red so set...
 - i. (1st dropdown) select "total_emp"
 - ii. (2nd dropdown) select "Top percent"
 - iii. (3rd dropdown) enter 25
 - iv. Background Color: Red
 - b. We want to show low number of employees as green so click **Add..** again and set...
 - i. (1st dropdown) select "total_emp"
 - ii. (2nd dropdown) select "Bottom percent"
 - iii. (3rd dropdown) enter 25
 - iv. Background Color: Green
 - c. We don't want to color code the summary occupation rows so set this to be a consistent looking header by clicking **Add...** again and set...
 - i. (1st dropdown) select "code" and append ".substr(3,4)" resulting in "`row["code"].substr(3,4)`"
 - ii. (2nd dropdown) select "Match"
 - iii. (3rd dropdown) enter "0000" - including the quotes
 - iv. Color: White

- v. Bold
 - vi. Background Color: Gray
13. To further conditionally style the rest of the rows click the row header that contains the data fields, click the **Highlights** tab in the **Property Editor**, click **Add..**
- a. We want to slightly differentiate each row so we'll add a light background for every other row
 - i. (1st dropdown) enter "row.__rownum % 2"
 - ii. (2nd dropdown) select "Equal to"
 - iii. (3rd dropdown) enter "0" - no quotes
 - iv. Background Color: choose a very light gray - RGB(238,238,238)
 - b. We also want to set the summary occupation rows be a consistent looking header by clicking **Add...** again and set...
 - i. (1st dropdown) select "code" and append ".substr(3,4)" resulting in "row["code"].substr(3,4)"
 - ii. (2nd dropdown) select "Match"
 - iii. (3rd dropdown) enter "0000" - including the quotes
 - iv. Color: White
 - v. Bold
 - vi. Background Color: Gray
14. Select the "salary" data field and in **Property Editor** select **Format Number** and set as **Currency**, 0 decimal places, check the **Use 1000s separator**, and select a "\$" as the **Symbol**.
15. Select the "total_emp" data field and in **Property Editor** select **Format Number** and set as **Fixed**, 0 decimal places, check the **Use 1000s separator**.
16. Style this table as you see fit. The following is how it was done as shown in this tutorial.
- a. Left align the "description" column
 - b. Right align the "total_emp" column
 - c. Right align the "salary" column
 - d. Double click the "description" header label and rename it to "Occupation Description"
 - e. Double click the "total_emp" header label and rename it to "Employees in Job"
 - f. Double click the "salary" header label and rename it to "Average Salary"
 - g. Click the row header of the row containing the data fields and In **Property Editor** give this row a 1px solid white top border
 - h. Resize the "salary" and "total_emp" columns to just fit their data values
 - i. Click the row header of the top header row with the column labels and give the row a black background color, white font color and bold font

Step 7 – Viewing Report

Watch the [video](#).

In this section, we let you explore on your own - page through the report, explore the various export formats such as, Microsoft Word, PowerPoint, Excel, PDF, etc. As shown in the image below you have many export options.

The screenshot shows the BIRT Report Viewer interface. The main report content includes a title "LOOK FOR MORE MONEY? FIND THE BEST OPPORTUNITY", a pie chart with four segments labeled with salary values: \$88,450 (pink), \$96,150 (blue), \$72,190 (yellow), and \$65,020 (green). Below the pie chart is a table titled "Average Salary" with the following data:

Occupation	Average Salary
Legal occupations	\$96,150
Management occupations	\$88,450
Business and financial operations occupations	\$72,190
Community and social services occupations	\$68,880
Healthcare practitioners occupations	\$65,020

An "Export Report" dialog box is open, showing options for "Export Format" (DOCX, Excel, OpenDocument Presentation, OpenDocument Spreadsheet, OpenDocument Text, PDF, PostScript, PowerPoint, PPTX, Word, xis_spudsoft, xlsx) and "Fit To" (Auto, Page). A bar chart at the bottom of the report shows salary ranges for various occupation categories, with the y-axis ranging from \$10,000 to \$100,000.

Now that you have successfully performed all of the steps in the procedure, you should have a good handle on what can be accomplished using Open Source BIRT, as well as reporting from data stored in the Hortonworks Sandbox's Hive Server.

Survey

Thank you for downloading and working through this tutorial. We're interested in your feedback. Please take a minute to answer a few questions in this [survey](#) so that we can continue to improve.

About Actuate - The BIRT Company™

Actuate founded and co-leads the [BIRT open source project](#), which is used by more than 2.5 million developers around the globe and serves as the foundation of Actuate's commercial offerings. Applications built with BIRT and BIRT iHub deliver more business and consumer insights to more people than all BI companies combined - ensuring organizations are ready for the exponential growth of Big Data and the proliferation of touch devices. Actuate's BIRT and BIRT iHub empower developers to rapidly develop custom, BIRT-based [business analytics](#) and [customer communications](#) applications. Applications built with one BIRT design can access and integrate any data, including unstructured sources. They provide a seamless user experience regardless of skill level and are supported by BIRT iHub for any cloud, hybrid, on-premise, web or touch device deployment. Headquartered in [Silicon Valley](#), Actuate has over 5,000 customers globally in a diverse range of business areas including [financial services](#), [technology](#) and the [government](#). Actuate is listed on NASDAQ under the symbol BIRT. For more information, visit www.actuate.com or engage with the BIRT community at www.birt-exchange.com.

About Hortonworks

Hortonworks develops, distributes and supports the only 100-percent open source distribution of Apache Hadoop explicitly architected, built and tested for enterprise grade deployments. Developed by the original architects, builders and operators of Hadoop, Hortonworks stewards the core and delivers the critical services required by the enterprise to reliably and effectively run Hadoop at scale. Our distribution, Hortonworks Data Platform, provides an open and stable foundation for enterprises and a growing ecosystem to build and deploy big data solutions. Hortonworks also provides unmatched technical support, training and certification programs. For more information, visit www.hortonworks.com. The Hortonworks Sandbox can be found at: www.hortonworks.com/sandbox.