

Big Data



IBM BusinessConnect
Vernetzter, intelligenter und informierter denn je



Big Data

Vom Hype – zur Realität

Wolfgang Nimführ

Big Data Business Development Executive

IBM Software Group



Something Profound is Happening

Continuous transformation is the new normal

Business models are under constant threat

- Shift of power to the consumer and the citizen
- Accelerating pressure to do more with less
- Great relationships trump great products

Convergence of technology disruptors creates opportunity



Not changing is the same as losing



The Big Data Conundrum

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

Just collecting and storing “Big Data” doesn’t drive a cent of value to an organization’s bottom line



Does this sound familiar?

Today we treat Aki like any other customer in her segment...

...but Aki is an individual



By using only our limited segmentation, we treat Aki like anyone else



Aki holds a mortgage and a savings account
U



Action	Impact on Retention	Impact on Customer Lifetime Value	Likelihood to respond positively to action
Cash Management Acct.			
Set meeting with Private Banking & Wealth Mgt. Advisor for a Portfolio Review			
Equity Bank Line / Secured Line-of-Credit			
Preferred Gold Credit Card			



Information helps us understand how Aki is different, but do we use it?



Aki holds a mortgage and a savings account with us

Last week Aki asked the Call Center about loan processing times

Aki has also posted property photos to Facebook asking friends to vote

This week, she checked mortgage rates on the Web Site three times

Aki's current credit score & profitability qualifies her for a preferred rate

And today she's tweeted a link to an article about buying a second home

By using all the information we can make our service unique to Aki



Aki holds a mortgage and a savings account

Action

Impact on Retention

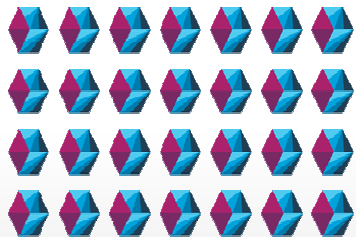
Impact on Customer Lifetime Value

Likelihood to respond positively to action



Action	Impact on Retention	Impact on Customer Lifetime Value	Likelihood to respond positively to action
Cash Management Acct.			
Preferred Gold Credit Card			
Equity Bank Line / Secured Line-of-Credit			
Mortgage special rate discount 25 basis points			

Volume



Data at Scale

Terabytes to petabytes of data

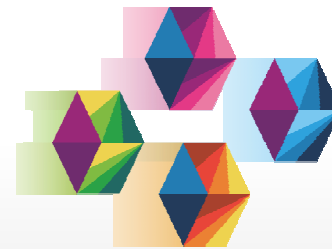
Variety



Data in Many Forms

Structured, unstructured, text, multimedia

Velocity



Data in Motion

Analysis of streaming data to enable decisions within fractions of a second.

Veracity



Data Uncertainty

Managing the reliability and predictability of inherently imprecise data types.

Velocity IS the game changer: It's NOT just how fast data is produced or changed, BUT the speed at which it must be received, understood, and processed.

Paradigma Shift

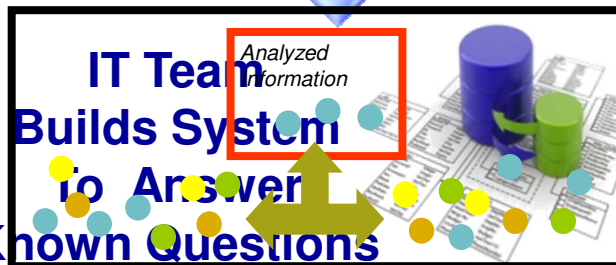


Traditional Analytics

Structured & Repeatable
Structure built to store data



Capacity constrained down sampling of available information



Carefully cleanse a small information before any analysis

Big Data Analytics

Iterative & Exploratory
Data is the structure



Whole population analytics connects the dots

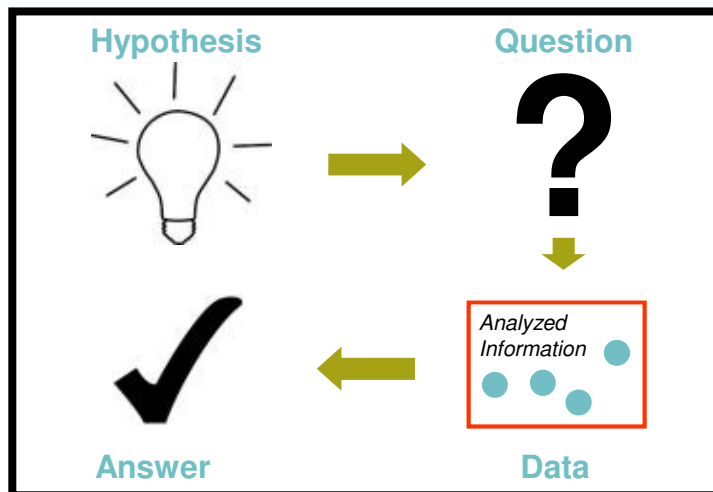


Analyze information as is & cleanse as needed & existing repeatable

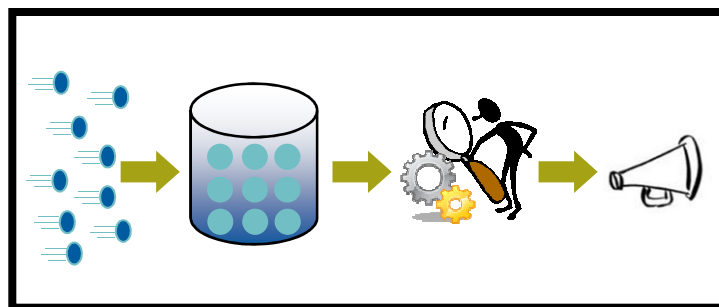
Paradigma Shift

Traditional Analytics

Structured & Repeatable
Structure built to store data



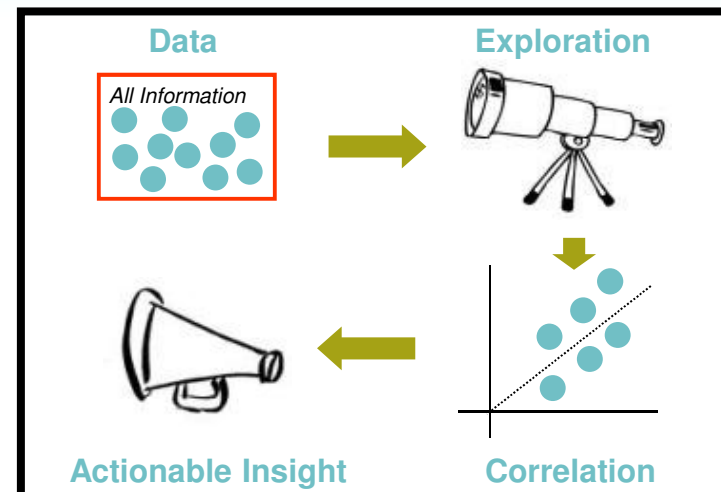
Start with hypothesis
Test against selected data



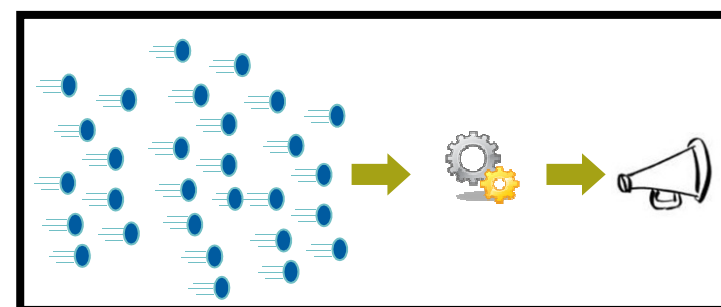
Analyze after landing...

Big Data Analytics

Iterative & Exploratory
Data is the structure



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



Analyze in motion...

Analytics Breadth to Enable Decisions

Big Data & Analytics adds Value at the Point of Impact



Business Value via Key Business-Driven Use Cases



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



Enhanced 360° View of the Customer



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
 Title: *Regional Sales Manager*

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 | Filter Feed

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 goes!" Salesforce - 2 days ago

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

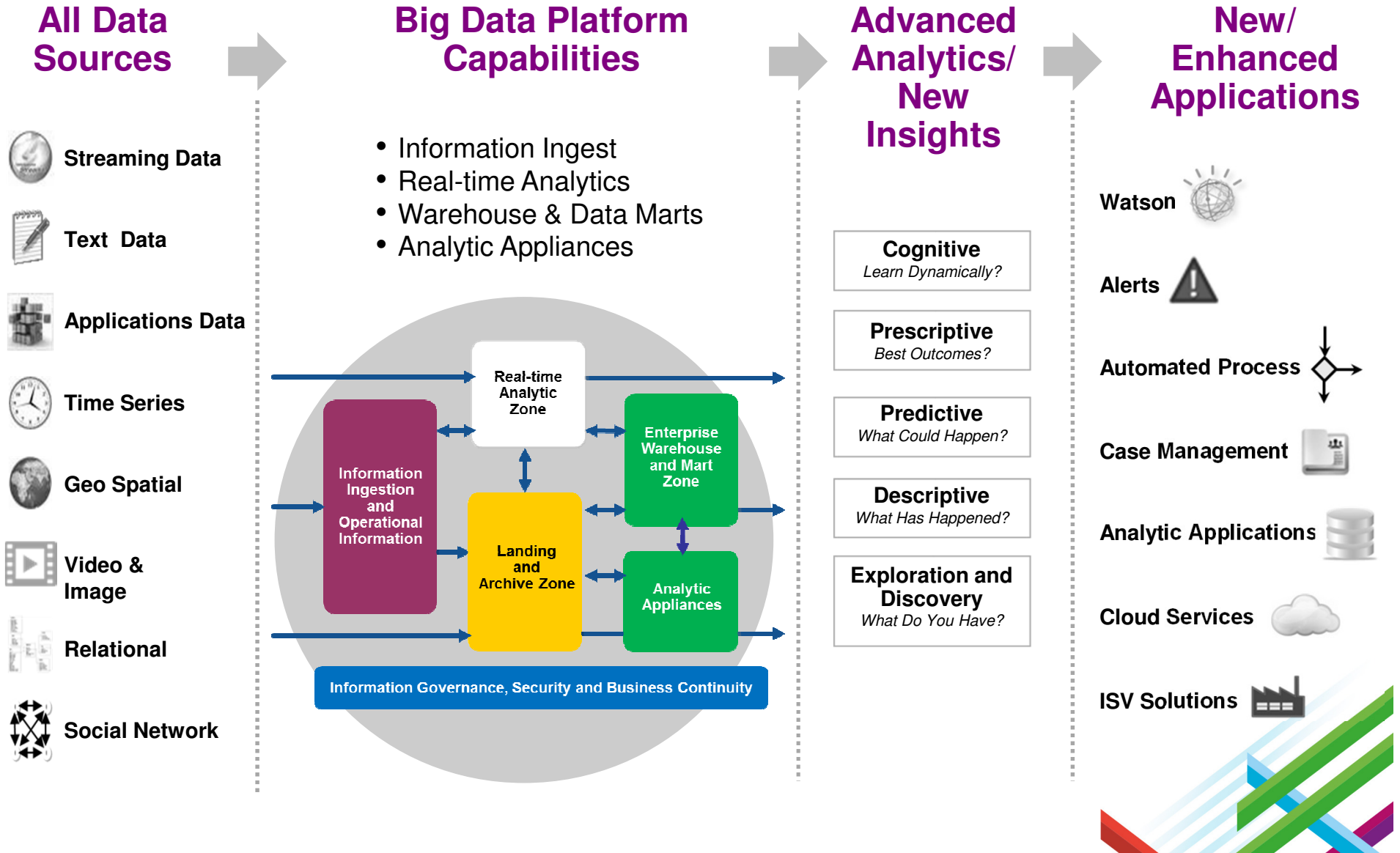
Notes: Janet is new to Wakefield but has been a wholesaler for 10 years New to Optimum Funds.
Salesforce - 2 days ago

Title: Optimum Introduction
 Welcome Janet to Wakefield. I wanted to introduce myself...
Salesforce - 2 days ago

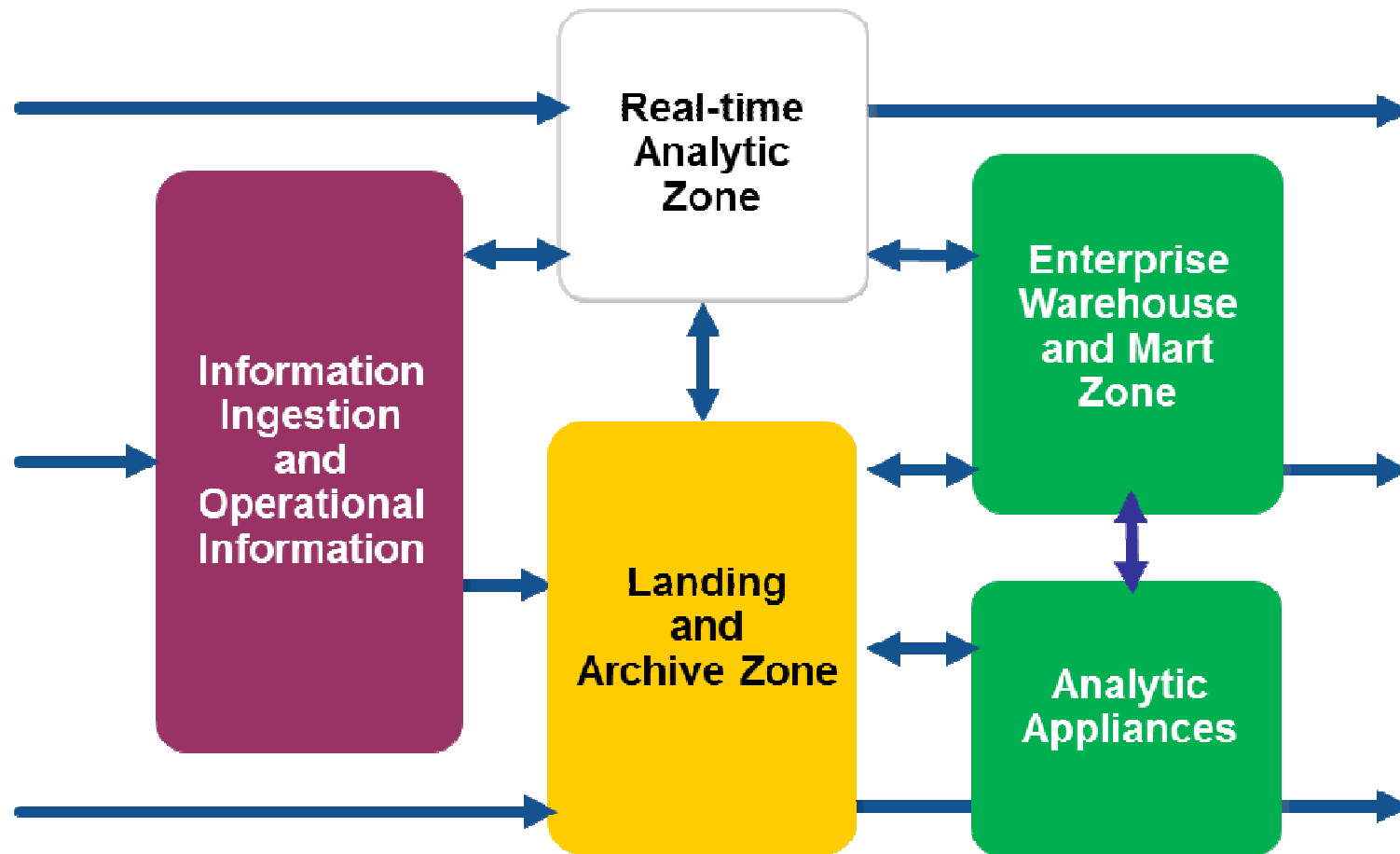
Unified View of Party's Information



Big Data & Analytics Architecture



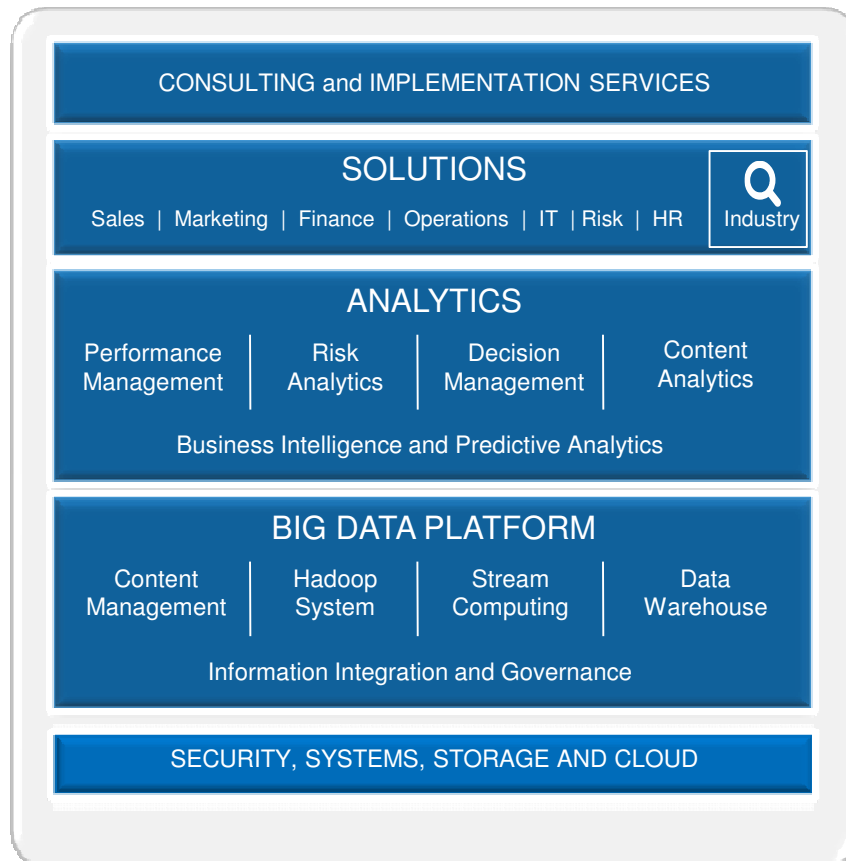
Big Data & Analytics Architecture



Information Governance, Security and Business Continuity



IBM Provides a Holistic and Integrated Approach to Big Data & Analytics



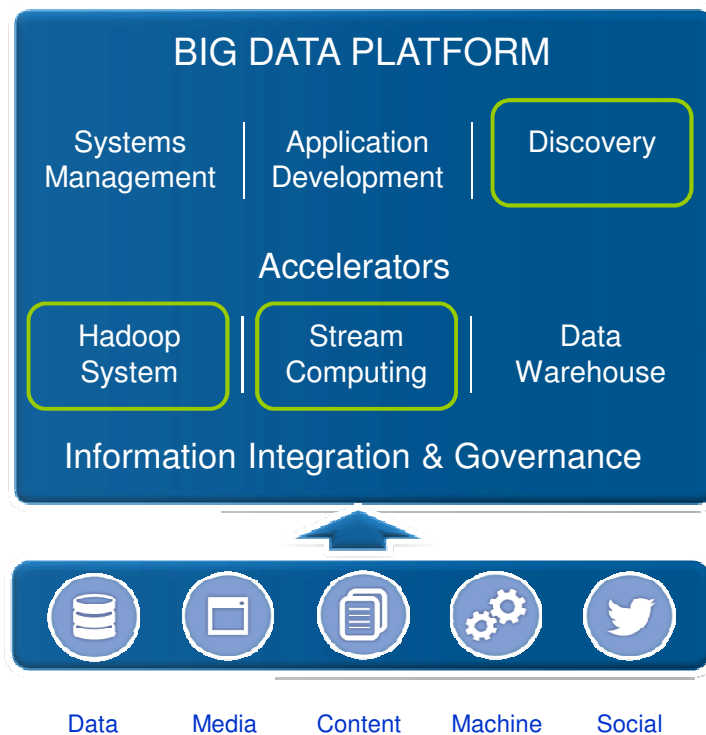
Enabling organizations to

- Assemble and combine relevant mix of information
- Discover and explore with smart visualizations
- Analyze, predict and automate for more accurate answers
- Take action and automate processes
- Optimize analytical performance and IT costs
- Reduced infrastructure complexity and cost
- Manage, govern and secure information





The Platform for New Insight and Applications



InfoSphere Data Explorer

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

InfoSphere BigInsights

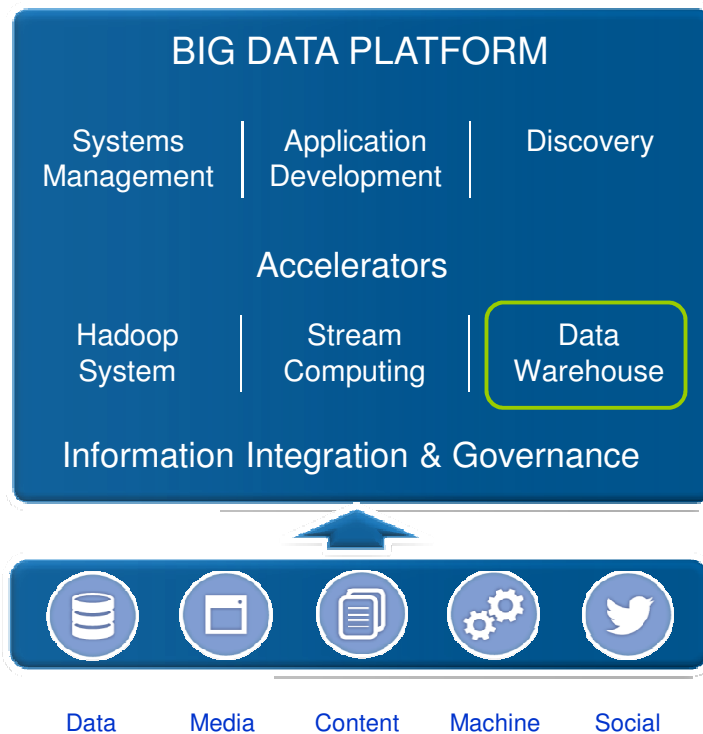
Cost-effectively analyze Petabytes of unstructured and structured data

InfoSphere Streams

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



Appliance Simplicity for Big Data



PureData System for Hadoop

Optimized system to accelerate big data analytics and online archive with appliance simplicity

PureData System for Analytics

*powered by Netezza technology
Optimized system delivering data services for analytics & reporting*

PureData System for Operational Analytics

Optimized system delivering data services for operational analytics

DB2 Analytics Accelerator

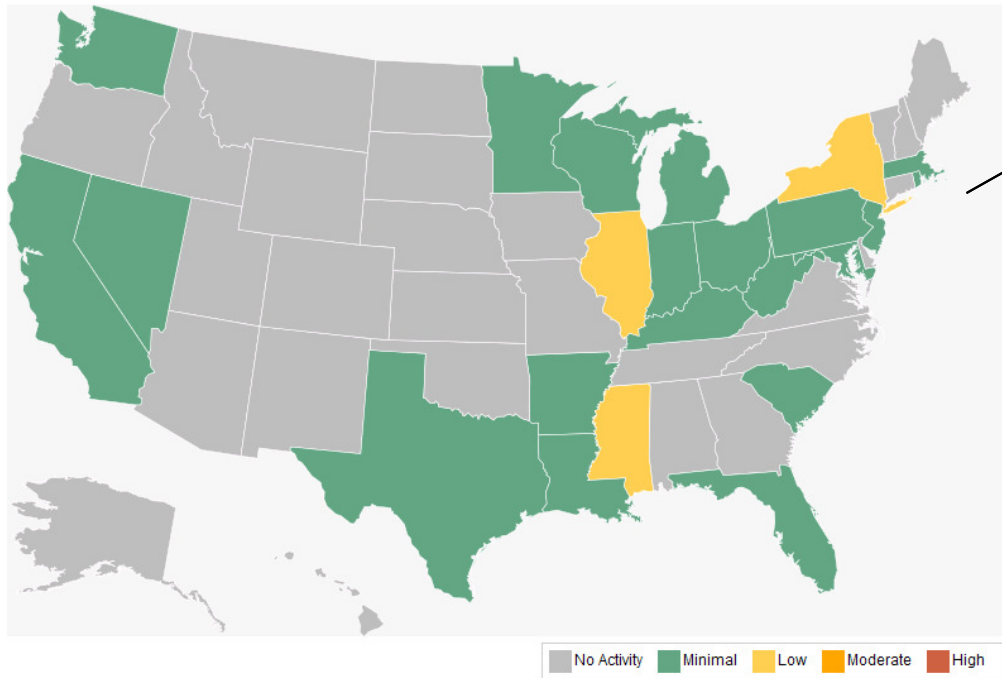
*(powered by Netezza technology)
Optimized appliance add-on enabling integration of business insights into operational processes*

PureData System for Transactions

Optimized system delivering data services for transactions

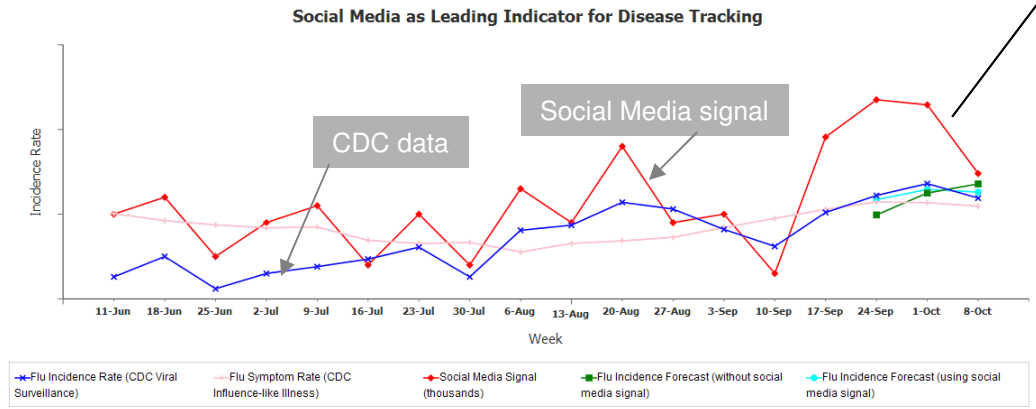


Example Application: Disease Tracking



Tracking of disease symptoms by consumer micro-segmentation

Clear correlation of flu onset trends from social media analysis a week ahead of when data is published by CDC



Example Application: Lead Generation to forecast Sales



Lead Generation Real Time Dashboard

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

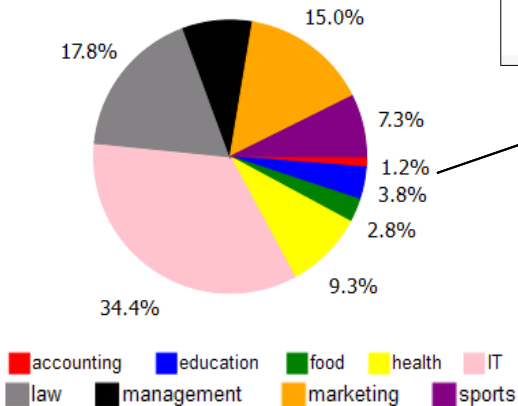
Name	Sources	Intent	Gender	Location	Interests	Tweet
	Twitter, Tumblr	Laptop	Female			I need a new laptop.
	Twitter, Facebook	Phone			Foodies	I need a phone.
Abdul Yab	Twitter, Tumblr	Car	Male	Cairo		Need a car.
kesha rose	Twitter	Trousers	Female	Bristol		omg i cant find any black trousers. i need some 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	@bdtrimberger 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
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

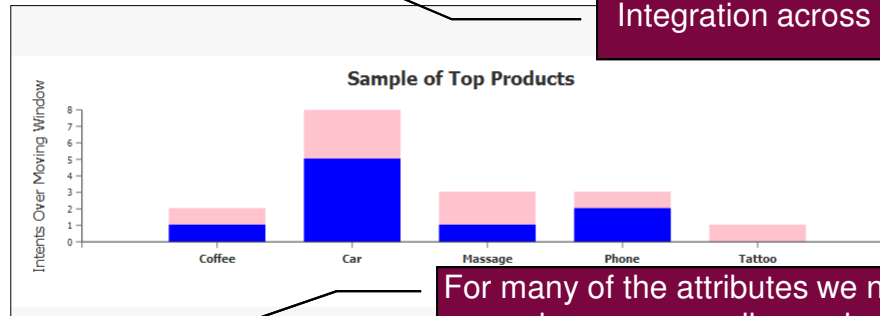
Entries contain promotional messages, wishful thinking, questions, etc

Micro-segmentation of product intents by occupation

Phone Intents by Occupation

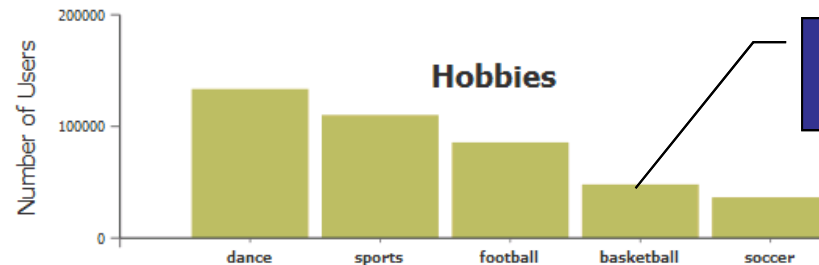


Integration across Social Media sites



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

Example Application: Institutional Risk Application

Annual Report

Loan Agreement

Insider Transaction

Proxy Statement

Counterparty Relationships

Legend:

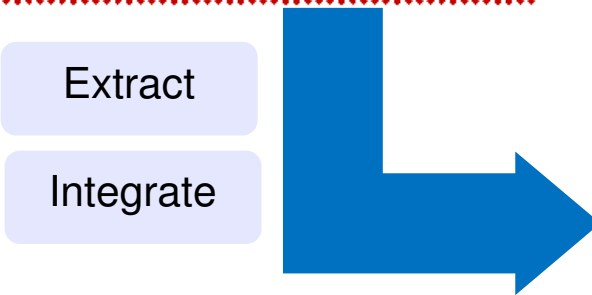
- Banking Institutions (4)
- Mutual Funds (17)
- 5% Beneficial Ownership (174)
- Outstanding Board Members / Officers (8)
- Institutional Holdings (120)

Loan Exposure

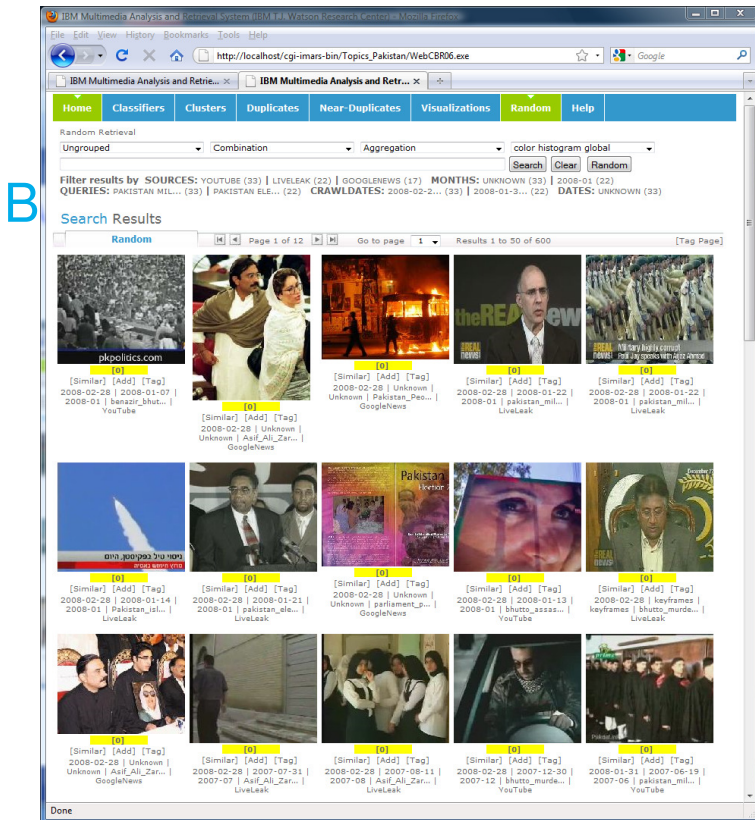
Loan Commitments in dollars

Number of joint loans

Loan Title	Co-Lender Information	Total Amount of Loan (\$)	Agreement Date
Credit Agreement (Primary borrower)	Co-Lender Company	200,000,000	2009-09-12
	Citigroup, N.A.		
	Bank of America, N.A.		
	Wells Fargo Bank, N.A.		
	Citibank, N.Y.		
	Bank of Montreal		
	JP Morgan Chase Bank		
	Bank of New York Mellon		
	State Street Bank and Trust		
	The Bank of New York		

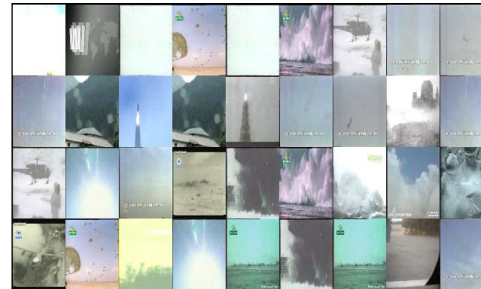


Example Application: Automatic Recognition of Scenes of Interest



B

Automatic scene recognition, clustering, duplicate and near-duplicate detection from image and video content



Air Attack



Demonstration



Wreckage



Explosion



Individual



Military



Large global property casualty insurance company improved profitability through reduced losses and more accurate risk-based pricing

Need

- Make more informed business decisions to respond faster to changes in the book of business and influence underwriting decisions much sooner and when to cap exposures

Benefits

- Quickly understand exposures and probable maximum losses faster and take action sooner
- Can now model catastrophe risk every month instead of 3 times per year
- Once data is refreshed, can create “what if” scenarios in hours, rather than weeks




Global aerospace manufacturer empowers staff with access to critical information

Need

- Improve operational efficiencies by providing a unified search, discovery and navigation capability to provide fast access to relevant information across the enterprise

Benefits

- Placed 50 additional aircraft into service worldwide during the first year without a staffing increase
- Saved USD36 million/year in supporting the 24/7 aircraft-on-ground program
- Provided supply chain visibility to reduce cycle time, saving millions of dollars on critical parts deliveries



Growing financial services firm supports merger with secure, single point of access to enterprise information

Need

- Integrate business applications and data repositories securely, standardize information governance and leverage expertise to facilitate collaboration across the organization

Benefits

- Deployed IBM® InfoSphere® Data Explorer in 12 weeks and achieved data transparency across 21 subsidiaries worldwide
- Managed all merger related information securely to support financial and technical audits and compliance requirements and to minimize risk
- Helped identify and evaluate the business value of data in acquired systems to determine the cost/savings of retaining or eliminating systems



Leading healthcare insurance provider call center enables 14,000 agents with single view of customer and product data

Need

- Inefficient access to huge volumes of siloed customer and product data reduced agent productivity and increased average call handle time. Agents needed faster access to information

Benefits

- Improved productivity for 14,000 agents, saving an average of 3 seconds on call handle time, and millions of dollars annually
- Helped ensure 99.999 percent uptime at every location, delivering a commanding query-per-second speed
- Improved application performance to support daily operations and business users at 180 sites



University of Ontario Institute of Technology (UOIT) Detects Neonatal Patient Symptoms Sooner

Need

- Performing real-time analytics using physiological data from neonatal babies
- Continuously correlates data from medical monitors to detect subtle changes and alert hospital staff sooner
- Early warning gives caregivers the ability to proactively deal with complications

Benefit

- Helps detect life threatening conditions up to 24 hours sooner
- Lower morbidity and improved patient care





Vestas optimizes capital investments based on 2.5 Petabytes of information

Need

- Model the weather to optimize placement of turbines, maximizing power generation and longevity

Benefits

- Reduce time required to identify placement of turbine from weeks to hours
- Reduces IT footprint and costs, and decreases energy consumption by 40 % -- while increasing computational power
- Incorporate 2.5 PB of structured and semi-structured information flows. Data volume expected to grow to 6 PB

Vestas

GTO: Contextual Enterprise

Access to a broad set of data and affordability of compute

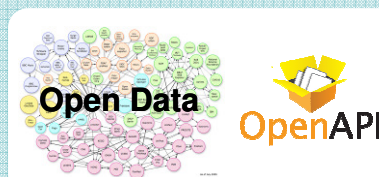
Smart phones and web browsers generate vast amounts of data about users and customers.



People freely share their sentiments, expertise, desires, and intentions via social media.



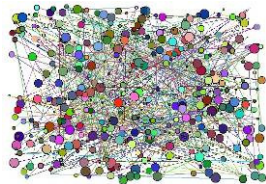
Open Data and Open APIs start delivering breakthroughs in access and integration.



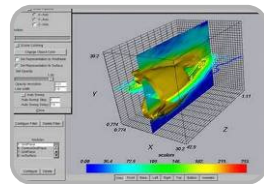
Semantic technologies like RDF enable automatic data discovery and ingestion.



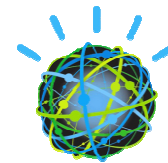
New databases emerge to store and manage web-scale graphs of contextual information.



Advanced analytics exploit rich context for dramatic new insights and predictions.



WATSON-style guidance is generated from massive stores of contextualized data.



Adaptive interactions and visualizations deliver value and insight to Markets of One.



Achieving the Contextual Enterprise will require innovation in specific areas

Gather: Collect All Relevant Data

Challenges:

- Maintain petabytes of historical data
- Terabytes of streaming updates
- Data-centric security, privacy, provenance
- Dynamic data sourcing
- Automatic discovery and ingest

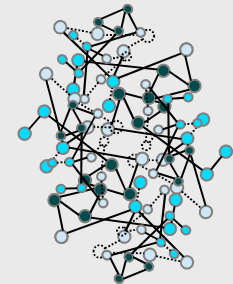


Value: Provide dynamic and diverse set of data sources for high value context discovery

Connect: Extract Features, Build Context

Challenges:

- Real-time extraction analytics and contextualization
- 10:1 context:data expansion
- Highly dynamic schema
- Uncertainty management
- Lifecycle management of context

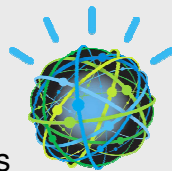


Value: Deliver enterprise-wide context awareness across all applications

Reason: Discover Hidden Implications

Challenges:

- Longitudinal analytics
- High dimensional data
- Continual analysis of ever changing context
- Adaptive analytics for dynamic input schemas



Value: Breakthrough insights, predictions, and optimal action selection

Adapt: Compose Appropriate Interactions

Challenges:

- Continual user context profiling
- Dynamic application evolution
- End-to-end latency management
- Adaptive multimodal input/output
- Context-based recommendations



Value: Optimal customer experiences and employee effectiveness



ibmbigdatahub.com - bigdatauniversity.com - youtube.com/user/ibmbigdata

The Big Data Hub



Understanding big data for the enterprise

Home

Blogs

Videos & Podcasts

Resources

Events

IBM Solutions

For Developers



Complete Book—Now Available!

Boost your big data IQ, learn about the most common use cases, discover how to deploy projects faster and with less risk—all from the authors of "Understanding Big Data." **Download now →**

Save the date

Information On Demand **2013**

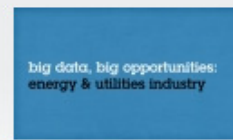
November 3 – 7
Mandalay Bay | Las Vegas, Nevada

Infographics & Animations

→ View All



Tuning Into Big Data as the Buzz Gets Louder



Big Data, Big Opportunities: Energy & Utilities



Big Data: The New Natural Resource

Featured Whitepapers & Reports

→ View All



Getting Big Value from Big Data: A Ventana Analyst Report



Big Data: New Insights Transform Industries



Analytics: The real-world use of big data

Next Events ...

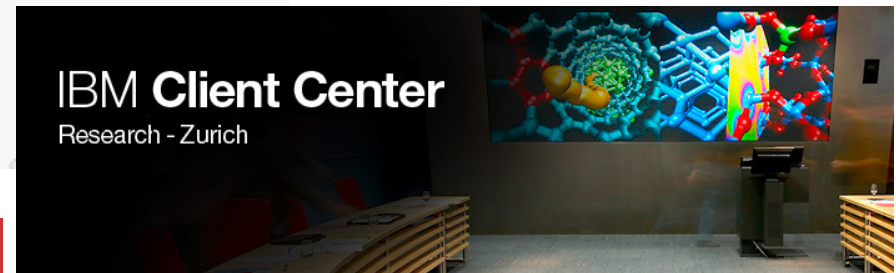


Information On Demand 2013

November 3 – 7

Mandalay Bay | Las Vegas, Nevada

Big data. Unique perspectives.
Register now. ▶



Big Data Developers in Zurich

s Photos Pages Discussions More

Hand's On BigData/Hadoop Analytics with IBM BigInsights

Edit Cancel Feature Copy Ticket Export Tell a friend Share

Thursday, November 21, 2013
9:00 AM to 5:00 PM

IBM Innovation Center Zurich, 8010 Zurich
Vulkanstrasse 106, Zürich (edit map)
IBM Innovation Center Zurich, Vulkanstrasse 106, 8010 Zurich, Switzerland



IBM BusinessConnect
Vernetzter, intelligenter und informierter denn je



Big Data

Von der Realität zur Umsetzung

Carlo Marchesi

PureData Champion CH/AT

IBM Software Group



Agenda

Big Data Analytics – the Benefits & Challenges
PureSystems
PureData System for Hadoop Overview
How to Leverage PureData System for Hadoop
Augmenting the Data Warehouse with Hadoop





Big Data Analytics

The Benefits and Challenges

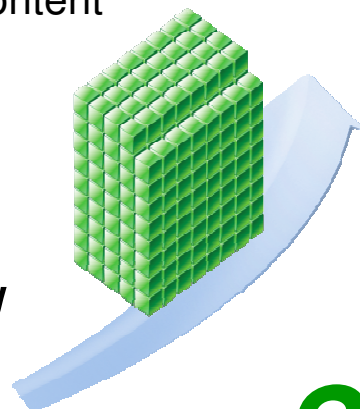


Information is at the Center of a New Wave of Opportunities...

44x

as much data and content over coming decade

2020
35 zettabytes



2009
800,000 petabytes

Volume
Variety
Velocity

80%

Of world's data is unstructured



... and Organizations Need Deeper Insights

1 in 3 Business leaders frequently make decisions based on information they don't trust, or don't have

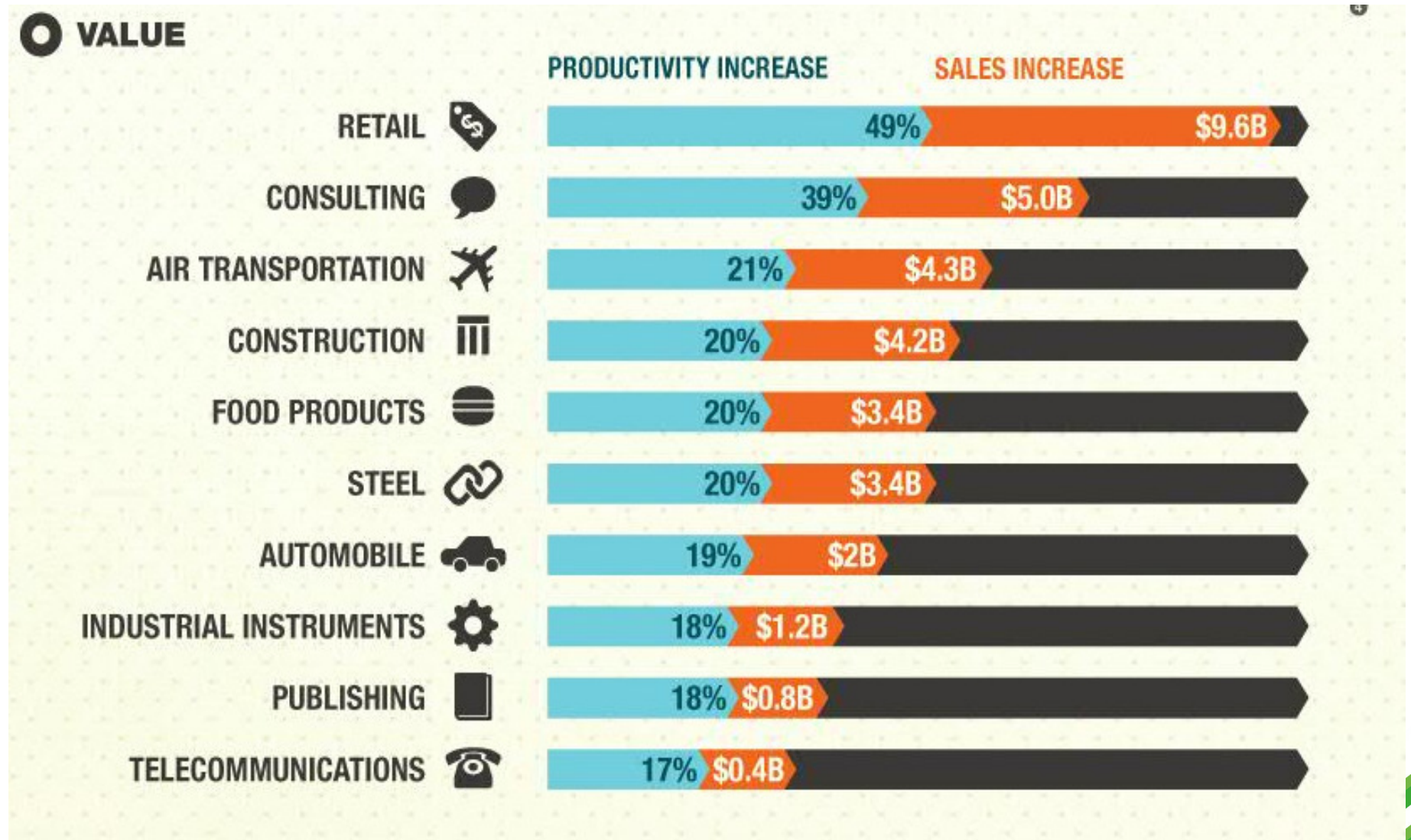
1 in 2 Business leaders say they don't have access to the information they need to do their jobs

83% of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

60% of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions



The Value of Big Data



Why are Big Data Analytics Important?

Big Data offers Better Insight for a Smarter Business and a Smarter World

60% increase

in retailers operating margins possible with big data¹

Potential annual consumer surplus

\$600 billion

from using personal location data globally¹

Potential annual value

\$300 billion

to US healthcare¹

Real opportunities for all industries

Big Data analytics changes how we have traditionally analyzed challenges

How to help a presidential secure re-election

How to help build better baseball teams

How to optimize our energy usage

¹ McKinsey Institute *Big data: The next frontier for innovation, competition, and productivity* May, 2011



The Challenges of Innovation

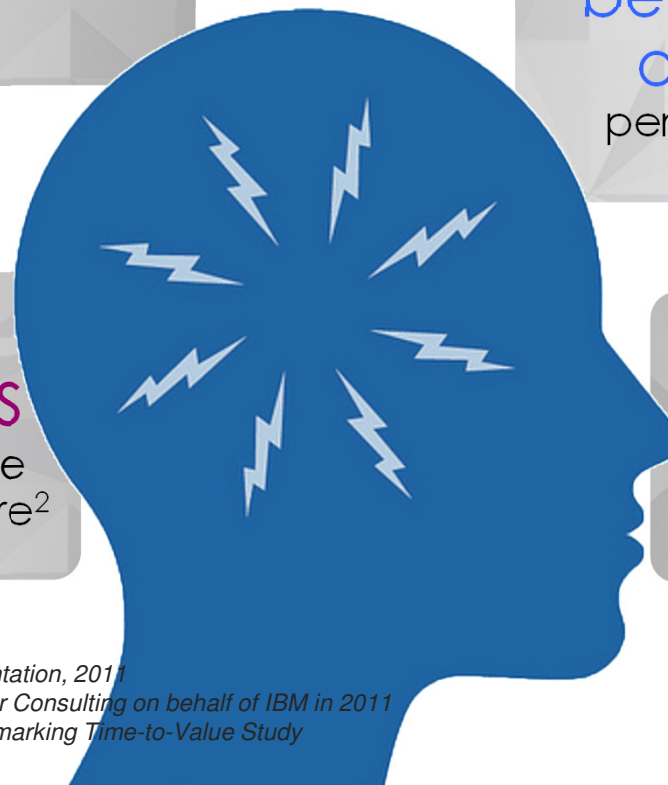
IT Professionals face unnecessary delays and costs as they deploy, maintain and update their capabilities

On average, more than
70% of IT budgets
are spent on operations and
maintenance¹

55% of IT professionals
experience downtime
that can last anywhere
**between minutes to
over a week** when
performing an infrastructure
upgrade²

It can take up to
4 to 6 months
just to establish hardware
and software infrastructure²

Nearly **66%**
of organizations fall behind
schedule when deploying
new IT capabilities³



¹IDC, Analyst Matt Eastwood, IDC Directions Presentation, 2011

²From a commissioned study conducted by Forrester Consulting on behalf of IBM in 2011

³IBM Market Insights Study – 2011 Business Benchmarking Time-to-Value Study



If this was easy, everyone would already be leveraging Big Data

Challenges of Leveraging Hadoop Technology

“Big Data offers big business gains but hidden costs and complexity present barriers that most organizations will struggle with”

- The Cost of Big Data, Eric Savitz, Forbes 5/2012

- Open source Apache Hadoop adaptation and implementation for enterprise usage is cumbersome
- Hadoop skills are in short supply
- Custom built solutions lack integrated cluster management, making system administration difficult
- Integration within the existing analytic ecosystem is critical and complex
- Time to value is elongated by the time to build and integrate everything
- Custom built solutions do not accelerate insight
- Most integrated solutions do not help with archival



Let's Simplify

From custom Big Data

...To organized simplicity



Designed to...

- Simplify the building, deploying and management of a Hadoop cluster
- Speed the time-to-value for Hadoop and unstructured data
- Maximize the overall analytic ecosystem
- Provide enterprise security and platform management



¹Based on IBM internal testing and customer feedback. "Custom built clusters" refer to clusters that are not professionally pre-built, pre-tested and optimized. Individual results may vary.





The Solution

IBM Pure System Family

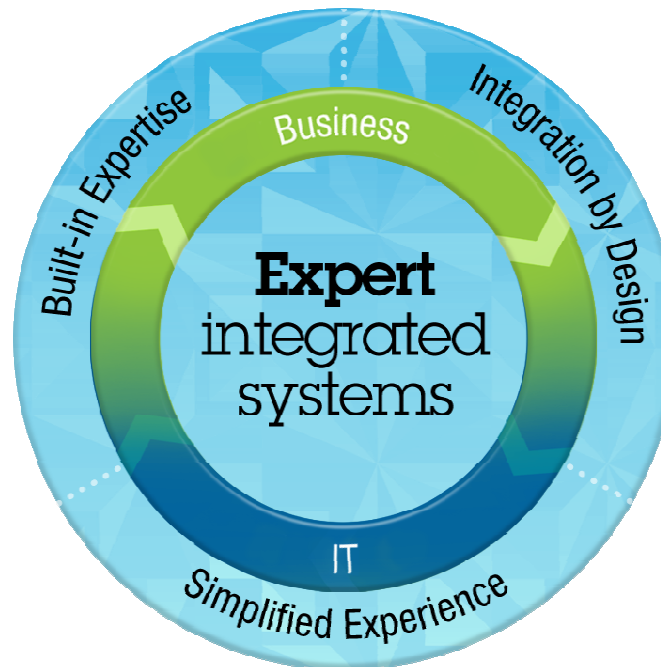


PureSystems™

Systems with integrated expertise and built for cloud

Built-in Expertise

Capturing and automating what experts do – from the infrastructure patterns to the application patterns



Integration by Design

Deeply integrating and tuning hardware and software – in a ready-to-go workload optimized system

Simplified Experience

Making every part of the IT lifecycle easier – with integrated management of the entire system and a broad open ecosystem of optimized solutions





The PureSystems Family Changes the Economics and Experience of IT and Accelerate Time to Value

PureFlex



Infrastructure

Delivering Infrastructure Services

PureApplication



Application Platform

Delivering Platform Services

PureData



New Model

Data Platform

Delivering Data Services



IBM PureData System

Meeting Big Data Challenges – Fast and Easy!



PureData^{New}
System for Hadoop

For Exploratory Analysis & Queryable Archive
Hadoop data services optimized for big data analytics and online archive with appliance simplicity

PureData
System for Analytics

For apps like Customer Analysis...
Data warehouse services optimized for high-speed, peta-scale analytics and simplicity

PureData
System for Operational Analytics

For apps like Real-time Fraud Detection...
Operational data warehouse services optimized to balance high performance analytics and real-time operational throughput

PureData
System for Transactions

For apps like E-commerce...
Database cluster services optimized for transactional throughput and scalability





The Solution

IBM PureData™ System for Hadoop



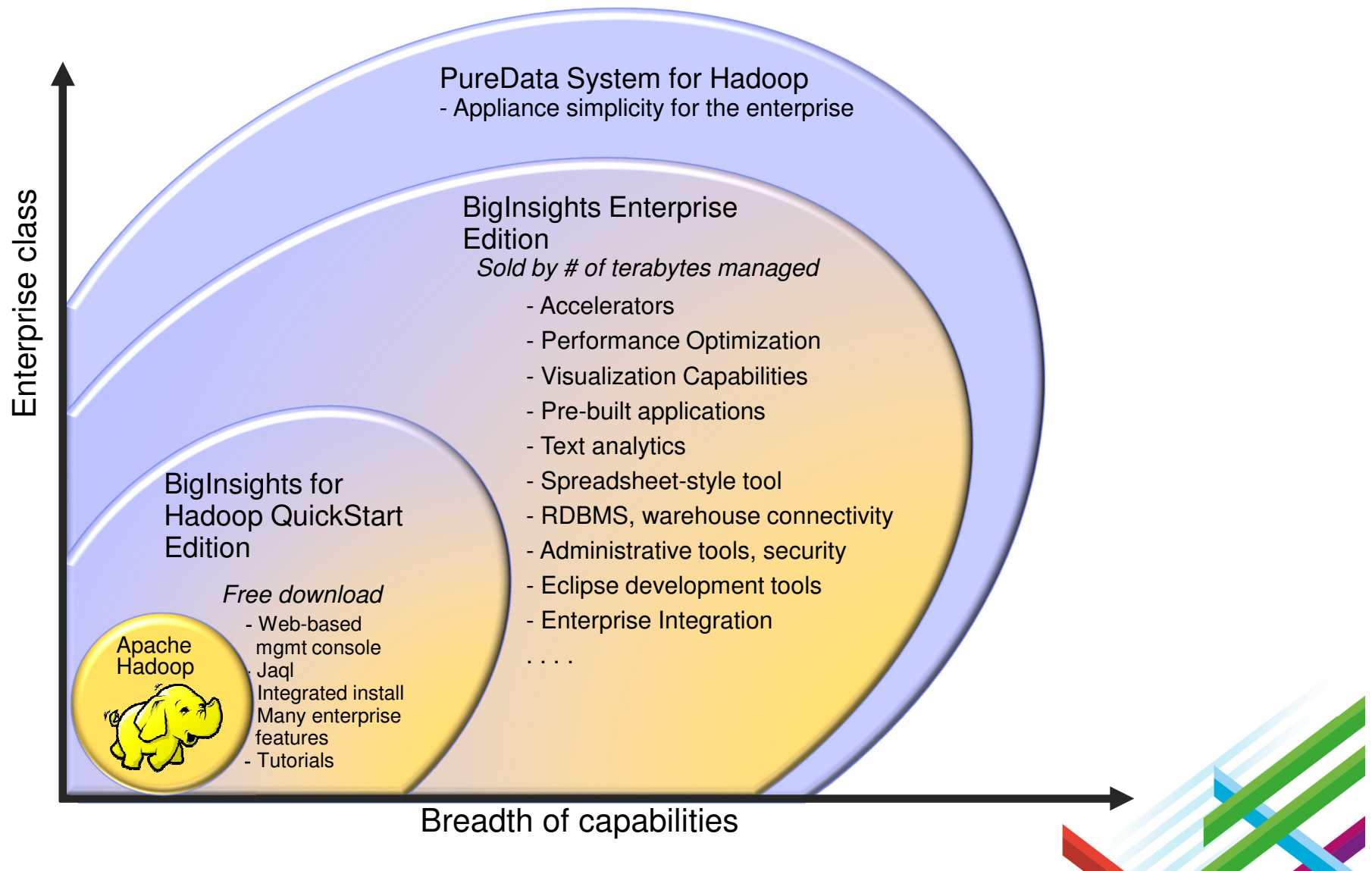
Announcing PureData System for Hadoop

Simplifying Big Data for the enterprise

- Accelerate time to value
- Accelerate time to insight
- Simplify big data adoption and consumption
- Extend the value of the data warehouse
- Implement enterprise class big data
- Minimize system setup and administration



From Getting Starting to Enterprise Deployment: InfoSphere BigInsights Brings Hadoop to the Enterprise





IBM PureData System for Hadoop

Accelerate Hadoop Analytics with Appliance Simplicity

PureData System for Hadoop

Speed

- Speed to insight with built-in analytics
- Speed to value with accelerated deployment

Simplicity

- Ready to load data in hours
- Integrated system management
- Appliance approach reduces complexity
- Single point of support

Smart

- Establish a cost efficient online data archive
- Easily leverage data across the big data platform
- Enterprise security, governance and high availability



Benefits of IBM PureData System for Hadoop

Exploring and analyzing more types of data



Built-In Expertise

*Accelerate
Big Data
Time to Value*

- **Deploy 8x faster**
than custom-built clusters¹
- **Built-in visualization**
to accelerate insight
- **Unlike big data appliances on the market, PureData System for Hadoop offers built-in analytic accelerators²**



Simplified Experience

*Simplify Big Data
Adoption & Consumption*

- **Single system console**
for full system administration
- **Rapid maintenance updates**
with automation
- **No assembly required, data load ready in hours**



Integration by Design

*Implement
Enterprise Class
Big Data*

- **Integrated Hadoop system with built-in archiving tools**
- **Delivered with more robust security**
than open source software
- **Architected for high availability**
- **Provides ability to load data at up to 14TB/hr**

¹Based on IBM internal testing and customer feedback. "Custom built clusters" refer to clusters that are not professionally pre-built, pre-tested and optimized. Individual results may vary.

²Based on current commercially available Big Data appliance product data sheets from large vendors.





What Makes PureData System for Hadoop Different?

Capabilities	Benefit
Accelerate Big Data Time to Value	<ul style="list-style-type: none">– IBM Accelerators for Text Analytics, Social Data, Machine Data – start analyzing your big data much more quickly and easily– IBM BigSheets – browser-based tool enabling business users to explore data stored in BigInsights clusters and create analytic queries without writing any code– IBM Big SQL – provides the ability to perform SQL queries across your data in Hadoop
Simplify Big Data Adoption & Consumption	<ul style="list-style-type: none">– Hardware Management – advanced web-based, hardware management and monitoring tool, including capabilities turning on and off nodes, rebooting nodes, monitoring hardware health, and hardware status alerts– Deploy system updates with ease
Implement Enterprise Class Big Data	<ul style="list-style-type: none">– High Availability – hardware HA capabilities: dual master nodes, redundant networking switches, power supplies and PDUs– Security – enterprise-level security built-in, including protected master nodes and isolated data nodes– Integrated archival software to help manage data growth– BigInsights Connectors – high-speed connectors for the IBM Netezza family of data warehouse appliances, IBM DB2®, IBM InfoSphere Warehouse and IBM Smart Analytics System– Netezza SQL-side Connectors – Netezza NPS-side read and write to and from HDFS & provides archiving functionality

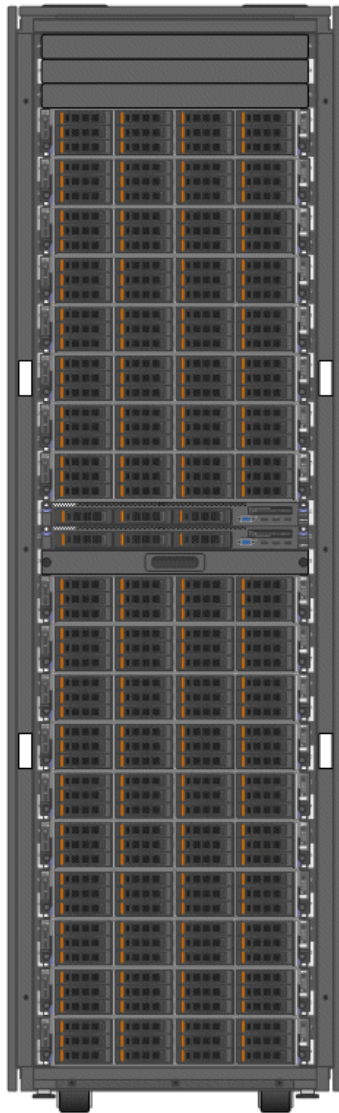


Robust Enterprise Readiness

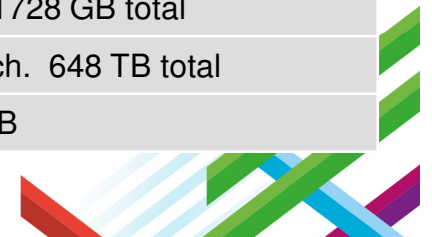
- **High Availability**
 - **Hardware:** Dual Master Nodes, Redundant Network Switches, Power Failover
 - **Software:** Customized Linux HA
- **Security**
 - Protected Master Node with secure port access
 - Data Node Isolation with secure port access
 - Integration with IBM Guardium - best in class security
- **Integrated Cluster Management**
 - Hardware monitoring and management via BigInsights Web Console



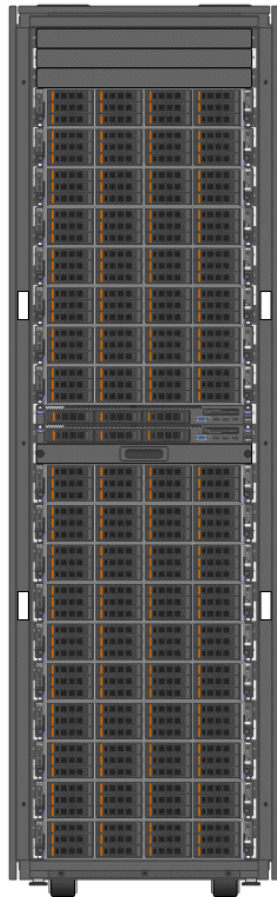
Key Features and Specifications



Key Features	
Hadoop Distribution	– InfoSphere BigInsights V2.1
Built-in Analytics/Accelerators	<ul style="list-style-type: none"> – IBM BigSheets – IBM Accelerator for Text Analytics – IBM Accelerator for Social Data – IBM Accelerator for Machine Data – IBM Big SQL
Development / Administration	<ul style="list-style-type: none"> – Eclipse-based Development Environment – Exposed Node Management
Enterprise Readiness	<ul style="list-style-type: none"> – Security – High Availability SW & HW – Hardware management & monitoring
Data Warehouse Integration	<ul style="list-style-type: none"> – Enterprise data warehouse connectors – Archival capabilities
Specifications	Full Rack
Management Nodes	1 primary, 1 standby (x3550 M4)
Data Nodes	18 (x3630 M4)
CPU Cores	216
Memory	96 GB per node, 1728 GB total
Raw Storage	216 drives, 3 TB each. 648 TB total
User Space	216 TB



Hardware Overview



1 Management Switch

- 48 x 1 Gbe

2 Rack Switches

- 48 x 10 Gbe, 4 x 40 Gbe

2 Management Nodes (Active/Passive)

- Dual 8 Core Intel 2.4 GHz
- 128 GB
- 3 x 3TB 3.5" Drives

18 Data Nodes

- Dual 6 Core Intel 2.2 GHz
- 96 GB memory
- 14 x 3 TB 3.5" Drives

- User Data Capacity: 216 TB uncompressed / 864TB compressed
- Power Requirements: 9.6 kW
- Cooling Requirements: 32,757 BTU/hr





How to Leverage IBM PureData™ System for Hadoop



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 Key Use Cases for PureData System for Hadoop



Big Data Exploration

- *Find new insights*
- *Visualize new data sources*
- *Understand data to improve decisions*

Big Data Exploration

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

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

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

Data Warehouse Augmentation

- *Queryable Archive*
- *Pre-Processing Hub*
- *Exploratory Analysis*

data for improved business results



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency



Use Case: Big Data Exploration

Provides an exploratory environment for data analysts to understand new and emerging data sources. Enables insight without the cost of modeling and structuring the data. Use analytic tools embedded within the platform to perform analysis on data in it's native form.



Use Cases

- Explore new data and previously untapped sources
 - New structured data sources
 - Video, audio, text documents, sensor data, log information, etc.
- Visualize and gain new insight with easy to use spreadsheet-style analysis
- Identify useful information that would add value when integrated with other data sources
- Used for data profiling to understand data before moving to other systems





Augmenting the Data Warehouse with Hadoop



Use Case: Pre-Processing Hub

Provides landing area for data from other sources. Enables aggregation, initial integration, visualization and exploration before data moves to other parts of the analytic ecosystem. Enables customer to aggregate data in single location. Landing Zone often serves and initial step to more advanced Hadoop analytics.



Use Cases

- Aggregation of data
- Pre-process cleansing
- Compliance requirements
- Simple analytics / exploration



Use Case: Queryable Archive

Enables an immediate storage alternative. Queryable Archive often serves and initial step to more advanced integration with their EDW and advanced Hadoop analytics.



Use Cases

- Immediate storage alternative of cold data
- Cost savings for cold data
- Compliance requirements
- Simple analytics / exploration
- When you find new correlations, go back and re-mine the archive data to gain additional insight

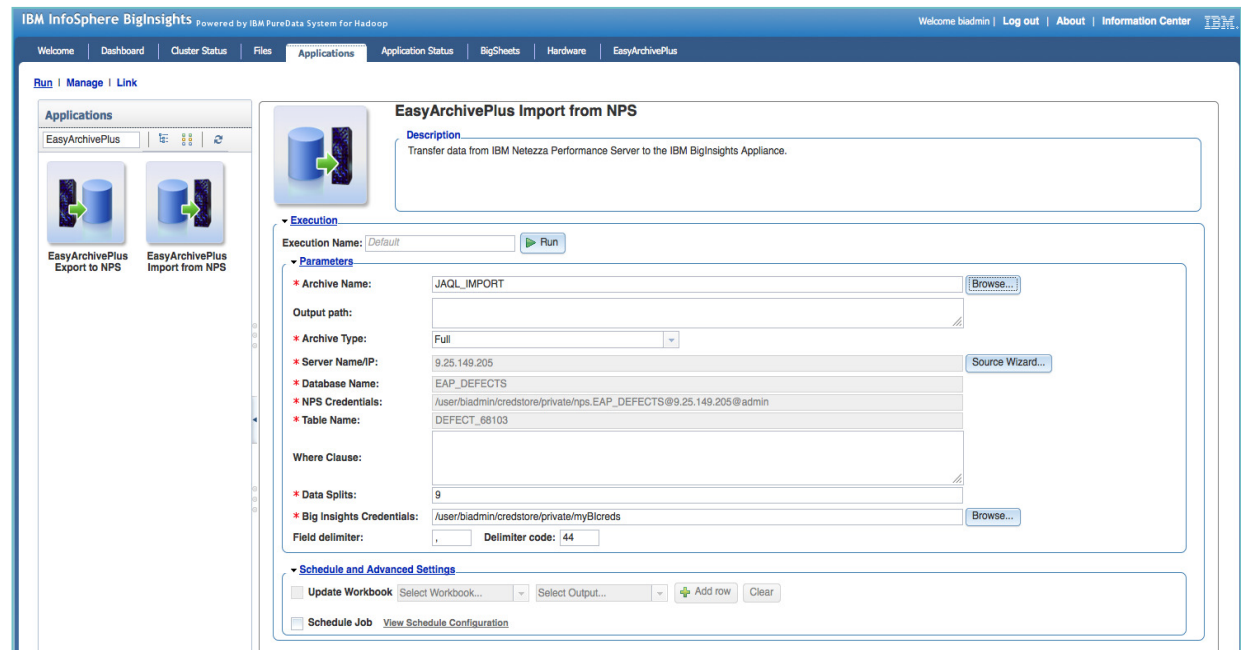




Optim EasyArchive: Overview

For Easy Data Offload from PureData System for Analytics Out-of-the-box

- Included application allows migration of data from PureData System for Analytics to PureData System for Hadoop at over 2TB/hr, out-of-the-box
- Provides simple, built-in user interface to allow users to migrate data between systems easily
- Enables quick configuration and scheduling of data migration
- Employs parallel processing between BigInsights and NPS using
- Leverages IBM-developed MapReduce programming for parallel processing
- Utilized Hive to allow for immediate access to migrated data



PureData System for Hadoop

Bringing Big Data to the Enterprise

PureData System for Hadoop provides value to the customer in several key ways:

For IT:

Accelerates Time-to-Value by allowing Hadoop clusters to be deployed 8x faster than homegrown clusters – saves time, money, and reduces complexity

For Business Users:

Accelerates Time-to-Insight through BigSheets, Big SQL and BigInsights analytics accelerators (advanced text, social data, machine data), – allows more immediate benefit of Big Data analytics

For Finance:

Provides ability to easily move cold data from DW to Hadoop which reduces cost of storing that data (\$10K/TB or greater for DW and ~\$500/TB for PDH/Hadoop) while keeping the data online for analytics



The new PureData System for Hadoop

A Smarter Hadoop Solution

Simplified Experience

- Designed for easy and quick deployment
- Built-in tools designed for users to derive value quickly
- Easy connectivity to common data warehouse systems

Built-in Expertise

- Enables ‘what-if analysis’ and advanced analytics
- Supports structured, semi-structured, and unstructured data
- Built-in text processing engine and library of annotators to analyze large volumes of text-based information

Integration by Design

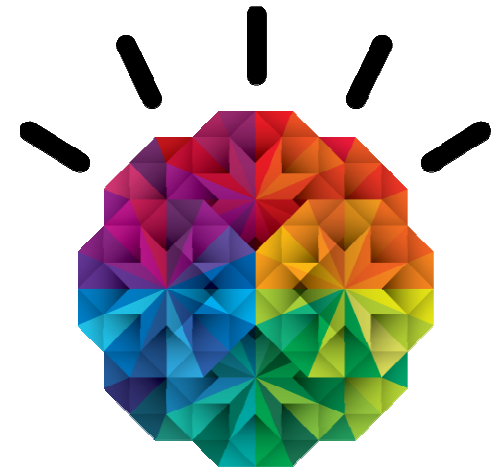
- InfoSphere BigInsights software, cluster management, and IBM System x® servers
- Automatic parallelization and resource optimization to scale economically
- Enterprise-class security and platform management



Take the Next Step...

Discover the value and begin your journey with IBM PureSystems:

- Visit ibm.com/puresystems to learn more
- Join the conversation about this new category of computing:
 - Twitter: @IBM PureSystems
Hashtag: #IBMPureSystems or #expertintsys
 - YouTube Channel: expertintegratedsys
 - Blog: expertintegratedsystemsblog.com
- Developers – Get started today with our no charge trial offerings!
ibm.com/developerworks/puresystems/try
- Explore PureSystem partner solutions
ibm.com/puresystems/centre
- Take a Test Drive
ibm.com/PureExperience



IBM BusinessConnect
Vernetzter, intelligenter und informierter denn je



Thank You!

