# 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*

? TBs of data every day

25+ TBs of log data every day

30 billion RFID tags today (1.3B in 2005)

76 million smart meters in 2009… 200M by 2014

4.6 billion camera phones world wide

100s of millions of GPS enabled devices sold annually

2+ billion people on the Web by end 2011

**Variety**

**Volume**

**Velocity**

# Why Big Data?

Reduce risk
  - Deeper understanding of market opportunities and threats

Lower cost
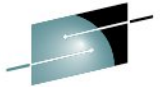  - Deliver goods and services smarter / more efficiently

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**
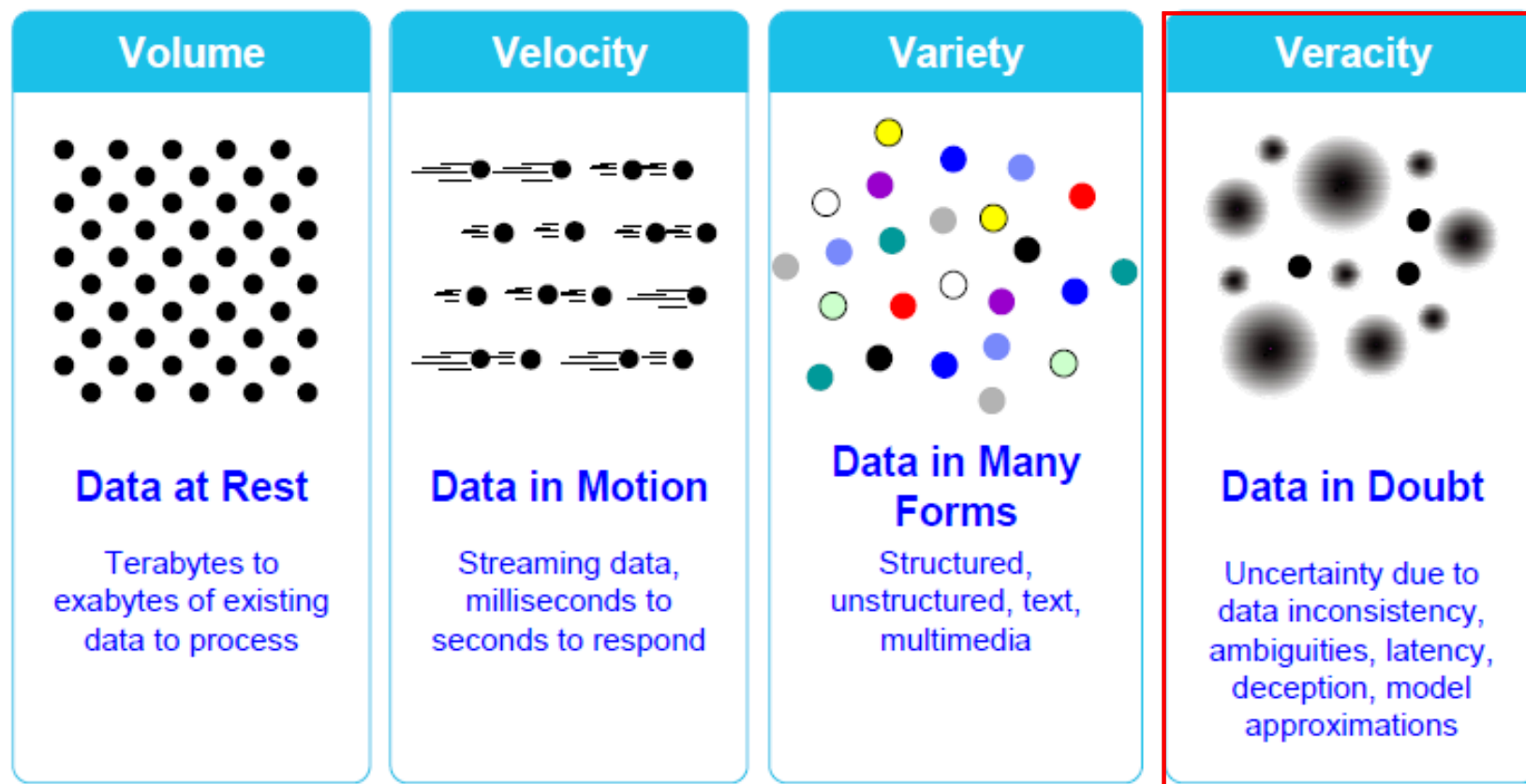
**Volume**

**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

SHARE 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**

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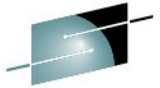
- What data can you manage / analyze today?



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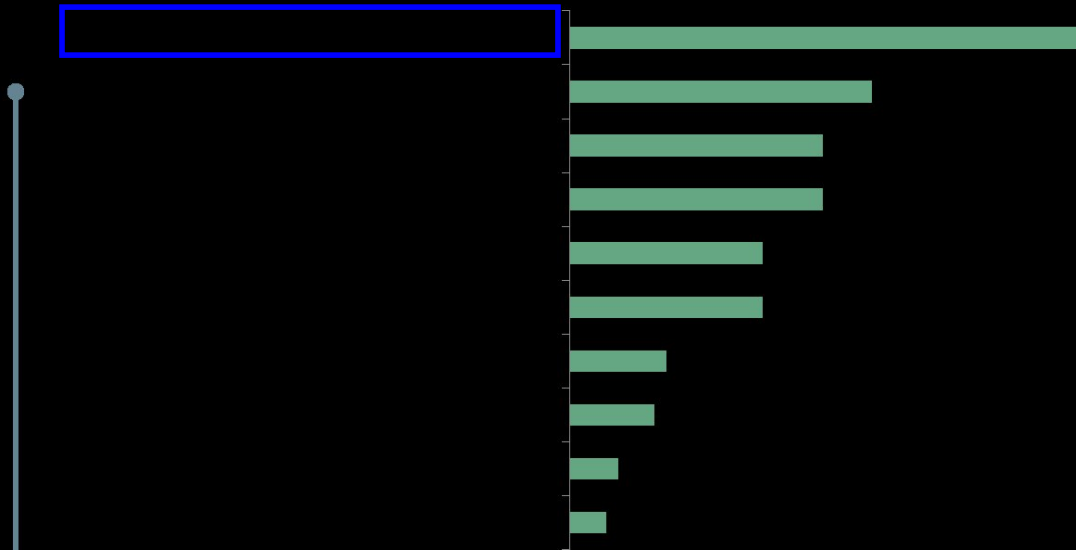
Most big data use cases hype its application for analysis of new, raw data from social media, sensors, and web traffic, but we found that firms are being very practical, with early adopters using it to operate on enterprise data they already have.

# 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
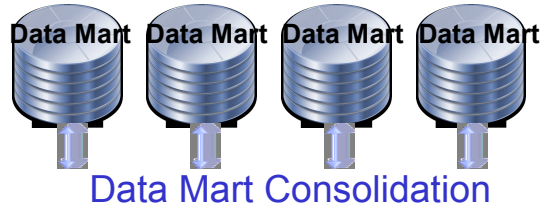
**System z PR/SM**
*Recognized leader in mixed
...tualization and workload isolation*

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*

# 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

# *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

*Data from Social Media sites analyzed with Text analytics*
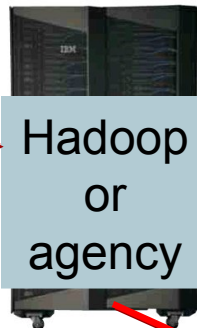
Refined Search parameters from OLTP environment

Data Warehouse

Integration

Business Analytics

DB2 for z/OS IMS

Information Governance

Hadoop or agency

Result Set for further processing
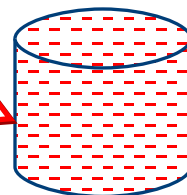
Data Warehouse + modeling applications

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
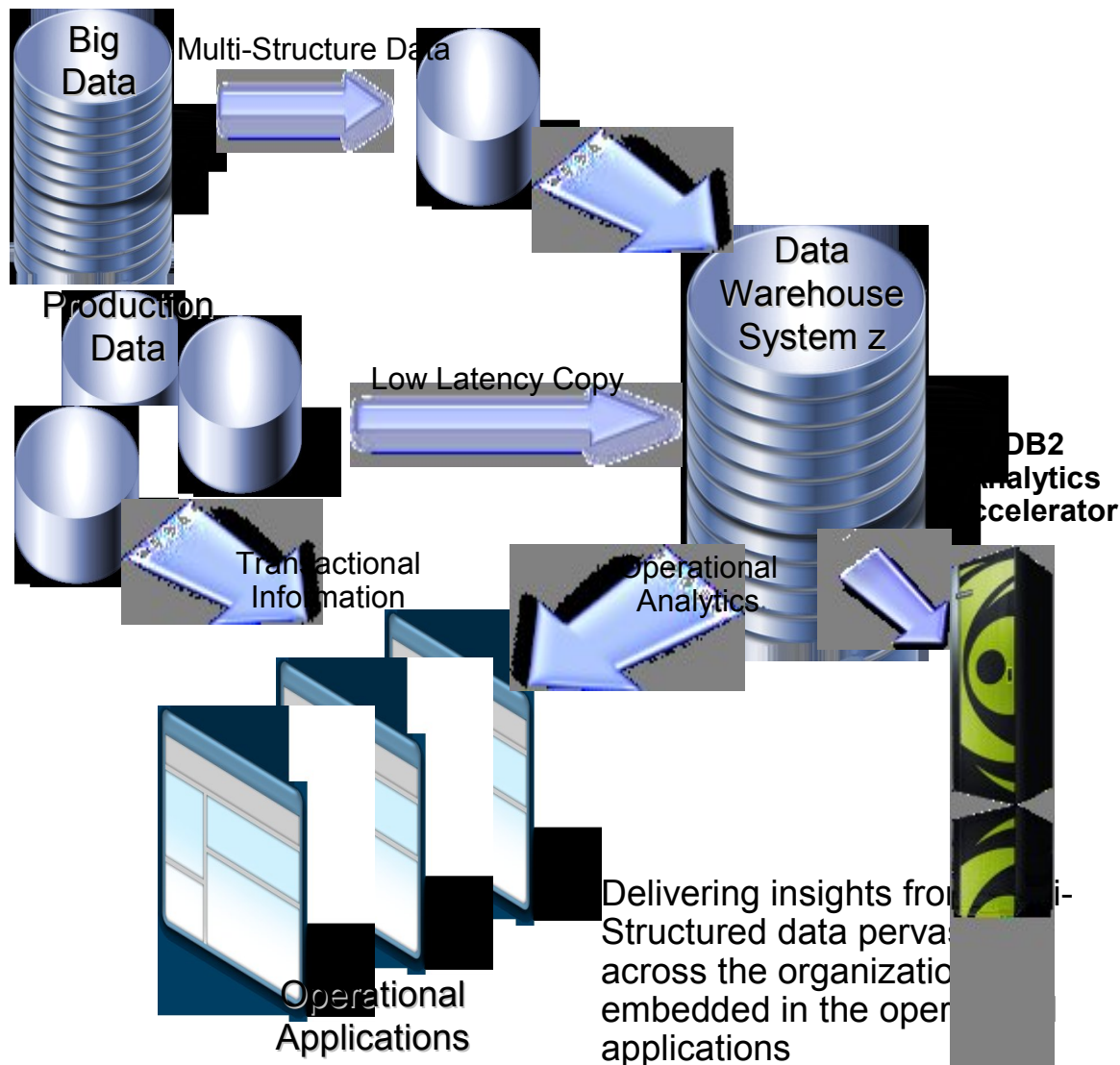
Traditional Sources
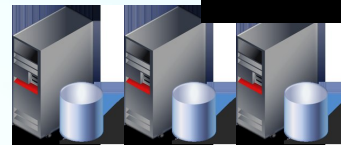
New Sources

in Anaheim
2012

# *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
- …
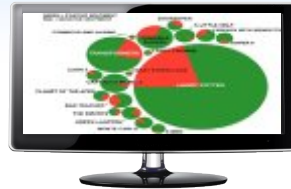
- Location-based marketing
- Smart meter analytics
- …

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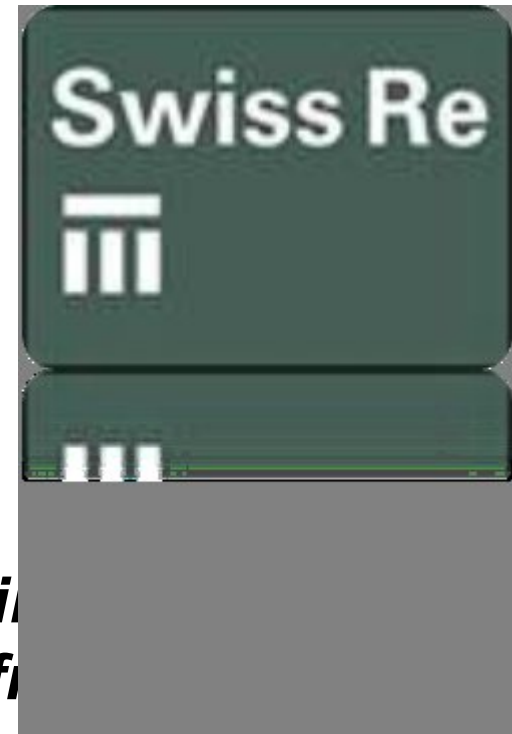
- *Insights i... policies fr... insurance companies*
- *Getting their reports as much as 70 percent faster*

# 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

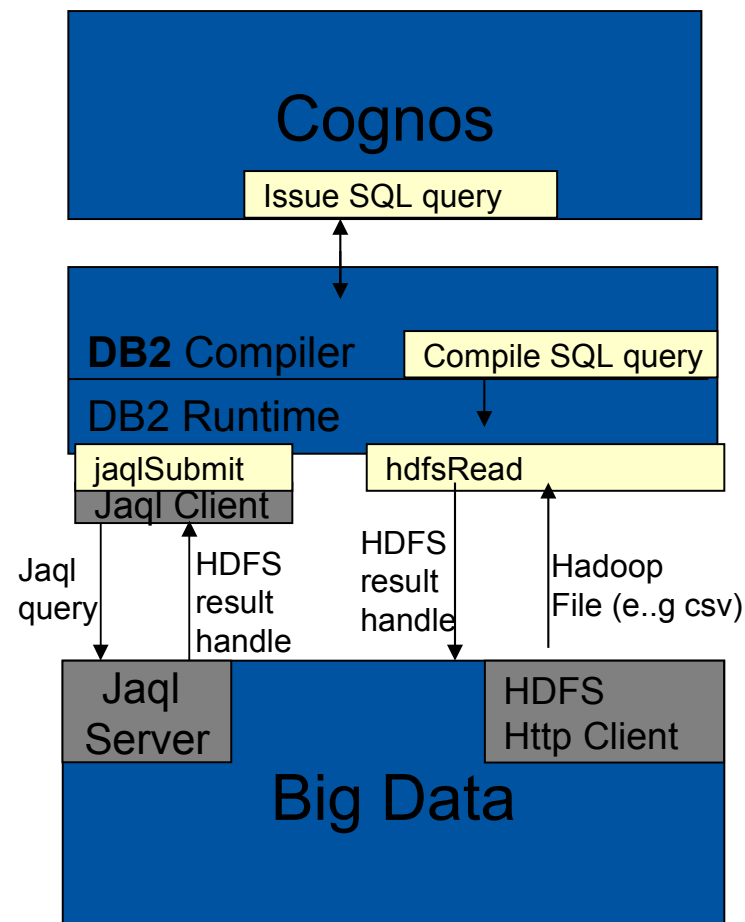Predictive

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**

# Connecting Big Data and DB2 for z/OS futures

Cognos

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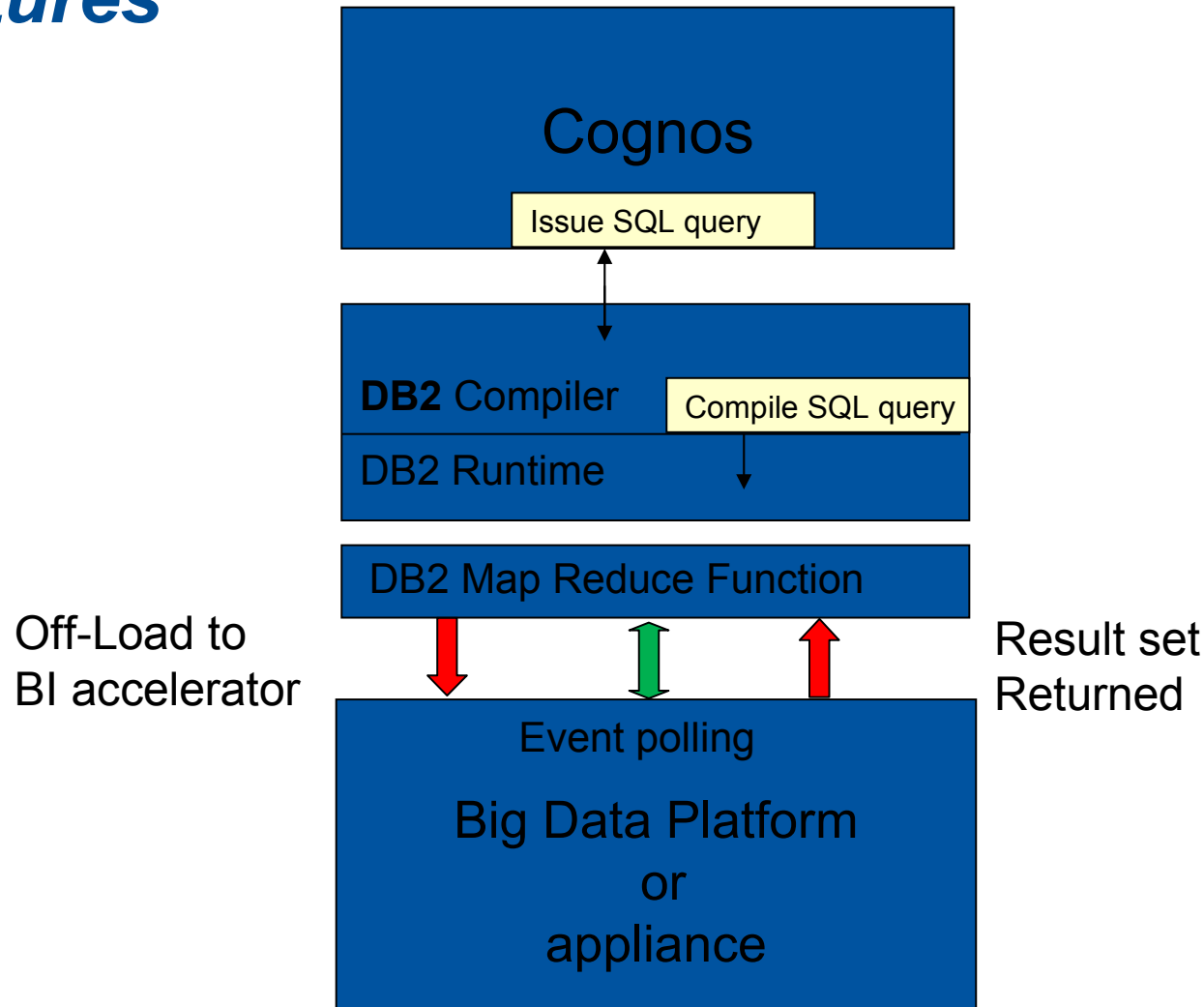
# *Direct Connect for Big Data and DB2 for z/OS Bulk Data movement*

**Read Input, Splits, Convert,
Partition**

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

| Big Data | | Map | | Map | | Map | | Shuffle/Sort | | Reduce | | Reduce | | Reduce | | zDB2 |
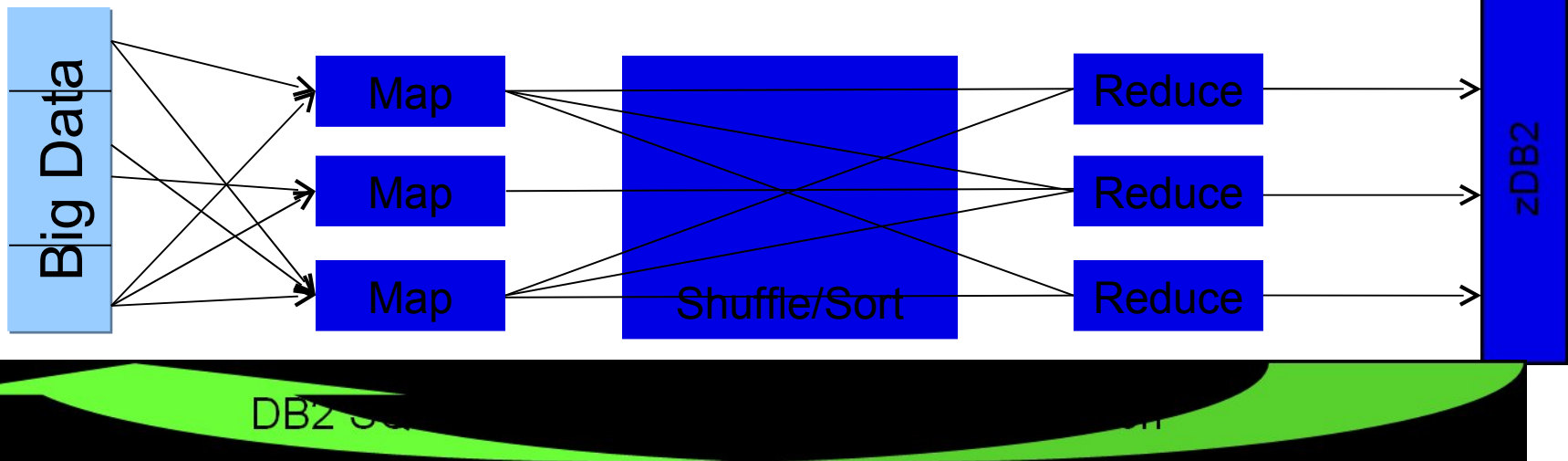
DB2 Catalog

- 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.
- BigData 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*
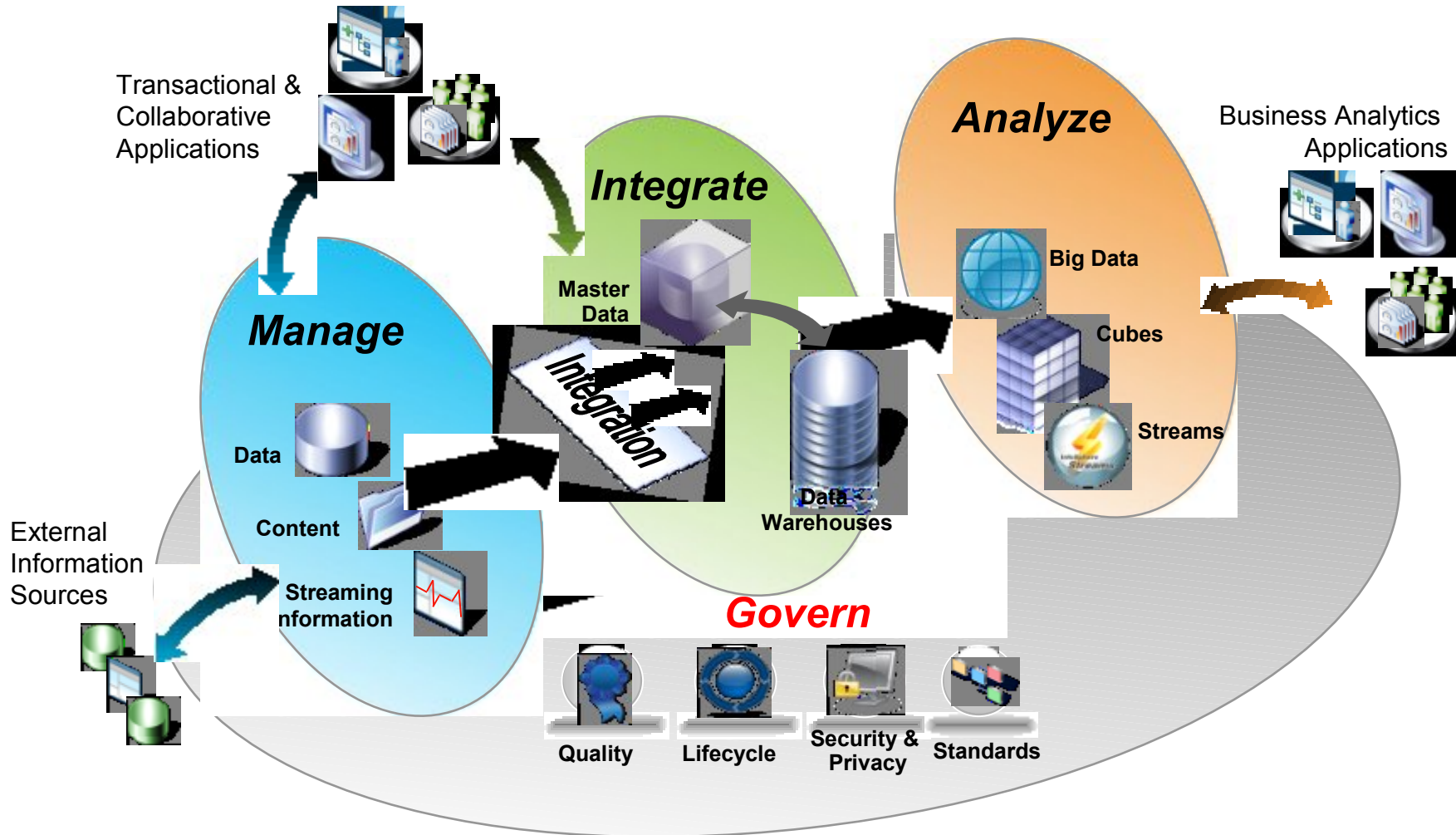
*Structured, analytical, log...*
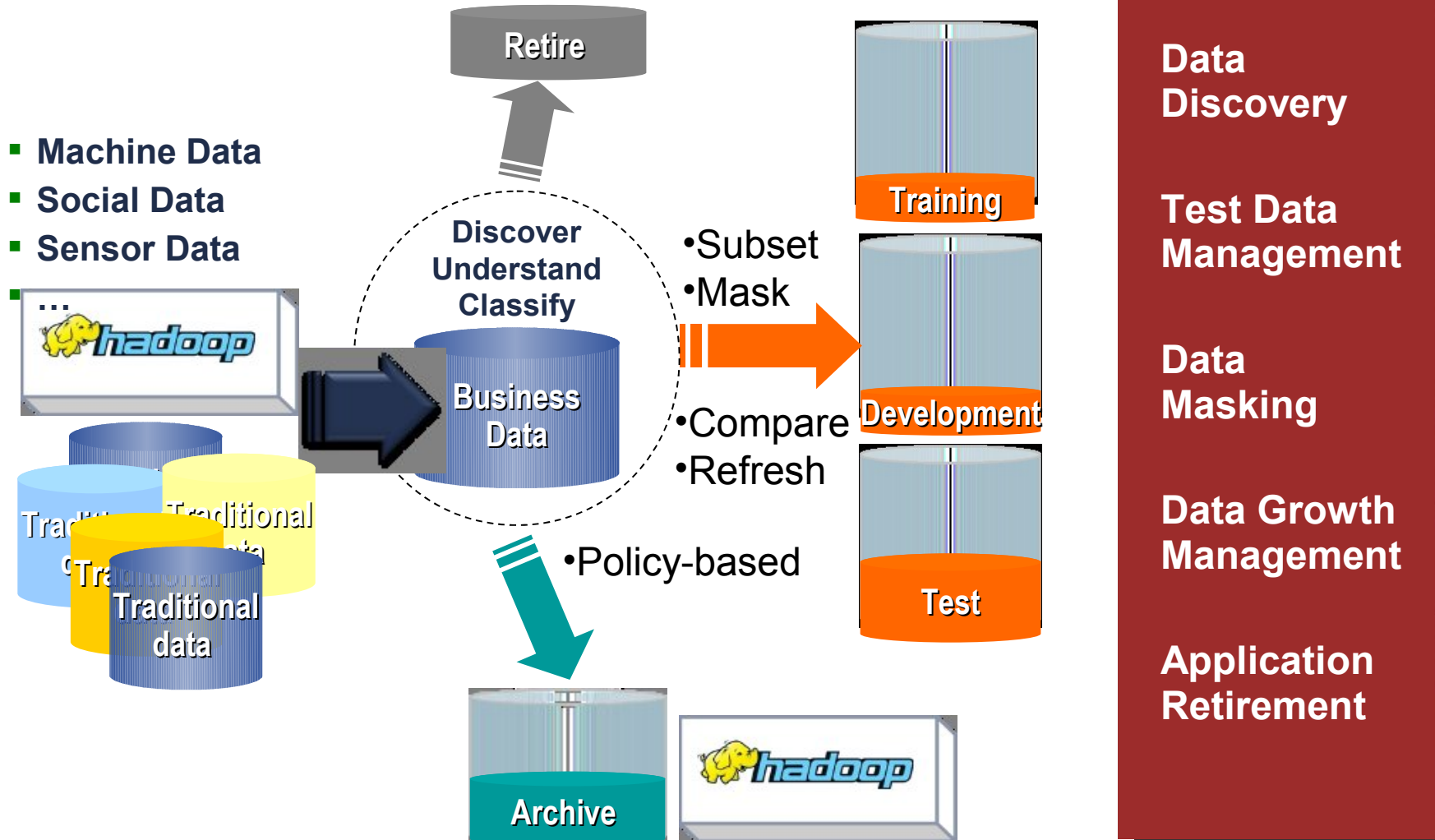
*New Approach*

*Creative, holistic thought, intuition*
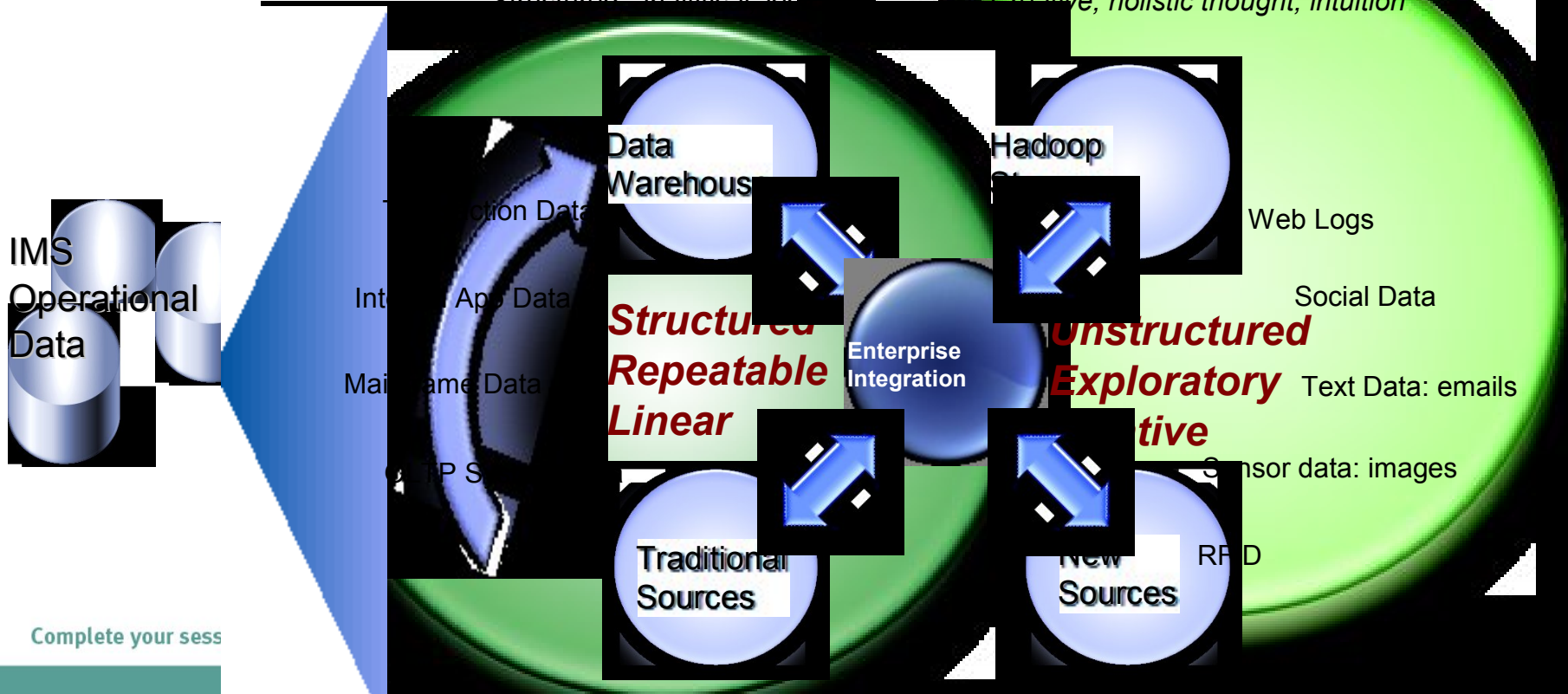
IMS Operational Data

Transaction Data

Internal Apps Data

Mainframe Data

OLTP S...

Data Warehous

Hadoop St...

Web Logs

Social Data

Text Data: emails

Sensor data: images

RFID

*Structured Repeatable Linear*

**Enterprise Integration**

*Unstructured Exploratory ...tive*

Traditional Sources

New Sources

# *Agenda*

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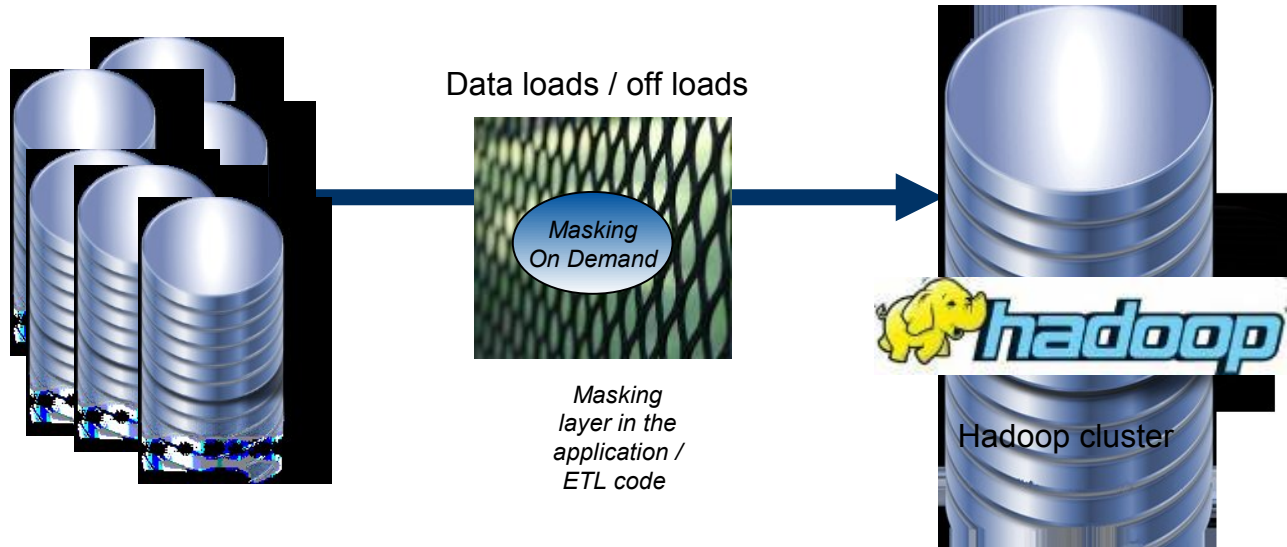  - **Data Management**
  - ***Information Governance***
- **Call to Action**

# Information Governance



Transactional & Collaborative Applications

Business Analytics Applications

**Analyze**

**Integrate**

**Manage**

Master Data

Big Data

Cubes

Streams

Data

Content

Streaming Information

Data Warehouses

External Information Sources

**Govern**

Quality

Lifecycle

Security & Privacy

Standards

# *Information Lifecycle Management (and Big Data)*

- **Machine Data**
- **Social Data**
- **Sensor Data**
- **...**

**Retire**

**Discover Understand Classify**

**Business Data**

Traditional data

Traditional data

Traditional 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**

Complete your sessions evaluation online at SHARE.org/AnaheimEval

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*
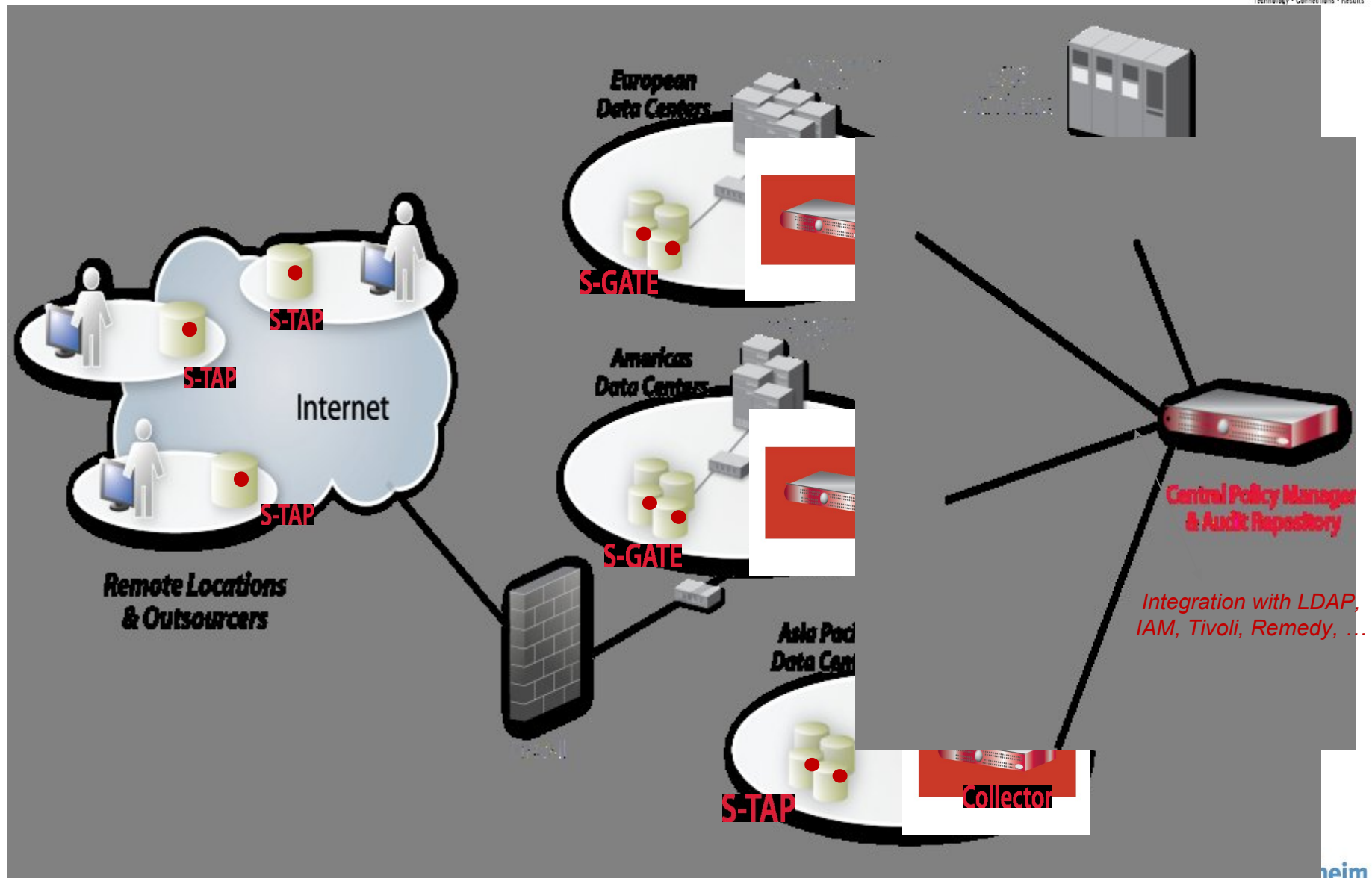
# Customer requirements for Hadoop support



- 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
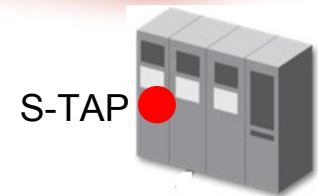
**"We want Hadoop Activity Monitoring"**
*Monitor and Audit Hadoop activity in real-time to support compliance requirements and protect data*

- 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*
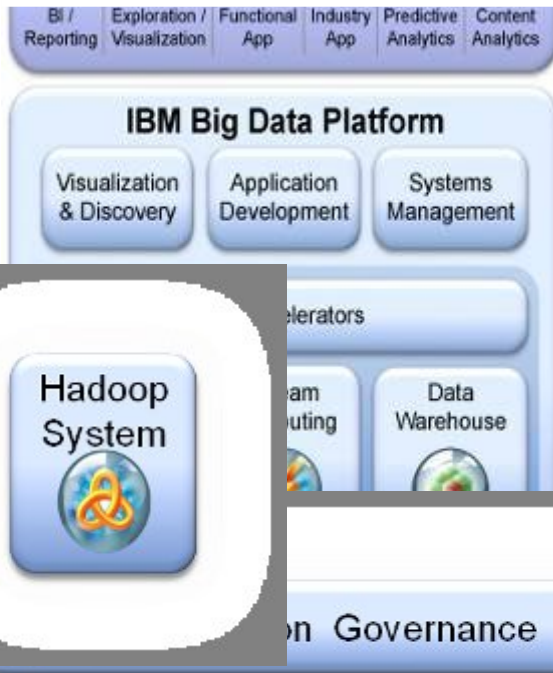


- **HDFS – Hadoop file system**

- 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**

SHARE
in Anaheim
2012

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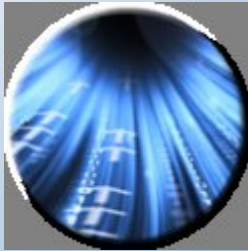
**1.** *Any type of data* — Manage and integrated any data types

**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
# Z