

REAL TIME DATA WITH JBOSS DATA VIRTUALIZATION AND DATA GRID

Erica Langhi
Solution Architect / Red Hat

STALE OR FRESH BREAD ?



DIGITAL TRANSFORMATION



Mobile



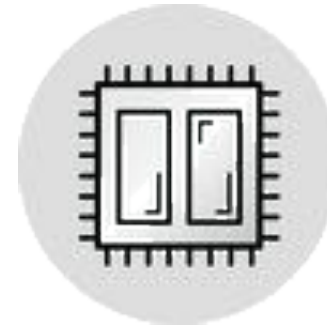
Big Data



Cloud



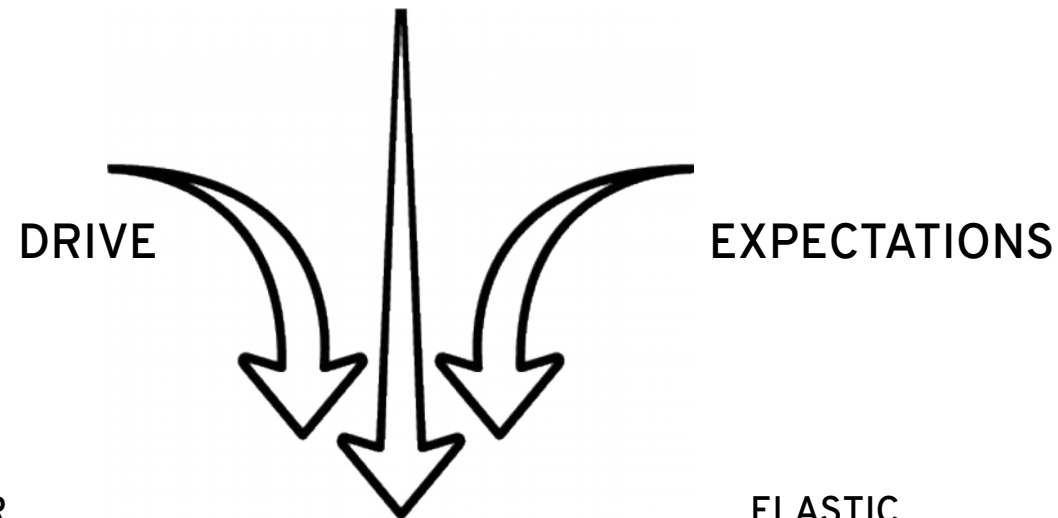
Social



Memory



Network



**BETTER
ANALYTICS**
Real-time analytics

**FASTER
PERFORMANCE**
Sub-seconds not hours

**ELASTIC
SCALABILITY**
On-demand Scale-out architecture

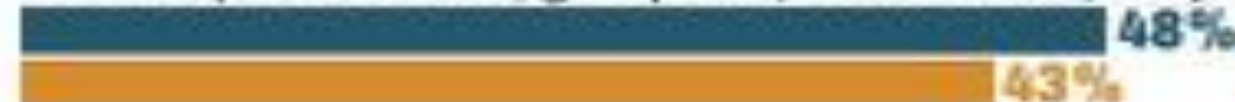
**HIGHER
AVAILABILITY**
Always available

DATA INITIATIVES

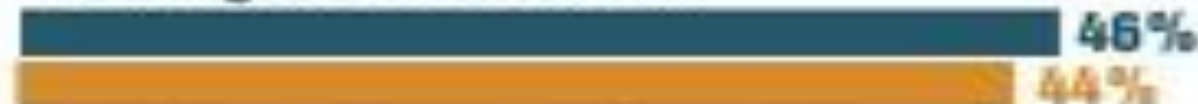
What data sources or challenges are driving, or would drive, your organization's interest in doing big data analysis?

■ 2015 ■ 2014

Finding correlations across multiple, disparate data sources (clickstreams, geospatial, transactions, etc.)



Predicting customer behavior



Predicting product or service sales



Predicting fraud or financial risk



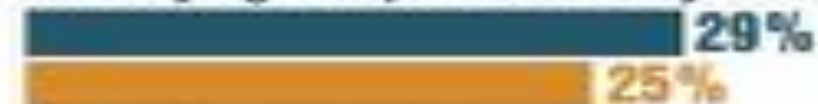
Analyzing social network comments for consumer sentiment



Analyzing high-scale machine data from sensors, web logs, etc.



Identifying computer security risks



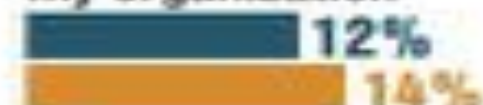
Analyzing web clickstreams



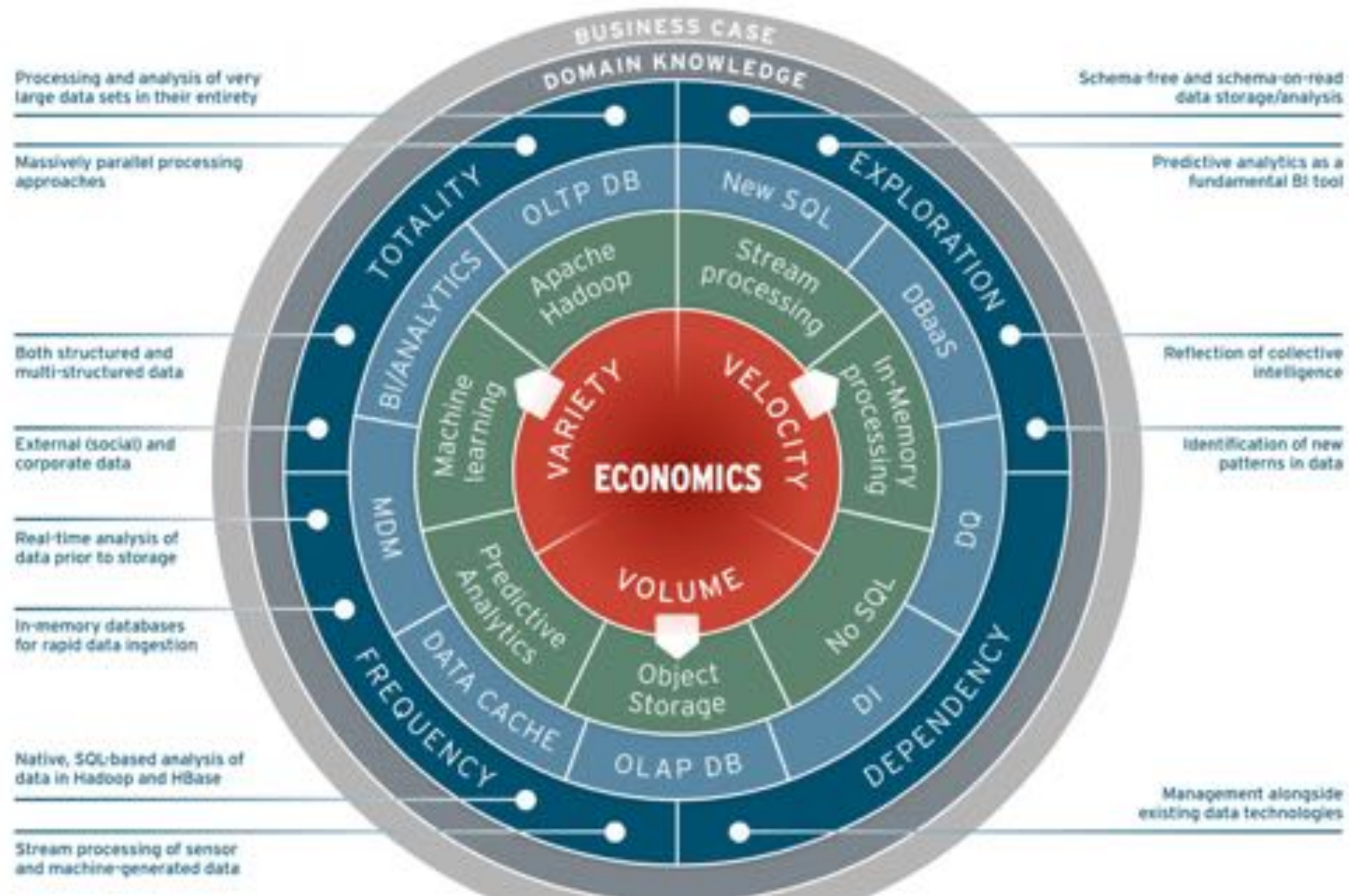
Other



Big data analytics is not of interest to my organization



TOTAL DATA



Source : 451 Research, 2016

DATA CHALLENGES

Big Data

- 3 Vs

Cloud

- Elastic and hybrid data

High Performance

- Data access and processing

No compromise

- Always available and secure

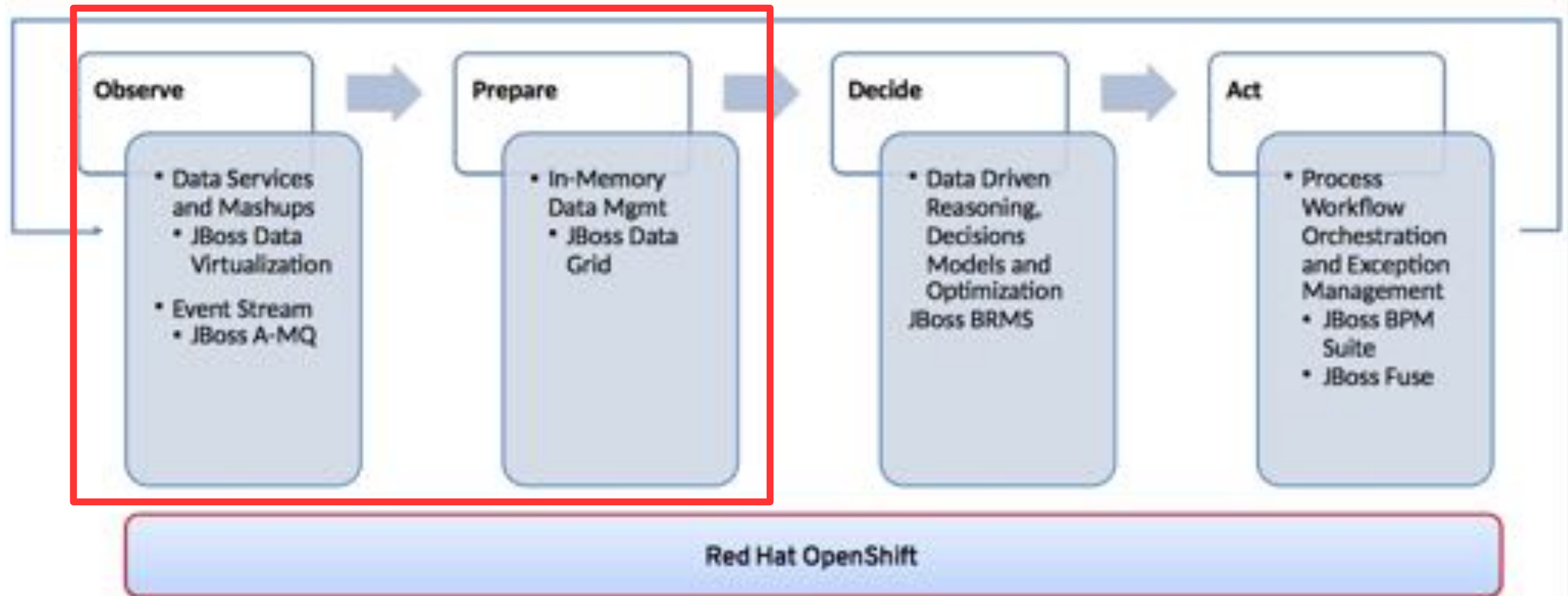
Reduction

- Cost, Risk and Data Sprawl

RED HAT PORTFOLIO



DATA FABRIC



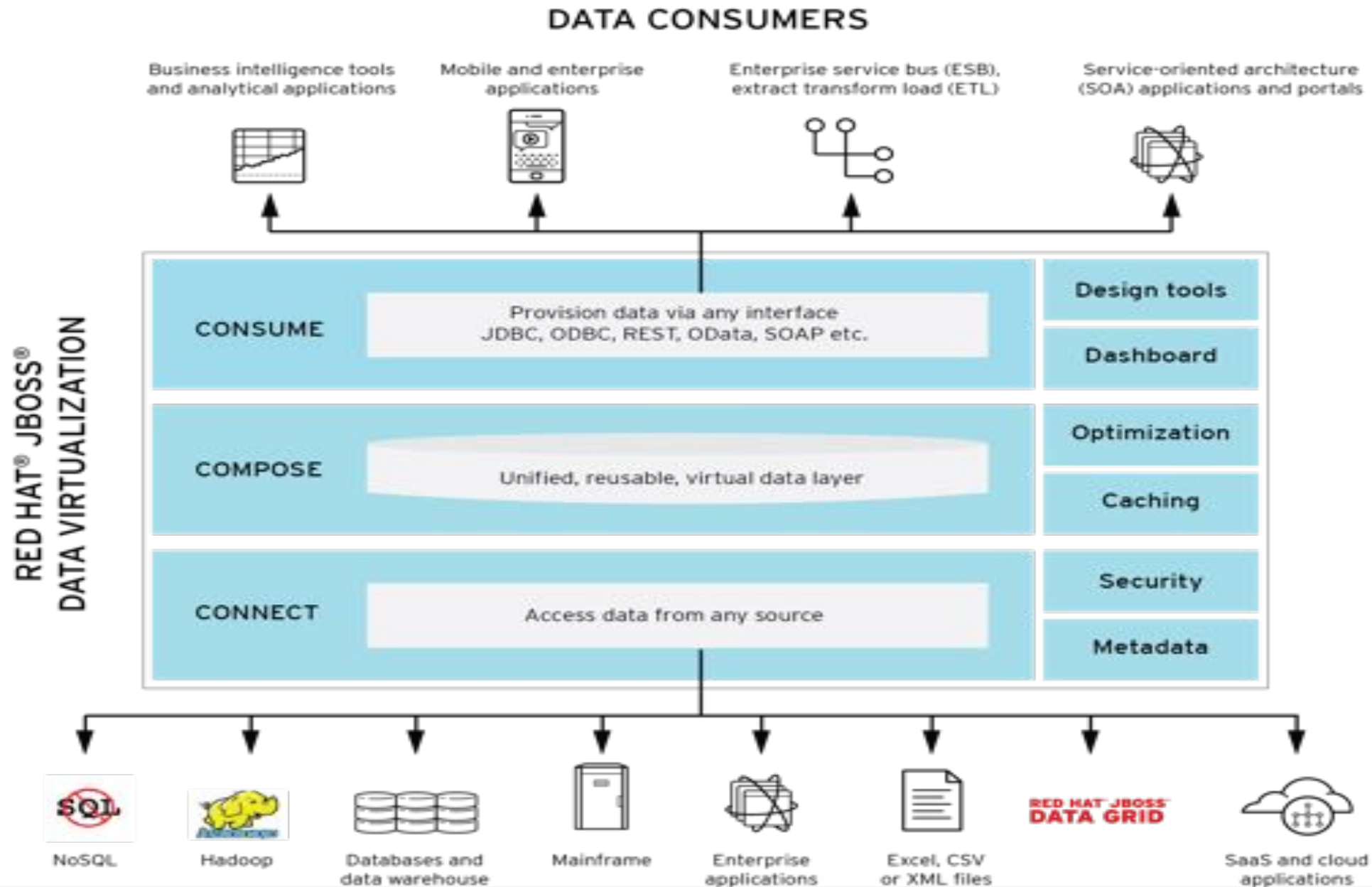
WORK SMARTER – INSIGHTS YOU NEED WITHIN THE APPLICATION YOU USE

JBoss DATA VIRTUALIZATION

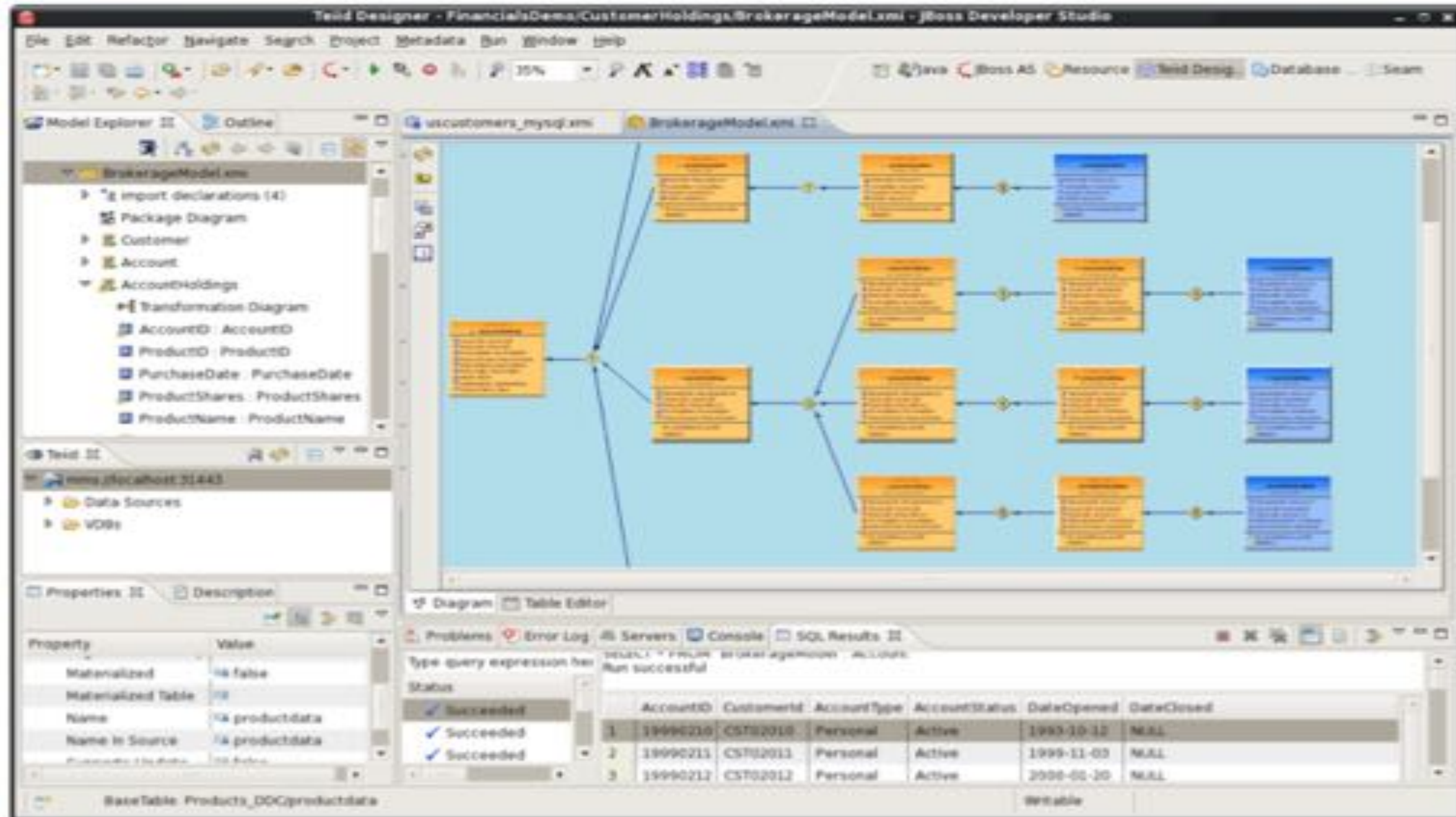
PREPACKAGED OR FRESHLY MADE SANDWICH?



DATA VIRTUALIZATION



MODEL DRIVEN DEVELOPMENT



VISUALISE YOUR DATA



ENHANCED DATA PROTECTION

Edit VDB Data Role
Select Finish to save data role

Name: ReadWrite

Description:
Allow ReadWrite operations to authenticated user

☐ Allow usage of temporary tables
☐ Apply this role to All Users

Mapped Role Names
superuser

Permissions

Model Conditions Masking Allowed Languages

Model	Security	Create	Read	Update	Delete	Execute	Alter
MarketData_csvFiles.xml	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
MarketDataView.xml	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
AccountHoldingsView.xml	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Accounts_MySQL.xml	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
AccountsView.xml	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

System Tables Access

☒ Allow this role to access SYSADMIN model

☐ READ ☐ EXECUTE

Cancel Finish

- Prevent unnecessary data exposure
- Apply role based data access to virtual database in addition to security at physical sources
- Auditing
- Row and column masking
- Centralised security policies management

PERFORMANCE OPTIMISATION



CACHING & MATERIALISATION

Multiple levels of caching to meet performance requirements and manage load on source systems:

- Result set Caching
- Code Table Caching
- Internal and external materialized Views

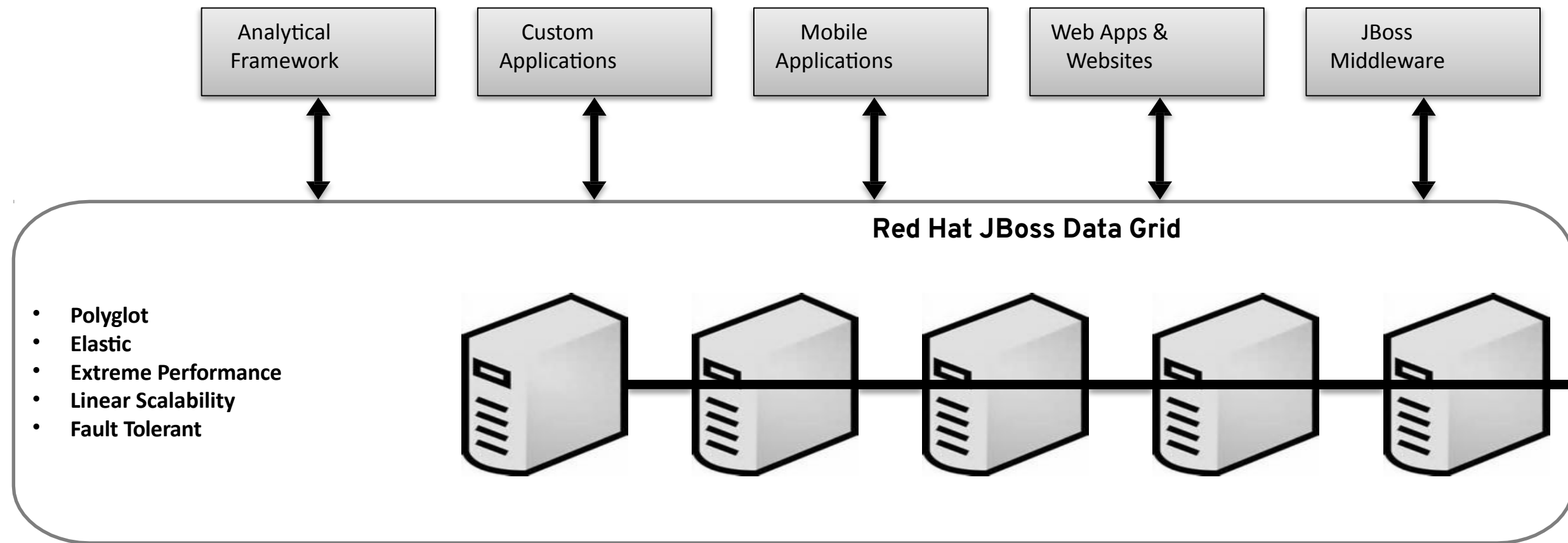
OPTIMISED QUERY ENGINE

- Access Patterns – criteria requirements on pushdown queries
- Pushdown – decompose user query into source queries
- Rule and cost based optimisation

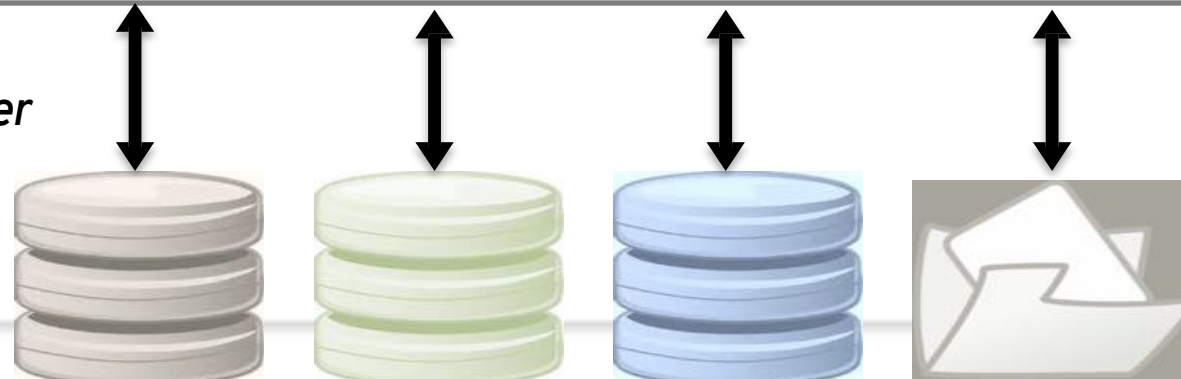
Performance benchmark at http://www.principledtechnologies.com/Red%20Hat/JDV_data_virtualization_1215.pdf

JBoss DATA GRID

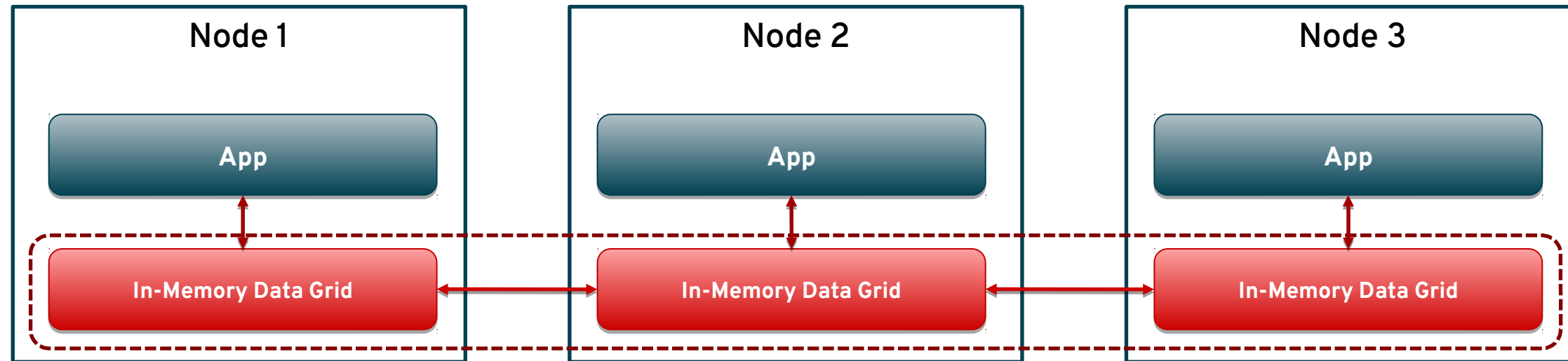
WHAT IS JBOSS DATA GRID?



IN-MEMORY DATA = all data needed is kept in memory
GRID = too big for one node, so data is distributed in cluster

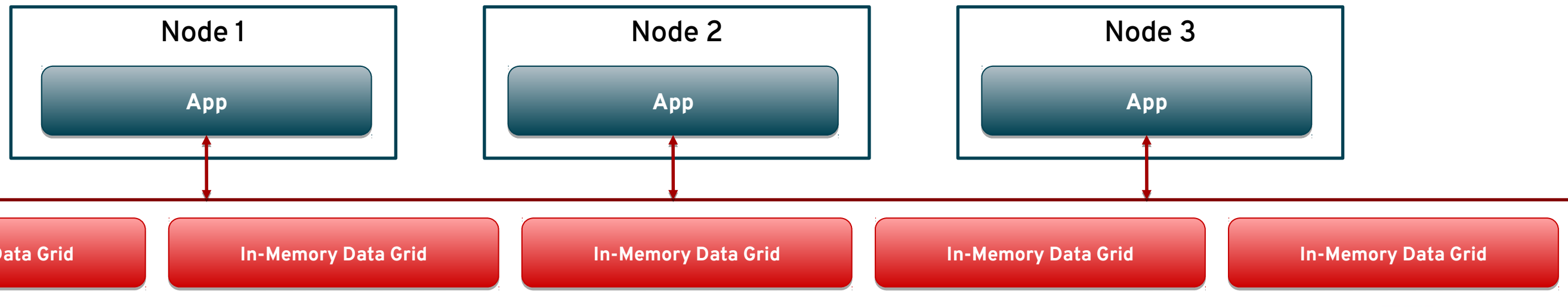


DEPLOYMENTS MODES



Embedded / Library Mode

Client / Server Mode



4 IN 1 PACKAGE

Distributed Cache

In-memory data store to keep the most frequently accessed data.

Transient, short-lived data storage

NoSQL Database

NoSQL Key-Value data store

Configurable transaction support

Event Broker

Listen and respond to data events throughout the data grid

Continuous queries ensure the latest result set

Big Data/IoT Analytics

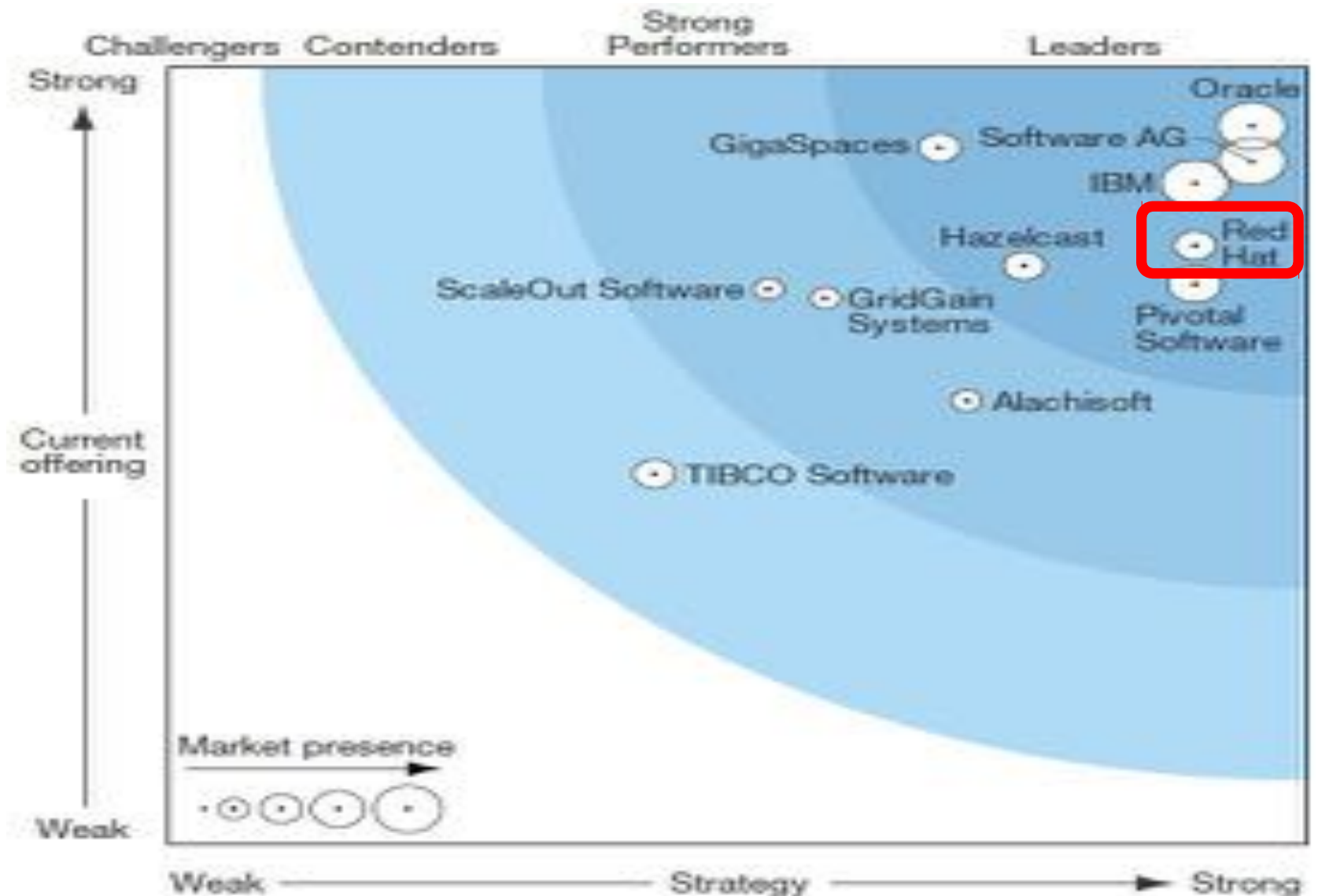
Simplified MapReduce with Java Stream API

Integration with Spark and Hadoop

LEADER IN FORRESTER WAVE™: In-Memory Data Grids, Q3 2015

Ahead in both evaluation dimensions vs. open source competitors

- 1) Current offering
- 2) Strategy and Vision



DATA GRID USE CASES

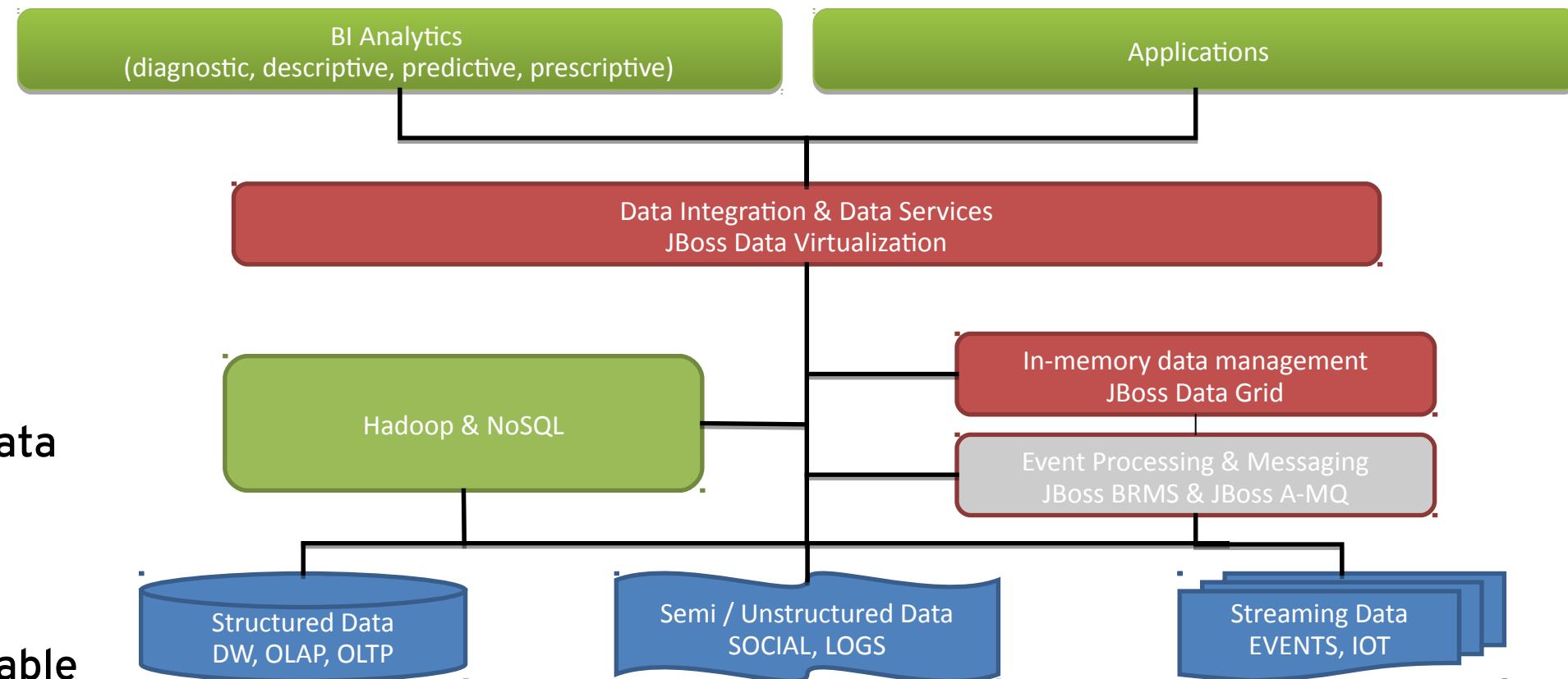
Scenario	
Horizontal	Web <ul style="list-style-type: none">• User-specific HTTP session and shared state across web farm• In-flight shopping carts for web retail• Enabling online self-service applications• Explicit storage of pre-computed or highly-accessed data
	LOB <ul style="list-style-type: none">• Enterprise-wide product catalog for analytics• Caching frequently used reference data from Enterprise applications
Verticals	Travel <ul style="list-style-type: none">• Aggregated flight pricing / availability retrieved from airlines
	Defense <ul style="list-style-type: none">• Sensor network data processing and threat detection
	Financial <ul style="list-style-type: none">• Per-user portfolio data and risk analysis, delayed quote storage for trading• Aggregate and process ticker stream for algorithmic trading, fraud detection

JBOSS DATA GRID MEETS DATA VIRTUALIZATION

JBoss Data Grid as:

- Federated data source for Data Virtualization

- High performance, high scalable materialisation target for Data Virtualization



CUSTOMERS STORIES

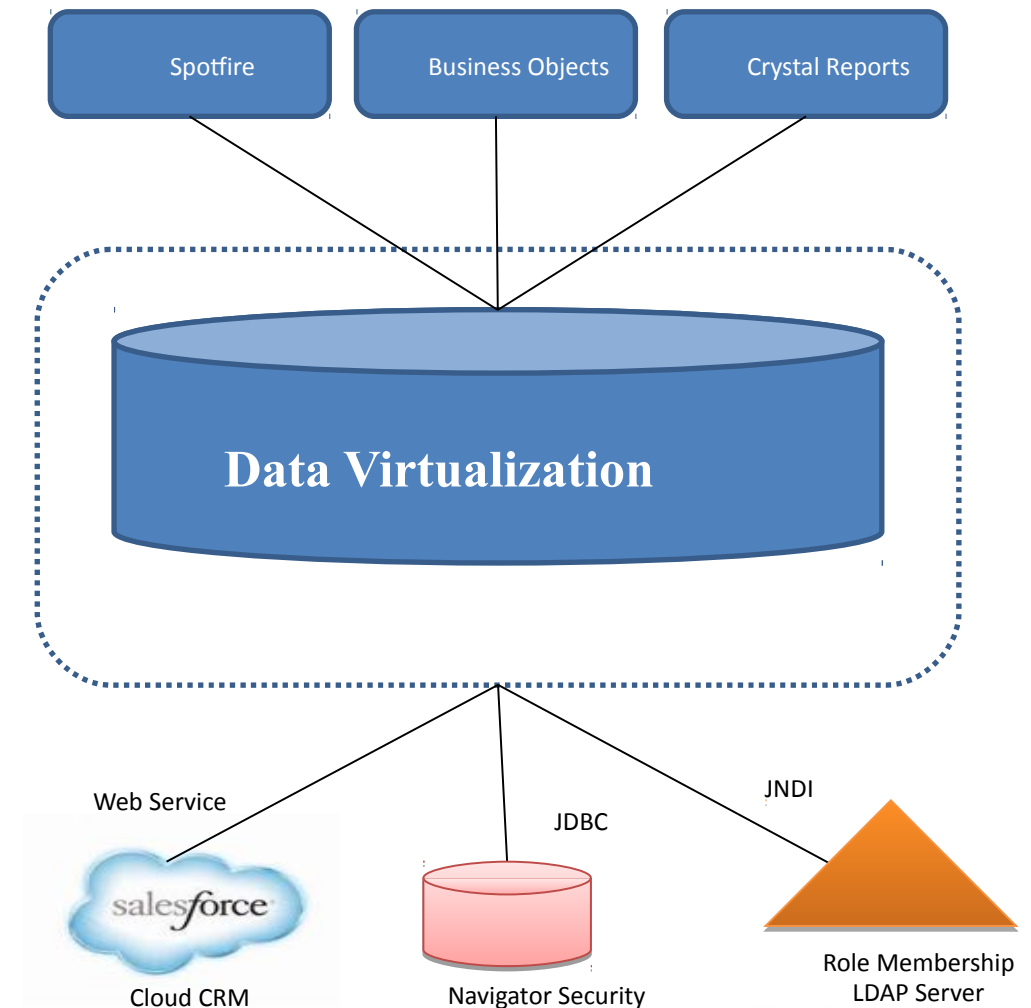
GLOBAL BIOTECH COMPANY : SELF SERVICE BI

Situation/Needs

- Needed to integrate cloud application data (salesforce.com) with on-premise, real-time data for operational reporting and monitoring
- Need to ensure HIPAA compliance

Solution

- Used Data Virtualization to provide unified interface to data to multiple BI tools
- Enabled business users to use the BI tools of choice while IT ensured better control of information
- Rapid development cycle with the use of common data models



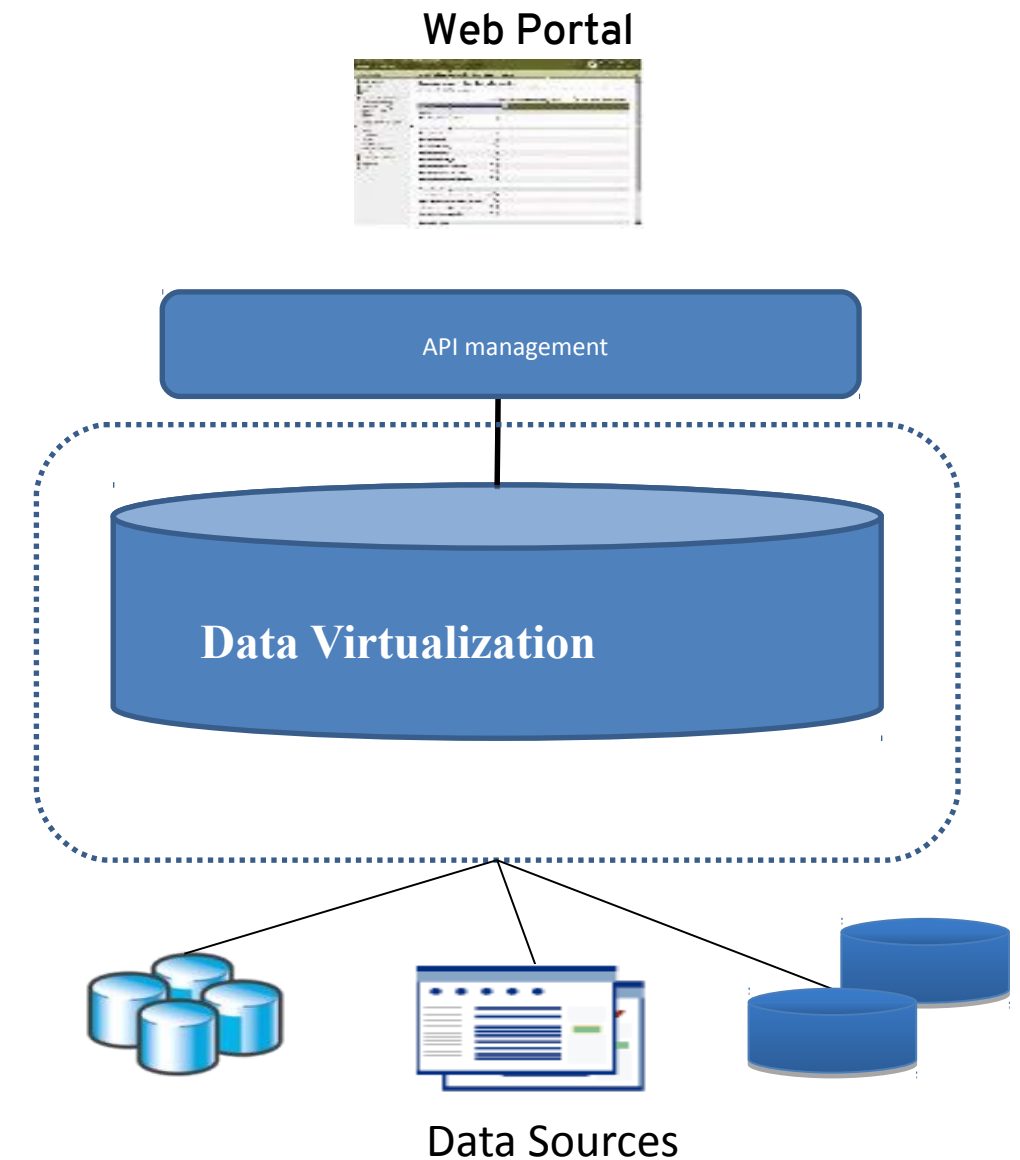
CONSULTANCY FIRM : DATA API

Situation / Needs

- Rethink traditional IT workloads
- Faster go-to-market, improving the consumers experience
- Create a platform to drive innovation, fostering collaboration
- Democratising access to data

Solution

- Combine historical and real-time data.
- Data abstraction is key to define a reusable, extensible architecture
 - Using model driven development new ideas and solution are implemented faster to respond to a rapid changing market



CHOOSE THE BREAD YOU WANT



CALL TO ACTION

- Align your data project to your business initiatives and make sure you can take decisions in close to real time
- Rethink your data integration strategy, check :



- <https://www.redhat.com/en/resources/re-think-data-integration-delivering-agile-bi-systems-data-virtualization>*
- Download the study on Modern Data Architectures:
<https://www.redhat.com/en/resources/modern-data-architecture-analyst-paper?intcmp=701600000000lhMNAAY>

MONTHLY TECH TALK SERIES

October 26th An introduction to 3Scale and API Management.

November 23rd EAP 7 and A-MQ 7. JEE and core

December 13th RHEL, RHEV, Atomic and OpenStack.

January 25th Software Defined Storage, Gluster, Ceph.

February 22nd Hybrid Cloud Architectures and Cloudforms

All @ Red Hat Monument Office – Morning and Evening sessions



redhat.