



# **IBM Big Data Forum - Think Big**

## **Büyük Veri Perdesini Aralıyoruz**

### **Harnessing the Power of Big Data**

Dirk deRoos

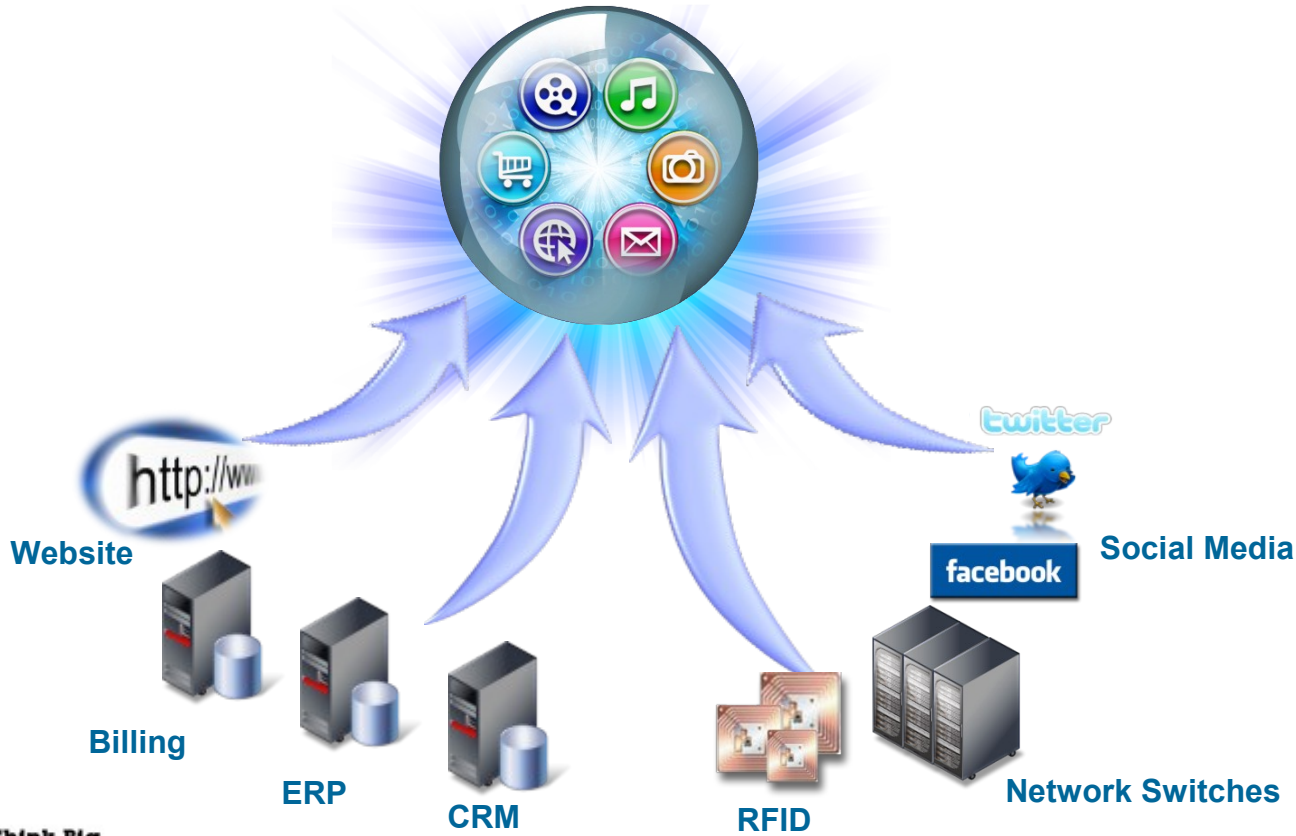
[dderoos@ca.ibm.com](mailto:dderoos@ca.ibm.com)

 [@Dirk\\_deRoos](https://twitter.com/Dirk_deRoos)

**World-Wide Technical Sales, IBM Big Data Platform**

# Big Data is a Hot Topic Because Technology Makes it Possible to Analyze ALL Available Data

Cost effectively manage and analyze *all available data, in its native form – unstructured, structured, streaming*



# The Big Data Conundrum

- The economics of deletion have changed....
- 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” over time

Data AVAILABLE to  
an organization

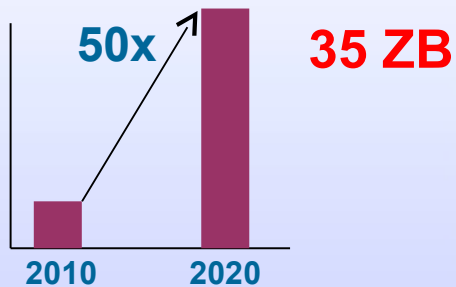
Signals  
and  
Noise



Data an organization  
can PROCESS

# The Characteristics of Big Data

Cost efficiently processing the growing **Volume**



Responding to the increasing **Velocity**



Collectively analyzing the broadening **Variety**



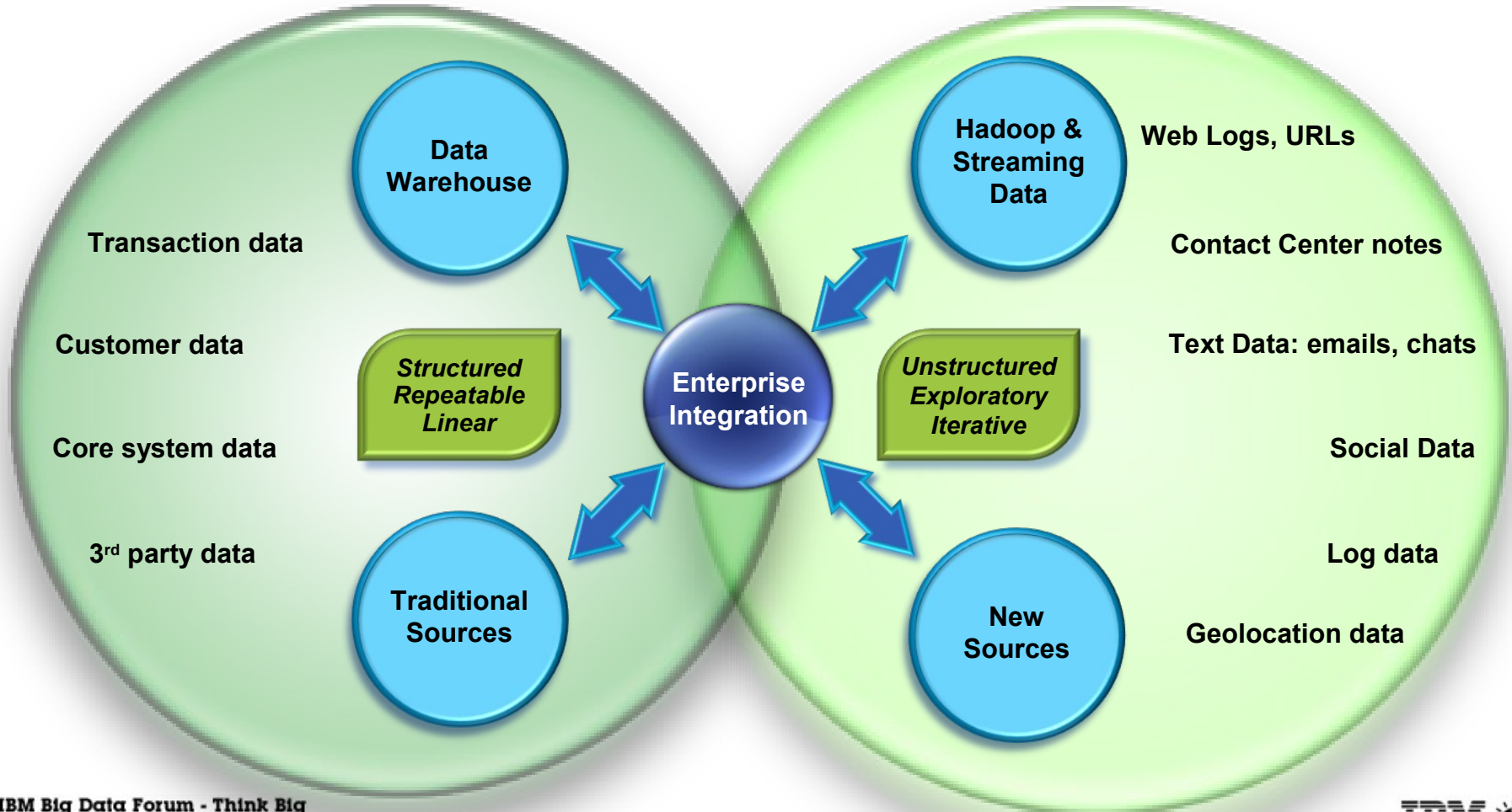
Establishing the **Veracity** of big data sources

**1 in 3** business leaders don't trust the information they use to make decisions

# Analytics is Expanding from Enterprise Data to Big Data

**Traditional Approach**  
*Structured, analytical, logical*

**New Approach**  
*Creative, holistic thought, intuition*





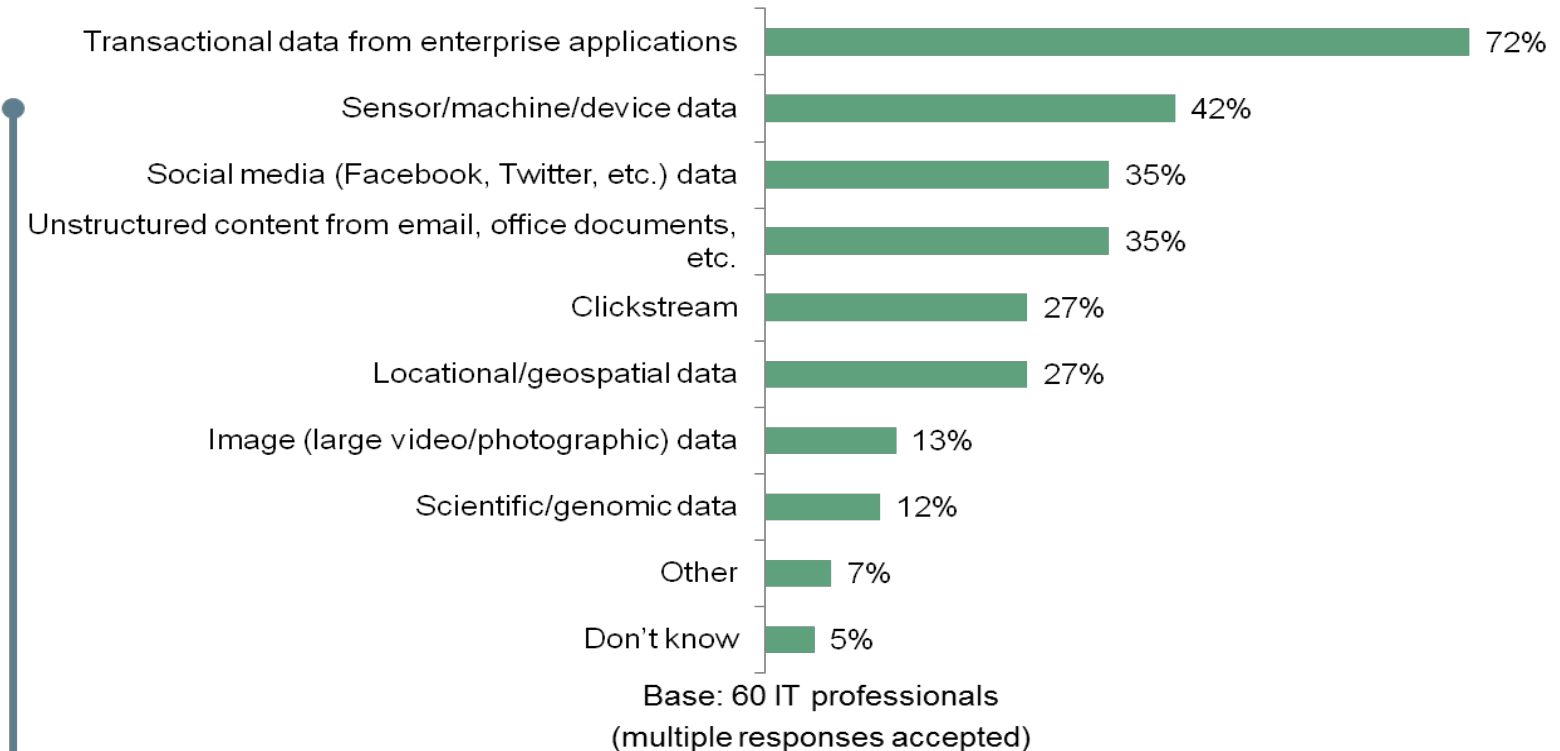
# Why Didn't We Use All of the Big Data Before?



# Requirement: Analyze Many Sources of Data

Big data: across diverse subject domains

“What types of data/records are you planning to analyze using big data technologies?”



Most big data use cases hype its application for analysis of new, raw data from social media, sensors, and web traffic, but we found that firms are being very practical, with early adopters using it to operate on enterprise data they already have.

# The 5 Key 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 by incorporating additional internal and external data 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



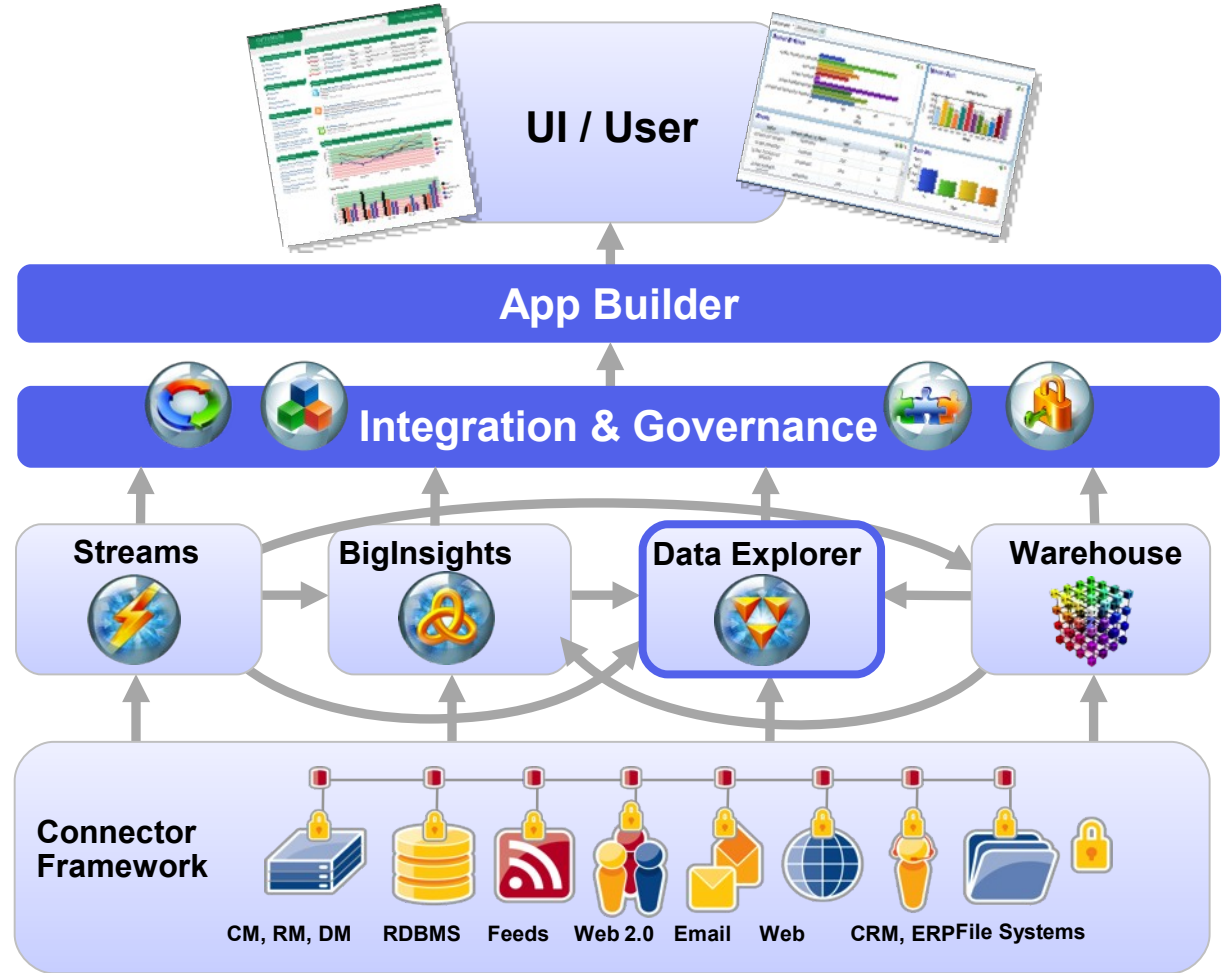
## Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency



# Big Data Exploration

- **Find, visualize and understand all big data to improve business knowledge**
  - Greater efficiencies in business processes
  - New insights from combining and analyzing data types in new ways
  - Develop new business models with resulting increased market presence and revenue



# Enhanced 360° View of the Customer

## SOURCE SYSTEMS



**CRM**  
Name: **J Robertson**  
Address: **35 West 15<sup>th</sup>**  
Address: **Pittsburgh, PA 15213**



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



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

Master Data Management



## 360° View of Party Identity


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



**OPTIMUM INVESTMENTS**

Home > Janet Robertson

Logged in as Frank Gelato | Help



**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...

Fulfillment

External Sources

Content Mgt.

Knowledge-bases

BigInsights

Support Ticketing

CRM

Email

Supply Chain

Streams

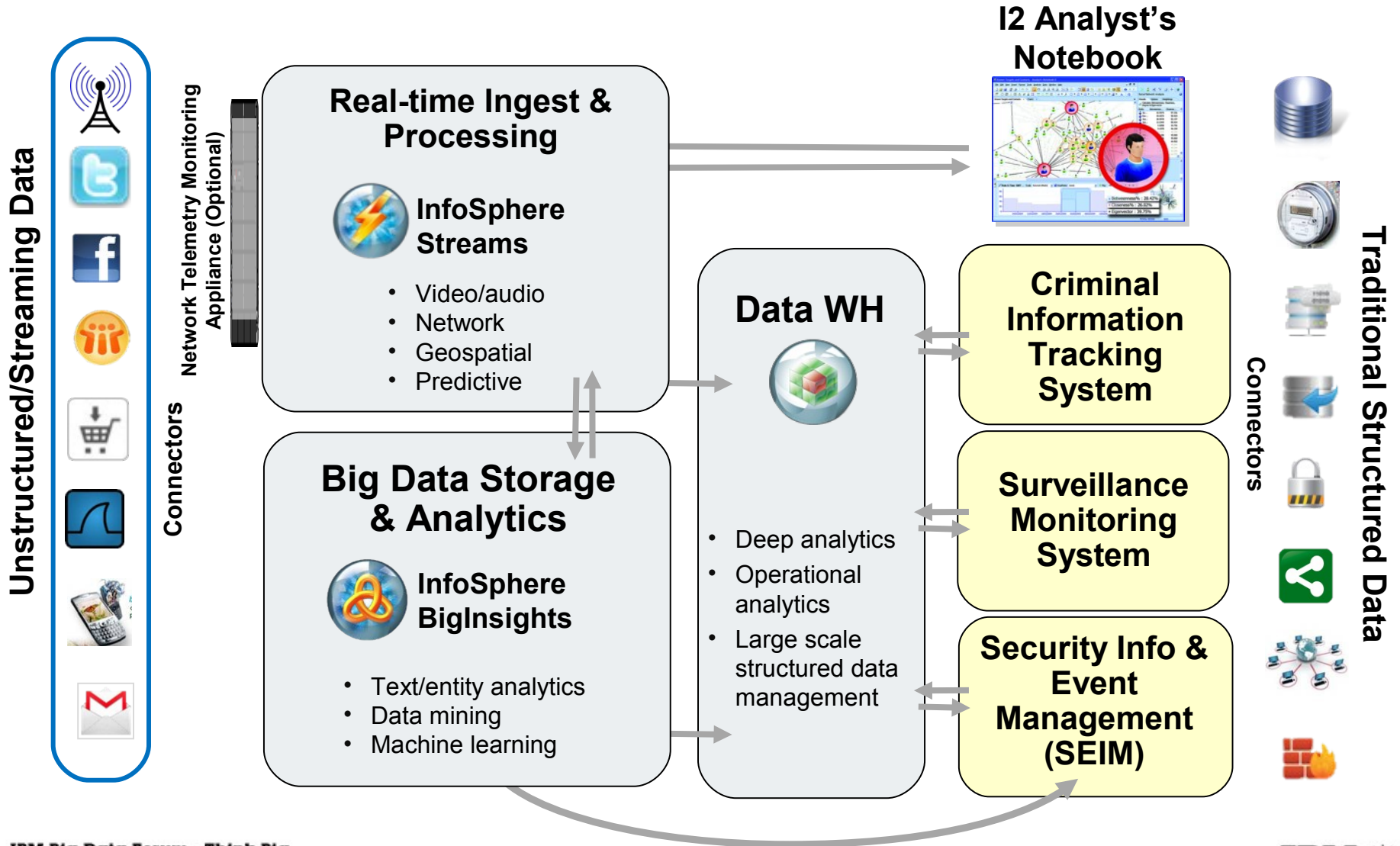
Warehouse

Unified View of Party's Information

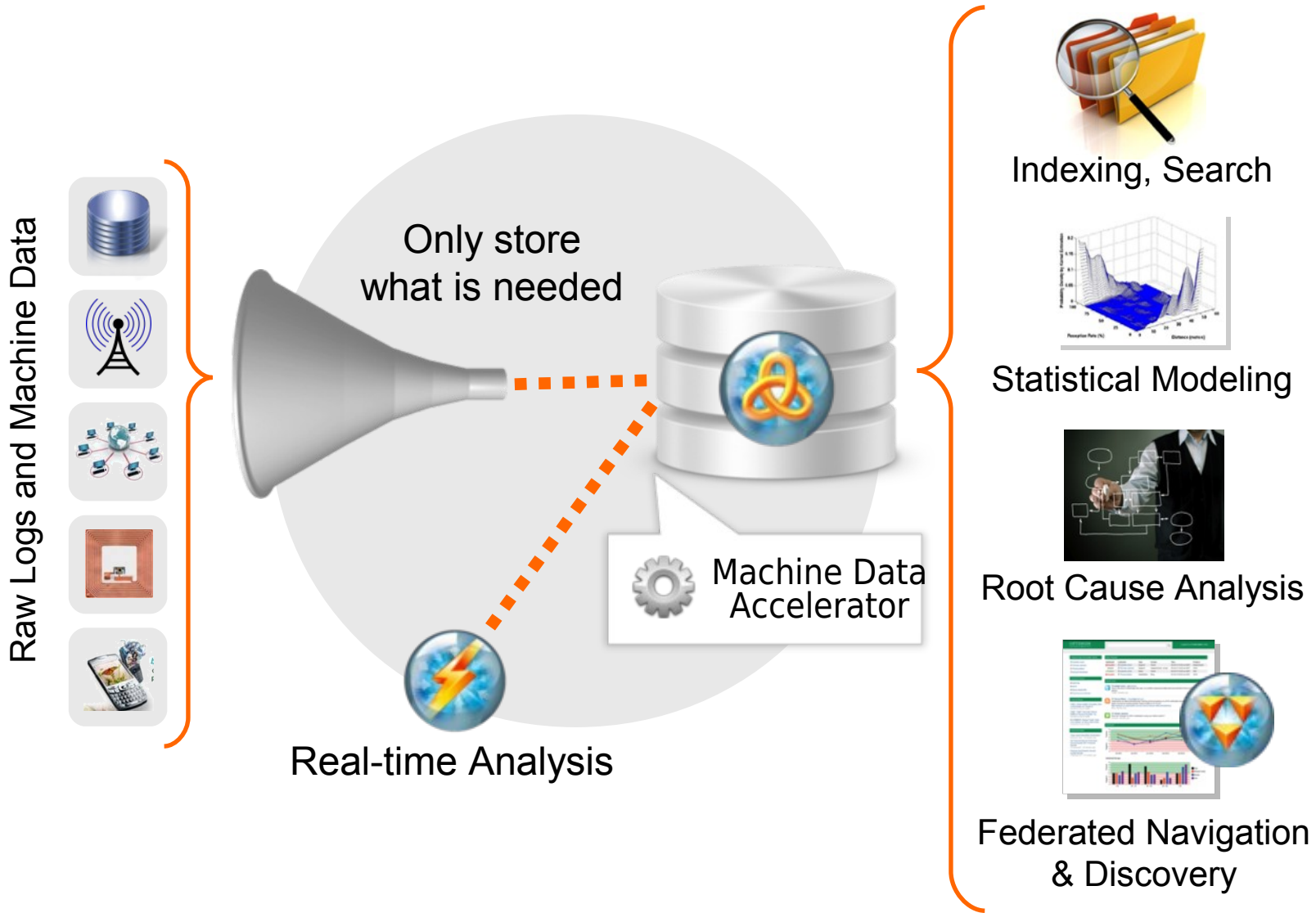
Büyük Veri Perdesini Aralıyoruz



# Security/Intelligence

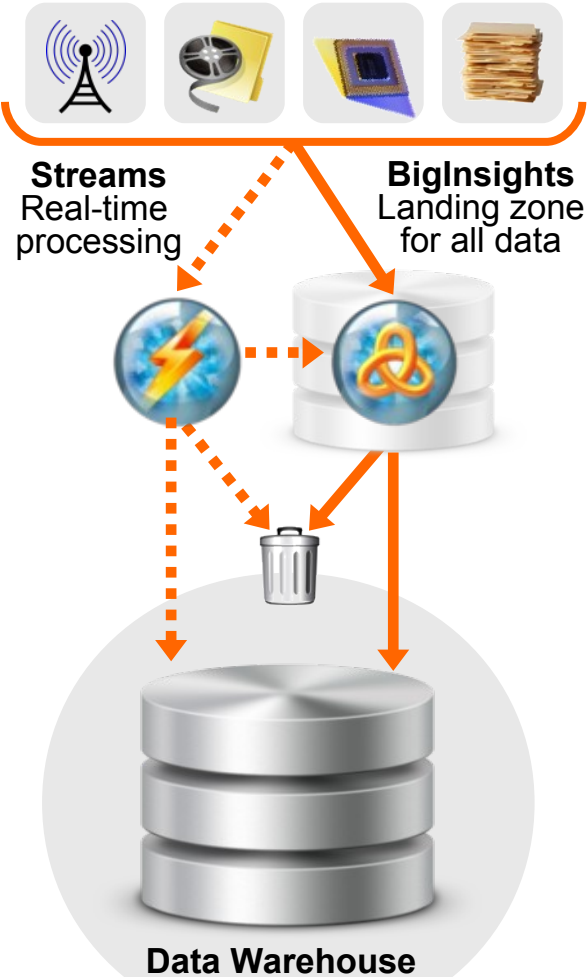


# Operations Analysis

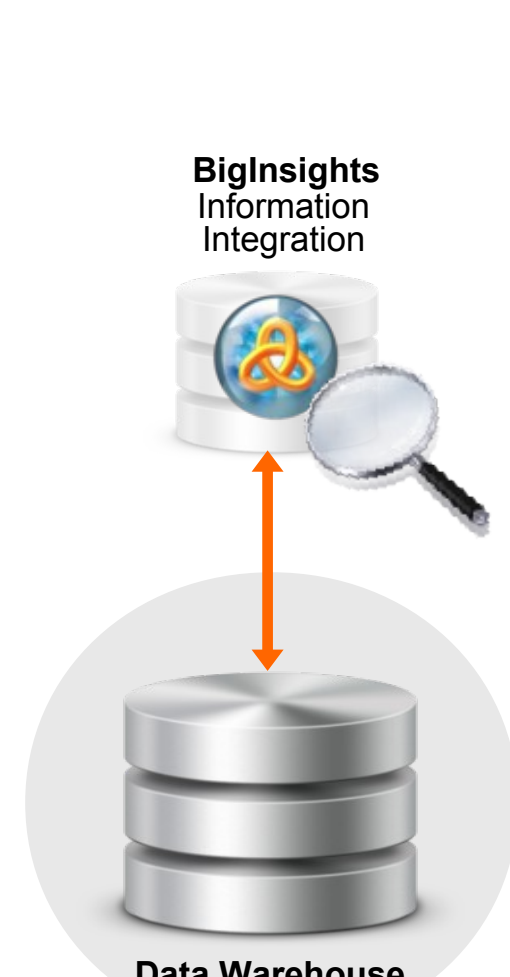


# Data Warehouse Augmentation

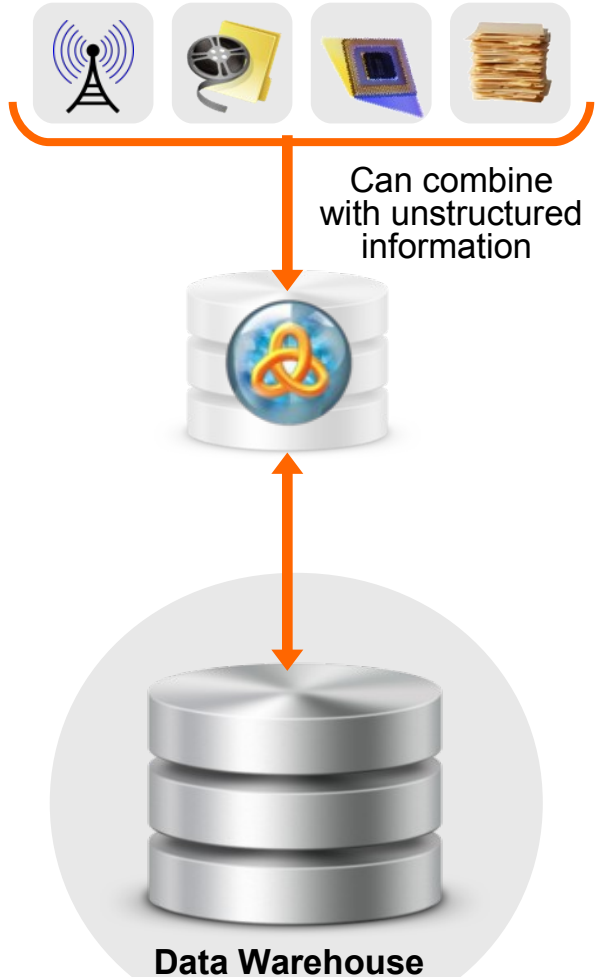
## ● Pre-Processing Hub



## ● Query-able Archive

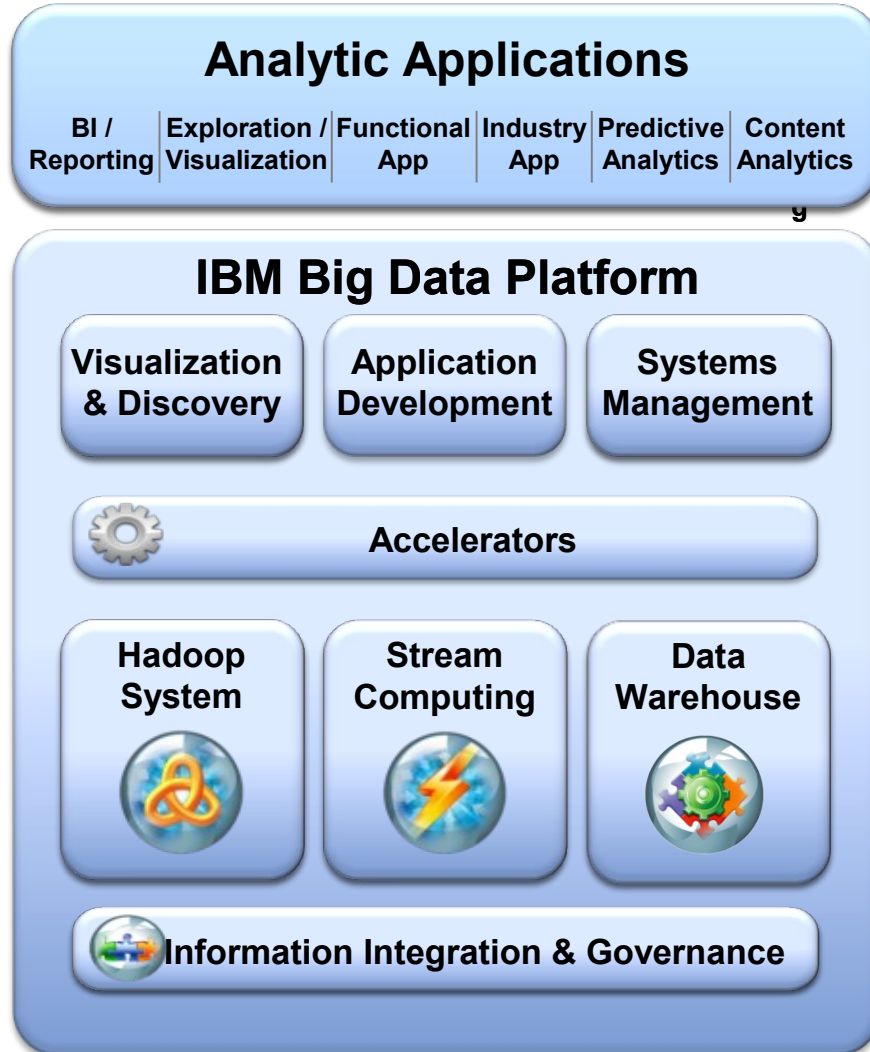


## ● Exploratory Analysis





# New Analytic Applications Demand a Big Data Platform



# Part of the IBM Big Data Platform

*Workload Optimized Solutions for All Your Analytic Needs*



*PureData System for Analytics*  
*PureData System for Operational Analytics*



# Backed by a Full Commitment to Open Source, IBM Significantly Extends Hadoop



- **Scalable**
  - New nodes can be added on the fly
- **Affordable**
  - Massively parallel computing on commodity servers
- **Flexible**
  - Hadoop is schema-less, and can absorb any type of data
- **Fault tolerant**
  - Through MapReduce framework



## IBM Innovation

- **Hadoop analytics for all users**
  - ‘App store’ interface, wizard-driven publication of apps, spreadsheet-based analytics, custom visualization
- **Enterprise integration**
  - Data exchange between enterprise data sources and analytics tools
- **Security and administration**
  - Single point of access, LDAP, audit logs, flexible admin GUI tools
- **Performance and reliability**
  - Adaptive MapReduce, Indexing, Compression, Flexible Scheduler

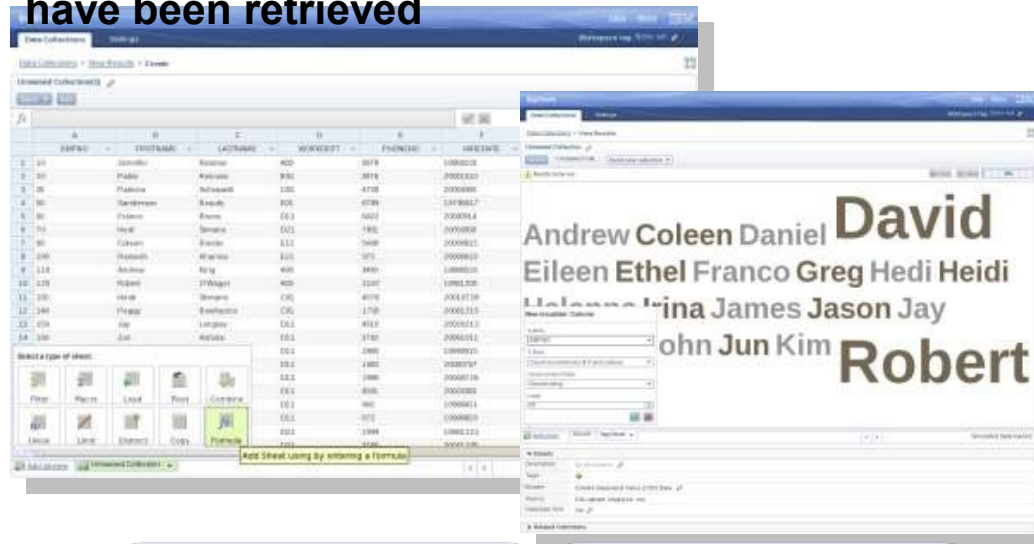
# Spreadsheet-style data visualization enables ad-hoc analytics and analysis of all data

 **InfoSphere BigInsights**

- Visualization & Exploration
- Development Tools
- Advanced Engines
- Connectors
- Workload Optimization
- Administration & Security
- Open source Hadoop components

## BigSheets Visualization and Exploration

- Web-based analysis and visualization for Users
- Familiar spreadsheet-like interface
- Define and manage long running data collection jobs
- Analyze content of the text on the pages that have been retrieved



**Gather**  
Crawl – gather statistically  
Adapter–gather dynamically

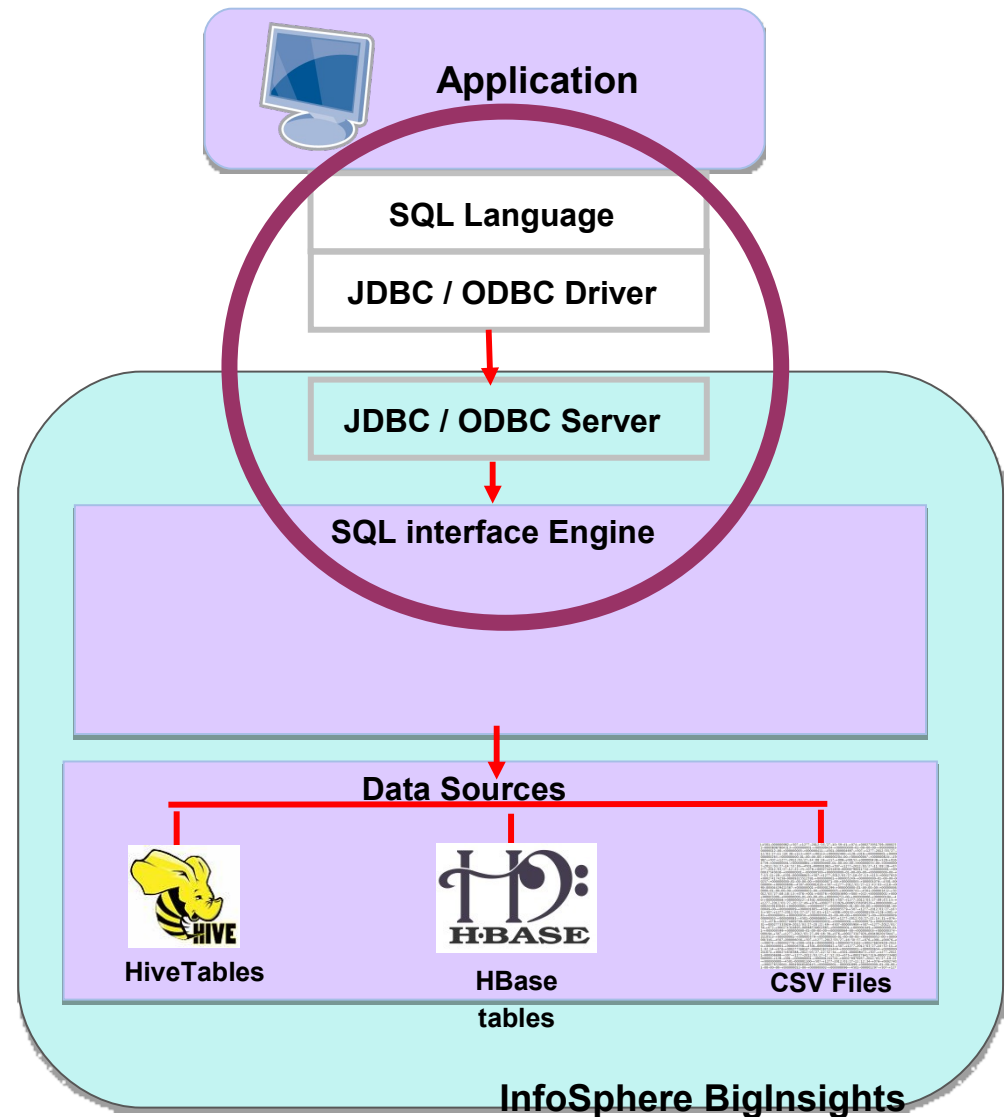
**Extract**  
Document-level info  
Cleanse, normalize

**Explore**  
Analyze, annotate, filter  
Visualize results

**Iterate**  
Iterate through any prior step

# BigInsights Future: The SQL interface . . . .

- **Rich SQL query capabilities**
  - SQL '92 and 2011 features
  - Correlated subqueries
  - Windowed aggregates
- **SQL access to all data stored in InfoSphere BigInsights**
- **Robust JDBC/ODBC support**
- **Take advantage of key features of each data source**
- **Leverage MapReduce parallelism OR achieving low-latency**





# IBM InfoSphere Streams v3.0

## A platform for real-time analytics on BIG data

### ■ Volume

- Terabytes per second
- Petabytes per day

### ■ Variety

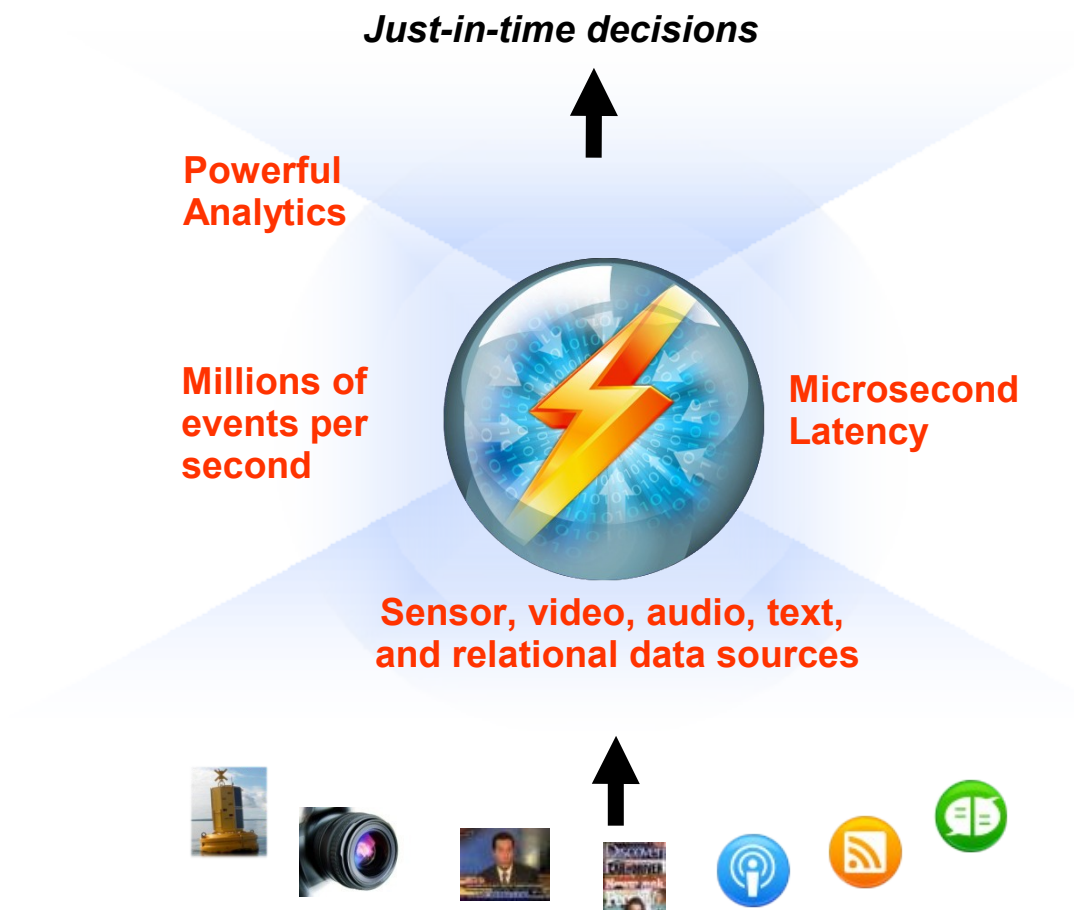
- All kinds of data
- All kinds of analytics

### ■ Velocity

- Insights in microseconds

### ■ Agility

- Dynamically responsive
- Rapid application development



# How Streams Works

- Continuous ingestion
- Continuous analysis





# What Are People Doing With Streams?



## Telephony

- CDR processing
- Social analysis
- Churn prediction
- Geomapping



## Transportation

- Intelligent traffic management



## Smart Grid & Energy

- Transactive control
- Phasor Monitoring Unit



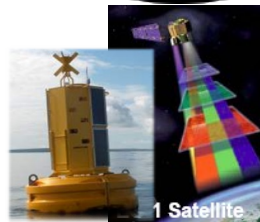
## Health & Life Sciences

- Neonatal ICU monitoring
- Epidemic early warning system
- Remote healthcare monitoring



## Stock market

- Impact of weather on securities prices
- Analyze market data at ultra-low latencies



## Natural Systems

- Wildfire management
- Water management

## Law Enforcement, Defense & Cyber-Security

- Real-time multimodal surveillance
- Situational awareness
- Cyber security detection



## Fraud prevention

- Detecting multi-party fraud
- Real-time fraud prevention



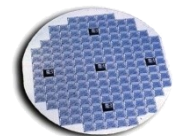
## e-Science

- Space weather prediction
- Detection of transient events
- Synchrotron atomic research

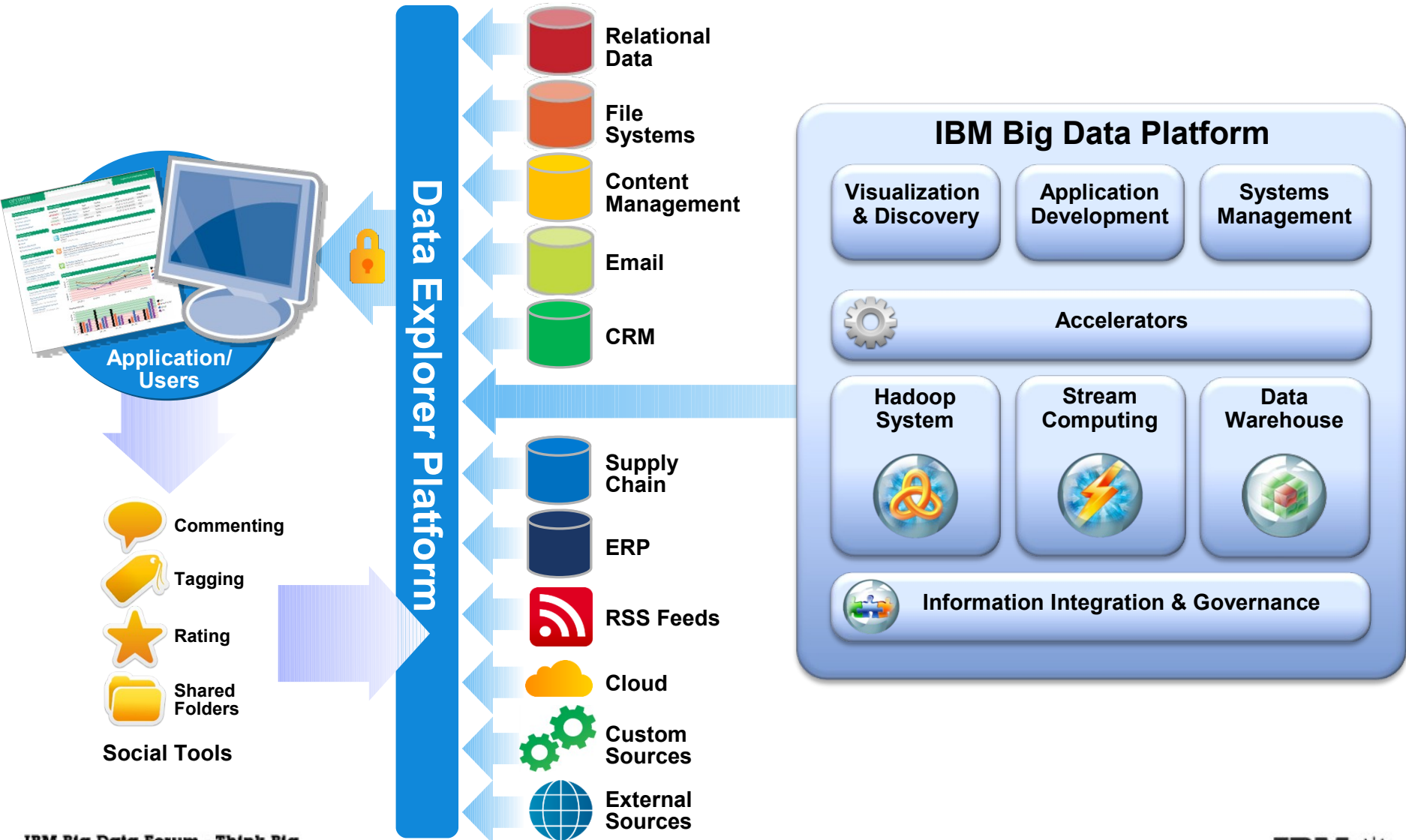


## Other

- Manufacturing
- Text Analysis
- Who's Talking to Whom?
- ERP for Commodities
- FPGA Acceleration



# Combining Enterprise Content with Big Data





Access across many sources

All Lotus Notes SharePoint Documentum Employees Trade Publications Desktop

Search: collaboration

Your query has been expanded. Show Expansions

Results 1-10 of about 449 Details

Dynamic categorization

Expertise location

Leveraging Structured and unstructured content

Enhanced by social collaboration

Organize content into virtual folders

Highly relevant, personalized results

Refinements based on structured information

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


- 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 (6)
- Analyst (5)
- Competitor (5)
- Velocity 6.0 Research (5)
- BI (3)
- Collaboration (2)
- Social Tagging (2)
- Euro (1)
- Events (1)

0 documents selected. Actions Select/deselect all on this page

**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)

1. Confirmation Package for the Gartner Portals, Content & Collaboration Summit

Fri, 9 Feb 2007 13:47:55 -0400 | From: Gartner Events | To: monarko@vivisimo.com

Gartner Portals, Content & Collaboration Summit CONFIRMATION PACKAGE Dear Attendee: Thank you for your registration to the Gartner Portals, Content & Collaboration at the Hyatt Regency Grand Cypress in Orlando, FL. I have enclosed the ... PACKAGE Dear Attendee: Thank you for your registration to the Gartner Portals, Content & Collaboration at the Hyatt Regency Grand Cypress in Orlando, FL. I have enclosed the ...

My Tags: Tradeshow Gartner PCC

In Marketing: Analyst Mobile

In Vivisimo: BI Competitor Security

200K - Lotus Notes - Rate result: ☆☆☆☆☆ - ignore

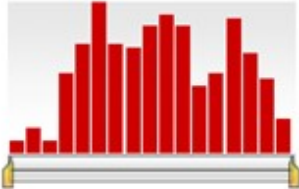
2. RE: Portals, Collaboration & Content Conference brochure deadline

Fri, 30 Jun 2006 10:43:01 -0400 | From: Kevin Scolio | To: Stacy Monarko

... Boston and in the Spring of 2007 to meet with many more portal, collaboration ...


Graphical Refinements

Date



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

Average Rating

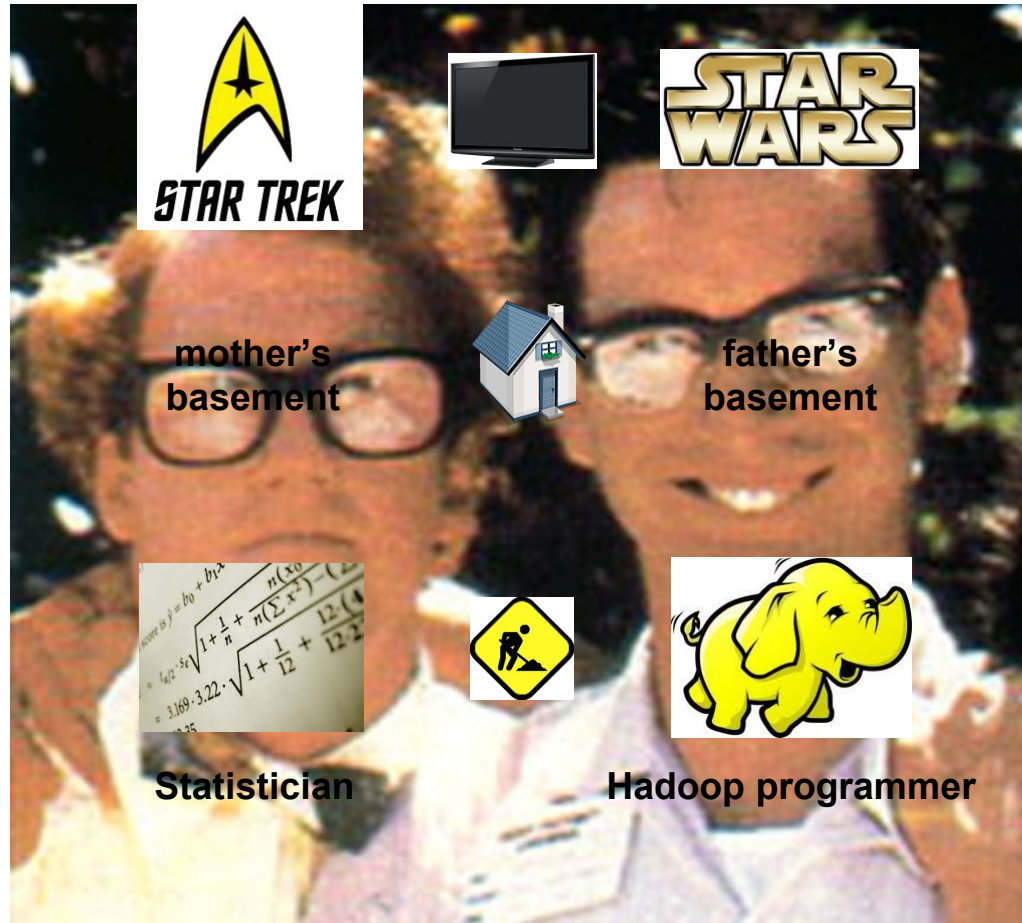


Selected: 1 - 4.5

Folders

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

# A Tale of Two Nerds



They're similar, but not the same.

Most quants don't have Hadoop skills.

**THINK**

**BIG**

**BIG**