Visualizing Big Data with Eclipse BIRT

Virgil Dodson
Director, BIRT Evangelists
Actuate Corporation
December 18, 2013

vdodson@actuate.com

Today's Agenda and Goals

- Introduction to Big Data
- Compare with 'Regular' Data
- Survey results
- Big Data connections in BIRT
- Live demo
- Deploying BIRT applications
- Questions

Big Data Manifesto

Providing effective business analytics tools and technologies to the enterprise is a top priority of CIOs and for good reason.

Effective business analytics – from basic reporting to advanced data mining and predictive analytics — allows data analysts and business users alike to **extract insights** from **corporate data** that, when translated into action, deliver higher levels of efficiency and **profitability** to the enterprise.

Underlying every business analytics practice is data.…

Source: Wikibon Community

Big Data Definition

Big data is a collection of data sets **so large and complex** that it becomes **difficult to process** using on-hand database management tools or traditional data processing applications

Web Logs	**RFID**	Sensors	**Social Networks**
Internet Text	Searches	Call Detail Records	Astronomy
Atmospheric Info	Genomics	Biogeochemical	Biological
Military Surveillance	Medical Records	**E-Commerce**	Photo/Video

Source: Wikipedia

Traditional Data vs. Big Data

Traditional Data	Big Data
Gigabytes to Terabytes	Petabytes to Exabytes
Centralized	Distributed
Structured	Semi-structured to Unstructured
Stable Data Model	Flat Schemas
Known Complex Interrelationships	Few Complex Interrelationships

Source: Wikibon Community

When to Use Big Data vs. Relational

Is Big Data a replacement for Relational Data?

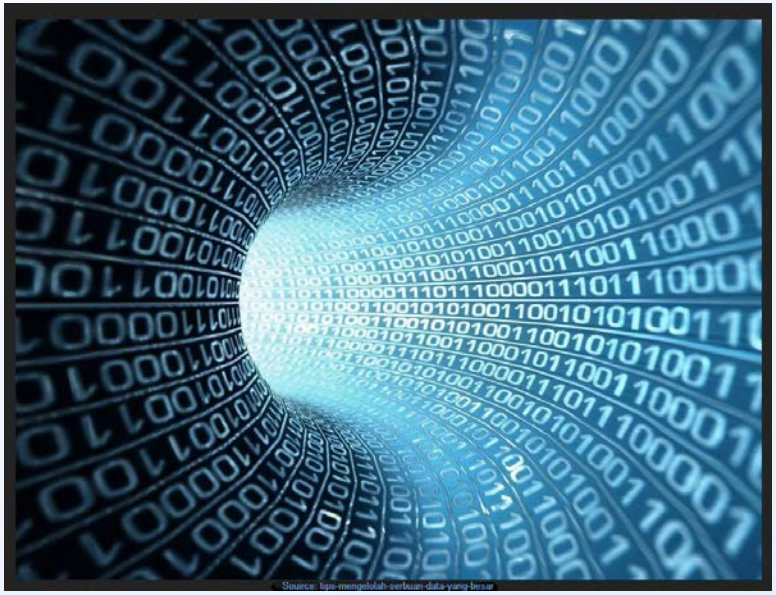
	Big Data	Relational
Analysis Type	Exploratory analysis to uncover value in the data	Operational analysis of what was discovered
Data Granularity	Store HUGE amounts of highly granular data	Store transform (sometimes) aggregated data
Timeframe	Data flows in BIG Data → "real-time" monitoring	Long term trending analysis

New Approaches To Big Data Processing And Analytics

ACTUATE
The BIRT Company

Traditional tools and technologies are straining

- New approaches to data processing
  - Commodity hardware to scale
  - Parallel processing techniques
  - Non-relational data storage capabilities
  - Unstructured, semi-structured data



Source: Wikibon Community

New Approaches To Big Data Processing And Analytics

ACTUATE
The BIRT Company

- Hadoop Approach
  - *Data broken into "parts"*
  - *Loaded into file system*
  - *Multiple nodes*
  - *MapReduce*
  - *Batch-style historical analysis*
- NoSQL
  - *Cassandra, MongoDB, CouchDB, HBase**
  - *Discrete data stored among large volumes*
  - *Higher performance than relational data sources*
  - *Categorized as Columnar, Document, Key-Value, Graph*
- Massively Parallel Analytic Databases
  - *Quickly ingest mostly structured data*
  - *Minimal data modeling*
  - *Scale to petabytes of data*
  - *Near real-time results to complex SQL*

Source: Wikibon Community

Big Data Growth Drivers

- Increased awareness of the Big Data benefits
  - Not just web, financial services, pharmaceuticals, retail
- Increased maturity of Big Data software
  - Data stores, analytical engines
- Increased availability of professional services
  - Supporting business use cases
- Increased investment in infrastructure
  - Google, Facebook, Amazon

Source: Wikibon Community

IDC 2013 Big Data Predictions

- The “Digital Universe” will expand to over 4 zettabytes… Over 50% growth from 2012
- The Big Data focus will shift “up the stack”, toward analytics and discovery, and analytic applications
- Spending will reach $10 billion in 2013, over $20 billion by 2016

Source: IDC, IDC Predictions 2013 presentation

Eclipse BIRT Survey – Oct/Nov 2012

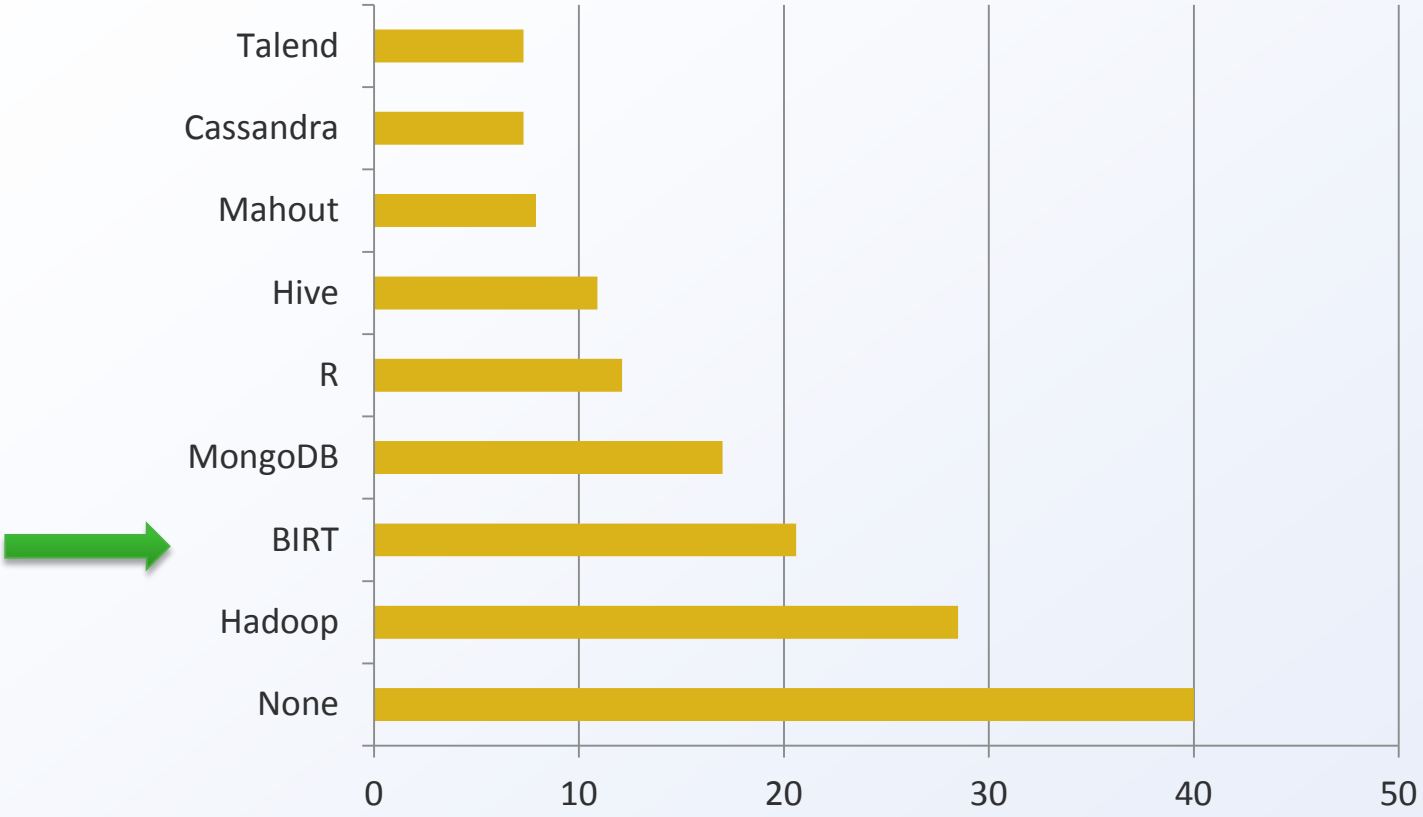
Big Data or Little Data - How Do You Display Yours?
The Eclipse Foundation would like to better understand how developers are using Eclipse with big data and reporting projects.

- We ran this survey to get the pulse of what technologies where in demand related to Eclipse/BIRT technologies.
- Eclipse Promoted the Survey.
- 60% of 518 responders claimed to be big data users

Eclipse BIRT Survey - Technology Choices

ACTUATE
The BIRT Company

What Big Data technologies are you using with Eclipse?

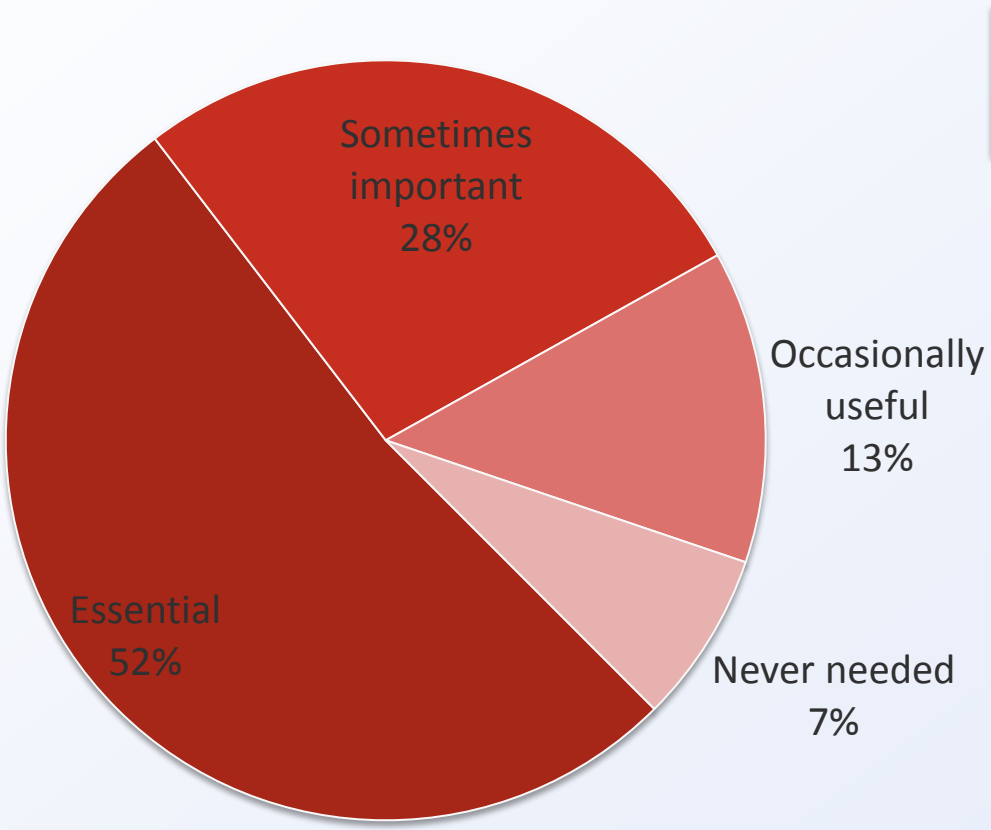


Note: Responders could choose more than one option

Eclipse BIRT Survey - Data Visualization

ACTUATE
The BIRT Company

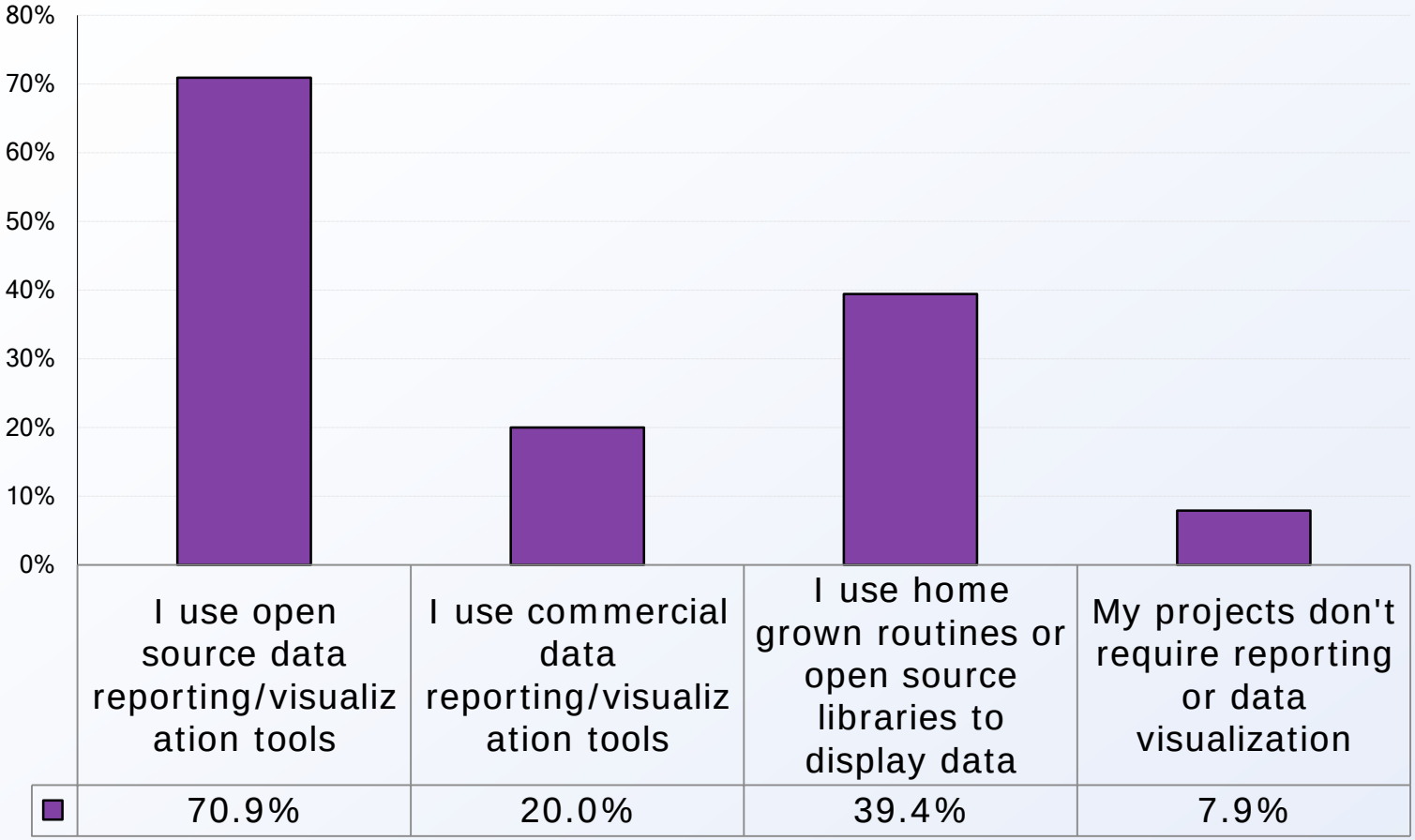
How important is Data Visualisation/Reporting to your project?



80%

Eclipse BIRT Survey - Data Visualization

How do you create/use data display tools in development?



	I use open source data reporting/visualization tools	I use commercial data reporting/visualization tools	I use home grown routines or open source libraries to display data	My projects don't require reporting or data visualization
■	70.9%	20.0%	39.4%	7.9%

Note: Responders could choose more than one option

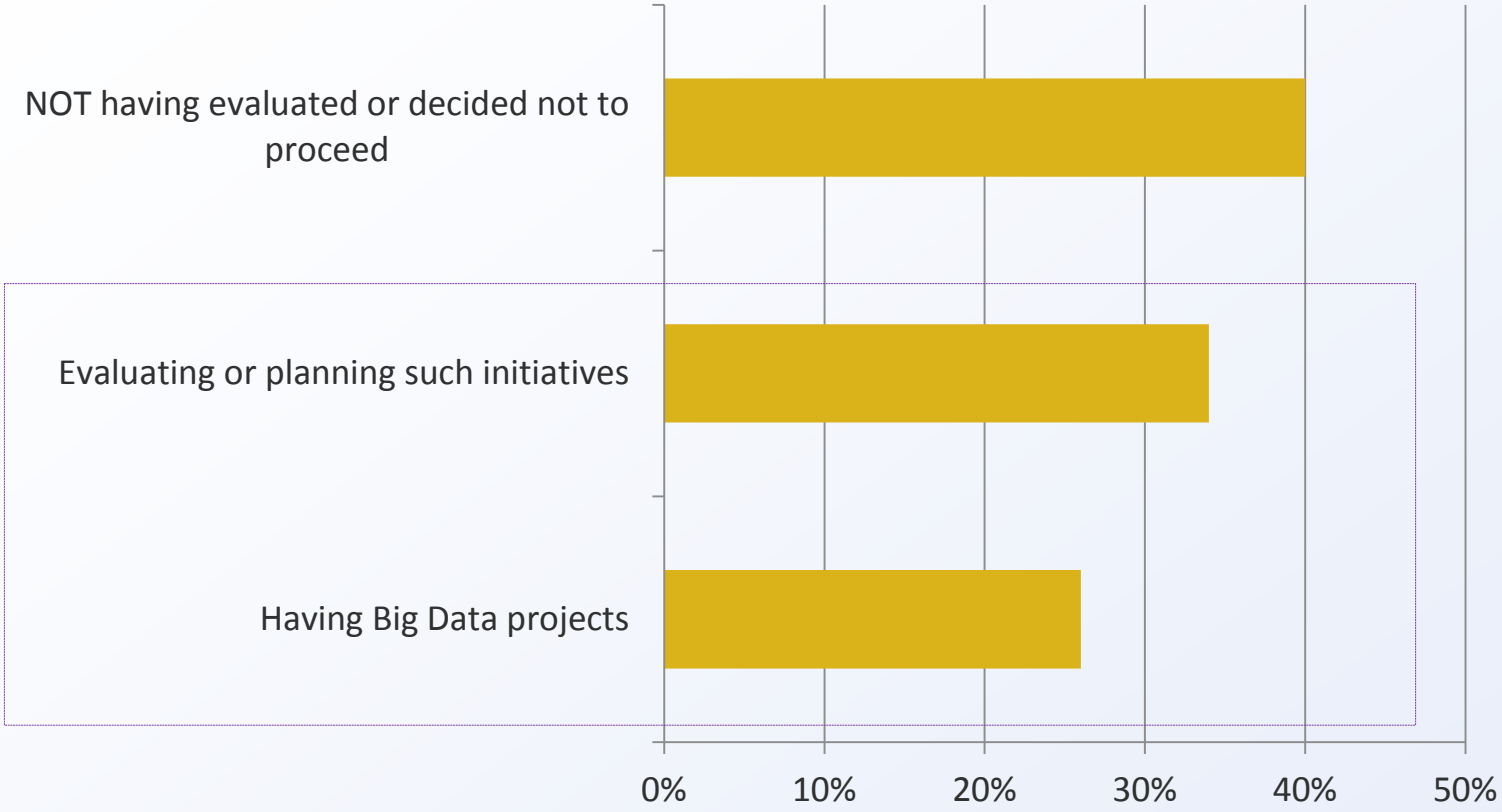
Independent Big Data Survey – Sept/Oct 2012

Goals:

- How many large firms (>$1B) are conducting Big Data projects
- What are such companies doing with their Big Data projects
- What are the expected benefits for those Big Data initiatives
- What are the inhibitors

- King Research received 516 surveys
  - 316 completed and 200 partially completed surveys
- Completed surveys were the primary source of analysis
- 32% of those who completed survey (98 respondents) work at companies with revenue of $1B or more

Independent Big Data Survey – Key Findings



60%

"Not enough staff with expertise" and "Expected cost of Big Data initiatives" are the major inhibitors

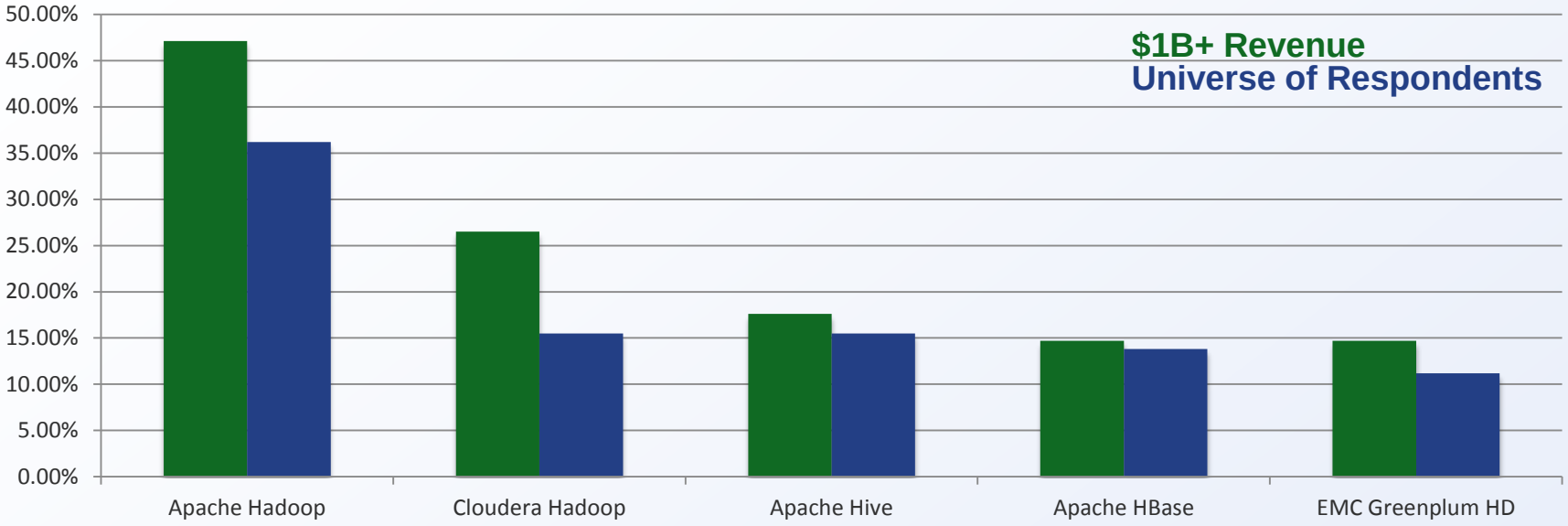
Independent Big Data Survey – Key Findings

Big Data Initiatives

Major Benefits *Expected*	Major Benefits *Realized*
Make better decisions, faster	Gain competitive advantage
Gain competitive advantage	Improve customer targeting
Improve efficiency	Make better decisions, faster
Improve customer targeting	Improve efficiency

Independent Big Data Survey – Big Data Technologies

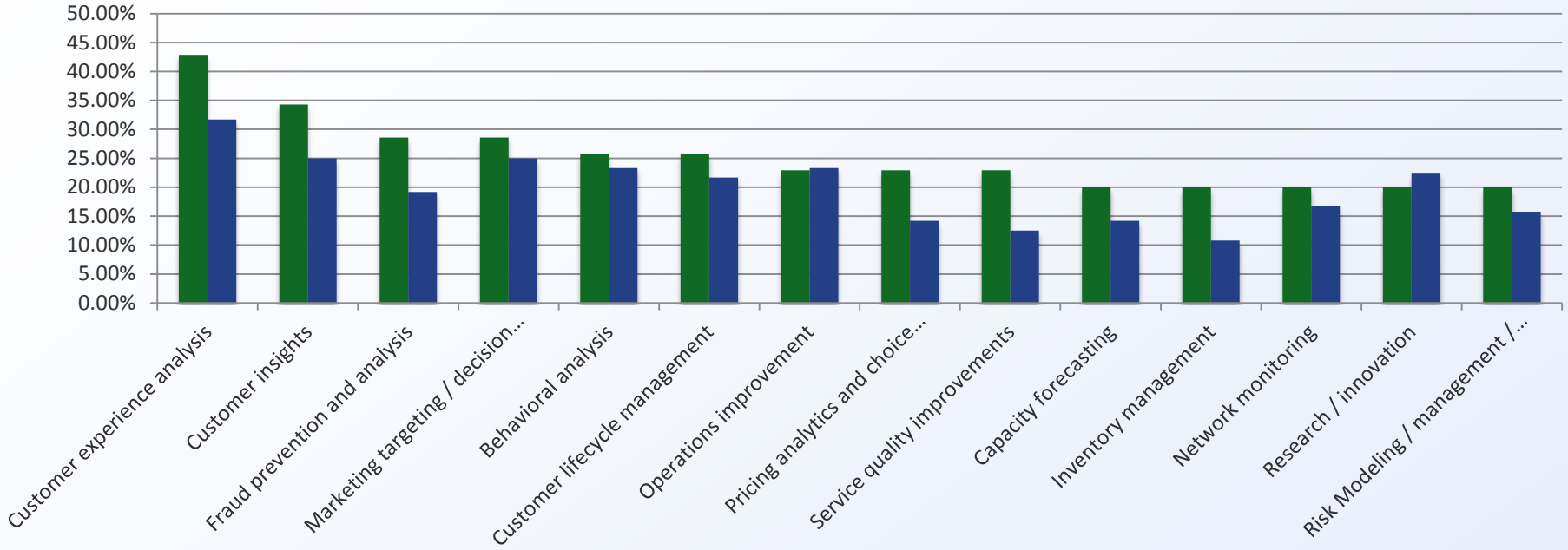
What Big Data technologies do you plan to use? (eval/planning)



- We asked about their planned use of 15 technologies, and the top 5, in descending order of frequency of mention are displayed above
- Other technologies planned for use at $1B+ organizations include: Apache Cassandra, 12%; Hortonworks Hadoop, 12%; Amazon DynamoDB, 9%; Apache CouchDB, 9%; VoltDB, 9%; HyperTable, 6%; **MongoDB**, **3%;** Datastax Cassandra, 3%

Independent Big Data Survey – Application Types

What are likely to be your Big Data applications? (responses from those who are evaluating or planning Big Data implementations)



- Our survey listed 23 frequently reported Big Data applications and when asked which of these they have evaluated or planned to use, they indicated an average 4.5 apps each.
- The most important application types are about customer analysis

Top Big Data Challenges

- **Data integration**
  - Top challenge
  - Integrating disparate data, different sources, different formats is difficult
- **Getting started with the right project**
  - Building the right team
  - Determine the top business problem
- **Architecting a big data system.**
  - High volume, high frequency data
  - Build unified information architecture
- **Lack of skills or staff**
  - Some hire externally / university hires.
  - Others try to re-train from within.
  - Cross pollinate skills from another part of the organization
  - Build centers of excellence that help with the training

Source:TDWI

Actuate Launches the BIRT Project

ACTUATE
The BIRT Company

Actuate proposed and started

BIRT

… a top-level Eclipse project

Professional open source

Primary development resources funded by Actuate

Contributions from many sources

IBM, Innovent Solutions and community

Actuate Joins Eclipse Foundation as Strategic Developer and Board Member

Adds BI and Reporting as Open Source Project

AUGUST

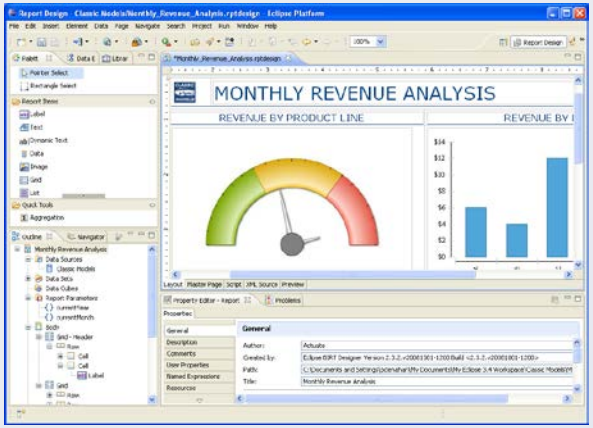
2004

What is BIRT?

ACTUATE
The BIRT Company

A New Generation of Data Visualization Technology

- Makes all data-driven content development easy
- Modern, web-page design metaphor
- Open and standards-based
- Flexible with rich programmatic control
- Full support for libraries and reuse
- Foundation for a range of solutions



Simplicity

that makes simple layouts easy

BIRT

Power

to create very complex layouts

BIRT Release History supports Big Data

September 2004		BIRT Project proposal accepted, and project launched
June 2005	1.0	Eclipse Report Designer, Report Engine, Chart Engine
December 2005	2.0	Support for a wide variety of common layouts
June 2006	2.1	Advanced parameters, **ability to join data sets,** …
June 2007	2.2	Dynamic crosstab support, web services data source, …
June 2008	2.3	JavaScript Debugger, BiDi Support, Charts in Crosstabs, …
June 2009	2.5	Page aggregates, Multiple drill-downs in Charts, …
June 2010	2.6	New charts, more chart control, developer productivity, …
June 2011	3.7	POJO Runtime, **Hive/Hadoop,** Open Office emitters…
June 2012	4.2	Maven Support, Excel Data Source, Relative Time Periods…
June 2013	4.3	POJO Data Source, **MongoDB/Cassandra support**, client JS

BIRT Data Access

- BIRT Offers many ways to get data
  - Standard Data Sources
    - *Flat File (CSV, TSV, SSV, PSV)*
    - *Hive Data Source (Hadoop)*
    - *Cassandra Scripted Data Source*
    - *MongoDB Data Source*
    - *JDBC Textual or Graphical*
    - *Web Service - XPath syntax*
    - *XML - XPath syntax*
    - *XLS/XLSX*
  - Scripted Data Source Written in Java or JavaScript
  - Open Data Access (ODA) DTP Project

Community Contributions
GoogleDocs
XML/A
Casandra
REST
MongoDB
Multi-Flat File
GitHub
Twitter JSON Search
Dropbox usage
YQL
Google Analytics
LinkedIn
Facebook FQL

BIRT Formatting Capabilities

- Listings, cross-tab, dashboard, pixel-perfect, charts
- Grouping, advanced aggregations, sub-totals, calculations
- Multi-section and sub-reports
- Conditional sections and logic
- Full programmatic control and scripting
- Embedded images
- Dynamic hyperlinking
- And much more!

Inventory Turn Trend

Jan 01, 08 to Dec 31, 08

Report Description : This report shows the inventory issue turn rate for each Business Unit and Facility by month.

Inventory Turn Trend by Month for Business Unit

Inventory Turn Trend by Month for Facilities of BU: Biocides

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
MAG	0	0	0	2	0	6	8	0	0	0	0	0	17
MFC	777	270	94	186	6095	3077	2003	35690	21033	16489	265	1878	87857
MHP	69	365	5076	36	122	5099	9214	4414	3950	42583	0	717	71644
MSF	35838	52	620	1412	40	345	926	489	32943	9155	4373	0	86194
MWG	287	460	276	19	338	319	124	664	223	210	0	0	2921
Total	36971	1148	6066	1655	6596	8847	12275	41256	58149	68436	4638	2595	248632

Classic Models, Inc
2207 Bridgepointe Parkway
San Mateo, CA 94404

Sales by Customer

Customer: **Atelier graphique - 103** Payment: **$22,314.36**

Order Number:	10123		
Order Date:	May 20, 2003 12:00 AM	Shipped Date:	May 22, 2003 12:00 AM

Product Code	Product Name	Orders	Price Each	Price
S18_1589	1965 Aston Martin DB5	26	$120.71	$3,138.46
S18_2870	1999 Indy 500 Monte Carlo SS	46	$114.84	$5,282.64
S18_3685	1948 Porsche Type 356 Roadster	34	$117.26	$3,986.84
S24_1628	1966 Shelby Cobra 427 S/C	50	$43.27	$2,163.50

Order Number:	10298		
Order Date:	Sep 27, 2004 12:00 AM	Shipped Date:	Oct 1, 2004 12:00 AM

Product Code	Product Name	Orders	Price Each	Price
S10_2016	1996 Moto Guzzi 1100i	39	$105.86	$4,128.54
S18_2625	1936 Harley Davidson El Knucklehead	32	$60.57	$1,938.24

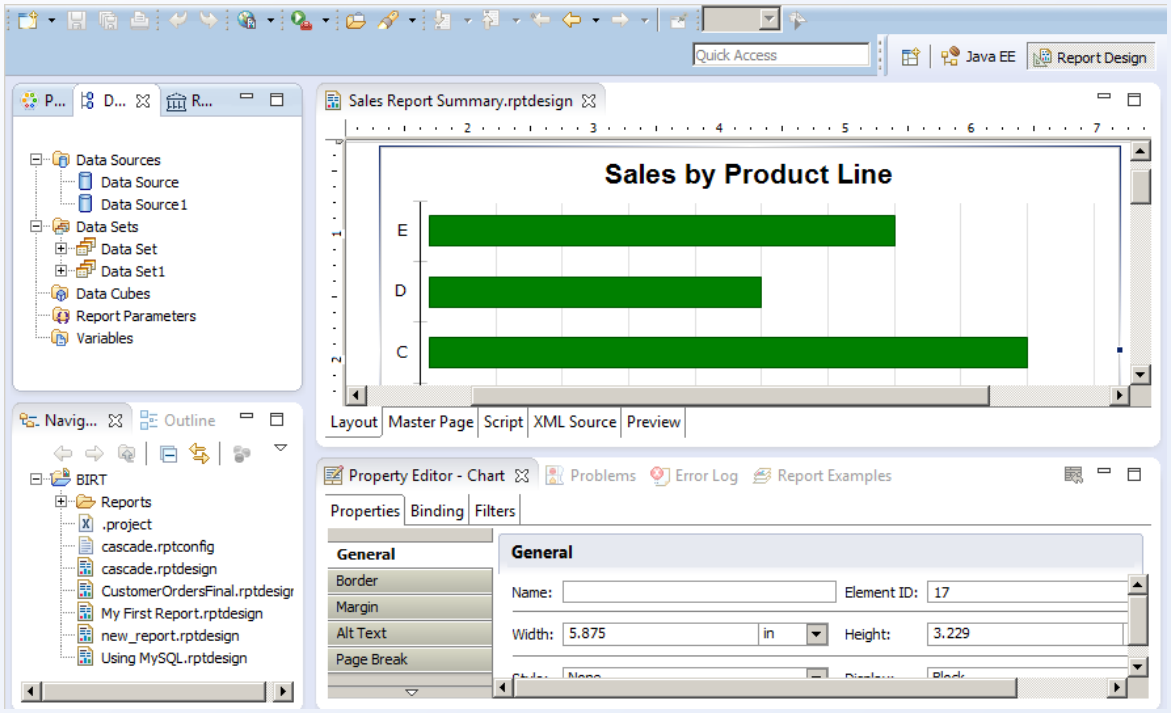
Order Number:	10345		
Order Date:	Nov 25, 2004 12:00 AM	Shipped Date:	Nov 26, 2004 12:00 AM

Product Code	Product Name	Orders	Price Each	Price
S24_2022	1938 Cadillac V-16 Presidential Limousine	43	$38.98	$1,676.14

Total Price: **$22,314.36**

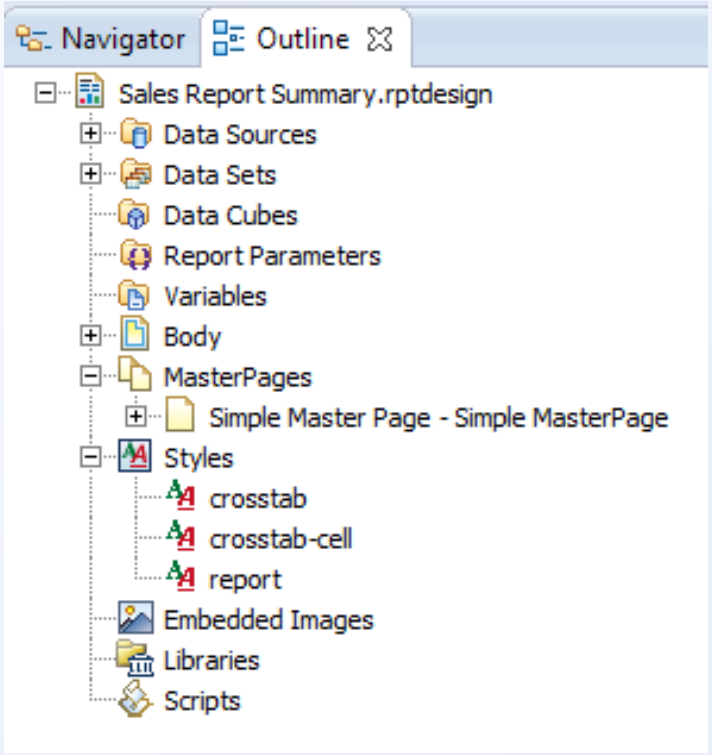
BIRT Productivity Aids

- Graphical layout and design
- Query and metadata editors
- Customizable cheat sheets and templates
- Formatting builder
- Grouping builder



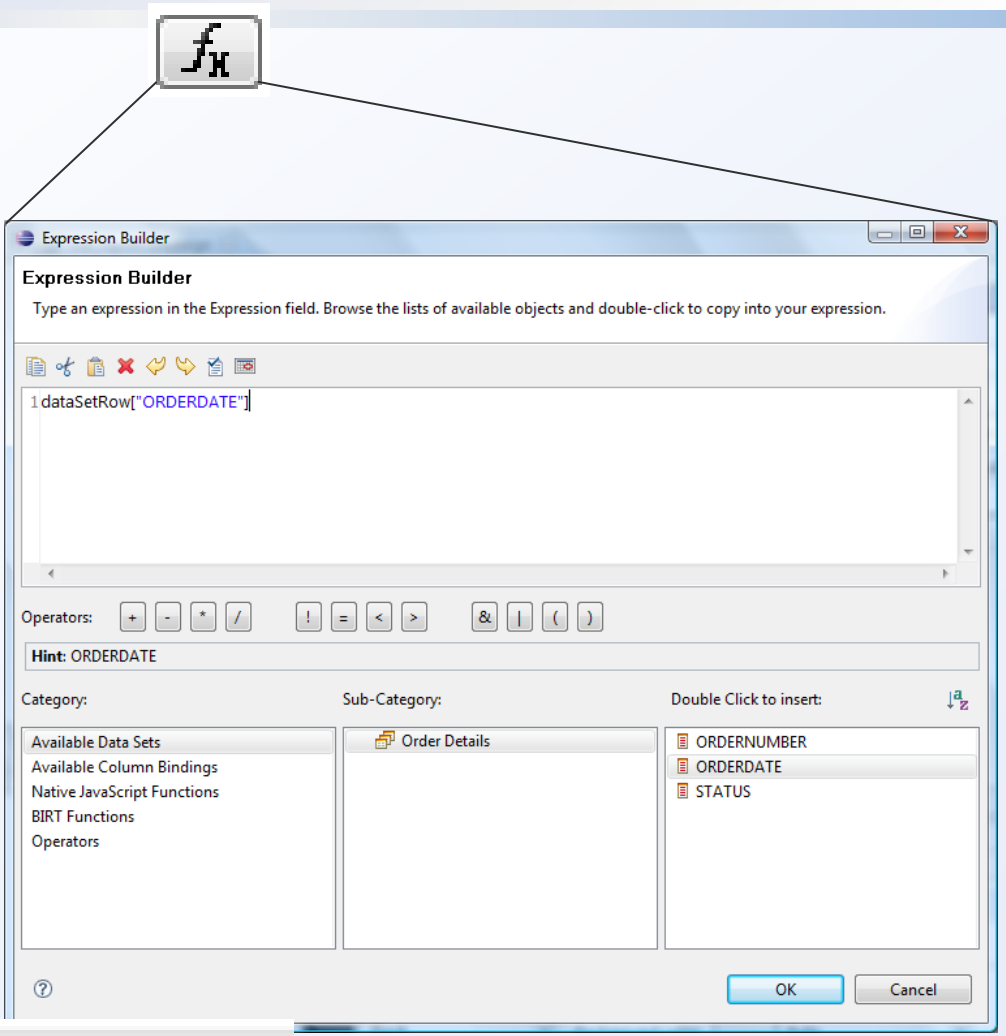
BIRT Re-use and Productivity

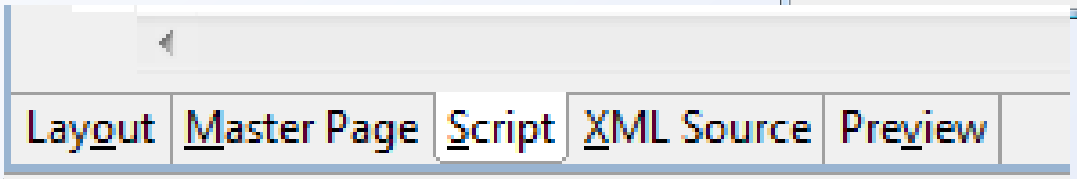
- Library support for publishing and sharing components
- Leverages common standards
  - SQL
  - Java
  - JavaScript
  - XML
- Cascading Style Sheets
- Integrated debugger



BIRT Expressions and Scripting

- Expressions are everywhere
  - Data Source\Set, Parameters
  - Report Items
    - *Data, Crosstab, Images*
  - Chart Elements
  - Visibility, TOC, Bookmarks
  - Events
- Helper Object References
  - Data Sets\Table Bindings
  - Native JavaScript Func's
  - BIRT Functions
  - Operators

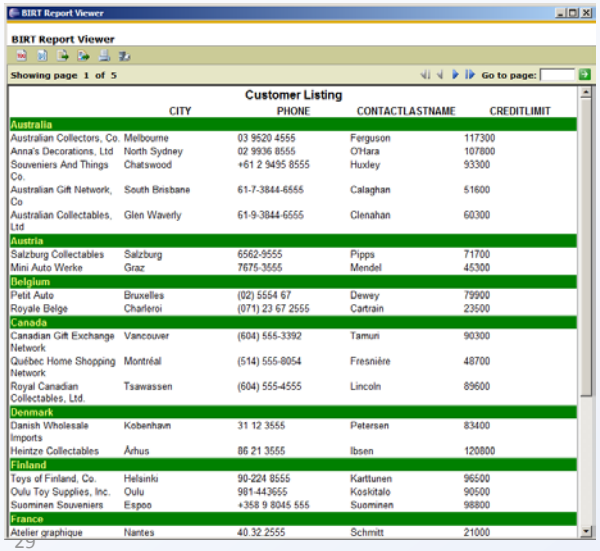


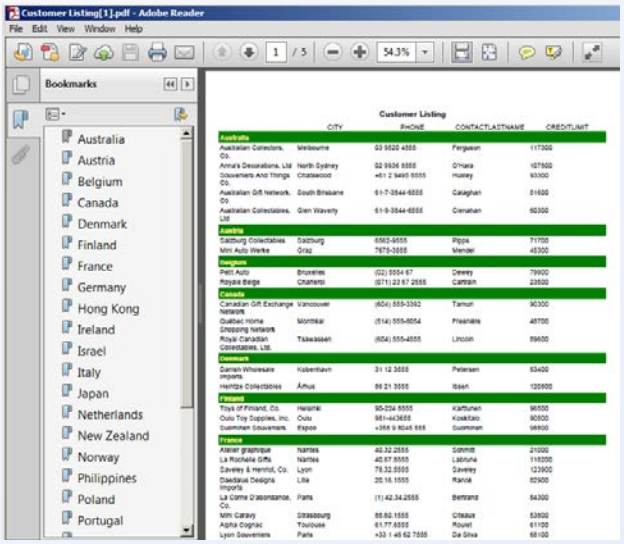


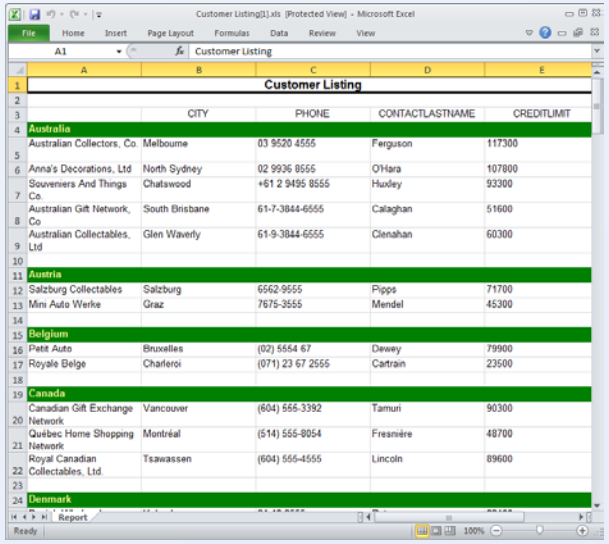
BIRT Output Formats

- HTML, PDF, MS Office, Open Office
- Custom emitters : CSV, Text, …
- Internationalization of labels and text
- Localization of date and numeric fields
- Support for Asian character sets and right-to-left display

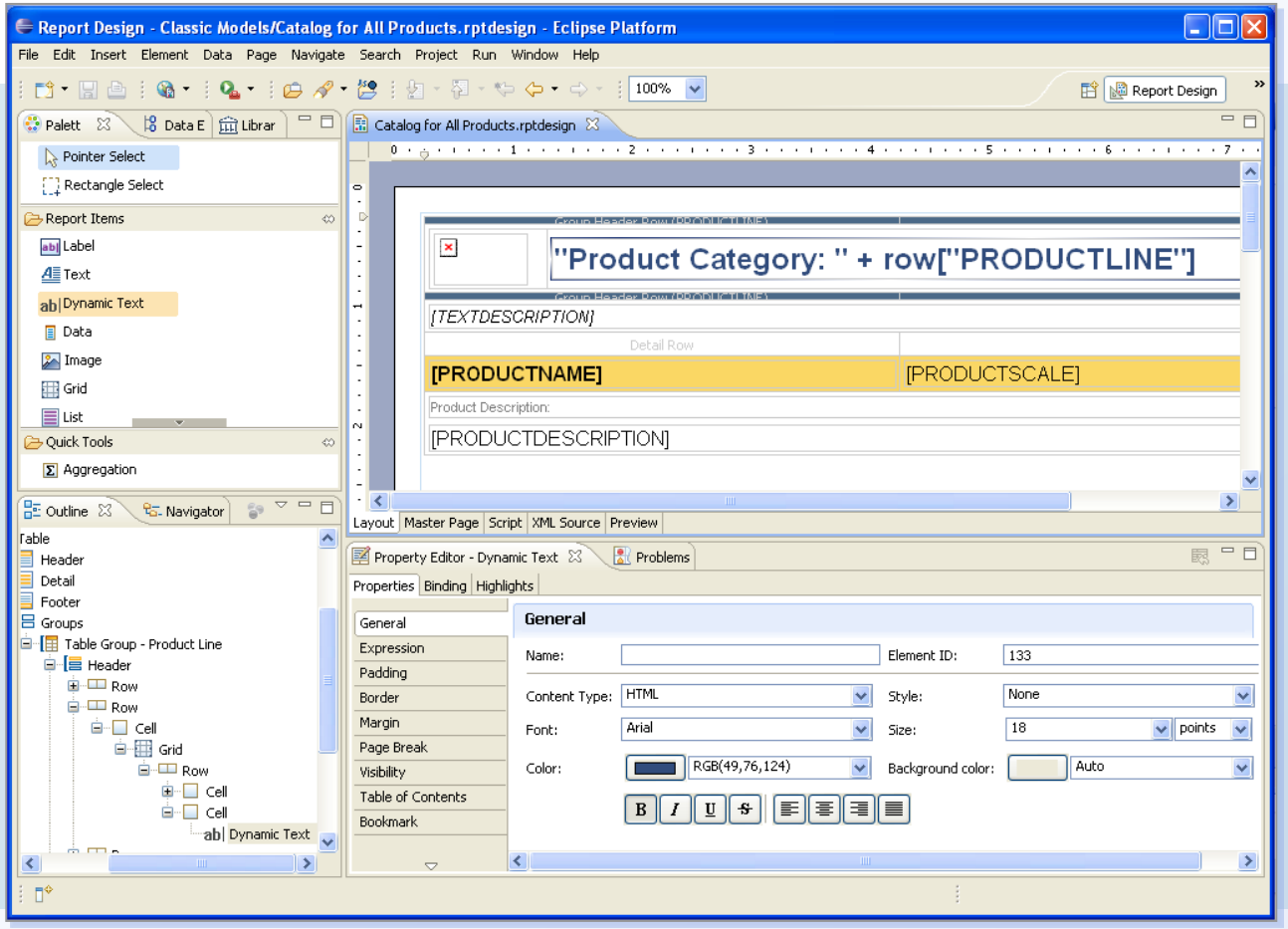
Web Viewer → **PDF** → **Excel**



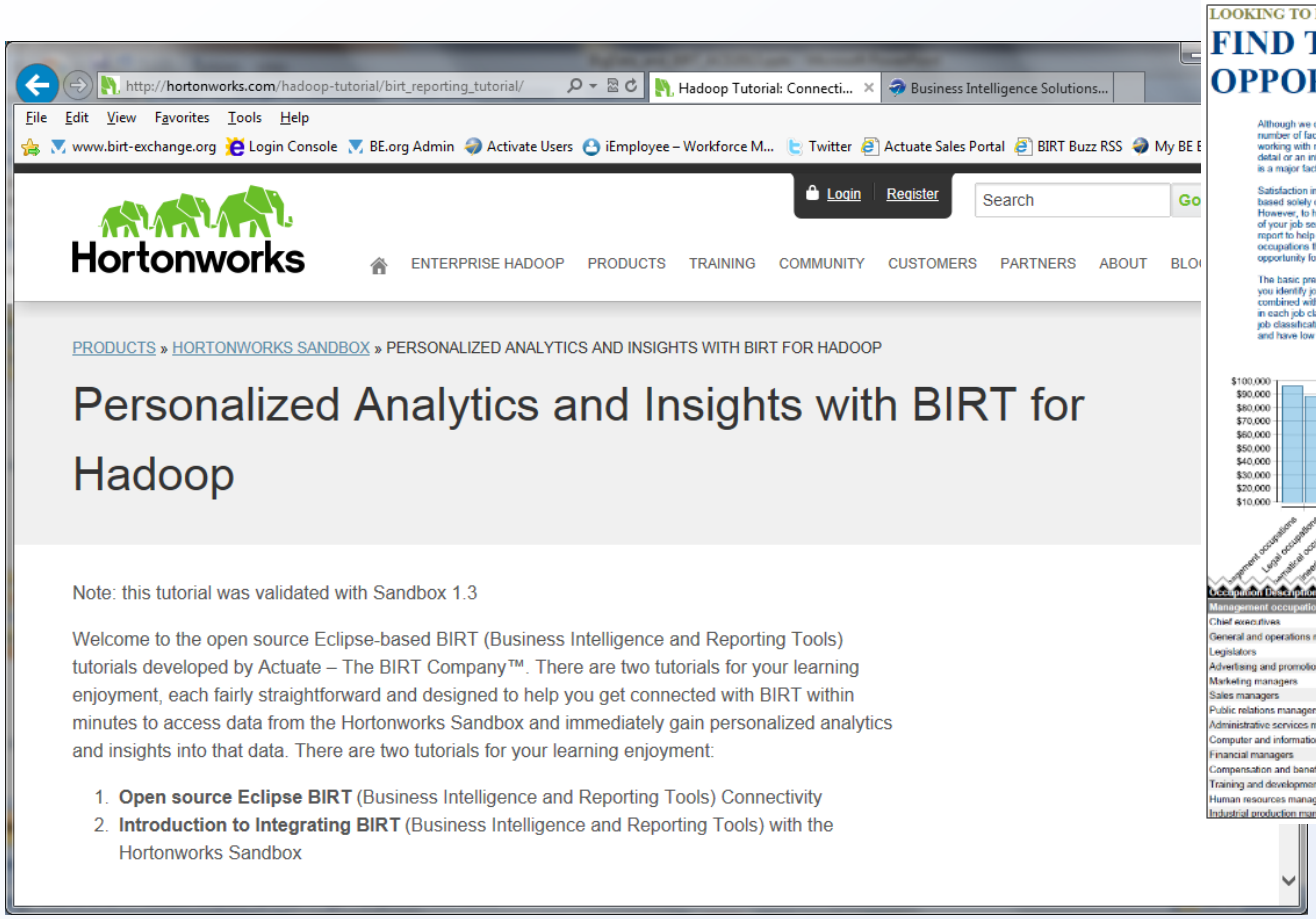


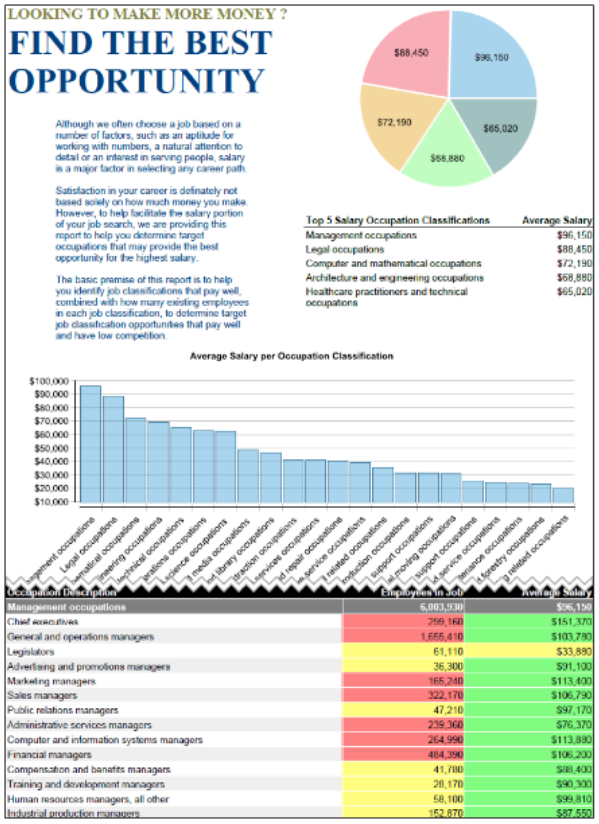


Live Demo – Connecting to Big Data



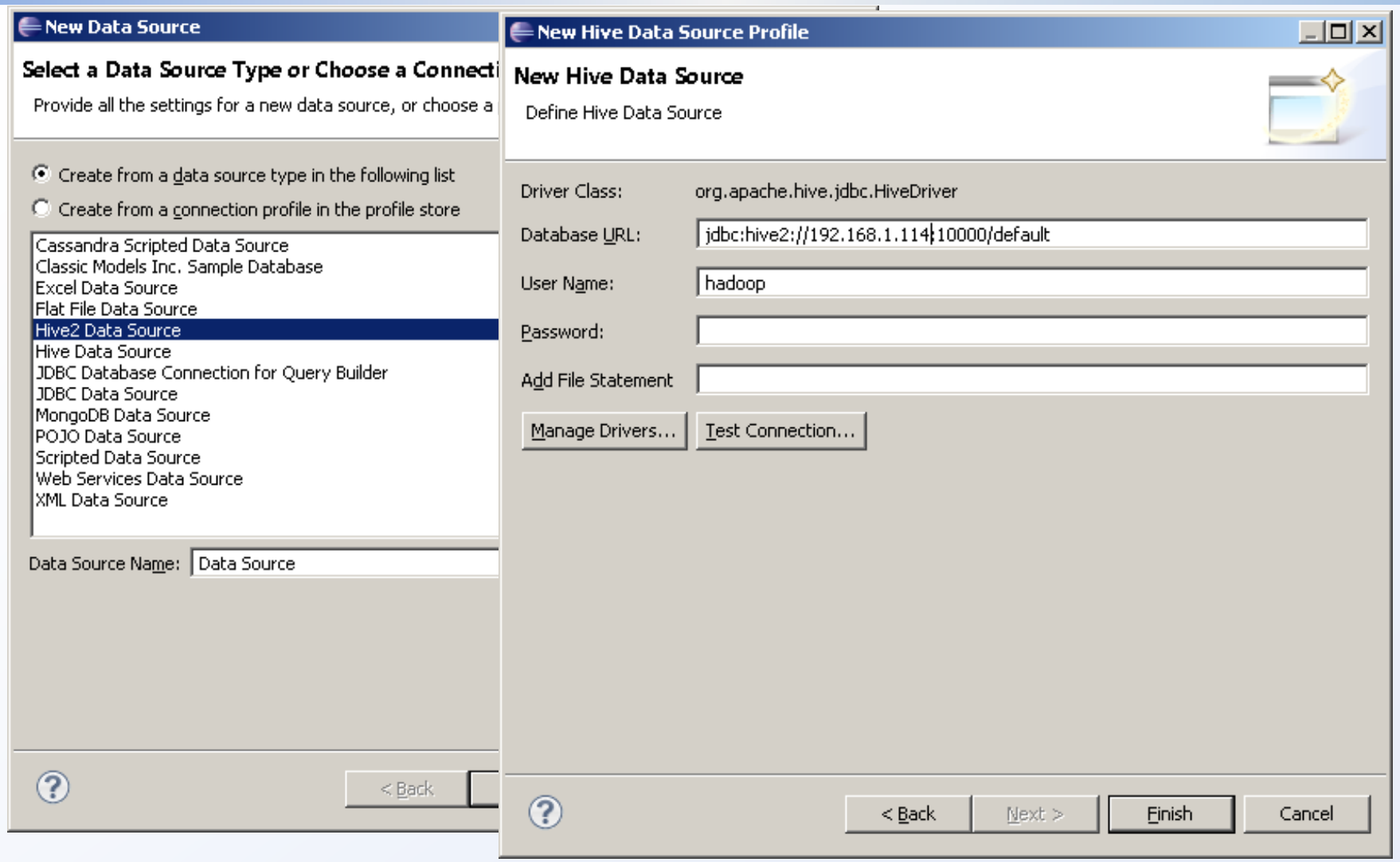
Connecting to Hadoop
Hortonworks Sandbox Tutorial



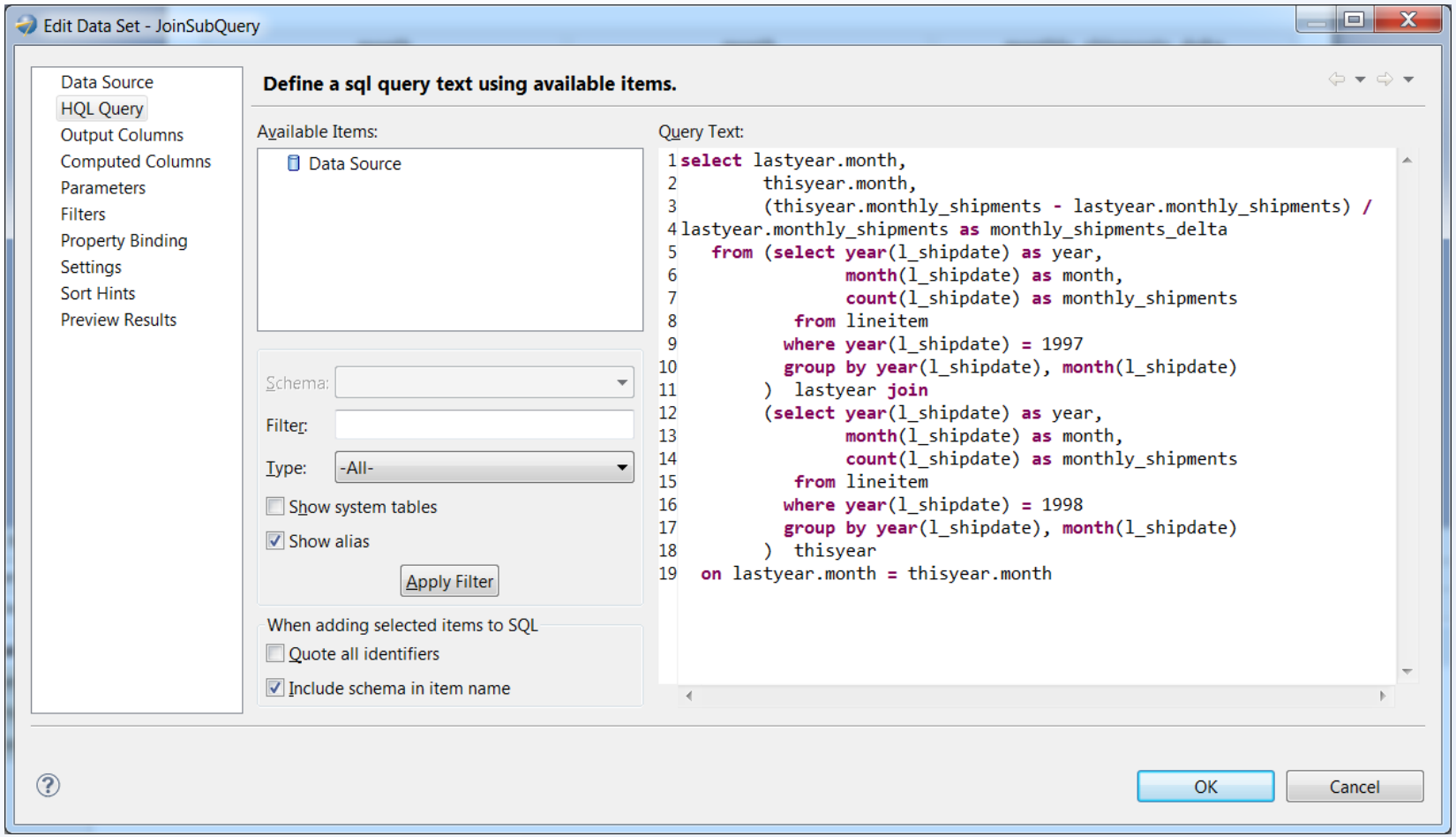


http://hortonworks.com/hadoop-tutorial/birt_reporting_tutorial/

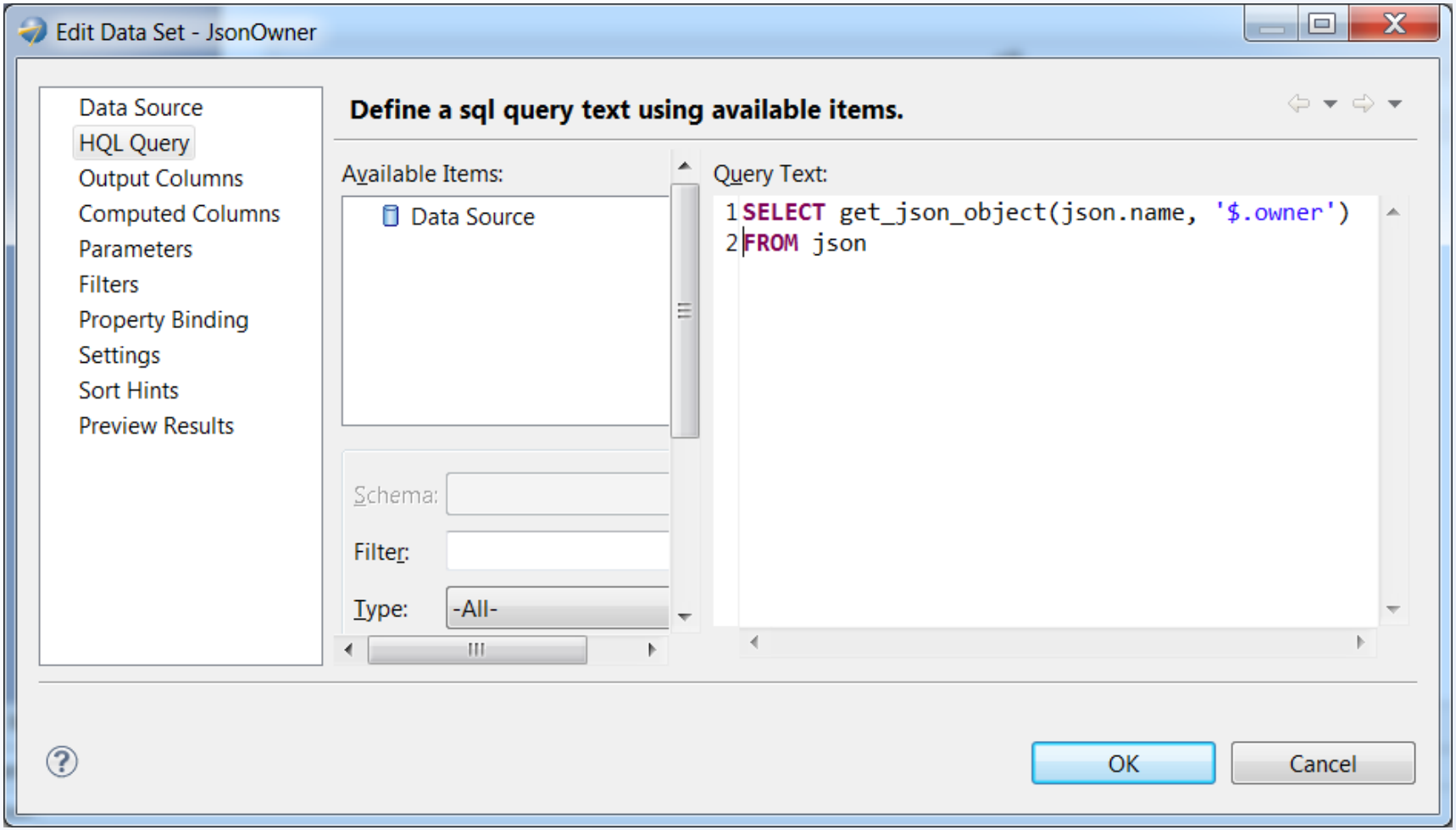
Connecting to Hadoop



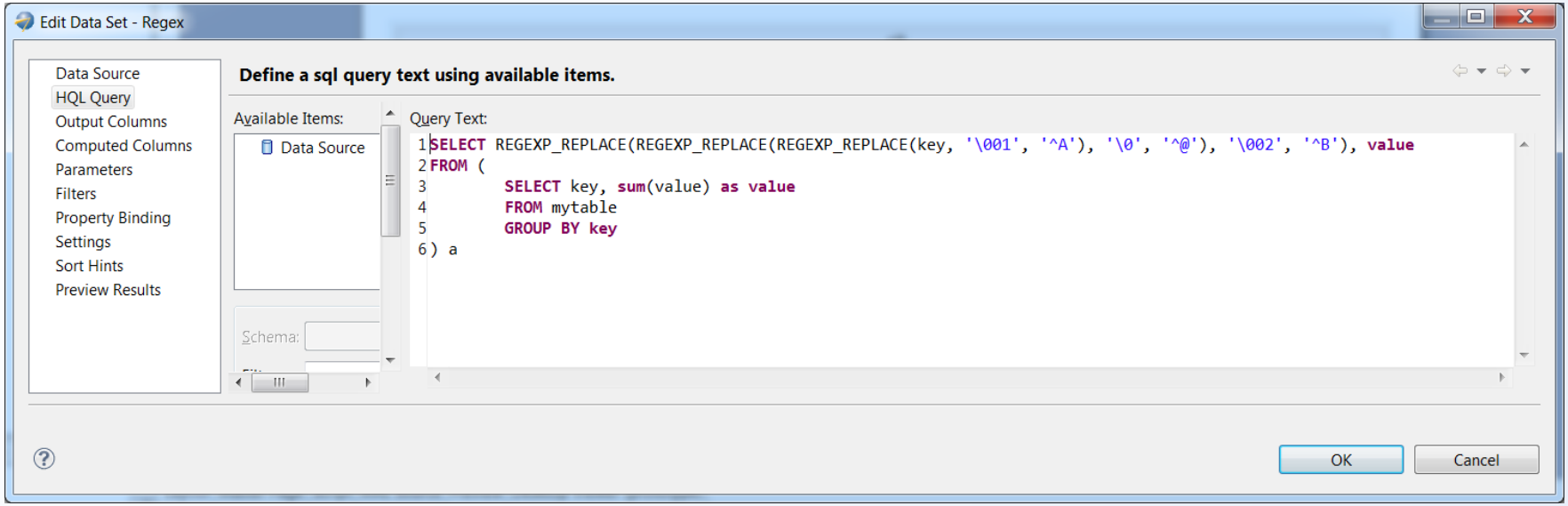
Hive JDBC – HQL Sub Query Example



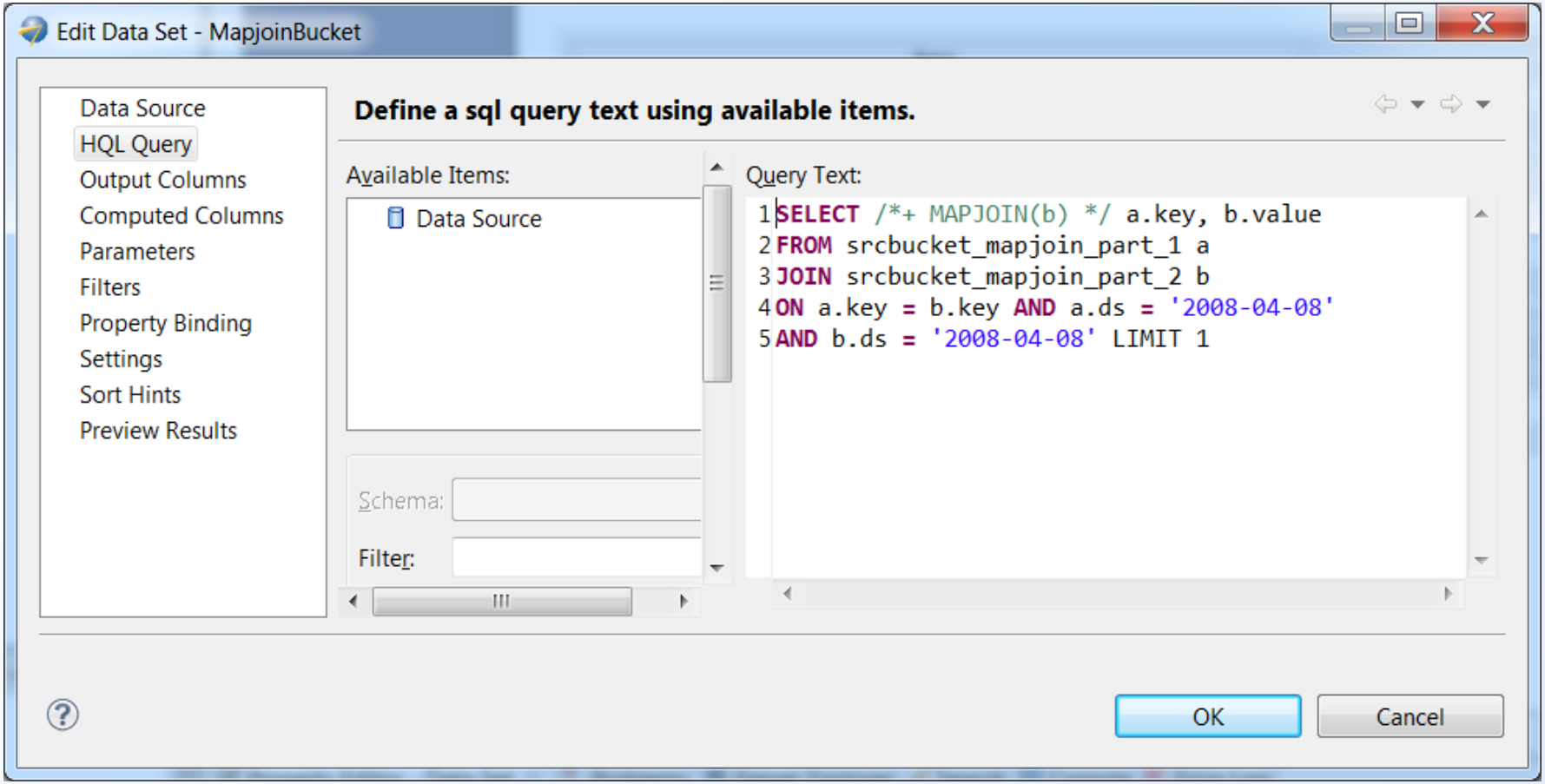
Hive JDBC – get_json_object UDF



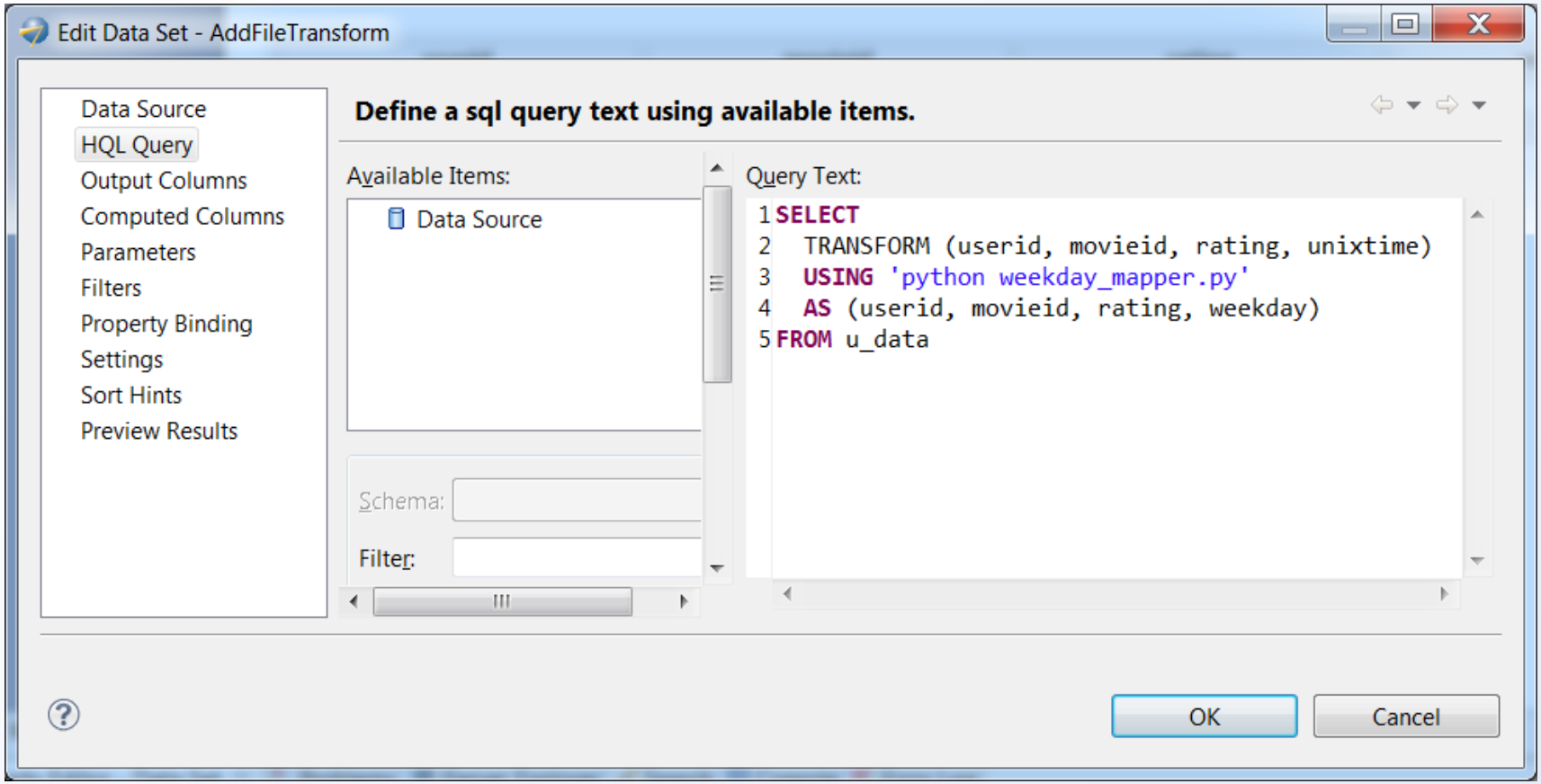
Hive JDBC – RegExP Example



Hive JDBC – HQL Hints example



Hive JDBC – Transform Example



Cassandra – Scripted Data Set Example

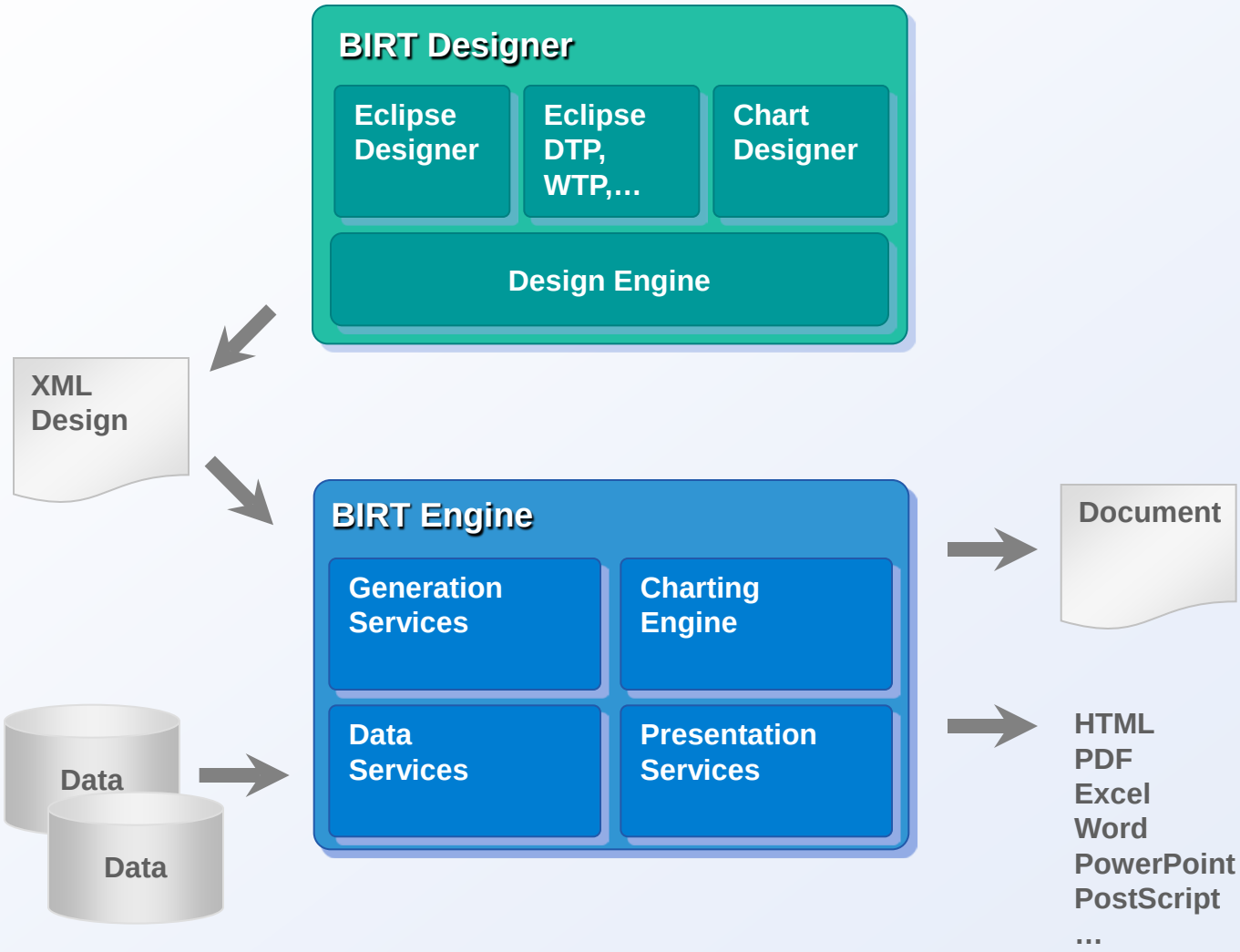
Script: open — open

```
importPackage(Packages.java.util);
importPackage(Packages.me.prettyprint.cassandra.serializers);
importPackage(Packages.me.prettyprint.cassandra.service);
importPackage(Packages.me.prettyprint.hector.api);
importPackage(Packages.me.prettyprint.hector.api.beans);
importPackage(Packages.me.prettyprint.hector.api.factory);
importPackage(Packages.me.prettyprint.hector.api.query);
importPackage(Packages.me.prettyprint.cassandra.model);

var cluster = HFactory.getOrCreateCluster("Test Cluster",new CassandraHostConfigurator("localhost:9160"));
var keyspace = HFactory.createKeyspace("users", cluster);



var cqlQuery = new CqlQuery(keyspace, StringSerializer.get(), StringSerializer.get(), StringSerializer.get());
cqlQuery.setQuery("select * from User");
var resultCQL = cqlQuery.execute();
rowsIterator = resultCQL.get().iterator();


```

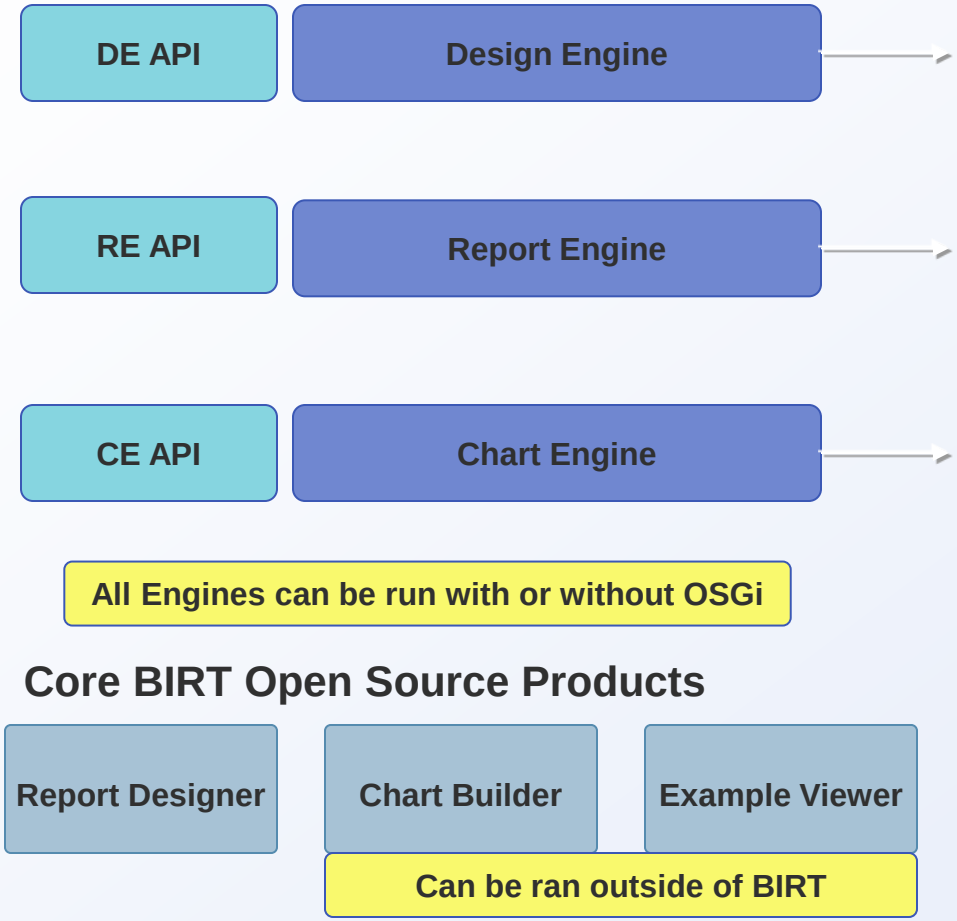
Script: fetch — open

```
if (rowsIterator.hasNext()) {
     var myrow = rowsIterator.next();
     var cols = myrow.getColumnSlice().getColumns();
     for( ii=0; ii < cols.size(); ii++ ){
        row[cols.get(ii).getName()] = cols.get(ii).getValue();
     }
     return true;
}else{
return false;
}
```

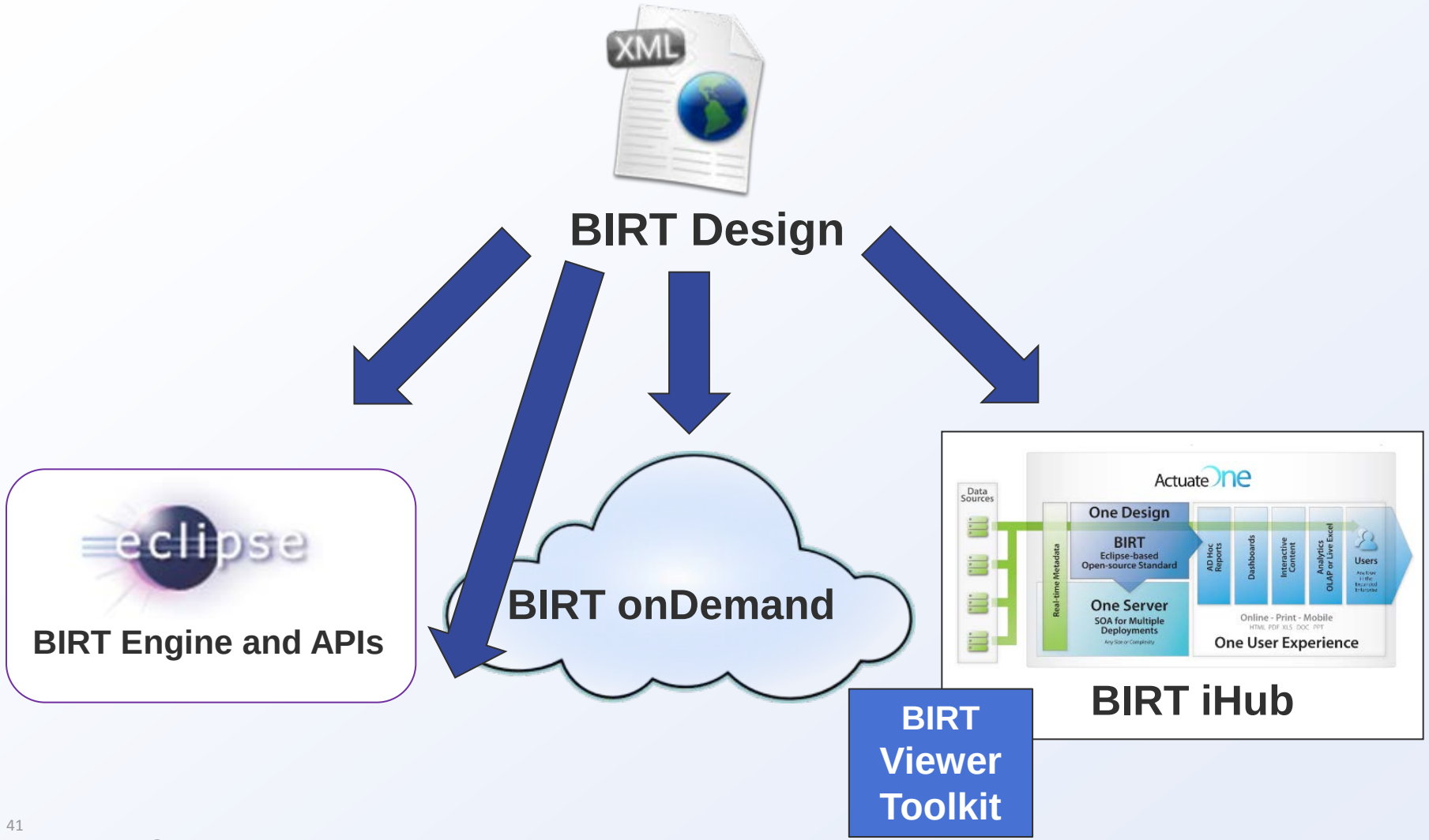
High-Level BIRT Architecture



High Level BIRT Architecture – APIs



Multiple Deployment Options



Report Engine API (REAPI)

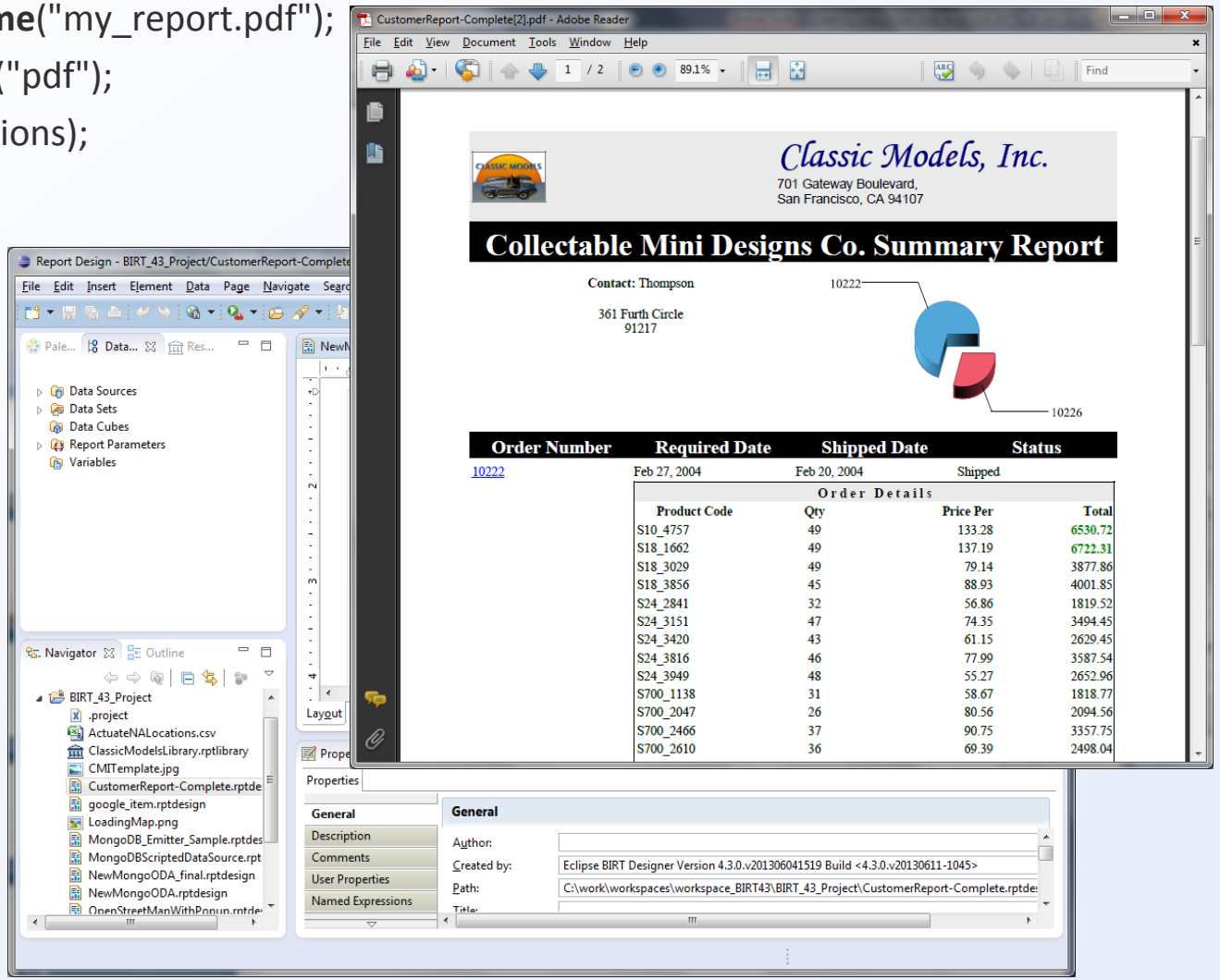
- Report Engine can be used in different ways
  - 3 tasks related to report execution and rendering
    - *RunTask*
    - *RenderTask*
    - *RunAndRenderTask*
  - Using RunTask and then RenderTask means multiple processes to generate and view a report.
  - RunAndRenderTask happens in single process so event firing order is different

Report Engine API (REAPI)

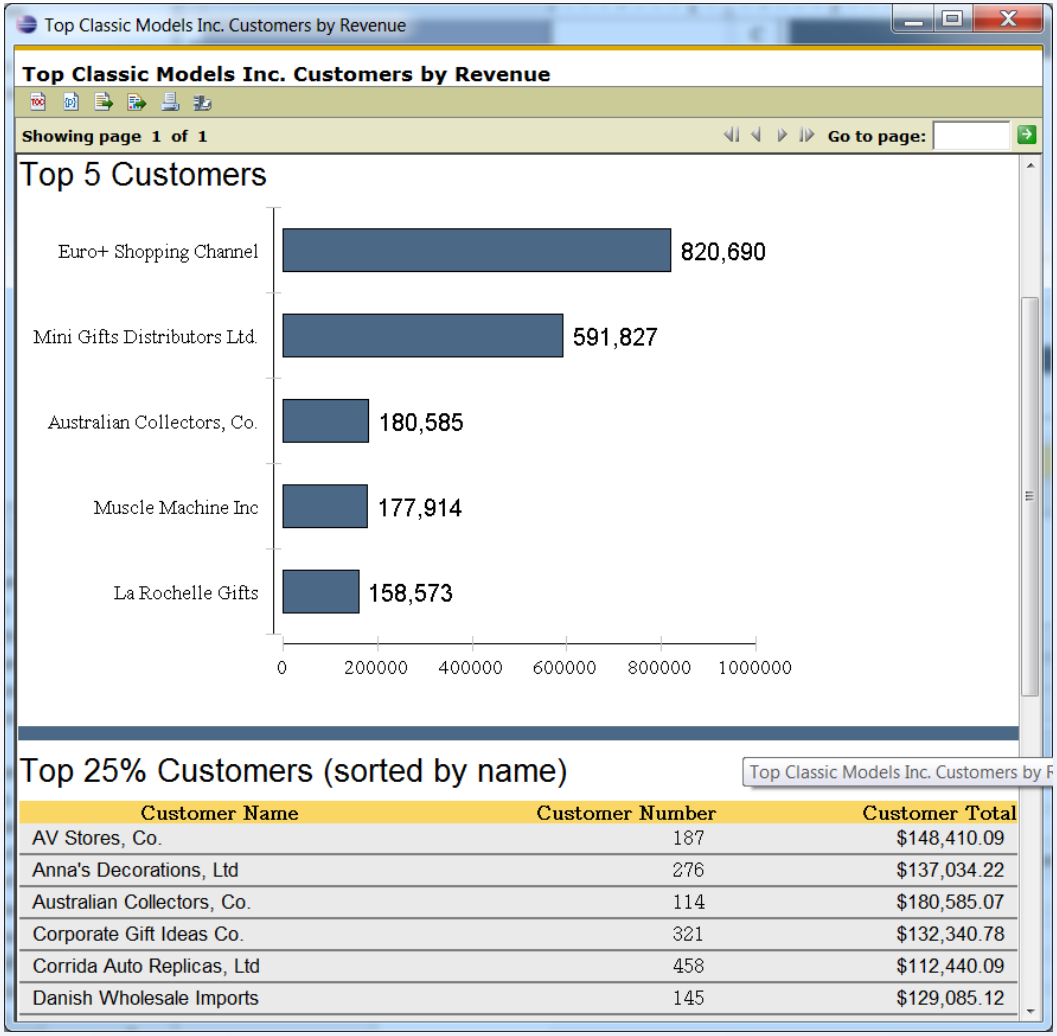
```
static void executeReport() throws EngineException
{
    IReportEngine engine=null;
    EngineConfig config = null;
try{
    config = new EngineConfig( );
    config.setLogConfig("C:\\BIRT_422\\logs", java.util.logging.Level.FINEST);
    Platform.startup( config );
    IReportEngineFactory factory = (IReportEngineFactory) Platform
 .   createFactoryObject( IReportEngineFactory.EXTENSION_REPORT_ENGINE_FACTORY );
    engine = factory.createReportEngine( config );
    IReportRunnable design = null;
    design = engine.openReportDesign("C:\\BIRT_43\\designs\\param.rptdesign");
    IRunAndRenderTask task = engine.createRunAndRenderTask(design);
    task.setParameterValue("ordParam", (new Integer(10101)));
    task.validateParameters();
 …
```

Report Engine API (REAPI)

```
    PDFRenderOption options = new PDFRenderOption();
    options.setOutputFileName("my_report.pdf");
    options.setOutputFormat("pdf");
    task.setRenderOption(options);
    task.run();
    task.close();
    engine.destroy();
}catch( Exception ex){
    ex.printStackTrace();
}
finally
{
    Platform.shutdown( );
}
```



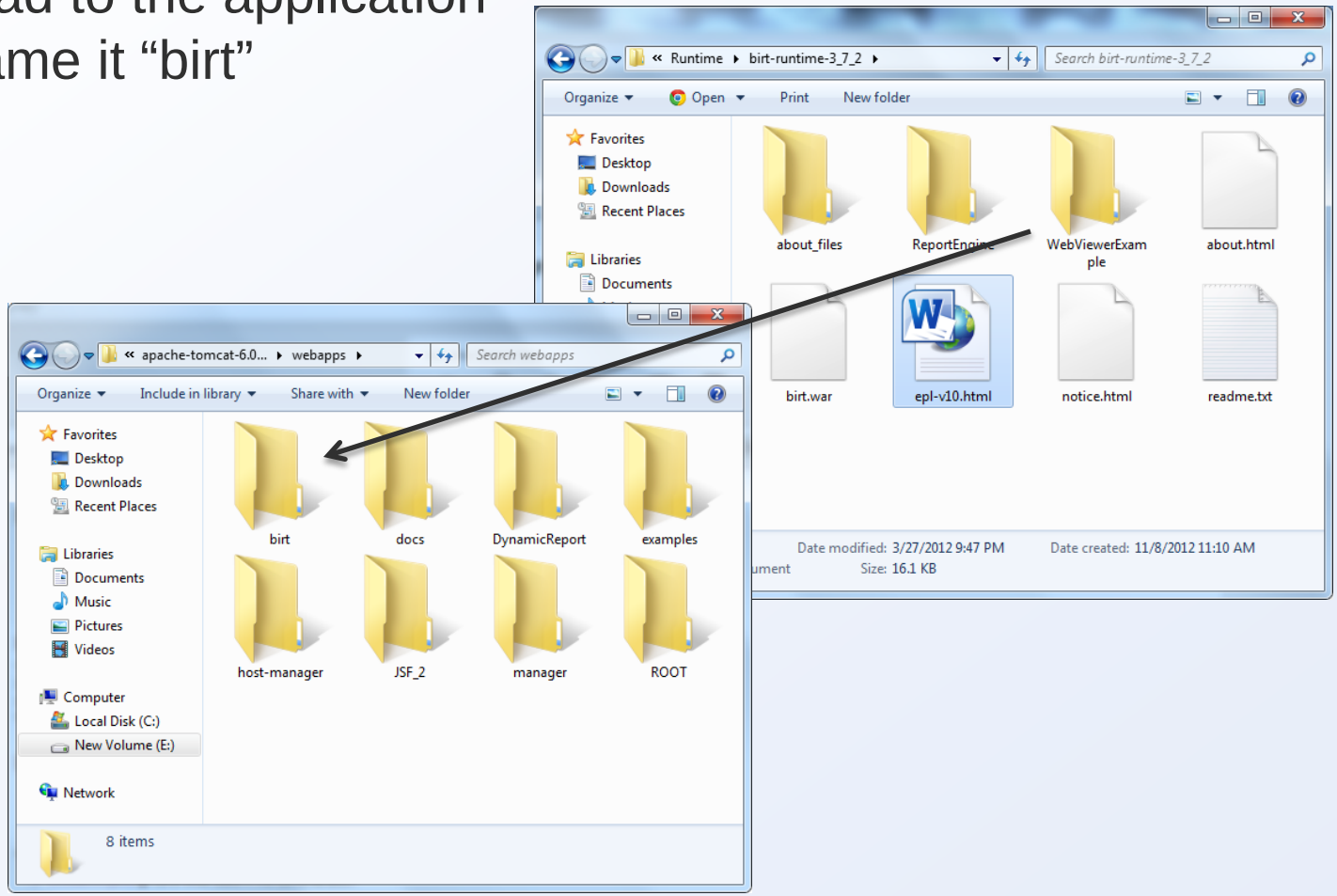
BIRT AJAX Based WebViewer Example



Customer Name	Customer Number	Customer Total
AV Stores, Co.	187	$148,410.09
Anna's Decorations, Ltd	276	$137,034.22
Australian Collectors, Co.	114	$180,585.07
Corporate Gift Ideas Co.	321	$132,340.78
Corrida Auto Replicas, Ltd	458	$112,440.09
Danish Wholesale Imports	145	$129,085.12

WebViewerExample in BIRT Runtime

Copy WebViewerExample from the runtime download to the application server and rename it “birt”



BIRT Web Viewer URLs

As a straight HTML output:

http://localhost:8080/birt/**run**?__report=reportName.rptdesign

With viewer toolbar, pagination, etc.:

http://localhost:8080/birt/**frameset**?__report=reportName.rptdesign

Other servlet mappings:

preview, document, output, extract

URL parameters include:

id	title	showtitle	toolbar	navigationbar	locale
format	report	document	parameterpage	pagerange	fittopage
svg	isnull	islocale	masterpage	resourceFolder	bookmark
rtl	page	overwrite			

BIRT Web Viewer URLs

Output to PDF:

http://localhost:8080/birt/frameset?__report=reportName.rptdesign&__**format=pdf**

To display a single report item as a reportlet:

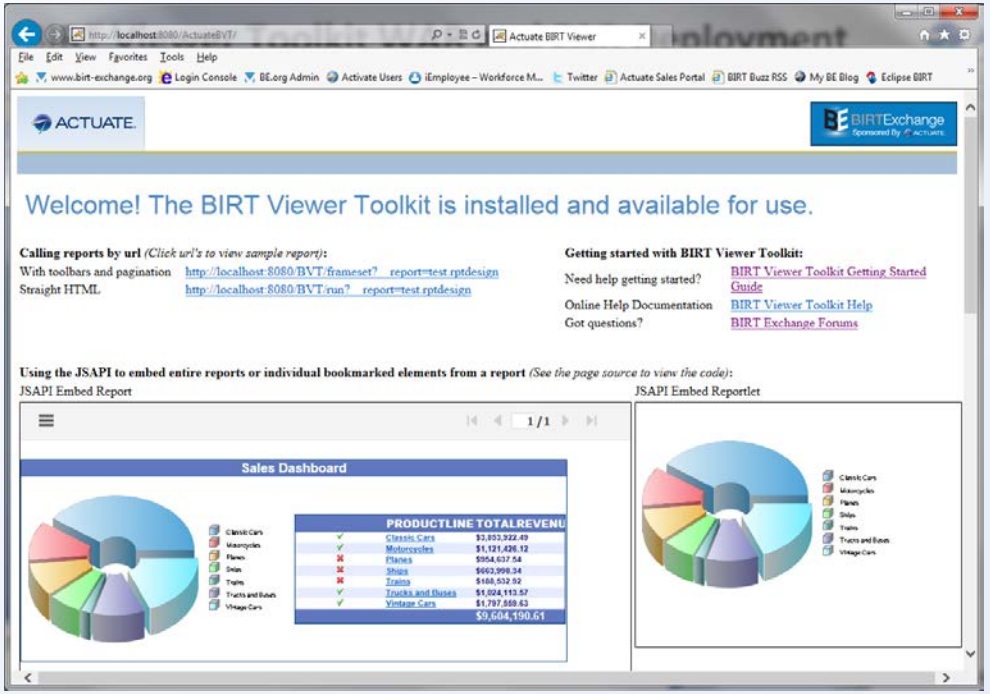
http://localhost:8080/birt/run?__report=SD.rptdesign&__**bookmark=myTable&__isreportlet=true**

BIRT Viewer Toolkit (BVT)

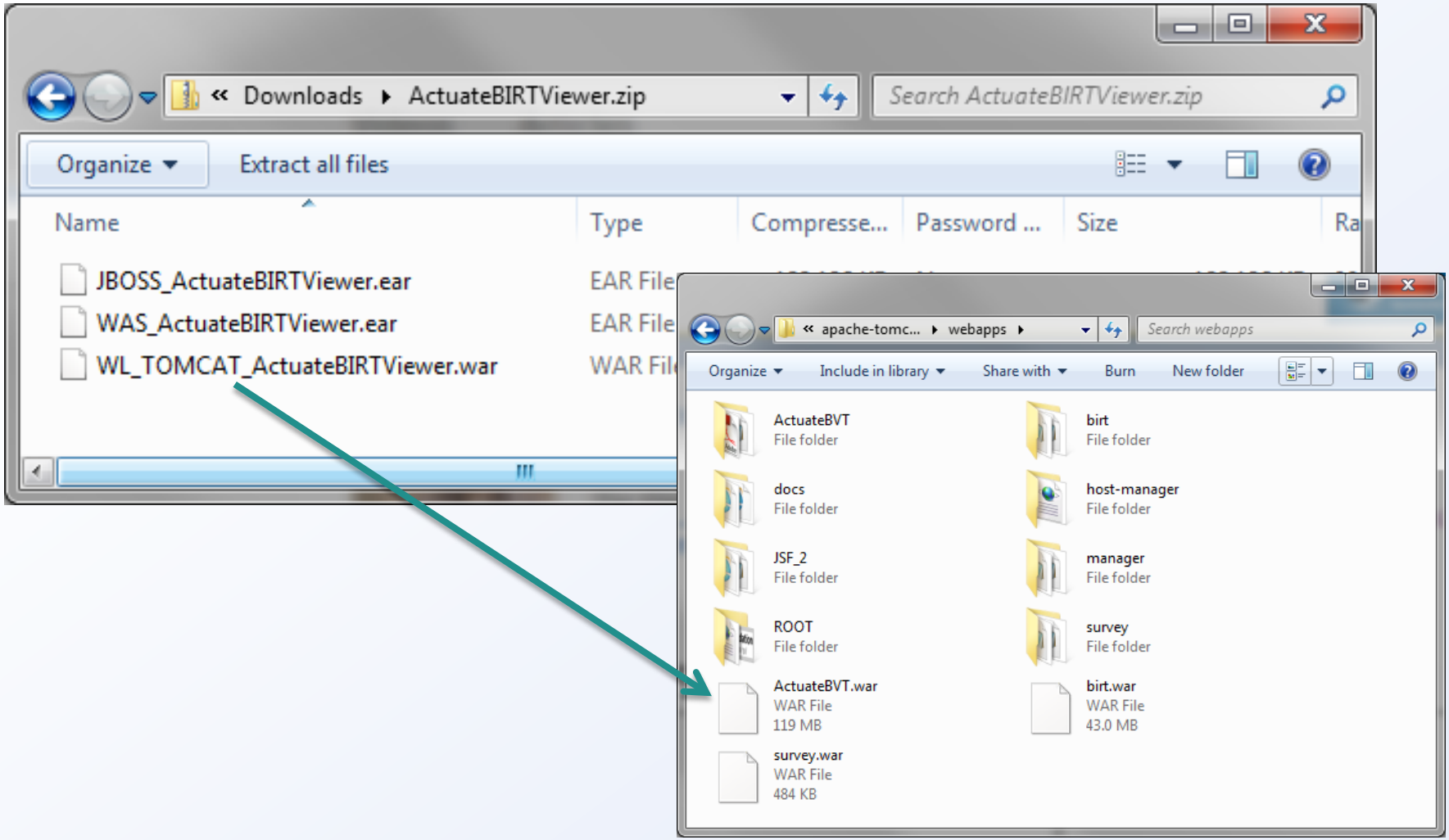
New product for use with open-source BIRT.

Key Features:

- **It's FREE!**
- It uses the Actuate commercial-grade Viewer (Built on OS Viewer)
  - More modern look and feel
  - More seamless parameter entry screen
  - Calendar date-picker
  - Progressive Viewing
- It's built on the latest BIRT version
- **Includes JavaScript API (JSAPI) AJAX library for easier integration into applications of all types.**



BIRT Viewer Toolkit WAR and EAR Deployment



http://localhost:8080/ActuateBVT

Running Reports with BVT

The URL's used with the OS viewer can still be used with BVT.

With viewer toolbar, pagination, etc.:

As a straight HTML output:
http://localhost:8080/ActuateBVT/run?__report=CustomerReport.rptdesign

Output to PDF:
http://localhost:8080/ActuateBVT/frameset?__report=SD.rptdesign&__format=pdf

To display a single report item as a reportlet:
http://localhost:8080/ActuateBVT/run?__report=SD.rptdesign&__bookmark=myTable

Parameterized report with date picker:
http://localhost:8080/ActuateBVT/frameset?__report=DynamicParameter.rptdesign

BVT and JavaScript (JSAPI)

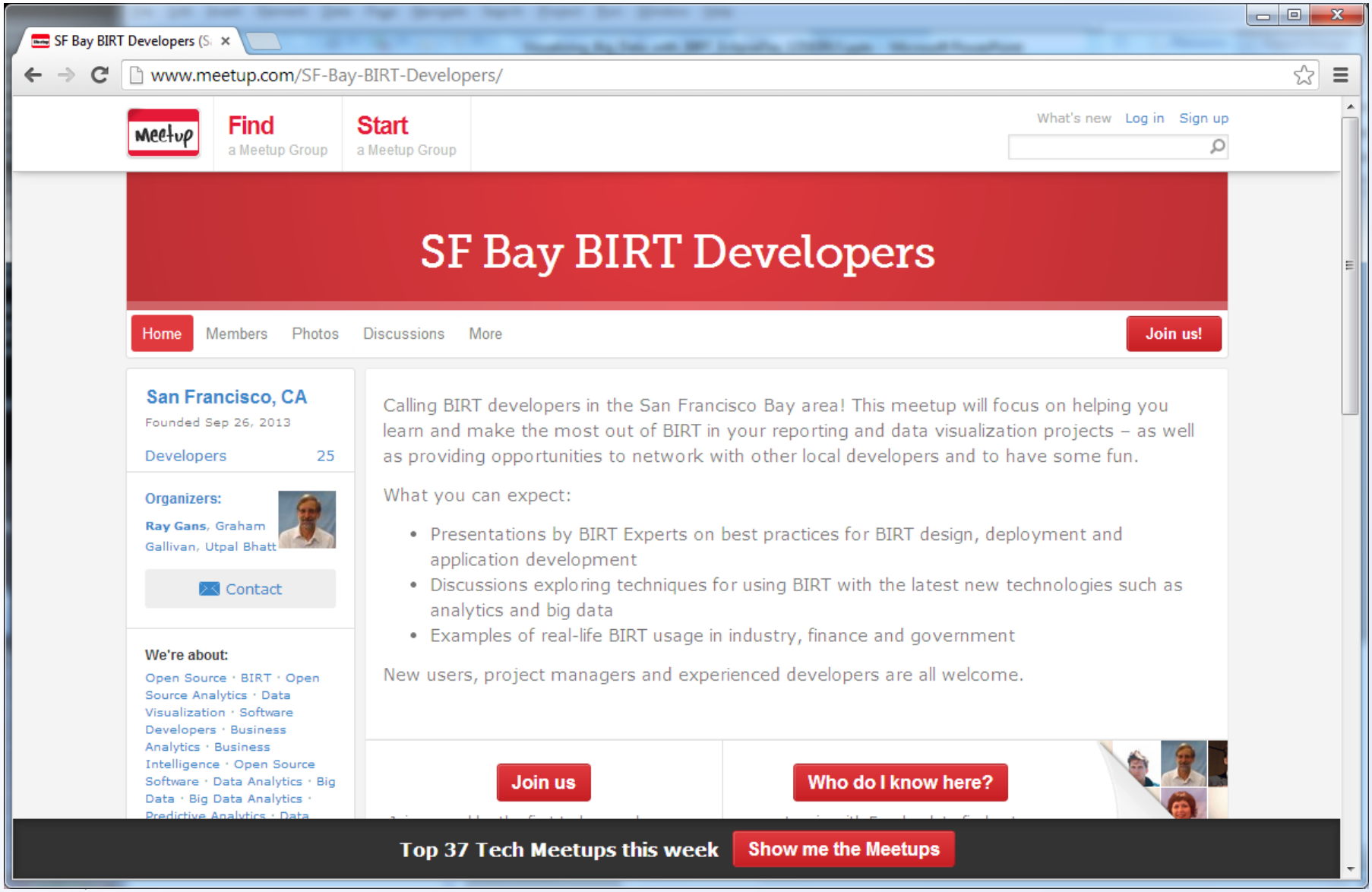
```
<html><head>
    <script type="text/javascript" language="JavaScript"
                src="http://localhost:8080/BIRTViewerToolkit/jsapi"></script>
</head>
<body onload="init();">
    <div id="myDivContainer" style="border-width: 1px; border-style: solid;"></div>
    <script type="text/javascript" language="JavaScript">
        var myViewer;
        function init() {
            actuate.load("viewer");
            actuate.initialize( "http://localhost:8080/ActuateBVT/", null, "administrator", "", createViewer );
        }
        function createViewer() {
            myViewer = new actuate.Viewer( "myDivContainer" );
            myViewer.setReportName("/reports/TableReport.rptdesign" );
            myViewer.submit();
        }
    </script>
</body></html>
```

Live Example: http://localhost:8080/ActuateBVT/SimpleReport.jsp

Sign up for SF Bay Area BIRT Meetups

Next Meeting in January

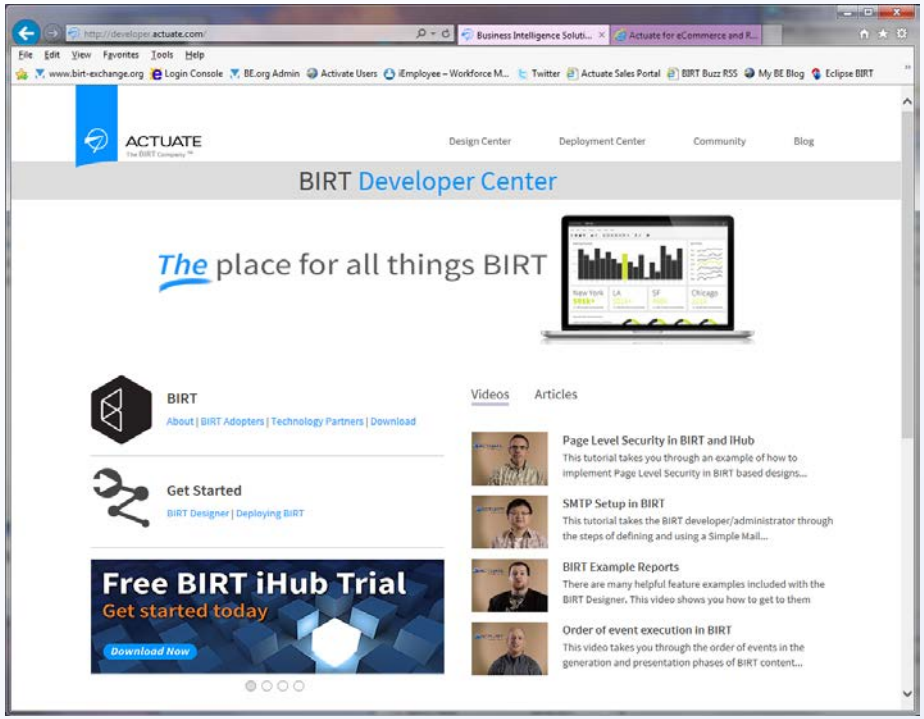
ACTUATE
The BIRT Company



BIRT Developer Center

The place for all things BIRT

- Access demos, tutorials, tips and techniques, documentation…
- Enables developers to be more productive and build applications faster



http://developer.actuate.com

Explore
- Search/sort
- Rate, comment
- Forums

Downloads
- Documentation
- Software
- Examples

Contribute
- BIRT designs, code
- Technical tips
- Applications