RED HAT
FORUM
Europe, Middle East & Africa

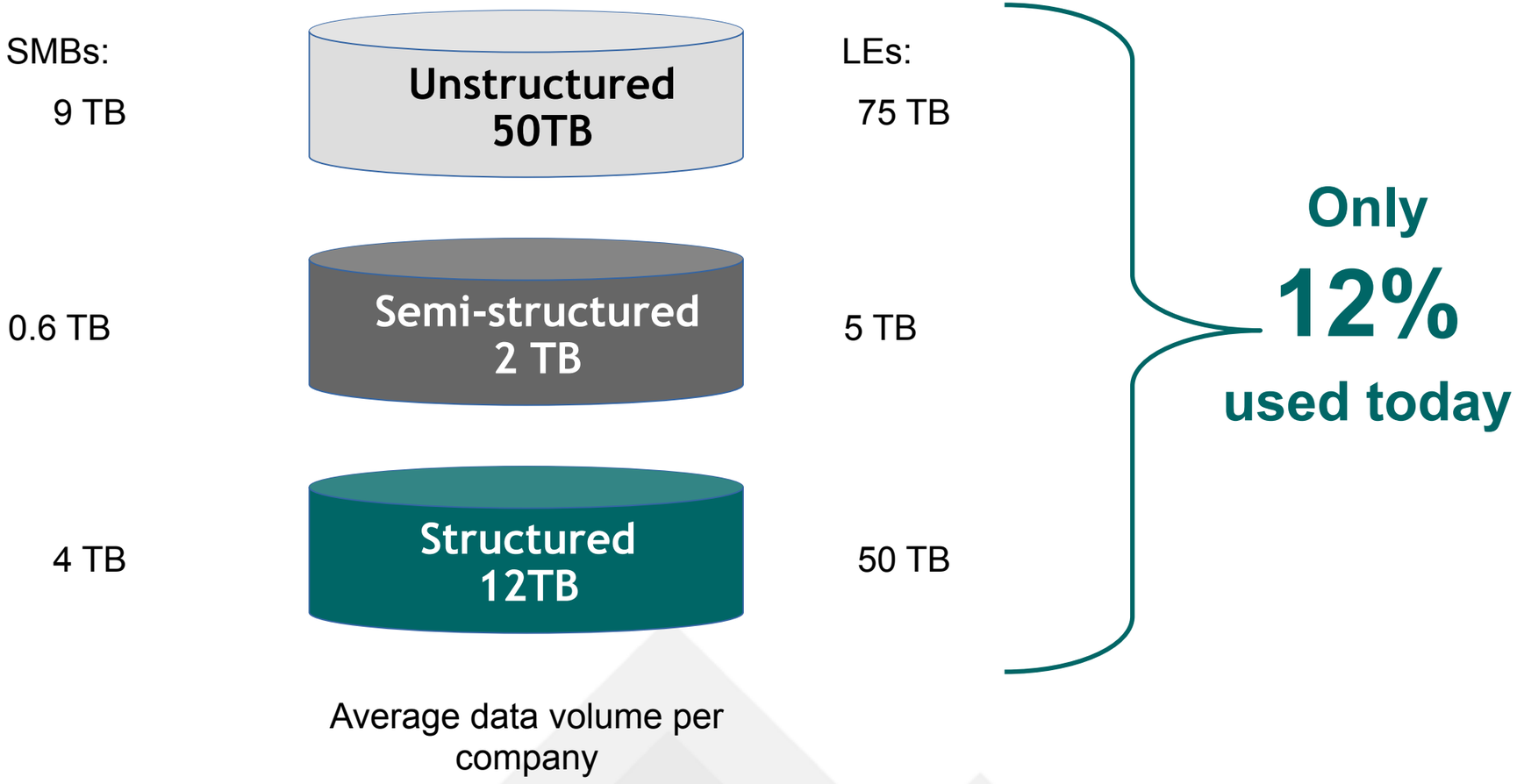
redhat.

Real-time data services with Red Hat JBoss Data Virtualization and Red Hat JBoss Data Grid

Emmanuel Bernard
Platform Architect
Consulting Software Engineer
Red Hat, Inc.

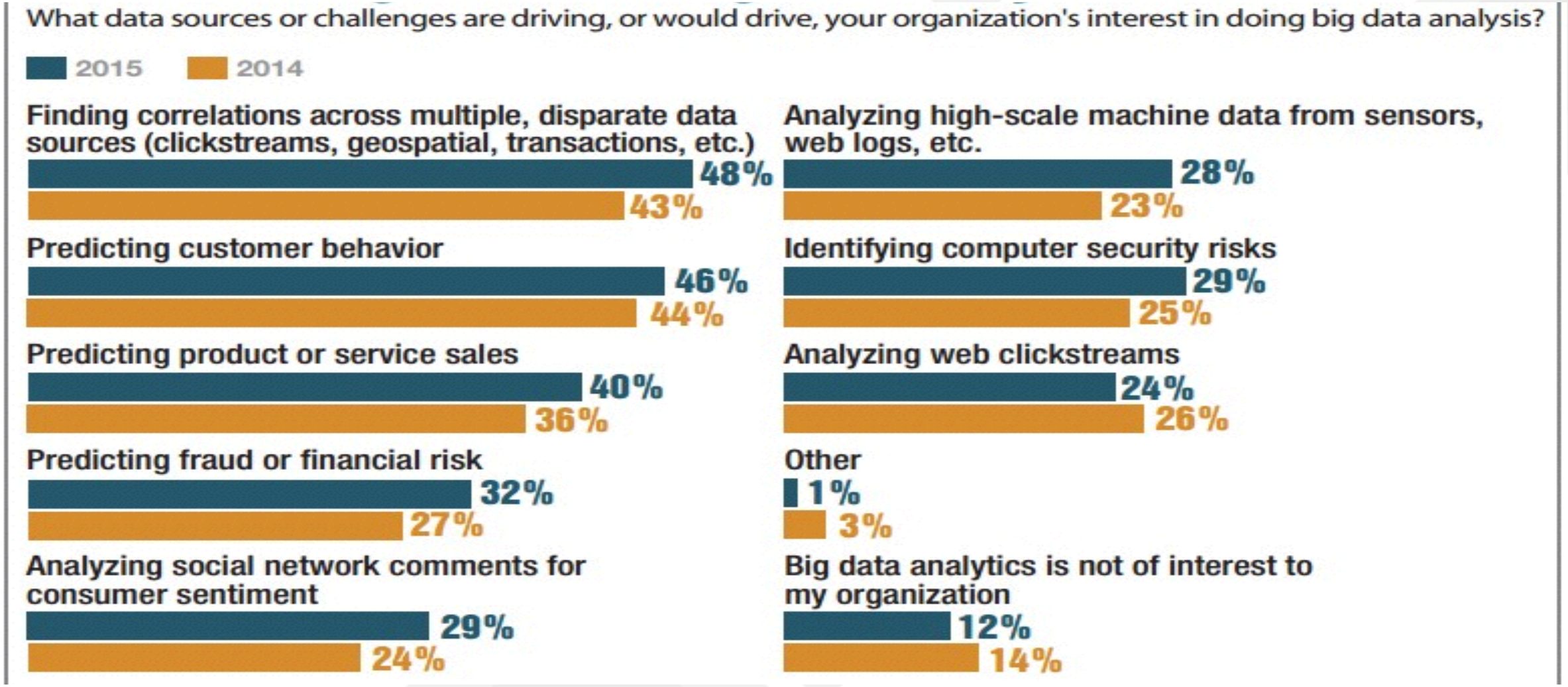
Malik Saheb
Senior Solutions Architect
Red Hat France

Companies use only a fraction of data



Source: Forrsights Strategy Spotlight: Business Intelligence And Big Data, Base: 634 business intelligence users and planners

DATA INITIATIVES



Source : Information week 2015 Analytics and BI survey

The Internet of Things Landscape

Connectivity and « smarter » devices driving change

DRIVE EXPECTATIONS

BETTER ANALYTCS
Real-time analytics

FASTER PERFORMANCE
Sub-seconds not hours

ELASTIC SCALABILTIY
On-demand Scale-out architecture

HIGHER AVAILABILITY
Always available

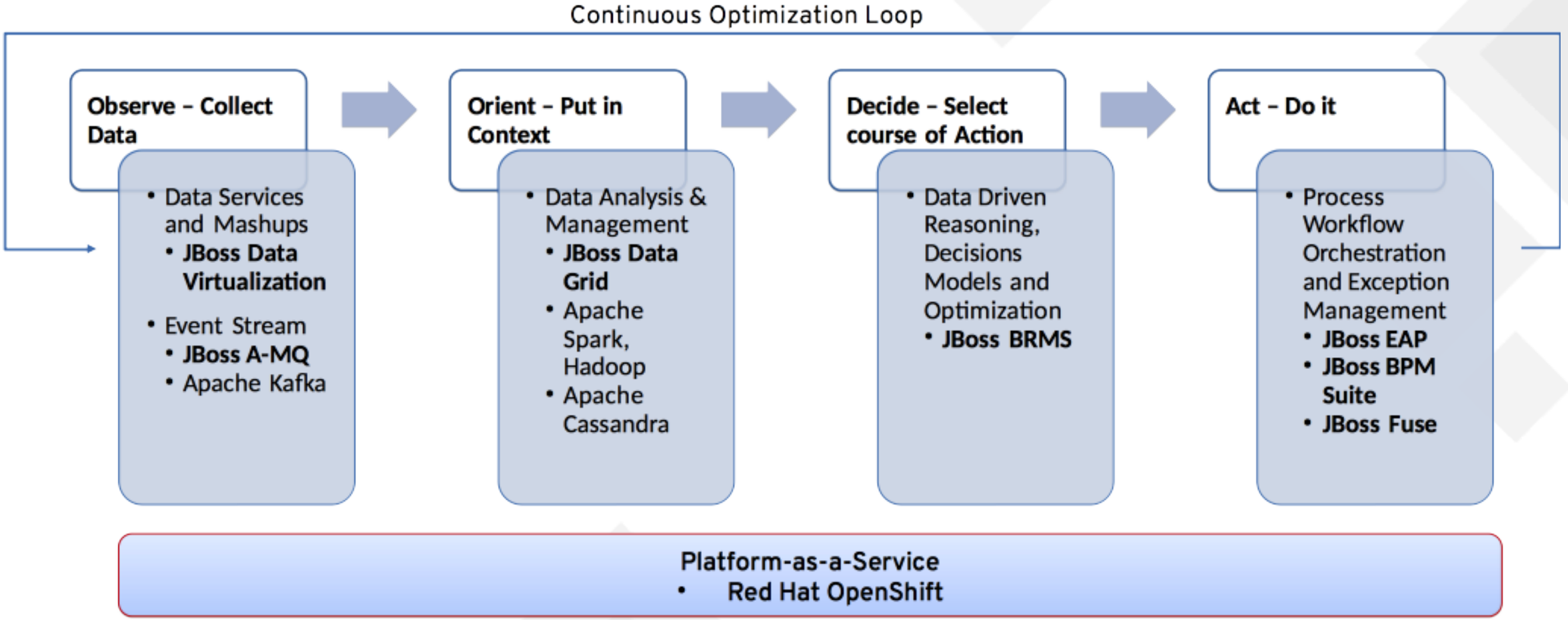
RED HAT
FORUM
Europe, Middle East & Africa

redhat.

How to use this data

WORK SMARTER

INSIGHTS YOU NEED, WITHIN THE APPLICATION YOU USE



INSIGHTFUL APPLICATIONS POWERED BY RED HAT

Open | Modular | Embeddable | Cloud Ready

Real time?

Manual steps

- Data scientist retrieves data
- Data scientist explores & cleans the data
- Data scientist build model based on data
- Developer implements the model and integrates into the app

How to gain speed

- Speed each process
- Automate

With what tools?

RED HAT
FORUM
Europe, Middle East & Africa

redhat.

Apache Spark

Apache Spark

Open Source parallel processing framework for large-scale data analytics

- Both batch and stream oriented

Processes data from HDFS, NoSQL and relational data stores

Claims 100x faster for in-memory processing and 10x faster on disk than MapReduce

Massive ecosystem (as data source and as libraries)

- Machine learning

Became an Apache top-level project in 2014

Over 1000 contributors from over 250 organizations

RED HAT
FORUM
Europe, Middle East & Africa

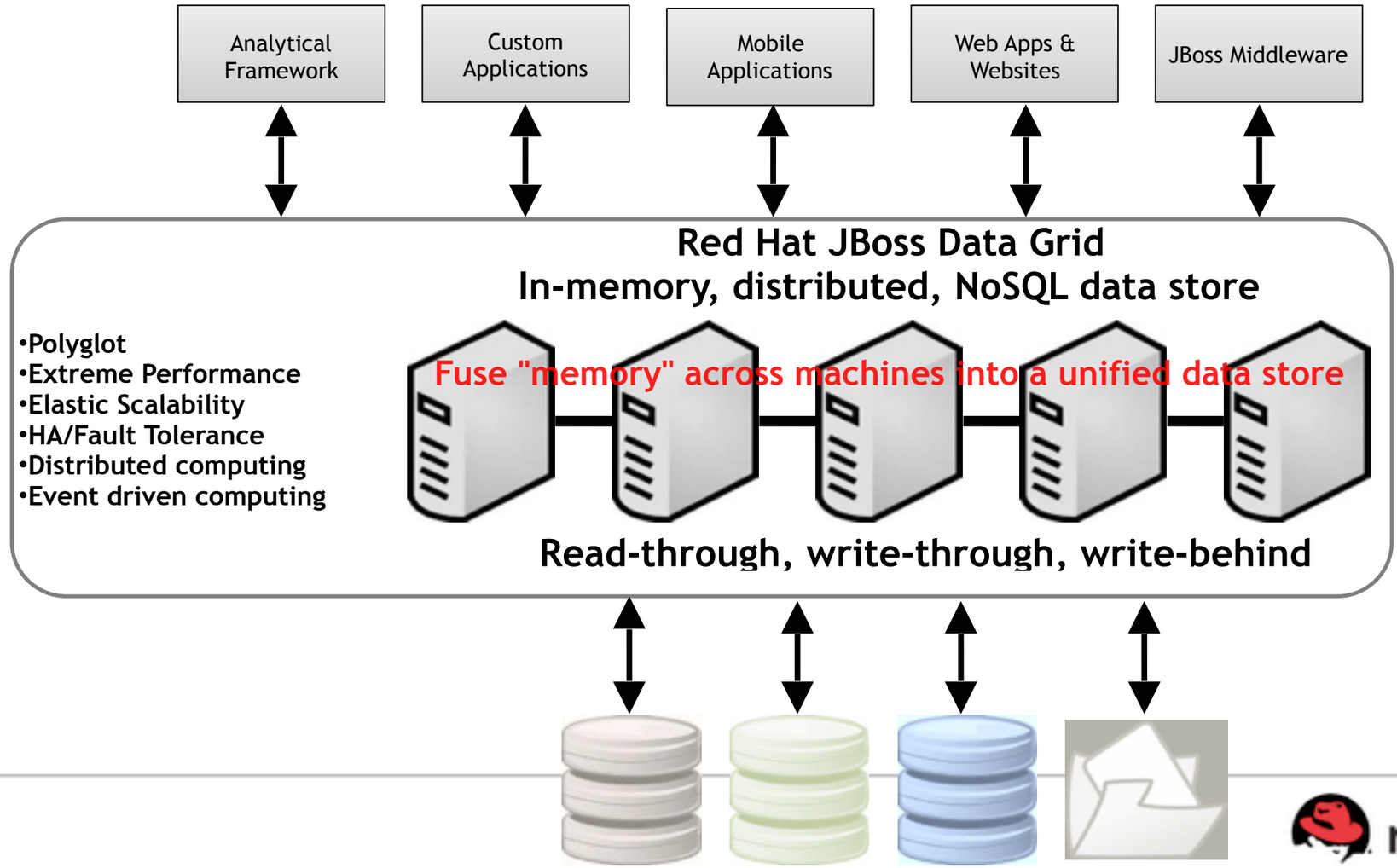
redhat.

JBoss Data Grid

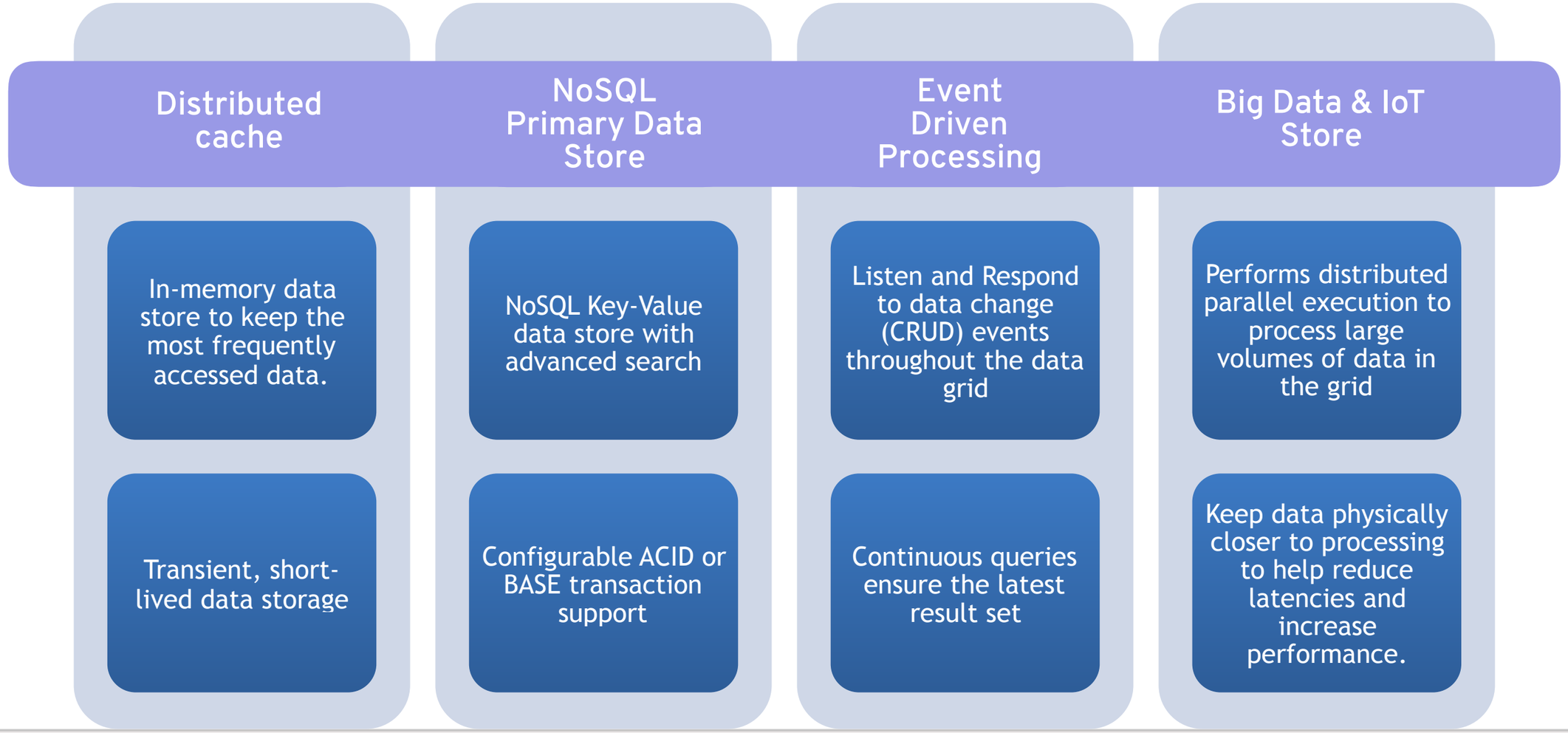
Red Hat JBoss Data Grid

IN-MEMORY DATA = all data needed is supposed to be kept in memory

GRID = too big for one node, so data is distributed in cluster



JBoss Data Grid 4-in-1 package



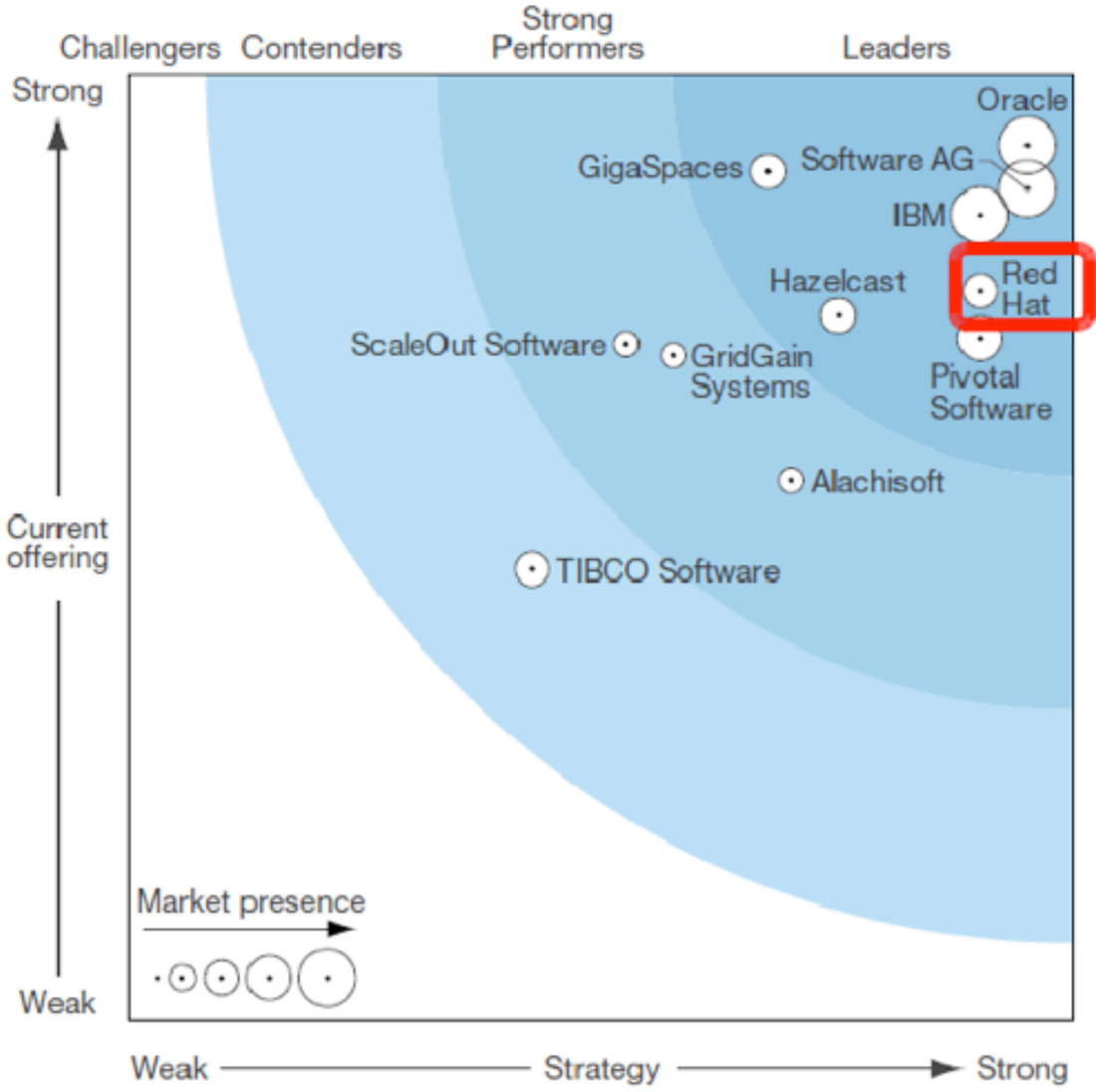
LEADER IN FORRESTER WAVE™ IN-MEMORY DATA GRIDS, Q3 2015

Ahead in both evaluation dimensions vs.
open source competitors

Current offering

Strategy and vision

Download free from https://engage.redhat.com/forrester-data-grid-s-201509240128



Compute features

Query

- Lookup
- Indexed and indexless OQL queries
- Full-text queries

Reactive / event based

- Clustered listeners
- Continuous queries

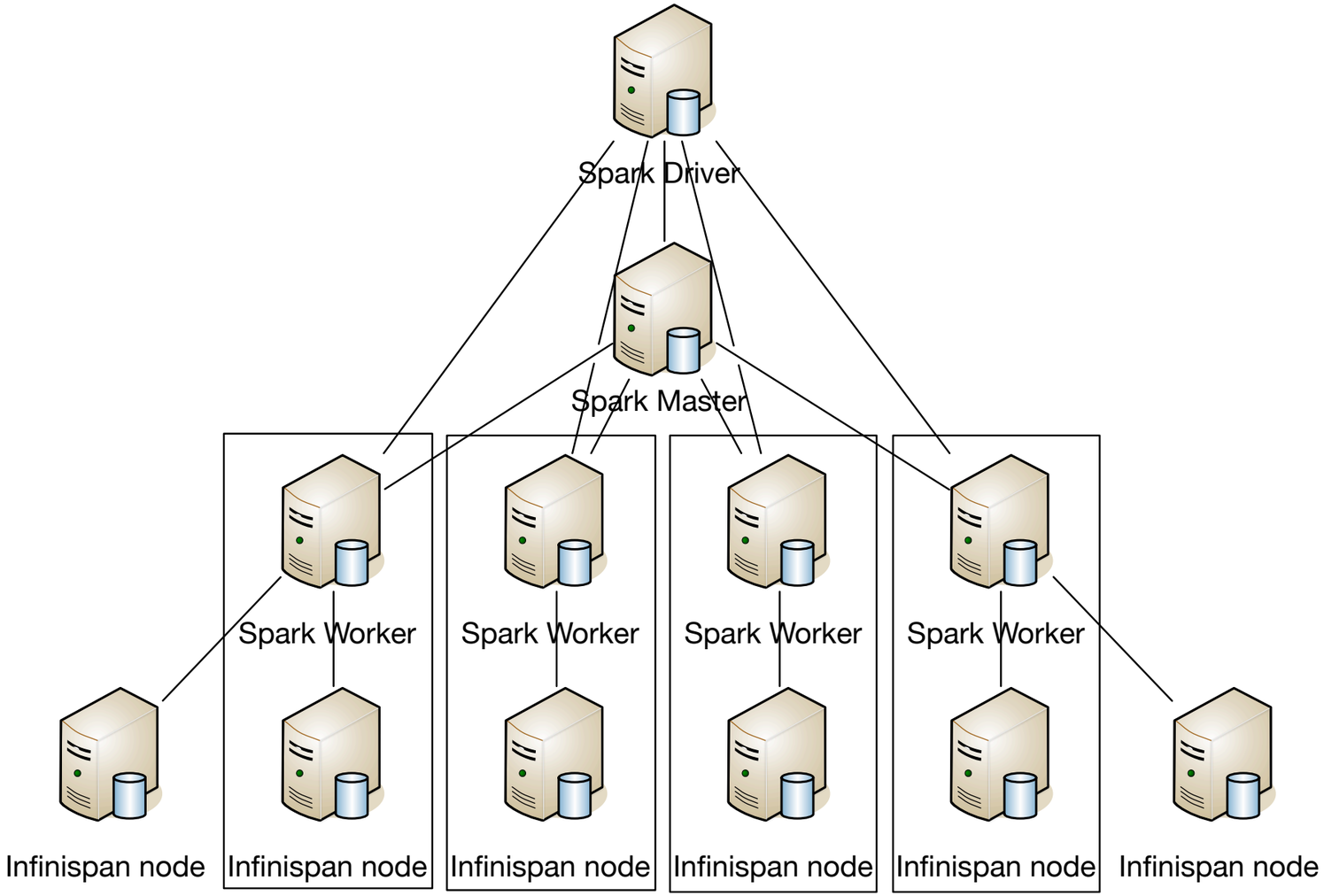
Move computation to the data

- Distributed streams
- Distributed execution framework / remote
- Task execution

Use other execution engines

- Hadoop (InputFormat / OutputFormat)
- Apache Spark

JBoss Data Grid & Apache Spark



Spark and JDG

use cases & features

Speed your Spark to/from

Source of data - amongst many

Machine learning / pattern detection

Computation output

Immutable RDD

Fast RDD caching

RDD and DStream (read & write)

Push down on filters

Make RDD or Stream out of a query

RED HAT
FORUM
Europe, Middle East & Africa

redhat.

JBoss Data Virtualization

"My analytics are becoming more difficult because of the variety and types of data sources (not just the volume)"

Source: Paradigm4 data scientist survey 2014

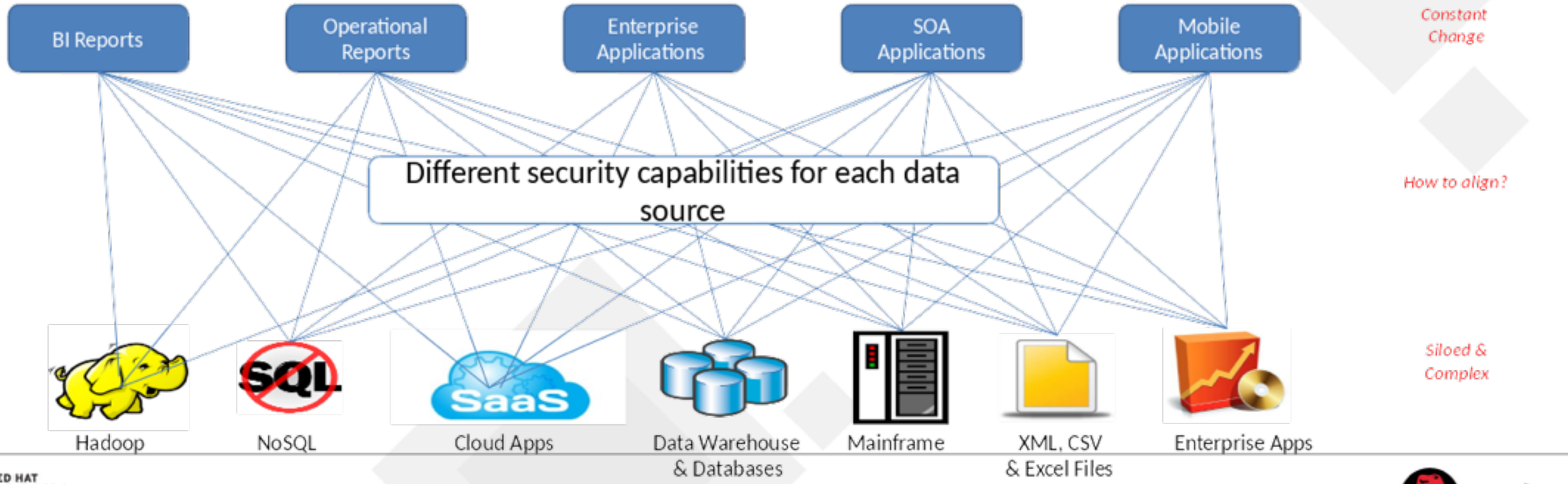
Work Data Control Challenges Getting Bigger with Big Data, Cloud, and Mobile

Security capabilities are tightly coupled to data sources

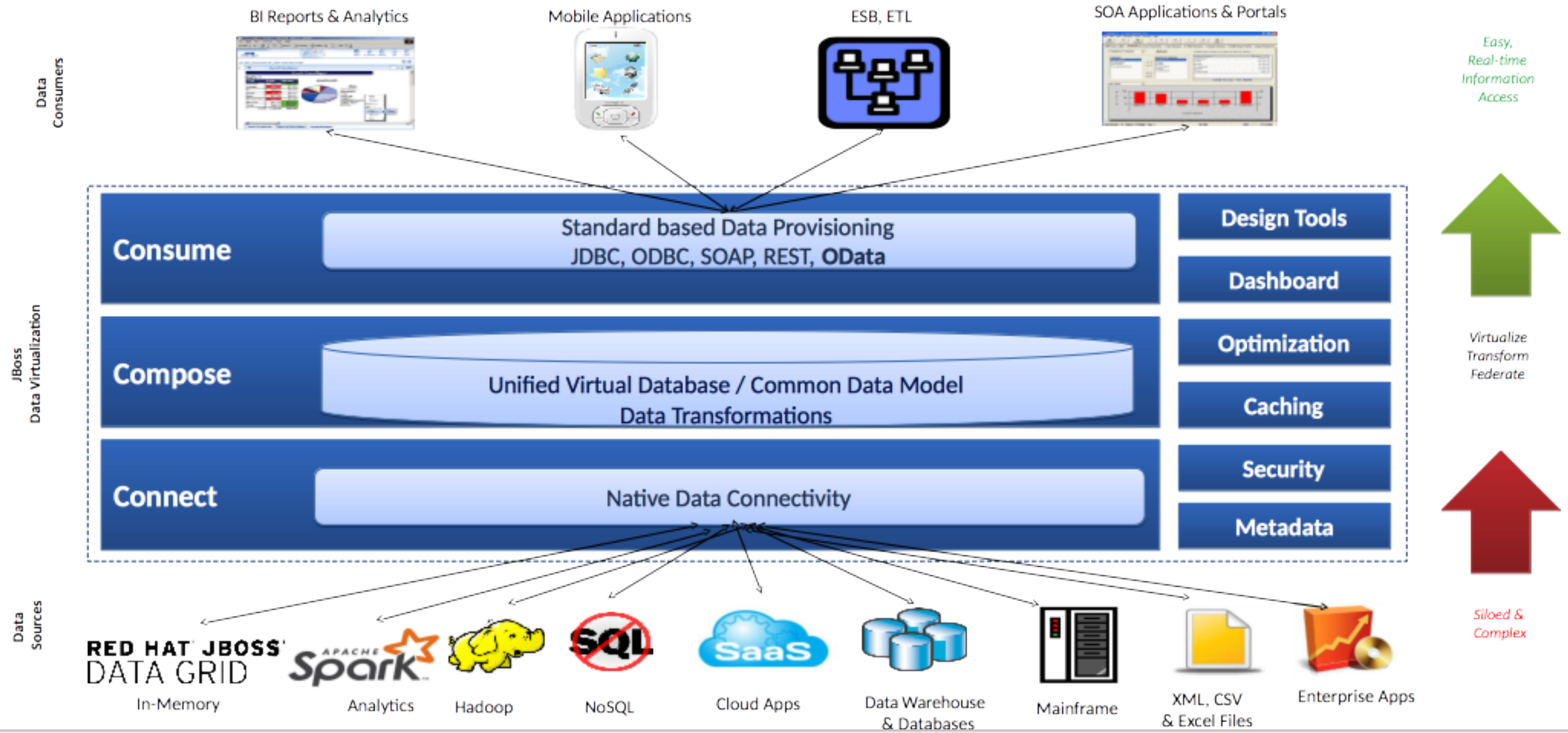
Extracting and moving data adds risk

Every project solves data access and integration in a different way

Inconsistent and decentralized control of data



Turn Siloed Data into Actionable Information with JBoss Data Virtualization



JBoss Data Virtualization Key Business Values

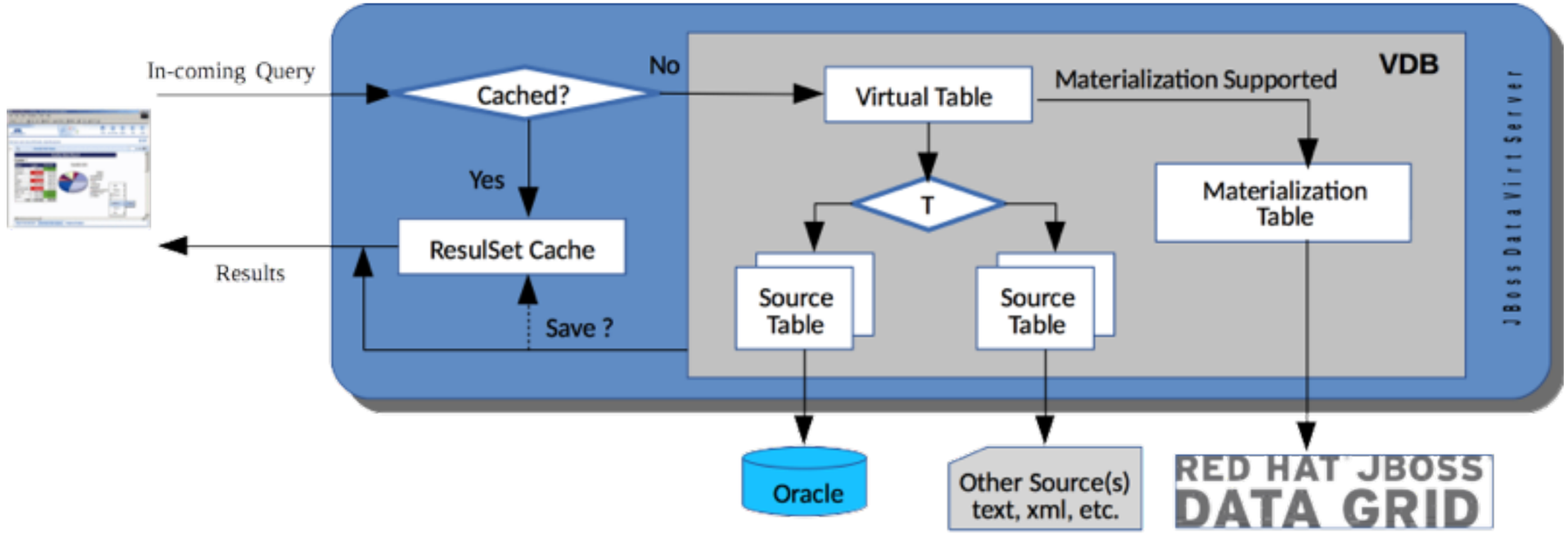
Improve Productivity	Boost Agility	Increase ROA	Better Information Control
• Better/faster than hand coding • Right data at the right time to the right people • Decision support, BI with a complete view of information	• Faster, less costly than batch data movement • Data virtualization provides loose coupling	• Improved utilization of data assets • Derive more value from existing investments • Complements existing systems	• Powerful security, Auditing, Data Firewall • Avoid data silo proliferation • Central data access and policy, Compliance

JBoss Data Grid ↔ JBoss Data Virtualization

JBoss Data Grid

As a federated read and write **data source** for JBoss Data Virtualization

As a high-performance, highly-scalable target for **external materialization** (materialization is a form of caching solution provided by JDV)



JBoss Data Virtualization ↔ Spark

Spark Consumes data from a JDV virtual database

- Centralized Security
- Optimizes performance with pushdown queries

JDV consuming data from Spark

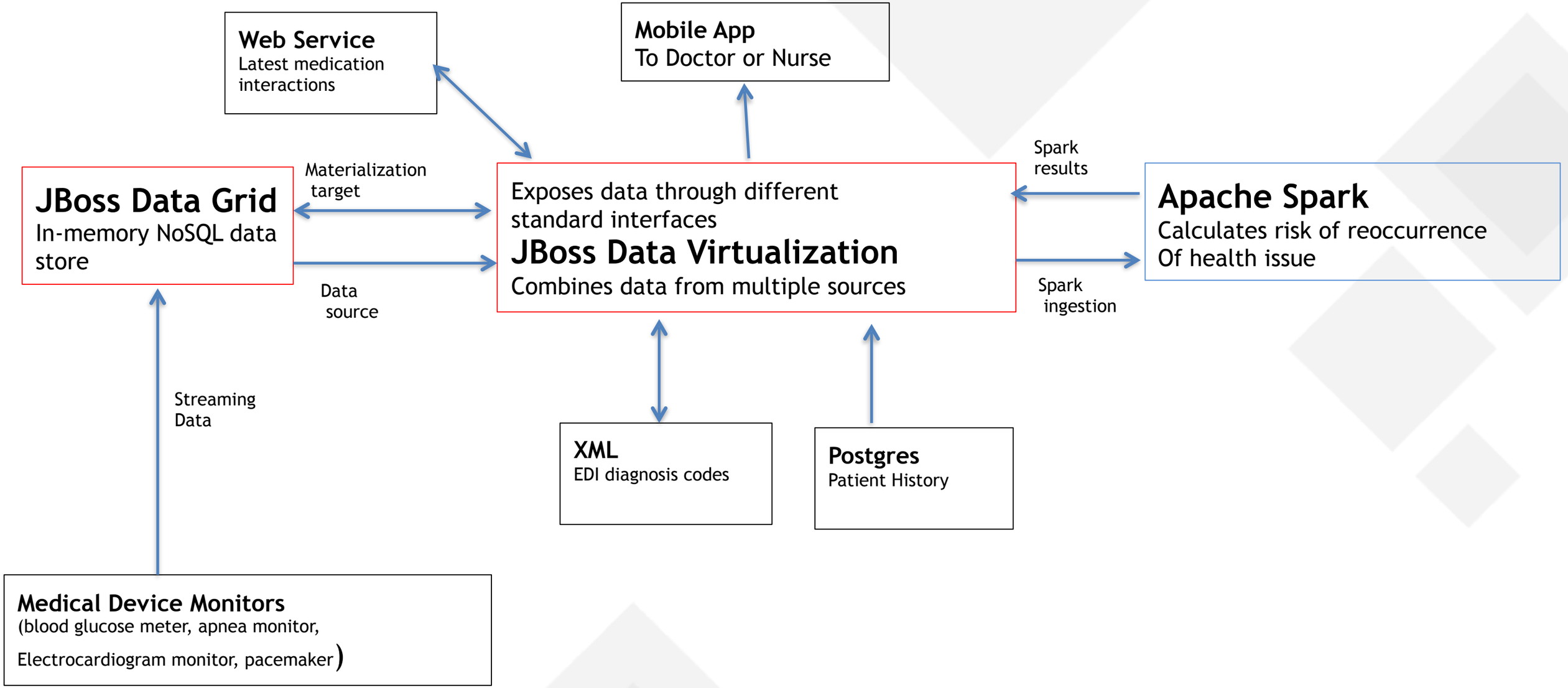
- Re-use: exposes same data through multiple different interfaces

RED HAT
FORUM
Europe, Middle East & Africa

redhat.

Piecing it all together

Healthcare example


Web Service
Latest medication interactions
Mobile App
To Doctor or Nurse
JBoss Data Grid
In-memory NoSQL data store
Materialization target
Data source
Exposes data through different standard interfaces
JBoss Data Virtualization
Combines data from multiple sources
Spark results
Spark ingestion
Apache Spark
Calculates risk of reoccurrence Of health issue
Streaming Data
XML
EDI diagnosis codes
Postgres
Patient History
Medical Device Monitors
(blood glucose meter, apnea monitor,
Electrocardiogram monitor, pacemaker)

Data is coming

Data is flowing

redhat®