

The **Big** Deal

About **Big** DATA

The Big Data Conundrum

- Just collecting and storing “Big Data” doesn’t drive a cent of value to an organization’s bottom line
- The percentage of available data an enterprise can analyze is decreasing proportionately to the available to that enterprise
 - Quite simply, this means as enterprises, we are getting “more naive” about our business over time

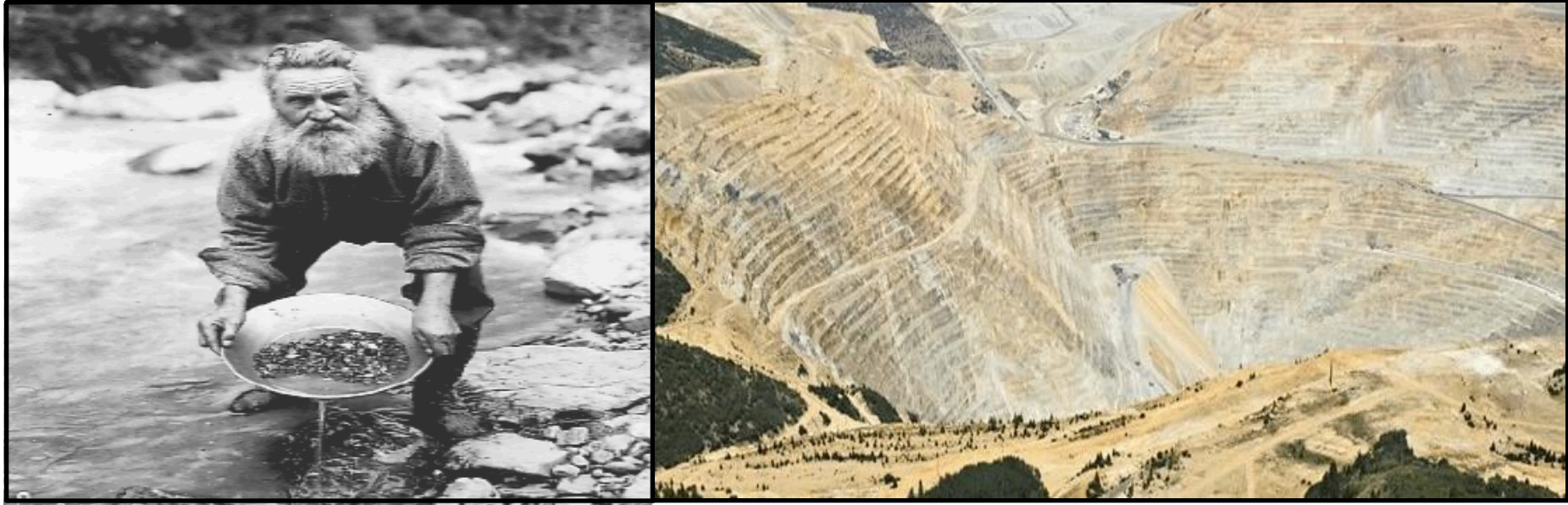
Data AVAILABLE to
an organization

Signals
and
Noise



Data an organization
can PROCESS

Why Didn't We Use All of the Big Data Before?



**Producing data is much easier
than making sense of it...**

The 5 Key Big Data Use Cases



Big Data Exploration

Find, visualize, understand all big data to improve decision making



Enhanced 360° View of the Customer

Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information sources



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis

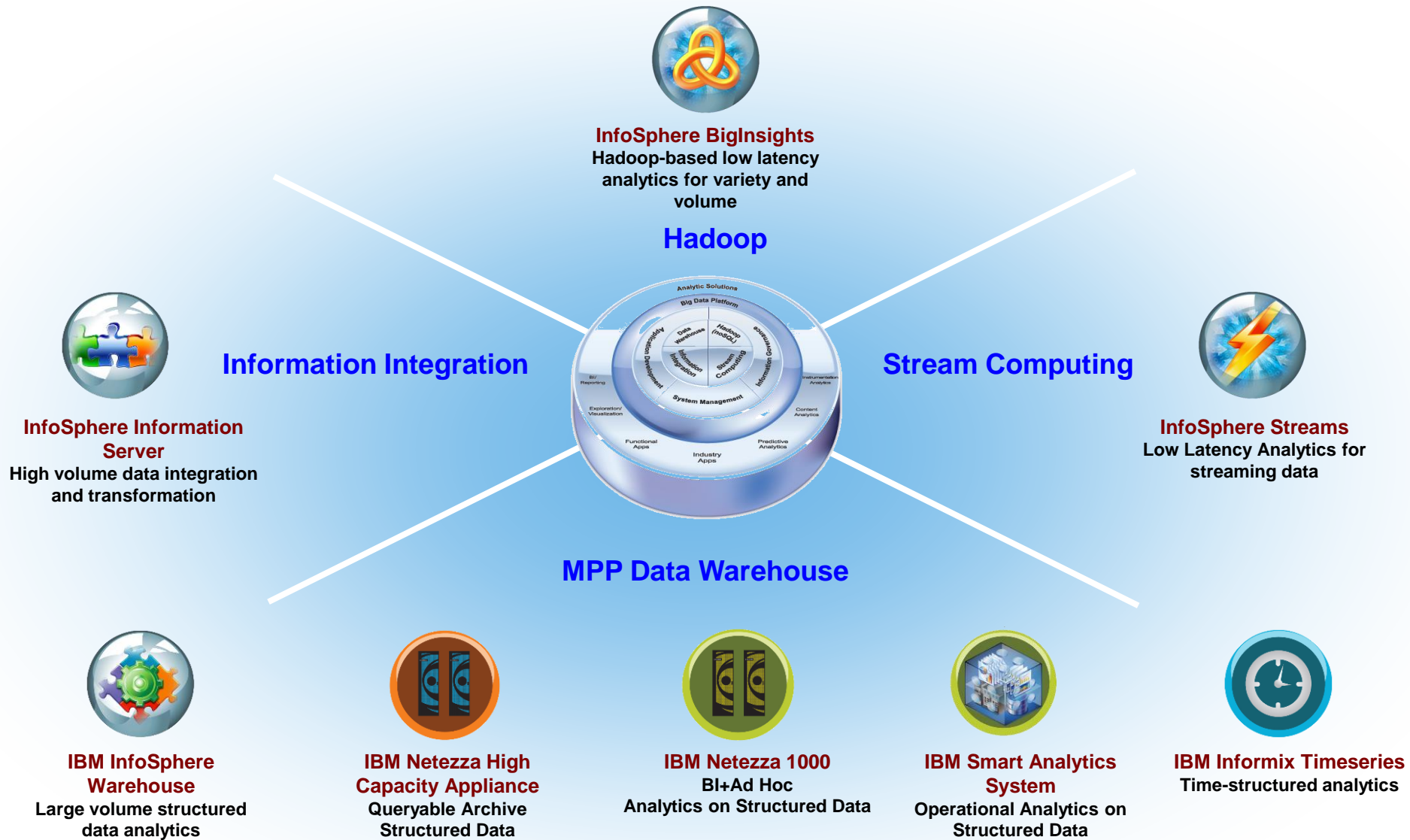
Analyze a variety of machine data for improved business results



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency

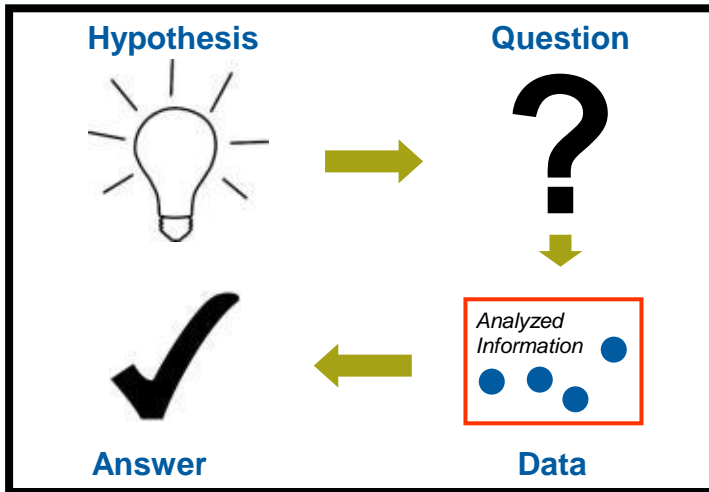
The IBM Big Data Platform



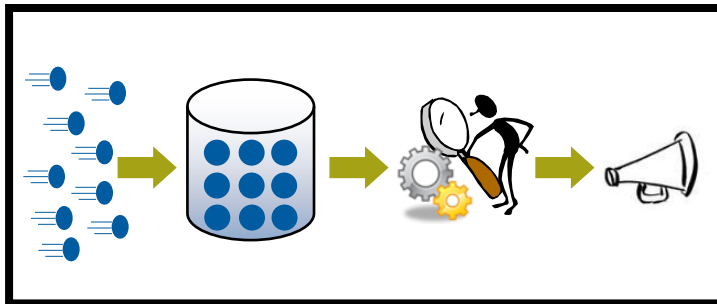
Warehouse Modernization Has to Themes

Traditional Analytics

Structured & Repeatable
Structure built to store data



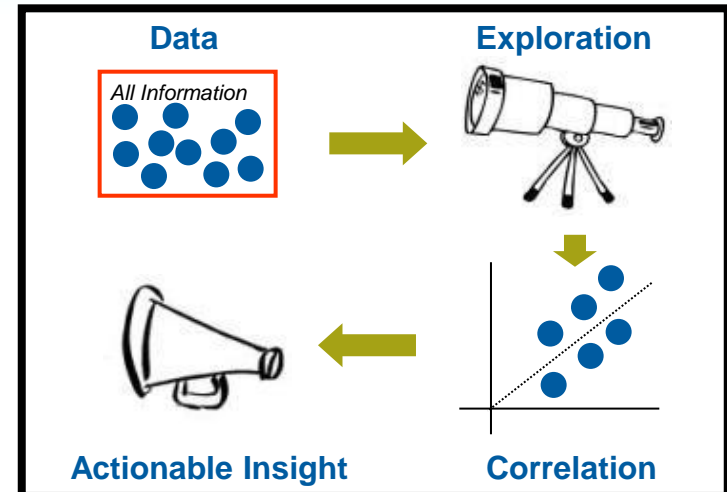
Start with hypothesis
Test against selected data



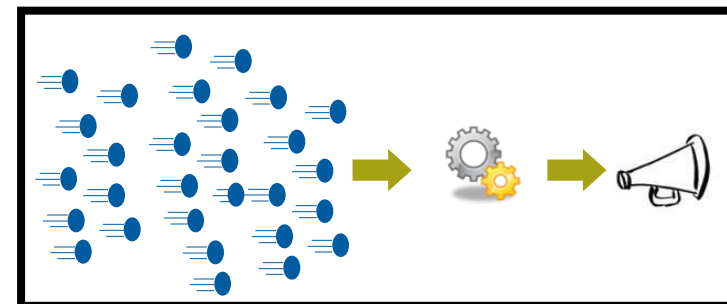
Analyze after landing...

Big Data Analytics

Iterative & Exploratory
Data is the structure



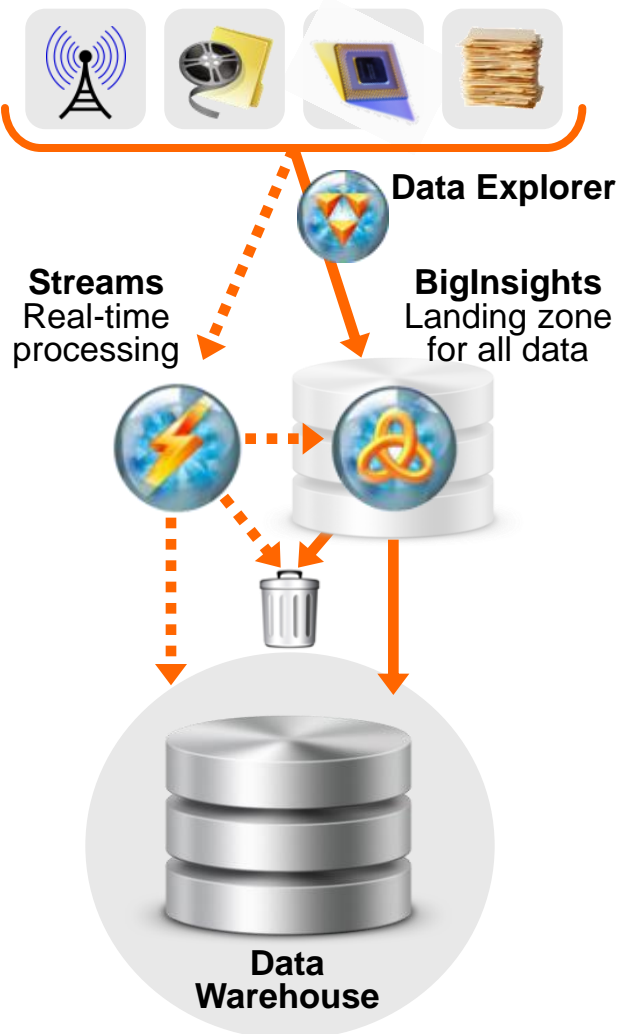
Data leads the way
Explore *all* data, identify correlations



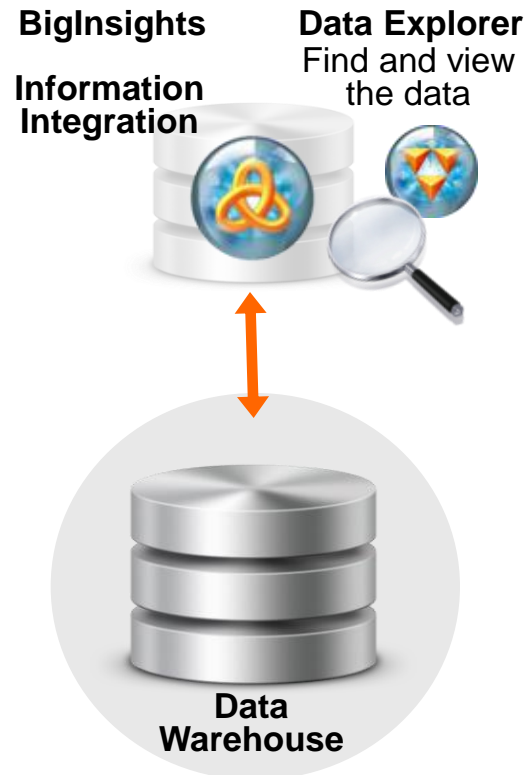
Analyze in motion...

Data Warehouse Augmentation: Value & Diagram

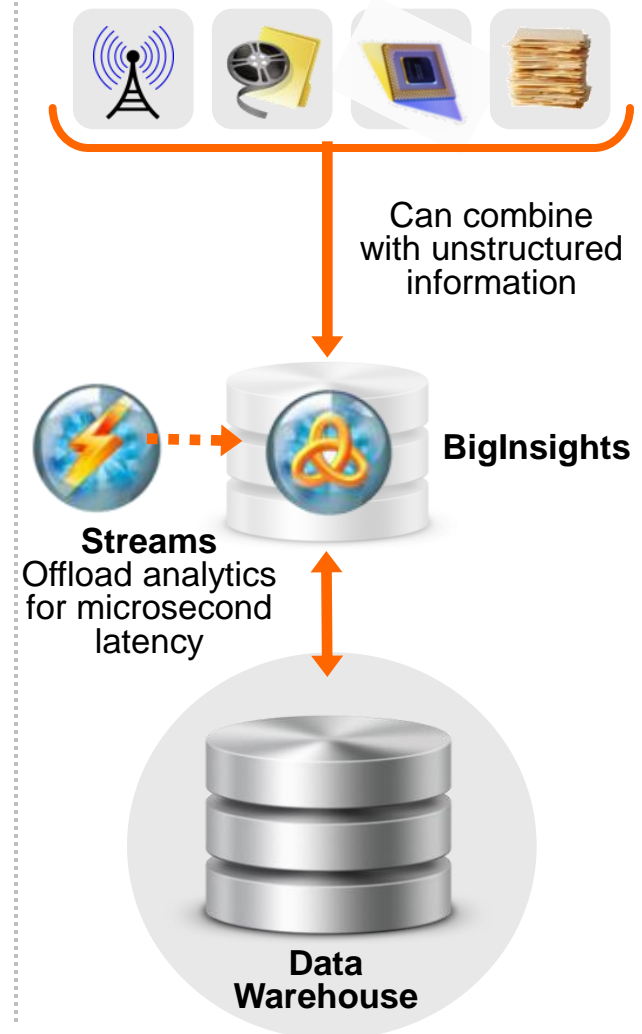
1 Pre-Processing Hub



2 Query-able Archive



3 Exploratory Analysis



Enhanced 360° View of the Customer: Value & Diagram

SOURCE SYSTEMS

CRM
Name: **J Robertson**
Address: **35 West 15th**
Address: **Pittsburgh, PA 15213**

ERP
Name: **Janet Robertson**
Address: **35 West 15th St.**
Address: **Pittsburgh, PA 15213**

Legacy
Name: **Jan Robertson**
Address: **36 West 15th St.**
Address: **Pittsburgh, PA 15213**

Master Data Management



360° View of Party Identity

First:	Janet
Last:	Robertson
Address:	35 West 15th St
City:	Pittsburgh
State/Zip:	PA / 15213
Gender:	F
Age:	48
DOB:	1/4/64



OPTIMUM INVESTMENTS

Logged in as Frank Gelato | Help

Home > Janet Robertson

Personal Information ?

Janet Robertson
Senior Regional Sales Manager
Department: Sales
Office: Pittsburgh, PA
jrobertson@wakefield.com
412.422.2499 x555

Stop Tracking

Associated Accounts ?

Wakefield Investments

LinkedIn History ?

Wakefield Investments
Title: Senior Regional Sales Manager
Years worked: Nov 2006 - Present
Morgan Stanley Smith Barney

Purchase History ?

Date	Amount	Fund
2011-04	23k	Western Asset U.S. Treasury Reserves
2011-01	25k	Optimum Short Duration Municipal Income Fund
2010-10	18k	Optimum Dividend Value Fund
2010-07	10k	Optimum Batterymarch S&P 500 Index Fund
2010-04	11k	Optimum ClearBridge Large Cap Growth Fund

Contact Activity Feed ?

Showing: **All Activity** | By Source | By Author

What's new with this contact?

Add Comment

Irene Tambolin updated the document Wakefield Contact Information. SAP - 1 days ago

Todd Williams updated the database Product Funds #322245 - Activity History Wakefield. Database - 3 hours ago

Chelle Kroll added the email address jrobertson@wakefield.com to contact Janet Robertson. Salesforce - 2 days ago

Frank Gelato commented to Janet Robertson: "Scheduled a meeting with Janet for next week. I'll let everyone know how it

Owned Products ?

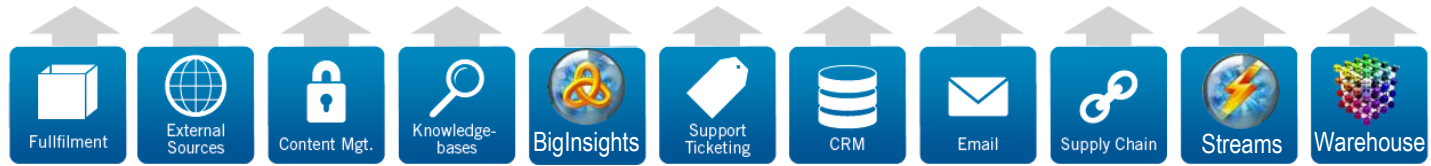
- Optimum Bond Fund
- Western Asset U.S. Treasury Reserves
- Optimum Western Asset Short Duration Municipal Income Fund
- Optimum Dividend Value Fund
- Western Asset New York Municipal Money Market Fund
- Optimum Western Asset Oregon Municipals Fund
- Optimum Special Equity Fund
- Optimum Western Asset New Jersey Municipals Fund

Recent Conversations ?

Email: Optimum Fund Management Overview
Janet, please find an overview of Optimum and the funds we offer. This is only a high level...
 Exchange - 3 hours ago

Notes: Janet is new to Wakefield but has been a wholesaler for 10 years New to Optimum Funds.
 Salesforce - 2 days ago

Title: Optimum Introduction
Welcome Janet to Wakefield. I wanted to introduce myself...



Unified View of Party's Information

Access across many sources

Topic Clusters
Top 192 Results remix
Portals, Content & Collaboration (16)
Marketing (28)
Mktgstrategysig
Teambuilding Exercises (12)
Social (16)
Speaking, Proposal (8)
Content Management (21)
Vivisimo (11)
Overview (14)
Communications (15)
Tech Choices (25)
more | all

Dynamic categorization



Stacy Monarko
Department: Pittsburgh, PA
Title: Director of Product Management
Extension: 971
Email: S.Monarko@vivisimo.com

Top Tags
collaboration (10), competitors (7), social networking (4), analysts (4), social networking (4)

Expertise location

Graphical Refinements

Date
Selected: 6/1/2005 - 2/25/2009

Average Rating
Selected: 1 - 4.5

Leveraging Structured and unstructured content

Highly relevant, personalized results

Windows (51)
Review (32)
Not Classified (24)
Rolling (8)
Idg News Service (5)
Rolling Review (5)
C.g. Lynch (4)

Folder

- Mobile (8)
- Sales Education (0)
- Analyst (5)
- Competitor (5)
- Velocity 6.0 Research (5)
- BI (3)
- Collaboration (2)
- Social Tagging (2)
- Euro (1)
- Events (1)

Enhanced by social collaboration

Refinements based on structured information

Organize content into virtual folders

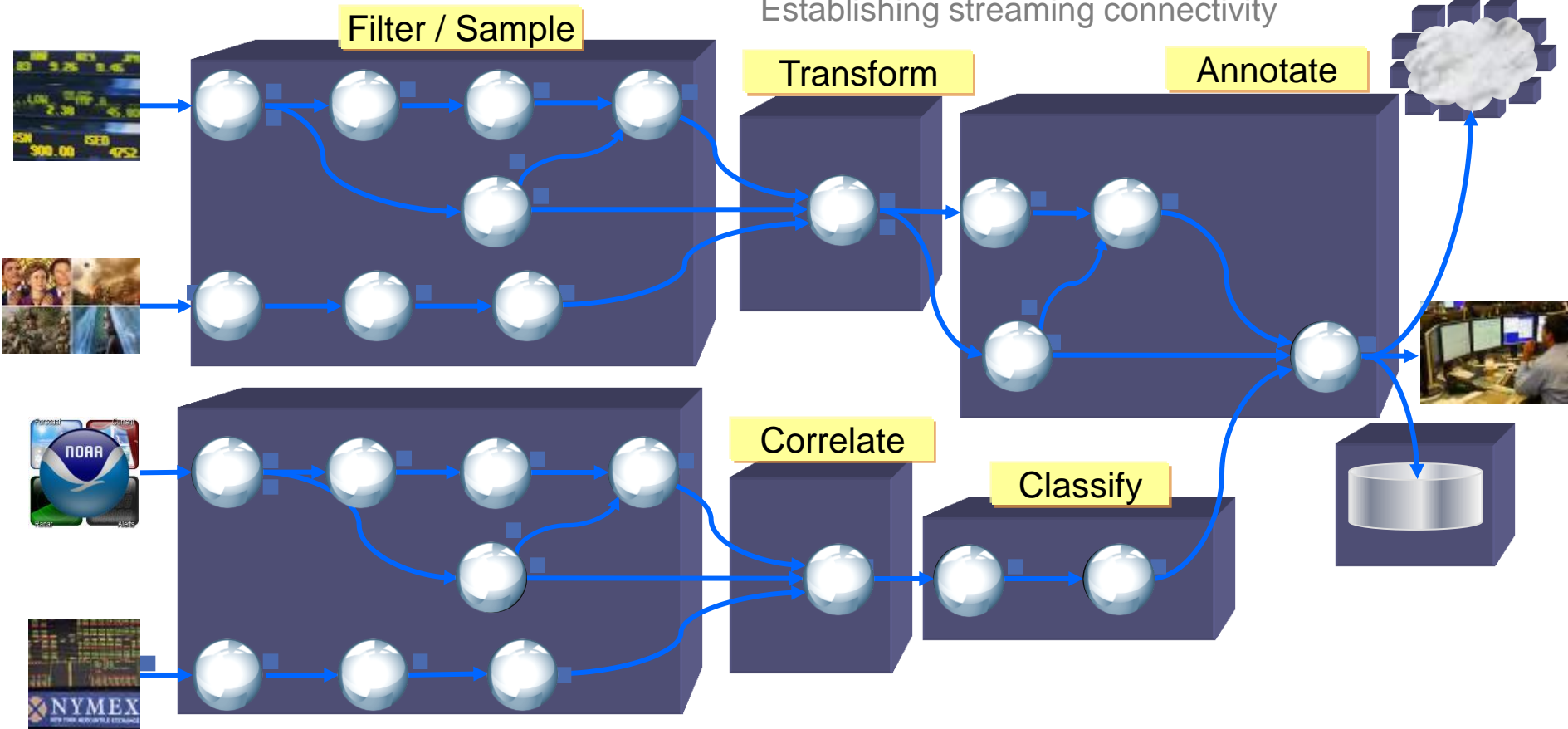
Folders

- Vivisimo
- All Results
- BI
- Competitor
- Sales Education
- Security
- Add folder
- Marketing
- All Results
- Analyst
- Mobile

How Streams Works

- Continuous ingestion
- Continuous analysis

Infrastructure provides services for
Scheduling analytics across hardware hosts,
Establishing streaming connectivity



Achieve scale:

- By partitioning applications into software components
- By distributing across stream-connected hardware hosts

Where appropriate:

- Elements can be *fused* together for lower communication latency



Massively Scalable Stream Analytics

Linear Scalability

- Clustered deployments – massive scalability

Automated Deployment

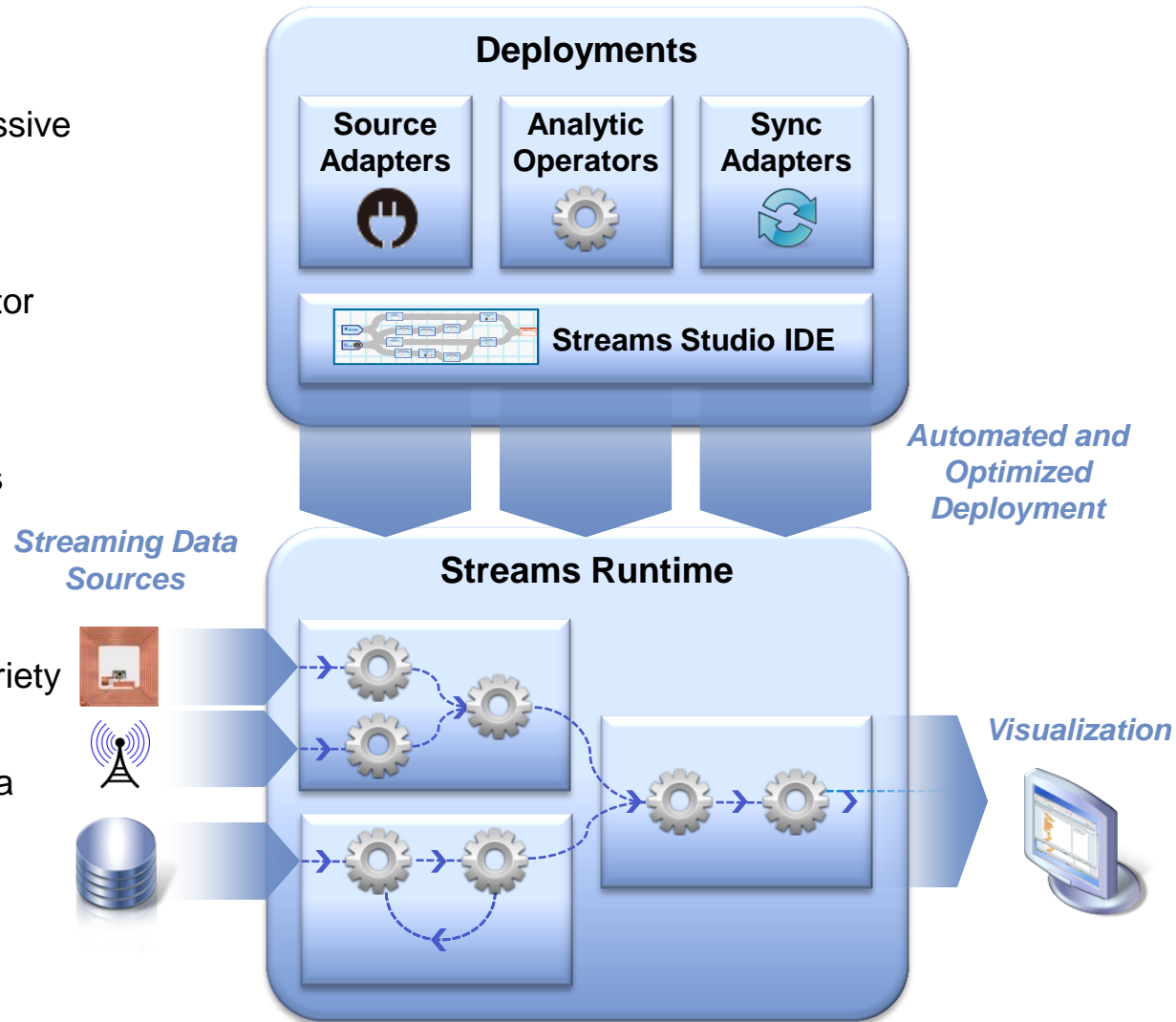
- Automatically optimize operator deployment across nodes

Performance Optimization

- Parallel & Pipeline operations
- Efficient multi-threading

Analytics on Streaming Data

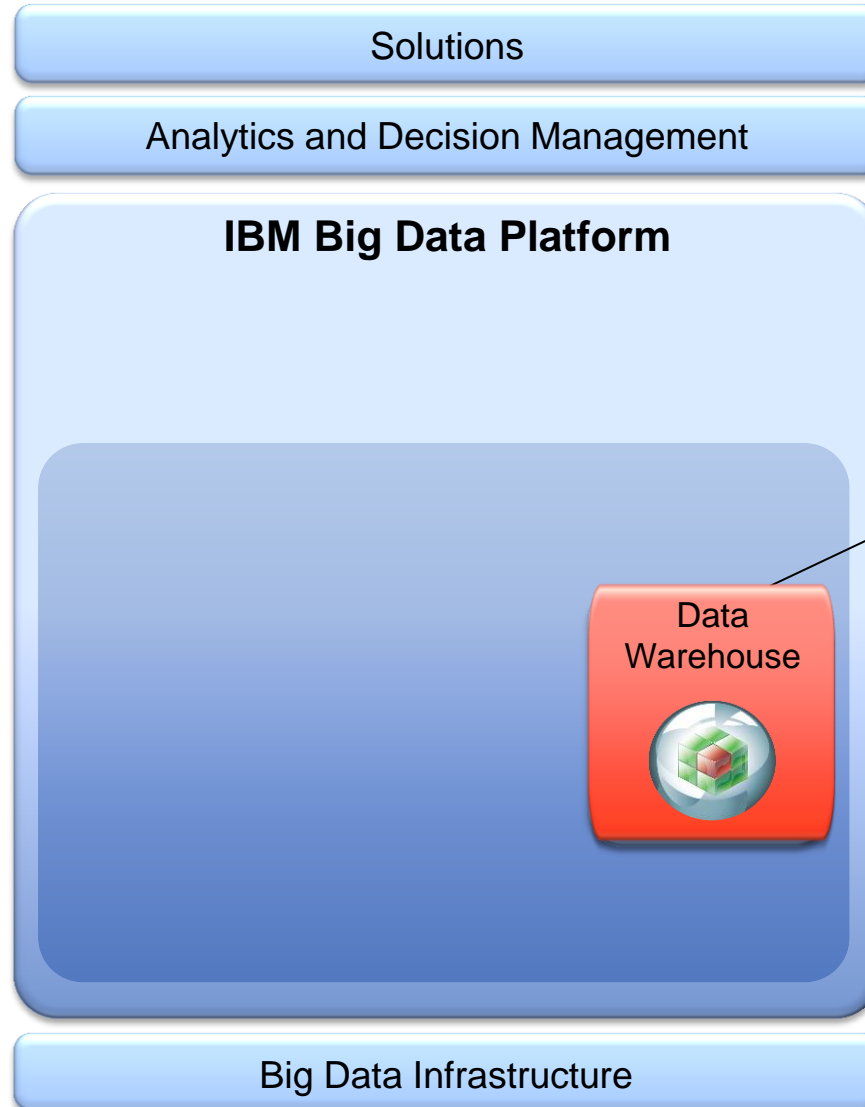
- Analytic accelerators for a variety of data types
- Ability to re-use C++ and Java
- Optimized for real-time performance



The Big Data Platform



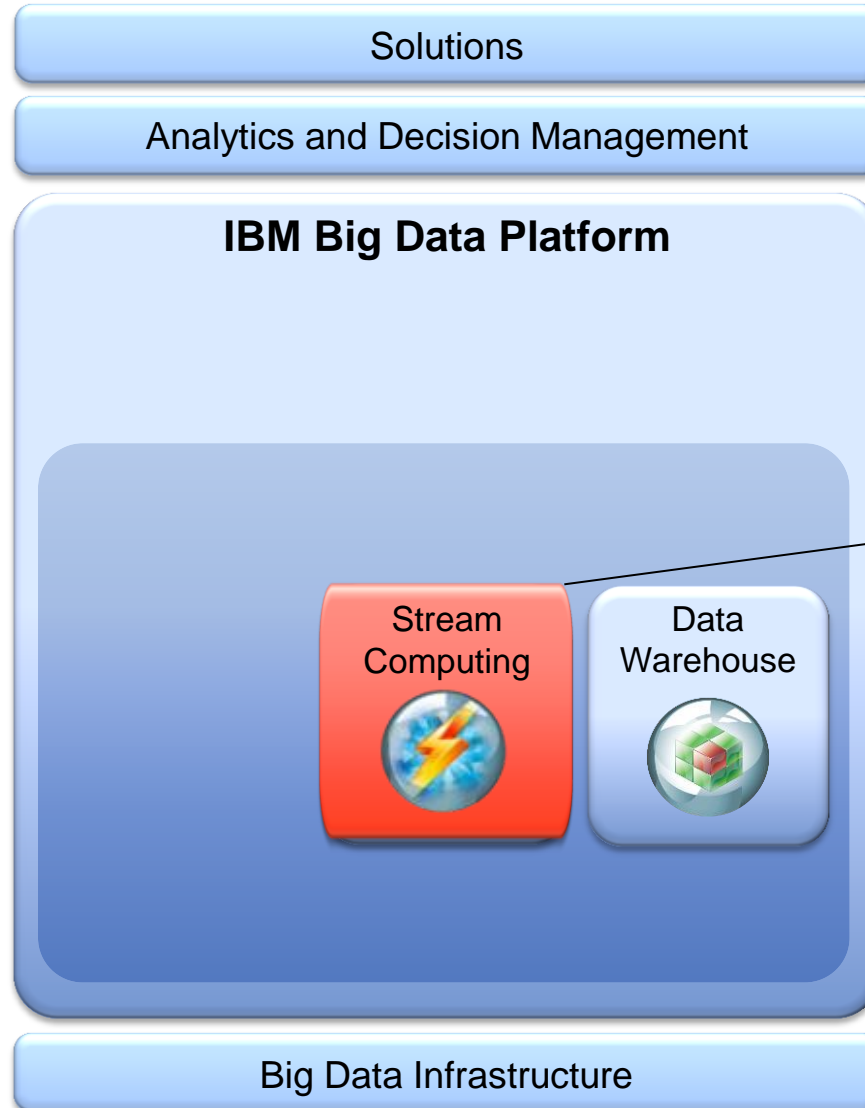
The Big Data Platform



Delivers deep insight with advanced in-database analytics & operational analytics

- **PureData for Analytics** – expert integrated systems to make advanced analytics faster & simpler
- **InfoSphere Warehouse** -- data warehouse software to access operational info in real time.

The Big Data Platform



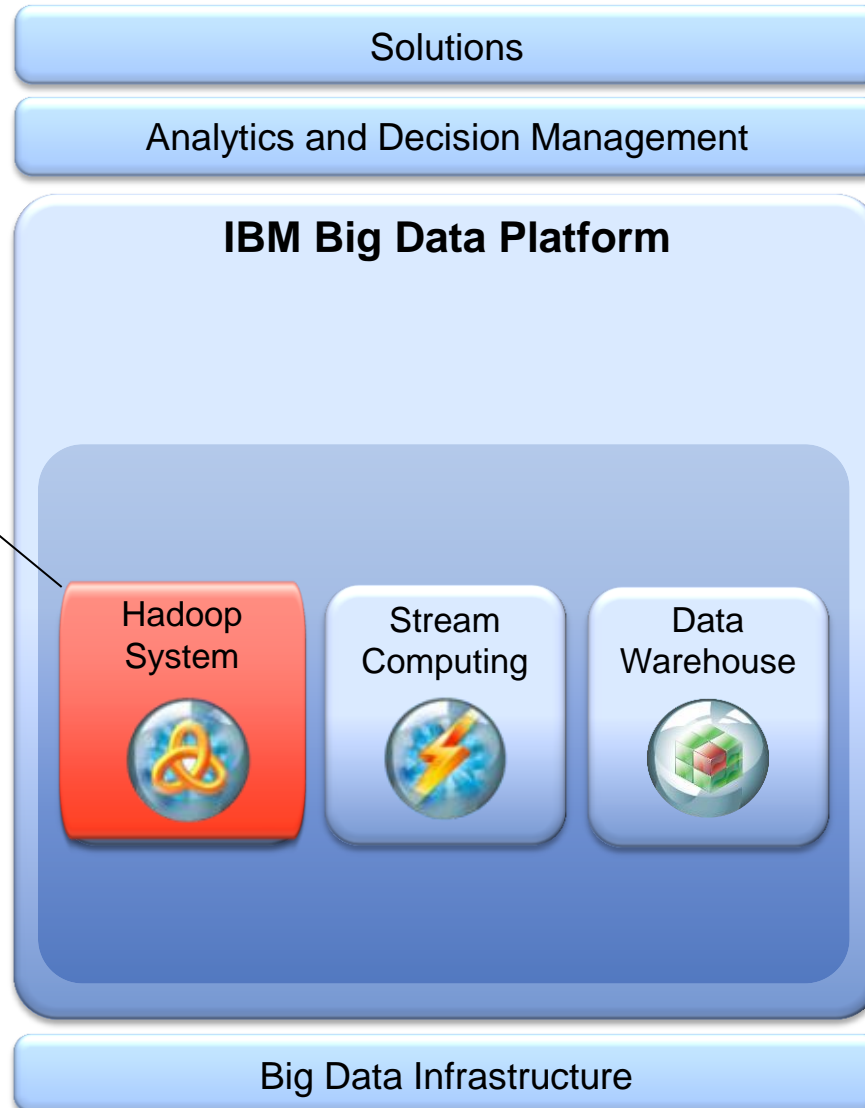
Analyze streaming data and large data bursts for real-time insights

- **InfoSphere Streams** – software enabling continuous analysis of massive volumes of streaming data with sub-millisecond response times

The Big Data Platform

Cost-effectively analyze Petabytes of unstructured and structured data

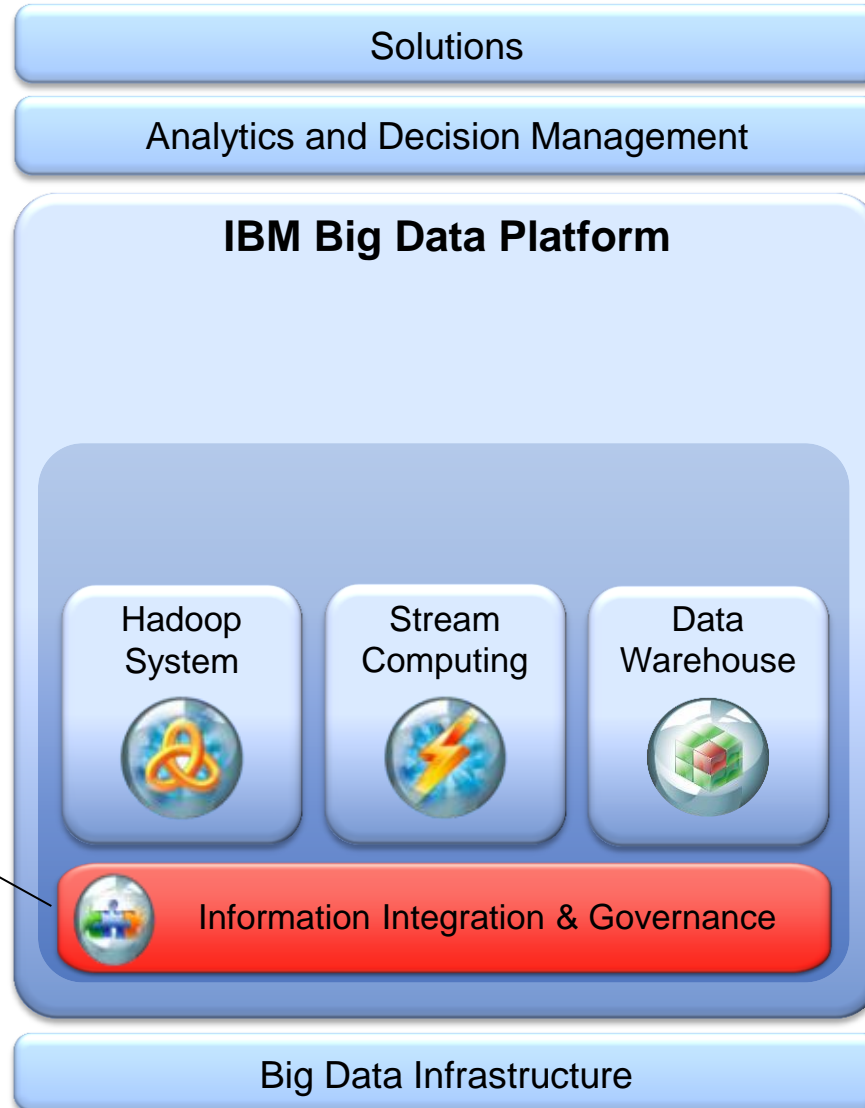
- **InfoSphere BigInsights** -- enterprise-grade Hadoop system enhanced with advanced text analytics, data visualization, tools, & performance features for analyzing massive volumes of structured and unstructured data.



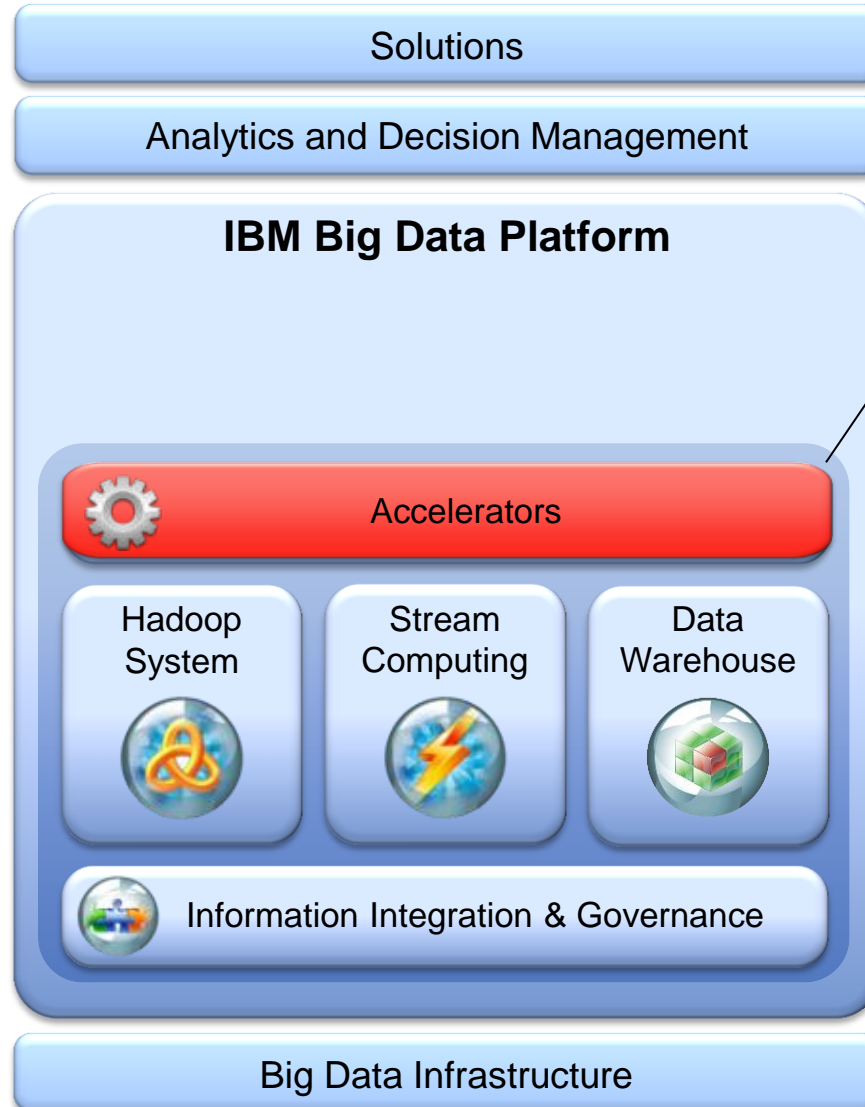
The Big Data Platform

Govern data quality and manage the information lifecycle

- **InfoSphere Information Server** –Cleanses data, monitors quality and integrates big data with existing systems
- **InfoSphere Optim** – manages business information throughout its lifecycle
- **InfoSphere Master Data Management** – manages and maintains trusted views of master and reference data
- **InfoSphere Guardium** – real-time database security and monitoring



The Big Data Platform



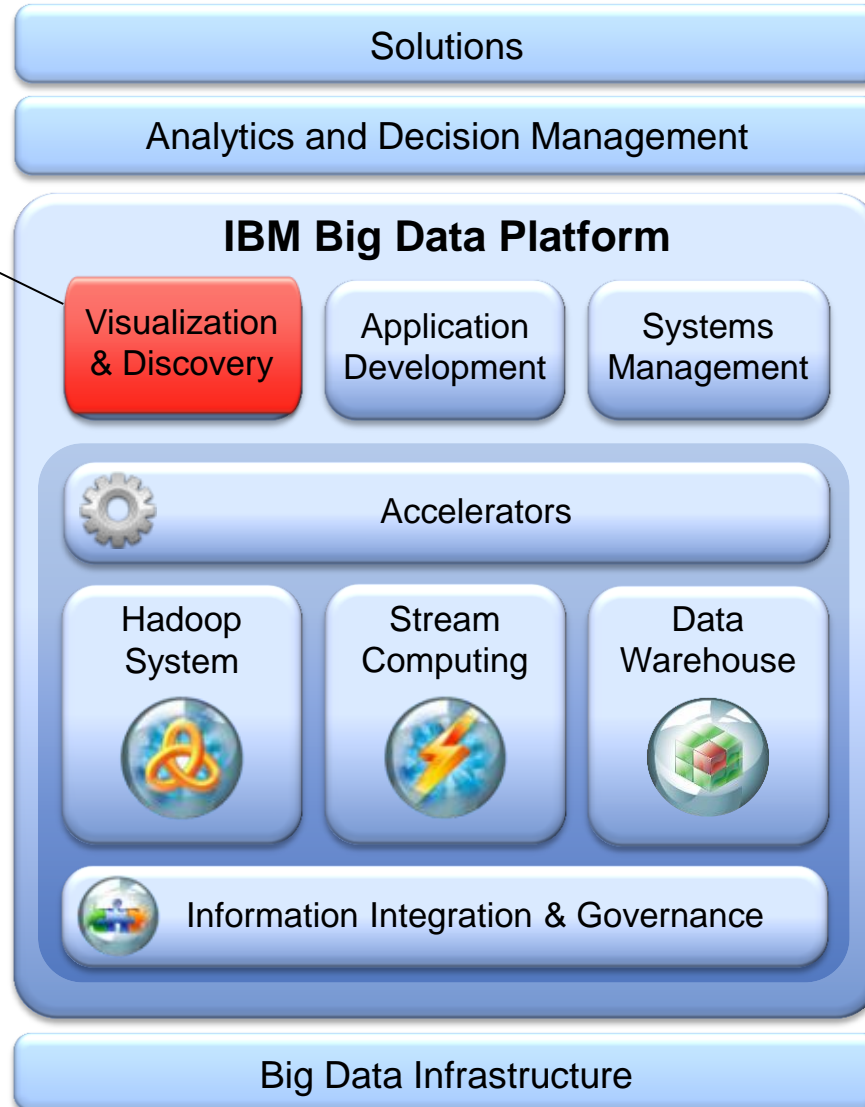
Speed time to value with analytic and application accelerators

- **Analytic Accelerators** – text analytics, geospatial, time-series, data mining
- **Application Accelerators** – financial services, machine data, social data, Telco event data
- **Industry Models** - comprehensive data models based on deep expertise and industry best practice

The Big Data Platform

Discover, understand, search, and navigate federated sources of big data

- **InfoSphere Data Explorer** – Discovery and navigation software that provides real-time access and fusion of big data with rich and varied data from enterprise applications for greater insight



The Big Data Platform

InfoSphere Data Explorer

Find, navigate, visualize all data

Accelerators

Speed time to value with analytic and application accelerators

InfoSphere BigInsights

Bringing Hadoop to the enterprise

InfoSphere Streams

Analytics for data in-motion exploration

PureData for Analytics & InfoSphere Warehouse

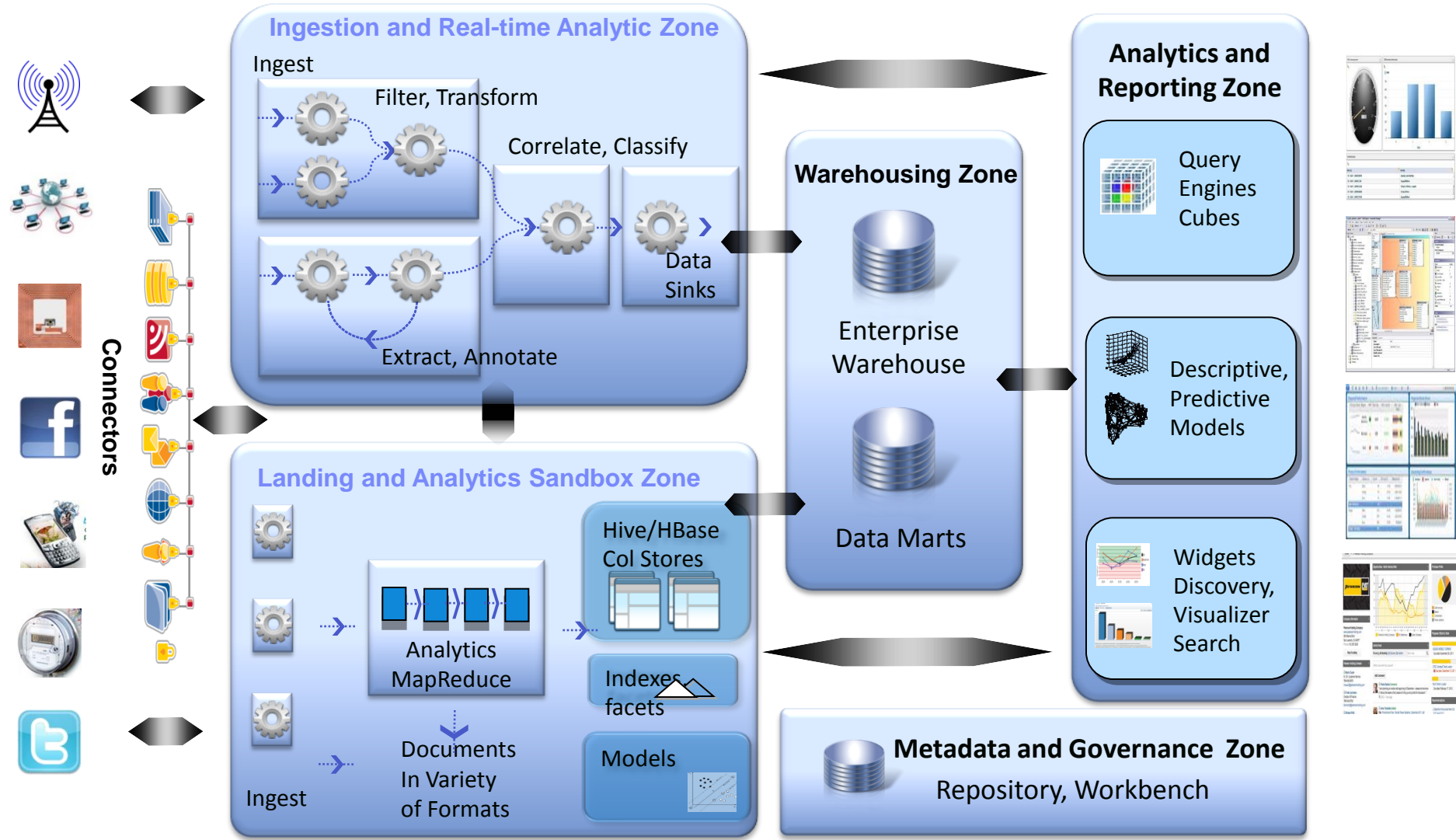
Delivers deep insight with advanced database analytics & operational analytics

Information Integration and Governance

Govern data quality and manage the information lifecycle

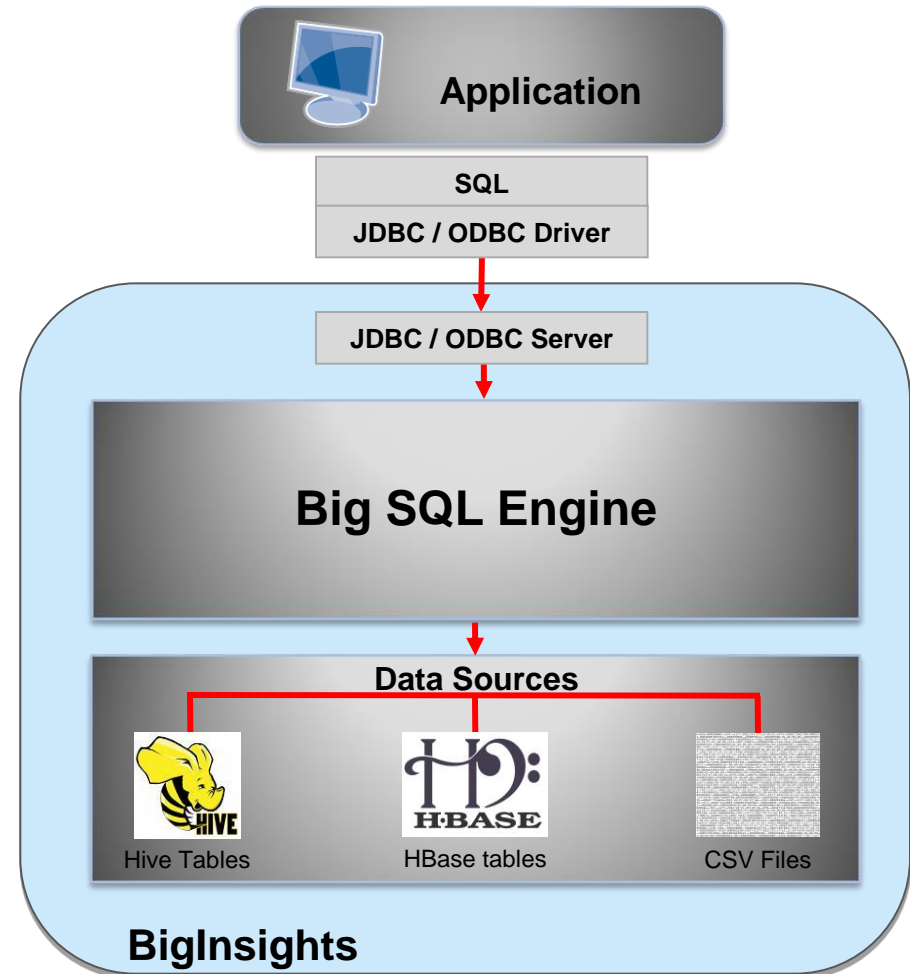


Emerging Big Data Implementation Pattern



Big SQL: Native SQL Query Access for Hadoop

- **Native SQL access to data stored in BigInsights**
 - ANSI SQL 92+
 - Standard syntax support (joins, data types, ...)
- **Real JDBC/ODBC drivers**
 - Prepared statements
 - Cancel support
 - Database metadata API support
 - Secure socket connections (SSL)
- **Optimization**
 - Leveraging MapReduce parallelism or...
 - Direct access for low-latency queries
- **Varied data sources**
 - HBase (including secondary indexes)
 - CSV, Delimited files, Sequence files
 - JSON
 - Hive tables

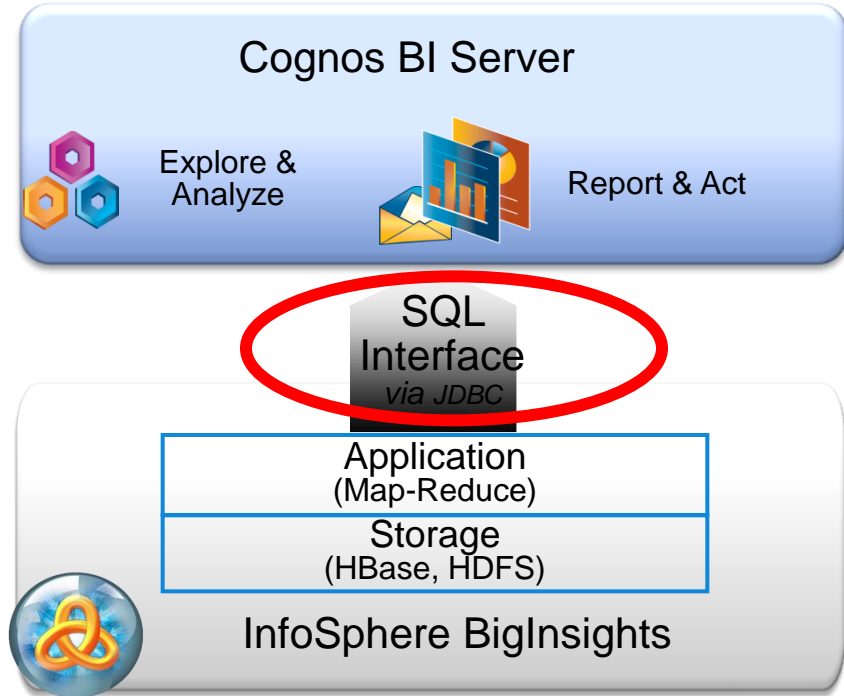


Cognos Business Intelligence Optimized for Big SQL

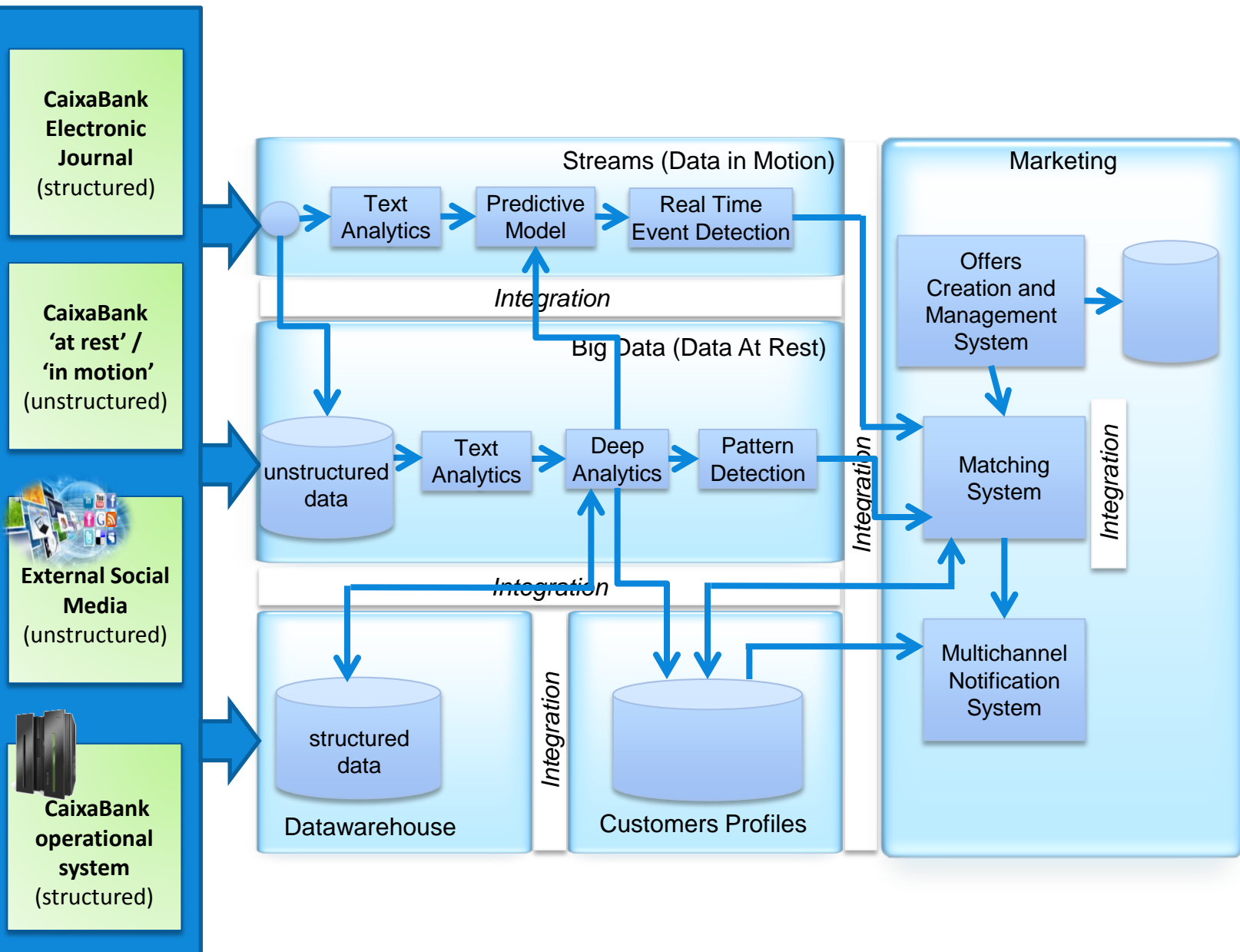
- **Cognos BI server can push down many computations to BigInsights**
 - Big SQL directs this processing to happen on BigInsights instead of the Cognos BI Server

- **Faster response times**
 - Increased opportunity for query processing to occur closer to the data

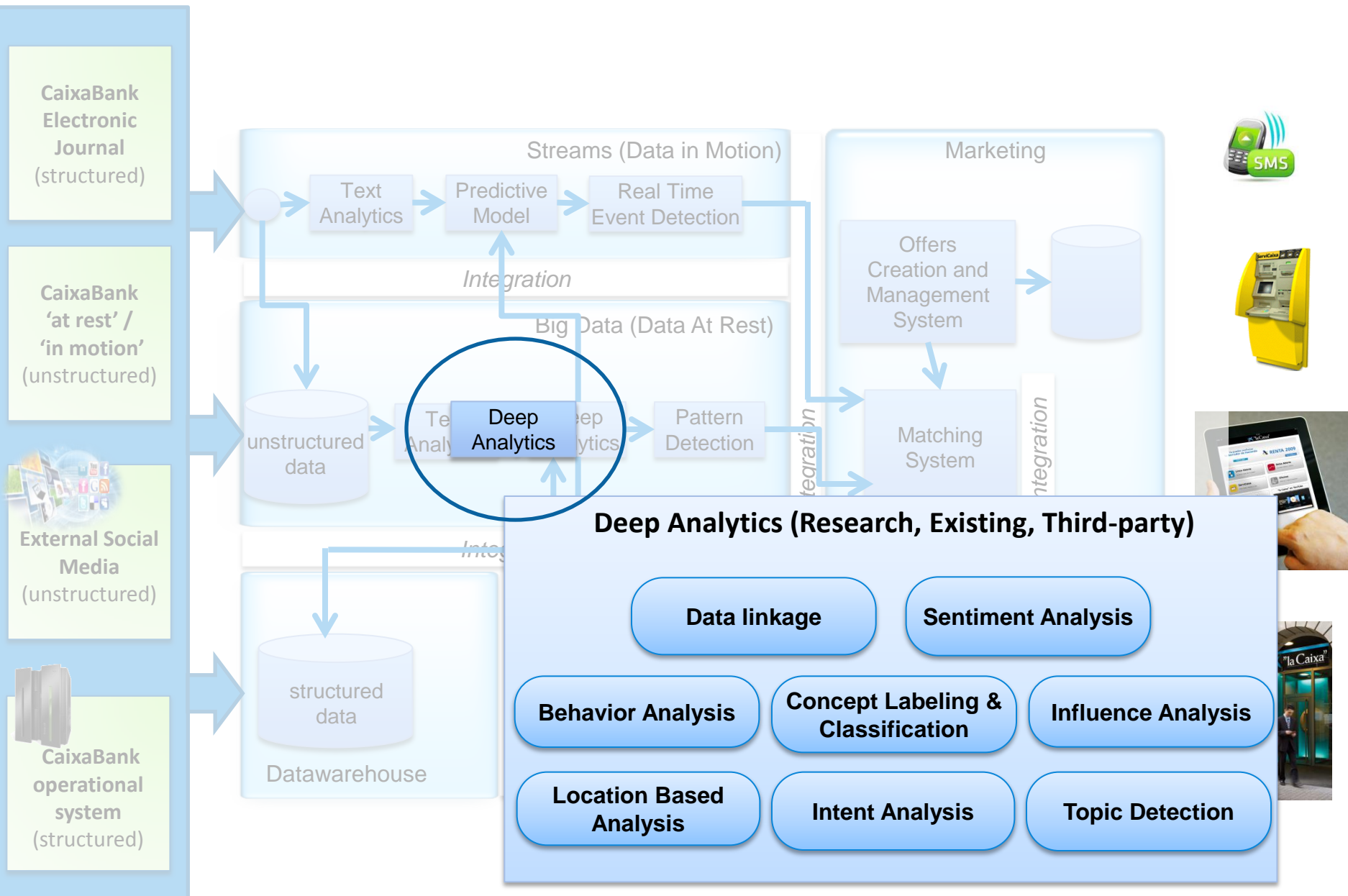
- **Free from the limitations of Hive (latency, SQL language support)**



Caixabank Big Data Reference Architecture



Caixabank Big Data Reference Architecture



A day in a BigData Life