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
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.

<table>
<thead>
<tr>
<th>Web Logs</th>
<th>RFID</th>
<th>Sensors</th>
<th>Social Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Text</td>
<td>Searches</td>
<td>Call Detail</td>
<td>Astronomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Records</td>
<td></td>
</tr>
<tr>
<td>Atmospheric Info</td>
<td>Genomics</td>
<td>Biogeochemical</td>
<td>Biological</td>
</tr>
<tr>
<td>Military Surveillance</td>
<td>Medical Records</td>
<td><strong>E-Commerce</strong></td>
<td>Photo/Video</td>
</tr>
</tbody>
</table>

# Traditional Data vs. Big Data

<table>
<thead>
<tr>
<th>Traditional Data</th>
<th>Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gigabytes to Terabytes</td>
<td>Petabytes to Exabytes</td>
</tr>
<tr>
<td>Centralized</td>
<td>Distributed</td>
</tr>
<tr>
<td>Structured</td>
<td>Semi-structured to Unstructured</td>
</tr>
<tr>
<td>Stable Data Model</td>
<td>Flat Schemas</td>
</tr>
<tr>
<td>Known Complex Interrelationships</td>
<td>Few Complex Interrelationships</td>
</tr>
</tbody>
</table>

Source: Wikibon Community
## When to Use Big Data vs. Relational

### Is Big Data a replacement for Relational Data?

<table>
<thead>
<tr>
<th>Analysis Type</th>
<th>Big Data</th>
<th>Relational</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exploratory analysis to uncover value in the data</strong></td>
<td>Operational analysis of what was discovered</td>
<td></td>
</tr>
<tr>
<td><strong>Store HUGE amounts of highly granular data</strong></td>
<td><strong>Store transform (sometimes) aggregated data</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Data flows in BIG Data ➔ “real-time” monitoring</strong></td>
<td><strong>Long term trending analysis</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Timeframe</strong></td>
<td><strong>Timeframe</strong></td>
<td><strong>Timeframe</strong></td>
</tr>
</tbody>
</table>
New Approaches To Big Data Processing And Analytics

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

- **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
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
What Big Data technologies are you using with Eclipse?

- None
- Talend
- Cassandra
- Mahout
- Hive
- R
- MongoDB
- BIRT
- Hadoop

Note: Responders could choose more than one option.
How important is Data Visualisation/Reporting to your project?

- Essential: 52%
- Sometimes important: 28%
- Occasionally useful: 13%
- Never needed: 7%

80%
How do you create/use data display tools in development?

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

Note: Responders could choose more than one option
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
“Not enough staff with expertise” and “Expected cost of Big Data initiatives” are the major inhibitors
### Big Data Initiatives

<table>
<thead>
<tr>
<th>Major Benefits <em>Expected</em></th>
<th>Major Benefits <em>Realized</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Make better decisions, faster</td>
<td>Gain competitive advantage</td>
</tr>
<tr>
<td>Gain competitive advantage</td>
<td>Improve customer targeting</td>
</tr>
<tr>
<td>Improve efficiency</td>
<td>Make better decisions, faster</td>
</tr>
<tr>
<td>Improve customer targeting</td>
<td>Improve efficiency</td>
</tr>
</tbody>
</table>
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%.
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 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*

ACTUATE

*The BIRT Company*

AUGUST

2004
What is BIRT?

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

Power to create very complex layouts
## BIRT Release History supports Big Data

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

Personalized Analytics and Insights with BIRT for Hadoop

Note: this tutorial was validated with Sandbox 1.3

Welcome to the open source Eclipse-based BIRT (Business Intelligence and Reporting Tools) tutorials developed by Actuate – The BIRT Company™. There are two tutorials for your learning enjoyment, each fairly straightforward and designed to help you get connected with BIRT within minutes to access data from the Hortonworks Sandbox and immediately gain personalized analytics and insights into that data. There are two tutorials for your learning enjoyment:

1. **Open source Eclipse BIRT (Business Intelligence and Reporting Tools) Connectivity**
2. **Introduction to Integrating BIRT (Business Intelligence and Reporting Tools) with the Hortonworks Sandbox**

http://hortonworks.com/hadoop-tutorial/birt_reporting_tutorial/
Connecting to Hadoop

New Hive Data Source Profile

- **Driver Class:** org.apache.hive.jdbc.HiveDriver
- **Database URL:** jdbc:hive2://192.168.1.114:10000/default
- **User Name:** hadoop
- **Password:**
- **Add File Statement:**

Manage Drivers...  Test Connection...
Hive JDBC – HQL Sub Query Example

```sql
select lastyear.month,
    thisyear.month,
    (thisyear.monthly_shipments - lastyear.monthly_shipments) /
    lastyear.monthly_shipments as monthly_shipments_delta
from (select year(l_shipdate) as year,
        month(l_shipdate) as month,
        count(l_shipdate) as monthly_shipments
    from lineitem
    where year(l_shipdate) = 1997
    group by year(l_shipdate), month(l_shipdate)
) lastyear
join
(select year(l_shipdate) as year,
    month(l_shipdate) as month,
    count(l_shipdate) as monthly_shipments
    from lineitem
    where year(l_shipdate) = 1998
    group by year(l_shipdate), month(l_shipdate)
) thisyear
on lastyear.month = thisyear.month
```
Hive JDBC – get_json_object UDF

```sql
1 SELECT get_json_object(json.name, '$.owner')
2 FROM json
```
Hive JDBC – RegExP Example

Define a SQL query text using available items.

```
SELECT REGEXP_REPLACE(REGEXP_REPLACE(REGEXP_REPLACE(key, '\001', '^A'), '\0', '^B'), '\002', '^B'), value
FROM (SELECT key, sum(value) as value
      FROM mytable
      GROUP BY key)
```
Hive JDBC – HQL Hints example

```sql
1 SELECT /*+ MAPJOIN(b) */ a.key, b.value
2 FROM srcbucket_mapjoin_part_1 a
3 JOIN srcbucket_mapjoin_part_2 b
4 ON a.key = b.key AND a.ds = '2008-04-08'
5 AND b.ds = '2008-04-08' LIMIT 1
```
Hive JDBC – Transform Example

Define a sql query text using available items.

1. SELECT
2. TRANSFORM (userid, movieid, rating, unixtime)
3. USING 'python weekday_mapper.py'
4. AS (userid, movieid, rating, weekday)
5. FROM u_data
Cassandra – Scripted Data Set Example

```java
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();

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

BIRT Designer
- Eclipse Designer
- Eclipse DTP, WTP, ...
- Chart Designer

Design Engine

BIRT Engine
- Generation Services
- Charting Engine
- Data Services
- Presentation Services

XML Design

Data

Document
- HTML
- PDF
- Excel
- Word
- PowerPoint
- PostScript
- ...
High Level BIRT Architecture – APIs

- **DE API**
  - Design Engine
  - Produces XML Report, Templates, and Library Designs

- **RE API**
  - Report Engine
  - Runs Reports and produces output – PDF, HTML, Doc, XLS, PS, PPT Etc

- **CE API**
  - Chart Engine
  - Consume Chart EMF model and produces Chart Output. Supports 14 Main types and many sub types. Outputs to PNG, JPG, BMP, SVG, PDF, SWT, and SWING

**All Engines can be run with or without OSGi**

**Core BIRT Open Source Products**

- Report Designer
- Chart Builder
- Example Viewer
  - Can be ran outside of BIRT
Multiple Deployment Options

- BIRT Design
  - XML

- BIRT onDemand
  - BIRT Engine and APIs
  - eclipse

- BIRT iHub
  - BIRT Viewer Toolkit
• 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
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();
        ...
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

Top 5 Customers

- Euro+ Shopping Channel: 820,690
- Mini Gifts Distributors Ltd: 591,827
- Australian Collectors, Co.: 180,585
- Muscle Machine Inc: 177,914
- La Rochelle Gifts: 158,573

Top 25% Customers (sorted by name)

<table>
<thead>
<tr>
<th>Customer Name</th>
<th>Customer Number</th>
<th>Customer Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV Stores, Co.</td>
<td>187</td>
<td>$148,410.09</td>
</tr>
<tr>
<td>Anna's Decorations, Ltd</td>
<td>276</td>
<td>$137,034.22</td>
</tr>
<tr>
<td>Australian Collectors, Co.</td>
<td>114</td>
<td>$180,585.07</td>
</tr>
<tr>
<td>Corporate Gift Ideas Co.</td>
<td>821</td>
<td>$132,340.78</td>
</tr>
<tr>
<td>Corrida Auto Replicas, Ltd</td>
<td>458</td>
<td>$112,440.09</td>
</tr>
<tr>
<td>Danish Wholesale Imports</td>
<td>145</td>
<td>$129,085.12</td>
</tr>
</tbody>
</table>
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
<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

SF Bay BIRT Developers

Calling BIRT developers in the San Francisco Bay area! This meetup will focus on helping you learn and make the most out of BIRT in your reporting and data visualization projects – as well as providing opportunities to network with other local developers and to have some fun.

What you can expect:

- Presentations by BIRT Experts on best practices for BIRT design, deployment and application development
- Discussions exploring techniques for using BIRT with the latest new technologies such as analytics and big data
- Examples of real-life BIRT usage in industry, finance and government

New users, project managers and experienced developers are all welcome.

Join us

Who do I know here?
The place for all things BIRT

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

Explore
- Search/sort
- Rate, comment
- Forums

Downloads
- Documentation
- Software
- Examples

Contribute
- BIRT designs, code
- Technical tips
- Applications

http://developer.actuate.com