TRANSFORM BIG DATA INTO
ACTIONABLE INFORMATION

Make Big Data Available for Everyone

Syed Rasheed Solution Marketing Manager

January 29th, 2014
Agenda

- Demystifying Big Data
- Data Challenges Getting Bigger
- Red Hat Big Data Strategy and Platform
- Data Virtualization Overview
- Customer Use Case for Big Data integration using Data Virtualization
- Demo
- Q&A
Demystifying Big Data

- Many definitions
  - 3 V’s = Volume, Velocity, Variety
  - 4 V’s = 3V’s + (Value or Veracity or Variability)
  - Forrester: Big Data is the frontier of a firm’s ability to store, process, and access all the data it needs to operate effectively, make decisions, reduce risks, and serve customers.
  - Gartner: Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.
A More Personal Definition

Big Data = All Your Data
Business Objective

*Data is becoming the new raw material of business: an economic input almost on a par with capital and labor. “Every day I wake up and ask, ‘how can I flow data better, manage data better, analyze data better?’”*

CIO - Wal-Mart
Data Challenges Getting Bigger

Existing Data Integration approaches are not sufficient

- Extracting and moving data adds latency and cost
- Every project solves data access and integration in a different way
- Solutions are tightly coupled to data sources
- Poor flexibility and agility

**How to align?**
Leading to Information and Agility Gap

Only 28% Users have any meaningful data access

Over 70% BI project efforts lies in Data Integration – finding and identifying source data

Need for Agility:
- 65% Constantly changing business needs
- 57% IT’s inability to satisfy new requests in a timely manner
- 54% The need to be a more analytics-driven organization
- 47% Slow and untimely access to information
- 34% Business user dissatisfaction with IT-delivered BI capabilities
Red Hat’s Big Data Strategy

- Reduce Information Gap thru cost effectively making **ALL** data easily consumable for analytics

![Data to Actionable Information Cycle Diagram](image-url)
Red Hat Big Data Platform
Better Data Integration Approach Required

Data Virtualization software virtually unifies data spread across various disparate sources; and makes it available to applications as a single consolidated data source.
Turn Fragmented Data into Actionable Information

**Data Consumers**
- BI Reports & Analytics
- Mobile Applications
- ESB, ETL
- SOA Applications & Portals

**JBoss Data Virtualization**
- Consume
  - Standard based Data Provisioning
  - JDBC, ODBC, SOAP, REST, OData
- Compose
  - Unified Virtual Database / Common Data Model
  - Unified Customer View
  - Unified Product View
  - Unified Supplier View
- Connect
  - Native Data Connectivity

**Data Sources**
- Hadoop
- NoSQL
- Cloud Apps
- Data Warehouse & Databases
- Mainframe
- XML, CSV & Excel Files
- Enterprise Apps

**Easy, Real-time Information Access**
- Design Tools
- Dashboard
- Optimization
- Caching
- Security
- Metadata

**Virtualize Abstract Federate**
- Siloed & Complex

**Virtualize Abstract Federate**
- Siloed & Complex

**Virtualize Abstract Federate**
- Siloed & Complex
JBoss Data Virtualization – Use Cases

- Simplifying BI development
- Transparent access to cloud-based data
- Easy development of data services (SOAP and REST)
- Untangling enterprise data with easy access
- Migration of data
- And many more …

DATA SOURCES
- Oracle DW
- SAP
- Hadoop
- Salesforce.com

DATA CONSUMERS
- BI Reports
- SOA Applications

Virtual Consolidated Data Source

Data Virtualization Software
- Consume
- Compose
- Connect

Easy, Real-time Information Access

Virtualize Abstract Federate

Siloed & Complex
# JBoss Data Virtualization: Supported Data Sources

<table>
<thead>
<tr>
<th>Enterprise RDBMS:</th>
<th>Hadoop:</th>
<th>NoSQL:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Oracle</td>
<td>• Apache</td>
<td>• JBoss Data Grid</td>
</tr>
<tr>
<td>• IBM DB2</td>
<td>• HortonWorks</td>
<td>• MongoDB</td>
</tr>
<tr>
<td>• Microsoft SQL Server</td>
<td>• Cloudera</td>
<td>• More coming...</td>
</tr>
<tr>
<td>• Sybase ASE</td>
<td>• More coming...</td>
<td></td>
</tr>
<tr>
<td>• MySQL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• PostgreSQL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Ingres</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enterprise EDW:</th>
<th>Office Productivity:</th>
<th>Technology Connectors:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Teradata</td>
<td>• Microsoft Excel</td>
<td>• Flat Files, XML Files,</td>
</tr>
<tr>
<td>• Netezze</td>
<td>• Microsoft Access</td>
<td>XML over HTTP</td>
</tr>
<tr>
<td>• Greenplum</td>
<td>• Google Spreadsheets</td>
<td>• SOAP Web Services</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specialty Data Sources:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• ModeShape Repository</td>
<td>• Mondrian</td>
</tr>
<tr>
<td>• Mondrian</td>
<td>• MetaMatrix</td>
</tr>
<tr>
<td>• MetaMatrix</td>
<td>• LDAP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enterprise &amp; Cloud Applications:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Salesforce.com</td>
<td>• More coming...</td>
</tr>
<tr>
<td>• SAP</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoSQL:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• JBoss Data Grid</td>
<td>• More coming...</td>
</tr>
<tr>
<td>• MongoDB</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Technology Connectors:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Flat Files, XML Files,</td>
<td>• SOAP Web Services</td>
</tr>
<tr>
<td>XML over HTTP</td>
<td>• REST Web Services</td>
</tr>
<tr>
<td></td>
<td>• OData Services</td>
</tr>
</tbody>
</table>
Better Together - Big Data and Data Virtualization
Hadoop should not be another Silo

- Data virtualization ...
  - unleashes the full value of Big Data for analytics
  - speeds up development on big data sources
  - offers an evolutionary adoption of big data
  - makes big data available to everyone – higher big data ROI
Better Together - Big Data and Data Virtualization
Capture, Process and Integrate Data Volume, Velocity, Variety

**Integrate & Analyze**
- BI Analytics (historical, operational, predictive)
- SOA Composite Applications

**Data Integration**
- JBoss Data Virtualization

**In-memory Cache**
- JBoss Data Grid

**Messaging and Event Processing**
- JBoss A-MQ and JBoss BRMS

**Structured Data**

**Streaming Data**

**Semi-Structured Data**

**Red Hat Storage**

**Red Hat Enterprise Linux & Virtualization**
Examples

RED HAT BIG DATA PLATFORM IN THE REAL WORLD
Big Data in Telco

Objective:
- Near Real-time Network Performance Visibility and Call Detail Record Analysis

Problem:
- Operator has a large amount of data being generated continuously in the network, but this data cannot be effectively utilized for performance issue detection
- Most analytics apps operate by post-processing Call Detail Records (CDRs) because the data cannot be collected in real-time

Solution:
- Cirries adapters with Red Hat messaging technologies can filter the information and route the right data to the right application as well as deposit it in a cloud-based data store/grid for data mining and predictive analytics
Big Data to Smart Data

Network Management System

Network Application

Analytic Systems

“Smart Data” Consumers

Data Route & Load Balance

Data Storage, Index & Search

Queuing & Distribution

Data Collect, Extract, Normalize & Filter

“Big Data” Network

Real-time Data Engine

RHEL/RHEV

Red Hat JBoss AMQ

Real-time Data Engine

RHEL/RHEV

Real-time Data Engine

RHEL/RHEV

3G

LTE

IP

Local / Regional

Cloud
Big Data in the Utilities

- Objective:
  - Combine data from smart meters on homes with data from electricity generation and transmission and make it available to power providers

- Problem:
  - The original smart grid project looked only at reading information from the meters on houses and now this data needs to be combined with generation and transmission data in a cost-effective way
  - The data points are all over the place: sensors on the lines, in the field, homes, etc.
  - The information must be accessible to multiple power providers through a common interface

- Solution:
  - Use Messaging to collect data from a variety of sources and route it to a CEP for initial filtering. Process with Hadoop map/reduce and BRMS and distribute data to Data Virtualization to be combined with other sources and consumed with BI tools, and/or to JDG for in-memory data caching and/or send to archive.
Smart Grid

Regulatory

Users

Compose

Authentication

Presentation

REST Exposure

Offline Storage

Data Virtualization

Cache

Normalization / MapReduce

NoSQL-Cassandra

PM Regional Translator / Scheduler

Adaptor Rules

Sensor Adaptor

Routing Function

Collector Sensors

Local Data Store

Collector Scada

Local Data Store

Collector Meter

Local Data Store

Transmission

Generation

Consumer

API Exposure & Portal Tier

Data Tier

Normalized Data Tier

Data Adaptation & Routing Tier

Element Connection Tier

Rules Creation / Updates

PM Data Schedule

PM Data Reports

PM Admin

Compose

Rules Creation / Updates

PM Data Schedule

PM Data Reports

PM Admin
Retail Customer Use Case
Gain Better Insight from Big Data for Intelligent Inventory Management

- Objective:
  - Right merchandise, at right time and price

- Problem:
  - Cannot utilize social data and sentiment analysis with their inventory and purchase management system

- Solution:
  - Leverage JBoss Data Virtualization to mashup Sentiment analysis data with inventory and purchasing system data. Leveraged BRMS to optimize pricing and stocking decisions.
Consider...

How would your organization change...

- If data were **readily reusable in place** rather than requiring significant effort to build new intermediary data tiers?
- If data could be repurposed **quickly** into new applications and business processes?
- If all applications and business processes could get **all of the information needed** in the form needed, where needed and when needed?
Red Hat JBoss Middleware

ACCELERATE

INTEGRATE

AUTOMATE

Development

Tools

- JBoss Developer Studio

User Interaction

- JBoss Portal

Business Process Management

- JBoss BRMS
- JBoss BPM Suite

Application Integration

- JBoss A-MQ
- JBoss Fuse
- JBoss Fuse Service Works

Data Integration

- JBoss Data Virtualization

Foundation

- JBoss EAP
- JBoss Web Server
- JBoss Data Grid

Management

Tools

- JBoss Operations Network
Why Red Hat for Big Data?

- Transform **ALL** data into actionable information
  - Cost Effective, Comprehensive Platform
  - Community based Innovation
  - Enterprise Class Software and Support

Data to Actionable Information Cycle

Data

Capture | Process | Integrate

Analytics
Red Hat Big Data Platform

- RHEL Platform Integration & Optimization
- Hadoop Integration
- JBoss Data Virtualization
- Hadoop Distributions
- Hadoop On Red Hat Storage
- Hadoop On OpenStack
- Fedora Big Data SIG

Middleware

Storage

Cloud / Virtualization
Thank You

Q&A