

Building Confidence in Big Data

Innovations in Information Integration & Governance for Big Data



Important Disclaimer

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY.

WHILST EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED.

IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE.

IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION.

NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, OR SHALL HAVE THE EFFECT OF:

- CREATING ANY WARRANTY OR REPRESENTATION FROM IBM (OR ITS AFFILIATES OR ITS OR THEIR SUPPLIERS AND/OR LICENSORS); OR
- ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.

Success With Big Data Comes Down to Confidence

**Confidence in Your
Data**



**...Before you act
on insight**

**Confidence in
Accelerating Value**



**...Before you start a big data
project**

**Confidence in Your
Skills**



**...To maximize the
value from big data**



Today's Announcement Builds Big Data Confidence

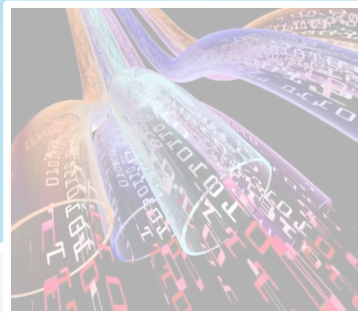
InfoSphere Integration & Governance for Big Data



Confidence in your Data

- Automated integration
- Visual context to understand data
- Agile governance to protect sensitive big data

PureData System for Hadoop



Confidence in Accelerating Value

- Appliance simplicity for Hadoop systems
- Appliance speed - Get up and running in hours

Big Data Stampede











Confidence in your Skills

- All the resources needed to get value from big data quickly
- Software, expertise and skills

IBM Big Data & Analytics Reference Architecture

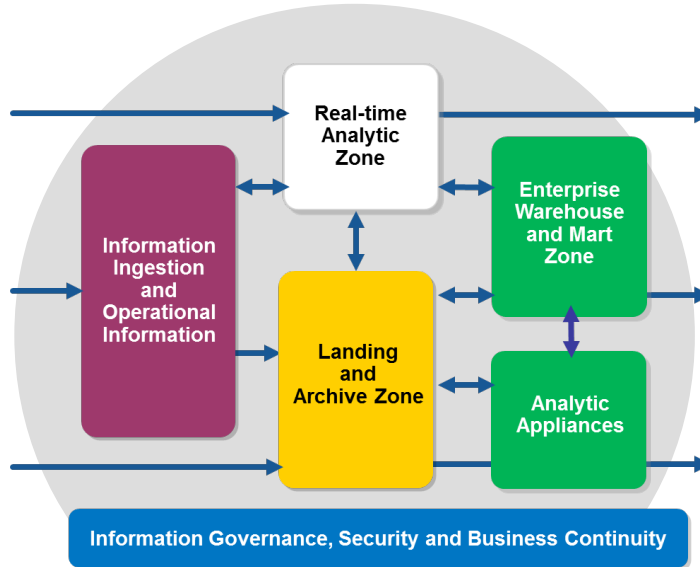


All Data Sources

-  Streaming Data
-  Text Data
-  Applications Data
-  Time Series
-  Geo Spatial
-  Video & Image
-  Relational
-  Social Network

Big Data Platform Capabilities








- Information Ingest
- Real-time Analytics
- Warehouse & Data Marts
- Analytic Appliances



Advanced Analytics/ New Insights

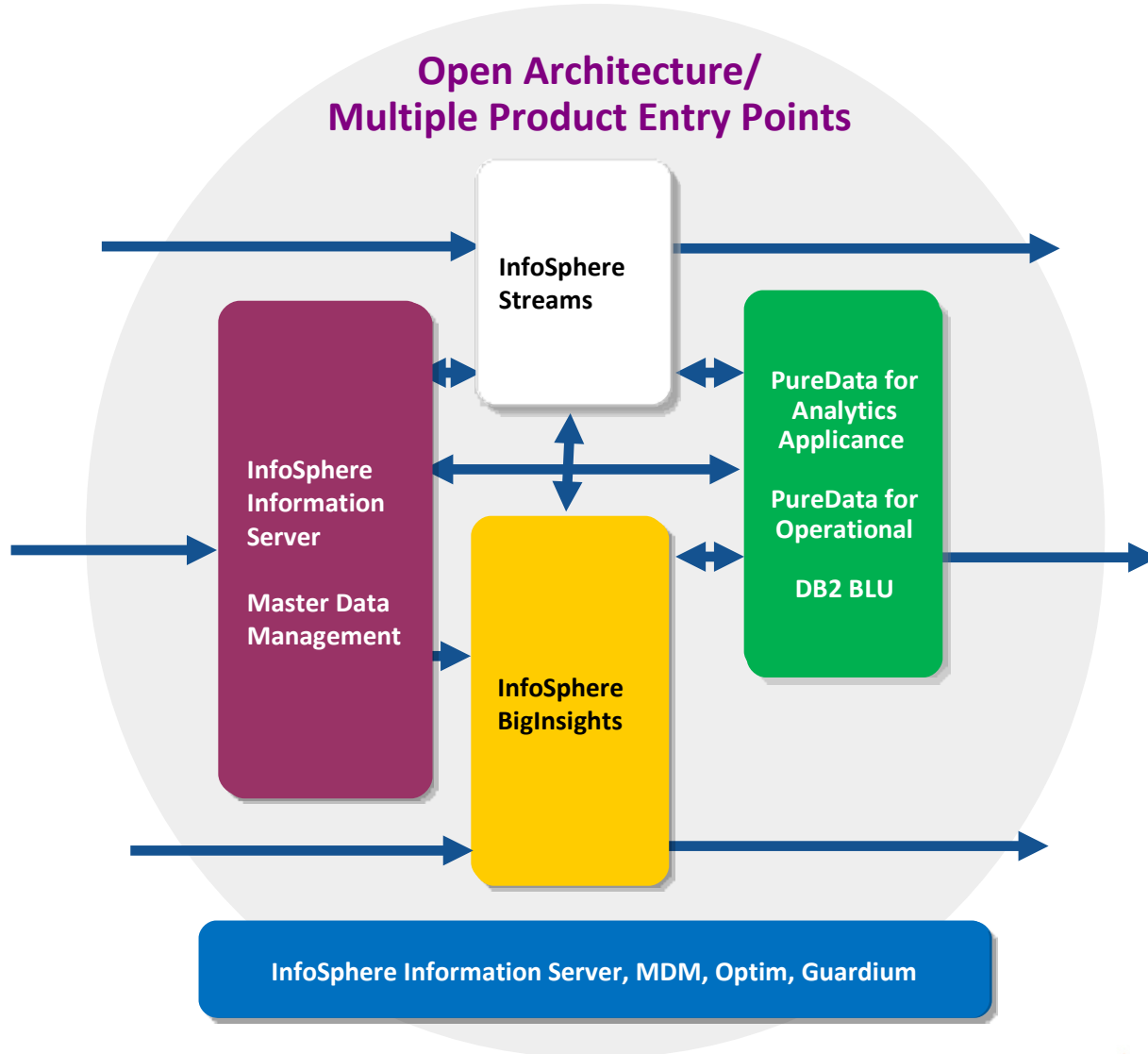
- Cognitive**
Learn Dynamically?
- Prescriptive**
Best Outcomes?
- Predictive**
What Could Happen?
- Descriptive**
What Has Happened?
- Exploration and Discovery**
What Do You Have?

New/ Enhanced Applications

- Watson** 
- Alerts** 
- Automated Process** 
- Case Management** 
- Analytic Applications** 
- Cloud Services** 
- ISV Solutions** 

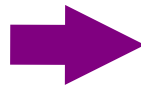


IBM Big Data & Analytics Reference Architecture



IIG Evolves to Address Big Data Challenges

How do I get access to new big data sources?

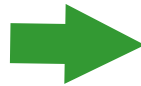


1

Automated Integration

Business users need rapid data provisioning among the zones

How do I digest all of this new information?



2

Visual Context

Categorize, index, and find big data to optimize its usage

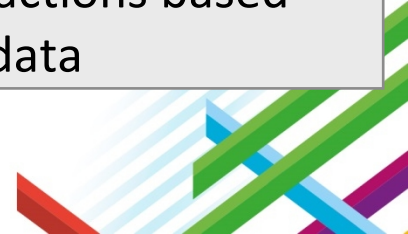
How do manage all of this new data?



3

Agile Governance

Ensure appropriate actions based on the value of the data



Six Innovations that Build Big Data Confidence

New

Automated Integration



Data Click

Self-service data provisioning for big data repositories

Big Match

Integration of master records from big data with probabilistic matching powered by Hadoop

New

Visual Context



Information Governance Dashboard

Visual context to give immediate status on governance policies

Big Data Catalogue

Categorization of metadata on all big data sources to find and leverage data

New

Agile Governance



Big Data Privacy & Security

Monitor and mask sensitive big data in Hadoop, NoSQL, and relational systems

Agile MDM for Big Data

Rapid onboarding and mastering of new big data sources and extension of 360° view with unstructured big data

InfoSphere Data Click Self-service Data Provisioning

New

**Automated
Integration**

Innovation

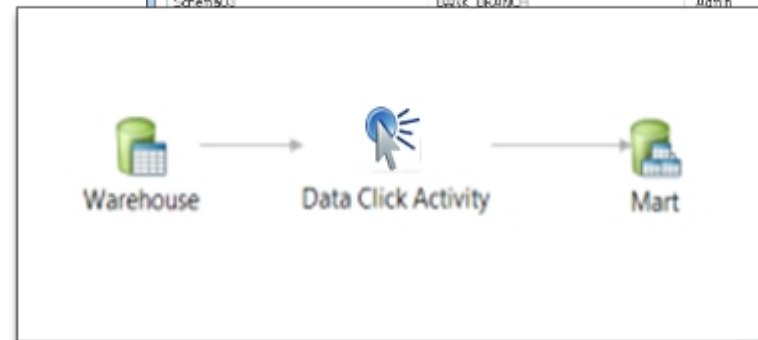
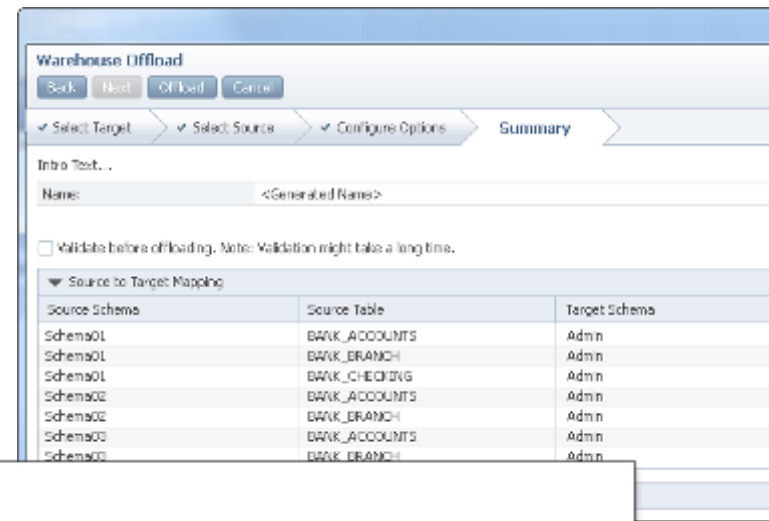
- Two-click data provisioning designed for business users
- Integration of more big data sources

Value

- Rapid provisioning of ad-hoc repositories
- Faster time to insight
- Self service to eliminate the IT bottleneck

Usage

- Enables rapid analysis of big data sources





Big Match: Find & Integrate Master Data in Big Data Sources



How It Works

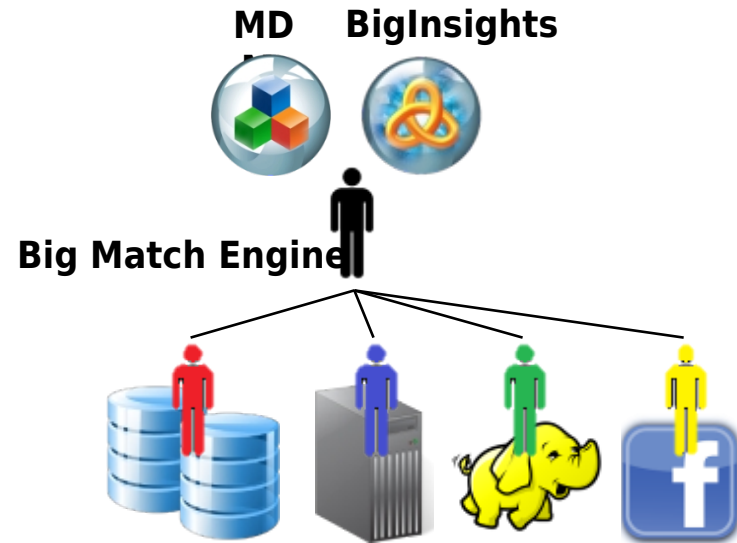
- Probabilistic matching on big data platform (BigInsights-Hadoop)
- Matching at a higher volume
- Matching of a wider variety of data sets

Client Value

- Find master data within big data sources
- Get an answer faster – enable real-time matching at big data volumes

Building Big Data Confidence

- Provides more context by detecting master entities faster



Information Governance Dashboard

Visualize and Control Governance

New

Visual Context

Innovation

- Indicators for policies and KPIs
- Rapid creation of tailored dashboards

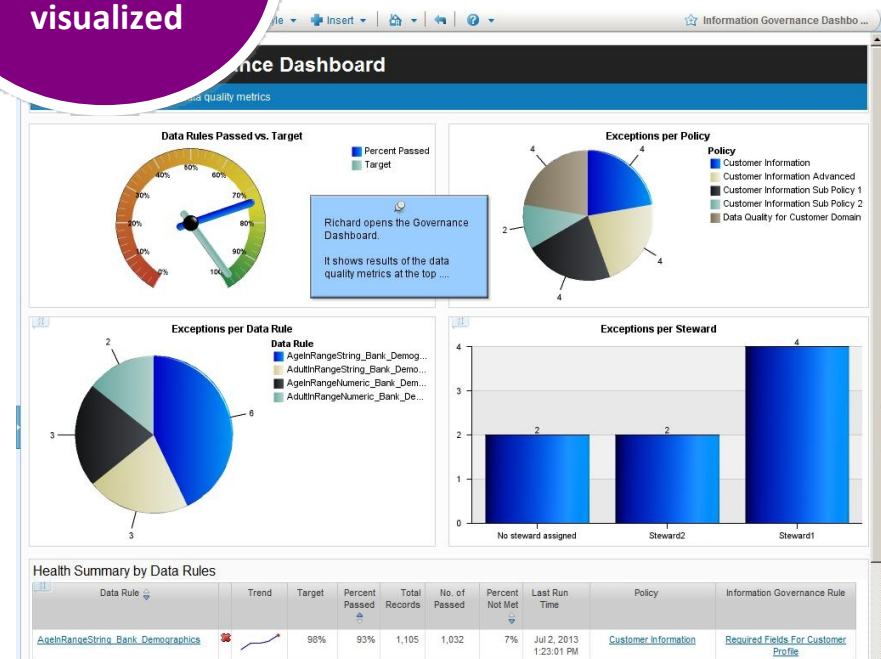
Value

- Immediate insight into governance policy status
- Interception of issues when they start, right at the source

Usage

- Raises data confidence with visual governance status

1000s of data points and policies visualized



Big Data Catalogue

Find Big Data More Easily

Innovation

- Stores metadata on every available big data source
- Provides structure to the Hadoop landing zone so data may be easily found and leveraged
- Classifies data (origin, lineage, source, value)

Value

- Find data more easily within a growing Hadoop landing zone and a complex zone architecture
- Rapidly leverage new big data sources

Usage

- Enables optimal usage of big data



Big Data Catalogue





Big Data Privacy and Security

Protect a Wider Variety of Sources

Innovation

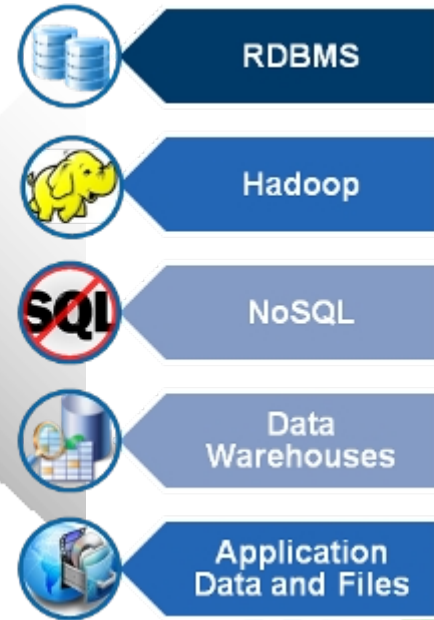
- Data activity monitoring of more NoSQL, Hadoop, and Relational Systems
- Masking of sensitive data used in Hadoop

Value

- Protection is a pre-requisite for the fundamental assumption of big data – sharing data for new insight
- Automation enables protection without inhibiting speed

Usage

- Ensures sensitive data is protected and secure





Agile MDM for Big Data: The Complete 360° View of Important Data



How It Works

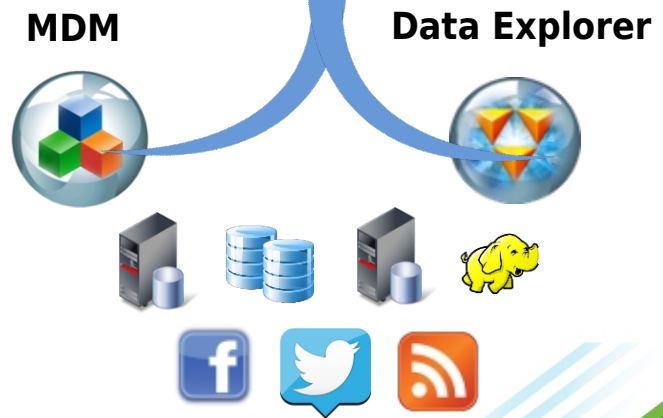
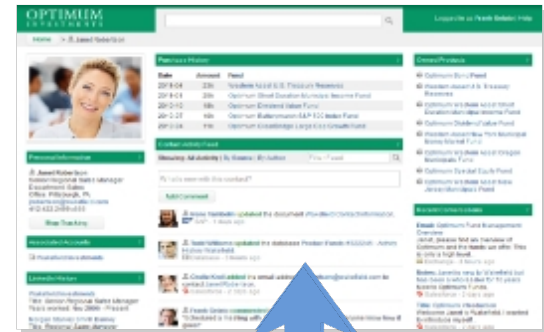
- Extend the master view with federated, unstructured big data
- Hybrid styles enable linking or consolidating source records based on confidence (Availability: July 2013)

Client Value

- Visualize every related data item in the 360° view
- Rapidly onboard new big data sources
- MDM adapts to the source

Building Big Data Confidence

- Provides a complete understanding of the customer or master entity



More Innovations that Build Confidence

New

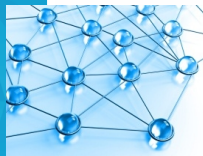
Automated Integration



- Real-time Hadoop replication
- Accelerated delivery for System z
- Certification on Hortonworks and Cloudera

New

Visual Context



- Enhanced performance for metadata ingestion
- Metadata APIs to enable application-specific dashboards & views
- Integration between InfoSphere Business Information Exchange and Data Explorer to identify and classify big data metadata

New

Agile Governance



- Hadoop integration with data lifecycle management—archive files on Hadoop
- Enhanced performance for monitoring reports
- Evolved governance practice—Agile process for the information governance community





Competitive Differentiators

✓ **Automated Integration**

- ✓ InfoSphere Data Click - No one else has delivered a feature that supports self-service data integration
- ✓ DB2 z optimization: InfoSphere Information Server - Full support for DB2 z is unique

✓ **Agile Governance**

- ✓ Hadoop based access to archived data - IBM is the only vendor who can generate immutable archive files & provide the same information in hadoop
- ✓ Masking confidential data and files for Hadoop projects - Most data security vendors don't have the breadth of capabilities and extensibility we offer

✓ **Visual Context**

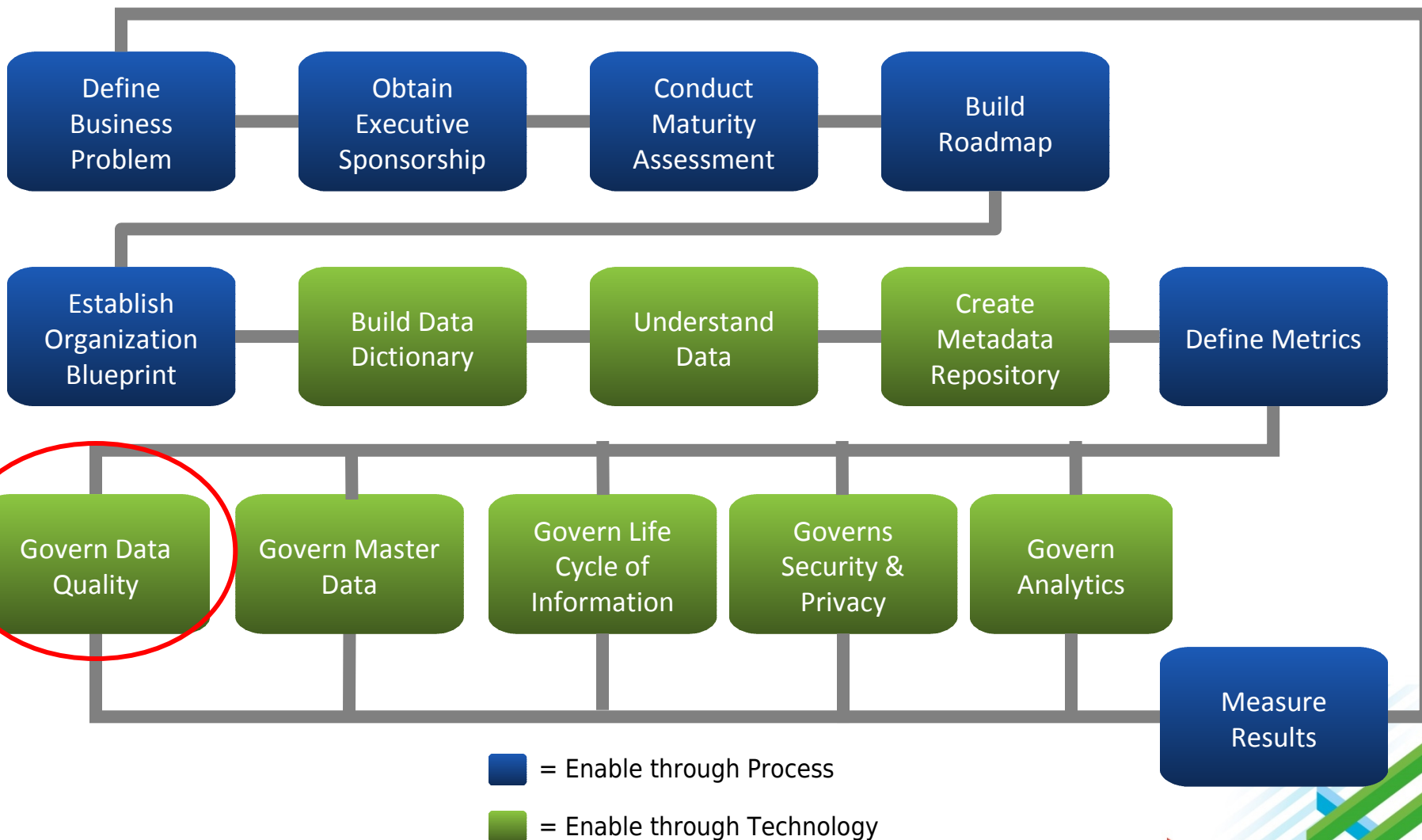
- ✓ Information Governance Dashboard - No other vendor offers a clear, complete view of governance policies and results
- ✓ MDM Matching on BigInsights - IBM is the only MDM vendor with integrated matching capabilities across traditional and big data platforms





IBM Information Governance Unified Process

Because Data Quality needs a repeatable Governance Process



IBM InfoSphere – Information Integration & Governance Platform for all Use Cases



Use Cases	InfoSphere Information Server	InfoSphere MDM	InfoSphere Optim	InfoSphere Guardium
Deliver Trusted Information for Data warehousing and business analytics	✓	✓	✓	✓
Build and manage a Single View	✓	✓	✓	✓
Consolidate and Retire Applications	✓		✓	✓
Make Enterprise Applications more Efficient	✓		✓	
Secure Enterprise Data and Ensure Compliance	✓	✓	✓	✓
Transform Enterprise Business Process & Applications with Trusted Information	✓	✓		
Integrate & Govern Big Data	✓	✓	✓	✓

Critical
 Complements

Consulting Firms talking....



It is already the case that stand-alone tools supporting specific data management disciplines are no longer good enough. What is needed now is a complete integrated suite of tools for end-to-end information governance that share metadata and that can be used together to systematically get information under control.

There is no question that IBM has understood the demand for this in the marketplace. They have invested heavily in information management over the last several years and have produced an integrated suite of tools in IBM InfoSphere Information Server and IBM InfoSphere Foundation Tools that lead the market in end-to-end structured information governance. In addition they have also integrated this platform with IBM MDM Server and IBM Cognos 10. This, coupled with shared metadata across the entire platform and full metadata lineage, makes IBM a certain shortlist contender for those companies trying to get their information under control.



**INTELLIGENT
BUSINESS
STRATEGIES**

IIG mastery helps transform the business

Figure 7

IIG Mastery Provides Business Transformation Potential For Big Data



Source: A commissioned study conducted by Forrester Consulting

In our research, 59 percent of organizations cited incompatible tools as one of the biggest barriers to a single, comprehensive version of business data.



V E N T A N A
R E S E A R C H

Gartner.

G00250592

Big Data Governance From Truth to Trust

Published: 18 July 2013

Analyst(s): Andrew White, Daniel Sholler

Business leaders recognize that big data can help deliver better business results through valuable insights. Without an understanding of the trust implicit in the big data (and applying information trust models), organizations maybe be taking risks that undermine the value they seek.

IBM BusinessConnect
Realize the art of the possible.



Thank You!



Andre De Locht

andre.de.locht@be.ibm.com

32-2-339-2554 M: 32-476 870 354

Brussels, BE IBM Sales & Distribution, Software Sales

Information Integration & Master Data Management Segment Leader - Europe,

Assistant: Christine Debecker