Cost-Effective Business Intelligence with Red Hat and Open Source

Sherman Wood
Director, Business Intelligence, Jaspersoft
September 3, 2009

presented by

Agenda

- Introductions

- Quick survey

- What is BI?: reporting, analysis, data integration

- BI is important for IT!

- BI Architectures with Red Hat: use cases, demonstrations, state of the market, open source options

    - Database: Marts and Warehouses

    - Data Integration and Map Reduce

    - Visualization: analytics, end user interactivity, in-memory, dashboards

    - Embedding BI in applications

- Q & A

    -

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT

Introductions and a Quick Survey

- Me, Jaspersoft

- Jaspersoft and Red Hat

- Platform provider (Ops, Sys admin, Uses RHEL) vs Software Vendors (Developer on RHEL) vs End Users

- Using enterprise software suites: SAP, Oracle

- Please ask questions as we go

  - Or wait for the Q&A at the end

- Anything you specifically want covered?

-

RED HAT :: CHICAGO :: 2009
SUMMIT

Priority of Business Intelligence

2009 CIO Technology Priorities
To what extent will each of the following technologies be a top five priority for you in 2009?

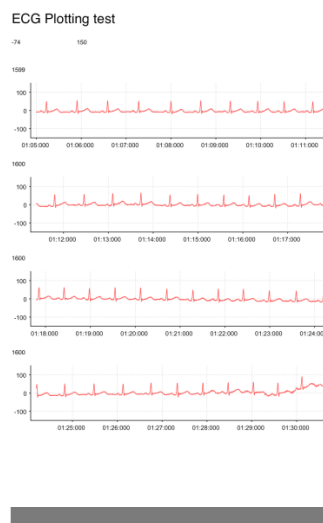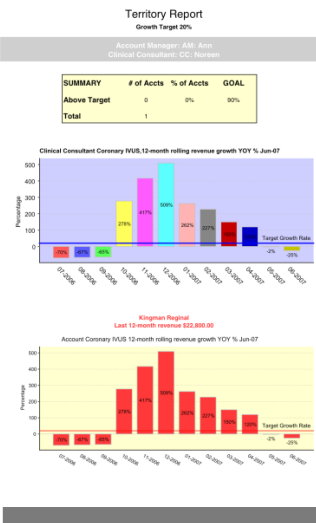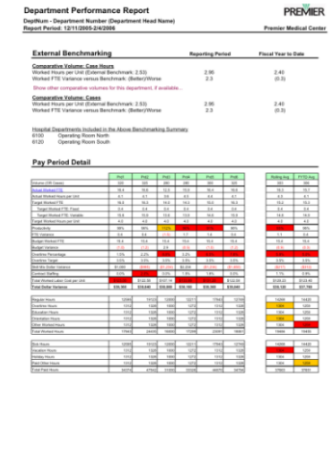|	Rank 2009		Rank 2008	Rank 2007
BI	1	←	1	1
Enterprise Applications (ERP, SCM and CRM)	2	←	2	2
Server and Storage Technologies (Virtualization)	3	←	3	5
Legacy Application Modernization	4	←	4	3
Collaboration Technologies	5	▲	8	10
Networking, Voice and Data Communications (VoIP)	6	▲	7	4
Technical Infrastructure	7	▼	6	8
Security Technologies	8	▼	5	6
Service-Oriented Applications and Architecture	9	▲	10	7
Document Management	10	▼	9	9

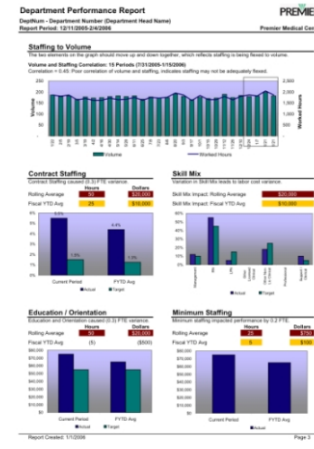Top-searched terms on gartner.com in 1Q09:

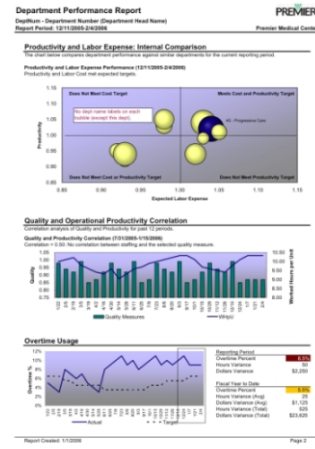1. "Magic Quadrant"

2. "Cloud Computing"

3. "Business Intelligence"

Note: Number of respondents = 1,500.
Source: 2008 Gartner Executive Programs' CIO Survey, January 2009

7

Gartner.

RED HAT :: CHICAGO :: 2009
SUMMIT

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

All Kinds of Reporting and Analysis

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

Demonstration

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT
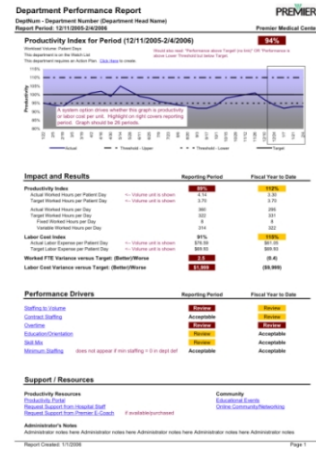
Traditional Business Intelligence

Reporting

Analytics

"BI"

Data Sources: Anything!

Data Integration

Clean Data (ODS)

Database

Data Mart/ Warehouse

RED HAT :: CHICAGO :: 2009
SUMMIT
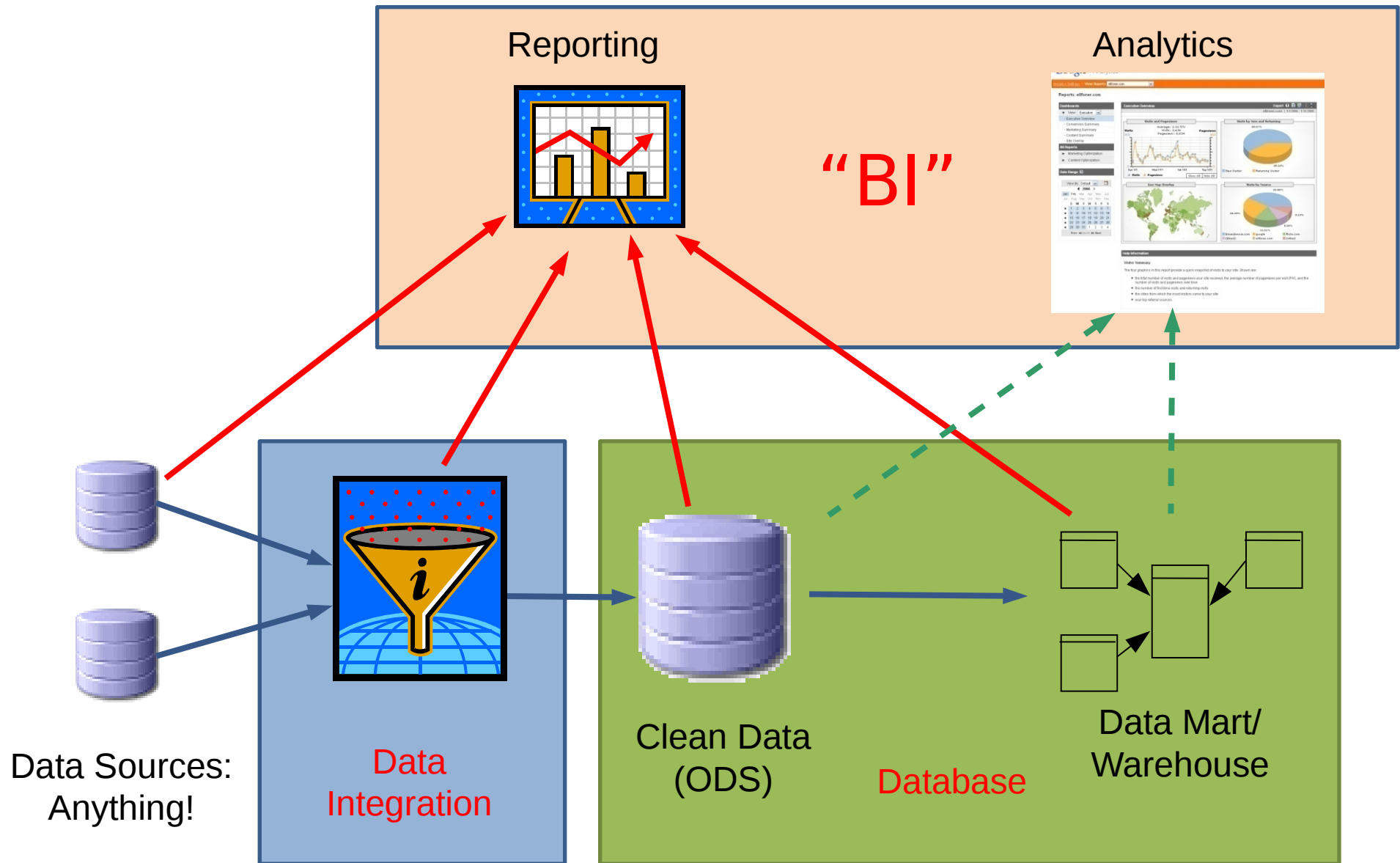
Embedded Business Intelligence

- Add BI to an existing application

- Aiming for seamless integration

    - Single sign on

- Often a single data source

- Typical for application developers and software vendors

- Examples: Red Hat Satellite, Virgin Money

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT

Stand Alone BI

- Within an organization, across groups/departments

- The analytic application

- Production reporting

- Often multiple data sources

- People come directly to the BI application

- Data warehouses and marts: star schema

- Typical for consultants/system integrators, internal IT groups
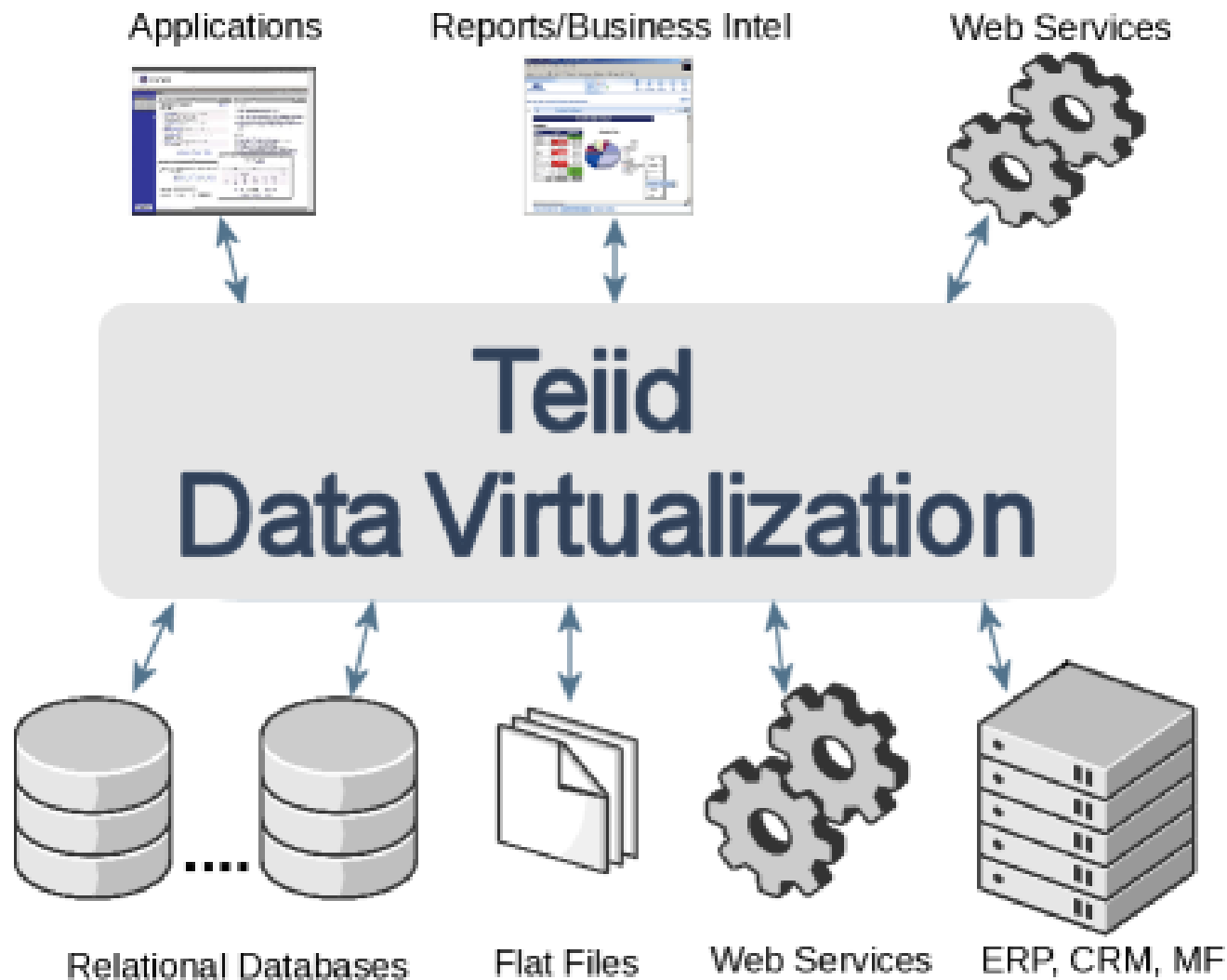
- Examples:

-

Red Hat and Open Source in BI

- Open source alternatives to the commercial vendors at all levels in the BI stack

- Linux is the OS of choice in non-Microsoft environments and RHEL is the leading commercial distribution

- Often see complete open source solutions: operating system, database, data integration, reporting and analysis

RED HAT :: CHICAGO :: 2009
SUMMIT

Open Source, "Free", Low Cost BI "Databases"

Optimized for data load and query

Family	
"Express" versions	• Oracle, Microsoft, DB2 • Limited capabilities – will need to upgrade
Postgres	• EnterpriseDB – commercial open source – distributed processing • Greenplum – commercial – compression, distributed processing
MySQL	• MySQL base engines – MyISAM – limited • Infobright – commercial open source - column store, compression • Kickfire - appliance
Vertica	• Column store, compression, distributed processing
Ingres	• VectorWise • Column store, compression, max out the CPU – not I/O • Out this year
Red Hat/JBoss	• Teiid • Virtual database – performance vs infrastructure overhead

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT

An Data Warehouse Alternative: JBoss Teiid

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT

Data Integration

- Commercial

  - Informatica

  - IBM - Ascential (DataStage)

- Open Source

  - Custom Scripting

  - Talend/Jaspersoft

  - Kettle

  - And Hadoop!

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT

Talend/Jaspersoft Data Integration

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft
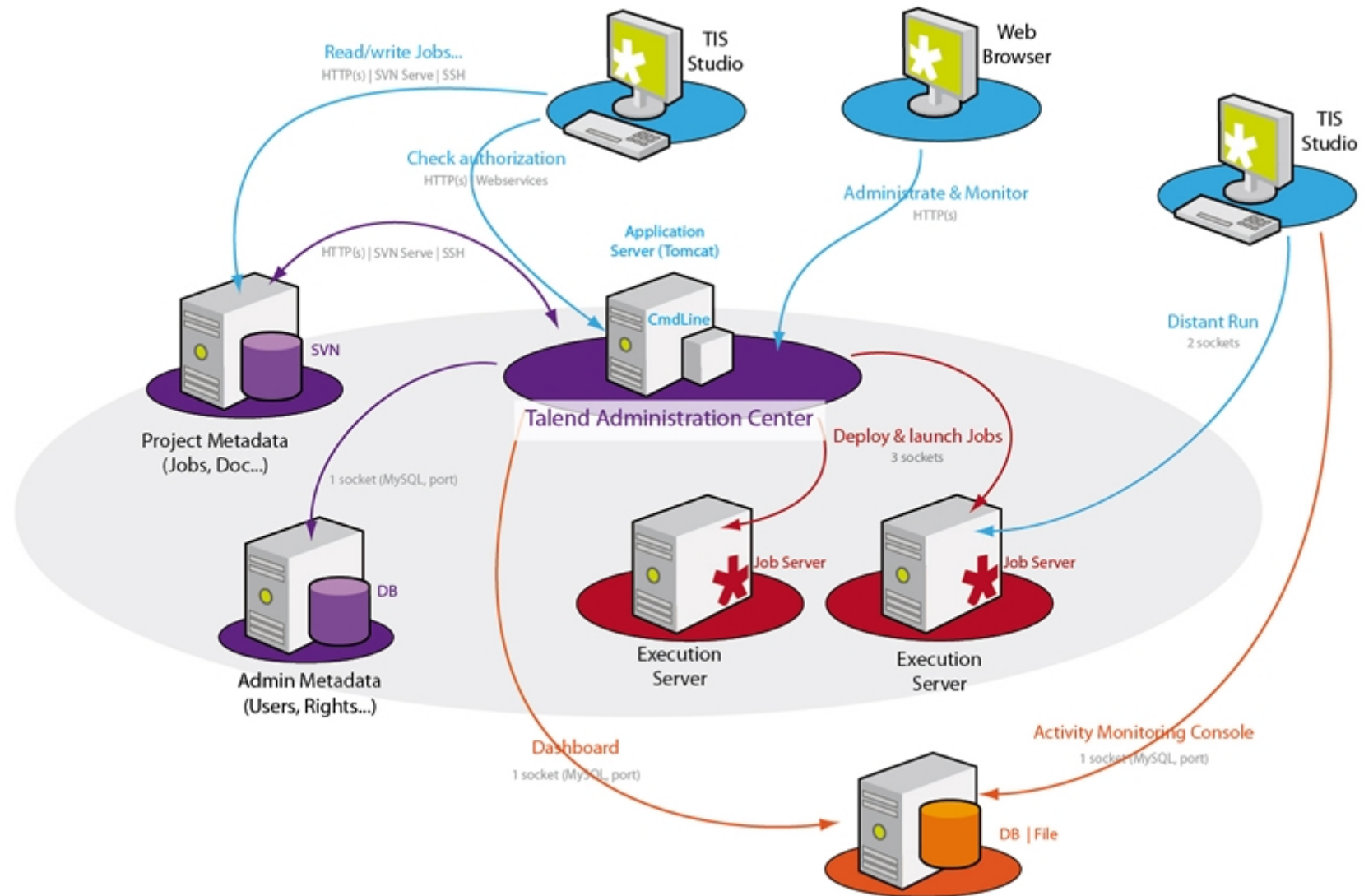
Map Reduce and Distributed File Systems

- Map Reduce algorithm: dealing with massive data

    - Google, Yahoo: how do I index the Web?

    - Batch processing: data analyst workbench to production

    - Different mind set to traditional data integration addressing a different class of problem

    - Originally the big web companies, now moving to broader use and embedding in databases: Asterdata, Greenplum, HadoopDB

- Hadoop provides open source frameworks and eco-system for Map Reduce

    - Hadoop file system (HDFS): extreme clusters with redundancy, failover

        - Kosmos File System too

    - Base shell scripting and Java for transformations

    - Higher level interfaces: Hive (SQL), Pig, Cascading

Map Reduce Examples

- Yahoo Hadoop, April 09: Sort 1 TB in 62 secs

    - Approx. 3800 nodes (in such a large cluster, some nodes are always down)

    - 2 quad core Xeons @ 2.0ghz, 8GB RAM per node

    - 4 SATA disks per a node

    - 1 gigabit ethernet on each node, 40 nodes per a rack, 8 gigabit ethernet uplinks from each rack to the core

    - Red Hat Enterprise Linux Server Release 5.1 (kernel 2.6.18)

    - Sun Java JDK 1.6.0 13-b03 (32 and 64 bit)

- Google does same in 63 secs with 1000 nodes, November 08

    - Google File System, not open source

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009 SUMMIT

BI and Map Reduce

- BI tools can access Map Reduce in several ways

- Push Map Reduce results into a traditional relational database

  - No different from any traditional BI/data warehouse process

  - Avoids Map Reduce performance issues for users

  - Limits flexibility of analysis

- Access Map Reduce directly

  - Access raw results on the file system or via APIs

  - Not useful for interactivity
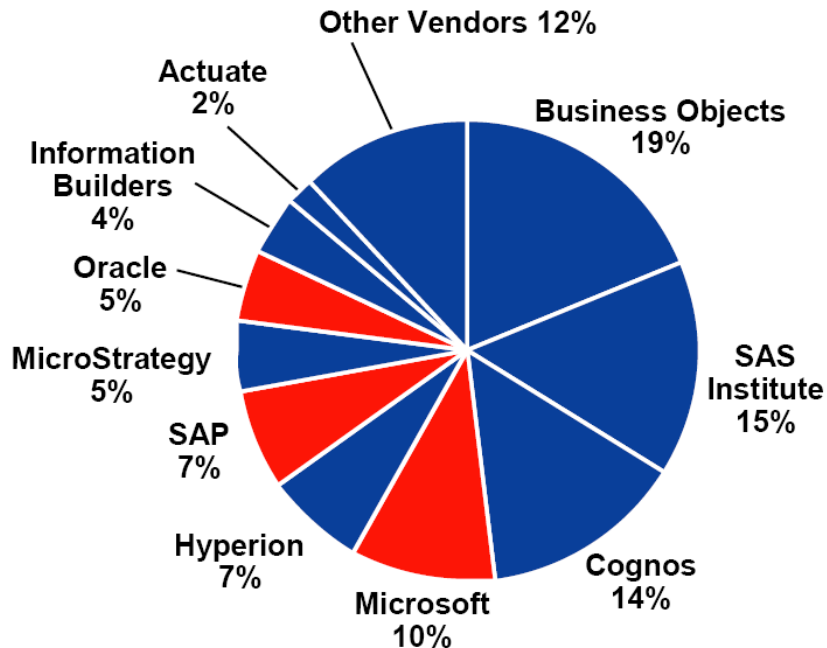
  - Data analysts have more flexibility

SOURCE: Google

RED HAT :: CHICAGO :: 2009
SUMMIT
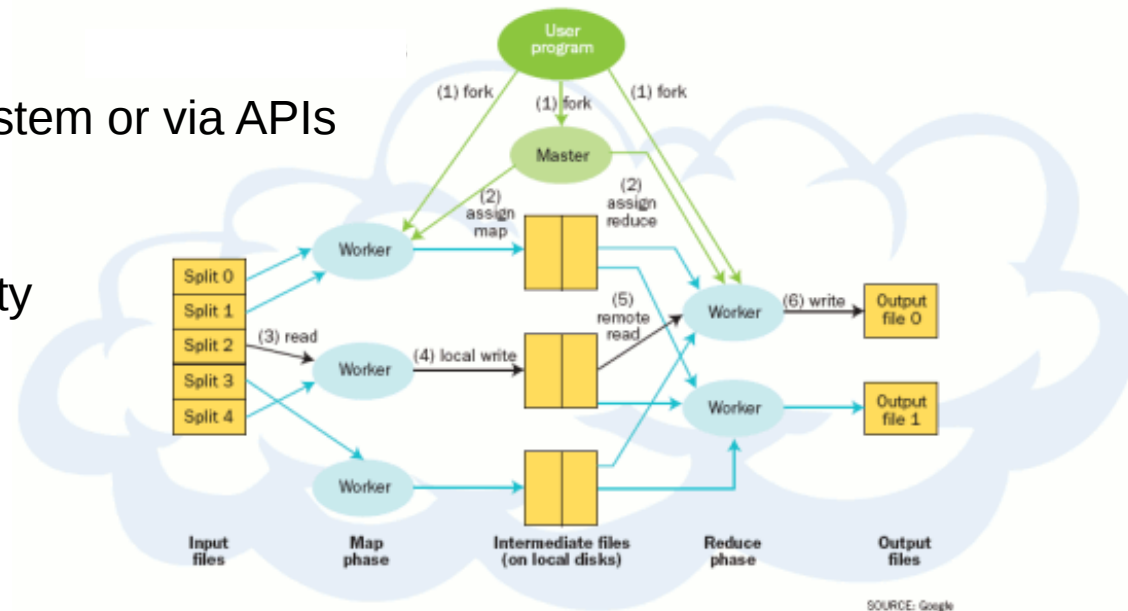
BI Market and Vendors

$5.74 Billion Market Worldwide (License and Maintenance) —
Market Share by Vendor Based on 2008 Total Software Revenue

Before 2007 Consolidation Activity

Stack Vendors Are Shown in Red

BI Platform Market Share at the Beginning of 2007

BI Platform Market Share, 2008

Other Vendors 12%
Actuate 2%
Information Builders 4%
Oracle 5%
MicroStrategy 5%
SAP 7%
Hyperion 7%
Microsoft 10%
Business Objects 19%
SAS Institute 15%
Cognos 14%

QlikTech 2%
Other Vendors 11%
Actuate 2%
Information Builders 3%
MicroStrategy 5%
Microsoft 11%
Oracle 11%
IBM 13%
SAP 28%
SAS Institute 14%

9

Gartner.

RED HAT :: CHICAGO :: 2009
SUMMIT
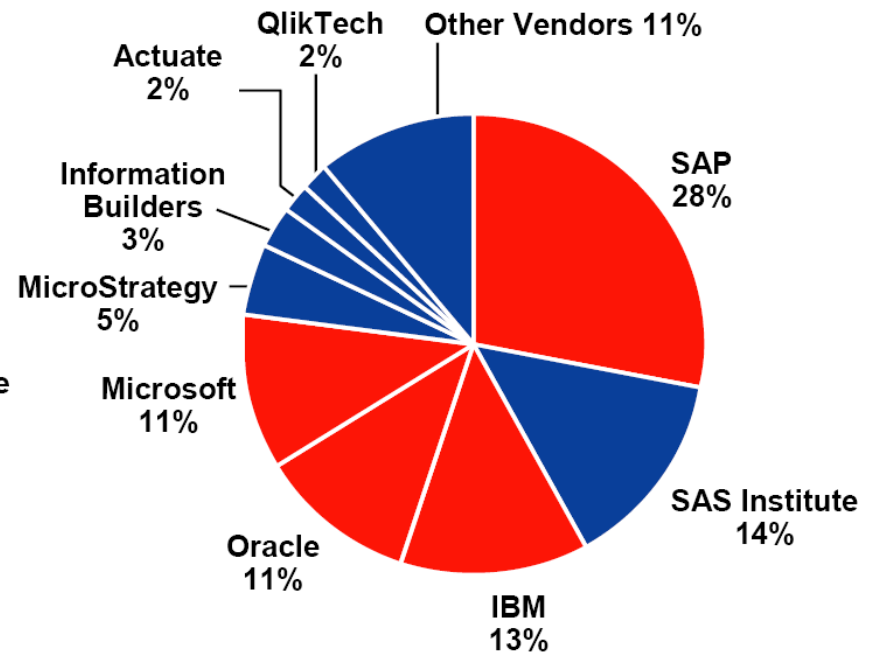
What drives cost of BI?

- Complexity of BI and DW technologies

    - Complicated BI tools drive high end user support costs

    - Most DWs require large complex/specialized hardware and constant tuning and configuration

    - Complex ETL processes prevent IT from keeping up with increasing data access needs

- Cost of software, hardware and per-user licenses

- Difficulty supporting new applications or decisions with limited resources

- RESULT: Only 15% of today's enterprise users have access to Business Intelligence tools

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT

Costs – Reporting and Analysis

- Commercial vendors tend to:

  - Charge per seat

  - Limit the number of users per server or CPU: 20 users in one case

  - High up front license cost

  - Pay for completely new licenses to add functionality

  - 22% annual maintenance fee!

  - Lock in, particularly now with all the industry consolidation

  - Don't play well with partners, which are often needed in BI

RED HAT :: CHICAGO :: 2009
SUMMIT

Reporting and Analysis License Cost Comparison

-
-
- Customer Comparison:
Jaspersoft vs. Business Objects
-
-
-
-
- 3-year TCO comparison from
BIScorecard

Capabilities	Business Objects Enterprise*	JasperServerPro
Operational Reporting Enterprise Reporting (no load balancing)	Upfront: $215,000 Annual: $47,300	Annual: $46,200
Operational Reporting Enterprise Reporting (load balancing)	Upfront: $417,500 Annual: $91,850	Annual: $46,200
Operational Reporting Enterprise Reporting Query and Analysis Dashboards (load balancing)	Upfront: $1,250,800 Annual: $275,176	Annual: $46,200

*Business Objects Enterprise estimates using CPU licenses for the production environment and Named User licenses for the preproduction environments. It does not include the cost for desktop licenses of the Report or Dashboard design tools.

Scenario	Traditional BI	JasperSoft Premium Support License
20-user SMB	$40,421	$42,000
100-user departmental BI	$216,680	$93,000
1000-user enterprise BI	$1,728,662	$186,000

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT

Costs – Data Integration

- Example ETL project: daily load of a data warehouse with data contained in three different RDBMS and a CRM application, and data provided by Web Services

-
-

Approach	1 year cost	5 year cost	Implementation Time
Manual Coding	$184,000	$284,000	24 weeks
Proprietary Data Integration	$581,500	$955,500	10 weeks
Open Source Data Integration	$62,000	$106,000	9 weeks

-
-

- Data integration sucks $s: 50-90% of budget and time for implementation

  - Can you use standard interfaces for applications and data?

RED HAT :: CHICAGO :: 2009
SUMMIT

Costs – Database

	Vertica 2.5	Oracle 10gR2 EE
Hardware	5 HP ProLiant DL 580 (16 cores, 2TB disk, 32 GB)	1 HP Integrity rx8640 2 EMC CX3-40
Server(s)	$79,085	$450,108
Storage	$0 (in servers)	$600,000
Data center – power, cooling	$4,800 / year	$43,200 / year
Data center - space	$6,600 / year	$6,600 / year
Implementation	$10,000	$49,000
Year 1 total	**$100,485**	**$1,148,908**

30TB warehouse for a telco. List prices. Vertica performed a magnitude faster than the Oracle configuration.

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT

Bottom Line

- Business Intelligence remains the highest priority for technology groups

- Vendor consolidation is making costs rise for commercial BI

- The data warehouse market has many new entrants and approaches for dealing with large data volumes, like:

  - JBoss Teiid

  - Compressed column data stores

  - Cloud and virtualized configurations

  - Map Reduce, massively distributed file systems

  - Open source alternatives

- Open source business intelligence, data integration and databases running on Red Hat provide excellent TCO compared with commercial alternatives

- Open source business intelligence and data Integration rely heavily on Java

RED HAT :: CHICAGO :: 2009
SUMMIT

QUESTIONS?

TELL US WHAT YOU THINK: REDHAT.COM/SUMMIT-SURVEY

Agenda

- Business Intelligence (reporting and data analysis) has been at the top of IT's to do list for over 10 years, but has remained expensive to purchase and deliver. This session will show configurations of Red Hat, Jaspersoft's commercial open source BI suite and other open source technologies to provide business intelligence services for organizations and individual applications at a fraction of the cost of proprietary options

-

RED HAT :: CHICAGO :: 2009

SUMMIT

Jaspersoft

- Over 8 million total open source project downloads

    - JasperReports – Java reporting library – 2001

    - iReport – graphical report designer – 2002

    - JasperServer/JasperAnalysis – business intelligence web server – 2005

- Estimated 100,000+ production deployments (WW)

- Over 90,000 registered developers in Community

- Worldwide development, offices in San Francisco and Dublin

- More than 9,000 paying customers (96 countries)

- 75% year-over-year growth

-

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

Reporting vs Analysis

- Reporting

  - Tends to be static, WYSIWYG/print focused

  - Same view over time: Production and scheduled reporting

- Analysis

  - Statistical and aggregate operations

  - Data exploration: interactivity

  - Data mining: predictive analysis, finding patterns

  - Convert analysis insights into actions with reports

- Process focused BI

  - Business/Corporate Performance Management: budget vs actuals

  - Business Activity Monitoring: "real time" analysis of processes

Cost Effective BI with Red Hat | Sherman Wood, Jaspersoft

RED HAT :: CHICAGO :: 2009
SUMMIT