

From Data to Intelligence in 17 Clicks

*Business Intelligence for the
Open Source Development Community*

Seth Grimes

Alta Plana Corporation

Alta Plana

Scott Rosenbaum

Innovent Solutions



March 26, 2009

Business Intelligence is Different

Most information technology
enables defined tasks

Business Intelligence
enables decisions

Business Intelligence

Information

drives business decision making

Process

events → data → intelligence → decision

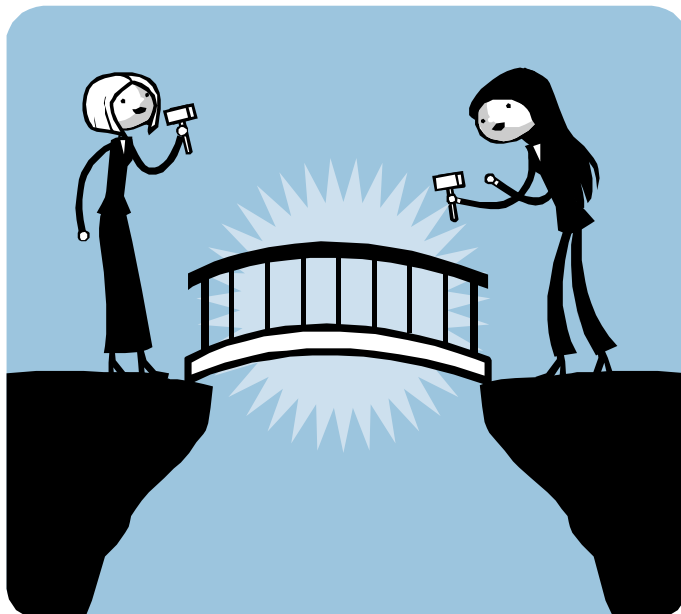
Software

tools to transform data into intelligence

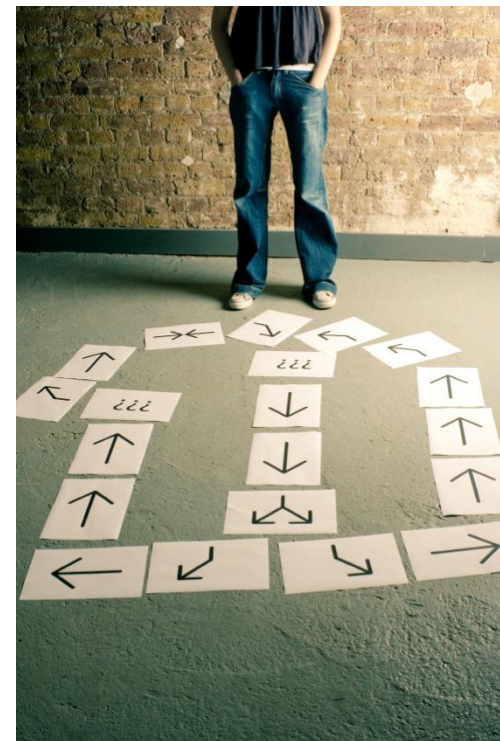
Bridging the Gap



Data



BI



Decisions

Make your data tell a story!

A BI Story...

A trade association had questions:

Which members at Top 500 firms fail to license the Association's SW?

goal = sales

How is Continuing Education course enrollment affected by changes to state CE requirements?

goal = planning, lobbying

What are the characteristics of members who *Cancel or Expire*?

goal = retention

A BI Story...

Our BI approach:

Gather data from source systems; assess quality; join.

Create data structures & analytical interfaces.

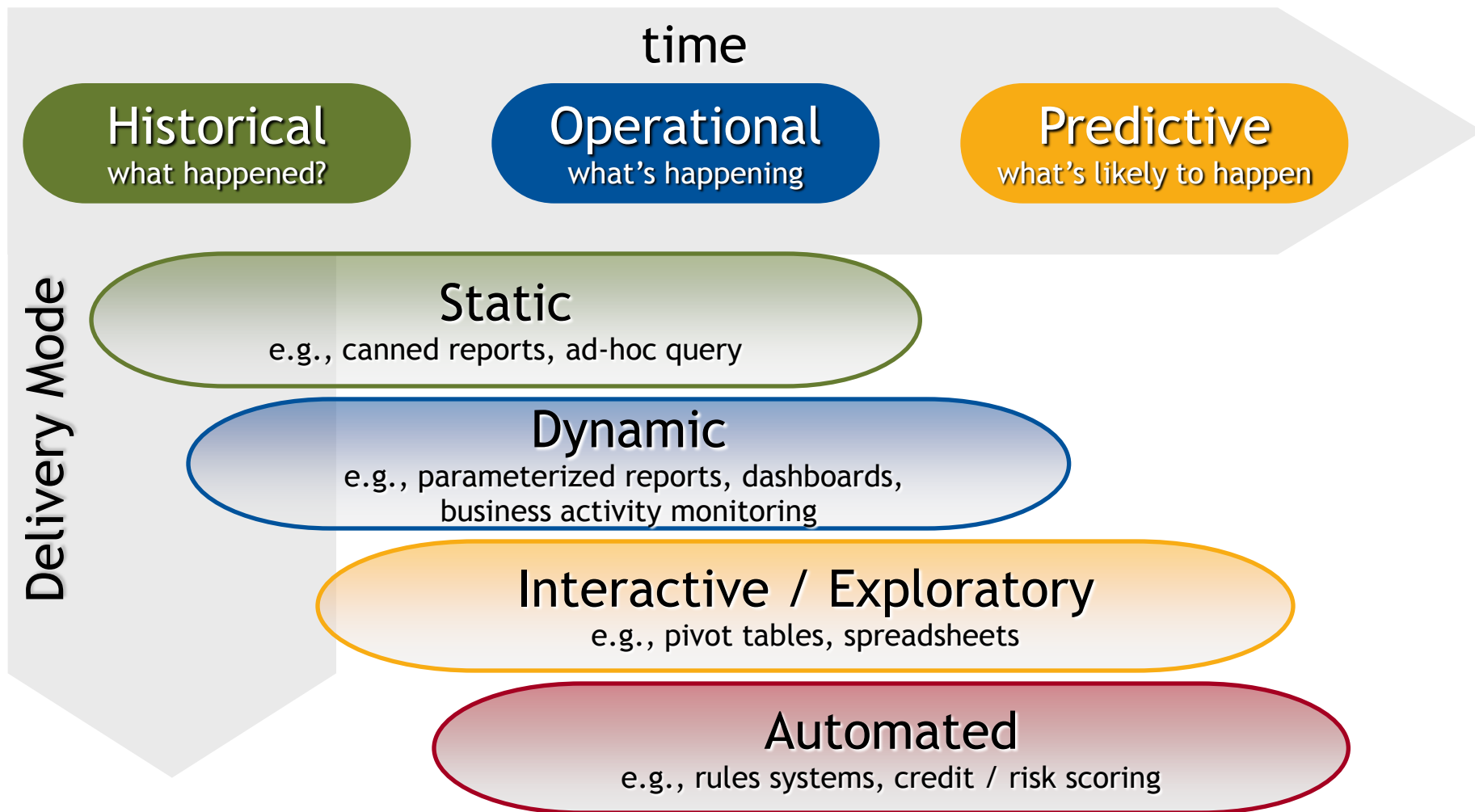
Web OLAP interface.

Query/data retrieval hooks into Excel & Access.

Prototype queries.

Knowledge transfer.

BI Modes



Software Platform Elements

Business Intelligence Platform

Information Delivery

- Reporting
- Dashboards
- Ad Hoc Query
- Microsoft Office Integration

Integration

- BI Infrastructure
- Metadata Management
- Development Environment
- Workflow and Collaboration

Analysis

- OLAP
- Visualization
- Predictive Modeling and Data Mining
- Scorecarding

Data Management

Governance, Quality, Integration, Security, Architecture

Source: Gartner (April 2007)



Pentaho Business Intelligence Platform Portal Demo

- Home
- Getting Started
- Reporting
- Business Rules
- Printing
- Exporting
- Widgets
- DataSource
- Secure
- Advanced

Filters

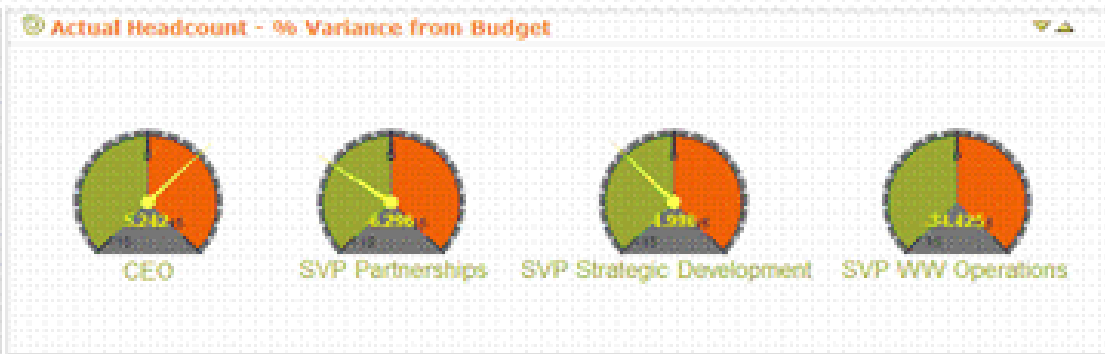
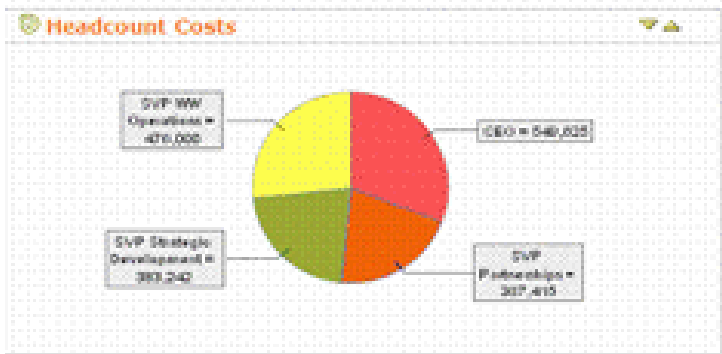
Select filters to apply to other controls on this page

REGION: Central

DEPARTMENT:

Headcount Data

Position	Actual	Budget	Variance
SVP Strategic Development	\$383,242	\$403,405	\$20,163
SVP Partnerships	\$367,415	\$392,100	\$24,685
CEO	\$549,625	\$522,250	-\$27,375
SVP WW Operations	\$476,000	\$725,887	\$249,887
Total	\$1,776,282	\$2,043,642	\$267,360



Actual Headcount - % Variance from Budget

			Measures			
Region	Department	Positions	Actual	Budget	Variance	Variance Percent
-All Regions	-All Departments	-All Positions	143,639,962.00	143,199,389.00	-440,593.00	-0.31%
		Executive Management	6,299,022.00	6,494,166.00	195,144.00	3.00%
		Finance	12,224,220.00	12,087,406.00	-136,814.00	-1.13%
		Human Resource	13,075,463.00	12,989,341.00	-86,122.00	-0.66%
		Marketing & Communication	13,910,753.00	13,770,267.00	-140,486.00	-1.02%
		Product Development	10,644,102.00	10,786,611.00	142,509.00	1.32%
		Professional Services	76,317,649.00	76,098,206.00	-219,443.00	-0.29%
		Sales	11,168,773.00	10,973,392.00	-195,381.00	-1.78%
Central	+All Departments	+All Positions	37,893,162.00	38,397,600.00	504,438.00	1.31%
Eastern	+All Departments	+All Positions	35,248,940.00	35,487,861.00	238,921.00	0.67%
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%

MDG Monitor: >MDG M

File Edit View History Bookmarks Tools Help

http://www.mdgmonitor.org/map.cfm?goal=7&indicator=0&cd=

AN INITIATIVE OF THE UNITED NATIONS

MDGMONITOR
TRACK. LEARN. SUPPORT

TRACKING THE MILLENNIUM DEVELOPMENT GOALS

Countdown to 2015: 6 years 287 days 16 hours 48 minutes 32 seconds

HOME | ABOUT | BROWSE BY GOAL | BROWSE BY LOCATION | MDG MAP | NEWS | PRESS ROOM | SUPPORT

STEP 1. SELECT GOAL INDICATOR

1. POVERTY

Population below \$1 (PPP) per day consumption, (%)

Children under 5 moderately or severely underweight, (%)

2. EDUCATION

Net enrollment ratio in primary education, both sexes

3. WOMEN

Seats held by women in national parliament, (%)

Gender Parity Index in primary level enrollment

STEP 2. SELECT LOCATION

Austria

Azerbaijan

Bahamas

Bahrain

Bangladesh

Barbados

Belarus

Belgium

Belize

Benin

STEP 3. SELECT YEAR

1990 - 1994 1995 - 1999 **2000 - 2004**

LEGEND Population below \$1 (PPP) per day consumption, percentage

40% or more 20% - 39% 10% - 19% 5% - 9% Less than 5% No data

* disclaimer

ABOUT THIS MAP | ADDITIONAL RESOURCES | TOPIC NEWS

© Copyright United Nations Development Programme, 2007. All Rights Reserved.
email: mdgmonitor@undp.org

Home | About | Browse by Goal | Browse by Location | MDG Map | News | Press Room | Support

BI Flow

Gather - BI Sources

evaluate and marshal data assets

Grind - BI Process

manipulate raw data into information

Deliver - BI Delivery

provide (secure) access to the information in readily usable formats

BI Sources

Sources are operational systems -

- Transactional databases.

- Stream and messaging sources.

Data from these sources is loaded to:

- Data warehouses.

- Data marts, i.e., local data stores.

And we have operational BI.

- Operational reporting isn't new.

- Operational BI - data analysis - is.

BI Processing

Apply Domain Knowledge

hierarchy, time, product line
measures

Manipulate The Data

Multi-dimensional cubes
KPI
charts

Communicate Findings

tell a story

BI Delivery

BI is delivered:

Via workbenches, dashboards, reports, maps, and other analytical UIs.

Embedded in the UIs of line-of-business (LOB) applications.

Behind the scenes, supporting automated decision making for LOB applications.

Credit and risk scoring and Web-experience personalization are examples.

As a Service.

Building the EclipseCon Program

90+
Projects

10,000+
Contributors

1,000+
Committers

18
Sponsors

180+
Member Companies

600+ hours of submissions

Monday

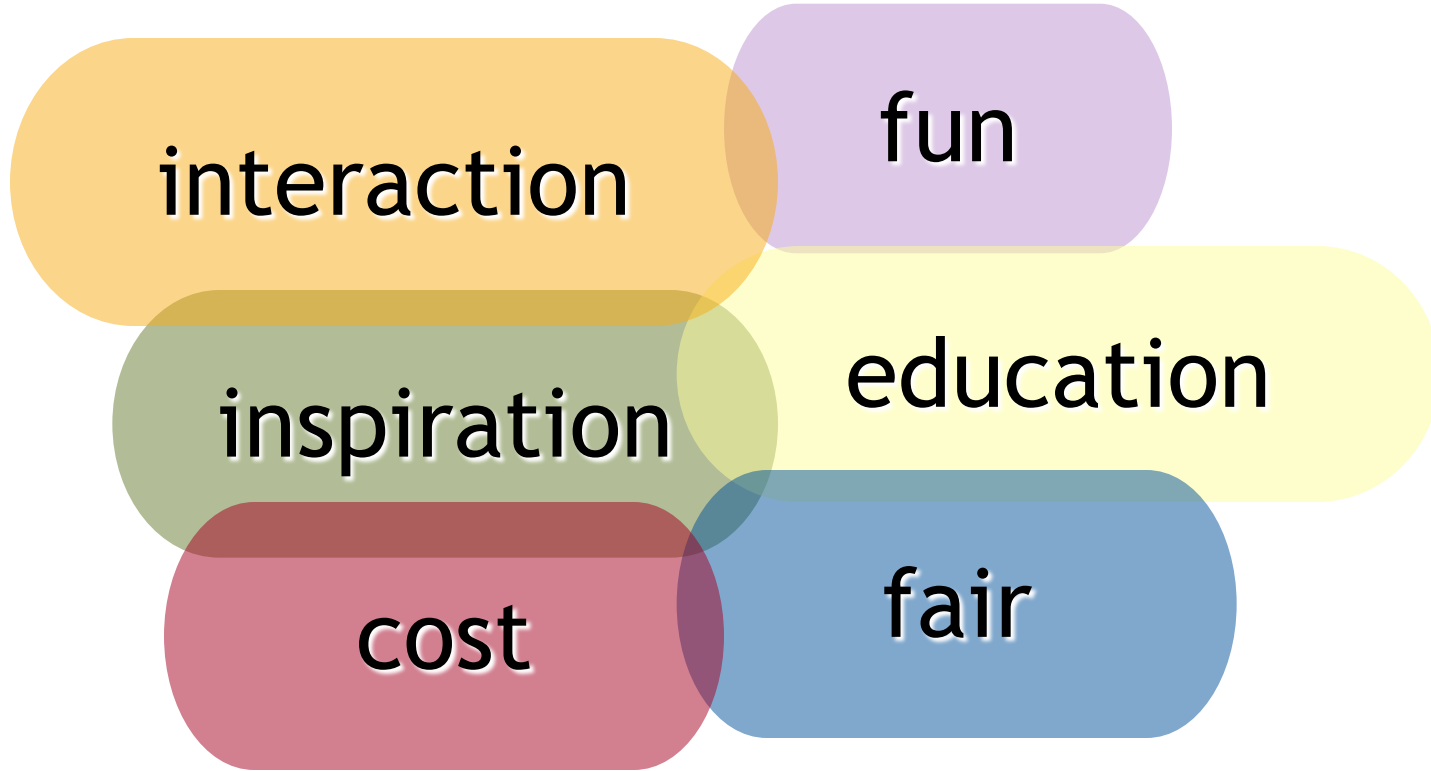
Tuesday

Wednesday

Thursday

254 Hours
available

Other Factors



Quality is Expected

Data Assets

Historical

registrations

evaluations

allocation history

Current

talk proposals

rfid attendance data

evaluations

twitter?

Can BI Help?

Speaker Quality

best speakers
worst speakers

Target Audience

what was proposed
what did people attend

Program Balance

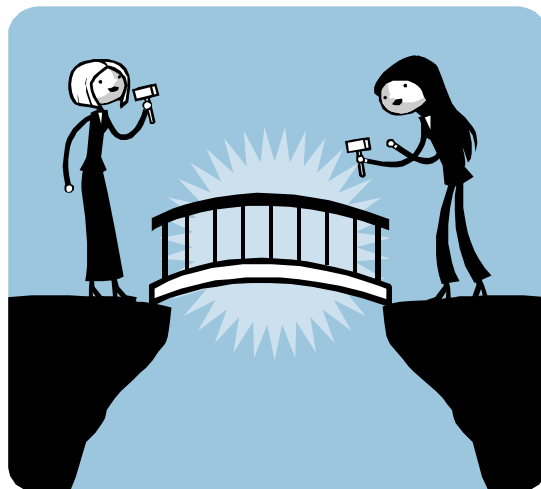
project coverage
company diversity

Difficulty

talk type
tutorial, long, short
talk difficulty
green, blue, black

17 clicks

registrations
evaluations
allocation history
talk proposals
rfid data
evaluations
twitter



speaker quality
target audience
program balance
difficulty

BI Scene

BIRT is only one piece of the puzzle

Mainstream Leaders

Specialists

Open Source

Upstarts

Hosted

Data Mining

Mainstream Leaders

The Big Four + One

SAP: Business Objects

IBM: Cognos

Oracle: Siebel Analytics, Hyperion

Microsoft

SAS

Others

Information Builders

MicroStrategy

Specialists

Visualization, performance management,
reporting, dashboards...

Actuate

QlikTech

Tableau

TIBCO Spotfire

OSBI Players



Data Mining

Angoss

KXEN

Megaputer

Salford Systems

SPSS

R *

WEKA *

* *Open Source*

Others

Hosted

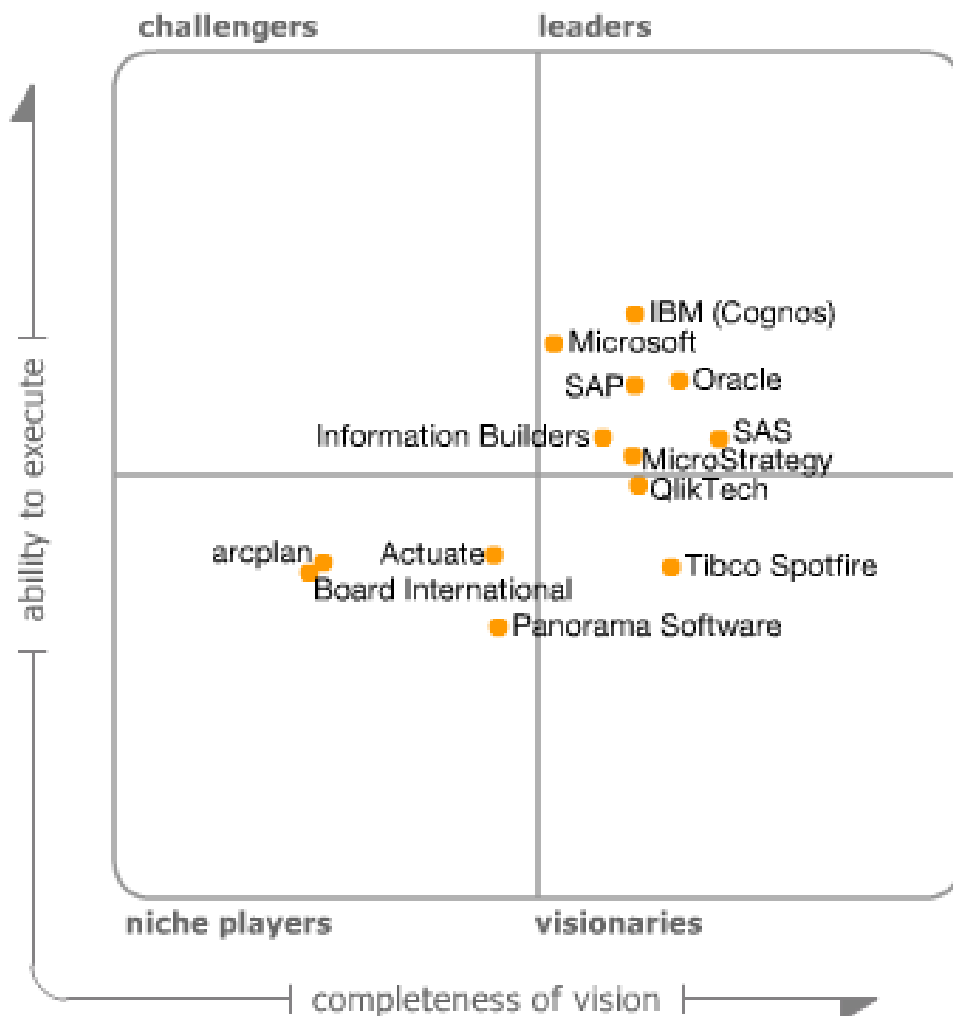
PivotLink

LucidEra

Upstarts

many

The BI World According to Gartner



<http://mediaproducts.gartner.com/reprints/sas/vol5/article8/article8.html>

As of January 2009

Where's Open Source BI?

Nigel Pendse is author of the BI Report - in '07:

Actually, I've been quite surprised at how little impact open source BI solutions seem to be having. I was expecting much more.

I guess there are two parallel universes: customers in OSW (open-source world) have decided for idealistic, economic or technical reasons that they must have an open-source solution, and don't even consider any proprietary options, while most other people ignore open-source solutions.

OSBI Category Error

My guru friend made a *category error*.

My Opinion:

Open source does not succeed (best) by merely replicating commercial, proprietary, closed source software and processes.

The most successful open source projects are not imitative, they are innovative.

OSBI is a preferred option for ISVs & developers.

OSBI is innovating in areas such as analytical workflows and streaming data.

OSBI has NOT (until recently) aimed to replace closed-source, commercial solutions.

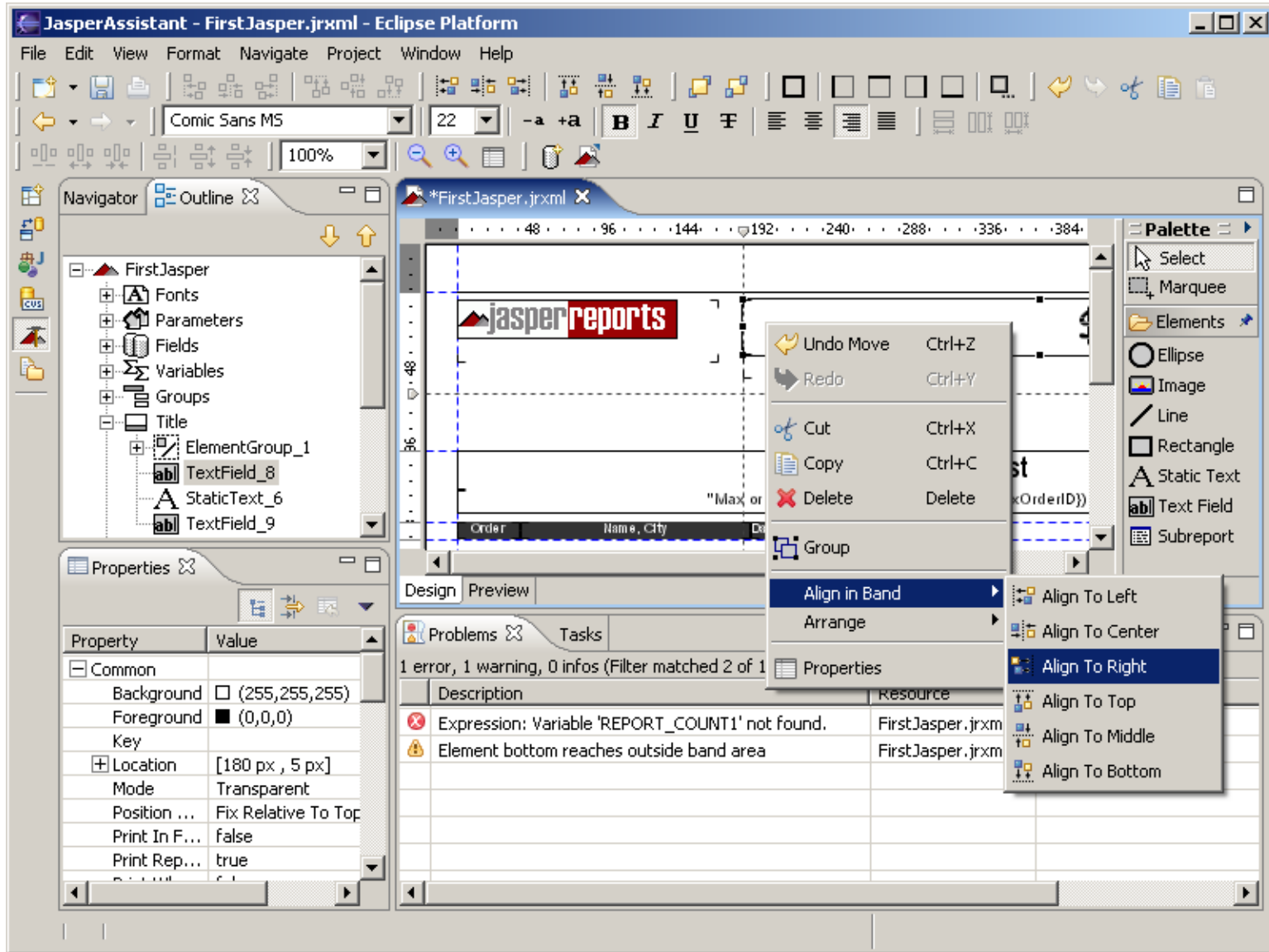
OSBI Players



The screenshot displays the Palo Client interface for a 'Sales' cube. The main window shows two pie charts representing data for 'Qtr. 1' and 'Qtr. 2'. The dimensions are set to 'Regions' (Europe) and 'Years' (2003). The measures are 'Variance' and 'Units'. A data table is visible on the right, showing values for different regions and products.

	West	East	South	North	
87.00	-15,242.00	-2,450.00	9,894.00	-6,039.00	
86.00	-18,781.00	3,516.00	-1,157.00	-4,514.00	
74.00	-16,036.00	-2,859.00	-6,531.00	352.00	
89.00	16,795.00	-2,287.00	17,942.00	-1,961.00	
61.00	16,489.00	146.00	15,993.00	1,523.00	
69.00	8,865.00	-2,257.00	1,181.00	-4,520.00	
81.00	-8,559.00	-176.00	768.00	1,036.00	

http://www.jpalo.com/en/products/palo_client.html



<http://www.jaspersoft.com/>

Talend Open Studio (1.1.1._r1413) | ocarbone@talend.com | TalendDemos

File Edit View Window Help

Repository Navigator

Job DataSimulation Job MailsJob

Business Models
Job Designs
DemoTOS
DataSimulation 0.1
MailsJob 0.1
PurgeJob 0.1
PurgeLog 0.1
selectLogTable 0.1
testGetDate 0.1
orchestration
t00_Prior_Test
t01_Components
t02_Concatening
t03_Routines
Copy_of_tFileOutputDelimited 0
efdscs 0.1
test 0.1
Code
Metadata

Designer Code

Properties PerlDoc RegExp Tasks Run (Job MailsJob) Problems (Job MailsJob) Modules Scheduler

Contexts

MailsJob

Default Context: Default

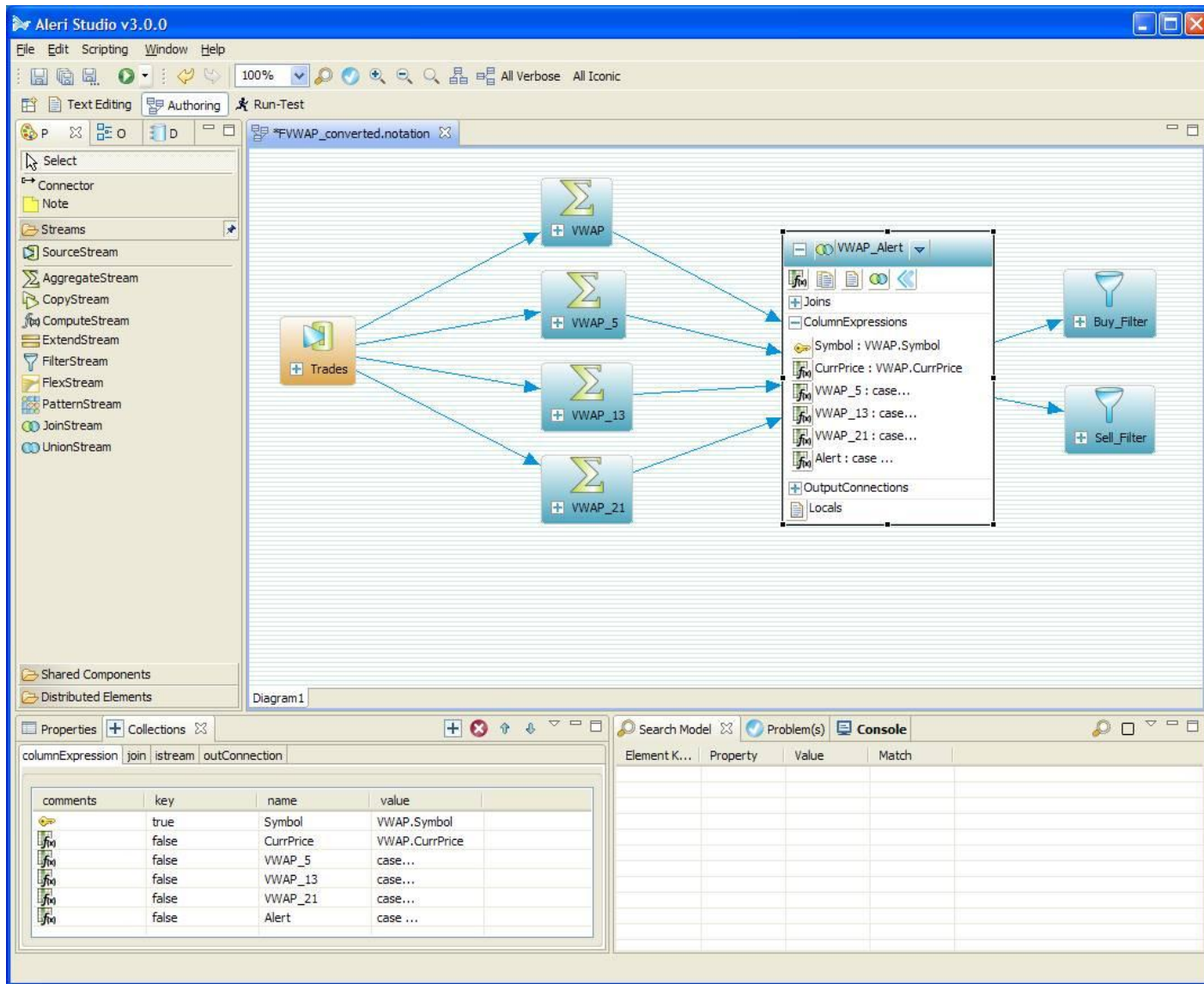
Default

Ask for Confirmation? Parameters: Add Remove Context: Copy Rename Remove

Prompt	Name	Prompt Message	Type	Default	Script code
<input type="checkbox"/>	RepertoireArchiveMail	D:\testFiles\demo\mails?	String	D:\testFiles\demo\mails	\$_context{RepertoireArchiveMail}
<input type="checkbox"/>	RepertoireLogs	D:\testFiles\demo\logs?	String	D:\testFiles\demo\logs	\$_context{RepertoireLogs}
<input type="checkbox"/>	NomFichierLog	log?	String	log	\$_context{NomFichierLog}
<input type="checkbox"/>	NbHeureRappelErreur	6?	String	6	\$_context{NbHeureRappelErreur}

Diagram description: The main workspace shows a Talend job design for 'Job MailsJob'. It features a 'RunBefore' sequence of tPOP_1, tPerl_1, and tDBSQLRow_1. An 'Iterate' loop contains tFileInputMail_1, tMap_1, emailOut (Main), tDBOutput_1, and tWarn_1. A 'Lookup' loop contains tDBInput_1 and row6 (Lookup). A 'Main' loop contains tCatcher_1, row3 (Main), tDBOutput_2, row4 (Main), tFileOutputDelimited_1, row5 (Main), tMap_2, row6 (Main), and tMsgBox_2. The 'Contexts' window is open, showing a table of context prompts and their corresponding script codes.

<http://www.talend.com/>



<http://www.aleri.com/>

The screenshot shows the Eclipse IDE with the SQLstream plugin. The interface includes a menu bar (File, Edit, Navigate, SQL, Window, Help), a toolbar, and a project explorer on the left. The project explorer shows a hierarchy: My SQLstream Instance > Schemas > SALES > WebFeed > Foreign Streams > DisasterTweets. The main workspace contains several data inspection windows:

- Inspect TwitterP...:** A table with columns 'handle' and 'tweet'.

handle	tweet
jennaoneia	Hair day at work today. On year.
nobleexperiment	@Nickf4rr So if I go emo at an AICPA convention, I can cry to you after realizing some bad quarterly results in a post!
Nasiap	Hispanics part of mature U.S. beer market: http://budurl.com/y66q
corysparks	@THEflyGIRL hey you
SonyEurope	Kaka is in the square. We're just about managing to stop him from being mobbed. The place is going wild!
- Inspect GoogleNewsAtom:** A table with columns 'ROWTIME', 'category', and 'title'.

ROWTIME	category	title
2008-12-04 18:00:46.217	U.S.	Ex-FBI agent faces 30 years to life for mob hit - CNN
2008-12-04 18:01:46.714	More Top Stories	Bill Richardson chalks up another Cabinet job for the resume - Los Angeles Times
2008-12-04 18:01:46.714	More Top Stories	Showdown in Hebron as settlers evicted - Jewish Telegraphic Agency
2008-12-04 18:01:46.714	World	Man said he warned Indian cops of activity - United Press International
2008-12-04 18:01:46.714	World	WHO Says Zimbabwe Cholera Estimates May Be Too Low - Voice of America
- Inspect SmallQuakesDay:** A table with columns 'mag', 'location', 'updated', 'point', 'elev', and 'id'.

mag	location	updated	point	elev	id
2.3	western montana	2008-12-03 16:33:10.000	44.7338 -111.1397	-7700	urn:earthq
2.5	Nevada	2008-12-03 19:29:47.000	40.7030 -119.3300	-7600	urn:earthq
2.8	Utah	2008-12-03 20:11:10.000	39.4128 -111.0942	-7000	urn:earthq
3.0	Puerto Rico region	2008-12-03 23:11:25.000	19.2588 -66.3607	-55100	urn:earthq
5.8	near the east coast of Honshu, Japan	2008-12-03 23:16:55.000	38.5488 142.8505	-14000	urn:earthq
3.2	Puerto Rico region	2008-12-03 23:33:15.000	19.4255 -66.3443	-45400	urn:earthq
- DisasterTweets (View):** Shows the SQL definition for the 'DisasterTweets' view in the 'WebFeed' schema.

The bottom of the IDE shows a 'Properties' window with a table:

Property	Value
description	monitor public timelines for tweets about earthquakes, hurricanes, terrorism, tsun...

<http://www.sqlstream.com/>

Market Analysis

OSBI projects/vendors are providing both tools and solutions.

OSBI is gaining end-user visibility.

There are now credible user stories showing robustness, scalability, reliability.

Alliances are breaking out of the open-source/small-shop world.

OSBI - and OS for BI - are happening!

Questions?

Thanks!

From Data to Intelligence in 17 Clicks

*Business Intelligence for the
Open Source Development Community*

Seth Grimes

Alta Plana Corporation

Alta Plana

Scott Rosenbaum

Innovent Solutions



March 26, 2009