

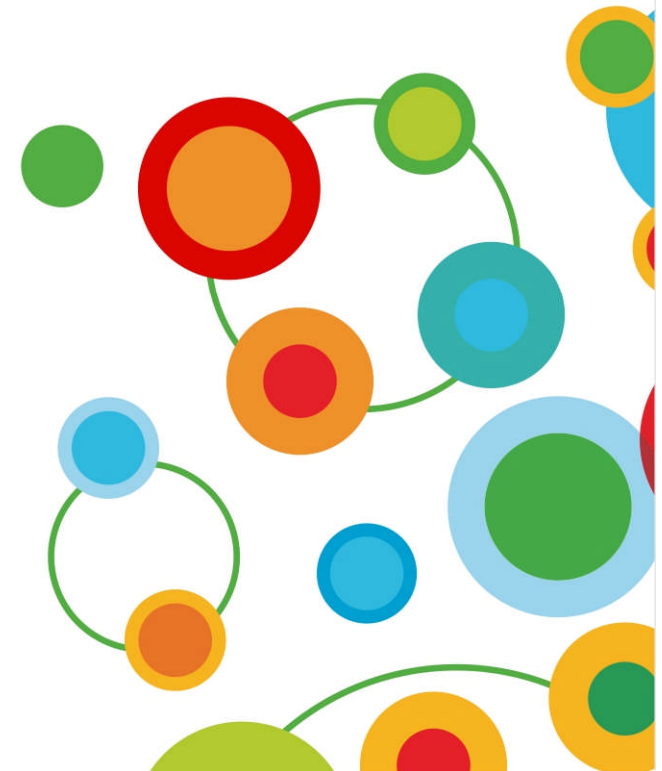


IBM Software

Information Management & Analytics Forum **2013**

Return on Information: The New ROI

**Bringing the power of Big Data to the
Enterprise – Part I**



The number of organizations who see analytics as a competitive advantage is growing.



analytics IQ
business initiative

Studies show that organizations competing on analytics outperform their peers

substantially outperform

220%

1.6x

Revenue
Growth


2.5x

Stock Price
Appreciation

2.0x


EBITDA
Growth

**IBM delivers a governable,
consumable Big Data platform that's
steeped in analytics for data in-motion
and data at-rest.**





C&A Brasil
about an hour ago via PostCron


Calças com a barra dobrada continuam sendo um sucesso entre os homens. Abuse! — with Michel Almeida.



Like · Comment · Share 35

 Giselle Magallanes, Cesar Santos and 213 others like this.

 Fuchs Erick e pra os que ficam falando de 'isso eh a ultima moda na europa', homophobic seal pra vcs ahuaahuahuaahua
31 minutes ago

 Giovane Farias a inveja tem facebook galera!
20 minutes ago



An increasingly sensor-enabled and instrumented business environment generates **HUGE** volumes of data with **MACHINE SPEED** characteristics...

1 **BILLION** lines of code
EACH engine generating 10 TB every 30 minutes!

The Social Layer in an Instrumented Interconnected World

Tweet data
generated daily
12+ TBs



? TBs of
data every day

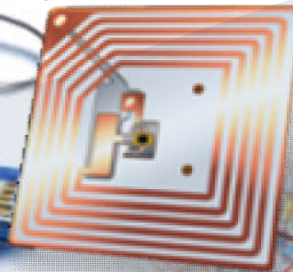


500+ TBs of
data per day

2.5B+ item shares
2.5B+ likes
300M photos

...
Facebook once
added a new user
every second for 3
years!

30 billion RFID
tags today
(1.3B in 2005)

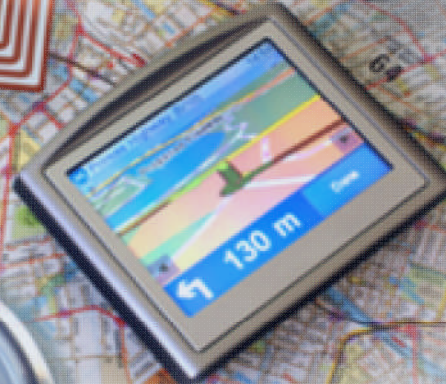


76 million smart
meters in 2009...
200M by 2014

4.6 billion
camera
phones
world
wide



*100s of
millions of
GPS
enabled
devices
sold
annually*



2+ billion
people
on the
Web by
end



2011

Beijing 2008™



83x

6,000,000 users on Twitter
pushing out 300,000
tweets per day

500,000,000 users on Twitter
pushing out 400,000,000
tweets per day

1333x

2011

1.8 ZB

2009

1 ZB

1 ZB=1T GB

2019
2020

35 ZB

1 ZB = 1B TB

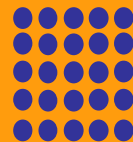
1 ZB = 1T GB



We've Moved into a New Era of Computing

12+ terabytes

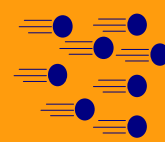
of Tweets
create daily.



Volume

5+ million

trade events
per second.



Velocity

100's

of different types of data.



Variety



Veracity

Only **1 in 3**

decision makers trust
their information.

5 Big Data Patterns



Big Data Exploration

Find, visualize, understand all big data to improve business knowledge



Enhanced 360° View of the Customer

Achieve a true unified view, incorporating internal and external 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

Applications for Big Data Analytics

Smarter Healthcare



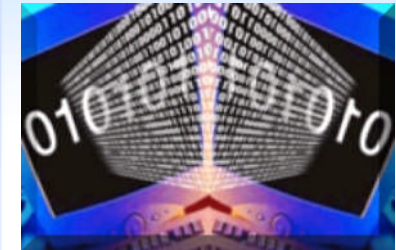
Multi-Channel



Finance



Log Analysis



Homeland Security



Traffic Control



Telecom



Search Quality



Manufacturing



Trading Analytics



Fraud and Risk



Retail: Churn, NBO



A Big Data Platform Manifesto

CONSUMABLE
(What not How/Patterns/Expert Systems, +++)

Understand and Navigate
Federated Big Data Sources



Federated Discovery
and Navigation

Manage and Store Huge
Volume of any Data



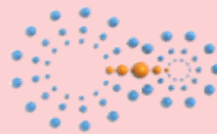
Hadoop File System
MapReduce

Structure and Control Data



Data Warehousing

Manage Streaming Data



Stream Computing

Analyze Unstructured Data



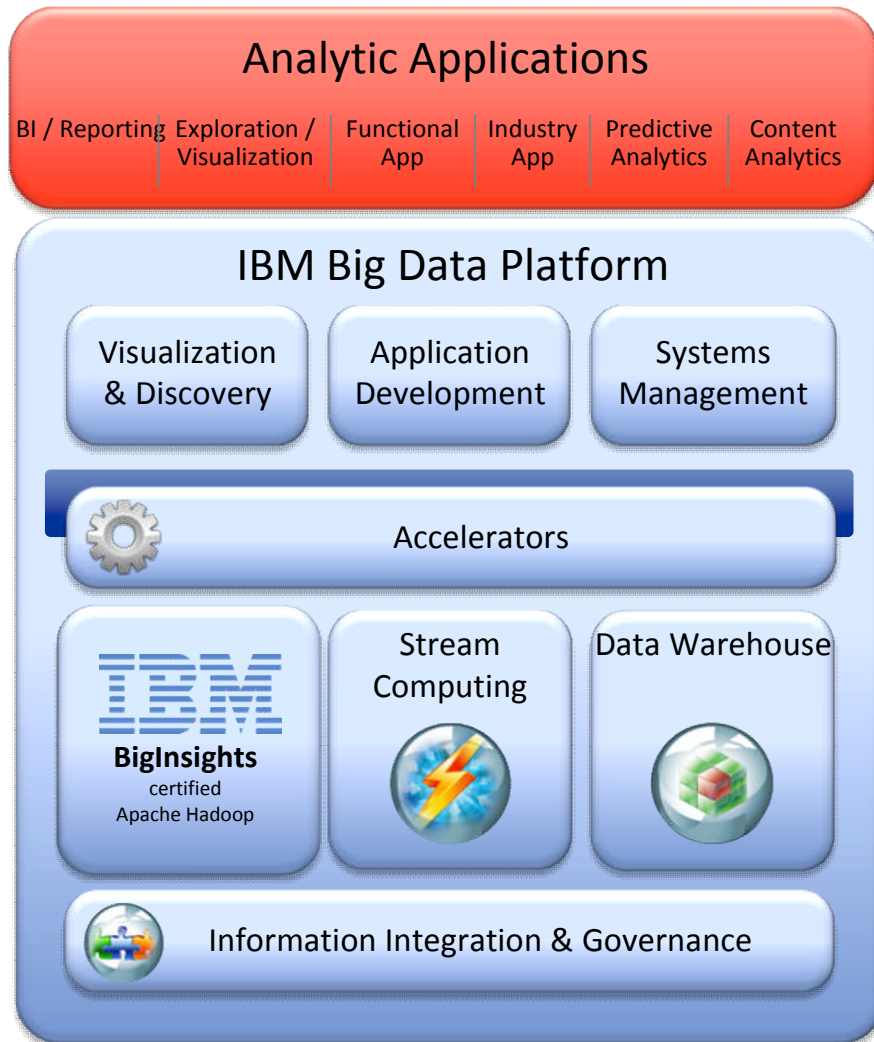
Text Analytics Engine

Integrate and Govern all
Data Sources



Integration, Data Quality,
Security, ILM, MDM

The IBM Big Data Platform Advantage



“IBM has the deepest Hadoop platform and application portfolio. IBM, an established EDW vendor, has its own Hadoop distribution; an extensive professional services force working on Hadoop projects; extensive R&D programs developing Hadoop technologies; connections to Hadoop from its EDW.” –The Forrester Wave™: 1Q12



“IBM InfoSphere BigInsights is a core capability of the most comprehensive Big Data analytics platforms out there right now...” – Krishna RoyLars

How to Analyze Unstructured Data

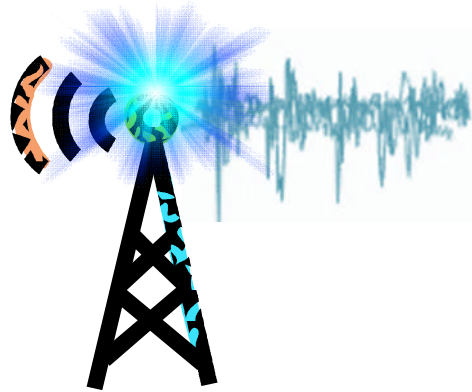
IBM Software

Information Management & Analytics Forum 2013

Return on Information: The New ROI



How Text Analytics Works

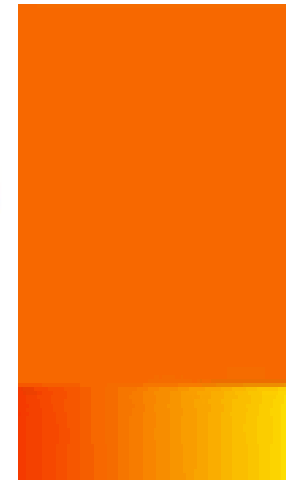
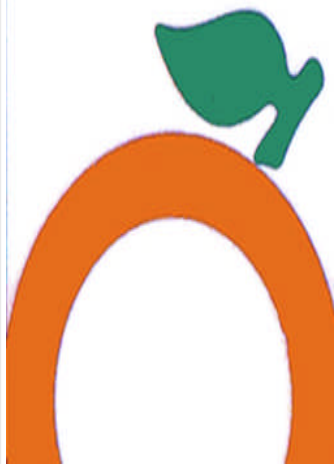


Football **World Cup 2010**, one team distinguished themselves well, losing to the eventual champions 1-0 in the Final. Early in the second half, **Netherlands' striker, Arjen Robben**, had a breakaway, but the **keeper** for **Spain**, **Iker Casilas** made the save. **Winger Andres Iniesta** scored for **Spain** for the win.

World Cup 2010 Highlights

Name	Position	Country
Arjen Robben	Striker	Netherlands
Iker Casilas	Keeper	Spain
Andres Iniesta	Winger	Spain

The Tricky Thing About Sentiment...



The Path to Consumability: Declarative Languages

Offline

Runtime

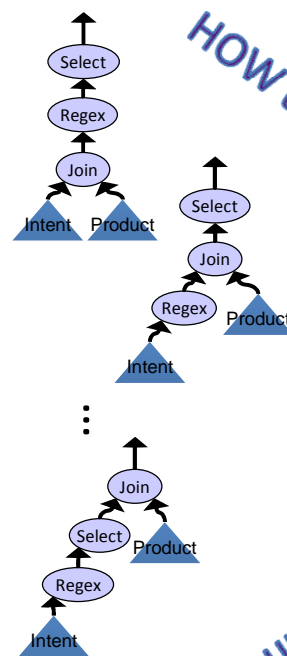
Development Environment

WHAT to EXTRACT

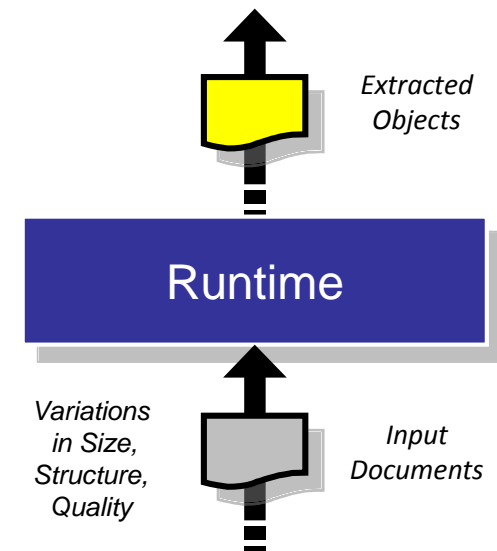
Declarative Language

```
create view MonetizableIntent as
select P.name as product,
       I.clue as strength
from   Intent I, Product P
where
  Follows(I.clue, P.name, 0, 20)
and   Not(ContainsRegex(/\b(not)\b/,
                        LeftContext(I.clue, 10)));
```

- Streams, Text Analytics, Machine Learning, SQL all are declarative, simple to learn languages
- All have strong development tooling and accelerators
- IBM has been optimizing declarative languages for decades, IN FACT, IBM INVENTED IT!

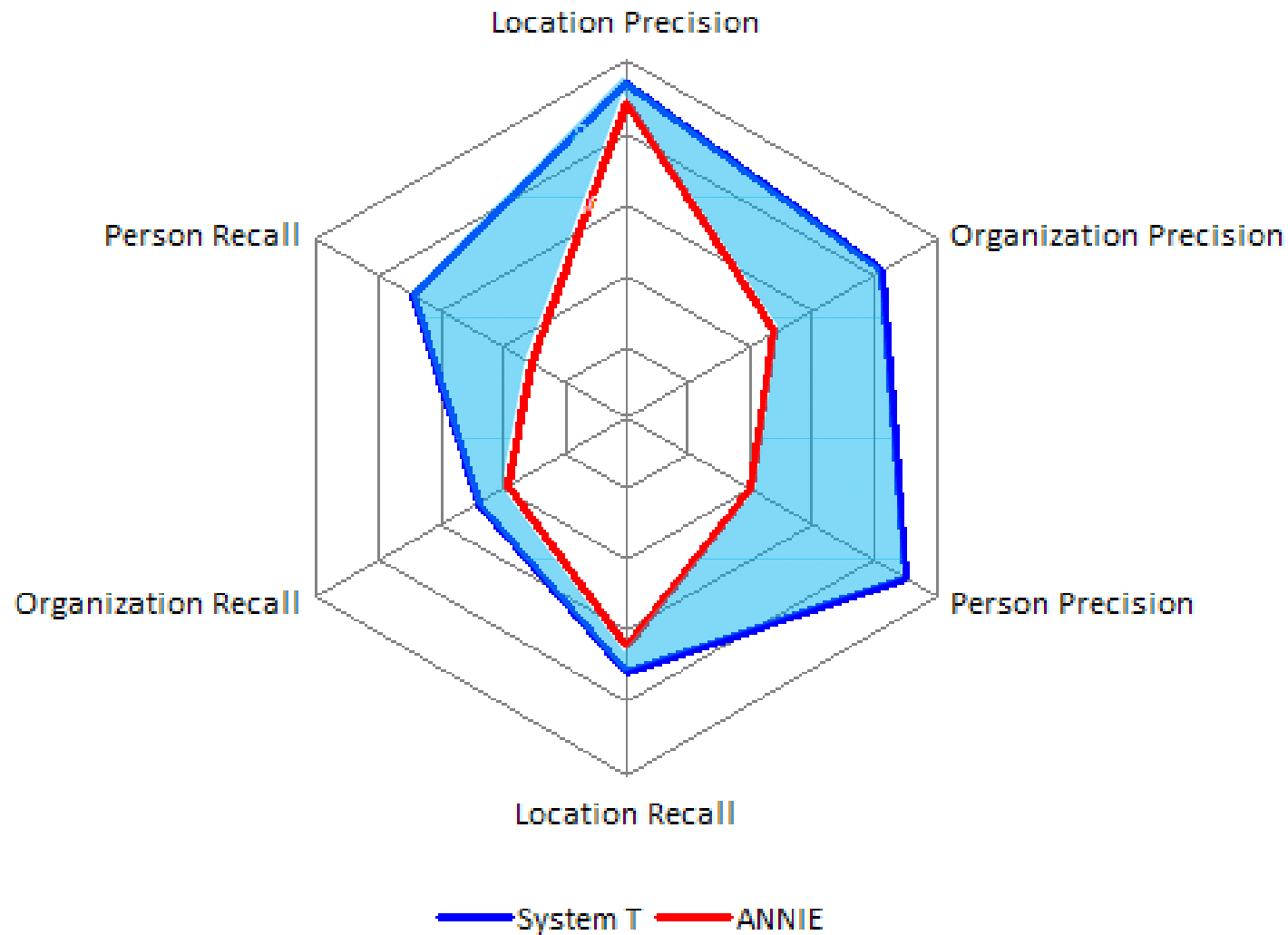


Ties resource to task at hand
(CPU optimized, size of document optimized, ++)



- High-throughput
- Small memory footprint
- Optimizes for tasks:
 - Example, Text Analytics needs CPU optimization

IBM Finds RIGHT Answers Better Than Anyone



Example Application: Lead Generation

Lead Generation Real Time Dashboard

Actions: [Resume](#) [Microsegmentation Information](#)

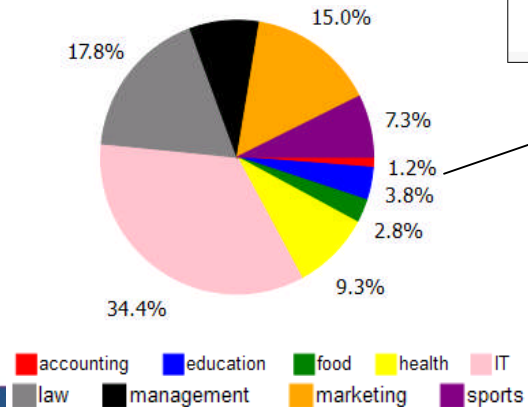
Name	Sources	Intent	Gender	Location	Interests	Tweet
	Twitter, Tumblr	Laptop	Female			I need a new laptop.
	Twitter, Facebook	Phone			Foodies	I need a phone.
Abdul Yab	Twitter, Tumblr	Car	Male	Cairo		Need a car.
kesha rose	Twitter	Trousers	Female	Bristol		omg i cant find any black trousers. i need some for my trial at john lewis tomorrow.
vanessa natal	Twitter	Conditioner	Female			I need to find a good conditioner...my hair is growing so thin...:(
Chris Castle	Twitter	Juicer		Seattle	Sports	@bdrimberger whoa! I've been thinking about getting a juicer. I might have to try yours out!
Matt Arlauckas	Twitter	Diner	Male	Rochester	Foodies	@shotbykim Holy Moley! I need a diner. STAT!
Jodi	Twitter	New Car	Female			Thinking about test driving cars this weekend. When's the best time of year to buy a new car? #testdrivepost
chris	Twitter	Shoes				@TGODcaponi nah just for chillin. i need a new pair of blue shoes.
Courtney Morrison	Twitter	Ice Cream				I really want some ice cream. I shoulda got that instead of these cookies.

Real-time product intents enriched with consumer attributes

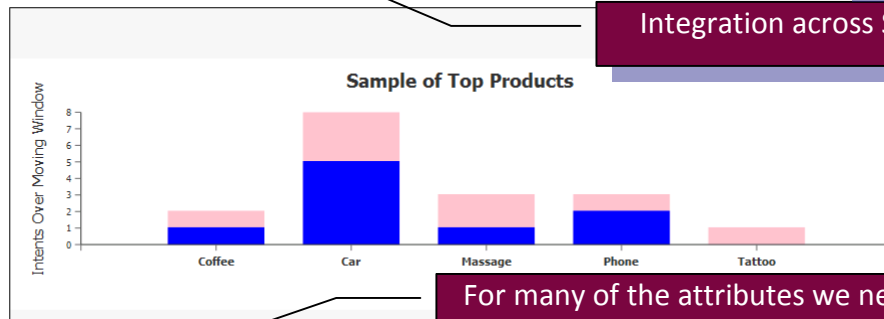
Entries contain promotional messages, wishful thinking, questions, etc

Micro-segmentation of product intents by occupation

Phone Intents by Occupation



Integration across Social Media sites



Real-time tracking by micro-segmentation

For many of the attributes we need to extract, cleanse, normalize and categorize




Micro-segmentation of consumers by hobbies

How to Analyze Data At Rest

IBM Software

Information Management & Analytics Forum 2013

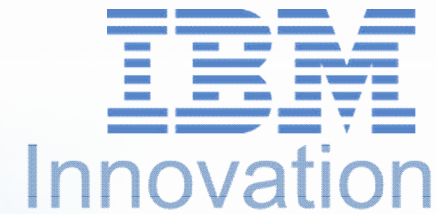
Return on Information: The New ROI

The footer features a dark blue background on the left with white text. To the right, there are three overlapping semi-circular shapes: a purple one, a green one, and a light blue one.

IBM Enriches Hadoop

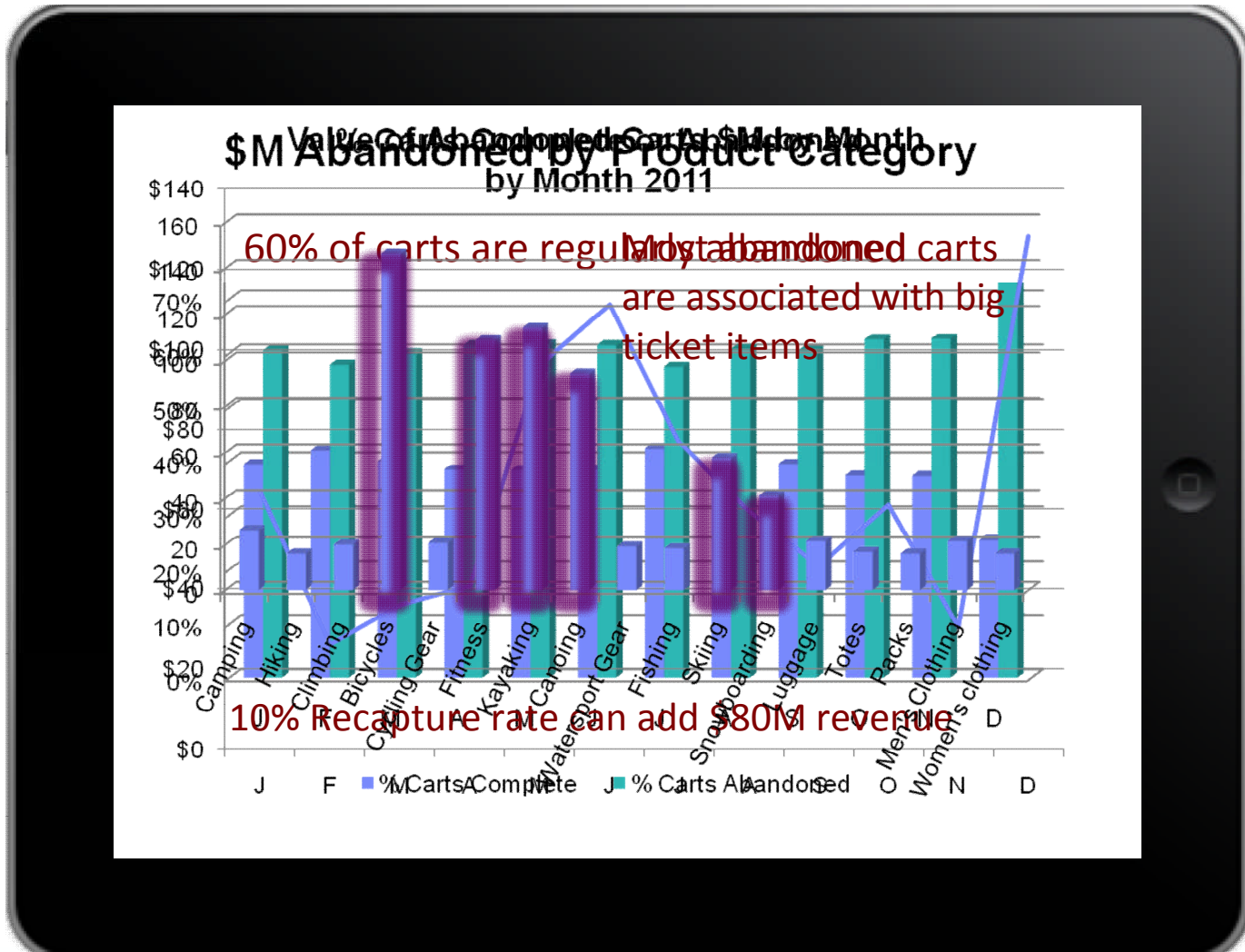


- **Scalable**
 - New nodes can be added on the fly
- **Affordable**
 - Massively parallel computing on commodity servers
- **Flexible**
 - Hadoop is schema-less, and can absorb any type of data
- **Fault Tolerant**
 - Through MapReduce software framework



- **Performance & reliability**
 - Adaptive MapReduce, Compression, Indexing, Flexible Scheduler, +++
- **Enterprise Hardening of Hadoop**
- **Productivity Accelerators**
 - Web-based Uis and tools
 - End-user visualization
 - Analytic Accelerators
 - +++
- **Enterprise Integration**
 - To extend & enrich your information supply chain

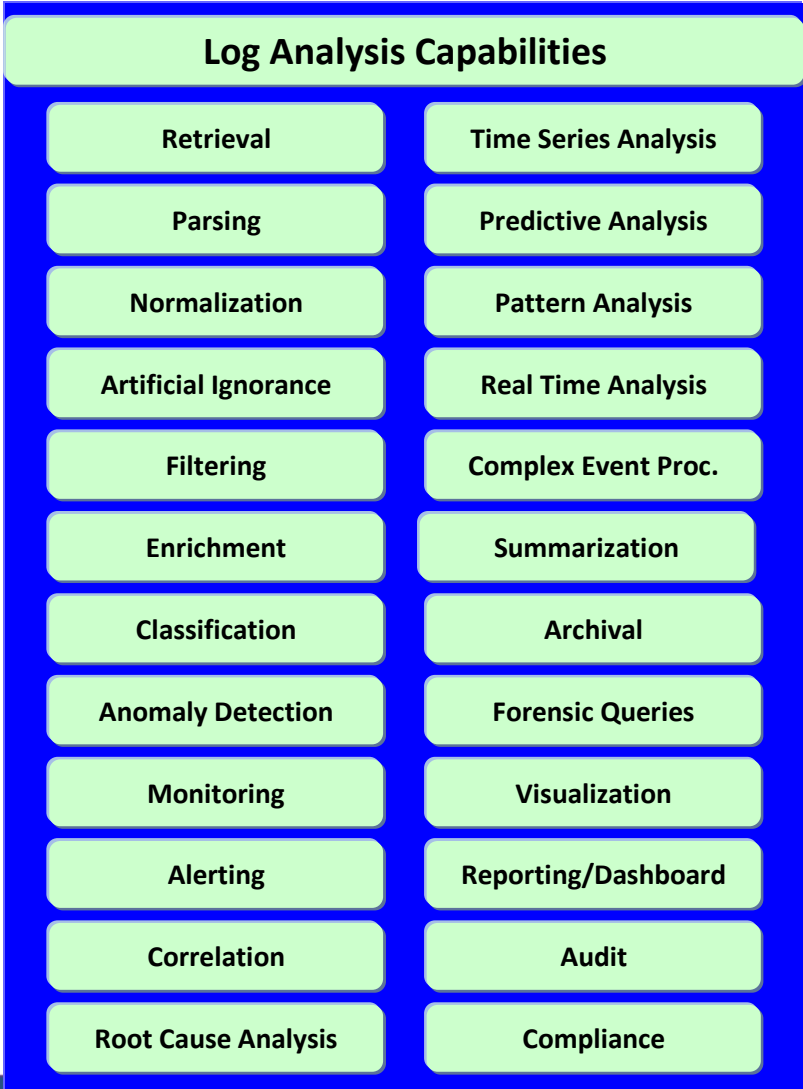
Imagine If You Could Generate Reports Like This...



The What is Log Analysis?



- App. Server Logs
- System Logs
- Database Logs
- Application Logs
- Messaging Logs
- Authentication Logs
- Authorization Logs
- Access Control Logs
- Intrusion Prevention
- Router Logs
- LAN Logs
- HDFS Logs
- Hardware Logs



- ### Outcomes
- Gain visibility into customer behavior and experience
 - Increase application availability
 - Proactive planning to increase operational efficiency
 - Troubleshoot issues and investigate security incidents
 - Monitor end-to-end infrastructure to avoid service outages or degradation

Click Stream Analysis and Next Click Optimization

- Easy to get pathing and attribution of the clickstream
 - Lightweight, low trauma solutions that can address immediate pain points with a strategic context that will move the organization towards an even greater ability to leverage Big Data
- Used for sessionization, path to purchase optimization, understanding of drop-out point, campaign take rates, pathing information
 - Input: Web log data (IP address, timestamp, URL, HTTP return codes)
- Goal: Determine number & position of abandoned shopping carts, campaign take rates, where abandoned, and dropped take rates: Where is “THE LAST MILE?”

Machine Data Accelerator

What is BigSheets?

Browser-based Big Data analytics tool for business users

Big Data Challenges...

- Business users need a no programming approach for analyzing Big Data
- Extremely difficult to find actionable business insights in data from multiple sources with different formats
- Translating untapped data into actionable business insights is a common requirement that requires visualization

How can BigSheets help?

- Spreadsheet-like discovery interface lets business users easily analyze Big Data with **ZERO PROGRAMMING**
- **BUILT-IN “readers”** can work with data in several common formats
 - JSON, CSV, DEL, TSV, Web, Hive, Line Reader, Nutch Web Crawler Output, +++
- Users can **VISUALLY** combine and explore various types of data to identify “hidden” insights

Path: /user/biadmin/sampleData/IBMWatson/news-data.txt

Name	Size	Block Size
news-data.txt	25.1 MB	1280 MB

Viewing Size: 10KB - Text Sheets

hdfs://localhost.localdomain:9000/user/biadmin/sampleData/IBMWatson/news-c

Line Reader

Ready

Select a reader:

- JSON Array
- Basic Crawler Data
- Character Delimited Data
- Comma Separated Value (CSV) Data
- Hive Reader
- JSON Array
- Line Reader
- Sheets Data
- Tab Separated Value (TSV) Data



ANALYST **CLERK**

DESIGNER

MANAGER **FIELDREP**

SALESREP

DESIGNER (10 occurrences)

OPERATOR

5	60	Franco	Bruno					
6	70	Hedi	Simane	D21				
7	90	Coleen	Rieder	E11	5456			
8	100	Ramesh	Khanna	E21	972			
9	110	Andrew	King	A00	3490	19880516		
10	120	Robert	O'Wager	A00	2167	19931205		
11	130	Heidi	Slimane	C01	4578	20010728	ANALYST	
12	140	Peggy	Bonifacino	C01	1793	20061215	ANALYST	18
13	150	Jay	Longley	D11	4510	20020212	DESIGNER	16
14	160	Jun	Ashida	D11	3782	20061011	DESIGNER	17
				D11	2890	19990915	DESIGNER	16
				D11	1682	20030707	DESIGNER	17
				D11	2986	20040726	DESIGNER	16
				D11	4501	20020303	DESIGNER	16
				D11	942	19980411	DESIGNER	17
				D11	672	19980829	DESIGNER	18
				D21	2094	19961121	CLERK	14
				D21	3780	20041205	CLERK	17

Select a type of sheet: **Built-in Sampling**

Filter	Macro	Load	Pivot	Combine
Union	Limit	Distinct	Copy	Formula


Add Sheet using by entering a formula

How to Analyze Data In-Motion

IBM Software

Information Management & Analytics Forum 2013

Return on Information: The New ROI



Data In Motion

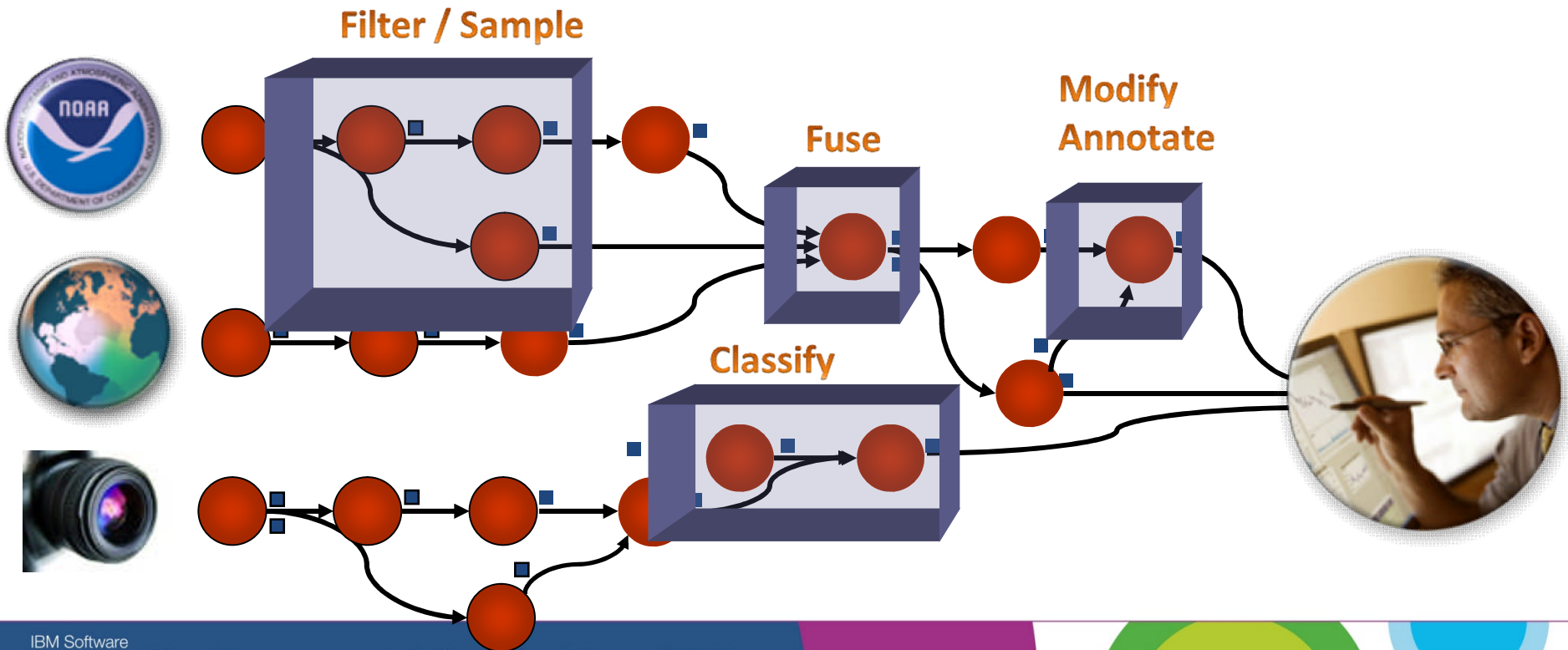
Algorithmic
Trading *Forum*

Analyzes and correlates 5M+ market messages/sec to execute algorithmic option trades with average latency of 30 micro-secs.

ipdr

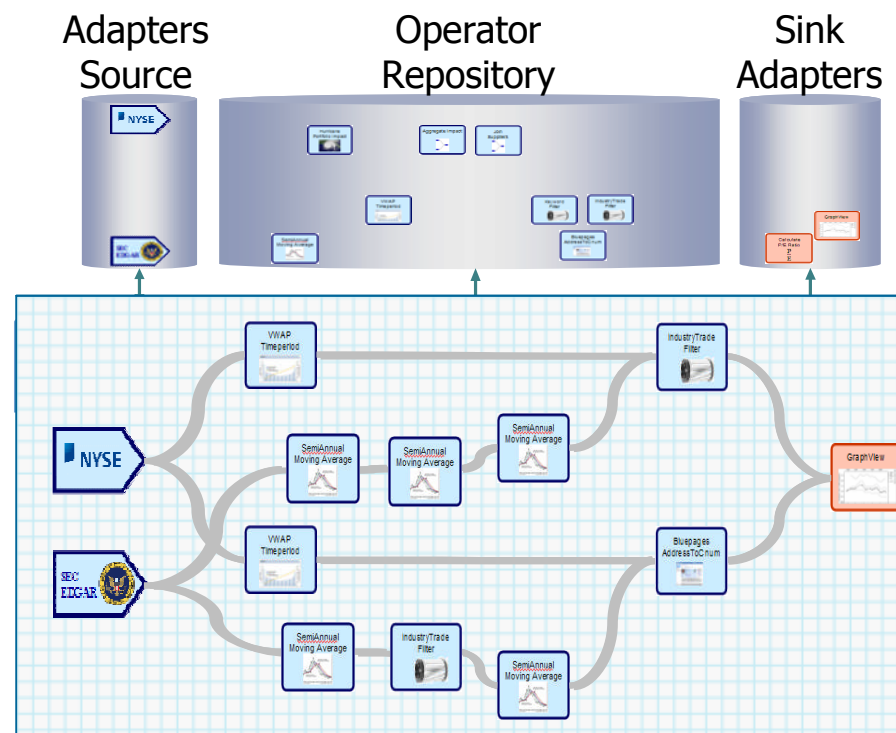
500K/sec, 6B+ IPDRs analyzed per day on more than 4 PBs/yr. sustaining 1GBps.

Consider: Data that is never stored, never has to be subjected to retention policies: COST SAVINGS



The Consumability of InfoSphere Streams

- Streaming analytic applications
 - Multiple input streams support advanced streaming analytics
 - Optimizing compiler automates deployment and connections
 - Extremely low latency
- Eclipse based IDE
 - Define sources, apply operators, define intermediary and final output sinks
 - User defined operators in Java or C++
 - Toolkits get you started faster
- Over 100 sample applications and toolkits with industry focused toolkits with 300+ functions and operators!



Declarative Language

```

composite NumberedCat {
  graph
  () as Sink = FileSink(Numbered) {param file: "result.txt";}
  stream<rstring contents> Lines = TCPSource() {
    param role: server;
    port: getSubmissionTimeValue("port");
  }
  stream<rstring contents> Numbered = Functor(Lines) {
    logic state: mutable int32 i = 0;
    onTuple Lines: i++;
    output Numbered: contents = (rstring) i + " " + contents;
  } }
    
```



UPPSALA
UNIVERSITET

"Streams allows me to again be a Space Physicist instead of a trying to be a Computer Scientist."

– Bó Thide, Professor

Social Media Analytics Architecture

Online flow: Data-in-motion analysis

Real time analytics.
Pre-defined views,
charts, workbooks,
dashboards, +++



Social Data Accelerator

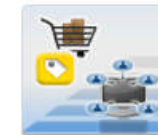


Pre-defined
Workbooks and
Dashboards

Social Data Accelerator



Offline flow: Data-at-rest analysis



Brand Management
Retail Configure -



Brand Management
Retail



Brand Management
Retail Local
Analysis



Brand Management
Retail Real-time
Analysis

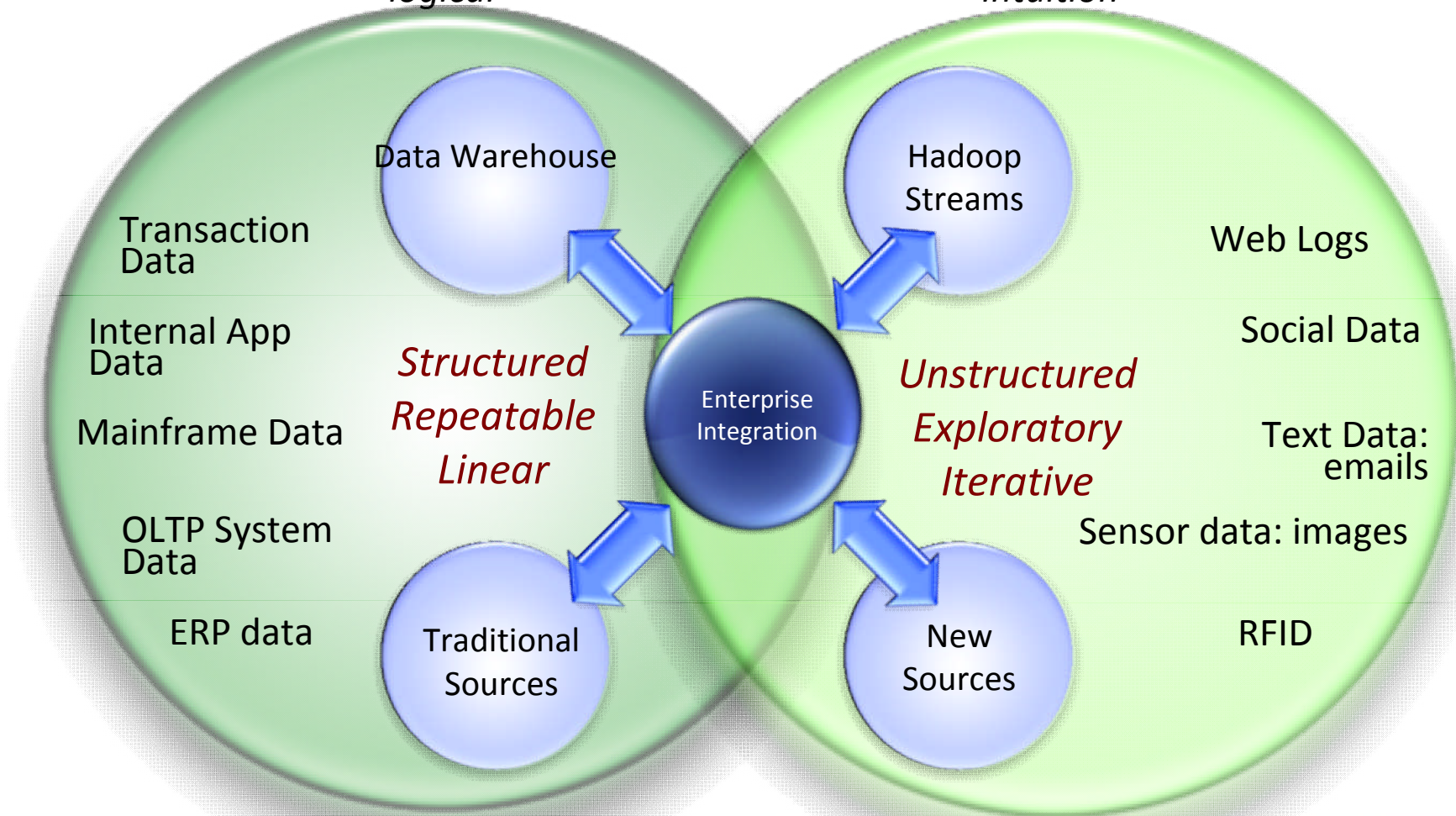
Social Data Accelerator

Optional: Indexed Search

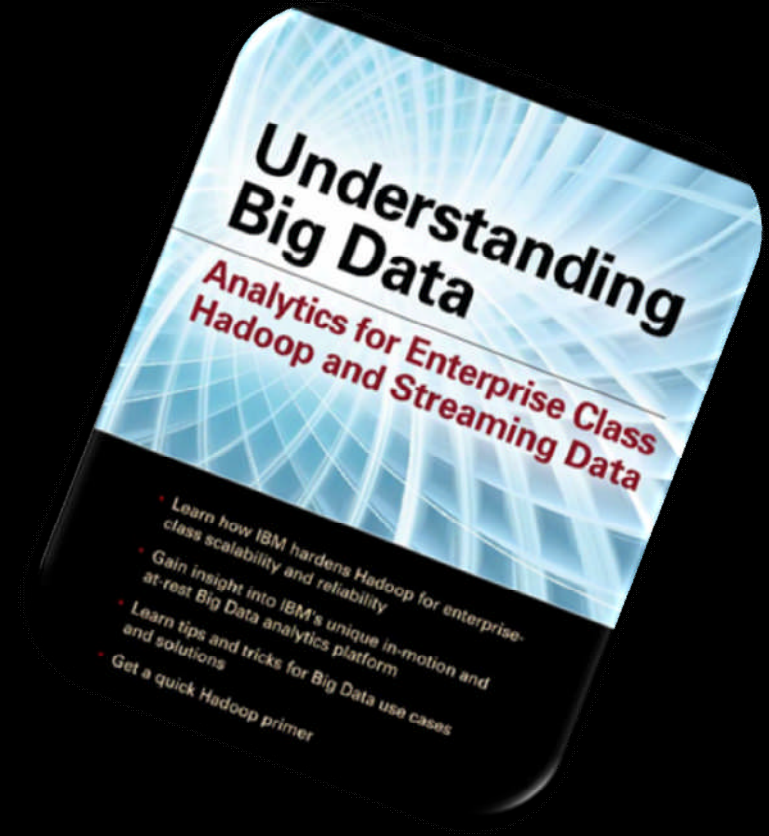
Complementary Analytics

Traditional Approach
Structured, analytical,
logical

New Approach
Creative, holistic thought,
intuition



THINK



BIG