#SHAREorg

SHARE
Technology · Connections · Results

Get the Latest on Big Data with IBM System z

Mark Simmonds
IBM Corporation

February 5, 2013
Session Number 12528

mtsimmon@us.ibm.com
linkedin.com/pub/mark-simmonds/11/596/14a/

SHARE in San Francisco 2013

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.

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

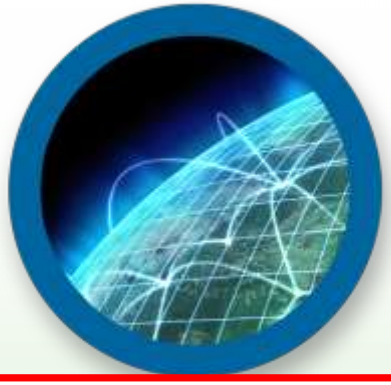
Manage and control the data you are responsible for

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



"Every day, we create 2.5 quintillion(10^{18}) bytes of data — so much that 90% of the data in the world today has been created in the last two years alone."

Volume

12 terabytes
of Tweets created daily

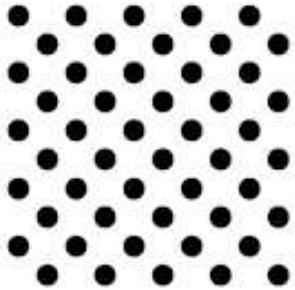
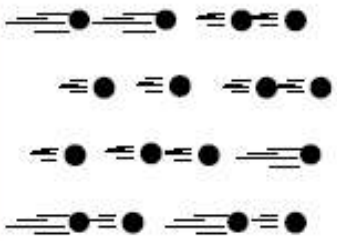
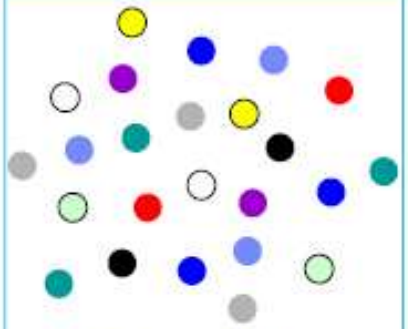
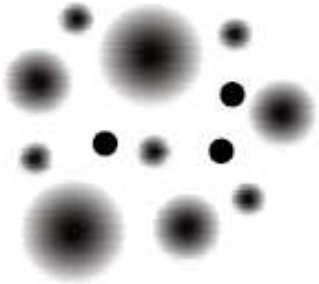
Velocity

5 million
trade events per second

Variety

100's video feeds
from surveillance cameras

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*
- 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

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

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 "How Forrester Clients Are Using Big Data, Forrester Research, Inc., September 20, 2011."

…and our webcast survey said…

- Have you already implemented or are you planning to implement any Big Data based initiatives within the next 6 months?

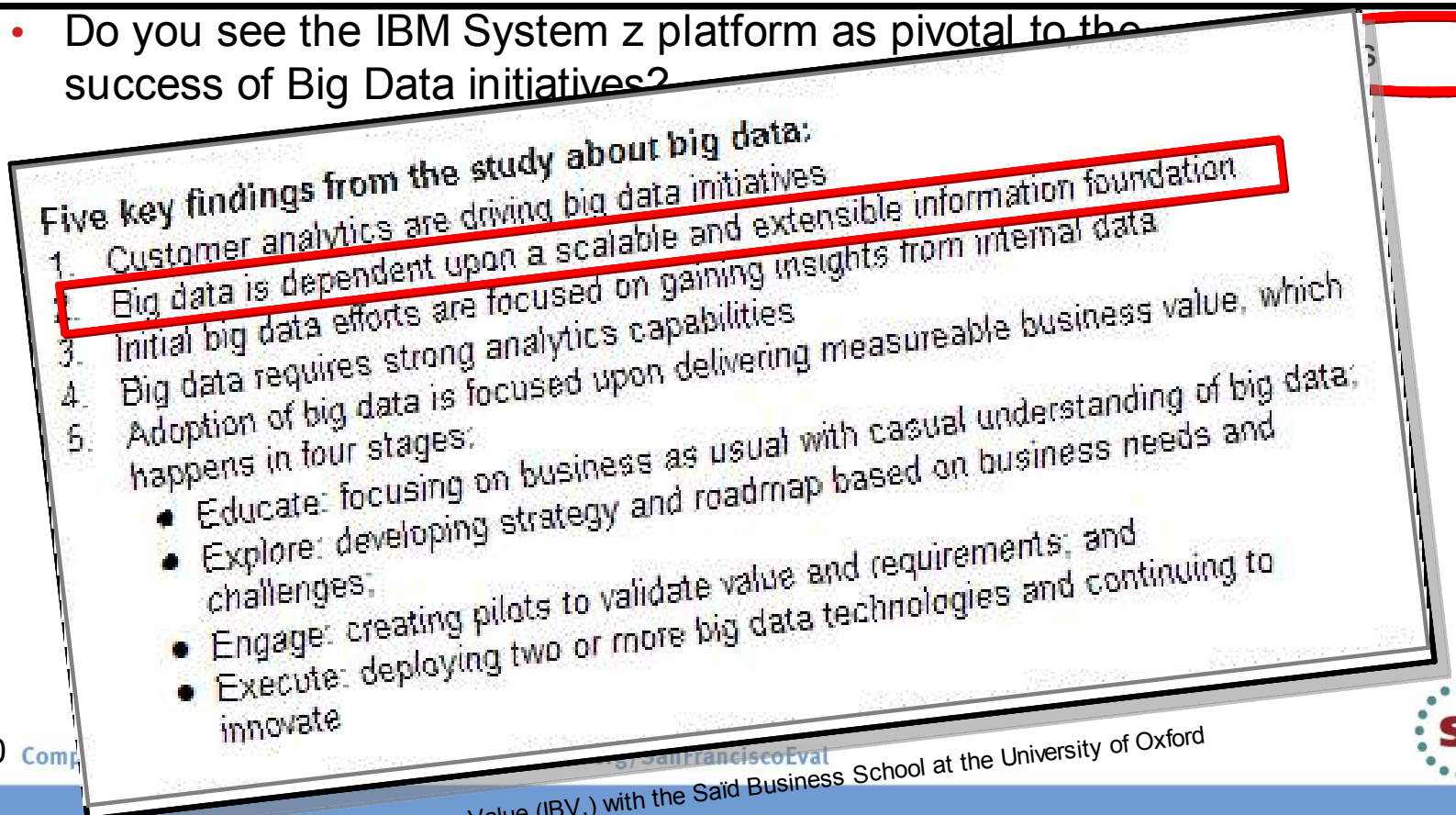
Yes	31%
No	69%

- How would you rate the value of being able to integrate insights from social media, telemetry, unstructured data into your analytics and decision making processes?

High	50%
Medium	36%
Low	14%

- Do you see the IBM System z platform as pivotal to the success of Big Data initiatives?

	90%
	10%



Five key findings from the study about big data:

1. Customer analytics are driving big data initiatives
2. Big data is dependent upon a scalable and extensible information foundation
3. Initial big data efforts are focused on gaining insights from internal data
4. Big data requires strong analytics capabilities
5. Adoption of big data is focused upon delivering measureable business value, which happens in four stages:
   - Educate: focusing on business as usual with casual understanding of big data;
   - Explore: developing strategy and roadmap based on business needs and challenges;
   - Engage: creating pilots to validate value and requirements; and
   - Execute: deploying two or more big data technologies and continuing to innovate

Value (IBV.) with the Saïd Business School at the University of Oxford

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 (Think inside the box!!!)

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

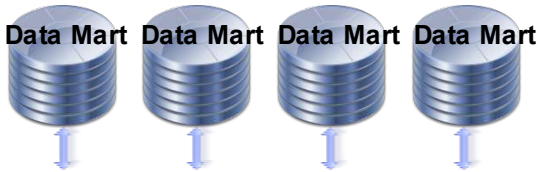
Unmatched end-to-end security

- 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.

zEnterprise, the Ultimate Consolidation Platform



Data Mart Consolidation

Transaction Systems (OLTP)

Data Warehousing
Business Intelligence
Predictive Analytics

System z PR/SM
Recognized leader in mixed virtualization and workload isolation

zBX
Applications & accelerators

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

IDAA
Recognized leader in cost-effective high speed deep analytics

Bringing it all together

- ***Better Business Response***
- ***Reduced Costs***
- ***More Available***
- ***More Secure***
- ***Reduced Data Movement***
- ***Reduced Data Latency***
- ***Reduced Complexity***
- ***Reduced Resources***

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

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

- Analytics and decision engines reside where the DWH / transaction data is
- "Noise" (veracity) 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"

Demand for <u>differently</u> 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

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

Deterrent for fraudsters - Cost Savings for the business

Make payment or investigate

zEnterprise

DB2 Accelator

Data Warehouse

Integration

Business Analytics

DB2 for z/OS IMS

Information Governance

Data from Social Media sites analyzed with Text analytics

Hadoop or agency

Refined Search parameters from OLTP environment

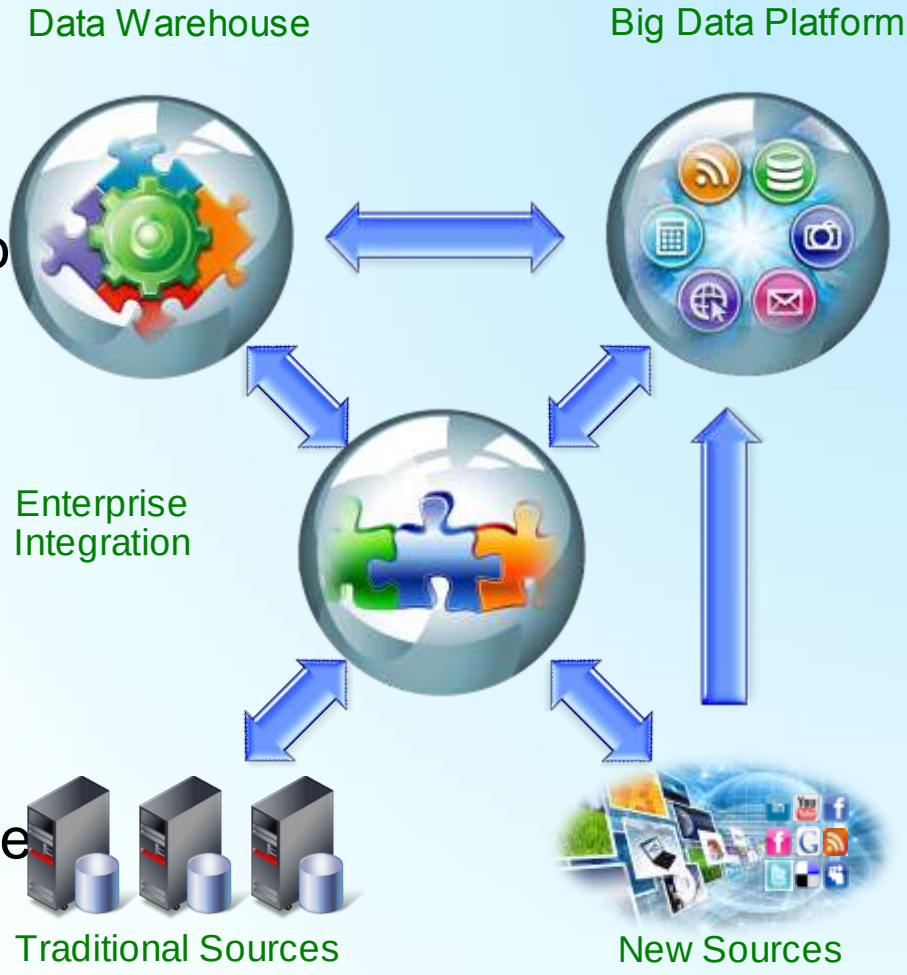
Result Set for further processing

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

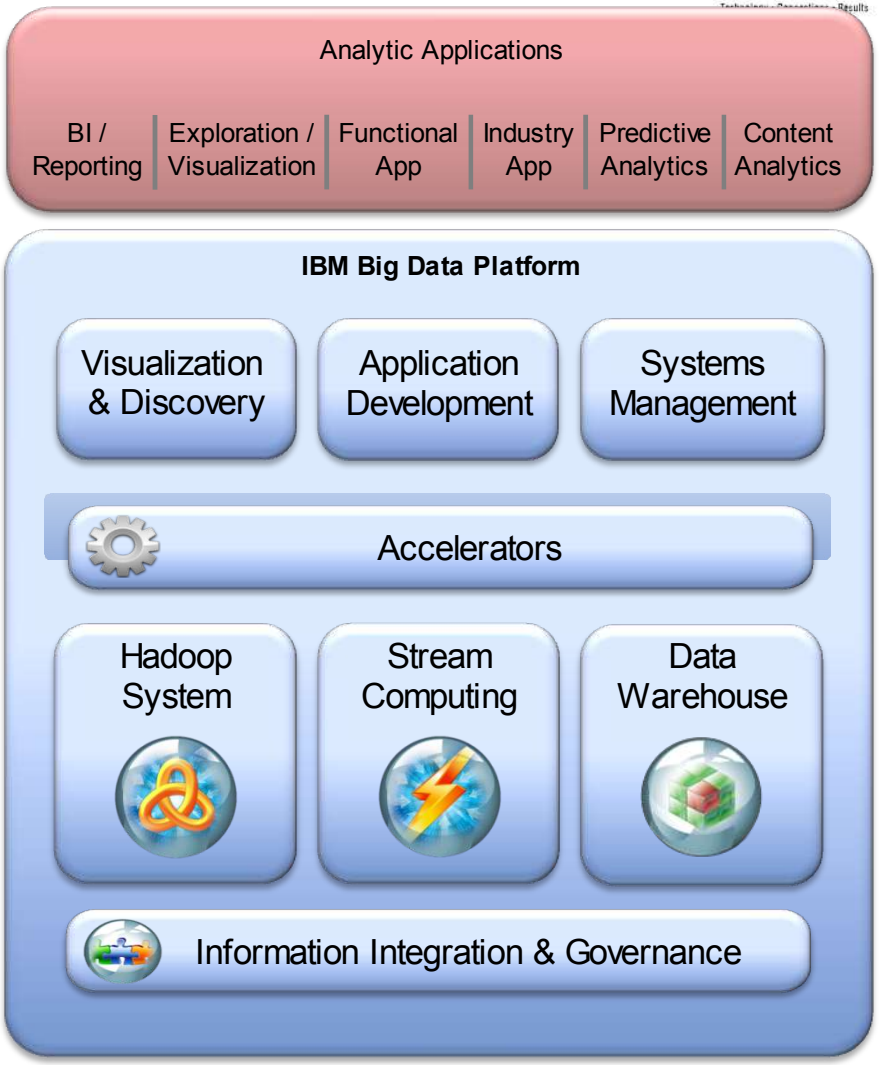
- **Information Integration**
  - Insights from Big Data must be incorporated into the warehouse and analytics/ decision engines
- **Information Governance**
  - Companies need to govern what comes in, and the insights that come out



IBM Big Data Strategy: Move the Analytics Closer to the Data

New analytic applications drive the requirements for a big data platform

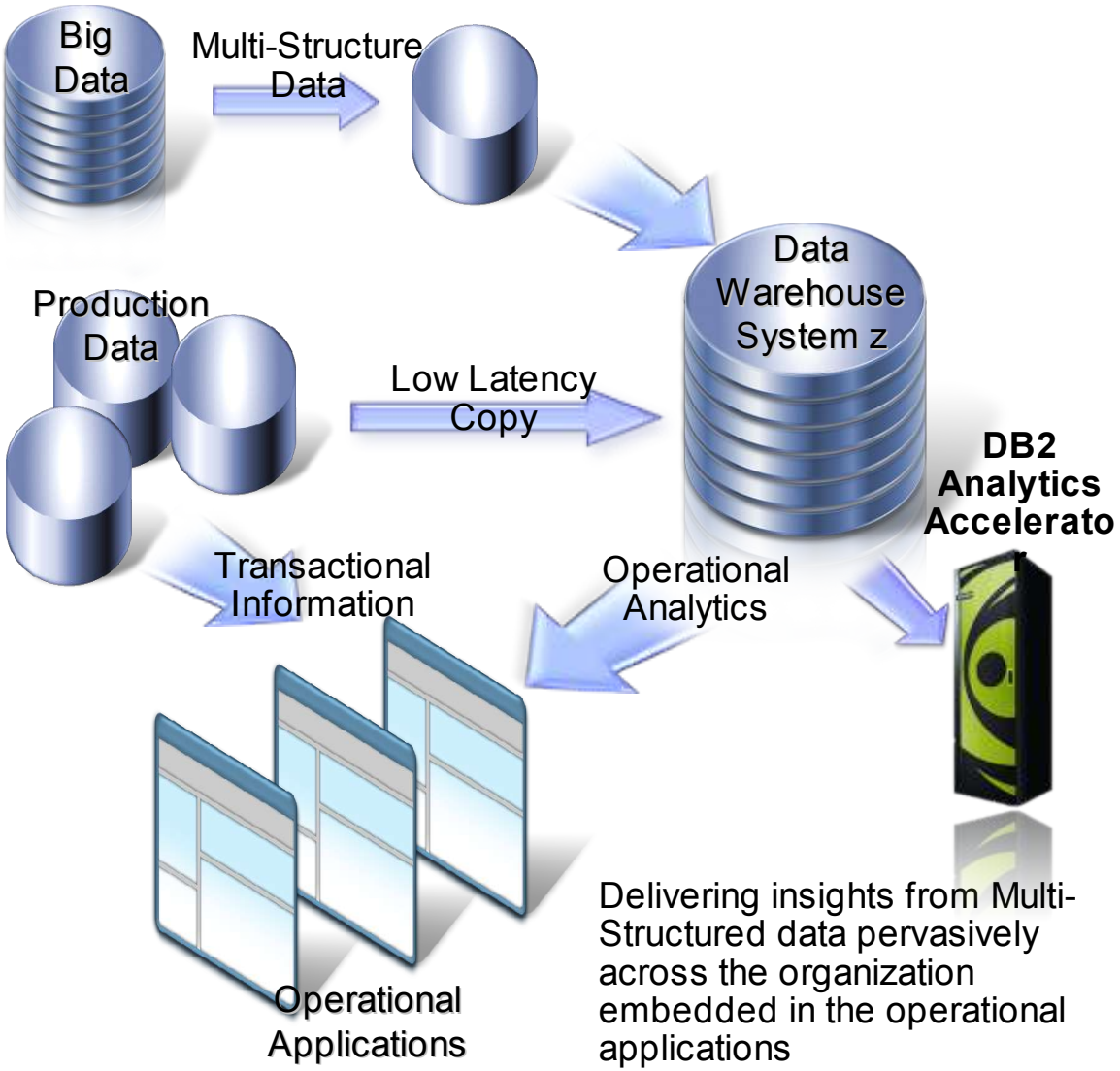
- Integrate and manage the full variety, velocity and volume of data
- Apply advanced analytics to information in its native form
- Visualize all available data for ad-hoc analysis
- Development environment for building new analytic applications
- Workload optimization and scheduling
- Security and Governance



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.



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.

Analytics-driven Organizations Can…

Identify Risk

…and immediately control it

- ***Insights into overlapping policies from multiple 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

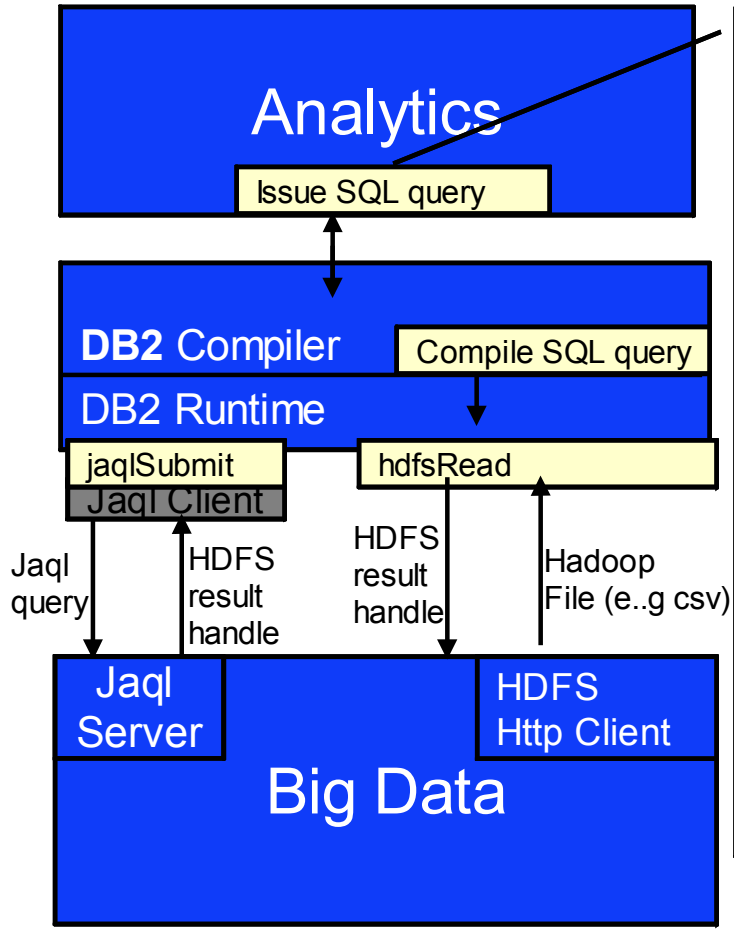

Mathematical

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

Agenda

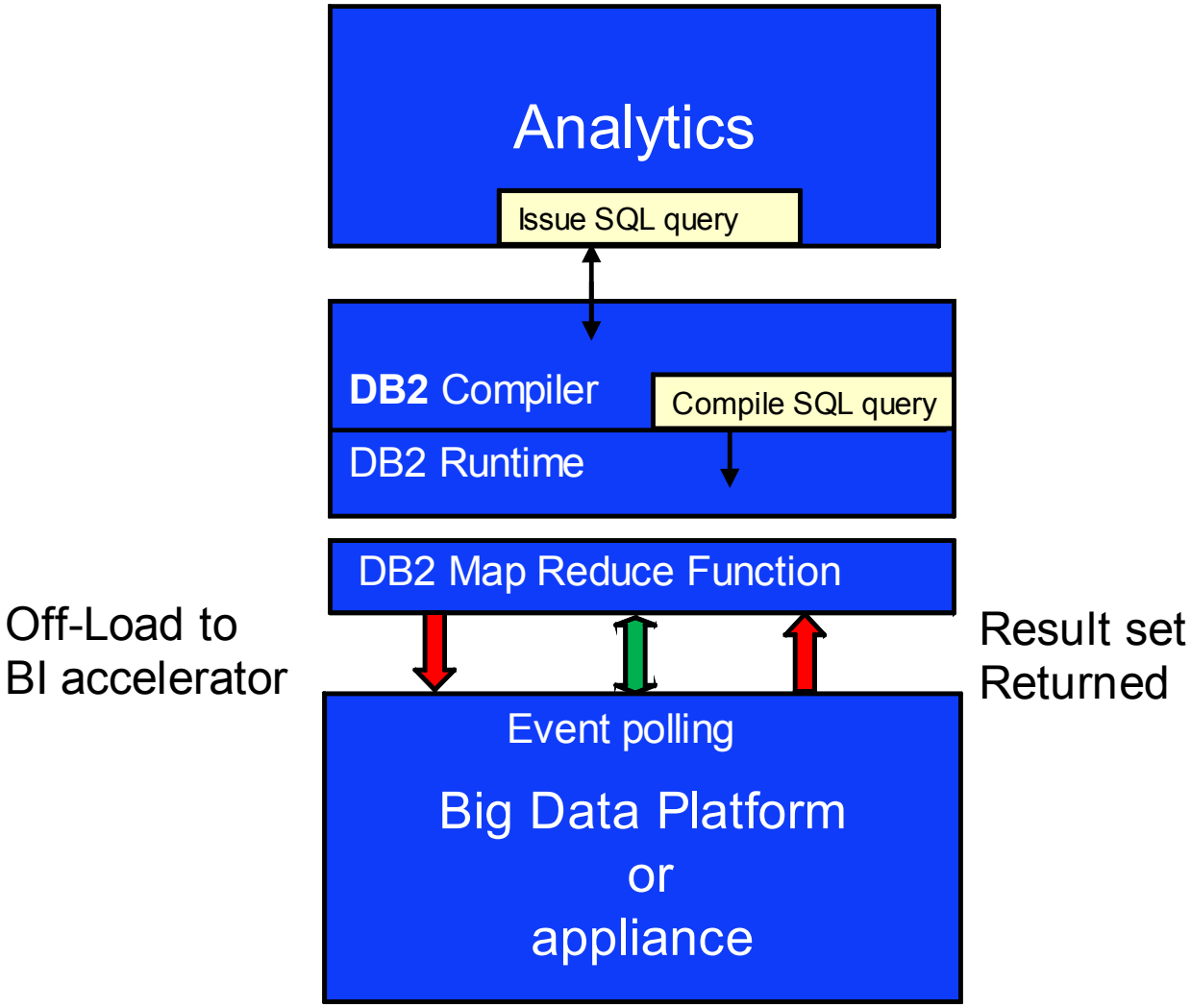
- Big Data – Why now?
- Start with System z
- Making Big Data a reality
  - Business analytics and Data Warehousing
  - *Data Management*
  - Information Governance
- Call to Action

Data management - Connecting Big Data and DB2: Phase 1

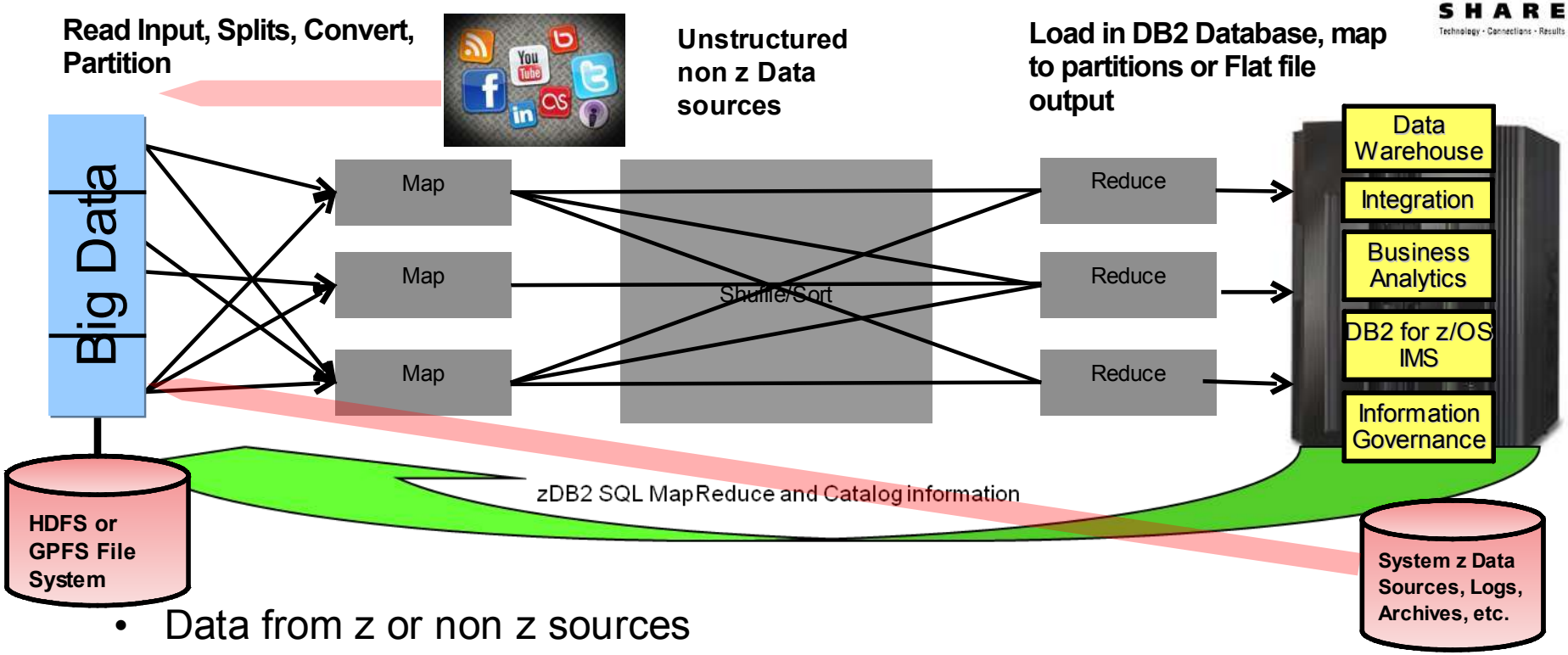



```
SELECT TX.*
FROM TABLE ( hdfsRead( jaqlSubmit('9.146.142.158:8080',
                                  'read(del( "/co2.dat", {
                                              fields: [ "station", "start", "lat",
                                                        "long", "alt",
                                                         "location", "co2" ],}))
                                  -> filter not isnull($.co2)
                                  -> group by $g = {$.station,
                                                    dateParts($.start){.
                                                    year,.month}}
                                              into {$g.station,$g.year, $g.month,
                                              lat: any($[].lat),long: any($[].long),
                                              alt: any($[].alt),
                                              location:  any($[].location),
                                              count: count($),minCO2: min($[].co2),
                                               maxCO2:  max($[].co2),
                                               avgCO2: avg($[].co2) }'',
'                                    ''["station", "year", "month", "lat", "long",
                                      "alt", "location", "count", "minCO2",
                                      "maxCO2", "avgCO2" ]''
                                    ) AS TX (station VARCHAR(35), year VARCHAR(4), month
    VARCHAR(2), lat VARCHAR(6), long VARCHAR(6), alt VARCHAR(8), loc VARCHAR
    (10), count VARCHAR(5), minCO2 VARCHAR(7), maxCO2 VARCHAR(7), avgCO2
    VARCHAR(20))) as TX
```

Connecting Big Data and DB2 for z/OS futures



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



- 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.

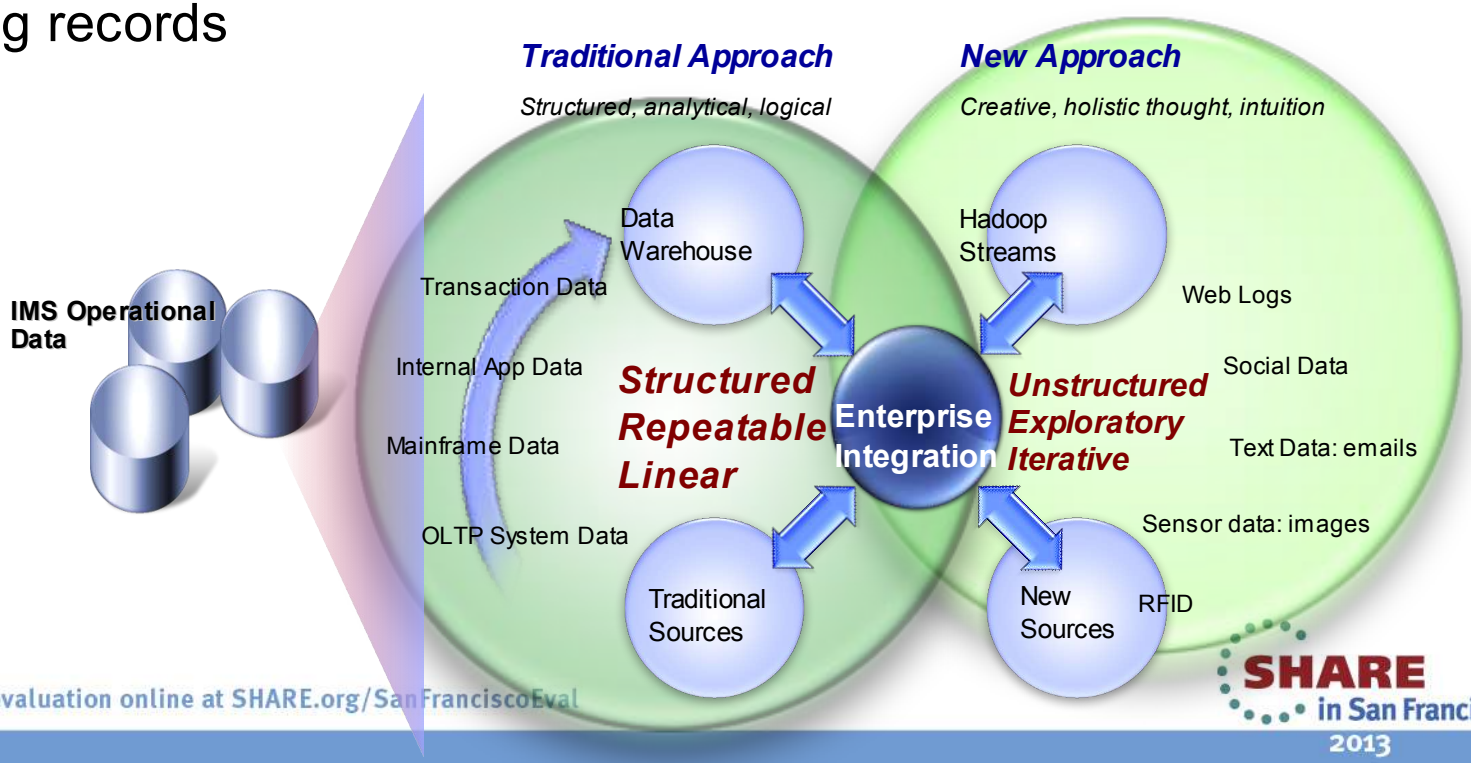
Big Data and CICS & IMS Apps

- Access to Hadoop analytics data stored in DB2 for z/OS
- z/OS Apps positioned as service requester of Big Data Analytics services
  - Based on numerous callout solutions: web services, MQ, …
  - Synchronously or asynchronously
- z/OS middleware logs analyzed with the Big Data Machine Data Accelerator (MDA)
  - Tremendous value in correlating and co-analyzing different types of logs from all sources
  - Correlate log records from off-platform application servers with IMS log records

Use Case	Customer Benefit	Description
Outage Prevention	Enhanced service levels and operational efficiency	Ad hoc analysis of IT log data to determine trends before system/service/application outage. Then using real time processing to detect such trends to prevent outage
Quality Improvement	Customer Satisfaction	Analyze machine data to identify areas of improvement from Quality perspective
Resource Prediction	Operational Efficiency	Analyze and model resource usage over time using machine/system data to predict resource requirements for future to increase operational efficiency

Big Data and IMS Databases

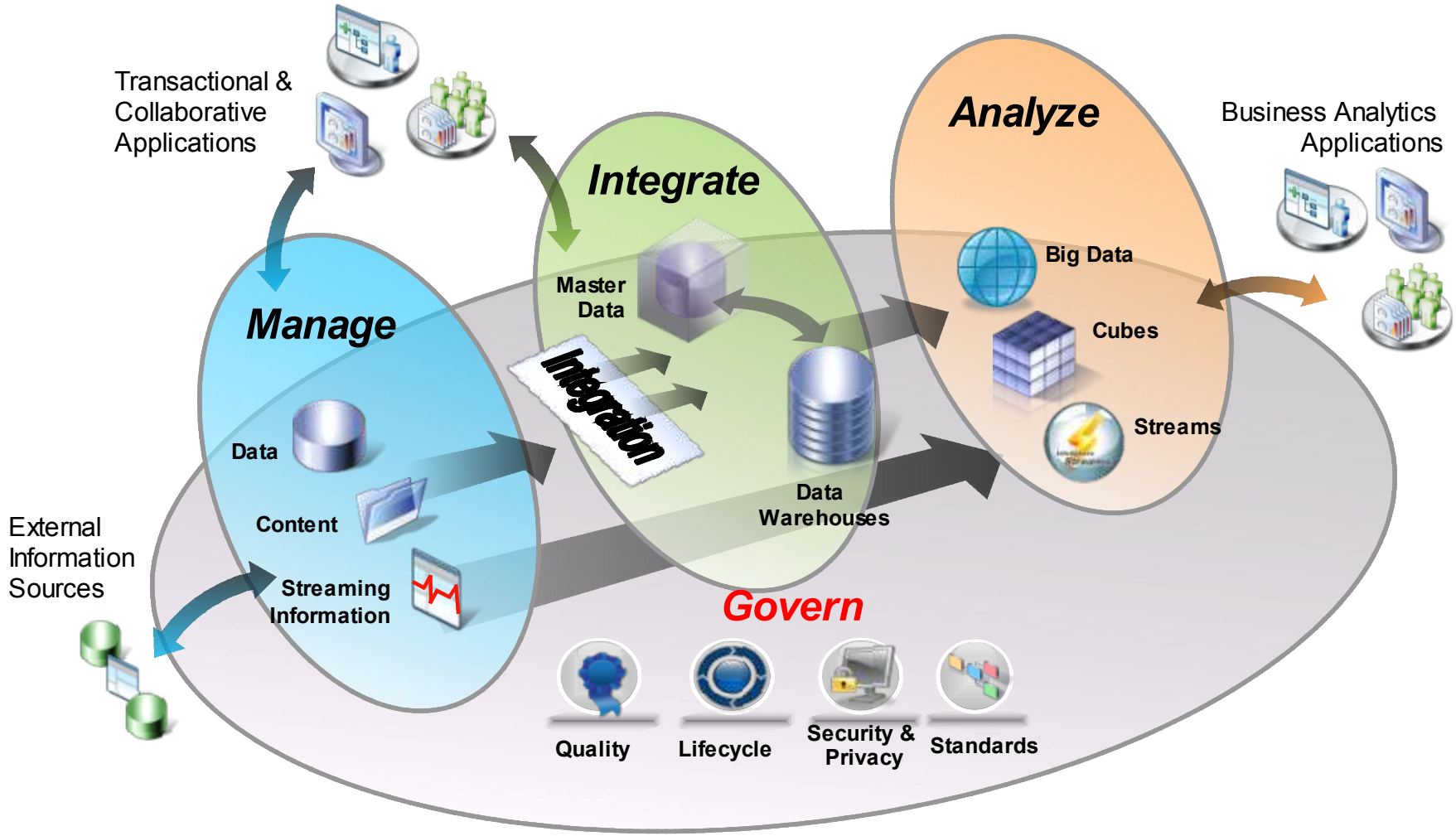
- IMS integration with the BigInsights application connectors
  - Merge trusted OLTP data with the Big Data platform
- Integrate IMS with the Big Data Machine Data Accelerator (MDA)
  - Correlate log records from off-platform application servers with IMS log records



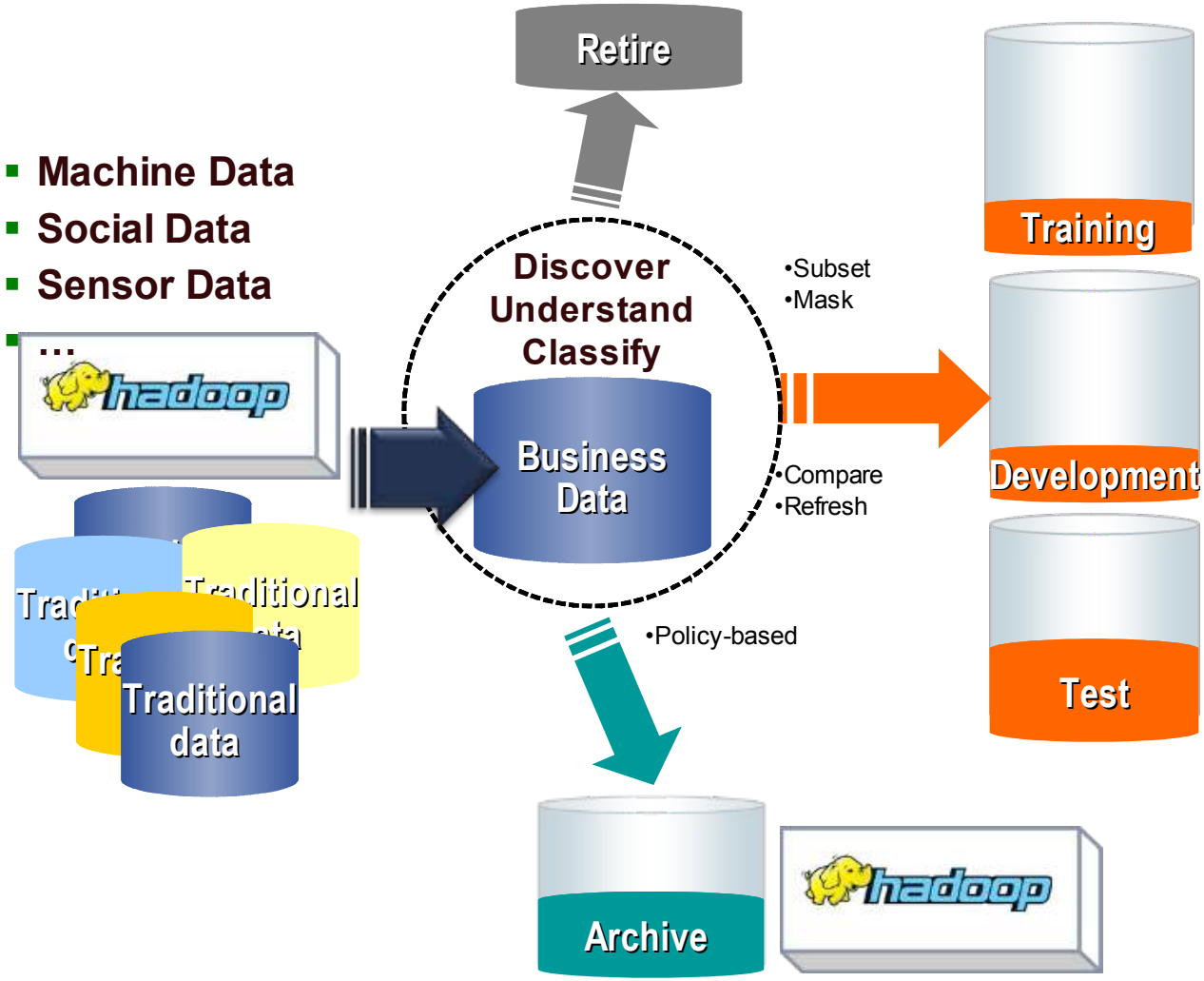
Agenda

- 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)



Data Discovery

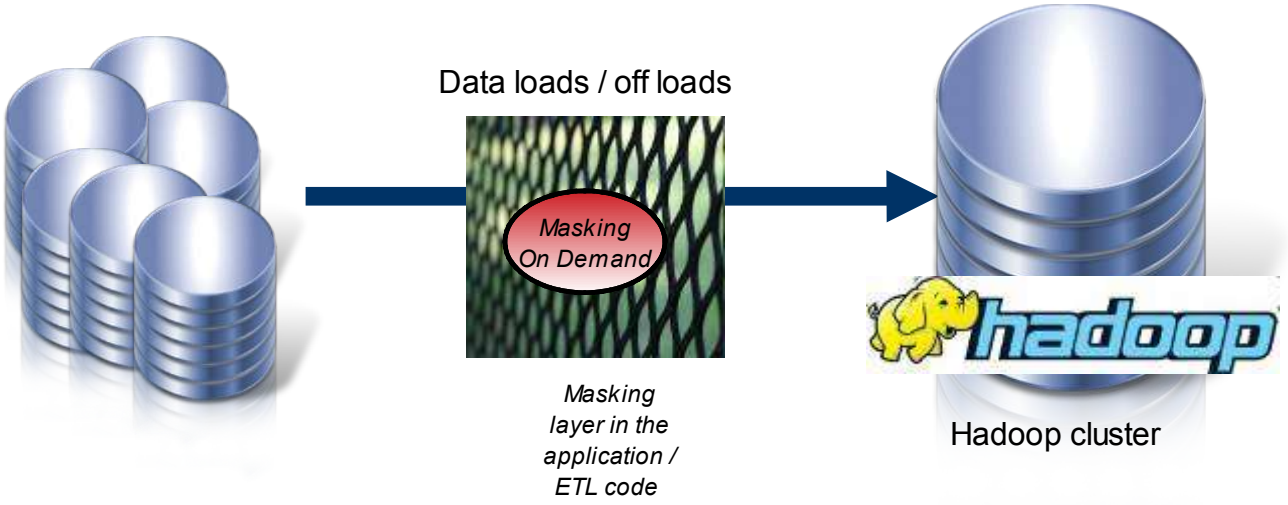
Test Data Management

Data Masking

Data Growth Management

Application Retirement

Ensuring Data Privacy and Security

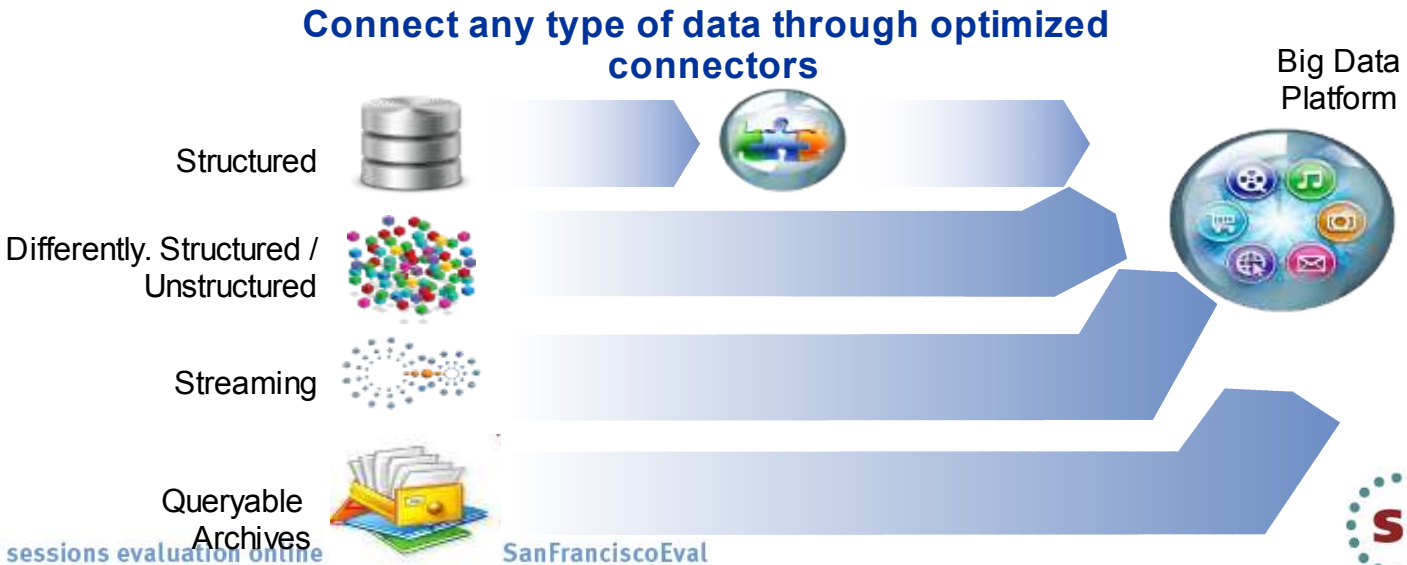


- **Usecase:**
  - Large scale analytics requires data from traditional sources to be combined with unstructured texual data to draw inferences
  - The analytics is predominantly trend analysis and individual data values are not extremely relevant.
- **Compliance risks:**
  - Data that is protected with 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.



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

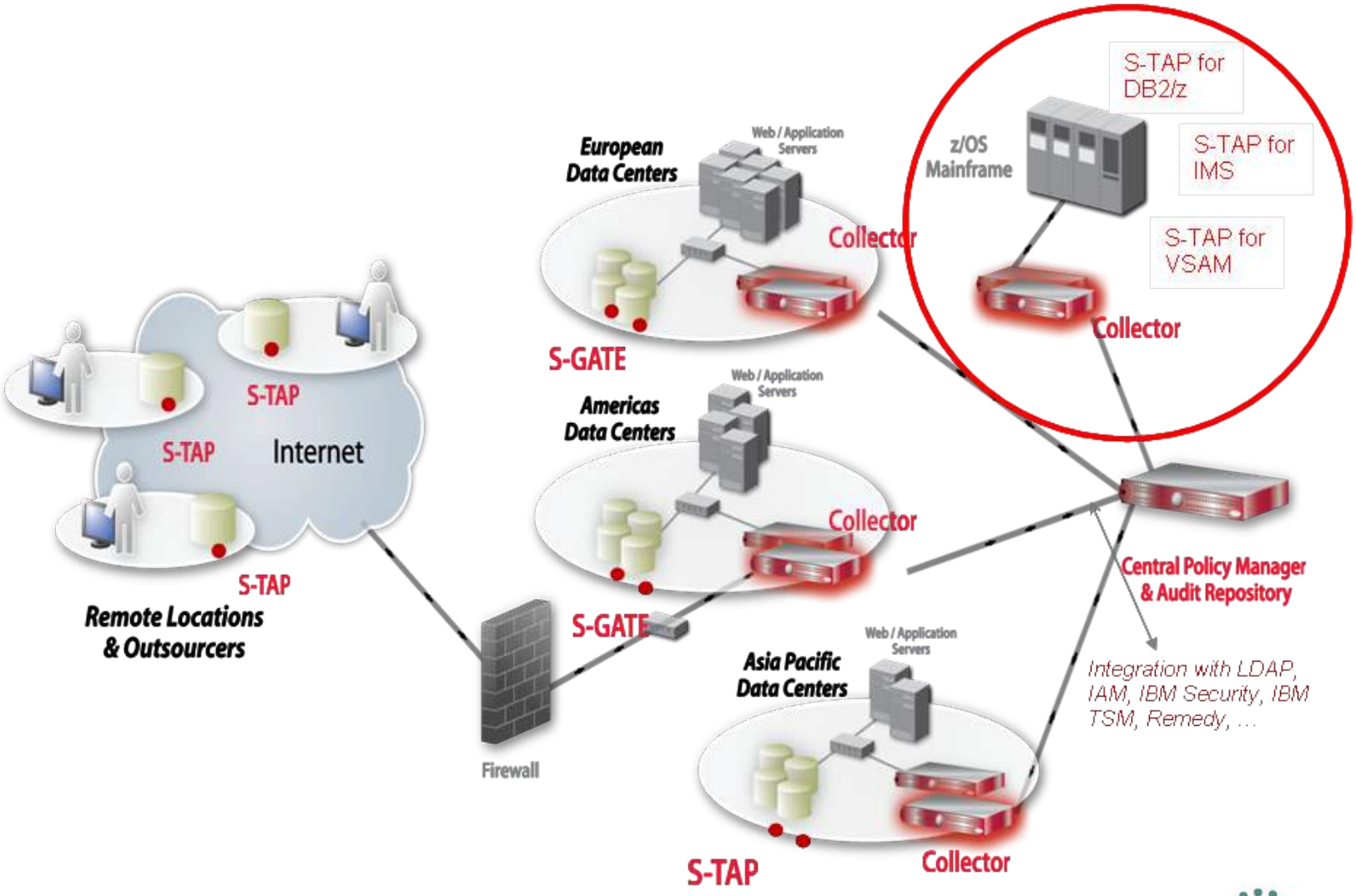
3.Test data management for big data projects

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

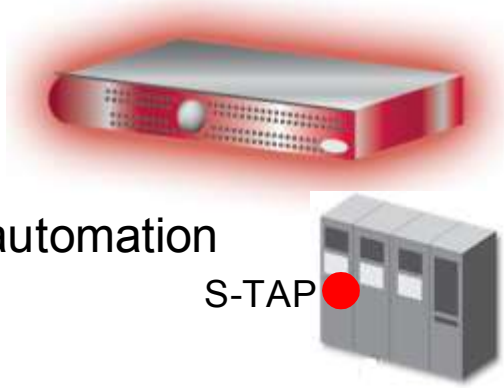
Advantages

- 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



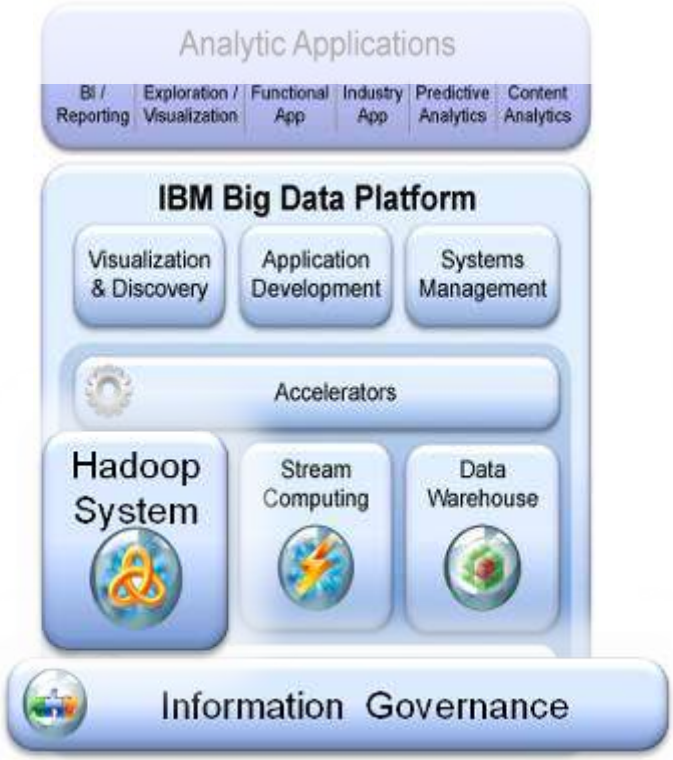
- Hardware or software appliances
  - Securely stores audit data collected by probes
  - Provides analytics, reporting & compliance workflow automation
    - Offloads audit data processing from mainframe
  - Integrated as part of the enterprise architecture
    - Centralized, cross-platform audit repository for enterprise-wide analytics and compliance reporting across System z & distributed environments

"We want Hadoop Activity Monitoring"

Monitor and Audit Hadoop activity in real-time to support compliance requirements and protect data

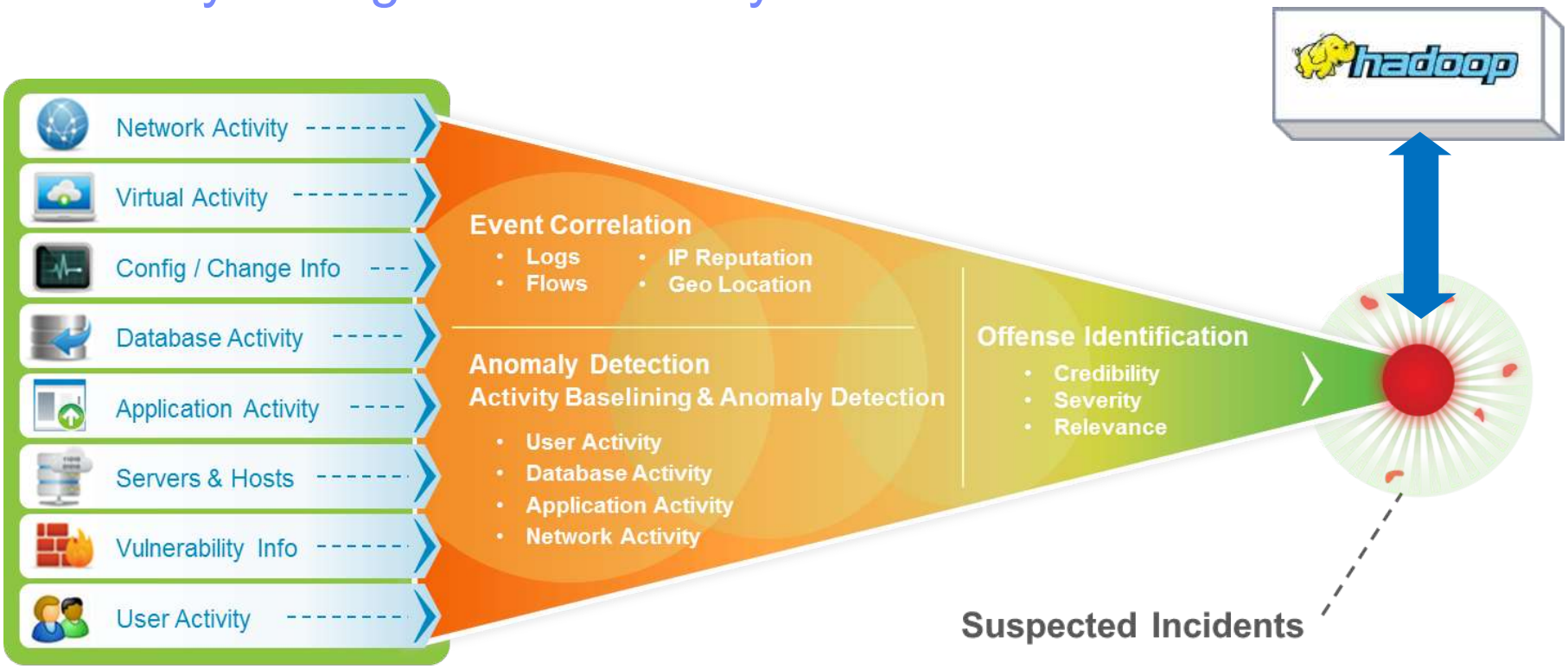
- Real time activity monitoring of HDFS and HBASE data sources
- Automated compliance controls
- Fully integrated with existing solution for database activity monitoring
- View Hadoop systems with other data sources

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**

Security Intelligence and Analytics vision



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*

- *Bance Carige Video*
  - www.youtube.com/watch?v=N7arHCI-CKk&list=PLIbvgyHNmNr1imFVsIskIpUuhBHkmNHr2&index=1

Big Data, IBM and System z – 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:**
  - IBM big data for z web site
  - Smarter Computing
  - Information Management System z
  - Information Governance Community
- **Education**
  - Free online education at bigdatauniversity.com
  - 20,000+ registered students
- **Further developments:**
  - Future webcast and announcements
- **Wanting to experiment on a big data integration project ?** ***Partner with IBM Silicon Valley Laboratory.***
- **Develop your own big data strategy** –Contact your local IBM sales representative to get started.

Useful URLs

- *www.ibm.com/bigdata/z*
- *www.ibm.com/smarterplanet*
- *www.ibm.com/software/os/systemz*
- *www.infogovcommunity.com*


THINK
BIG

THINK

#SHAREorg

SHARE
Technology · Connections · Results

Get the Latest on Big Data with IBM System z

Mark Simmonds
IBM Corporation

February 5, 2013
Session Number 12528

mtsimmon@us.ibm.com
linkedin.com/pub/mark-simmonds/11/596/14a/

SHARE in San Francisco 2013