Get Ready for Big Data with IBM System z

Product strategy

SHARE 2012, Anaheim

Mark Simmonds

System z Information Management – Product Marketing

Disclaimer

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.   Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Complete your sessions evaluation online at SHARE.org/AnaheimEval

Agenda

- **Big Data – Why now?**
- **Start with System z**
- **Making Big Data a reality**
  - **Business analytics and Data Warehousing**
  - **Data Management**
  - **Information Governance**
- **Call to Action**

What is Big data?

Ability to Process, Integrate, Understand data from anywhere.

The challenges :

How and which data to leverage for better business outcomes

Manage and control the data you are responsible for

12+ TBs of tweet data every day

Zettabytes of data in databases

30 billion RFID tags today (1.3B in 2005)

4.6 billion camera phones world wide

? TBs of data every day

25+ TBs of log data every day

100s of millions of GPS enabled devices sold annually

76 million smart meters in 2009… 200M by 2014

2+ billion people on the Web by end 2011

Variety

Volume

Velocity

Why Big Data?

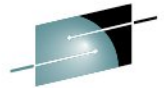**Reduce risk**
- Deeper understanding of market opportunities and threats

Lower cost
- Deliver goods and services smarter / more efficiently

Increase revenue opportunities
- Help predict customers' / your next move

Lower the costs and risks of making more money

New era of computing requires

Information from Everywhere

Radical Flexibility

Extreme Scalability
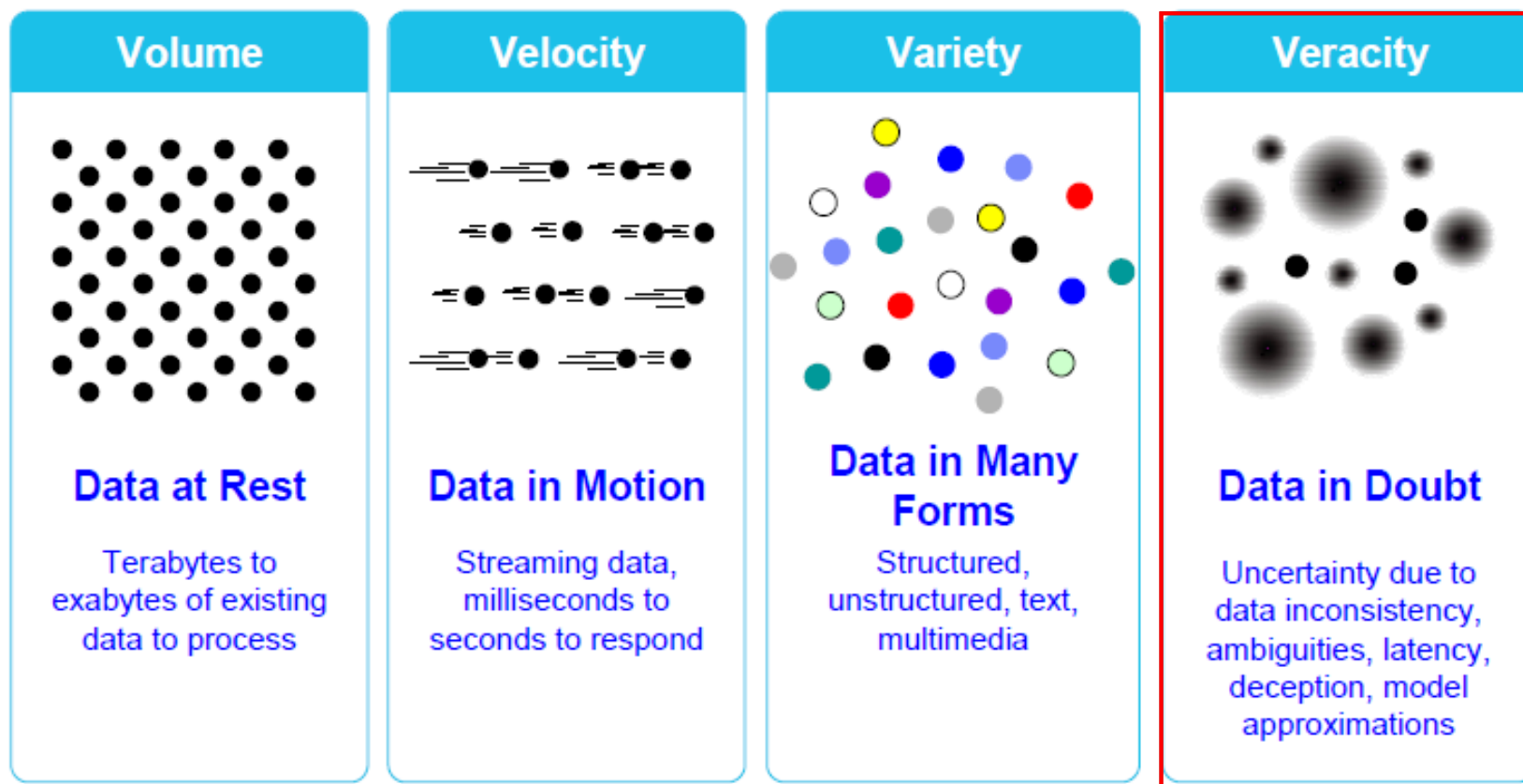
Volume

12 terabytes
of Tweets created daily

Velocity

5 million
trade events per second

Variety

100's video feeds

from surveillance cameras

SHARE in Anaheim

2012

The fourth dimension of Big Data: Veracity – handling data in doubt

| Volume	Velocity	Variety	Veracity
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia	Uncertainty due to data inconsistency, ambiguities, latency, deception, model approximations

Constructing context by combining data from many sources minimized uncertainty
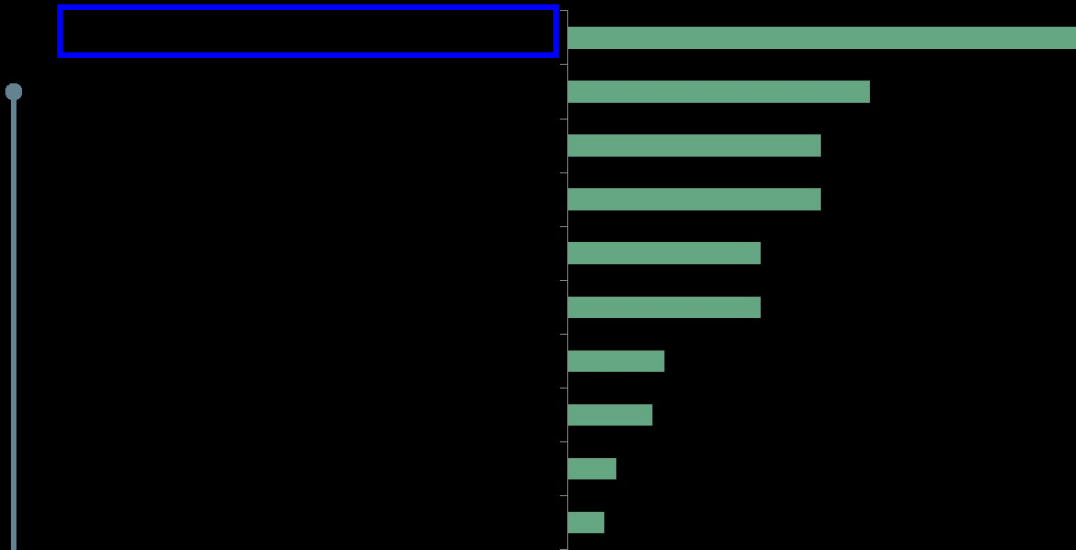
Agenda

- **Big Data – Why now?**
- ***Start with System z***
- **Use Case - Making Big Data a reality**
  - **Business analytics and Data Warehousing**
  - **Data Management**
  - **Information Governance**
- **Call to Action**

Where to start - Reality check…
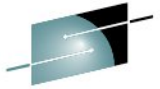
- What data can you manage / analyze today?

Big data: across diverse subject domains

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

System z Data – core to Big Data projects
THE platform for Enterprise Mission Critical transaction processing and data

DB2: Top 66 banks in the world   *DB2: 9 of the top 10 global life/health insurance providers*   *DB2: 24 of the top 25 US retailers*

UPS runs DB2 for z/OS to support the world's largest known peak database workload - 1.1 Billion SQL statements per hour!

24x7 ATM Deposits & Withdrawals

Runs the world's stock exchanges & banking networks

Reserves airline seats

Tracks the world's packages

8 of every 10 of the largest retail banks in Australia, Germany, Japan, and the United States use IMS for their core banking

$3 trillion/day transferred through IMS by one customer   *95% of top Fortune 1000 companies use IMS*   *Over 15 billion GBs of production data in IMS…*

System z Platform – Lowering cost and risks

Highest availability on the planet

- Continuous availability during trading periods
- Non-disruptive upgrades of HW, z/OS, and subsystems, including DB2
- Built-in system redundancy (memory, cooling, power…)
- Comprehensive multi-site disaster recovery

System-level mixed workload management with full resource utilization

- System-level WLM manages all resources
- 100% utilization, 24 hours a day
- Most cost effective SLA

Real-world scalability with performance

- Superior in the industry
- Scale out with absolute access during business trading periods

Unmatched end-to-end security

- From logon through data encryption
- Never been hacked

The most cost effective platform to manage and maintain

With it's unique architecture and deep integration with System z, DB2 for z/OS is the undisputed leader in total system availability, scalability, security and reliability.

The Ultimate Consolidation Platform

Data Mart **Data Mart** **Data Mart** **Data Mart**

Data Mart Consolidation
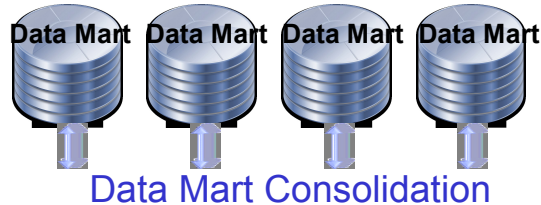
System z PR/SM
*Recognized leader in mixed
[vir]tualization and workload isolation*

Transaction Systems
(OLTP)

Data Warehousing
Business Intelligence
Predictive Analytics

z/OS:
*Recognized leader in
mixed workloads with
security, availability
and recoverability*

Netezza:
*Recognized leader in
cost-effective high
speed deep analytics*

Together:
*Destroying the myth that transactional and decision support
workloads have to be on separate platforms*

Bringing it all together

- *Better Business Response*

- *Reduced Costs*

- *More Available*

- *More Secure*

- *Reduced Data Movement*

- *Reduced Data Latency*

- *Reduced Complexity*

- *Reduced Resources*

SHARE
in Anaheim
2012

Majority of today's analytics based on relational / "Structured" Data

- Analytics and decision engines reside where the DWH / transaction data is

- "Noise" (veractity) surrounds the core business data

  - Social Media, emails, docs, telemetry, voice, video, content

- What data are you prepared to **_TRUST?_**

- Where do you put your trusted Data?

"Circle of trust"

- Data Warehouse
- Integration
- Business Analytics
- DB2 for z/OS IMS
- Information Governance

Demand for _differently_ structured data to be seamlessly integrated, to augment analytics / decisions

- Analytics and decision engines reside where the DWH / transaction data is

- "Noise" (veractity) surrounds the core business data

  - Social Media, emails, docs, telemetry, voice, video, content

- Multi-source streams enhance "corporate knowledge"

  - Lower risk and cost

  - Increased profitability

  "Circle of trust" widens

Data Warehouse

Integration

Business Analytics

DB2 for z/OS IMS

Information Governance

ARE in Anaheim
2012

Agenda

- **Big Data – Why now?**
- **Start with System z**
- *Use Case - Making Big Data a reality*
  - **Business analytics and Data Warehousing**
  - **Data Management**
  - **Information Governance**
- **Call to Action**

Imagine the Possibilities of Analyzing *All* *available data*
Solve key issues completely by analyzing "big" and OLTP data

Faster, More Accurate, Less Expensive

Real-time Traffic Flow Optimization

Precise fraud & risk detection

Understand and act on customer sentiment
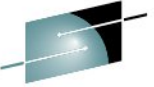
Accurate and timely threat detection

Predict and act on intent to purchase

Low-latency network analysis

Fraud Detection – Claiming disability allowance.

"Unable to work"

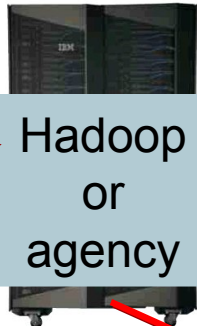**Work Status**

"Dude – awesome vacation"

Facebook Post

Investigation

Deterrent for fraudsters - Cost Savings for the business

Make payment or investigate

zEnterprise

DB2 Accelator

Refined Search parameters from OLTP environment

Data from Social Media sites analyzed with Text analytics

| Data Warehouse |
| Integration |
| Business Analytics |
| DB2 for z/OS IMS |
| Information Governance |

Hadoop or agency

Result Set for further processing
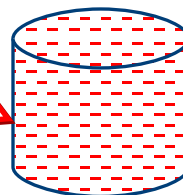
Data Warehouse + modeling applications

Result set uploaded or directly imported into OLTAP DBMS

Enterprise Integration and Governance – the key to success of incorporating Big Data

- **Information Integration**
  - Insights from Big Data must be incorporated into the warehouse and analytics/decision engines

- **Information Governance**
  - Companies need to govern what comes in, and the insights that come out
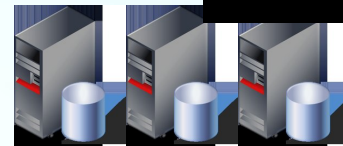
Data Warehouse

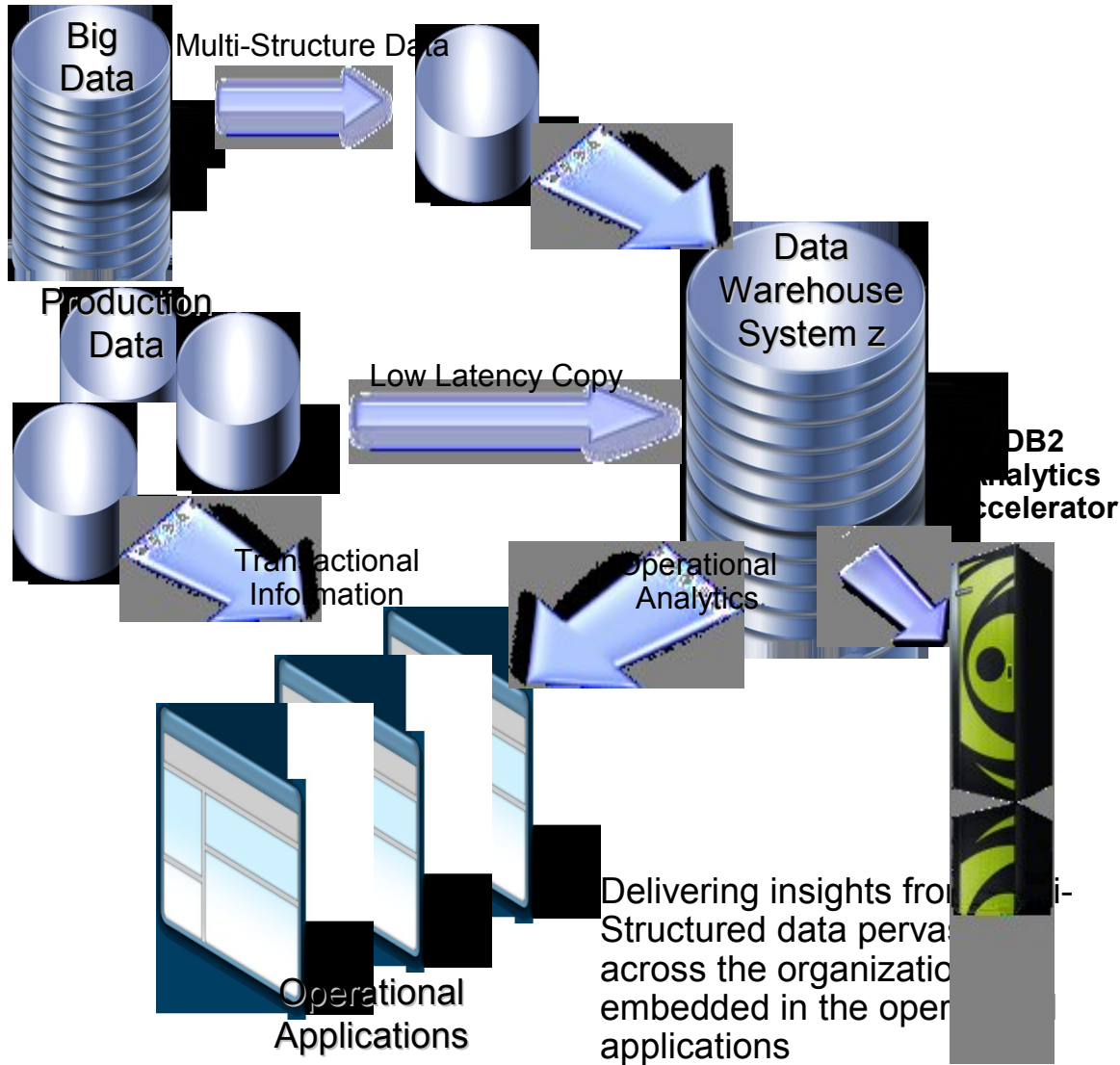Big Data Platform

Enterprise Integration

Traditional Sources

New Sources

Agenda

- **Big Data – Why now?**
- **Start with System z**
- **Making Big Data a reality**
  - *Business analytics and Data Warehousing*
  - **Data Management**
  - **Information Governance**
- **Call to Action**

Business Analytics and Data Warehouse - bringing it all together for better business outcomes.

Big Data

Multi-Structure Data

Production Data

Low Latency Copy

Data Warehouse System z

DB2 Analytics Accelerator

Transactional Information

Operational Analytics

Operational Applications

Delivering insights from multi-Structured data pervasive across the organization embedded in the operational applications

Benefits

- Deliver new insights from multi-structured data such as sensor, social, and clickstream to make fact-based decisions

- Combine multi-structured data with historical data warehouse information to increase understanding

- Provide analytic information at the point of decision enabling fact-based decisions

- Pervasively enable decision makers and other end users across the organization

- Accelerate long running DB2 for z/OS queries from minutes to seconds for greater business value with Analytics Accelerator.
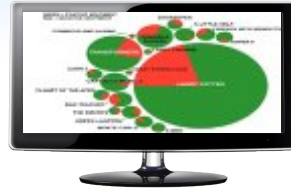
New era of analytic applications – finer grained insights

Advanced Analytic Applications

- Customer churn
- Risk management
- ...

- Location-based marketing
- Smart meter analytics
- ...

## Big Data Platform
Process and analyze any type of data

Accelerators

- Analyze data in motion
  - Non-relational data analytics
- MapReduce / noSQL
  - Machine Learning
    - Text analytics

- Visualization and exploration
- Scalability for large data volumes
- Hardware-based query acceleration
- Stream computing

- Integrate and manage the full variety, velocity and volume of data

- Apply advanced analytics to information in its native form

- Visualize all available data for ad-hoc analysis

- Development environment for building new analytic applications

- Workload optimization and scheduling

- Security and Governance

Analytics-driven Organizations Can…

Identify Risk
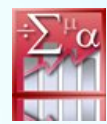
…and immediately control it

Swiss Re
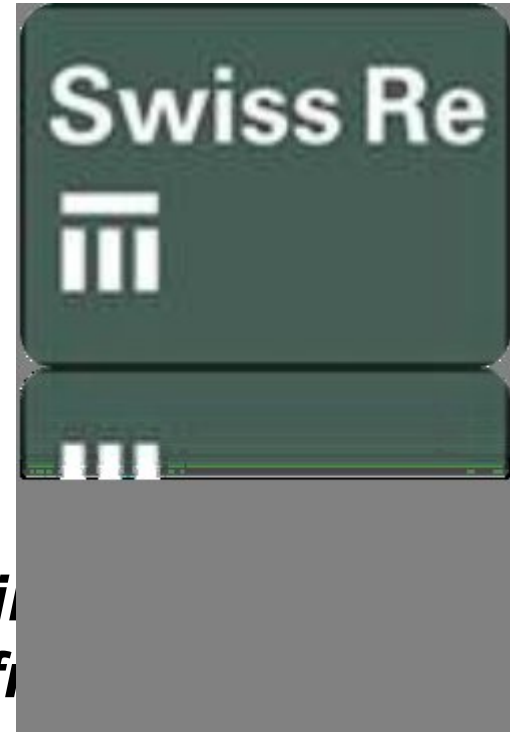
- *Insights i... policies fr... insurance companies*
- *Getting their reports as much as 70 percent faster*

SHARE in Anaheim 2012

Analytics for V⁴ – Built-for-Purpose, Built-for-Variety

- Leading analytics from IBM Research
- Built-for-purpose to analyze data in its native format

Text   Statistics

Image & Video   Mining

Acoustic   Predictive
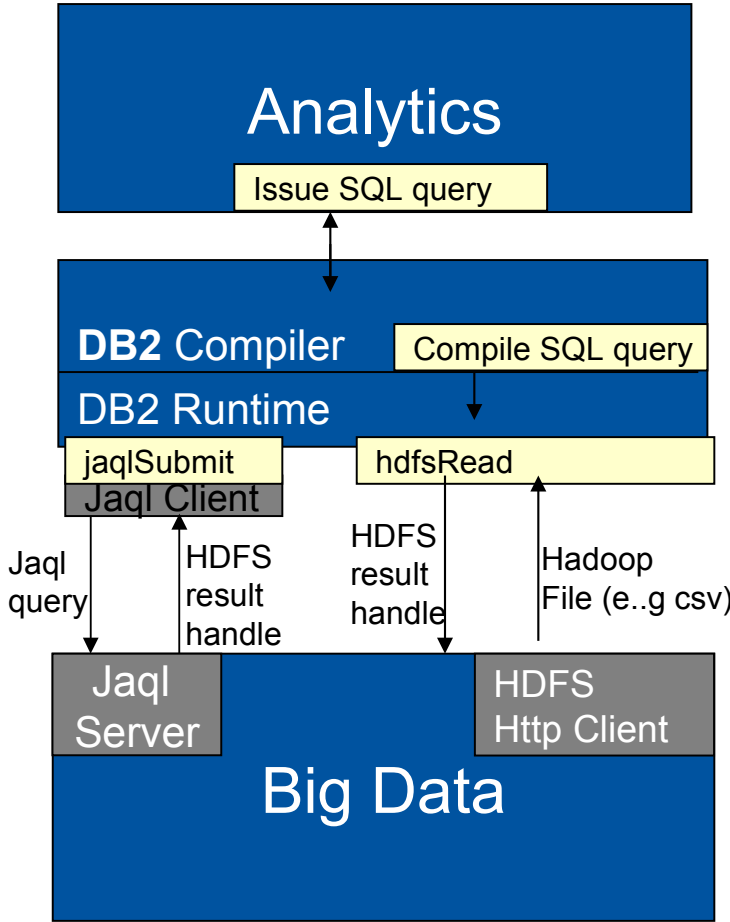
Financial   Geospatial

Times Series   Mathematical

IBM Differentiator – significant research investment in analytics; designed for use with Big Data.
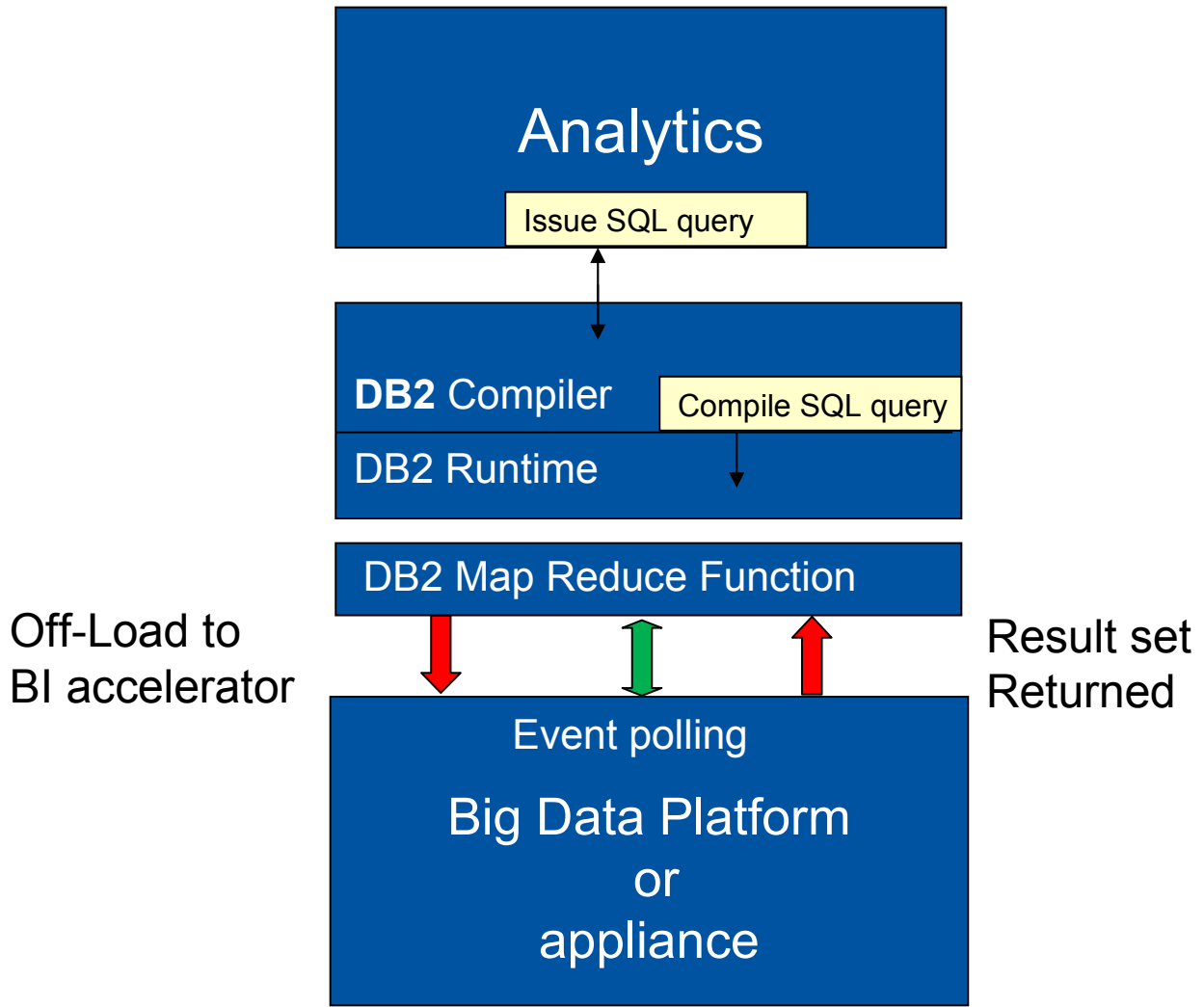
Agenda

- **Big Data – Why now?**
- **Start with System z**
- **Making Big Data a reality**
  - **Business analytics and Data Warehousing**
  - ***Data Management***
  - **Information Governance**
- **Call to Action**

Data management - Connecting Big Data and DB2: Phase 1

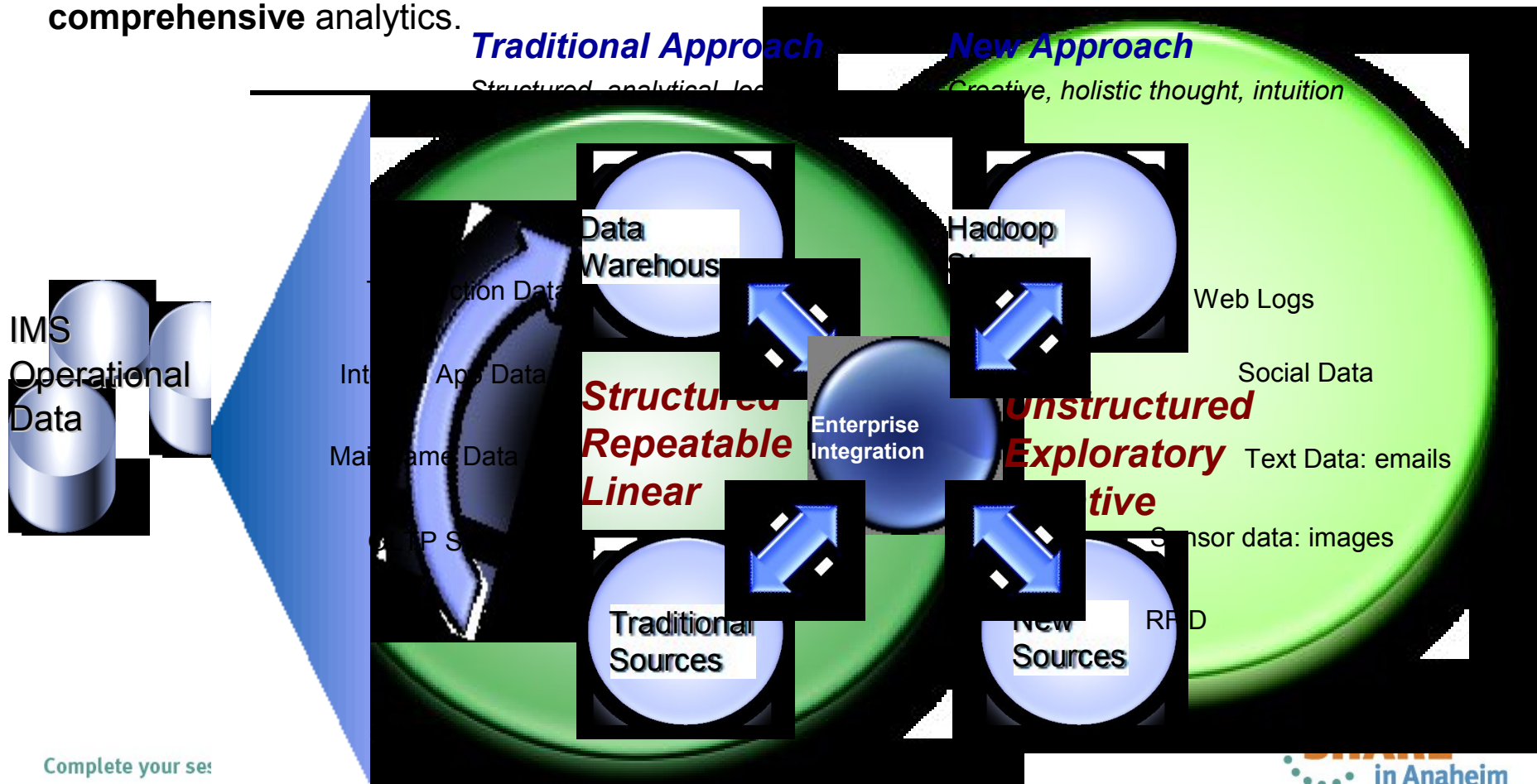# *Connecting Big Data and DB2 for z/OS futures :Phase 2*

Analytics

Issue SQL query

DB2 Compiler

Compile SQL query

DB2 Runtime

DB2 Map Reduce Function

Off-Load to
BI accelerator

Result set
Returned

Event polling

Big Data Platform
or
appliance

Direct Connect for Big Data and DB2 for z/OS Bulk Data movement

Read Input, Splits, Convert, Partition

Unstructured non z Data sources

Load in DB2 Database, map to partitions or Flat file output

Big Data

Map	
Map	
Map	Shuffle/Sort

Reduce
Reduce
Reduce

Data Warehouse
Integration
Business Analytics
DB2 for z/OS IMS
Information Governance

zDB2 C...

HDFS or GPFS File System

System z Data Sources, Logs, Archives, etc.
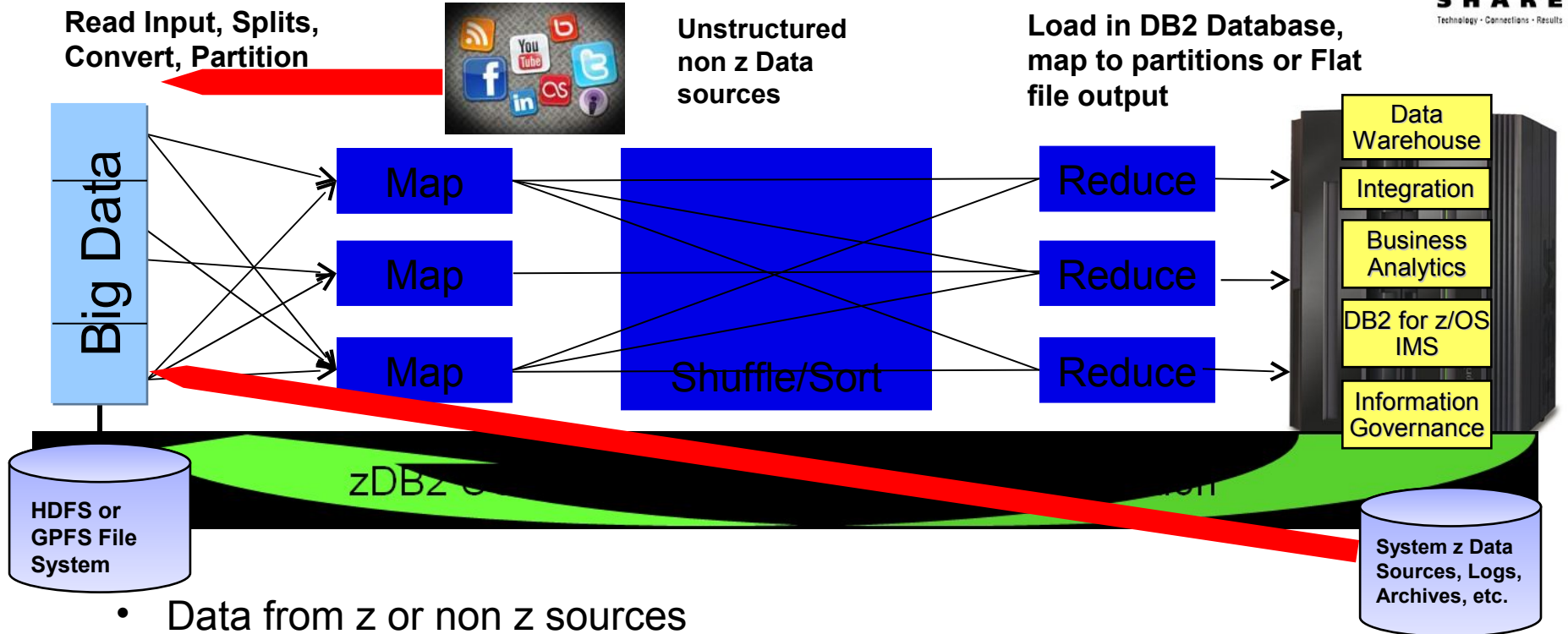
- Data from z or non z sources
- Direct Loading from Hadoop into DB2 for z/OS Partitions
- Can be output files for later load
- Hadoop needs DB2 Catalog information for data format and to match partitions to threads.
- BigInsights to perform data conversion.
- High Speed Data movement off network via zDDB feature of DS8800 Requires both sides to use zDDB feature API.
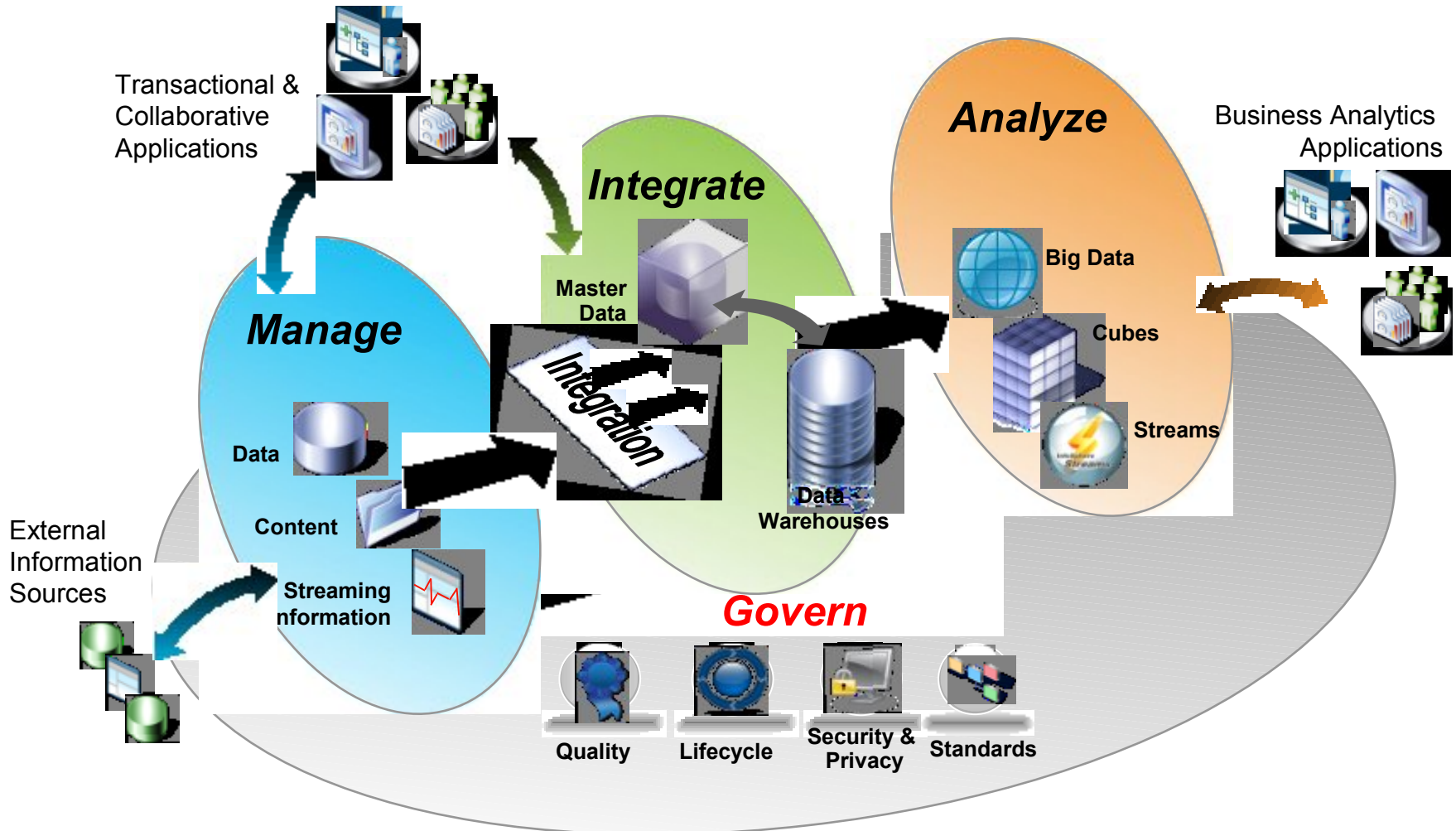
IMS and Big Data

- IMS applications for Big Data Analytics include Finance, Manufacturing, Telecom, Retail, Log Analysis, Fraud and Risk.
- IMS manages a high percentage of the world's operational mission-critical data.
- Integrate IMS structured data with new forms of unstructured data for more **comprehensive** analytics.

Traditional Approach

Structured, analytical, le...

New Approach

Creative, holistic thought, intuition

IMS Operational Data

Transaction Data

Int... App. Data

Mai...ame Data

...RP S...

Data Warehous...

Hadoop St...

Web Logs

Social Data

Structured Repeatable Linear

Enterprise Integration

Unstructured Exploratory ...tive

Text Data: emails

Sensor data: images

Traditional Sources
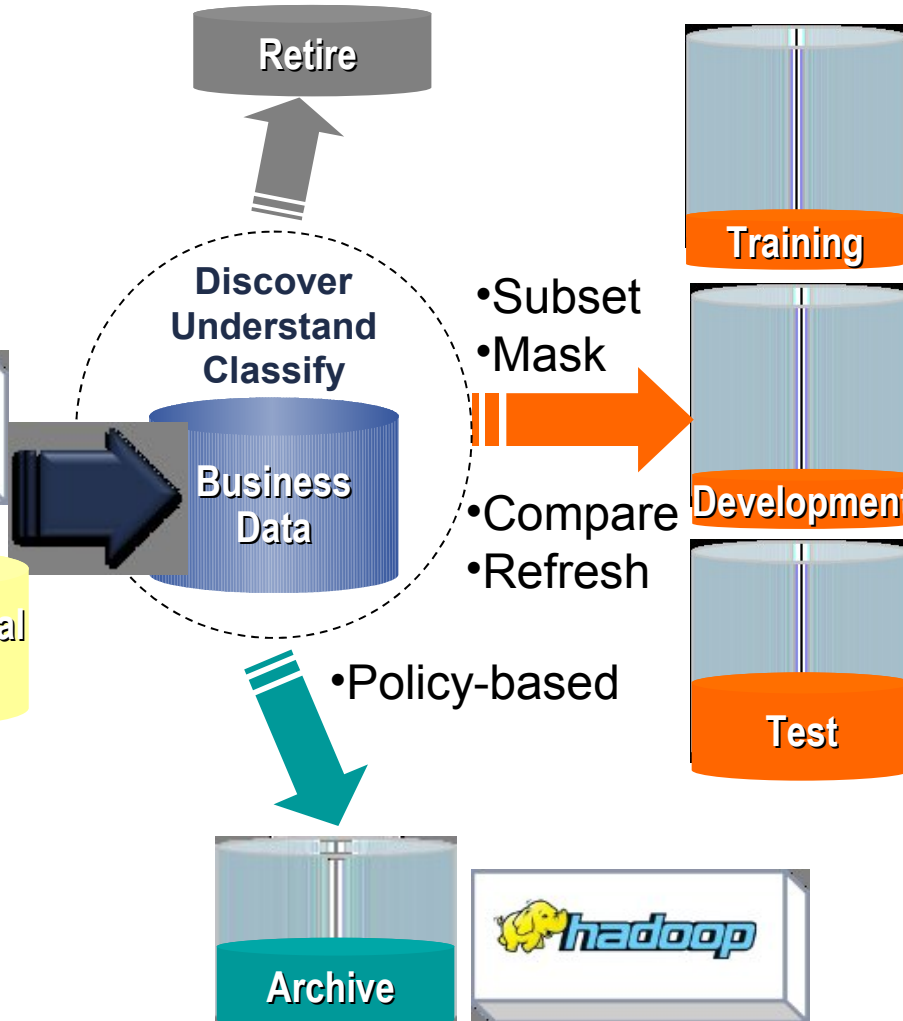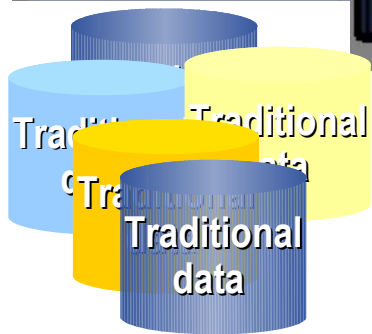
New Sources

RF...D

Agenda

- **Big Data – Why now?**
- **Start with System z**
- **Making Big Data a reality**
  - **Business analytics and Data Warehousing**
  - **Data Management**
  - ***Information Governance***
- **Call to Action**

Information Governance

Information Lifecycle Management (and Big Data)
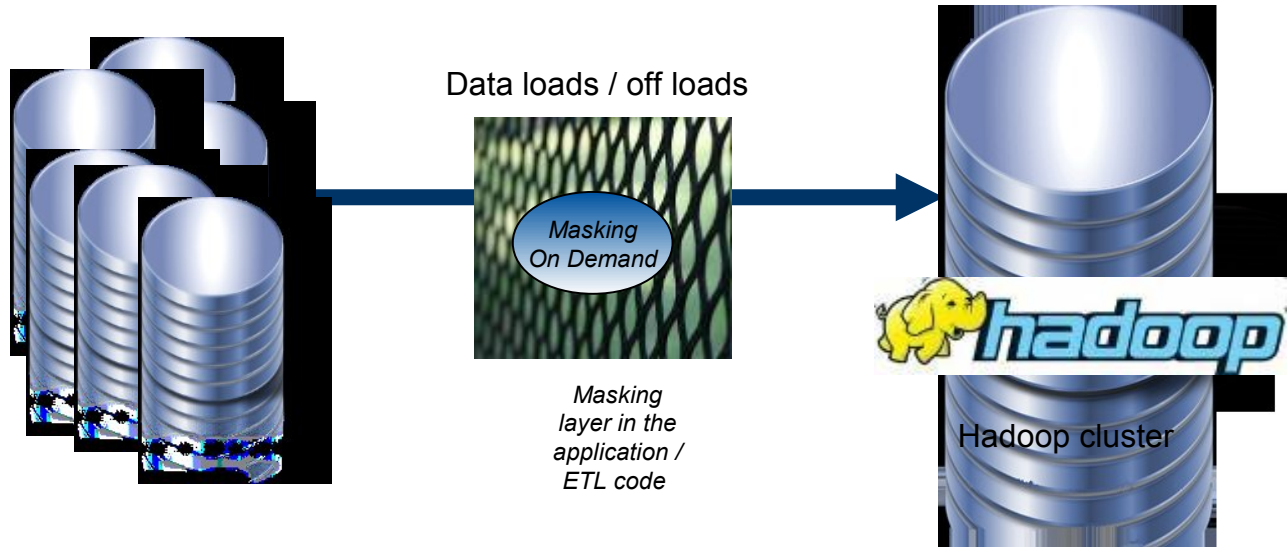
- **Machine Data**
- **Social Data**
- **Sensor Data**
- **…**

Retire

Discover Understand Classify

Business Data

Traditional data

•Subset
•Mask
•Compare
•Refresh

•Policy-based

Training

Development

Test

Archive

Data Discovery

Test Data Management

Data Masking

Data Growth Management

Application Retirement

in Anaheim

2012

Ensuring Data Privacy and Security

Data loads / off loads

Masking On Demand

Masking layer in the application / ETL code

Hadoop cluster

- **Usecase:**
  - Large scale analytics requires data from traditional sources to be combined with unstructured texual data to draw inferences
  - The analytics is predominantly trend analysis and individual data values are not extremely relevant.
- **Compliance risks:**
  - Data that is protected with masking and encryption in traditional sources is moved to the hadoop clusters exposing the enterprise to data leaks and legal exposure
- **Solution:**
  - Existing Data Privacy solution - Masking on Demand functionality could provide real-time means to mask the data as its being loaded to ensure compliance

Big Data and Data Archiving

- **Usecase:**
  - Customers expect to move data that are not actively using for day to day operation but would like to
    - A) Keep the data for compliance reasons and
    - B) Would like to use the data in big data analytics practices
- **Compliance concern:**
  - Ensuring compliance with industry, government and business regulations while drawing out key insights from the data during analytics.
- **Solution:**

  Extend current capabilities to create archives for storage on the Hadoop platform with active usage characteristics – aka-Queryable archives. These archives can now be used as part of big data's analytics while ensuring governance expectations of the enterprise are met.

Connect any type of data through optimized connectors

Big Data Platform

Structured

Differently. Structured / Unstructured

Streaming

Queryable Archives

# *Data Lifecycle Management (for Big Data)*

Customers asking for:

1. Masking on demand while loading to bigdata environments

2. Queryable archives on hadoop for analytics

3. Test data management for big data projects

4. Masking and redaction of unstructured content in big data.

5. Archiving of the big data environments to provide a point in time snapshot of the analytics process

Advantages

- Fuller governance functionality available for Big Data.
- Uniform masking technology across the enterprise.
- Ensure regulatory compliance of big data.
- Part of the Full-Stack IBM Solution.

Protecting all data across multiple platforms
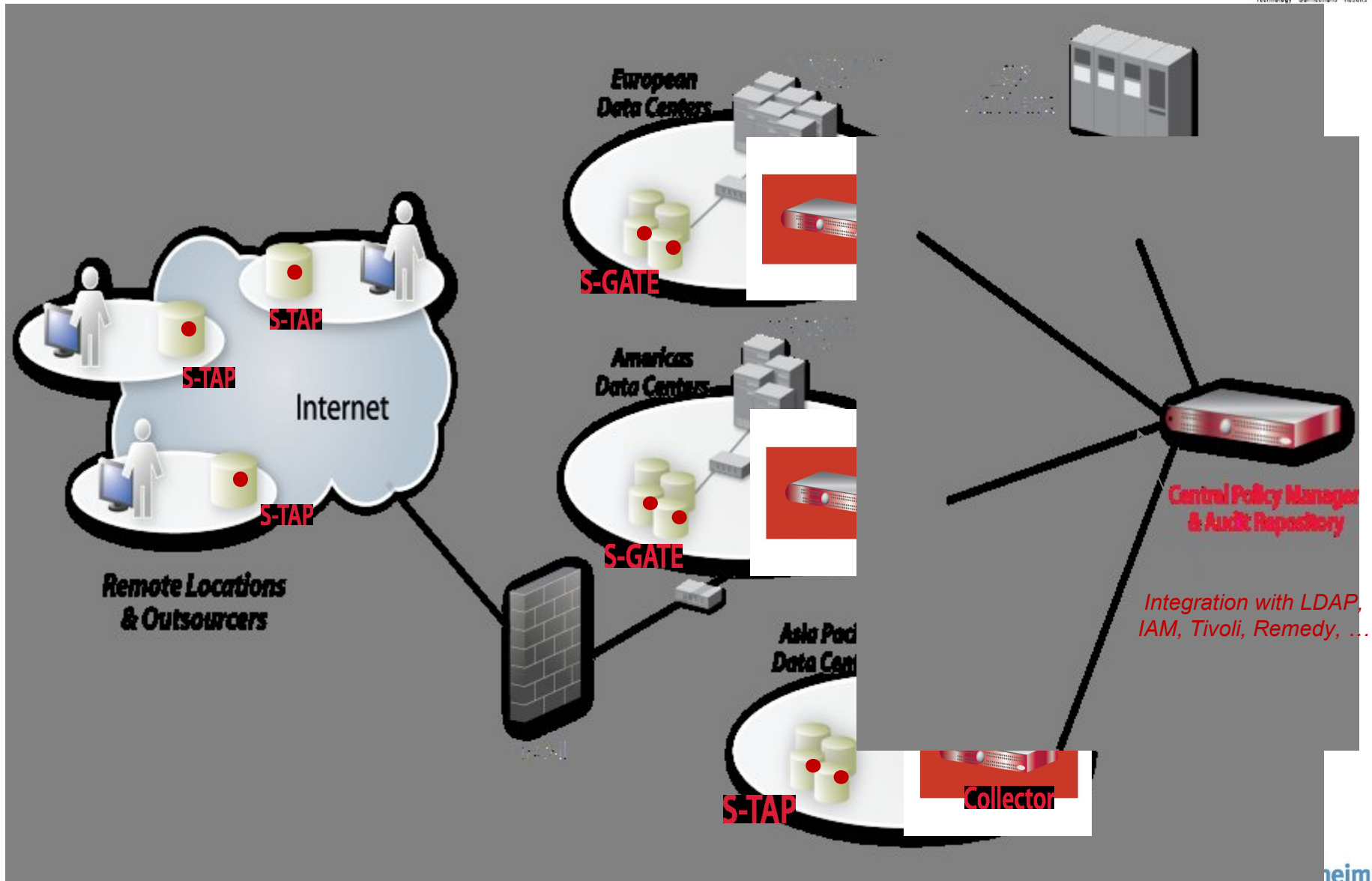
Customer requirements for Hadoop support

- Hardware or software appliances
  - Securely stores audit data collected by probes
  - Provides analytics, reporting & compliance workflow automation
    - Offloads audit data processing from mainframe
  - Integrated as part of the enterprise architecture
    - Centralized, cross-platform audit repository for enterprise-wide analytics and compliance reporting across System z & distributed environments
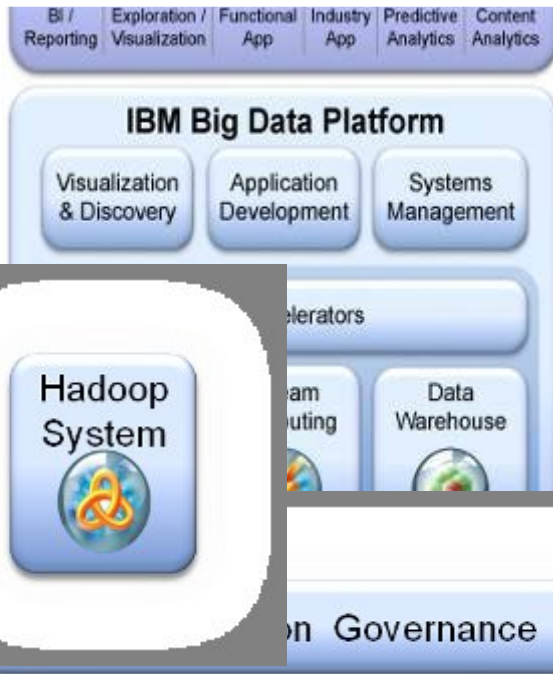
S-TAP
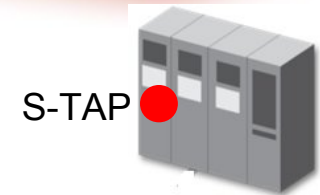
"We want Hadoop Activity Monitoring"
Monitor and Audit Hadoop activity in real-time to support compliance requirements and protect data

- Real time activity monitoring of HDFS and HBASE data sources
- Automated compliance controls
- Fully integrated with existing solution for database activity monitoring
- View Hadoop systems with other data sources

Monitoring of Hadoop

IBM Big Data Platform

BI / Reporting	Exploration / Visualization	Functional App	Industry App	Predictive Analytics	Content Analytics

Visualization & Discovery	Application Development	Systems Management

...elerators

Hadoop System

...eam ...uting | Data Warehouse

...n Governance

HDFS – Hadoop file system

▪Capture HDFS activity

    ▪User + IP address

    ▪Action: Open, Create, Delete, Rename, Set Owner, Set Permission, ListStatus, etc. etc.

    ▪Source and target of actions

    ▪Related Permissions

MapReduce – A processing framework

▪Capture MapReduce activity

    ▪Operation

    ▪Target

    ▪Permissions and description

Oozie – Hadoop workflow engine
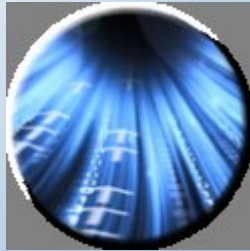
▪Capture Oozie activity

    ▪Jobid

    ▪Appname

    ▪Operations and parameters

▪And Stream them for policy evaluation and auditing in real time

Agenda

- **Big Data – Why now?**
- **Start with System z**
- **Making Big Data a reality**
  - **Business analytics and Data Warehousing**
  - **Data Management**
  - **Information Governance**
- **Call to Action**

For Big Data, IBM and System z is the clear choice

1. *Any type of data* — Manage and integrated any data types

2. *Derive better and faster insights* — Analytics built for variety, with most accurate analytic engines

3. *Enterprise Class* — Reliable, Available, Secure, Scalable

4. *Information Governance* — Comprehensive Information Governance technology, integrated with Big Data

Take Action Now!

Next steps:

- **For additional information including whitepapers and demos, please visit:**
  - Bringing Big Data to the Enterprise
  - Smarter Computing
  - Information Management System z
- **Education**
  - Free online education at bigdatauniversity.com
  - 20,000+ registered students
- **Further developments:**
  - SHARE Feb 2013
  - Future webcast and announcements

- **Develop your own big data strategy – Contact your local IBM sales representative to get started.**
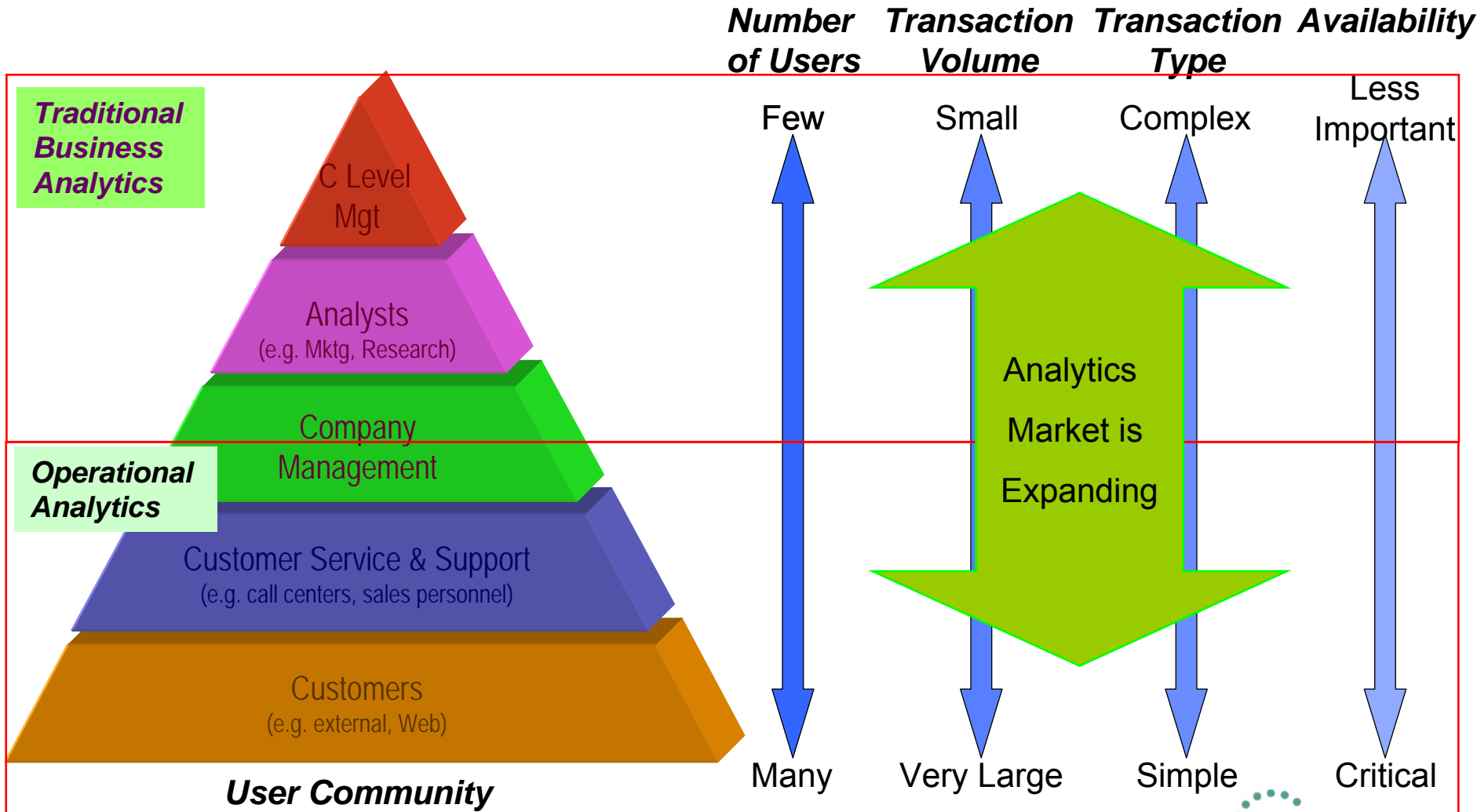
THINK
BIG

THINK
Z

IBM Business Analytics with the Big Data Platform … what's new and exciting!

Mike Biere

IBM Corporation

8/7/2012

Session Number: 11394

Current Analytics Market is expanding

	Number of Users	Transaction Volume	Transaction Type	Availability
Traditional Business Analytics	Few	Small	Complex	Less Important
C Level Mgt				
Analysts (e.g. Mktg, Research)				
Company Management		**Analytics Market is Expanding**		
Operational Analytics				
Customer Service & Support (e.g. call centers, sales personnel)				
Customers (e.g. external, Web)	Many	Very Large	Simple	Critical

User Community

Getting analytics out to frontline workers is more critical than ever…

- **More informed customer interaction = higher customer satisfaction**

- **Higher customer satisfaction = improved business performance**

- A dissatisfied consumer will tell between 9 and 15 people about their experience. About 13% of dissatisfied customers tell more than 20 people.
  Source: White House Office of Consumer Affairs, Washington, DC

- 86% of consumers quit doing business with a company because of a bad customer experience, up from 59% 4 years ago
  Source: Harris Interactive, Customer Experience Impact Report

- For every customer complaint, there are 26 other customers who have remained silent
  Source: Lee Resource Inc

- Happy customers who get their issue resolved tell about 4 to 6 people about their experience.
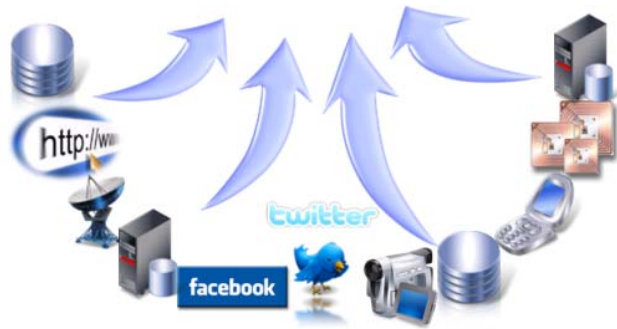 *Source: White House Office of Consumer Affairs, Washington, DC*

- Attracting a new customer costs 5 times as much as keeping an existing one
  Source: Lee Resource Inc.

What is "BIG DATA"?

All kinds of data
Large volumes
Valuable insight, but difficult to extract
Often extremely time sensitive

Where is this data coming from?

12 TB of tweets being created each day.

Every day, the New York Stock Exchange captures 1 TB of trade information.

5 Billion mobile phones in use in 2010. Only 12% were smartphones.

Every second of HD video generates > 2,000 times as many bytes as required to store a single page of text.

More than 30M networked sensor, growing at a rate >30% per year.

What is your business doing with it?

BigData is a platform and an eco-system that emerged from Web 2.0 companies like Google, Yahoo and Facebook who were responding to the need to collect and analyze data to service consumers over the world.

Internet response time is an expectation, there is no alternative to analyzing data "as-is" without lengthy cleaning, transforming or struc to data "as-accepted-by-our-analytical-tools".

Background

- **Before "big data"** Data Warehouse BI/BA queries were traditionally run on OLAP SQL data bases BUT SQL isn't suited to searches across log files and unstructured data.
- Furthermore the **complex OLAP queries that have been used by more traditional business intelligence applications aren't fast enough** to provide the sort of response time now needed with the inclusion of big data sources.
- This issue **led to the development of the MapReduce framework**. It allows for the breaking down of analysis tasks that would be ponderous with SQL into distributed computing tasks that can be defined with much less code.
- To make MapReduce more palatable, database vendors have started integrating it into their products. DB2 LUW offers a way to embed MapReduce calls within its database, as does Oracle, HP and Teradata. Biginsights has connectors that can be used for communication

- *Imperative that if system z wants to continue to position itself as a competitive BI/BA platform that we provide a methodology to more closely integrate Traditional OLAP with Big Data analytics and as such DB2 for z/OS, SQL with MapReduce and by definition the Hadoop file system If these MapReduce functions can be executed through SQL queries, they can be accessed through standard business intelligence tools.*

$\mathcal{H}$

SHARE
in Anaheim

2012

What can you do with big data?

Act on Deeper Customer Insight

- Social media customer sentiment analysis
- Promotion optimization
- Segmentation
- Customer profitability
- Click-stream analysis
- CDR processing
- Multi-channel interaction analysis
- Loyalty program analytics
- Churn prediction

Create Innovative New Products

- Social Media - Product/brand Sentiment analysis
- Brand strategy
- Market analysis
- RFID tracking & analysis
- Transaction analysis to create insight-based product/service offerings

Optimize your Operational Processes

- Smart Grid/meter management
- Distribution load forecasting
- Sales reporting
- Inventory & merchandising optimization
- Options trading
- ICU patient monitoring
- Disease surveillance
- Transportation network optimization
- Store performance
- Environmental analysis
- Experimental research

Prevent Fraud and Reduce Risk

- Multimodal surveillance
- Cyber security
- Fraud modeling & detection
- Risk modeling & management
- Regulatory reporting

Proactively Maintain your Assets

- Network analytics
- Asset management and predictive issue resolution
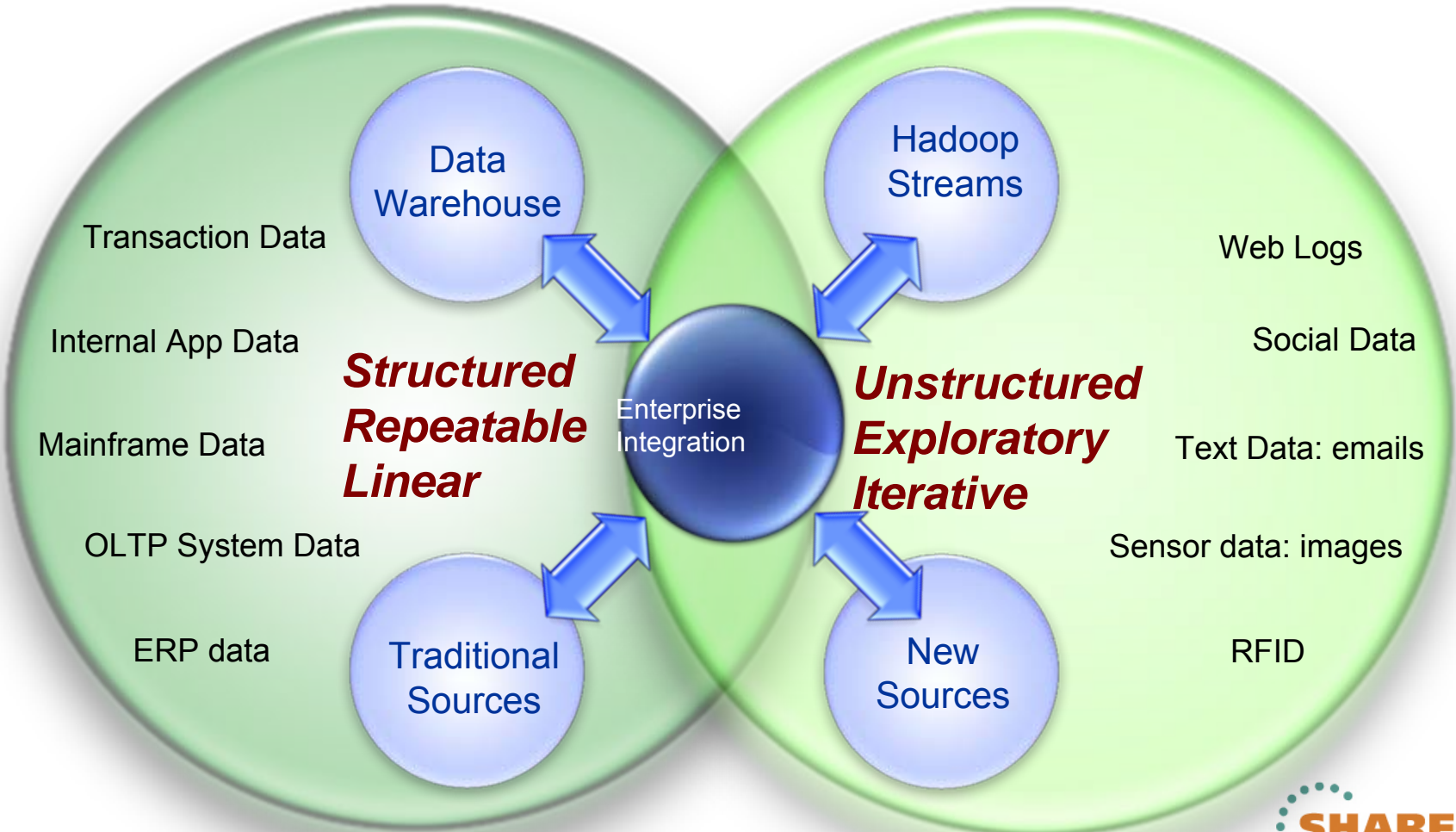- Website analytics
- IT log analysis

Complementary approaches … different use cases

Traditional Approach
Structured, analytical, logical

New Approach
Creative, holistic thought, intuition

Transaction Data

Internal App Data

Mainframe Data

OLTP System Data

ERP data

Data Warehouse

Traditional Sources

Structured Repeatable Linear

Enterprise Integration

Unstructured Exploratory Iterative

Hadoop Streams

New Sources

Web Logs

Social Data

Text Data: emails

Sensor data: images

RFID

SHARE
in Anaheim
2012

Pains Addressed by a Big Data Platform

- High cost of storing and analyzing data combined with data growing volumes

- Cost and performance of enterprise data warehouse - single DW cannot meet everyone's needs

- Inability to exploit new sources of data – need to explore, prove value, and extract it cost effectively

- Loss of fidelity and huge time/cost to convert unstructured data (video, audio, textual content) to structured format for analysis

- Inability to act and high cost of acting on data in real-time leads to lost opportunities

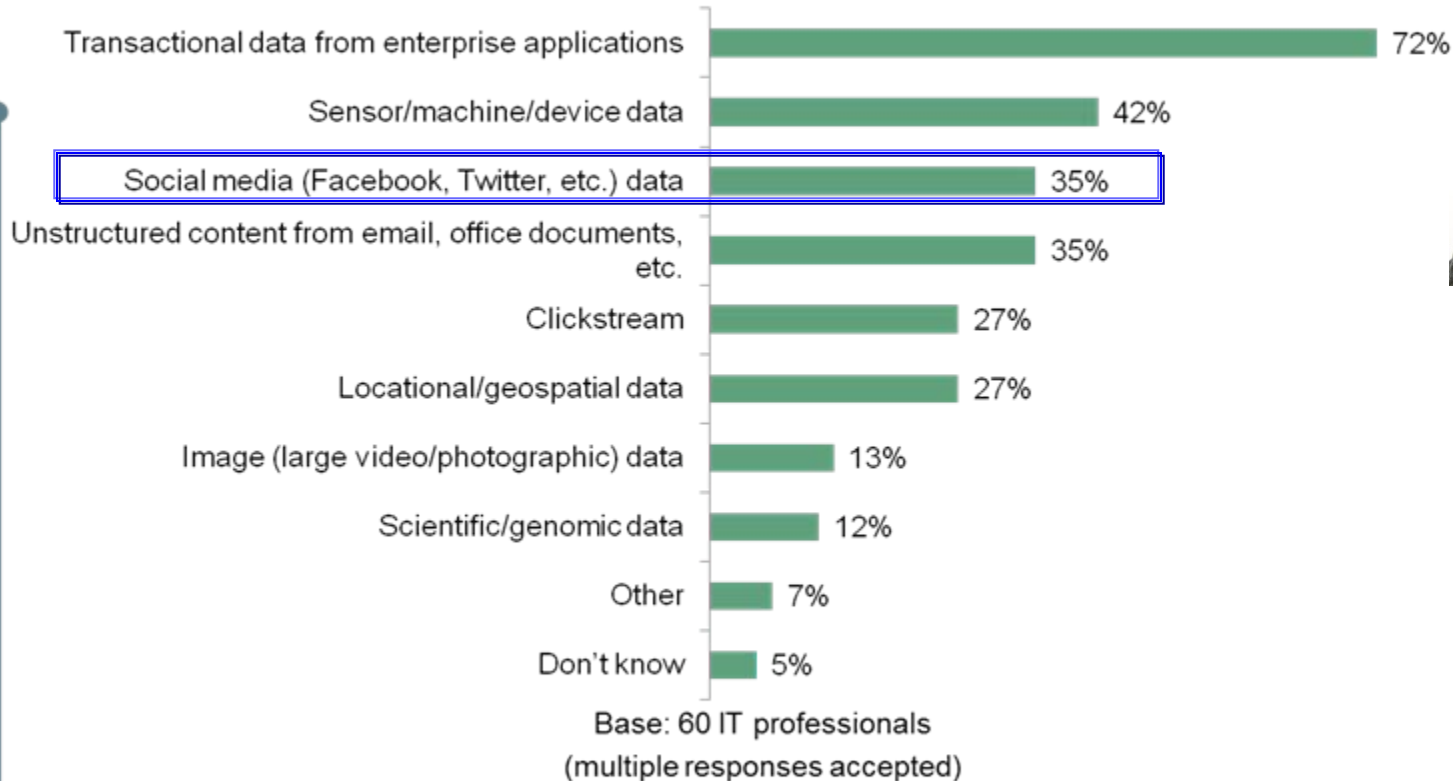- High cost to maintain data online when it could exist in an online archive – query-able archive
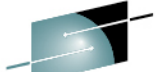
9

SHARE
in Anaheim
2012

Where to start - Reality check…

- What data can you manage / analyze today?

Big data: across diverse subject domains

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

Type	%
Transactional data from enterprise applications	72%
Sensor/machine/device data	42%
Social media (Facebook, Twitter, etc.) data	35%
Unstructured content from email, office documents, etc.	35%
Clickstream	27%
Locational/geospatial data	27%
Image (large video/photographic) data	13%
Scientific/genomic data	12%
Other	7%
Don't know	5%

Base: 60 IT professionals
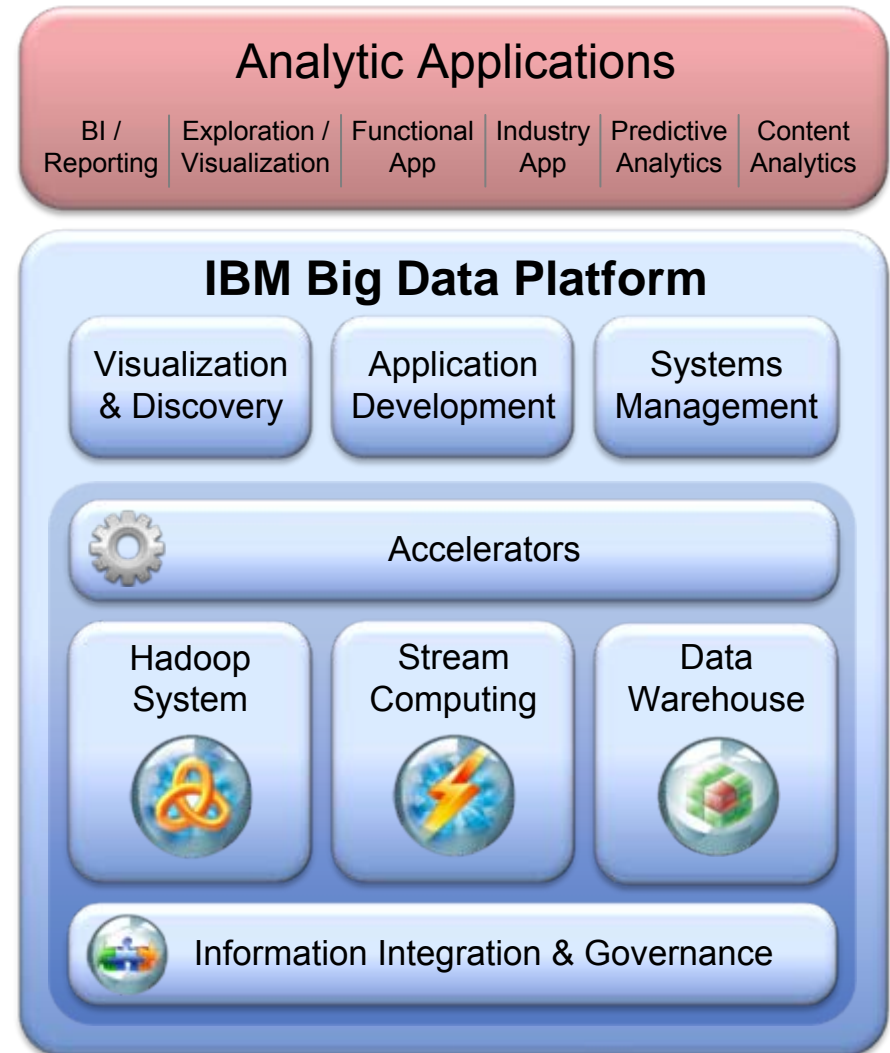(multiple responses accepted)

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

Source: June 2011 Global Big Data Online Survey

IBM Big Data Strategy: *move the analytics closer to the data*

New analytic applications drive the requirements for a big data platform

- Integrate and manage the full variety, velocity and volume of data
- Apply advanced analytics to information in its native form
- Visualize all available data for ad-hoc analysis
- Development environment for building new analytic applications
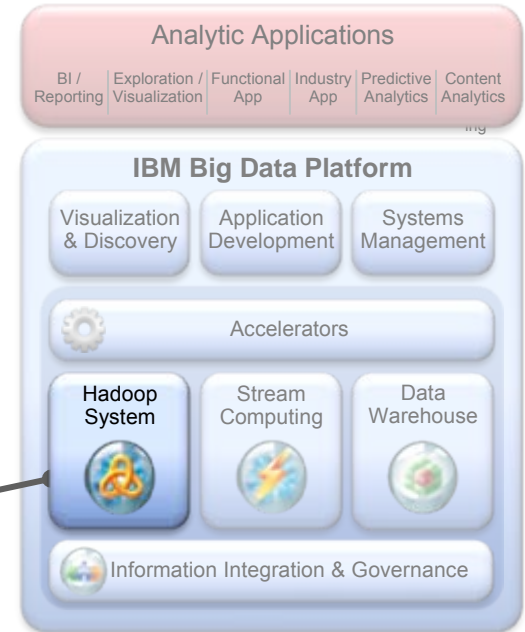- Workload optimization and scheduling
- Security and Governance

Analytic Applications

BI / Reporting	Exploration / Visualization	Functional App	Industry App	Predictive Analytics	Content Analytics

IBM Big Data Platform

Visualization & Discovery	Application Development	Systems Management

Accelerators

Hadoop System	Stream Computing	Data Warehouse

Information Integration & Governance

IBM's Big Data platform will support open source distributions

InfoSphere BigInsights

Enterprise capabilities

- Advanced Engines
- Indexing
- Connectors
- Workload Optimization
- Administration & Security

Open source based components

IBM certified Apache Hadoop *or* **cloudera** *or* *Other open source distributions (future)*

Analytic Applications

BI / Reporting | Exploration / Visualization | Functional App | Industry App | Predictive Analytics | Content Analytics

IBM Big Data Platform

Visualization & Discovery | Application Development | Systems Management

Accelerators

Hadoop System | Stream Computing | Data Warehouse

Information Integration & Governance

1. Hadoop (file system)
2. Map Reduce (parallel processing)
3. Hbase (database)
4. Oozie (workflow)
5. Zookeeper (distributed coordination)

The data challenge for business analytics development in IBM

- How to effectively transition from 'traditional' data sources to embrace the new formats?

- Can it be done efficiently or will it alter the basic design points of the solution?

- Which data sources are most critical to embrace today per our customer feedback?
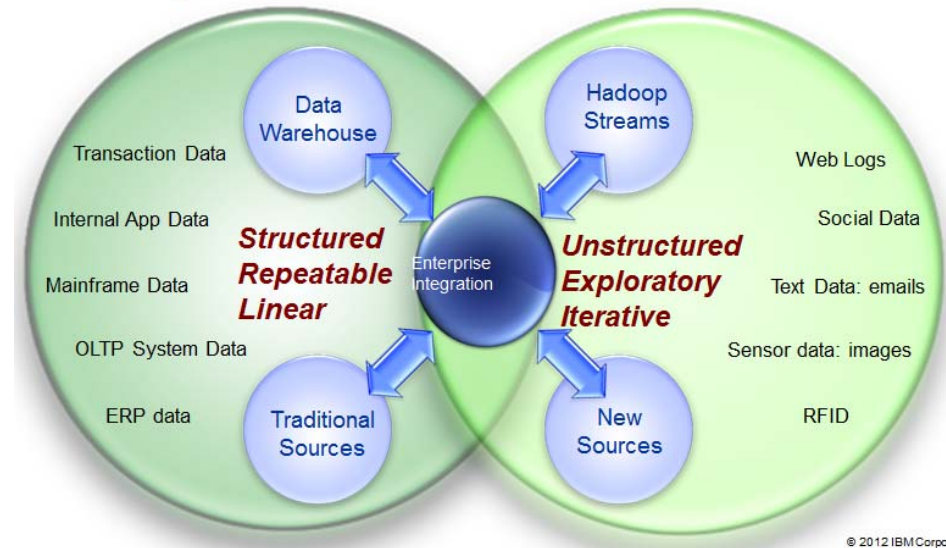
Big data: across diverse subject domains

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

Type	Percentage
Transactional data from enterprise applications	72%
Sensor/machine/device data	42%
Social media (Facebook, Twitter, etc.) data	35%
Unstructured content from email, office documents, etc.	35%
Clickstream	27%
Locational/geospatial data	27%
Image (large video/photographic) data	13%
Scientific/genomic data	12%
Other	7%
Don't know	5%

Base: 60 IT professionals
(multiple responses accepted)

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

Source: June 2011 Global Big Data Online Survey

13

in Anaheim
2012

IBM Business Analytics for Big Data

- **IBM Big Data Business Analytics portfolio**
  - **IBM Cognos for Linux on z/OS ***
  - **IBM Cognos for z/OS ***
  - **IBM Query Management Facility ***
  - IBM BigInsights
  - **IBM SPSS ***
  - IBM iLog
  - InfoSphere Streams
  - IBM Vivisimo
  - **IBM Collaboration and Mobile solutions (part of the Lotus brand) ***
    * These are my domain and I will cover today

- **BA product Suite will be supported with a set of solutions that run on System z as well as in a hybrid environment to support:**
  - Structured data analysis
  - Semi-structured data analysis (e.g. Lotus Notes applications)
  - Unstructured data (emails, video, and more)
  - Business analysis information delivery to a wide range of devices (e.g. PCs, Laptops, iPads, and other mobile devices)

- **Purpose: to apply modern BA solutions to the many data formats and sources within the Big Data infrastructure. An evolutionary process as not all tools support all data formats/sources within the Big data initiative.**

14

Continuous enhancements for IBM's Business Analytics solutions … for example

Configuring IBM Content Analytics to generate IBM Cognos BI reports

If you use IBM® Cognos® Business Intelligence (IBM Cognos BI) with IBM Content Analytics, users can generate reports about text mining results, such as results that include facet values, frequency counts, and correlation values, and open the reports with IBM Cognos BI.

The generated reports are saved on the IBM Cognos BI server and they are accessible from the Reports view of the text miner application. The generated reports have a hypertext link that allows users who analyze the report to return to the text miner application with the search conditions that were applied when the report was created. The user can continue to mine the content of the collection or, for example, open documents that match the search conditions.

Real time scoring with DB2 for z/OS and SPSS Modeler 15 – GA on 6/15/12

- **Delivers better, more profitable decisions, using the latest data, at the point of customer impact**

  - Enables more informed customer interaction

  - Improves fraud identification and prevention

- **With improved accuracy, speed and performance while reducing cost and complexity**

  - Improves accuracy by scoring new and relevant data directly within the OLTP application

  - Scales to large data volumes to improve accuracy of data models

  - Delivers the performance needed to meet and exceed SLAs of OLTP applications

  - Minimizes demand on network, HW, SW and resources

Customer Interaction

Data In

Real-Time Score/ Decision Out

Part of an extensive
Business Analytics solution
on System z!

SHARE
in Anaheim
2012

In-database Scoring of DB2 for z/OS represents a unique opportunity

Customer Interaction

Business Application - Operational/Transactional System

- OLTP System built on DB2 for z/OS
- On-line transaction processing (OLTP) systems, often referred to as transactional systems are designed to process small, quick, interactive workloads for which users expect fast response times.
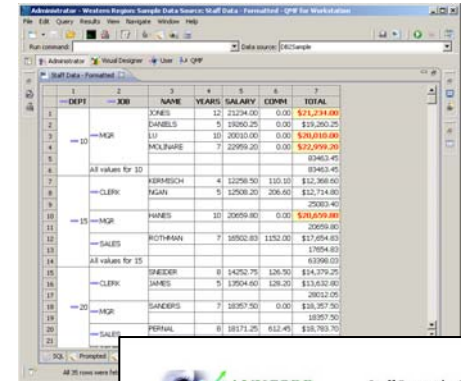
Data Repository

- Data Warehouse
- Operational Data Store
- Data Mart

*Data moved to data warehouse for long term analysis

SHARE in Anaheim

2012

QMF 10 – Benefits

- Business Analytics is a key area of interest to any enterprise - QMF now offers:
  - Graphical queries, reports
  - Dashboards and KPIs
  - Advanced analytic functions and analysis
  - JDBC access to any RDBMS
  - Multi-dimensional analysis
  - Operational Business Intelligence
  - Heterogeneous database access
  - Federated data
  - Mobile device support
  - Unstructured data project (in progress)

- Enhanced and easier installation, maintenance, administration.

- Enhanced ease of use and compatibility with existing QMF infrastructure, objects and workflows.

Evolution of Target User Base

Database Administrators
- QMF for TSO
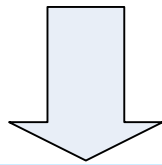
Technical Users
- QMF for Windows
  – Extends QMF to the desktop
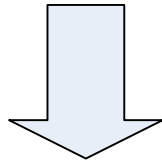
Data Analysts, IT
- QMF for Windows/WebSphere V8
  – Extends QMF to the web browser
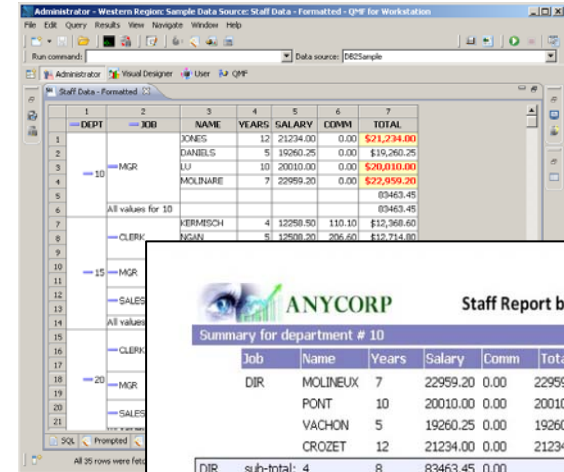 – Introduces graphical reporting
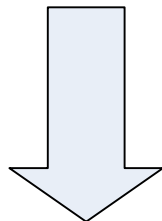
All Business Users
- QMF for Workstation/WebSphere V10
  – Intuitive visual solutions
  – Personalized, non-technical GUI
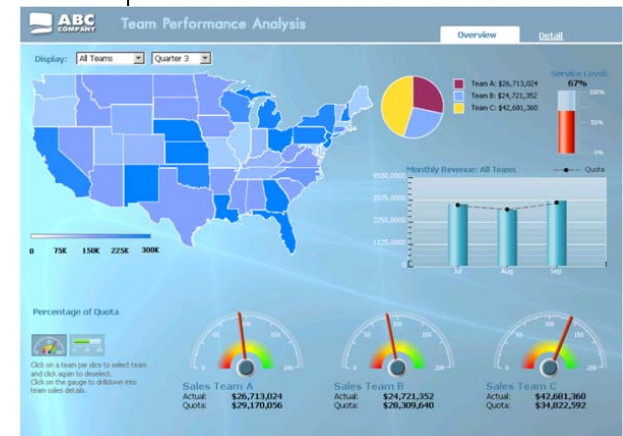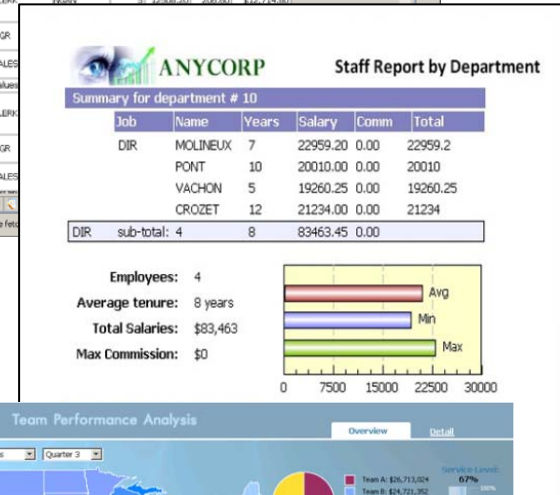 – Enhanced security
  – Support for OLAP & variety of DBs

All users across the enterprise

SHARE in Anaheim
2012

QMF for Workstation and WebSphere

- Ad-hoc & prepared queries
  - Create/edit/re-use QMF queries
  - Apply groupings, aggregations and conditional formats and more
  - One-click export to Excel, data files or DB table
  - One-click generation of reports from the data

- Tabular and graphical reports
  - Generate QMF tabular reports or graphical *visual reports*
  - Highly customizable page-based layouts
  - Draw data from any number of data sources

- QMF Dashboards
  - Customized, interactive data visualizations
  - Rapid authoring model

Taking QMF to the Web

- QMF V10's web application allows content to be readily distributed across the organization
  - Create, run and share queries on the Web
  - Run and share reports on the Web
  - Deploy dashboards to business end users as stand-alone solutions or embedded within web pages and web applications
- Ability to render content in
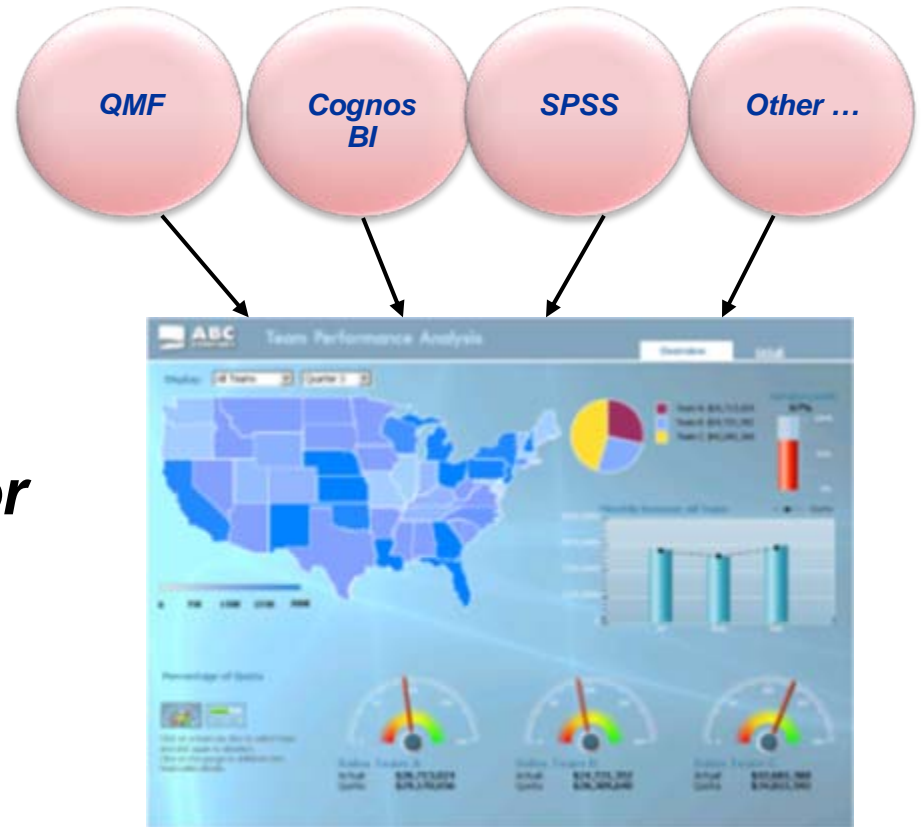 - Pure HTML & JavaScript
  - Adobe Flash Player
  - PDF

One dashboard or 20?

Should every solution provide its own dashboard and analytics?

… or

Complete your sessions evaluation online at SHARE.org/AnaheimEval

Deep Analytics Appliance – revolutionizing analytics

Purpose-built analytics appliance

Speed: 10-100x faster than traditional systems

Simplicity: Minimal administration and tuning

Scalability: Peta-scale user data capacity

Smart: High-performance advanced analytics

Dedicated High Performance Disk Storage

Blades With Custom FPGA Accelerators

SHARE in Anaheim
2012

DB2 for z/OS –BigInsights integration work items

- Function Shipping
  - REST based DB2 scalar udf to kick off BigData job returning status from scheduler
  - REST based DB2 table udf to read a BigData file given the table schema at table udf definition time
    - Different BigData Analytic tasks return different number and types of columns.
    - This approach could require the user to create up to one table udf per analytic function defined on DB2 that is run on BigData
  - REST based DB2 table udf to read a BigData file given the table schema at table udf invocation time
    - The same table udf may be invoked by different analytic functions defined on DB2, the schema on the DB2 is specified at the time of query
    - Inability to handle analytic algorithms where the schema is known only on execution time, e.g., dimension reductions, where the reduced number of dimensions are known only after the analytics is run.
  - REST based DB2 table udf to read a BigData file given the table schema during table udf execution time
- Job shipping
  - Define a BigData job in DB2 using the emerging SQL/MapReduce standard and push entire MapReduce jobs from DB2 to BigData
- Efficient Bulk Data Transfer
  - A mechanism for BigData and DB2 to understand each other's metadata, facilitate movement of bulk data, online and or offline efficiently in parallel

REST - REpresentational State Transfer

B

SHARE
in Anaheim
2012

IBM Mobile Strategy
Enabling users to participate in social business on the move

- A **comprehensive solution** for social business delivered as first-class mobile **platform** experiences

- Made **available on leading devices** through the device platforms' associated distribution channel or enterprise 'app stores'.

- *Wide support for business analytics capabilities such as event detection and notification, reports, charts, dashboards etc.*

- Supported by **application development tools** to help partners and customers reach their mobile user base by mobilize their information and applications

- Complete with **enterprise governance capabilities** that are easily managed on premises, hosted, or in the cloud

25

IBM Connections 3.0.1 - Mobile Application
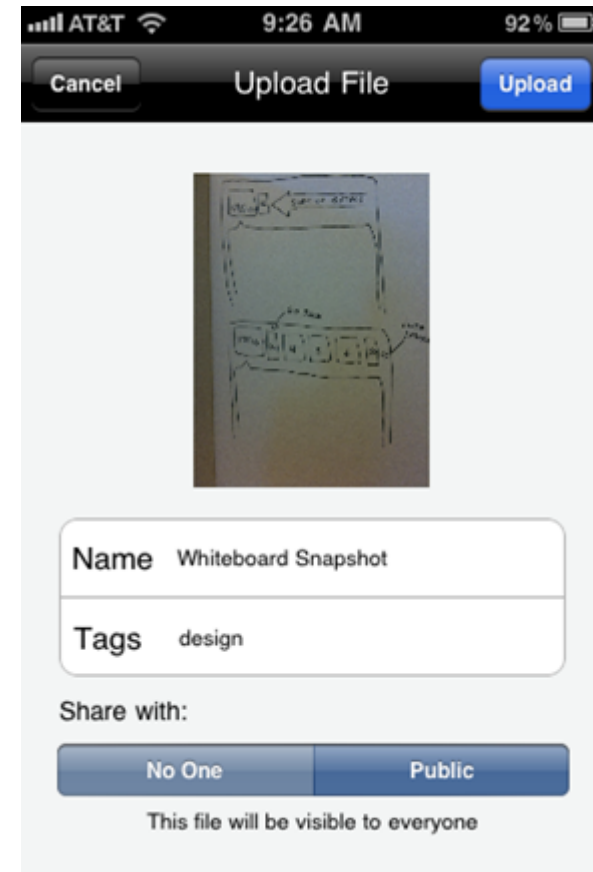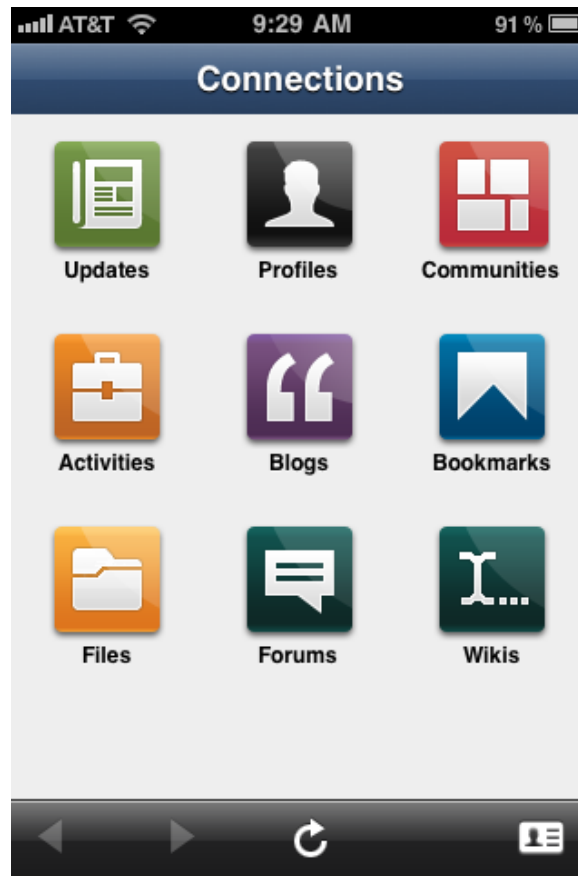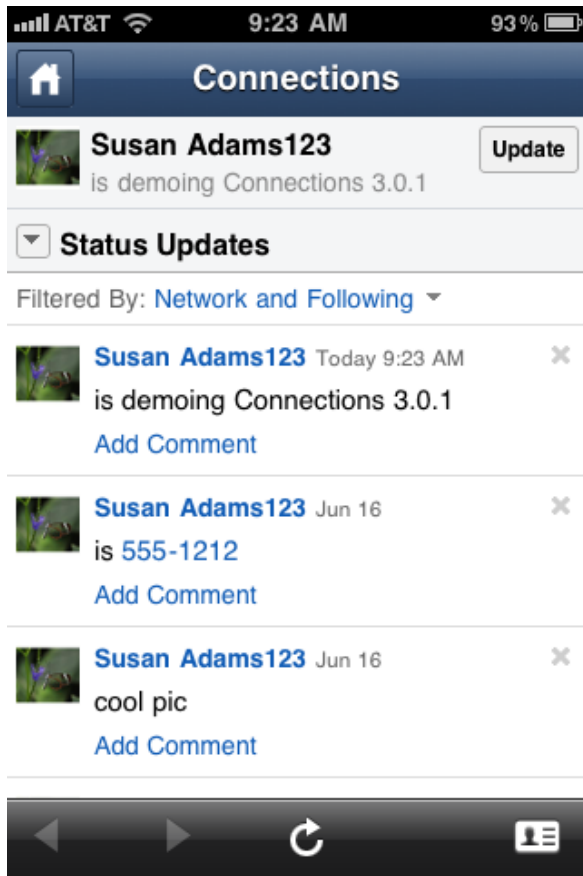
IBM Sametime Mobile Meetings

- ## Access to online meetings on mobile devices
  - Participate in online meetings
  - View content, add comments

- Available today
  - Lotuslive meetings today (RIM Blackberry and iOS)

  - Presentation, Desktop App Sharing (Zoom in/out)

  - Sametime Meetings 8.5.2

    - RIM BlackBerry mobile application

    - Web browser meetings access from iPhone/iPad/Android devices

Vivisimo Overview *New … you may find it interesting*
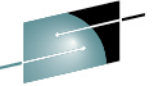
Value Proposition
- Accuracy – more relevant results due to position-based indexing
- Security – respects the security rights of underlying systems
- Scalability – scales to trillions of records

Differentiators

- **Unique Federated Discovery and Navigation Technology**
    - Position-based vs. vector-based index
    - Clustering and faceting to navigate data results
- **Scalable Architecture**
    - Fully distributed, fault-tolerant , unlimited scalability
- **Advanced On-the-Fly Analytics**
    - State-of-the-art real-time text and meta-data analytics
- **Secure Connectivity**
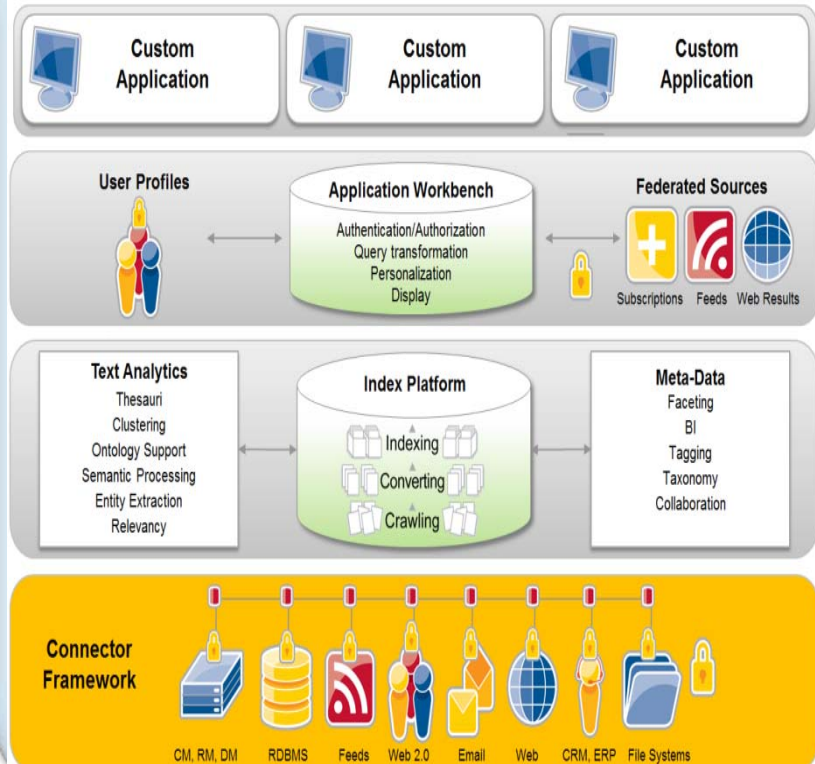   - Secure data integration of multiple repositories in complex IT environments
- **Powerful Development Tools**
    - Easy-to-deploy applications across varied and large data sets & sources
- **Fast Time to Value**
    - Rapid deployments from POCs to production

Vivisimo Product Overview

28

InfoSphere BigInsights on zEnterprise

Big data

Cloud

Optimized
systems

Exploration:

IBM Software Group, responsible for Big Data Development, is actively working alongside zEnterprise Development in Poughkeepsie and Technical Sales Support to explore and validate a suitable IBM InfoSphere BigInsights offering that will run within the zEnterprise BladeCenter Extension (zBX) frame on virtualized HX5 Blades.

SHARE
in Anaheim
2012

Summary

- ✓ IBM business analytics are in transition as we embrace more of the Big Data initiative
  - ✓ Access more data in new forms
  - ✓ Analyze
  - ✓ Auto-analyze & predict
  - ✓ Collaborate
  - ✓ Deploy
- ✓ Solution elements that offer traditional query, reporting, charts, and dashboards are working on extensions to their data access (e.g. Cognos, QMF)
- ✓ IBM DB2 for z/OS is being significantly enhanced for a variety of new analytic uses (IDAA, in-DBMS scoring, temporal support)
- ✓ New deployment options are being added and extended such as mobile device support for QMF
- ✓ … and much more to come

30

SHARE
in Anaheim
2012