IBM SolutionsConnect 2013
Turning Opportunity into Outcomes.

From Operational Data to Trusted Big Data

IBM Information Governance Platform

Andre De Locht
Information Integration
Sr Business Consultant
☎ +32 476 870 354
✉ andre.de.locht@be.ibm.com

Information you can trust is everywhere....
But context matters !

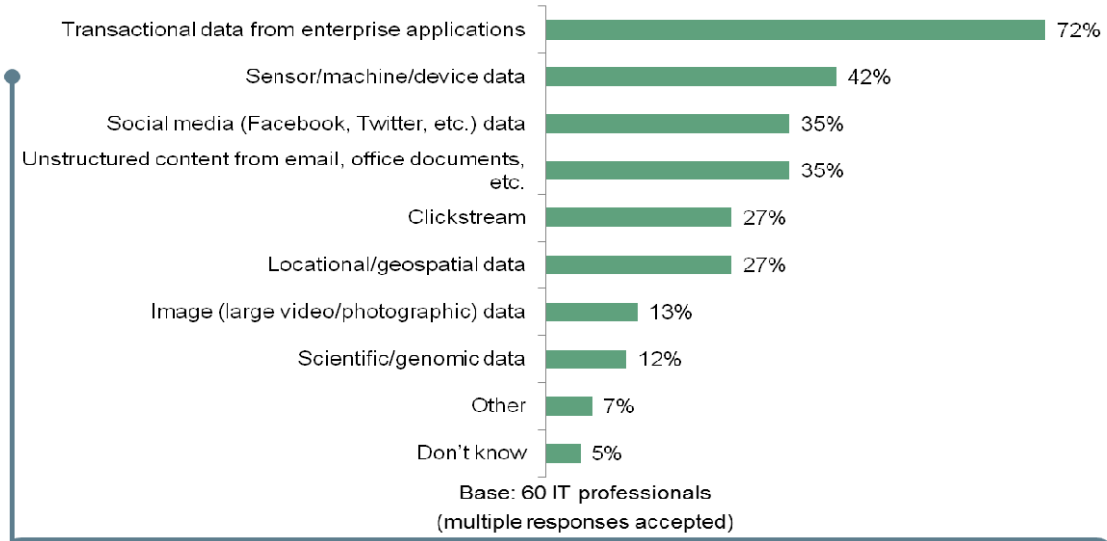
1. *Why are we here?*
2. *The Information Supply Chain*
3. *Talking about Business Value...*
4. *Trick or Treat!*

Why Are We Here?

Big Data is HOT, and Information Management is top of mind for clients…

Big data: across diverse subject domains

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

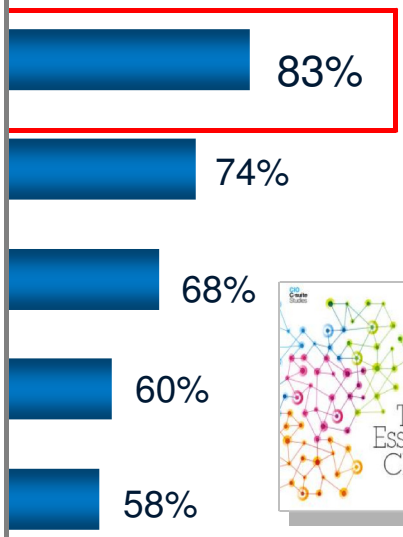


Base: 60 IT professionals
(multiple responses accepted)

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

and Market Growth



A recent Institute for Business Value study highlights how organizations are adopting big data in four phases

Big data adoption

Educate	Explore	Engage	Execute
Focused on knowledge gathering and market observations	Developing strategy and roadmap based on business needs and challenges	Piloting big data initiatives to validate value and requirements	Deployed two or more big data initiatives and continuing to apply advanced analytics
Percentage of total respondents 24%	Percentage of total respondents 47%	Percentage of total respondents 22%	Percentage of total respondents 6%

When segmented into four groups based on current levels of big data activity, respondents showed significant consistency in organizational behaviors

Total respondents n = 1061

Totals do not equal 100% due to rounding

“Data is the new Oil”

In its raw form, oil has little value. Once processed and refined, it helps power the world.

thenewfederalist.eu
Webzine of the Young European Federalists

Home page > Current Affairs > **Neelie Kroes: Information is the new oil!**

Neelie Kroes: Information is the new oil!

Interview of the European Commissioner for the Digital Ag

Wednesday 4 January 2012, by Laurent Nicolas

FINANCIAL TIMES
World business newspaper

easingly, businesses are applying
tics to social media such as
book and Twitter, as well as to
uct review websites, to try to
erstand where customers are

“Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”

Forb
“…now V
digesting
research,
clinical pr
outcomes
treating c

CNBC

“Data is the new oil.”
Clive Humby

If data is the new oil, we need a bigger drill

Posted on April 15, 2010 by Simon Kendrick

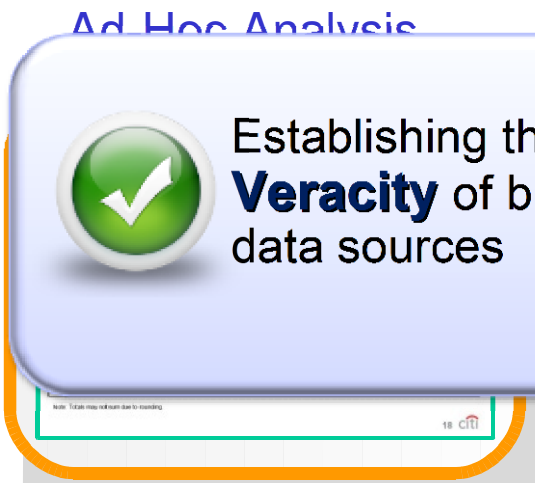
IBM delivers a governable, consumable Big Data platform that's steeped in analytics for data in-motion and data at-rest.

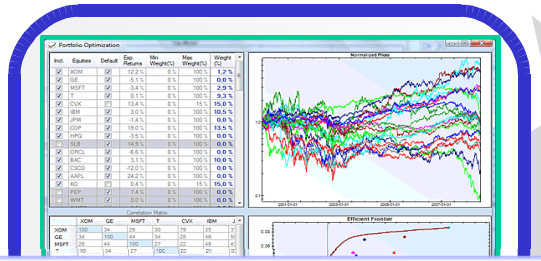
Every organization is on an analytics journey

Optimization

Predictive Analytics

BI Reporting and Ad Hoc Analysis





Establishing the **Veracity** of big data sources

1 in 3 business leaders don't trust the information they use to make decisions

- What happened?
- When and where?
- How much?

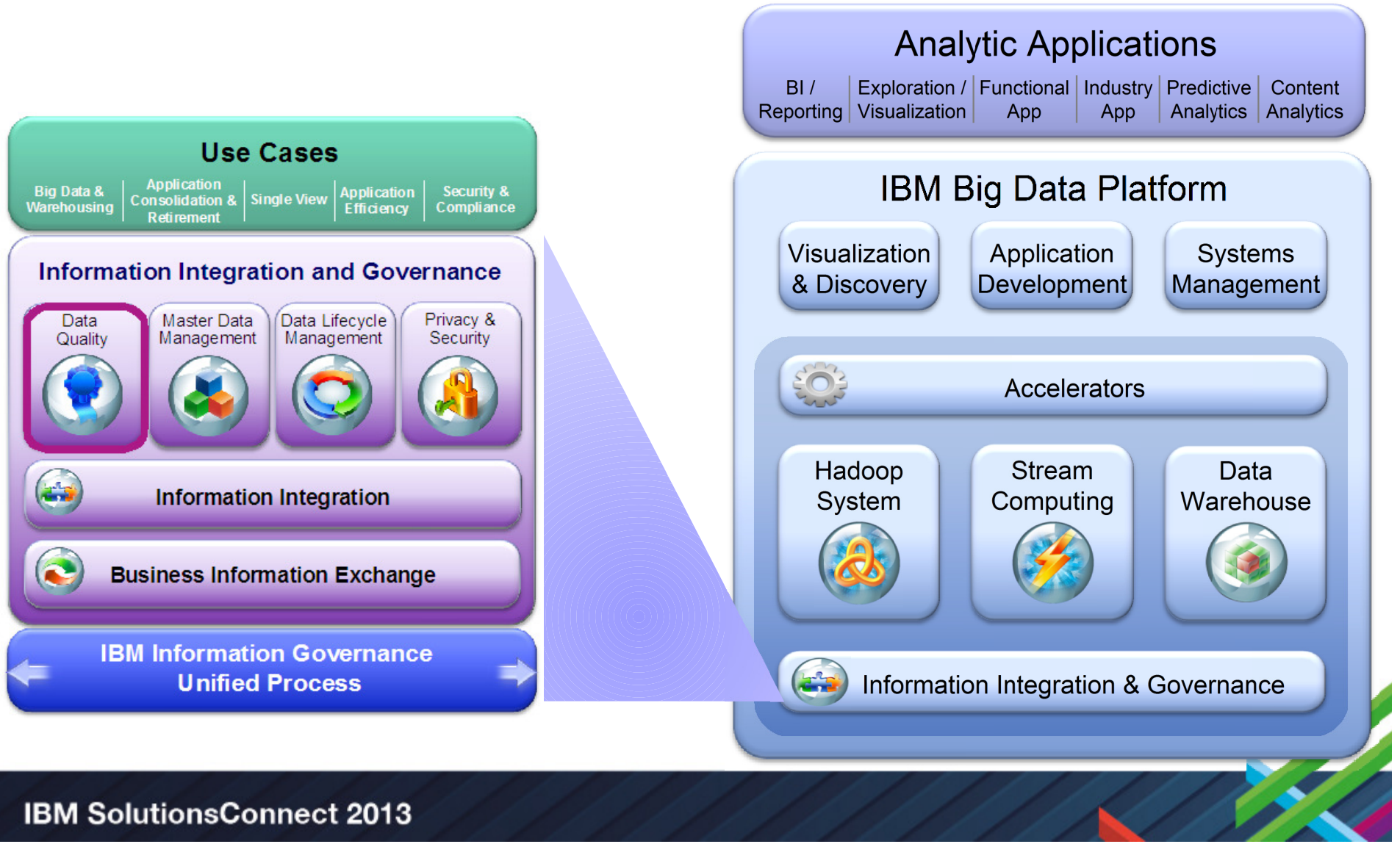
- What will happen?
- What will the impact be?

Cost efficiently processing the growing **Volume**

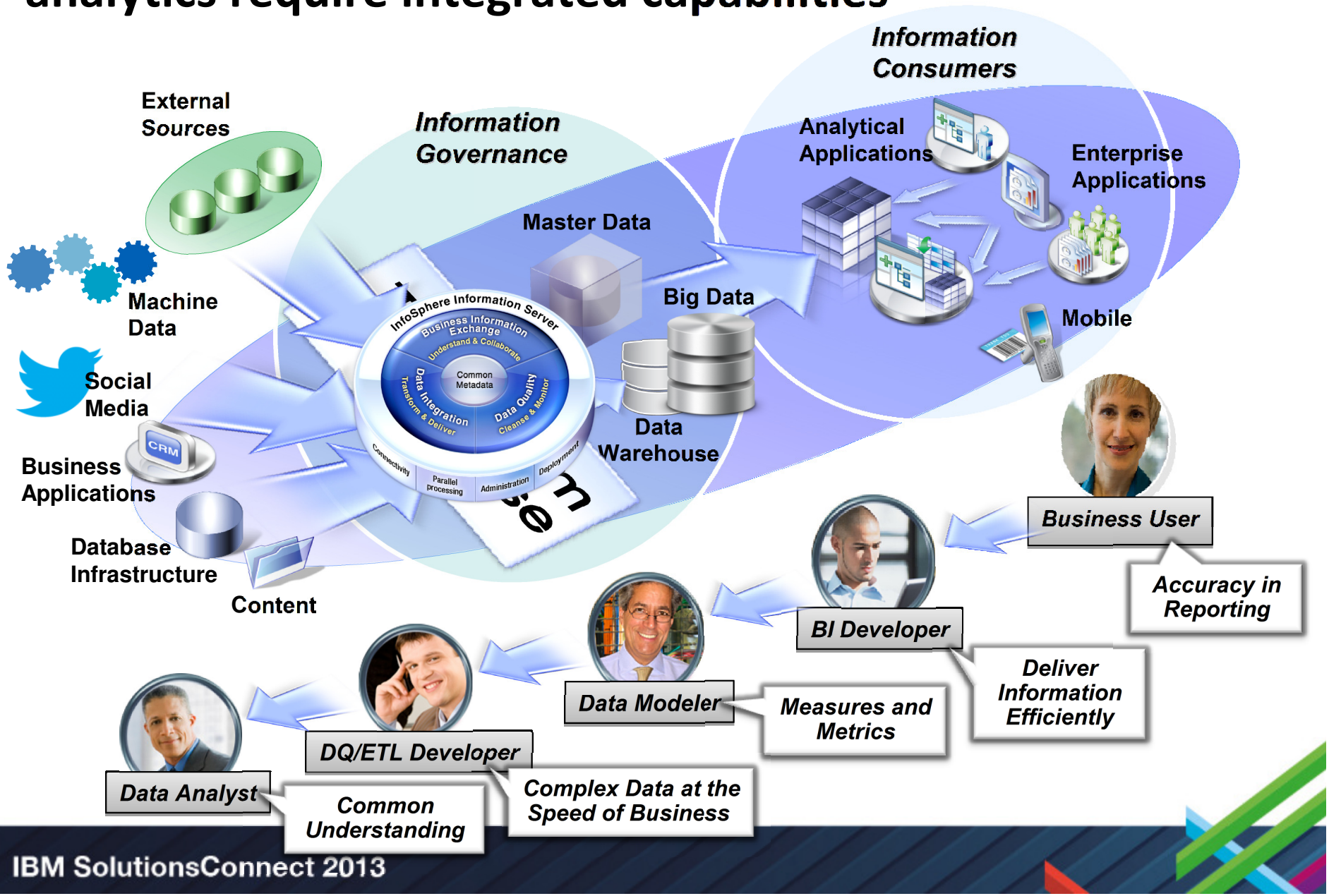
Responding to the increasing **Velocity**

Collectively analyzing the broadening **Variety**

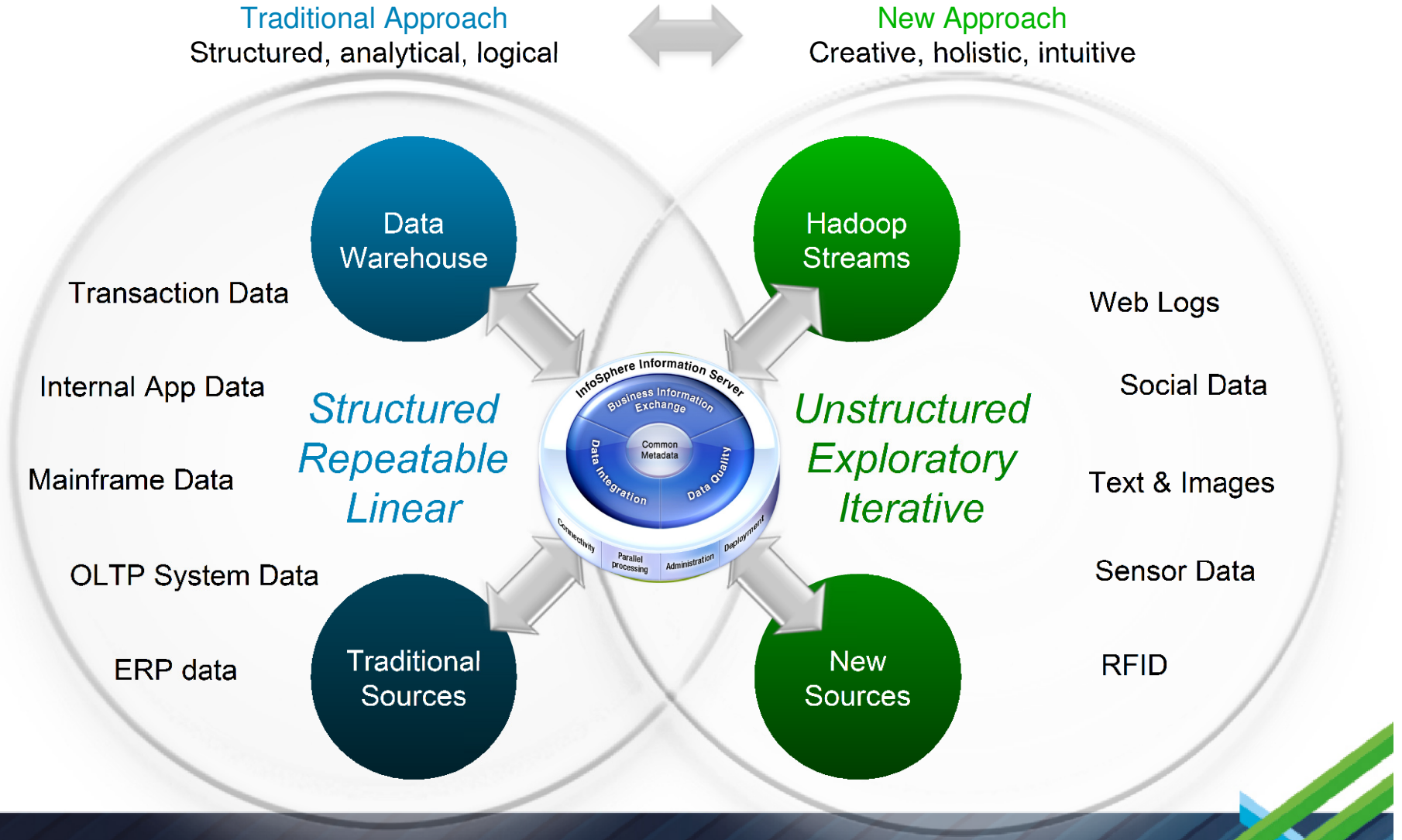
InfoSphere creates trusted information for use in key initiatives



Information Supply Chain - Because Enterprise analytics require integrated capabilities

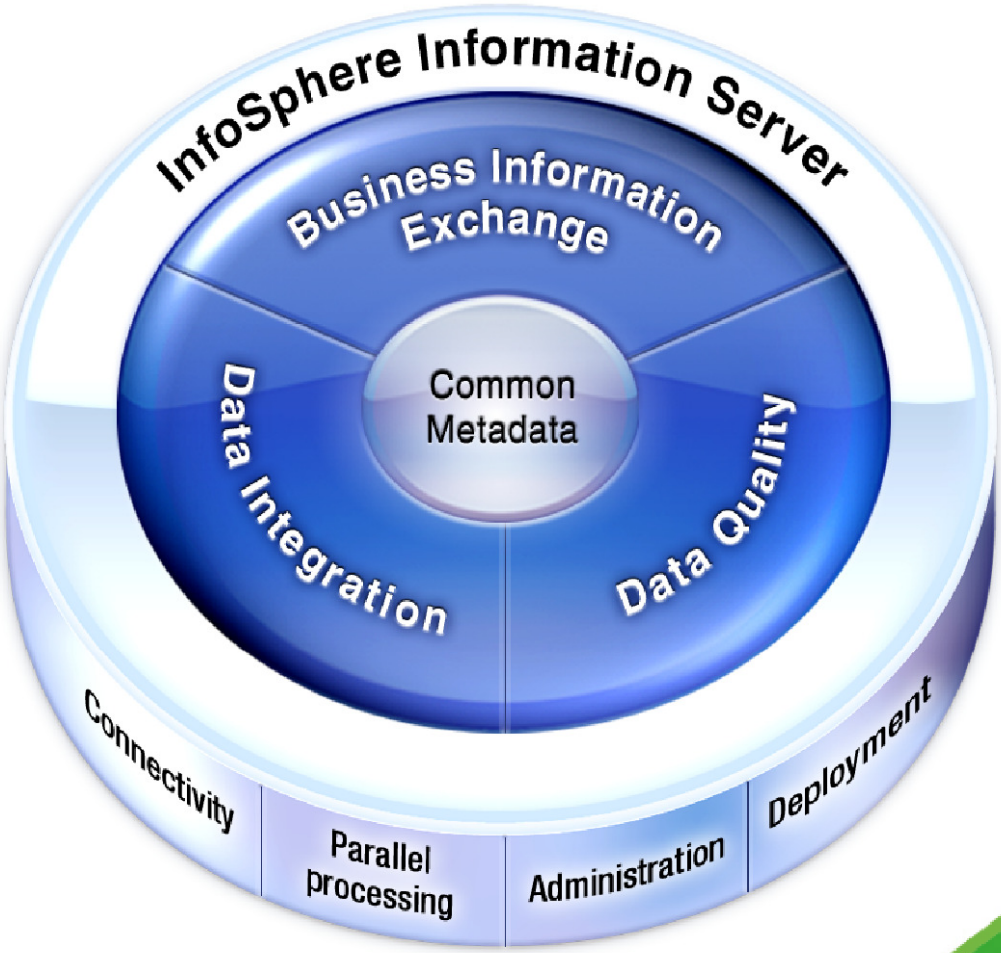


But Information Integration Requirements Are Becoming More Complex and Demanding on Infrastructure



The IBM approach to address each of the Requirements for Information Integration Integration and Governance: InfoSphere Information Server

- Integrating and transforming data and content
- to deliver accurate, consistent, timely and complete information
- on a single platform unified by a common metadata layer



Business Information Exchange: Understand Your Information, Transforming it into an Enterprise Asset

Three dimensions of understanding and governance:

Business

Gain and manage business perspective about information and align with IT

- leading technology for business-friendly access and pre-packaged terms
- time-saving Industry Models for warehouses in key industries

Process

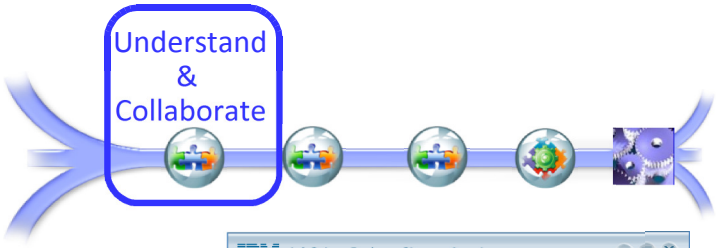
Guide projects with best practices to achieve goals with reduced risk

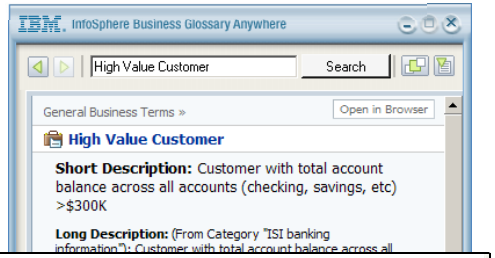
- unique capability to architect information projects with embedded methodology that can be tracked

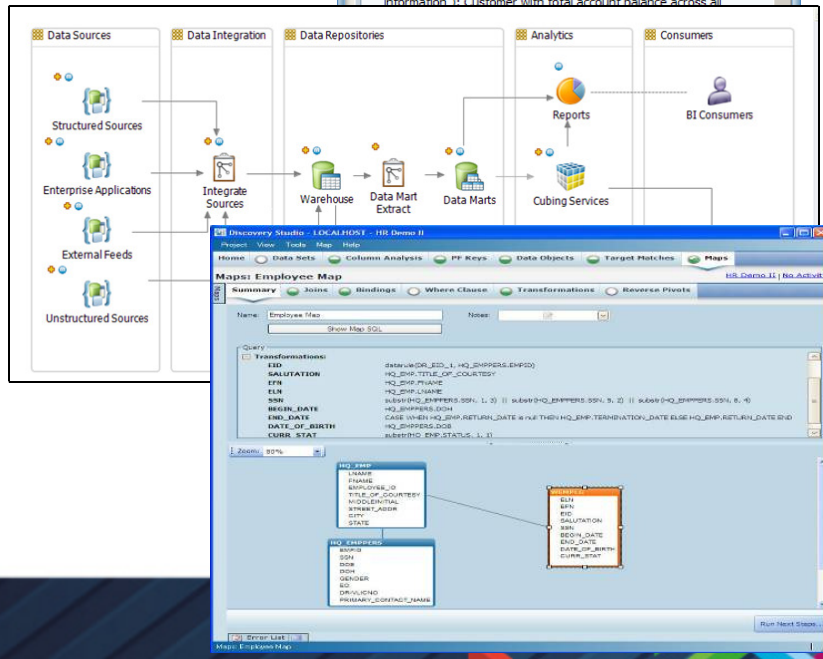
Technology

Discover data structures and understand your lineage to manage compliance

- unique capability for discovering business objects







Data Quality: Cleanse Data and Monitor Quality, Turning Data Assets into Trusted Information

Analyze data, cleanse data and control data quality

Analyze

Use source system analysis to understand your issues

- automated discovery of critical data and hidden data relationships

Cleanse

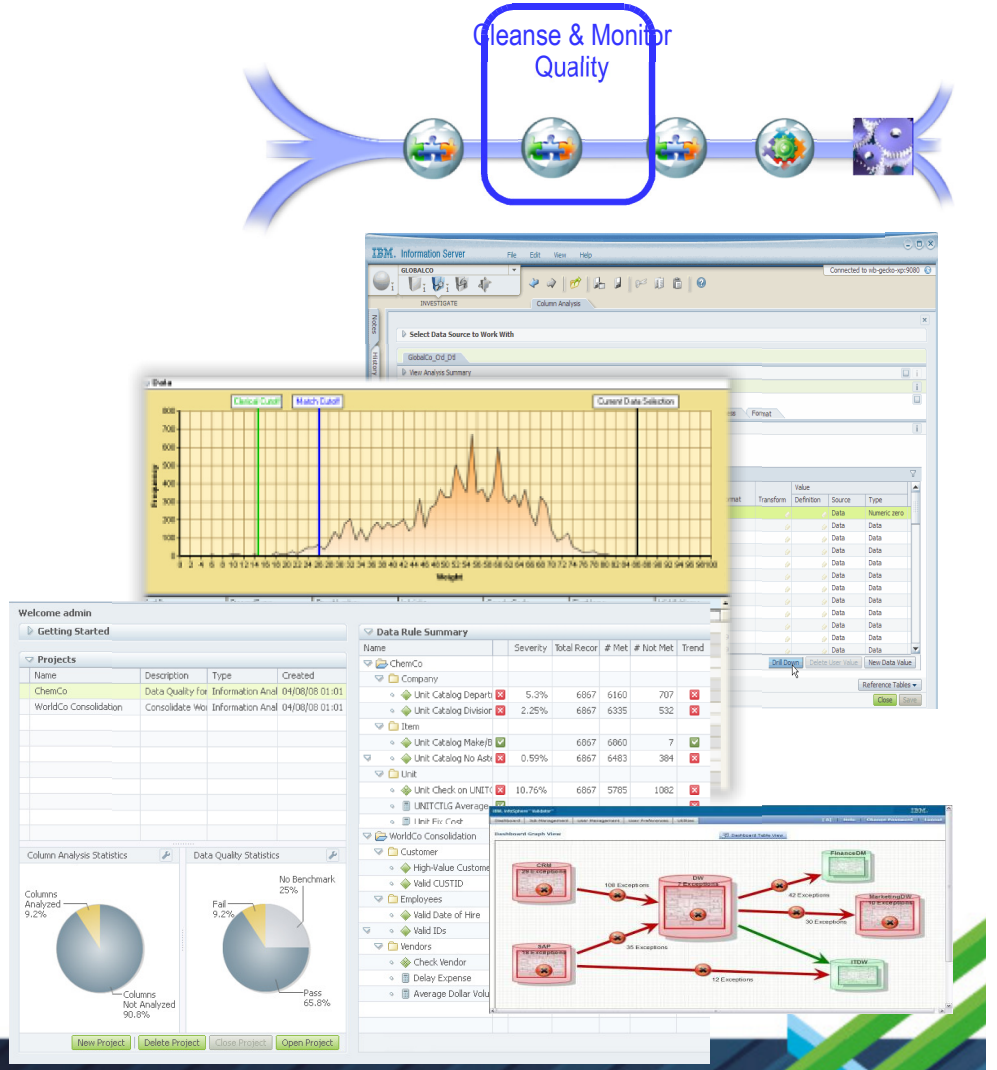
Investigate, standardize, match and survive data

- most comprehensive & customizable solution

Control & monitor quality

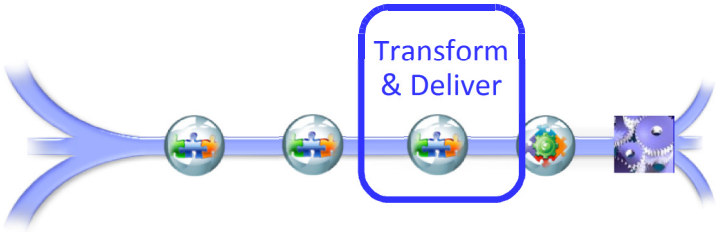
Assess and monitor the quality of your data in any place (database/or data flow) and across systems

- unique capability to align DQ metrics with business & governance objectives



Data Integration: Transform and Deliver Data to Any System, Improving Time to Value

Integrate and transform data on demand across multiple sources and targets …



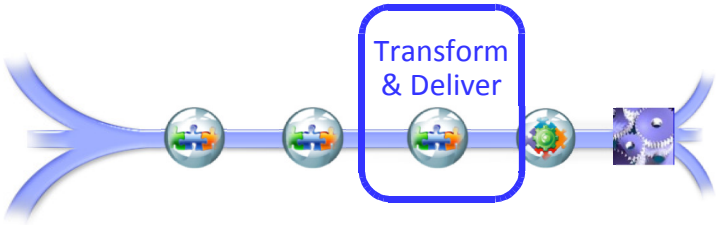
Transform

Satisfy the most complex transformation requirements with the most scalable runtime available

- Transform and aggregate any data volume
- Benefit from hundreds of built-in transformation functions
- Leverage metadata-driven productivity and enable collaboration
- Use a simple, web-based dashboard to manage your runtime environment
- Manage your requirements for transformation activities to align with the business

Data Integration: Transform and Deliver Data to Any System, Improving Time to Value

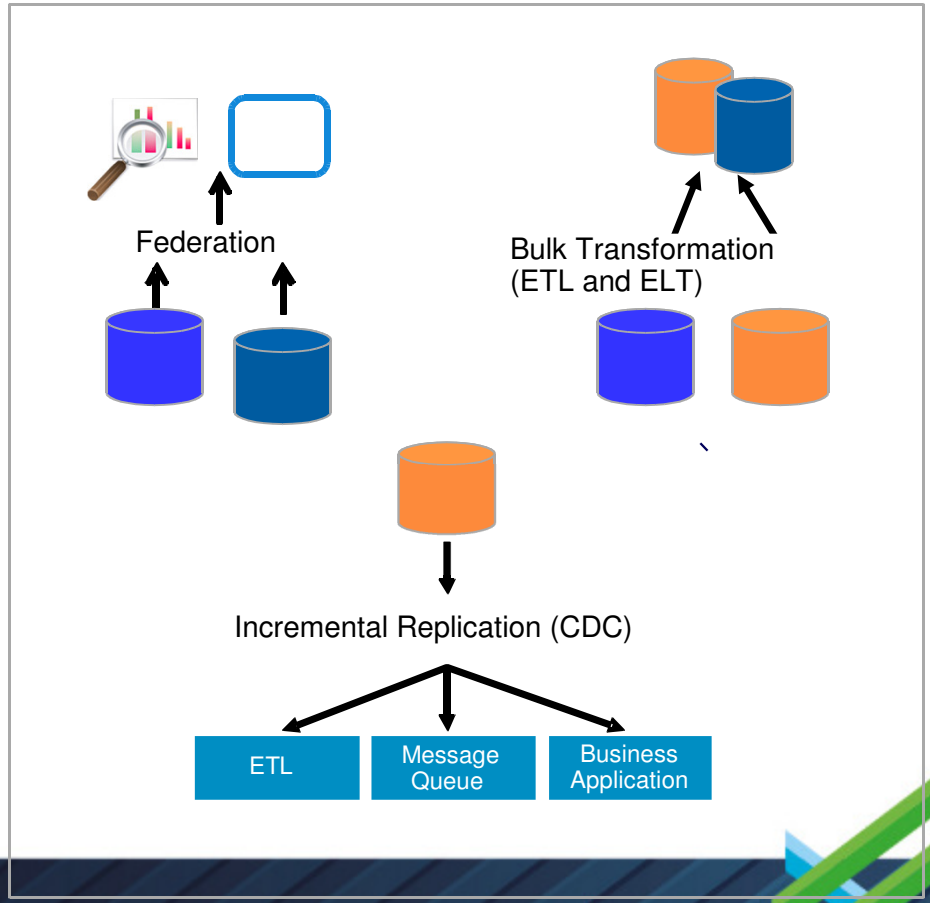
Deliver data efficiently according to your business requirements



Deliver

Leverage unique capabilities to:

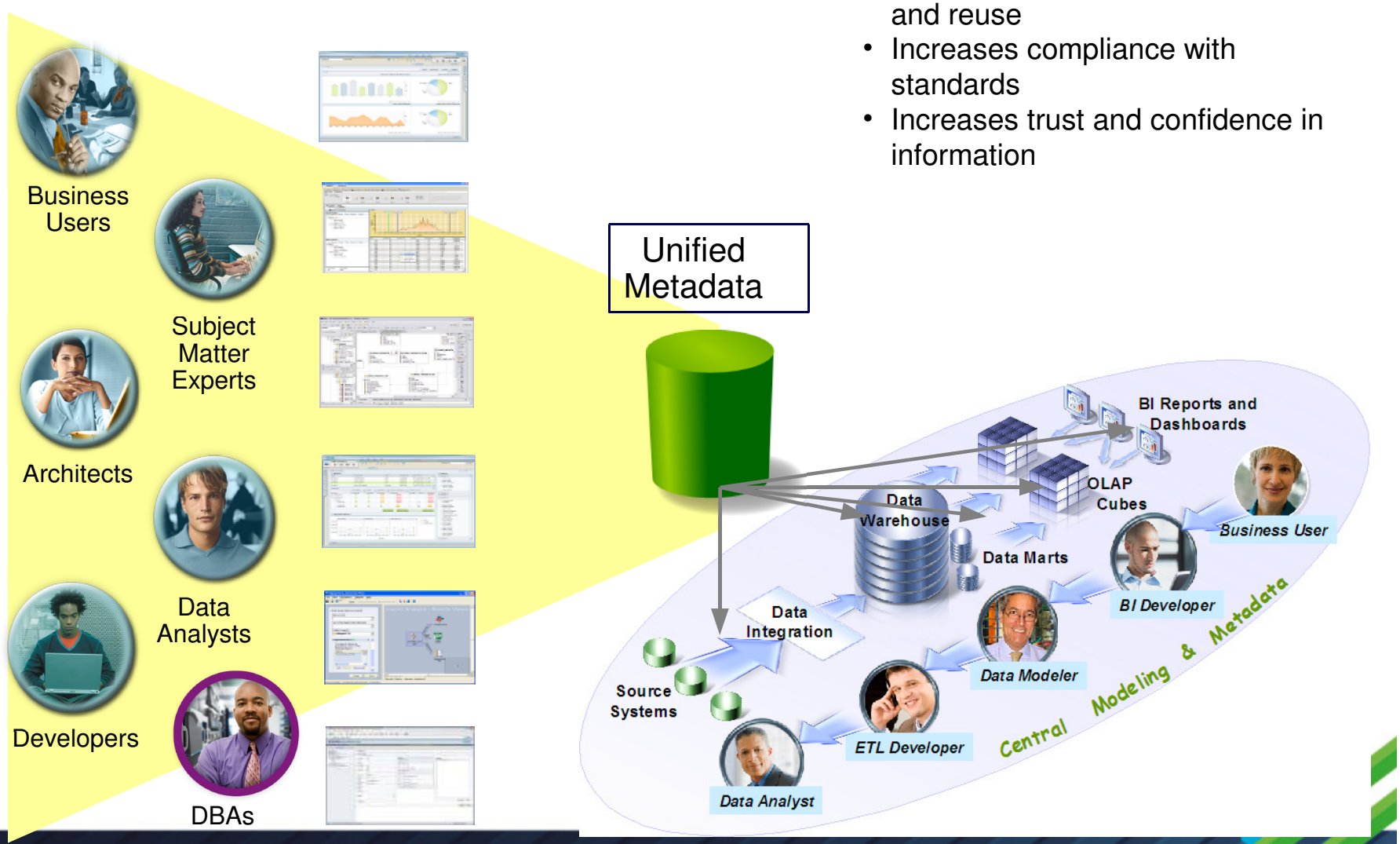
- Read from low-impact DBMS logs
- Apply any level of complex data integration/data quality logic
- Guarantee delivery through two-phased commit to one or more DBMS or MQ targets
- Support the most complex & challenging real-time integration requirements



Common Metadata Layer - Why It's Important

- Simplifies integration
- Facilitates change management and reuse
- Increases compliance with standards
- Increases trust and confidence in information

Different roles create and consume different metadata



InfoSphere Information Server v9.1: Anywhere Integration for a New Era of Computing

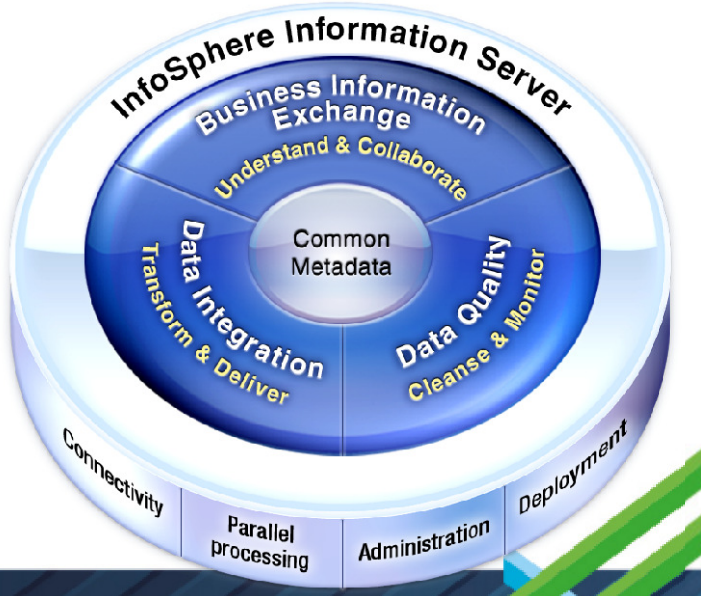
1 **Agile Integration** – *for faster, more flexible integration*

2 **Business-Driven Governance** – *for consistent rules and policies*

3 **Sustainable Quality** – *for greater accuracy and adaptability*

… to help our customers achieve:

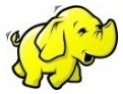
- Faster Time to Value
- Reduced Risk
- Lower TCO



“Anywhere” Integration Capabilities - Greater Agility to Get the Most from Your Environment

Unstructured Stage support
Provides native support for Excel spreadsheets data sources

InfoSphere Data Click
Achieve greater business agility in just a few clicks with simple self-service data integration capabilities on our rich governance framework



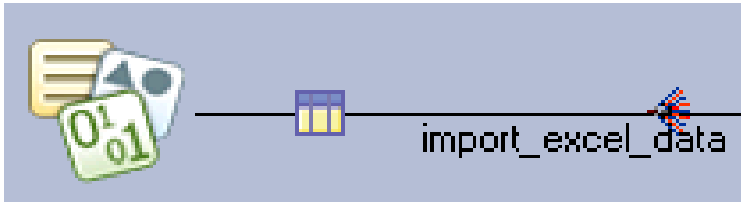
Enhanced support for big data
Integrate extreme variety, volume and velocity for both big data at-rest (Hadoop-based) as well as big data in-motion (stream-based)

Create business-driven information governance rules and policies
What is a customer? How do we manage customer data?
How do we ensure customer data complies with our information governance policies in North America? In Europe?

Assign terms to business processes
Drag and drop terms in your business process modeling environment to create elements and/or assign terms

Native Excel Access

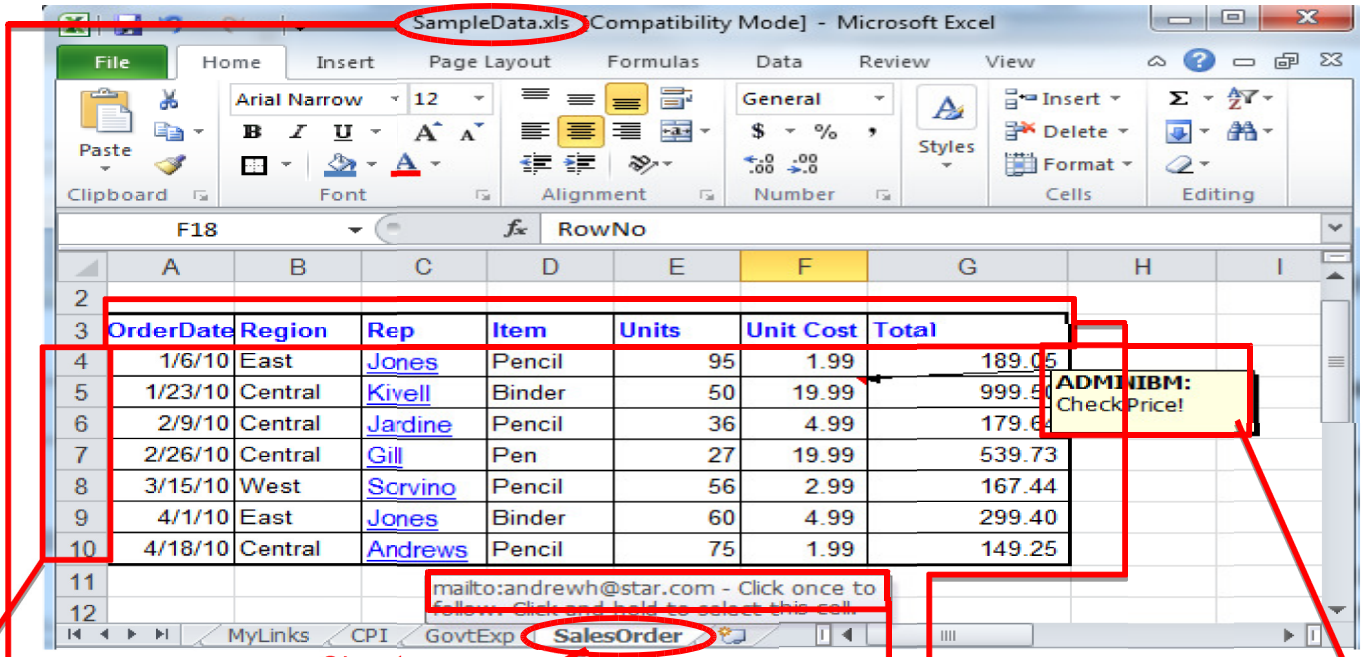
New Unstructured Stage provides….



- **3-Step Configuration** – enter file name, select data range and setup column name mapping
- **Native Excel File Support** – reads data from any supported OS (Windows, Linux, Unix)
- **XLS Version Variants** - includes both Excel 97-2003 OLE2 (.xls) and 2007 OOXML (.xlsx) file format
- **Simplified Access** - no need to define ODBC DSN for each excel file
- **Password Protected Files** – read data from password encrypted files
- **Multiple File Support** – ability to simultaneously process multiple files in different nodes
- **Multiple Sheet Support** - extract data from multiple sheets at a time
- **Runtime Column Propagation** – define column metadata
- **Excel Field Extraction** – ability to extract row numbers, comments, hyperlinks, formulas, document Properties, etc...

Excel Access

Excel



	A	B	C	D	E	F	G
3	OrderDate	Region	Rep	Item	Units	Unit Cost	Total
4	1/6/10	East	Jones	Pencil	95	1.99	189.05
5	1/23/10	Central	Kivell	Binder	50	19.99	999.5
6	2/9/10	Central	Jardine	Pencil	36	4.99	179.64
7	2/26/10	Central	Gill	Pen	27	19.99	539.73
8	3/15/10	West	Sorvino	Pencil	56	2.99	167.44
9	4/1/10	East	Jones	Binder	60	4.99	299.40
10	4/18/10	Central	Andrews	Pencil	75	1.99	149.25

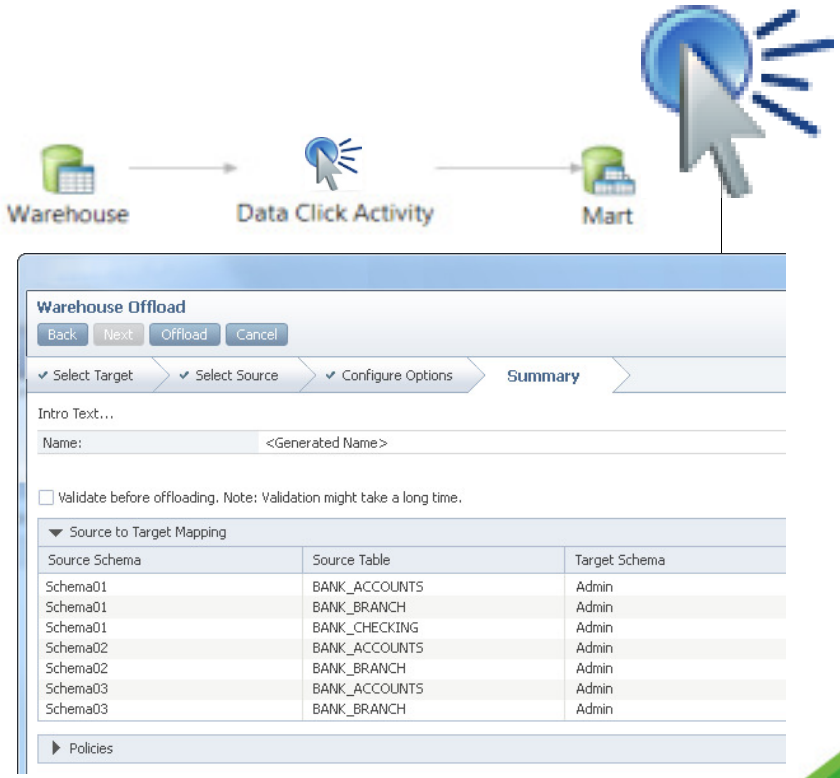
DataStage

RowNo	File	Sheet	OrderDate	Region	Rep	Email	Items	Unit	Unit Cost	Comment	Total
4	SampleData.xls	SalesOrder	1/6/10	East	Jones	mailto:jonesp@star.com	Pencil	95	1.99		189.05
5	SampleData.xls	SalesOrder	1/23/10	Central	Kivell	mailto:Kivell@star.com	Binder	50	19.99	Check Price!	999.5
6	SampleData.xls	SalesOrder	2/9/10	Central	Jardine	mailto:jardinej@star.com	Pencil	36	4.99		179.64
7	SampleData.xls	SalesOrder	2/26/10	Central	Gill	mailto:pgill@star.com	Pen	27	19.99		539.73
8	SampleData.xls	SalesOrder	3/15/10	West	Sorvino	mailto:sorvino@star.com	Pencil	56	2.99		167.44
9	SampleData.xls	SalesOrder	4/1/10	East	Jones	mailto:jonesp@star.com	Binder	60	4.99		299.4
10	SampleData.xls	SalesOrder	4/18/10	Central	Andrews	mailto:andrewh@star.com	Pencil	75	1.99		149.25

Agile Integration Capabilities - Introducing InfoSphere Data Click

Business Value: Speeds time to value and increases business agility by shrinking the time required to complete tasks – from days and weeks to minutes and hours.

- **Enables novice users to perform data provisioning, with just two 'clicks'**
  - Simple UI choices - automated, without coding
  - Both design and operational metadata for built-in governance
  - Optimized batch and real-time runtimes
- **First release focuses on accelerating PureData for Analytics deployments**
- **IBM is the only vendor to tackle the data provisioning problem with end-to-end best-of-breed capabilities & built-in governance**

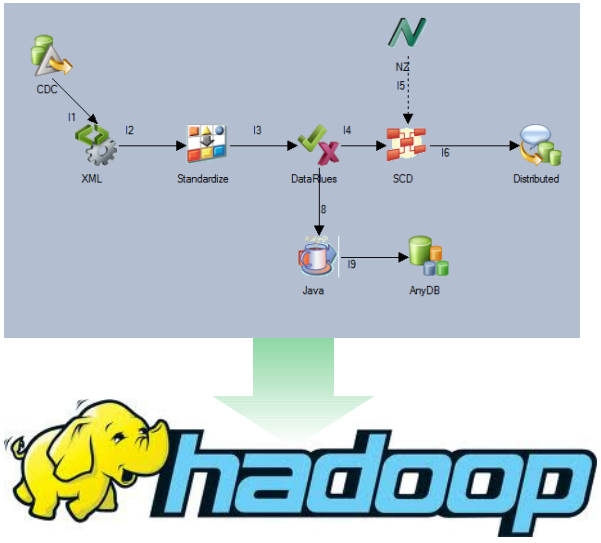


Faster Time to Value

Agile Integration Capabilities – Enhanced Big Data Support

Business Value: Allows organizations to accelerate their "massive scale analytics" agenda by combining our MPP integration platform with the power of Hadoop

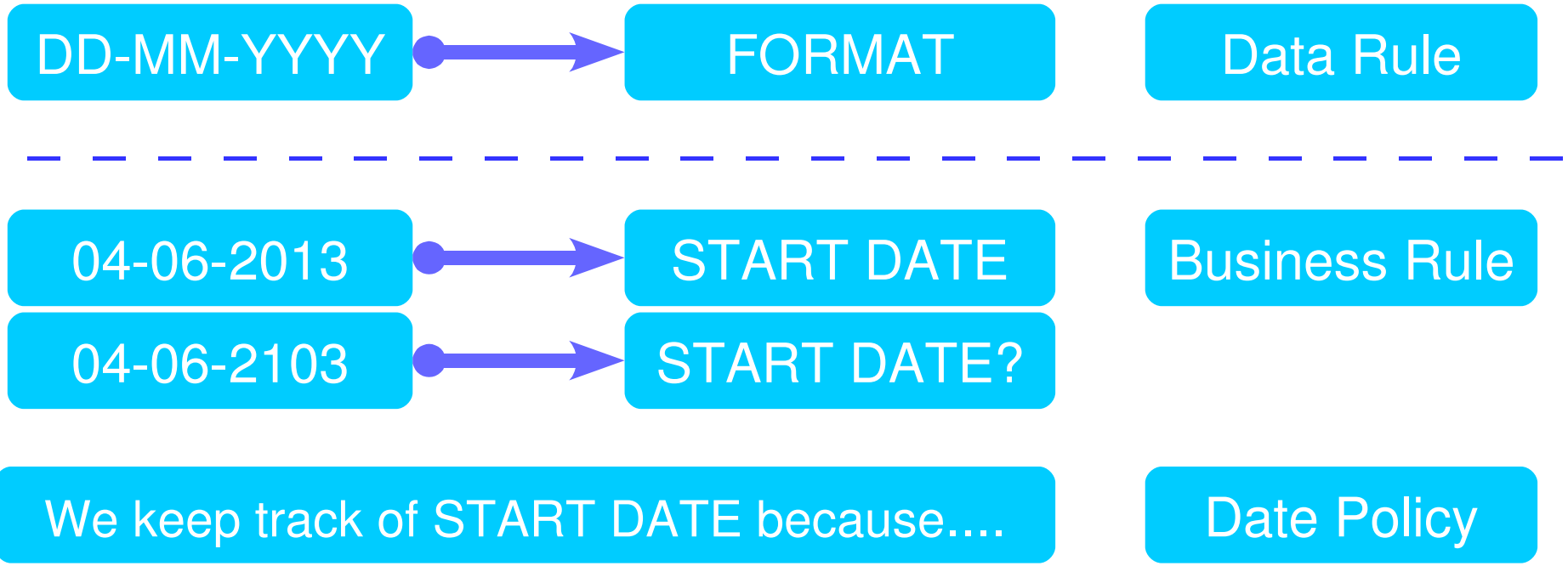
- **Combined workflows**
  - Easily mix traditional data integration workflows with analytical activities happening in Hadoop
- **Balanced Optimization for Hadoop**
  - Build data integration jobs for Hadoop in the same way as traditional ETL to leverage the locality of information processing for big data
- **Real-time analytics processing support**
  - Direct integration with InfoSphere Streams to combine the power and reach of both platforms
- **Big Data governance**
  - Provide lineage metadata information for big data file system



IBM is the only vendor with such a broad set of capabilities for big data at-rest and in-motion.

Lower TCO

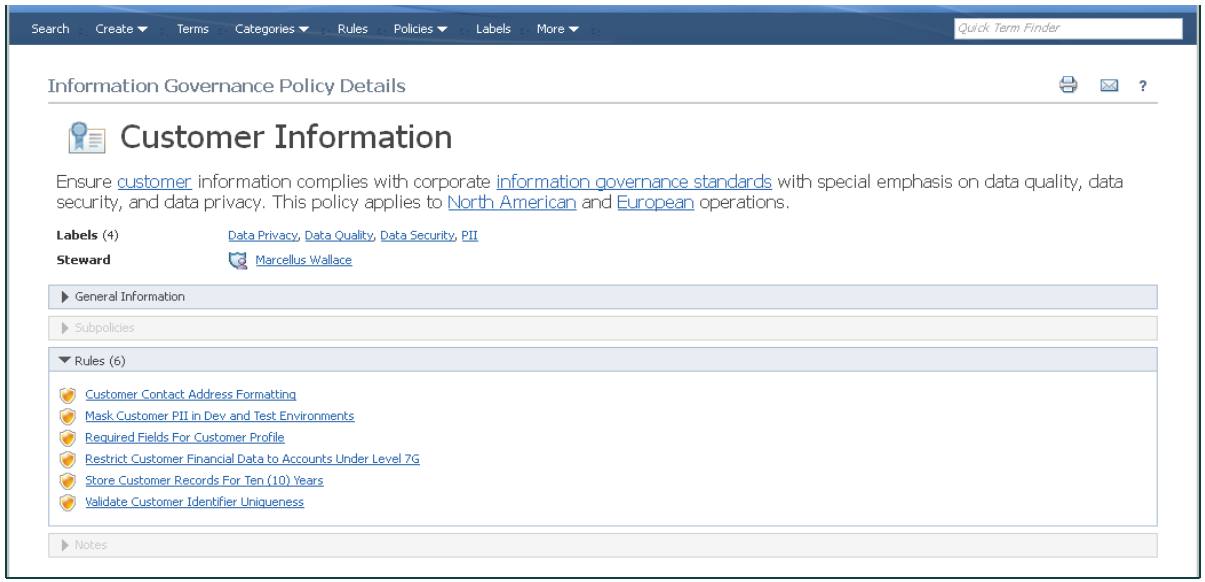
Business-Driven Governance Capabilities - Why Is Business-Driven Governance Important?



Business-Driven Governance Capabilities - Create Business-Driven Policies and Rules

Business Value: Gain greater control of and have more confidence in the information that powers your business.

- **Policies define areas of information governance**
- **Information governance rules describe business expectations on information**
- **This capability provides a practical starting point for information governance initiatives**
- **IBM is the only vendor that offer deeply integrated business policies & IT rules**

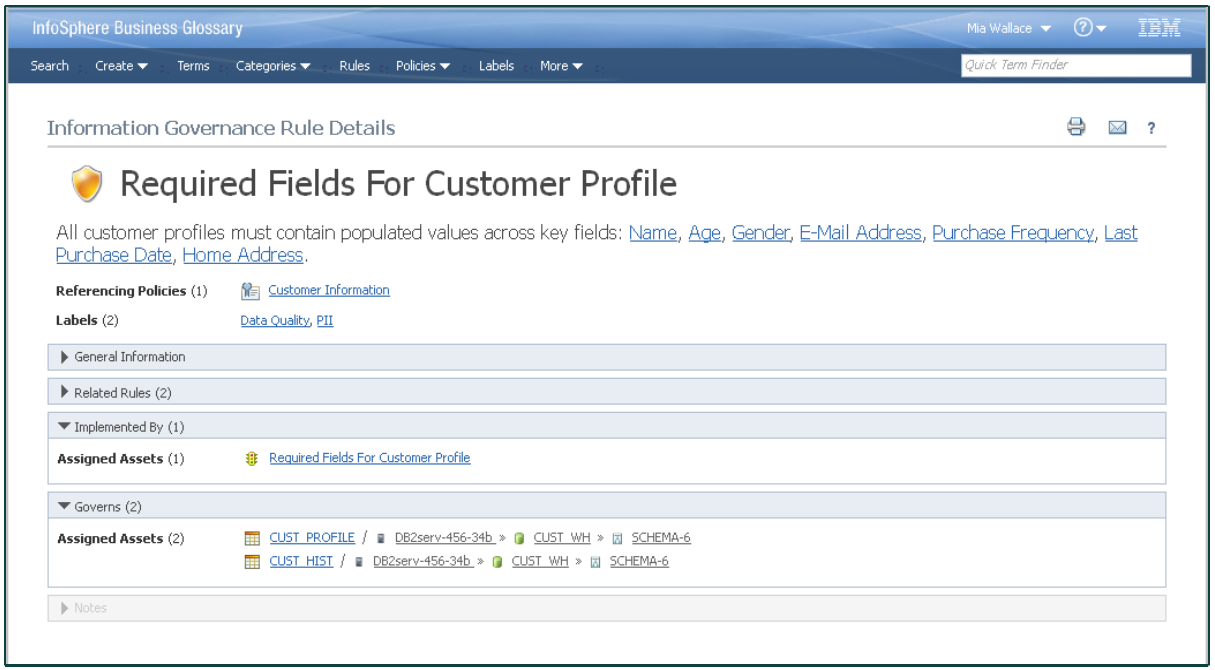


Reduced Risk

Business-Driven Governance Capabilities – Align Implementations with Business Requirements

Business Value: More collaboration and alignment between LOB and IT teams increases accuracy, lowers risk, and ensures compliance.

- **Implemented By: Indicates the operational assets that implement this rule**
- **Governs: Indicates that the data sources should comply with this rule**

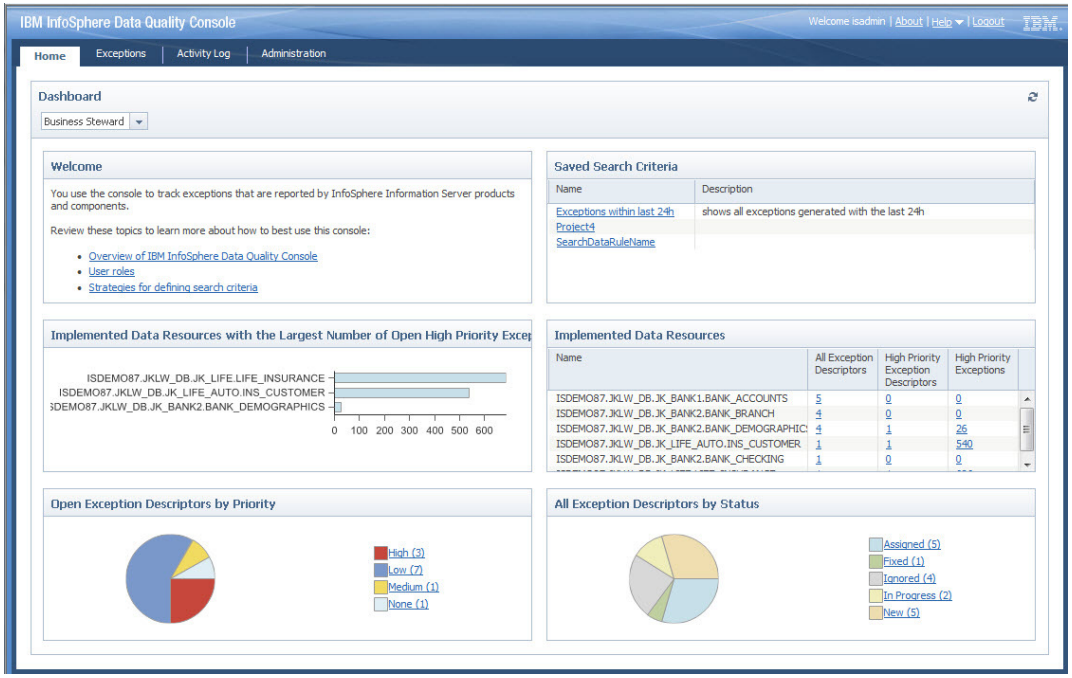


Reduced Risk

Sustainable Quality Capabilities - InfoSphere Data Quality Console Provides At-a-Glance Views

Business Value: Increases data quality awareness & provides greater flexibility and insight for managing and maintaining data quality

- **Provides unified environment to assess and monitor data for data stewards, analysts, etc.**
- **Data quality summary charts direct users to most critical information**
- **Includes comprehensive search, filtering and drill down capabilities**



IBM is the vendor with the broadest capabilities to monitor data rule exceptions within and across sources.

Reduced Risk

Sustainable Quality Capabilities: Ensure Accuracy and Quickly Adapt to Changes in the Environment

New Information Governance Rules & Policies

Extended platform support

Data Validation Rule Impact Analysis

Data Validation Rule Sequencing

New Address Verification Module

New Standardization Rules Designer

New Data Quality Console

define objectives

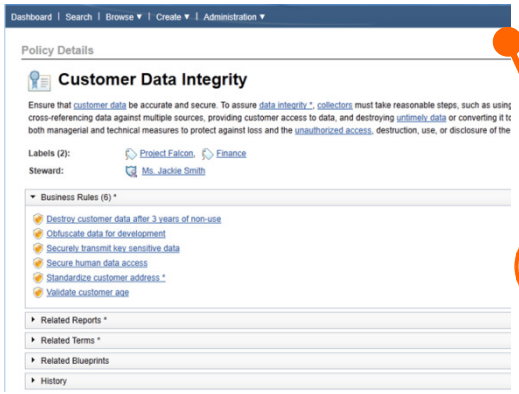
discover

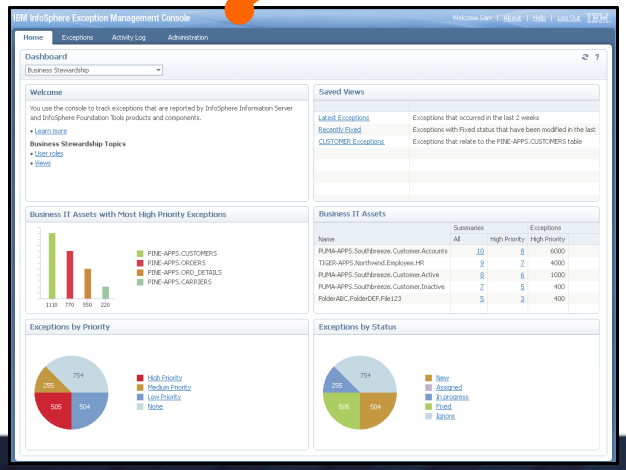
validate

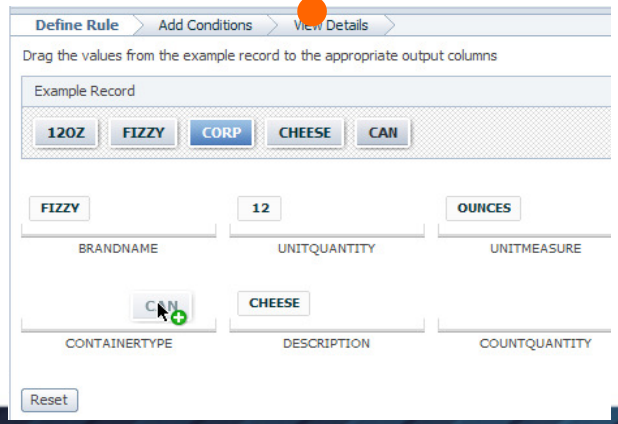
cleanse & enrich

monitor / track

"master"

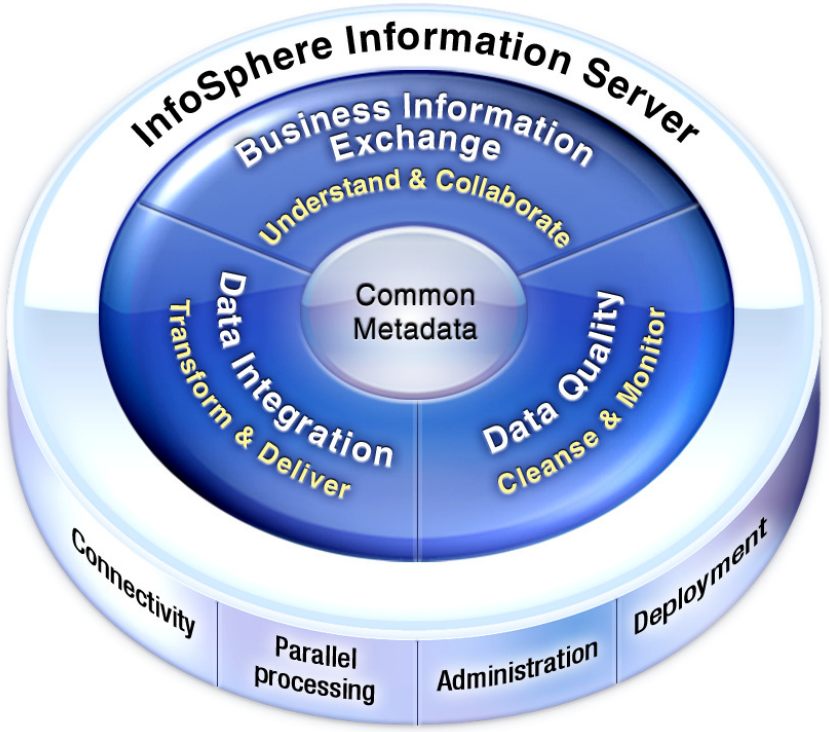






InfoSphere Information Server v9.1

- New capabilities focused on business-driven governance
- Unique features for agile integration
- Enhanced usability for more sustainable data quality
- Close collaboration with customers and partners via early release programs to accelerate success
- Simplified packaging to address key use cases



What's next ?

Market Leadership?

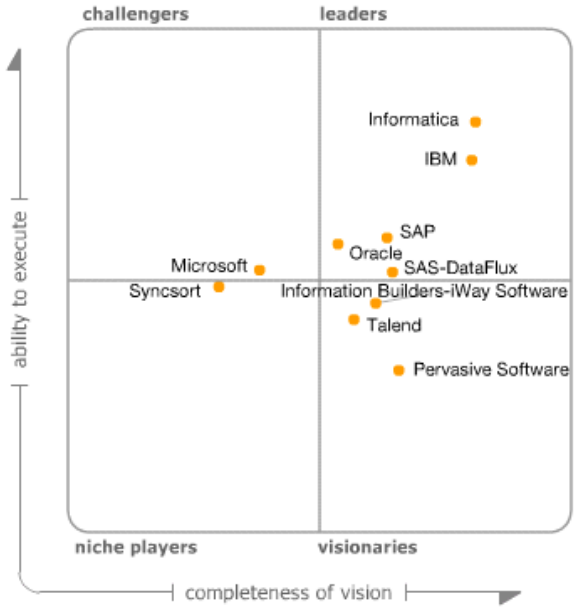
IDC Analyze the Future

TABLE 1

Worldwide Data Integration and Access Software Revenu
2009–2011 ($M)

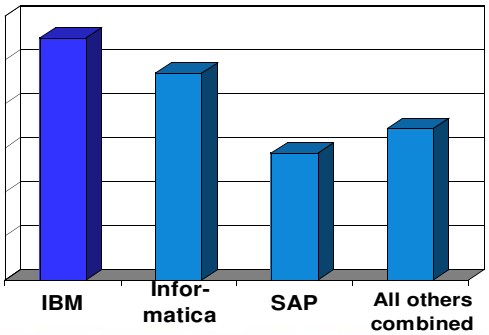
	2009	2010	2011
IBM	599.1	720.0	797.1
Informatica	400.8	455.0	549.6
SAS	433.6	455.7	510.8
SAP	338.9	407.2	493.8

Data Integration

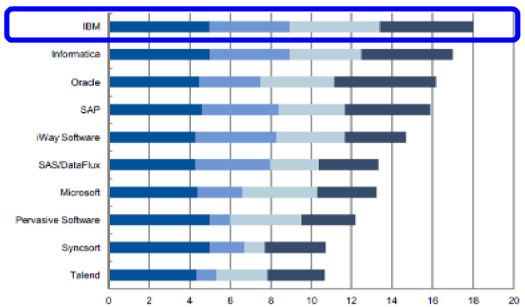


As of October 2012

Gartner: IBM Leads Data Integration Tools Market*



Gartner: Critical Capabilities for Data Integration, 2012



Source - Critical Capabilities for Data Integration Tools: Common Data Delivery Styles, Gartner, 2012

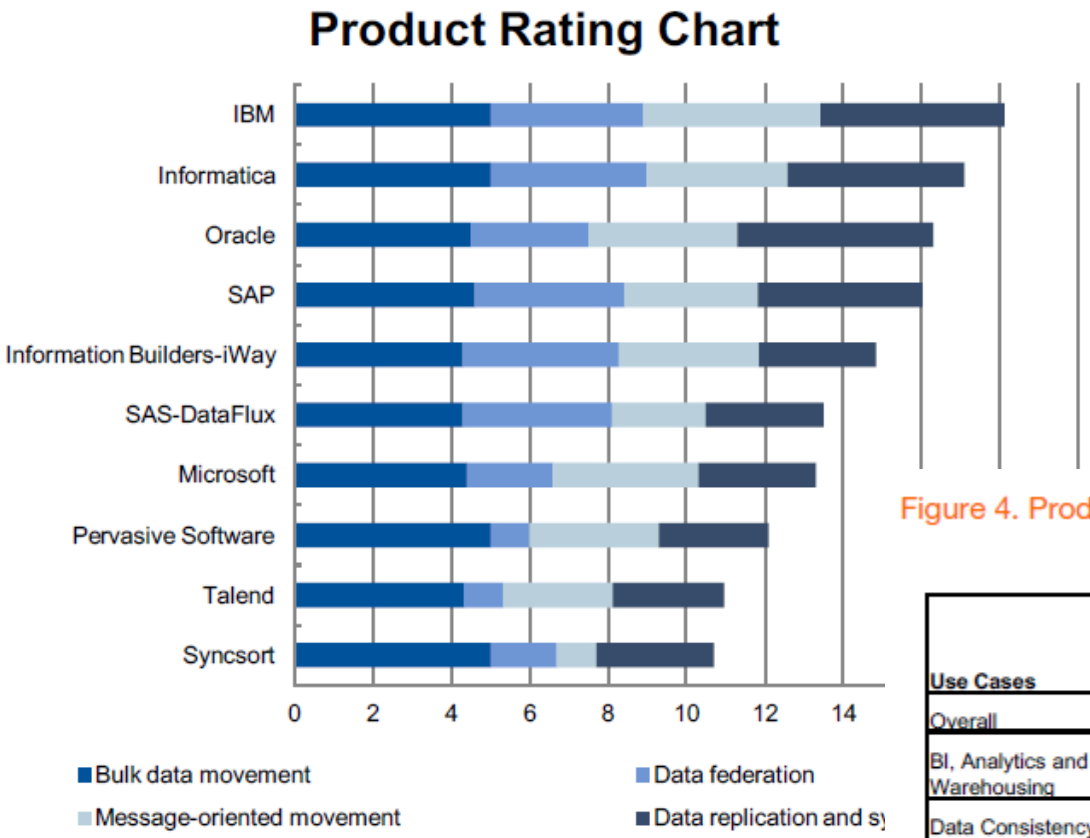
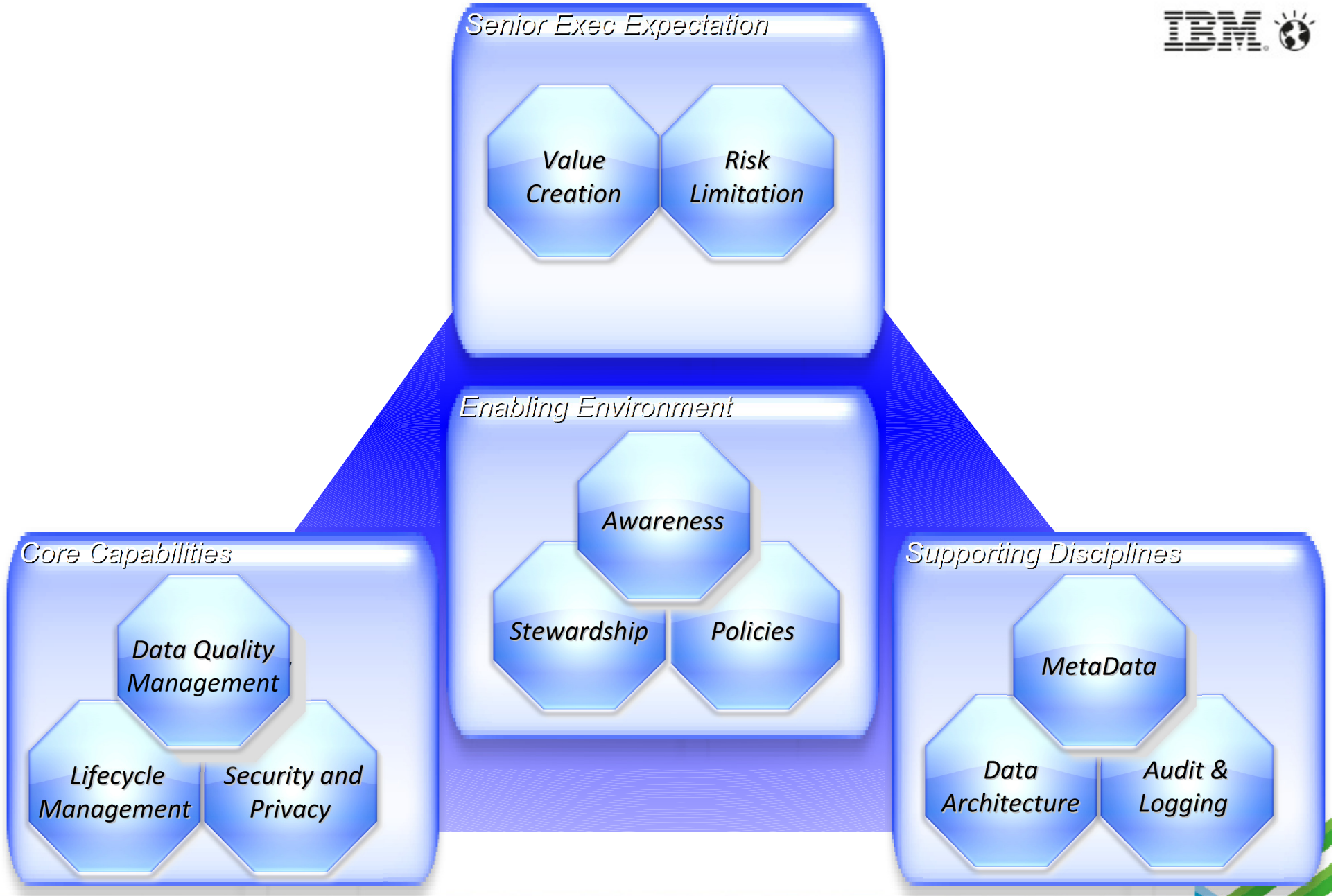


Figure 4. Product Score in Use Cases

Use Cases	IBM	Informatica	Oracle	SAP	SAS-DataFlux
Overall	4.7	4.6	4.3	4.3	3.8
BI, Analytics and (Logical) Data Warehousing	4.7	4.7	4.2	4.3	3.9
Data Consistency Between Operational Applications	4.9	4.7	4.5	4.3	3.7
Data or System Migrations and Consolidations	4.8	4.7	4.6	4.3	3.7
Master Data Management	4.7	4.6	4.2	4.3	3.9
Interenterprise Data Acquisition or Sharing	4.7	4.5	4.2	4.2	3.6

Source: Gartner (December 2012)


Senior Exec Expectation
Value Creation
Risk Limitation
Enabling Environment
Awareness
Stewardship
Policies
Core Capabilities
Data Quality Management
Lifecycle Management
Security and Privacy
Supporting Disciplines
MetaData
Data Architecture
Audit & Logging

Flow and timing of typical Starter Kit engagement:

First Find the Pain, then Ask About Consequences

2 hrs

Briefing
- agreement on scope (information domain – process domain - ....)
- agreement on participants LoB+IT
- agreement on business sponsorship

½ day

Information Gathering/part 1
- plenary session (LoB+IT)
- data quality and availability on agreed scope
- application domain & data usage grid

Information Gathering/part 2
- IT participants only
- architect view
- CVE model (metrics gathering)

2 days

TEAM IBM

Prepare recommendations & Business Case
- Kickstart templating
- Benchmarking – Reference cases
- Architect high level integration view
- IBM Solution mapping
- Business Case – TCO/ROI details
- Conclusion & Next Steps

2 hrs

Delivery Meeting
- presentation of recommendations & Business Case
- presentation of IBM Solution components
- agreement on next steps

Data Profiling Lab

- Do you dare to profile YOUR data sample ?
- 10 participants from this audience
  - First come – First serve base
- Data Profiling Lab
  - Hands-on : bring your own data on the USB stick provided here
  - will take place 1H of September 2013 in IBM Client Center, Brussels