End-to-End Open Source Business Intelligence, ETL, and Data Warehouse for MySQL

Bob Zurek, CTO, Infobright
Brian Gentile, CEO, Jaspersoft
Agenda

- BI Market Forces
- Joint Jaspersoft and Infobright solution
- Case Studies
- Demo
- How to Learn More
**Disruption**  
**A Changing BI Market**

**Consolidation**  
Oracle + Hyperion  
SAP + Business Objects  
IBM + Cognos

**Result:** Hegemony & Opportunity!  
Cum. Mkt Cap = $300+ B  
Cum ’08 Rev = $130+ B

**End-User Involvement**  
“Desktop” Tools / Interactivity  
Collaboration / Workgroup  
“Consumerization” of information

**Result:** Ease-of-Use is Imperative

**Technology Shifts**  
Software-as-a-Service  
Web Services (SOA)  
Search  
In-Memory Analytics

**Result:** Much Simpler Delivery & Integration is Required

**Besieged CIO**  
CEO wants growth/innovation  
Information explosion confounds  
BI still tops CIO/IT “Must Do” list  
80% of IT budget goes to maintenance

**Result:** ‘09 Ave. IT budget = No growth
Evolution of the BI Market
Cost of Ownership

• 5+ years ago, multi-million dollar budgets were required
  • Teradata or Netezza for a data warehouse
  • Informatica PowerCenter or IBM DataStage for ETL
  • SAP Business Objects, Microstrategy, or IBM Cognos for BI

• Open source BI and data warehousing changes everything!
  • Businesses of every size can use reporting and analytics to compete more effectively
  • Especially in today’s economic climate, businesses can emerge further ahead of their competitors after the downturn by investing in powerful open source software and enabling business initiatives to proceed and outperform the competition
Barriers to Information Success

- Complexity of BI tools and interfaces
- Cost of BI software and per-user licenses
- Difficulty accessing relevant, timely, or reliable data
- Insufficient IT staffing or excessive software requirements for IT support
- Difficulty identifying applications or decisions that can be supported by BI

Source: September, 2008, InformationWeek Research Business Intelligence Survey of 385 business technology professionals using Business Intelligence tools.
The Market: Size & Growth

- **8.1% CAGR**
  - 6.5% Stand-Alone Growth
  - 13.2% Embedded Growth

- **Key Trends:**
  - Vendor consolidation
  - Continued “CIO must have”
  - Less capital-intensive options become more popular

Open Source Software Market

- **2006**
  - $1.8B
- **2011**
  - $5.8B


- **26% Market CAGR**
- Growth accelerates over forecast period
- Increased customer interest
  - Greater choice
  - New innovation
  - Comfort with subscription model
  - Larger enterprise adoption
- Macro-economics
Why Open Source Business Intelligence?

- The Power of Open Source – customize & extend
- The Power of Community – contributions & testing
- Affordability (lowest TCO) – prove value then subscribe
- Modern Architecture – designed for integration

![Image of chart showing most important technologies for 2008](image-url)
Overview of the end-to-end architecture for BI, ETL, and data warehouse environments
Joint Architecture

Sources:
- Source 2 RDBMS
- RDBMS
- POJO, EJB, XML, MDX, Custom

JasperETL

DW / Data Integration Developer

Data Warehouse Professional

Column Orientation
Knowledge Grid Intelligence

INFOBRIGHT

BI Power User

JasperAnalysis

BI User

JasperServer

Analysis

Reporting

Dashboards

©2009 Jaspersoft Corporation. Proprietary and Confidential
Community & Commercial Success
- Nearly 9 million total product downloads
- Estimated 100,000+ production deployments (WW)
- Over 98,000 registered developers in Community
- 10,000 commercial customers (96 countries)
- ~80% year-over-year sales growth

Open Source BI Market Leader
- Community and Professional Editions of each product within the Jaspersoft BI Suite
- The leading BI suite that helps small, mid-market and enterprise customers implement simple, powerful BI solutions throughout the organization

The Most Widely Used BI Software in the World!
Jaspersoft Business Intelligence Suite
Advanced Reporting and Analysis

iReport
Complex Report Design (via desktop)

JasperServer
Interactive Reporting Server (via web browser)

JasperAnalysis
Interactive OLAP Data Analysis (via web browser)

JasperETL
High Performance Data Integration (Transformation Inside or outside the Database)

World’s Most Popular Reporting Library

BI Developers

Business Users

Developers & DBAs
Jaspersoft Advantage
Reporting & Dashboards

Reporting
• Print-ready pixel perfect production reports
• Web 2.0 drag & drop ad hoc query and reporting
• Access to any data source: relational, XML, Hibernate, EJB, POJO, custom
• Connect directly or via rich business metadata layer

Dashboards
• Web 2.0 drag & drop dashboard designer
• Live component-level refresh
• Mash-ups with external content
Jaspersoft Advantage
Integrated Analysis

Analysis
• In-memory analysis directly in reports
• Standards-based and scalable ROLAP
• Web, Excel, mobile user interfaces

Data Integration
• Wide range of connectors & transforms
• Intuitive graphical user interface
• High performance generated code
• Activity Monitoring Console

BI Suite
• Modular, shared services
• SaaS-enabled / multi-tenancy
• Small-footprint, 100% pure Java
• Designed for integration / embedding
• Internationalized and localized
• Extensible and customizable source code
Jaspersoft Advantage
Simple, Powerful BI

Sophisticated charts

Integration with R for data mining

Drag & Drop Dashboards & Reports

Mash-Ups For Real-Time Updates
Infobright Momentum

- **Technology Innovation**
  - First commercial open source data warehouse
    - Community & Enterprise Editions
  - Lowest cost technology in the market
  - Simplest and easiest to build and manage
  - Powerful & scalable MySQL
  - Technology unlike ANYTHING in the industry
Infobright Advantage

- **Strong Momentum & Mature Product**
  - > 60 Customers in 10 Countries
  - > 33 Partners on 6 continents
  - A vibrant open source community
    - +1 million visitors
    - Approaching 10,000 downloads
    - 2,000 active community participants
Infobright Advantage

Knowledge Grid – statistics and metadata “describing” the super-compressed data

Data Packs – data stored in manageably sized, highly compressed data packs
Data compressed using algorithms tailored to data type

Smarter architecture
- Load data and go
- No indices or partitions to build and maintain
- Knowledge Grid automatically updated as data packs are created or updated
- Super-compact data footprint can leverage off-the-shelf hardware
Infobright Advantage
Column Orientation

<table>
<thead>
<tr>
<th>EMP_ID</th>
<th>FNAME</th>
<th>LNAME</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Moe</td>
<td>Howard</td>
<td>10000</td>
</tr>
<tr>
<td>2</td>
<td>Curly</td>
<td>Joe</td>
<td>12000</td>
</tr>
<tr>
<td>3</td>
<td>Larry</td>
<td>Fine</td>
<td>9000</td>
</tr>
</tbody>
</table>

**Row Oriented** (1,Moe,Howard,10000; 2,Curly,Joe,12000; 3,Larry,Fine,9000;)

- Works well if all the columns are needed for every query.
- Efficient for transactional processing if all the data for the row is available

**Column Oriented** (1,2,3; Moe,Curly,Larry; Howard,Joe,Fine; 10000,12000,9000;)

- Works well with aggregate results (sum, count, avg.)
- Only columns that are relevant need to be touched
- Consistent performance with any database design
- Allows for very efficient compression
**Data Packs**
- Each data pack contains 65,536 data values
- Compression is applied to each individual data pack
- The compression algorithm varies depending on data type and data distribution

**Compression**
- Results vary depending on the distribution of data among data packs
- A typical overall compression ratio seen in the field is 10:1
- Some customers have seen results have been as high as 40:1
Infobright Advantage
Knowledge Grid

Data Pack Nodes (DPN)
A separate DPN is created for every data pack created in the database to store basic statistical information.

Character Maps (CMAPs)
Every Data Pack that contains text creates a matrix that records the occurrence of every possible ASCII character.

Histograms
Histograms are created for every Data Pack that contains numeric data and creates 1024 MIN-MAX intervals.

Pack-to-Pack Nodes (PPN)
PPNs track relationships between Data Packs when tables are joined. Query performance gets better as the database is used.

This metadata layer = 1% of the compressed volume
Infobright in Action
Online Analytics

Before

- 20 billion online transactions/month
- Previous database technology unable to quickly produce the detailed statistics needed
- In some cases, could not resolve the queries at all
- Constantly changing click stream data required continual re-building and tuning of database

After

- Loaded >3.2 billion rows of data at >300,000 rows per second
- Achieved 30x data compression
- Much faster data loading and query performance
- No indexing or tuning needed
- Running on single industry standard server ($12.5K cost)

“Infobright performed extremely well... It handled huge volumes of data and resolved queries faster than I have ever witnessed after years in the industry – and did so using standard hardware”

– Mats Johansson, Lincube (SI)
### Before
- Goal to transition from regional to national player
- Scalability issues
- Need to replace first generation solution with data warehouse
- Overcome deployment obstacles for larger market

### After
- SaaS BI, multi-tenancy featuring...
- Sun products: MySQL, NetBeans, GlassFish
- Jaspersoft BI Suite including analysis, reporting, and ETL
- MySQL DW: 180 GB; Sources 700GB
- 70 franchise organizations representing more than 2,000 stores. About 4500 end users.

“Jaspersoft added a lot of ‘wow factor’ ...we've gone from regional provider to competing with industry giants and are being considered by household brands for global deployments.”

– **Bruce Belvin**, Monolith Software Solutions
Joint Open Source Project and Virtual Machine

Project
• Integrated examples
• Quick start guide

Virtual Machine image
• VMware
• Open Source Editions:
  • Jaspersoft CE
  • Infobright (ICE)

• Joint Project:
http://jasperforge.org/projects/infobright
• Joint VMware image:
http://www.infobright.org
Where to learn more

Jaspersoft

Dan Houston
E-mail: dhouston@jaspersoft.com
Phone: 415-348-2361

Or info@jaspersoft.com

Infobright

Roger Cummings
E-mail: roger.cummings@infobright.com
Phone: 630-308-4200

Or info@infobright.com