



IBM Big Data & Business Analytics



Sevilay Kurt, Client Technical Architect – sevilay@tr.ibm.com
Ayhan Önder, Client Technical Professional – ayhano@tr.ibm.com



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

44x

as much Data and Content

Over Coming Decade

35 zettabytes

2020

Velocity

Variety

Volume

2009 800,000 petabytes 80%

Of world's data is unstructured



... And OrganizationsNeed 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%

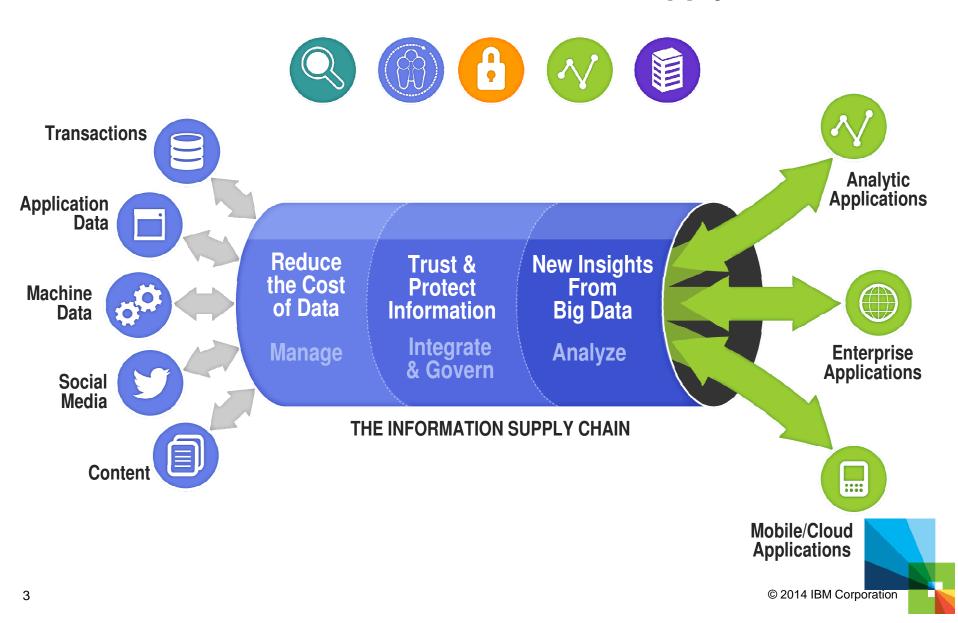
of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

60%

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



From Natural Resources to Information Supply Chain





Meeting today's challenges means thinking differently about architecture

Reduce complexity

Each system tuned for the different needs of different analytic and transaction processing

Accelerate time to value

Data management expertise is built in each system and ready for immediate use

Improve IT economics

The entire system lifecycle is simplified from acquisition to retirement



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 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





Big Data Use Study

Big data adoption stages



Respondents were asked to identify the current state of big data activities within their organizations. Percentage does not equal 100% due to rounding. Total respondents=1061



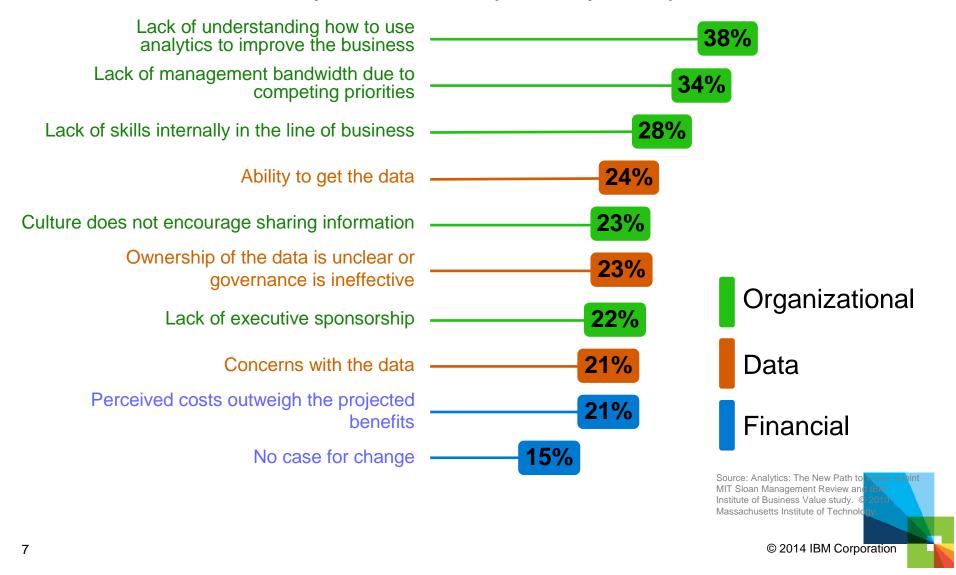
2012 Big Data @ Work Study surveying 1144 business and IT professionals in 95 countries

Why Analytics Matters



Organizational, not data or financial concerns, are holding back adoption

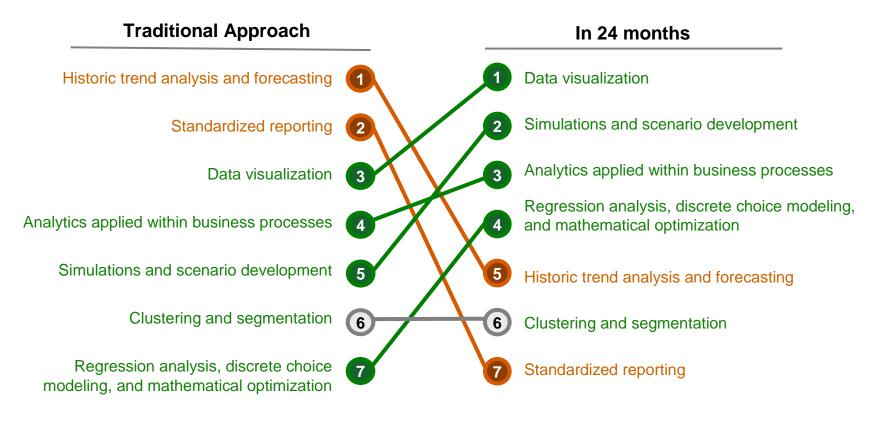
Primary obstacles to widespread analytics adoption





Organizations want to "see" insights more clearly – and act on them

Analytic techniques that provide the most value



Source: Analytics: The New Path to Value, a joint MIT Sloan Management Review and IBM Institute of Business Value study. © 2010 Massachusetts Institute of Technology.

Increased or sustained value

O Decreased in value

Why Analytics Matters



Three segments emerged based on analytics capabilities with clear patterns illustrating common priorities, activities and roadblocks

	Brand and market management	Transformed (21%)
	General management	 Highest levels of analytics prowess and experience
	Workforce planning and allocation	Seeking targeted revenue growth
	Customer experience	 Feel the most pressure to do more with analytics
	Risk management	
Р	roduct research and development	Experienced (46%)
	Customer service	 Seeking to grow revenue with focus on cost efficiencies
Strategy and business development		 Seeking to expand ability share information and insights
Sales and marketing		Aspirational (33%)
Operations and production		 Turn to analytics for ways to cut costs
Financial	Focused on analytics at point of need	

Commercial HPC > Business Analytics

Mathematical-Analytics Landscape



Degree of Complexity

Increasing prevalence of compute and data intensive parallel algorithms in commercial workloads driven by real time decision making requirements and industry wide limitations to increasing thread speed

Competitive Advantage



Merging the Traditional and New Data Approaches

Traditional Approach
Structured & Repeatable Analysis

Business Users

Determine what question to ask





IT

Structures the data to answer that question



Monthly sales reports
Profitability analysis
Customer surveys

Big Data Approach

Iterative & Exploratory Analysis



IT

Delivers a platform to enable creative discovery



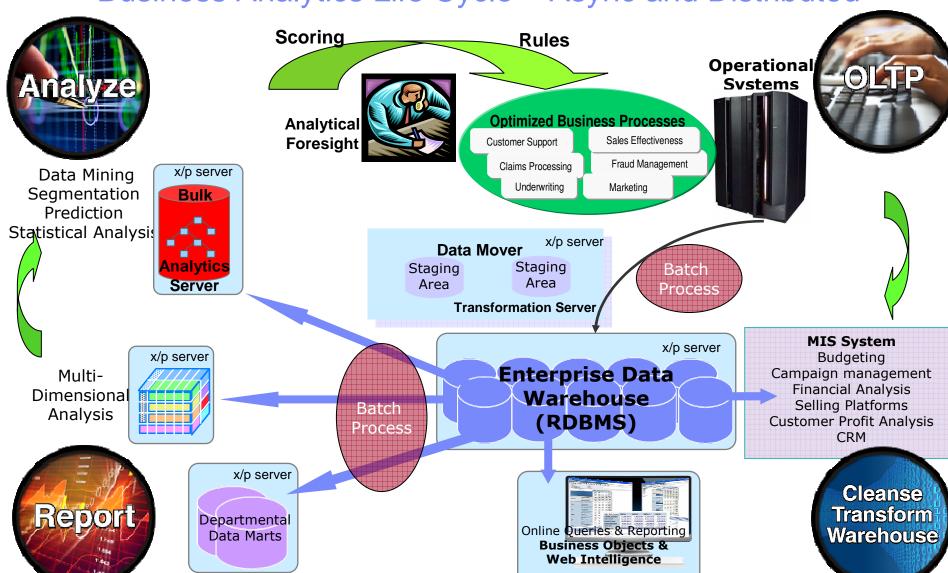
Business

Explores what questions could be asked

Brand sentiment
Product strategy
Maximum asset utilization



Business Analytics Life Cycle – Async and Distributed

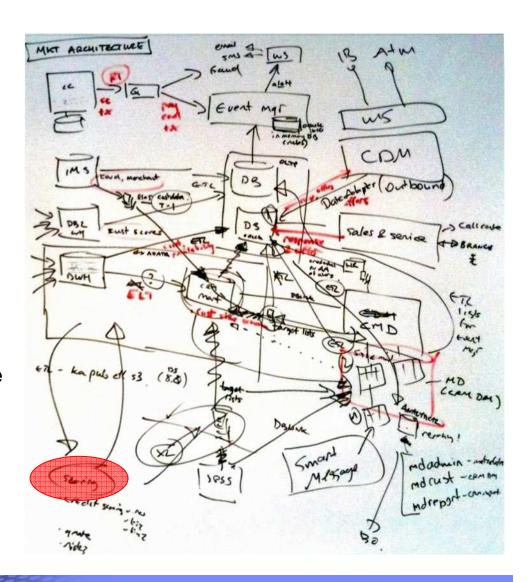




What usually happens when all views of data are **not** co-located

Client key concerns:

- Cannot deliver real-time analytics
- Inadequate performance
- Governance model
- Data latency
- Data completeness
 - Not all in one source
 - Lack access to fine-grained data
 - Lack "customer intent" e.g. cancelled transactions
- Multi-day workshop captured the complexity of the current architecture
- The picture does not show all the steps before the data gets to the off-platform warehouse

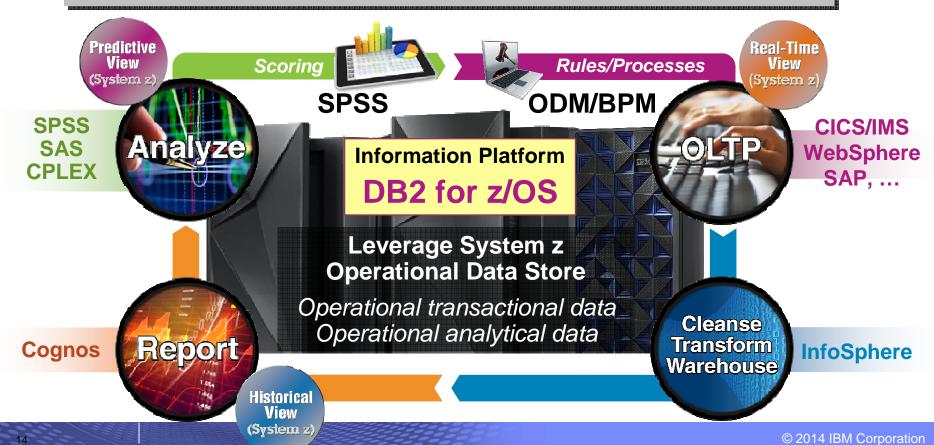




The System z strategy integrates transactions and business critical analytics into one streamlined, end-to-end data lifecycle

Differentiate DB2 for z/OS and System z to integrate analytics with real time OLTP Superior end/end analytics lifecycle integration

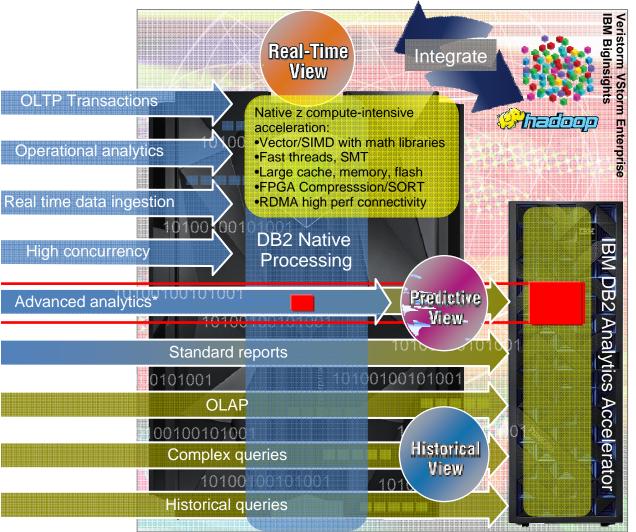
Better business response,
Reduced data movement, reduced complexity, reduced configuration resources,
More accurate, more secure, more available





System z: A Hybrid Transaction and Analytics Processing platform

A single workload-optimized system for accelerating decisions to the speed of business *Everything is online – analytics in the right place*



More timely insights from data

- High-speed analytics easily integrated into operational applications
- Historical views are quickly analyzed for more train-of-thought analysis
- Decision makers can perform business analysis they never dared in the past
- Secured environment for highly sensitive data
- Speeds batch reporting cycle to meet stricter service level agreements

Operational benefits

- Configuration simplification
- Single point of entry
- ■Reduced data movement
- High fidelity data
- Dynamic routing for most efficient fit for purpose execution architecture
- •Single environment for security, logging, back-up, and recovery
- Competitive price/performance



Real-time, data-oriented solutions transform these questions into actionable insights



How do I target and retain my best customers?

ex.: churn management

Work from a single, current view of the customer

Next Best Action solutions



How do I reduce fraud?

•ex.: real-time fraud identification/prevention

Drive insights directly into payment systems

Anti-Fraud solutions



How do I manage risk?

•ex.: operational and financial risk visibility

Develop a real-time enterprise view of risk-related data Governance, Risk and Compliance solutions



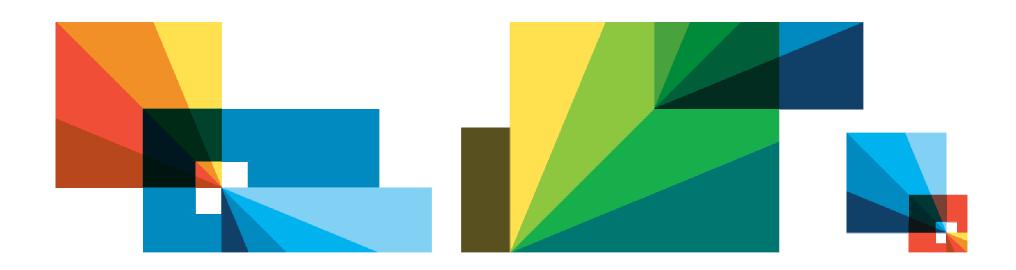
How do I focus limited resources where they will be most effective?

ex.: supply-chain optimization

Know where everything and everyone is right now **Resource Optimization** solutions



IBM Big Data Technology



A Big Data Platform Manifesto

CONSUMABLE "I'm a scientist that just wants to run my applications faster"	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	In-Memory Analytics 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



Systems Application Discovery Accelerators Hadoop System Computing Data Warehouse Information Integration & Governance

Watson 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



Systems Application Discovery Accelerators Hadoop Stream Computing Data Warehouse Information Integration & Governance

Watson 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



Systems Application Discovery Accelerators Hadoop Stream Computing Data Warehouse Information Integration & Governance

Watson 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

BigInsights: Value Beyond Open Source



Enterprise Capabilities

Visualization & Exploration

Development Tools

Advanced Engines

Connectors

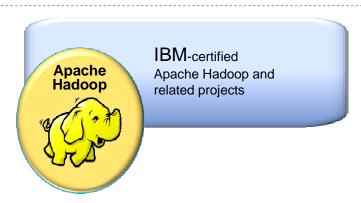
Workload Optimization

Administration & Security

Key differentiators

- Built-in text analytics
- Enterprise software integration
- SQL support
- Spreadsheet-style analysis
- Integrated installation of supported open source and other components
- Web Console for admin and application access
- Platform enrichment: additional security, performance features, GPFS (alternative file system), . . .
- World-class support
- Full open source compatibility

Open source components



Business benefits

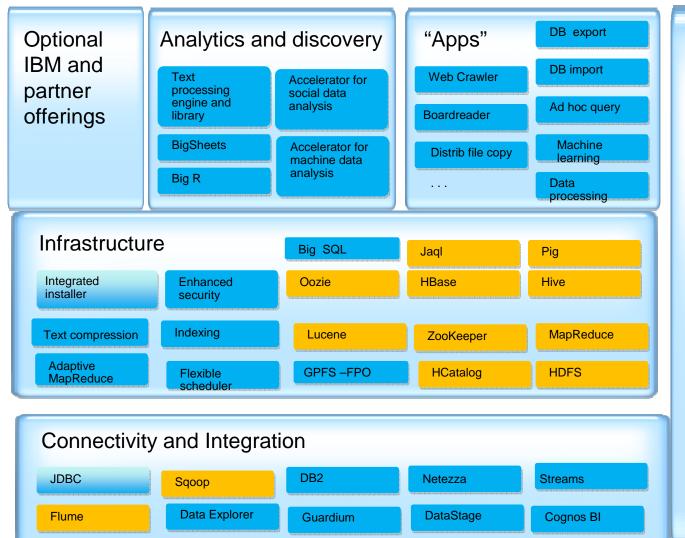
- Quicker time-to-value due to IBM technology and support
- Reduced operational risk
- Enhanced business knowledge with flexible analytical platform
- Leverages and complements existing software



BigInsights Enterprise Edition

Open Source

IBM



Administrative and development tools

Web console

- Monitor cluster health, jobs, etc.
- Add / remove nodes
- Start / stop services
- Inspect job status
- Inspect workflow status
- Deploy applications
- Launch apps / jobs
- Work with distrib file system
- •Work with spreadsheet interface
- Support REST-based API
- Create / view alerts
- . . .

Eclipse tools

- Text analytics
- MapReduce programming
- Jagl, Hive, Pig development
- BigSheets plug-in development
- Oozie workflow generation



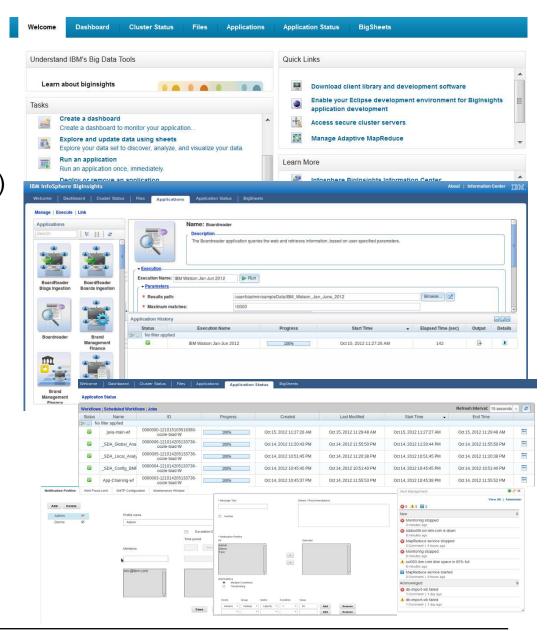
Integrated Web Console

Manage BigInsights

- Inspect /monitor system health
- Add / drop nodes
- Start / stop services
- Run / monitor jobs (applications)
- Explore / modify file system
- Create custom dashboards
- **–** . . .

Launch applications

- Spreadsheet-like analysis tool
- Pre-built applications (IBM supplied or user developed)
- Publish applications
- Monitor cluster, applications, data, etc. Create / view event alerts.





Big SQL

- Standard SQL syntax and data types
 - Joins, unions, aggregates . . .
 - VARCHAR, decimal, TIMESTAMP, . . .

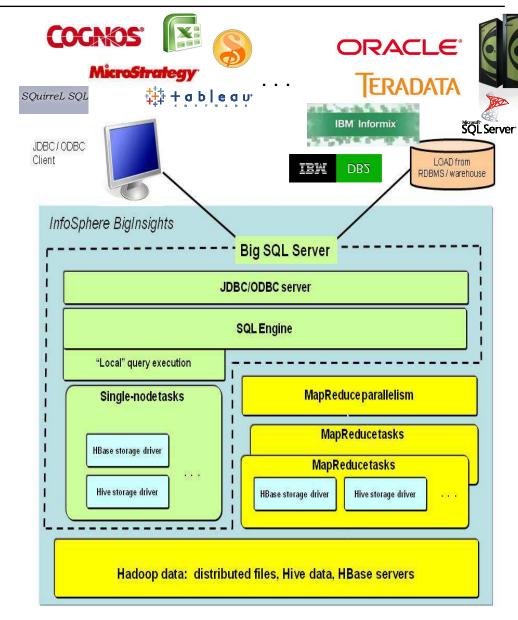
JDBC/ODBC drivers

- Prepared statements
- Cancel support
- Database metadata API support
- Secure socket connections (SSL)

Optimization

- MapReduce parallelism or...
- "Local" access for low-latency queries
- Varied storage mechanisms appropriate for Hadoop ecosystem
- Integration
 - Eclipse tools
 - DB2, Netezza, Teradata, etc. (via LOAD)
 - Cognos Business Intelligence

- , , ,





IBM is named a leader in The Forrester Wave™: Big Data Hadoop Solutions, Q1 2014

Figure 2 Forrester Wave™: Big Data Hadoop Solutions, Q1 '14



FORRESTER®

"IBM has offerings in grid computing, databases, and many other data management technologies that it can bring to a comprehensive Hadoop solution."

"IBM has advanced analytics tools, a global presence, and implementation services, so it can offer a complete big data solution that will be attractive to many customers."

"IBM has more than 100 Hadoop deployments, some of which are fairly large and run to petabytes of data."

Source: Forrester Research, Inc. The Forrester Wave is copyrighted by Forrester Research, Inc. Forrester and Forrester Wave are trademarks of Forrester Research, Inc. The Forrester Wave is a graphical representation of Forrester's call on a market and is plotted using a detailed spreadsheet with exposed scores, weightings, and comments. Forrester does not endorse any vendor, product, or service depicted in the Forrester Wave. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change.

IBM InfoSphere BigInsights For Hadoop



Audited STAC Report™
Securities Technology Analysis Center

IBM InfoSphere BigInsights for Hadoop





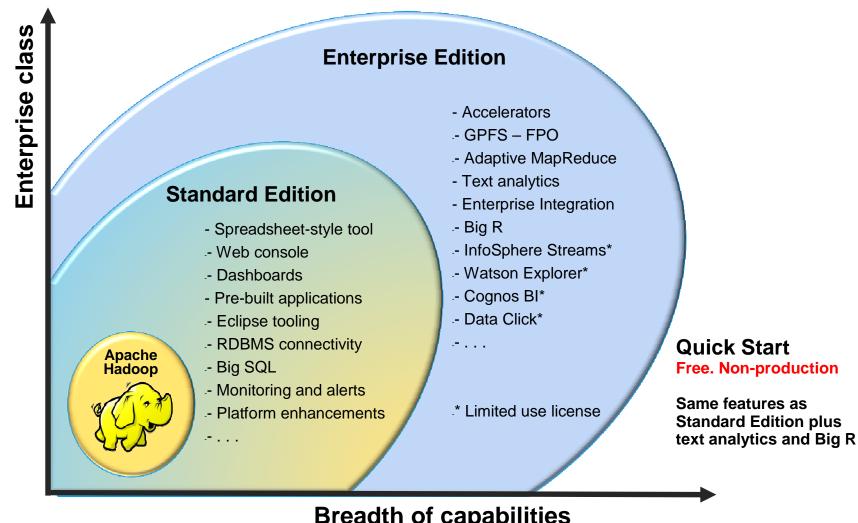
Open Source



1. 4x is approximate value. See the STAC Report[™] at http://www.stacresearch.com/node/15370. Testing involved the SWIM benchmark (https://github.com/SWIMProjectUCB/SWIM) and jobs derived from production workload traces. Testing was conducted in controlled laboratory conditions.



From Getting Starting to Enterprise Deployment: **Different BigInsights Editions For Varying Needs**



Breadth of capabilities



Systems Application Discovery Accelerators Hadoop Stream Computing Data Warehouse Information Integration & Governance

Watson 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



IBM InfoSphere Streams

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





How Streams Works

→ Continuous ingestion

scheduling analytics across hardware hosts, → Continuous analysis establishing streaming connectivity Filter / Sample **Transform Annotate** Correlate Classify

Infrastructure provides services for

Achieve scale:

By partitioning applications into software components

By distributing across stream-connected hardware

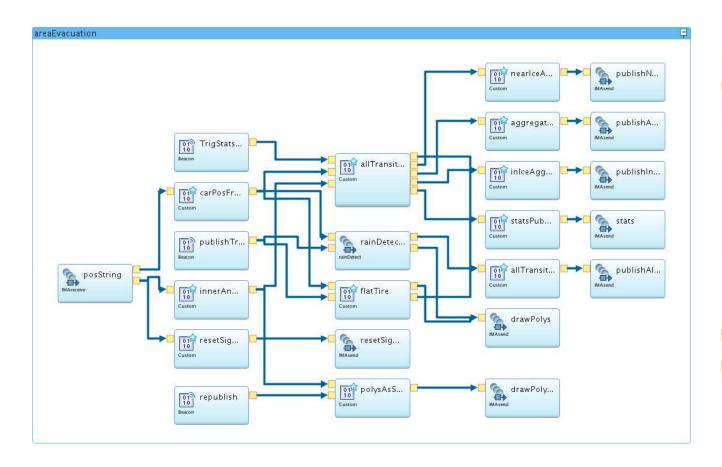
Where appropriate:

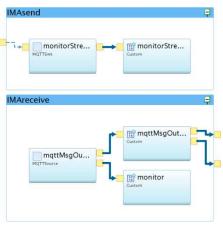
Elements can be *fused* together

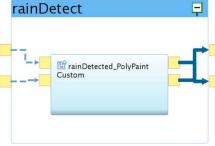
for lower communication



Sample Application Graph







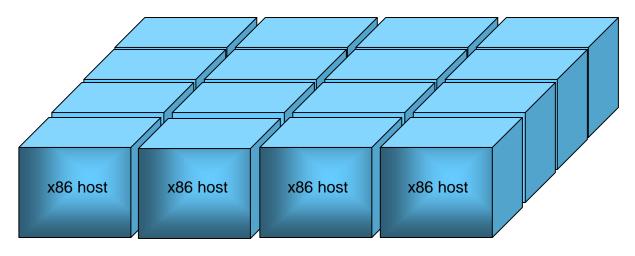


Streams Runtime Illustrated



Optimizing scheduler assigns jobs to hosts, and continually manages resource allocation

Commodity hardware – laptop, blades or high performance clusters





Streams Runtime Illustrated

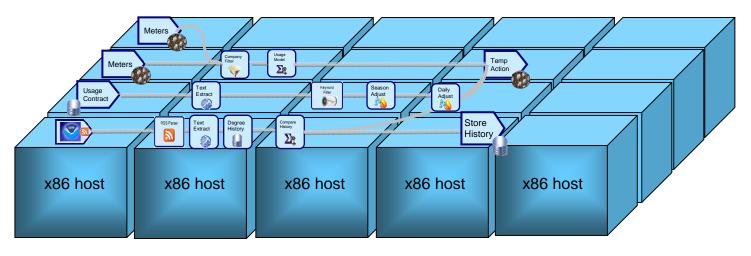


Optimizing scheduler assigns PEs to hosts, and continually manages resource allocation

Commodity hardware – laptop, blades or high performance clusters

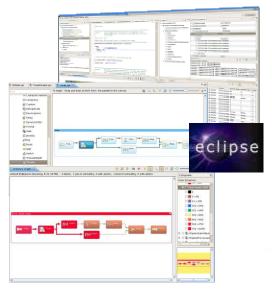
Dynamically add hosts and jobs

New jobs work with existing jobs





Comprehensive Tooling



- Eclipse IDE, remote development
- REST API
- Web console
- Drag & Drop editor
- Instance graph
- Streams visualization
- Streams debugger

Sophisticated Analytics with Scale-out Architecture Toolkits & Accelerators



- Clustered runtime for nearlimitless capacity
- RHEL, CentOS, SLES support
- X86 & Power multicore hardware
- InfiniBand support
- Ethernet support



- Big Data, CEP, Database, Data Explorer, DataStage, Finance, Geospatial, Internet, Messaging, Mining, R, SPSS, Standard, Text & Timeseries toolkits
- Telco & Social Media accelerators



Toolkits and Accelerators to Speed Up Development

Standard Toolkit

Relational Operators

Filter Sort Functor Join

Punctor Aggregate

Adapter Operators

FileSource
FileSink
UDPSource
UDPSink
UDPSink
Export
TCPSource
Import

TCPSink MetricsSink

Utility Operators

Custom Split

Beacon DeDuplicate

Throttle Union

Delay ThreadedSplit
Barrier DynamicFilter

Pair Gate
JavaOp Switch
Parse Format

Decompress CharacterTransform

XML Operator

XMLParse

IBM Supported Toolkits

Database DataStage

Big Data Data Explorer

Messaging Internet
Text Analytics Mining
SPSS CEP

Time Series Geospatial

Financial R

Open-Source Toolkits

JSON HTTP/REST

OpenCV Accumulo

HBase ...

Big Data Accelerators

Telco Event Data Analytics

Social Data Analytics

User-Defined Toolkits

Extend the language by adding user-defined operators, types,

and functions

37 © 2013 IBM Corporation



IBM's Key Platform Capabilities

Systems Management Application Development Discovery Accelerators Hadoop System Computing Data Warehouse Information Integration & Governance

Watson 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



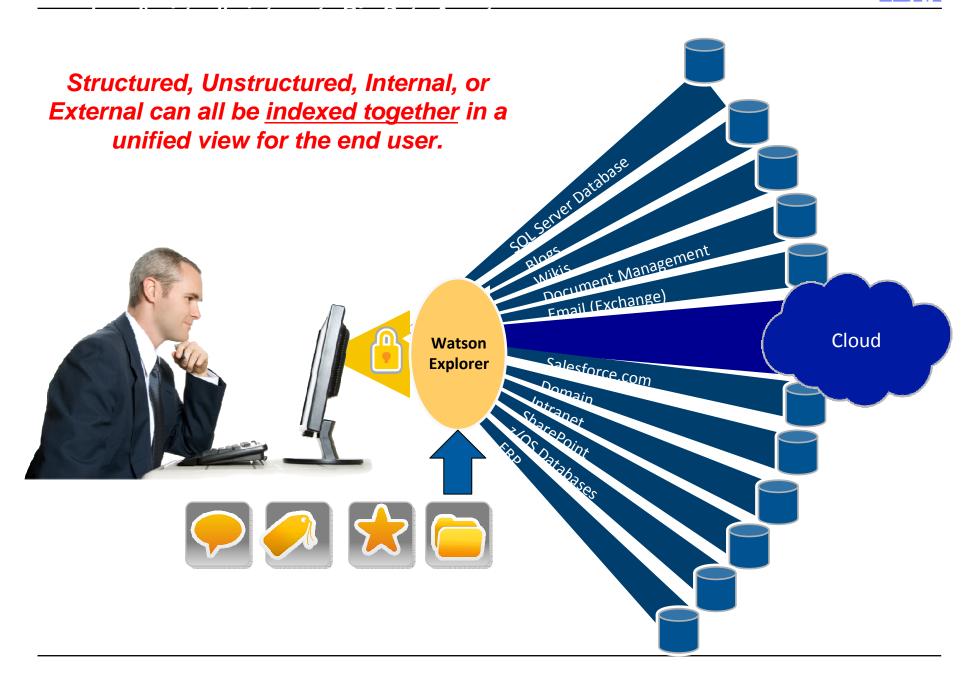
IBM Watson Explorer

Consolidates and visualizes information across enterprise applications and big data assets – helping organizations discover, analyze & integrate their data in a unified view.

Value:

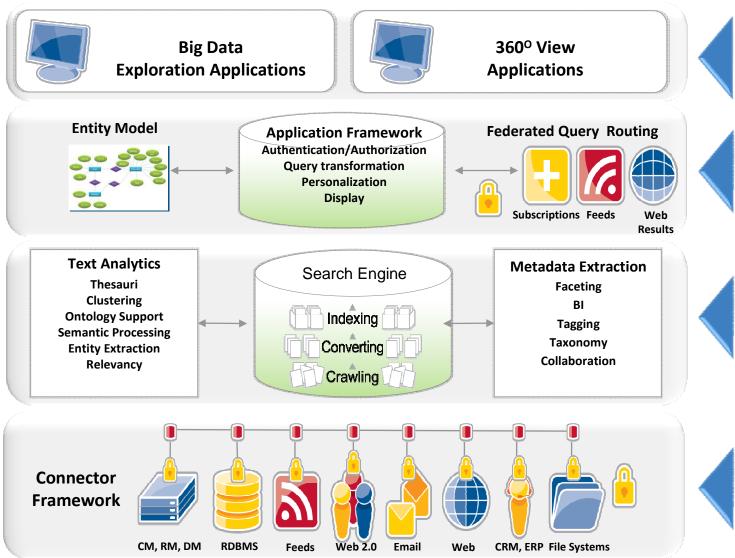
- Better visibility into available data assets
- •Better use and re-use of information
- •Better decision-making at all levels
- Deeper understanding of customers
- More efficient operations
- •Improved compliance and risk reduction





Watson Explorer application architecture





Add folder

TEM

Topic Clusters

Top 193 Results

Marketing (27)

Tech Choices (25)
 Not Classified (17)

. Content Management (20)

. Communications (15)

. Portais, Content &

* Social (16)

more I all

Refinements

▼ Category

Management (141)

Not Classified (24)

Rolling Review (5)

C.g. Lynch (4)

Competitor (9)

Mobile (6)

Hadoop (4)

Analyst (3)

Queries (3)

more

▼ Fliatype

Competitor 2 (7)

Sales Education (5)

Velocity 6.0 Research (4)

In Marketing:

Analyst X

Meeting with Journo (3)

Idg News Service (5)

Microsoft (105)

Windows (51)

Review (32)

→ Rolling (8)

▼ Folder

BI (10)

Mobile (55)

* Practices (14)

Overview (14)
 Vivisimo (10)

Collaboration (16)

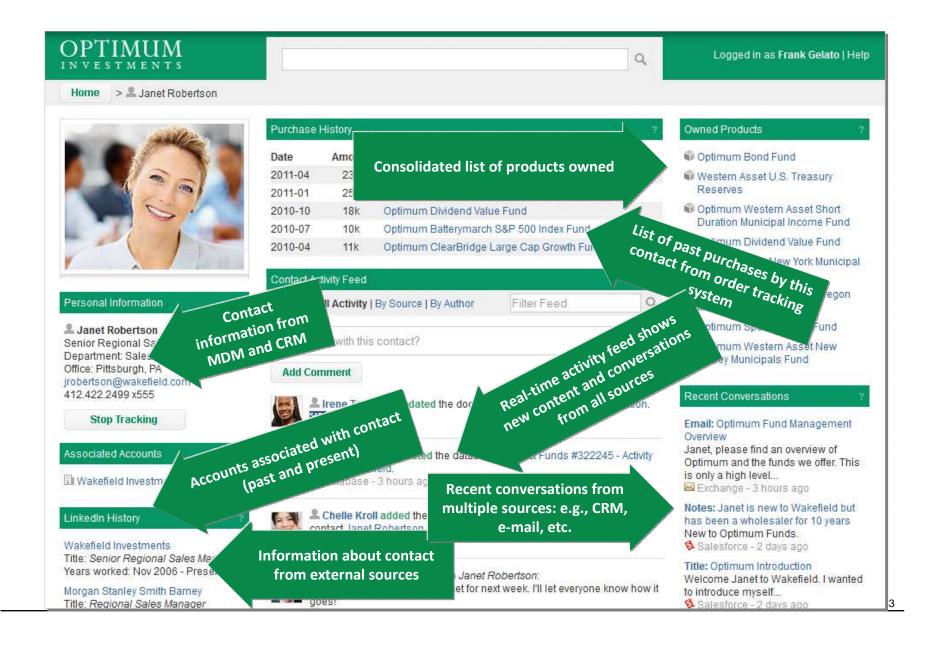
Sort by: Date Relevance

remix

Dynamic Categorization harePoint About I Help I Advanced I Profile I Logout - gelato Trade Publications Q Results 1-10 of about 452 DOM: Your query has been expanded. Show Expansions Save Query 0 documents selected. Actions * Select/deselect all on this page Graphical Refinements Employees Stacy Leidwinger Department: Pittsburgh, PA Top Tags Senior Director of Product Management collaboration (10), competitors (7), social tagging (4), analysts (4), social networking (4) Title: Extension: Email: S.Leidwinger@vivisimo.com Leveraging Selected: 6/1/2005 - 2/25/2009 Structured and Average Rating unstructured content 1. Rolling Review: Web 2.0 Tools Demand A Cautious Approach ... staff, many are investigating wikis, file-sharing services, and other consumer technologies to deliver Web-based collaboration inexpensively. Bringing these tools into a corporate environment presents thorny issues, how ing visibility and control, all in an easy-tonavigate application. But are there cost-effective collaboration t In this Holling Review, we' ... Highly relevant, 8 Vivisimo-demos: This was useful for ABC. personalized 9 Vivisimo-demos: This works here, but over there. results + add new comment show all 9 comments Selected: 1 - 4.5 All Tags: BCP COLLABORATION DDA [show all 10] **Enhanced** My Tags: add/edit tags by social Folders Competitor X RI X Security X B Hadoop X Gueries X Competitor 2 X collaboration Edit * IBM Refinements based All Results - 78K - Trade Publications - Avg rating: ***** Avg rating 4. (5 ratings) on structured Analyst Meeting BI information tter for Professional Networking, Job Hunting and Collaboration? Competitor Preston Gralia and Jake Widman March 06, 2008 — Computerworld — Social networking is no longer the Next Big Thing; it's now as much part of our Competitor 2 Web experience as search engines. Previously considered the province of kids who wanted to keep up with class gossip, social networking services are Hadoop being co-opted ... 1 Tucker: this is useful for Biorn Jennin + add new comment Mobile Organize content All Tags: Big Insights COLLABORATION Queries into virtual folders Reportes Octubre 2013 My Tags: Big Insights add/edit tags Security

Enterprise-wide discovery enables a new category: 360 degree information application







InfoSphere BigInsights Quick Start is the newest member of the BigInsights family

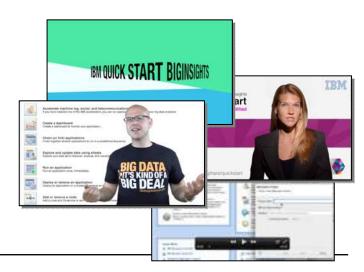
What is BigInsights Quick Start?

- No charge, downloadable edition that allows you to experiment with enterprise-grade Hadoop features
- Simplifies the complexity of Hadoop with easy-to-follow tutorials and videos
- No time or data limitations to allow you to experiment for a wide range of use cases



IBM InfoSphere BigInsights
Quick Start







IBM InfoSphere Streams Quick Start: Real Time Analytic Processing at your Fingertips

What is Streams Quick Start?

- No charge, downloadable edition to allow you to experiment with stream computing
- No time or data limitations for use on your unique use cases in non-production systems
- Sophisticated analytics for streaming data quickly ingest, analyze and correlate data
- Comprehensive development tools and scaleout architecture to get up and running quickly, support available through forums & communities**



IBM InfoSphere Streams

Quick Start





Big Data Stampede: Leading the Charge for Big Data Success

IBM Expertise

Removes the guesswork and delivers savings in time and cost

Big Data Platform

Provides the use of unmatched capabilities

Skills & Knowledge Transfer

Ensures client self sufficiency and big data capabilities

 Standard Roadmap
 Research
 Use Case Selection
 Product Selection
 Education & Training
 Services

Stampede

- Tutorials/Training
- BigInsights Quick Start
- Expert/BVA

Business Value



Time to insights



Business Analytics Initiative

March 2014





Business Analytics Initiative

- High Level Strategy Presentation (March)
- Identification of Initial Focus Areas
 - DWH, ETL and Reporting
 - Next Best Action
- Initial Workshop for DWH, ETL & Reporting (April)
 - As-is architecture
 - Challenges & Requirements





HALKBANK – IT CHALLENGES

- Data Latency (T-1, T-2..), Data Redundancy
- Management of two DWH overhead Development & Maintainance
- Huge growth trend of reporting system
 - 80K, 100K, 140K query@PX13 (in last 3 months)
 - ~ 900 batches/day in total. 1-1.5TB data transfer
 - Halkbank will build Loan & Deposit DM Management System (additional load)
- DWH-PX13 resources are limited
 - New reports are on PDA
 - ~ 5000 reports, last ~2 years moved
 - %70 of cpu usage offloaded onto PDA
 - %70 İstihbarat, The rest is for BO Reports & CDC Replication
 - Application code changes required to move to PDA
- ETL process problems
 - End-to-end batch /ETL process, management & monitoring problems
 - BO, ETL, SAS, SPSS, PL1/JCL integration diffuculties
- Building test environments for BO & SPSS & improving ADLC, Security and

Governance - ****Audit



HALKBANK – IT REQUIREMENTS

- Architectural Review, Evaluation of New Architecture
 - Eliminate multiple DWH Systems Maintenance Effort
 - End to End Architecture & Flow Review
 - Reference Architecture Best Practices
 - Optimization of Information Life Cycle Management
 - Proper Test Environment Setup
 - Addressing Governance Security Issues
- Real Time Data Infrastructure
 - Minimize bulk data transfer
 - Accessing Real Time Data from Business, Applications & Reports
- Reporting System Enhancements
 - Business Object BO Reporting Efforts
 - Improve Time to Market, Simplification, Access from End User
 - Building BO Report Inventory
 - Elimination of Duplicate Reports
 - Approx. 5000 active reports. (10K?) / In-house inventory system, not effective

Three types of analytics programs deliver productivity or growth benefits

Areas of Benefits

Analytic Solution Domains

Benefits

Infrastructure **Productivity**

Take-out cost and improve efficiency

Information **Management** Foundation (IMF)

- Eliminate multiple data silos
- Enhance level of data trust and accuracy
- Broaden organizational access to key data
- Reduce repository footprints and data model objects
- Reduce number of data integration programs and tools
- Rationalize disparate reporting tools and maintenance
- Reduce analysts data gathering time
- Reduce cost and complexity of storing large amounts of data

Business Productivity

Improve control. bottom line and stop losses

Risk, Fraud and **Finance** (RFF)

Supply Chain

- Lower Finance costs as a % of revenue
- Transactional activity cost reduction
- Headcount reduction
- Higher productivity of finance FTEs via IT enablement
- Maximize ROI tactical and strategic cash improvement opportunities

Increased customer lifetime value

Improved advertising reach and

- Planning and Forecasting
- Risk management

effectiveness

Reduced attrition

Improved demand visibility Better customer channel and management management

Focus on growth

Intelligent profitable growth

Customer, Marketing and Sales (CMS)

Human Capital

Increased sales

headcount

- Eliminated unproductive marketing spend
- Reduced non-productive customer contacts

 - More effective deployment of
- Optimized sales activities

- More effective service agents



"New path to value" is a five-point approach to operationalizing analytics

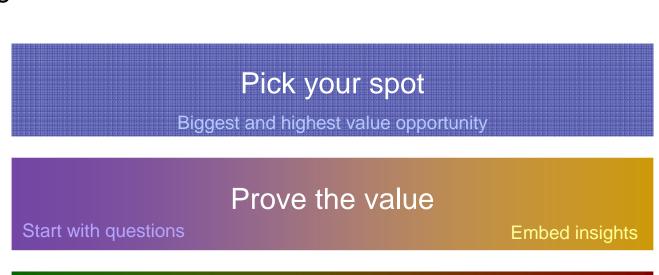


Embed insights to drive actions and deliver value





How to get started



Roll it out over time Add capabilities Information agenda



IBM BAO Jumpstart Methodology



BAO Jumpstart 2.0 addresses the top obstacle to analytics adoption – lack of understanding how to use analytics to improve the business



Recommendation 1:
Focus on the biggest and
highest value opportunities



Recommendation 2: Within each opportunity, start with questions, not data



Recommendation 3: Embed insights to drive actions and deliver value



Recommendation 4:
Keep existing capabilities
while adding new ones



54

Recommendation 5:
Use an information agenda to plan for the future

BAO Jumpstart 2.0

Identify highest value opportunities to be addressed with analytics

Assess ability to operationalize analytics opportunities

Determine actionable next steps and the fastest path to value

Potential Follow-on Activities

- Analytics Proof-of-Value
- PM Diagnostic
- Analytics Center of Excellence
- Information Management Foundation Phase 0
- Targeted projects identified during BAO Jumpstart

© 2014 IBM Corporation