

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!*

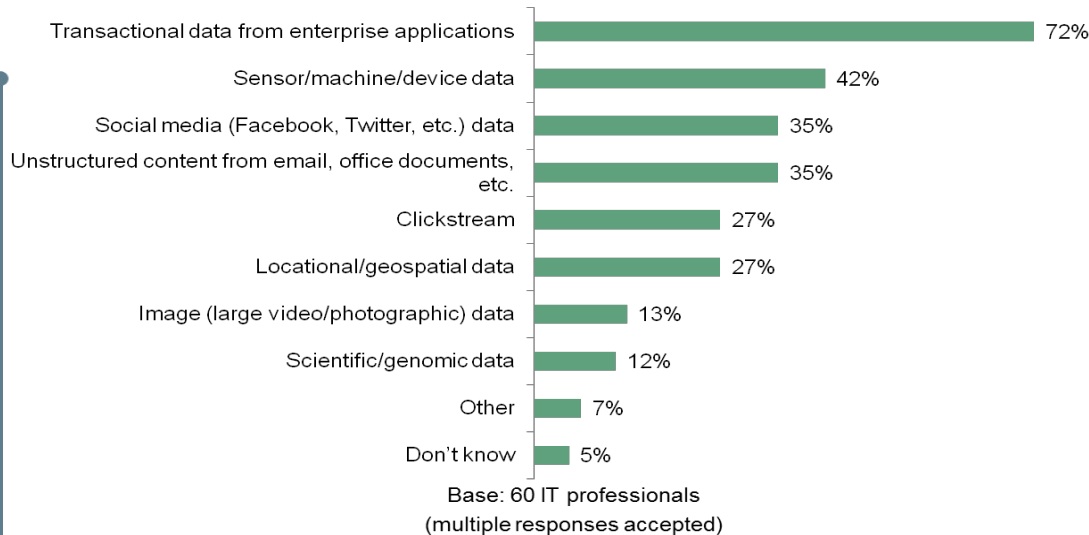
Why Are We Here?



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

Big data: across diverse subject domains

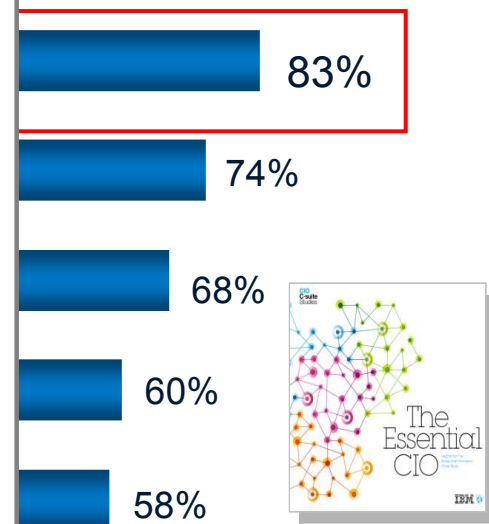
“What types of data/records are you planning to analyze using big data technologies?”



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

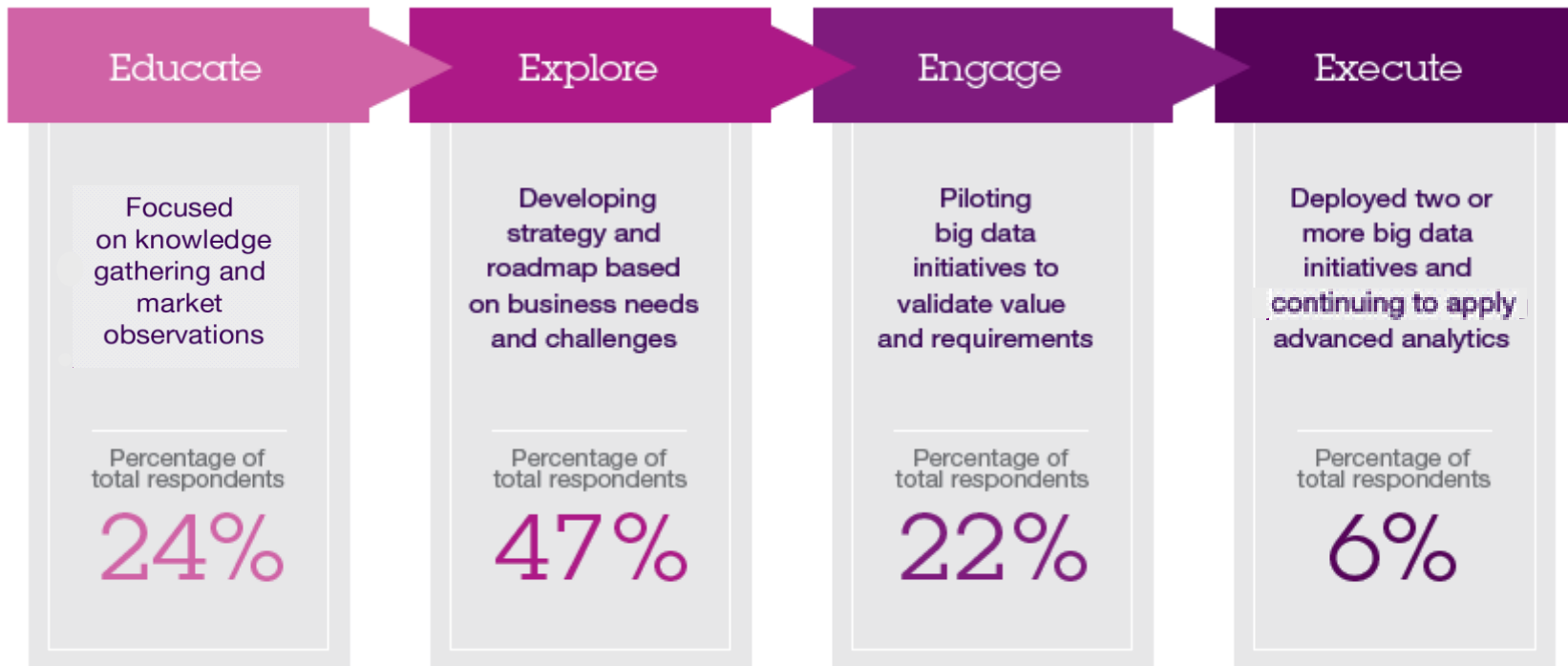
and Market Growth



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



Big data adoption



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 Agenda

Wednesday 4 January 2012, by Laurent Nicolas

“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.”



“Data is the new oil.”
Clive Humby

FINANCIAL TIMES
World business newspaper

...easingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to understand where customers are

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



BI Reporting and
Ad Hoc Analysis

Predictive
Analytics

Optimization



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?

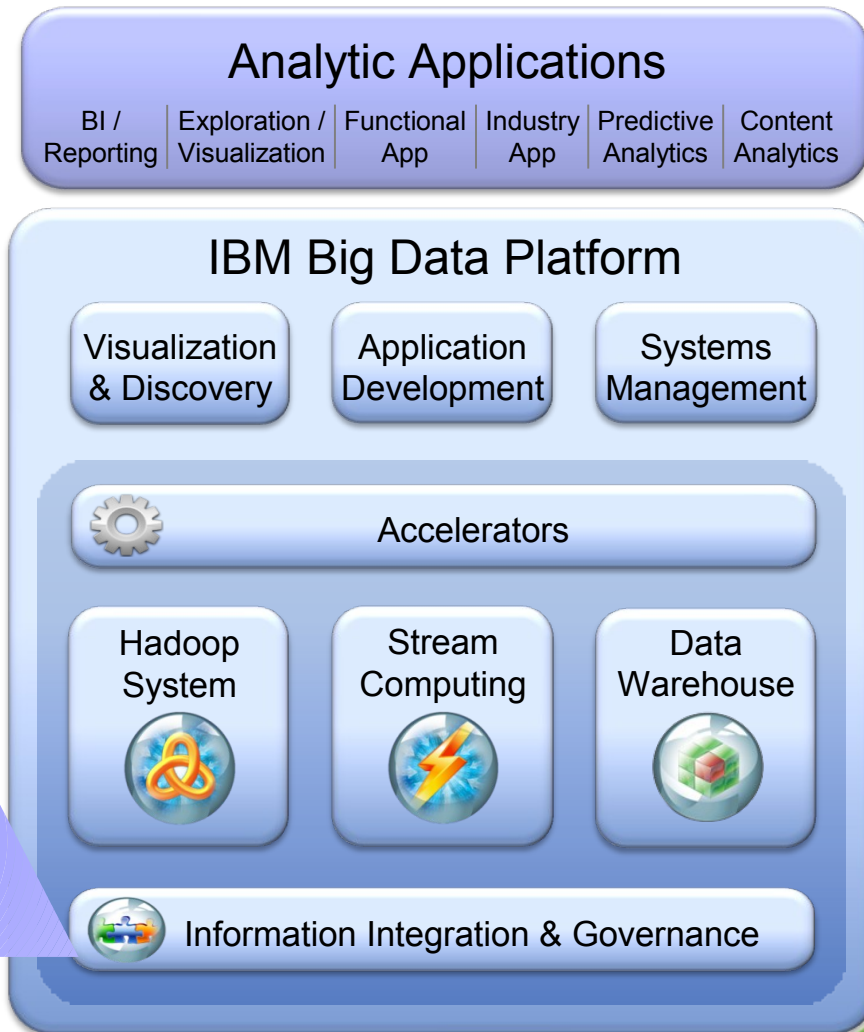
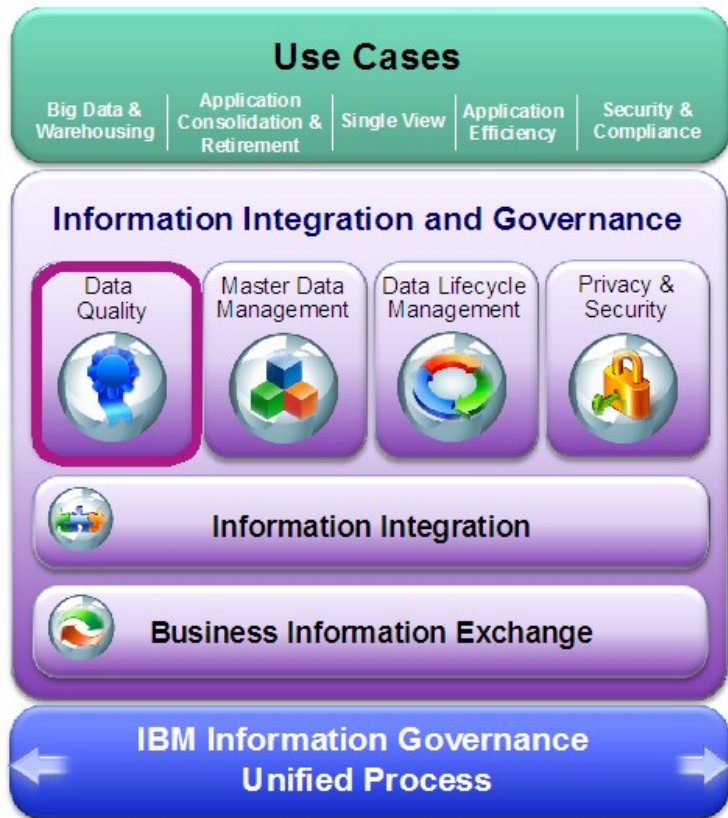
- What will happen?
- What will the impact be?

Cost efficiently processing the growing **Volume**

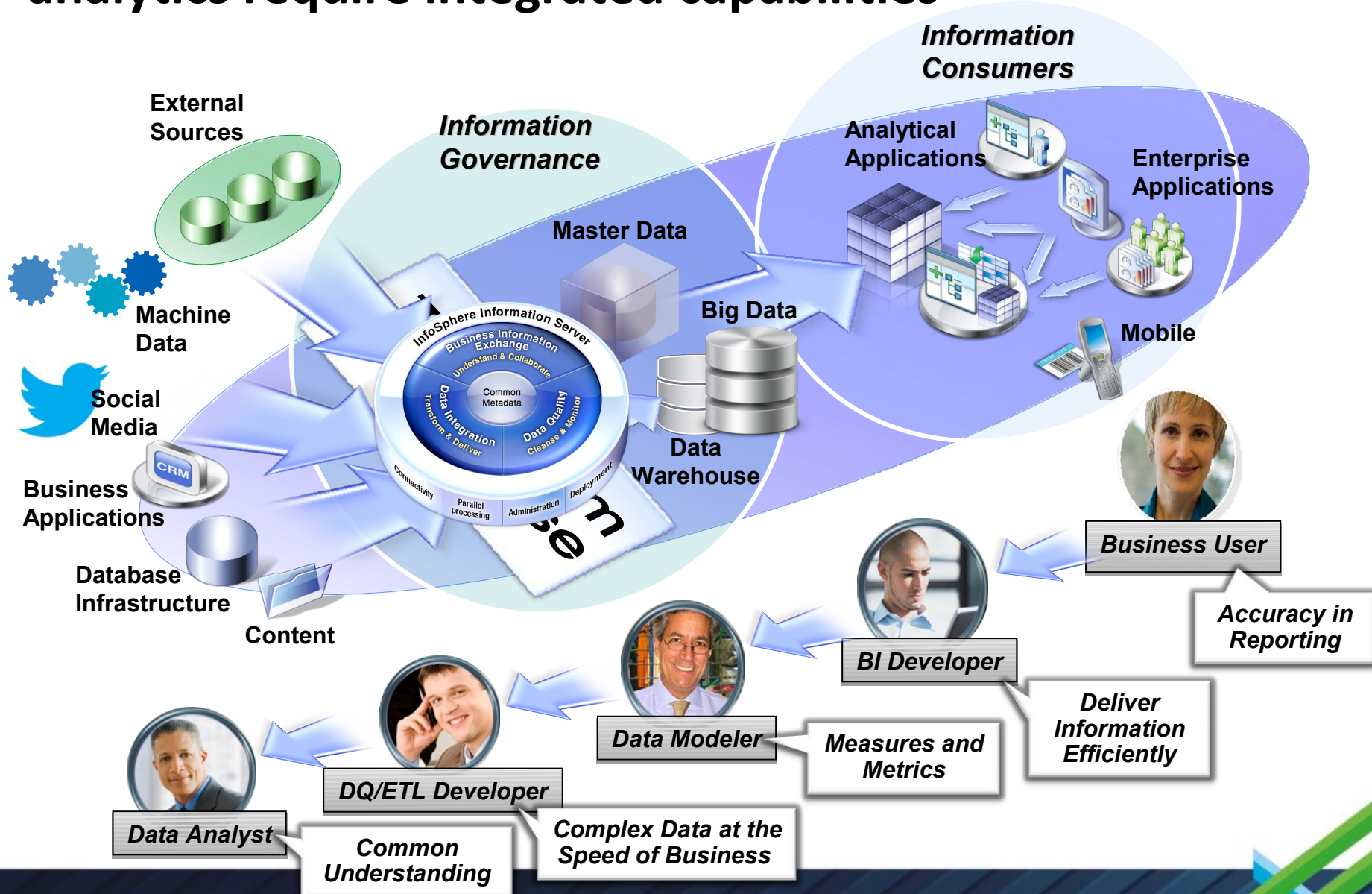
Responding to the increasing **Velocity**

Collectively analyzing the broadening **Variety**

InfoSphere creates trusted information for use in key initiatives

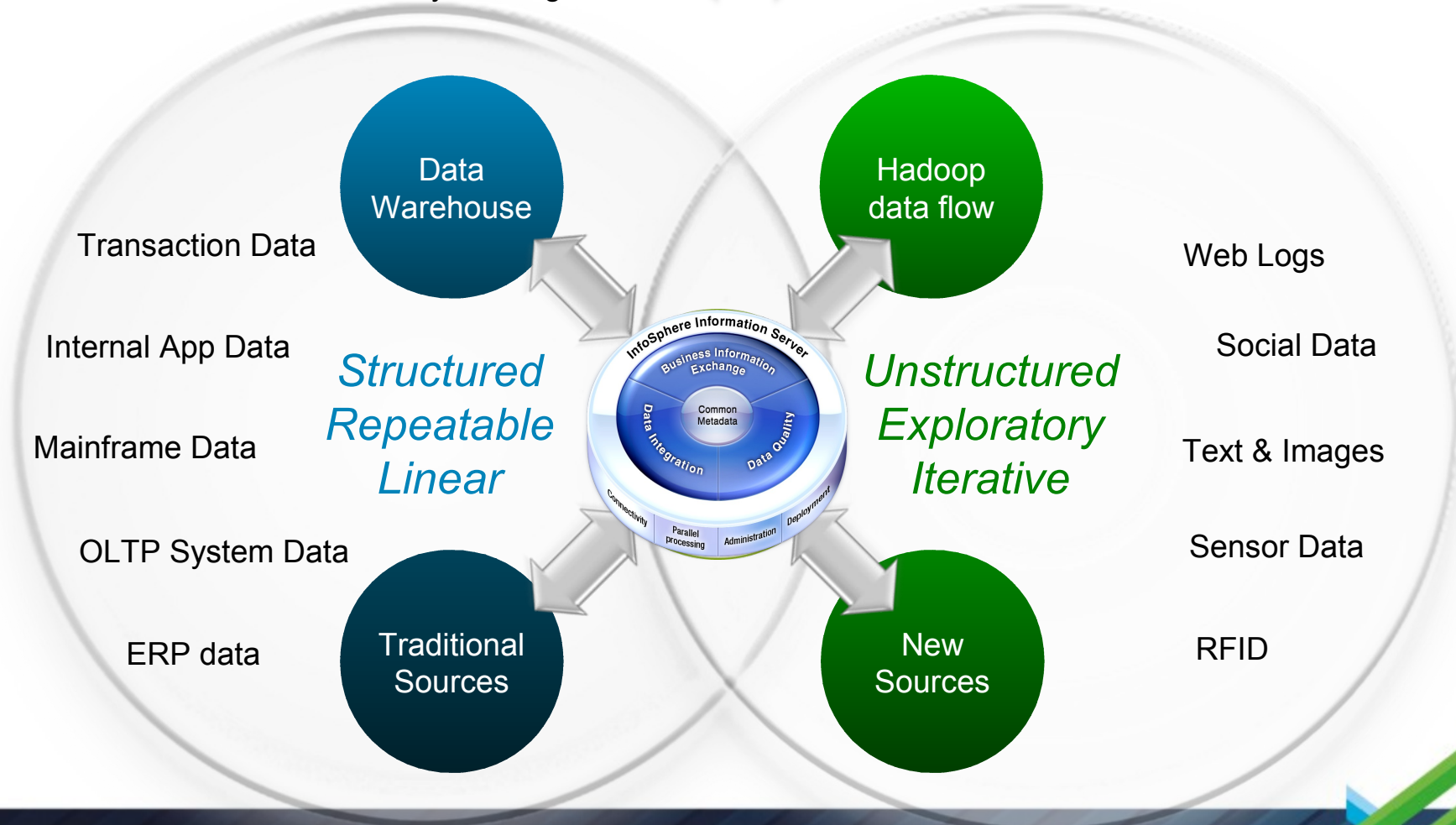


Information Supply Chain - Because Enterprise analytics require integrated capabilities



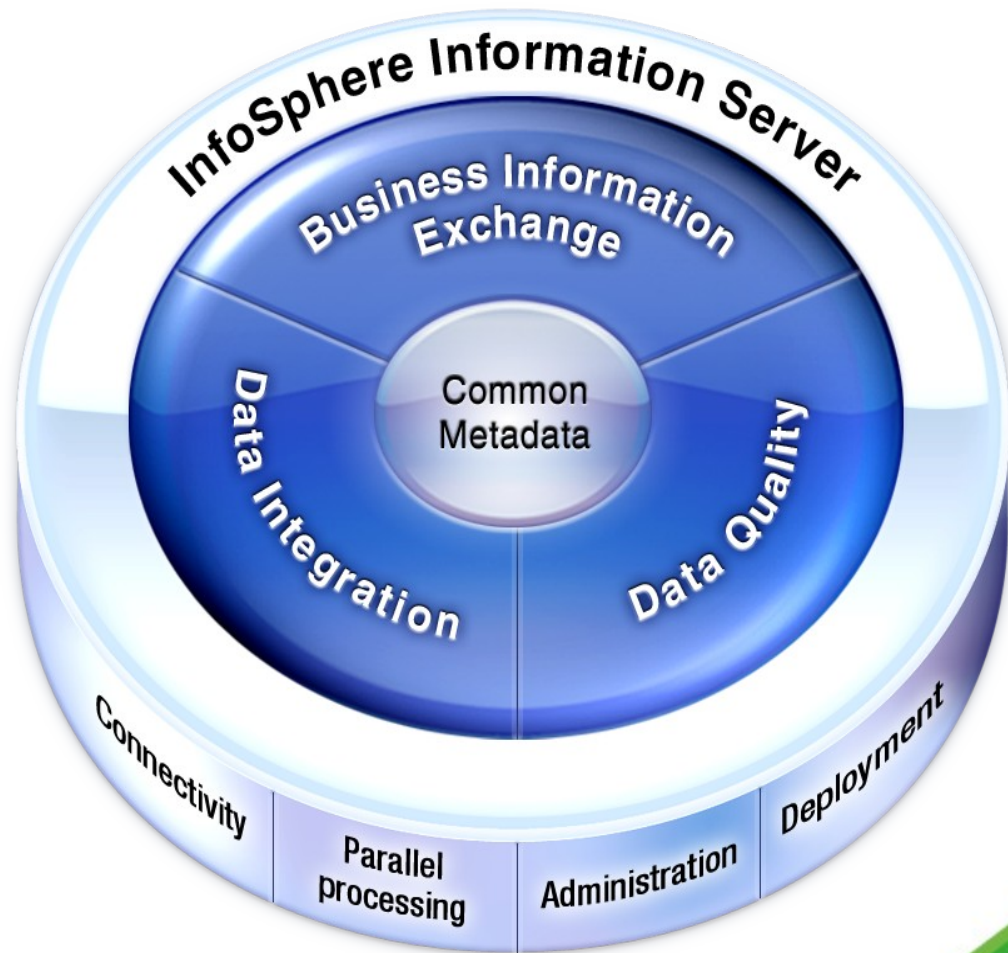
But Information Integration Requirements Are Becoming More Complex and Demanding on Infrastructure

Traditional Approach
 Structured, analytical, logical
 ↔
New Approach
 Creative, holistic, intuitive



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

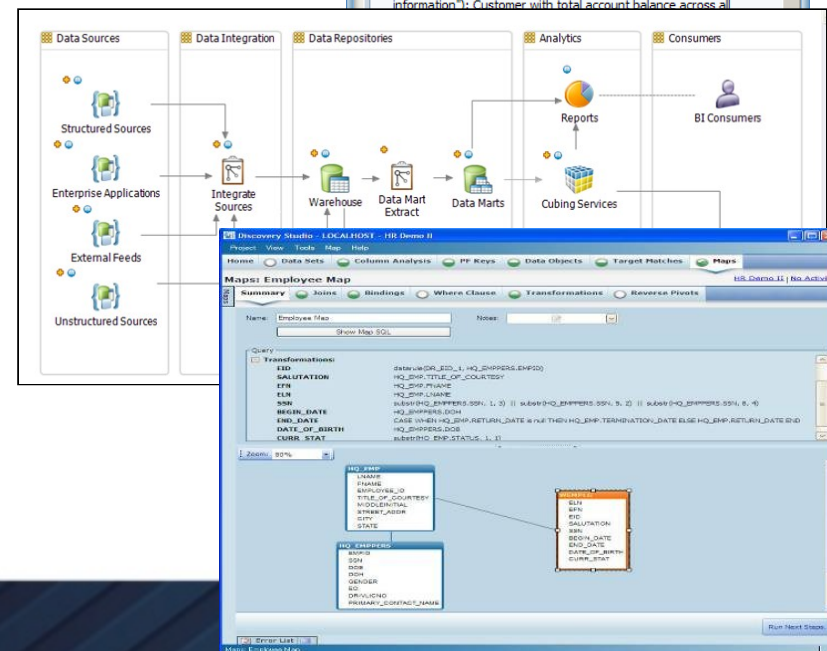
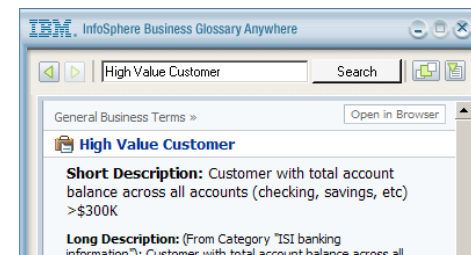
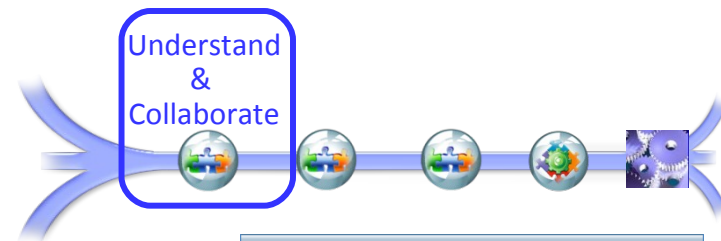
Guide projects with best practices to achieve goals with reduced risk

- 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

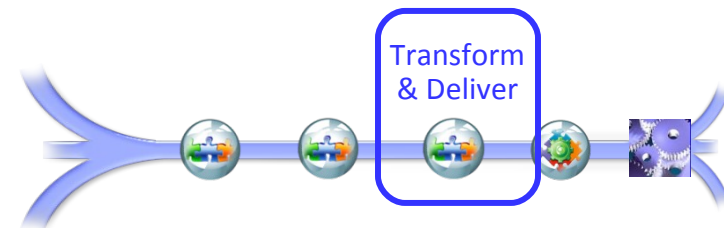
- unique capability to align DQ metrics with business & governance objectives



Name	Severity	Total Record	# Met	# Not Met	Trend
Company					
Unit Catalog Depart	5.3%	6867	6160	707	✗
Unit Catalog Division	2.25%	6867	6335	532	✗
Item					
Unit Catalog Make/B		6867	6860	7	✓
Unit Catalog No Ast	0.59%	6867	6483	384	✗
Unit					
Unit Check on UNITY	10.76%	6867	5785	1082	✗
UNITCTLG Average					
Unit Fix Cost					
WorldCo Consolidation					
Customer					
High-Value Customer					
Valid CUSTID					
Employees					
Valid Date of Hire					
Valid IDs					
Vendors					
Check Vendor					
Delay Expense					
Average Dollar Volu					

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

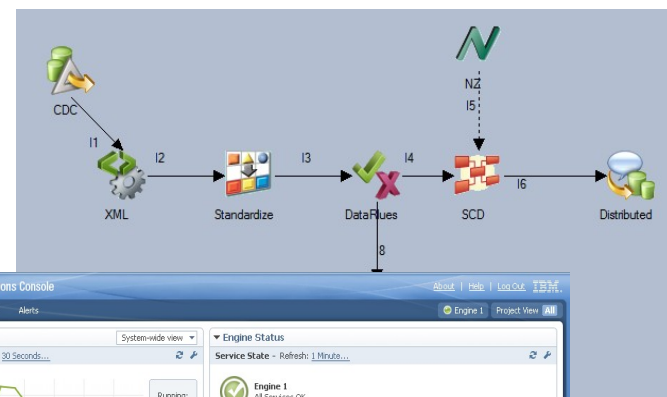
Integrate and transform data on demand across multiple sources and targets ...



Transform

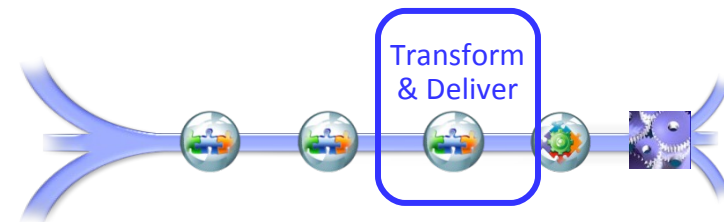
Satisfy the most complex transformation requirements with the most scalable runtime available

- 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

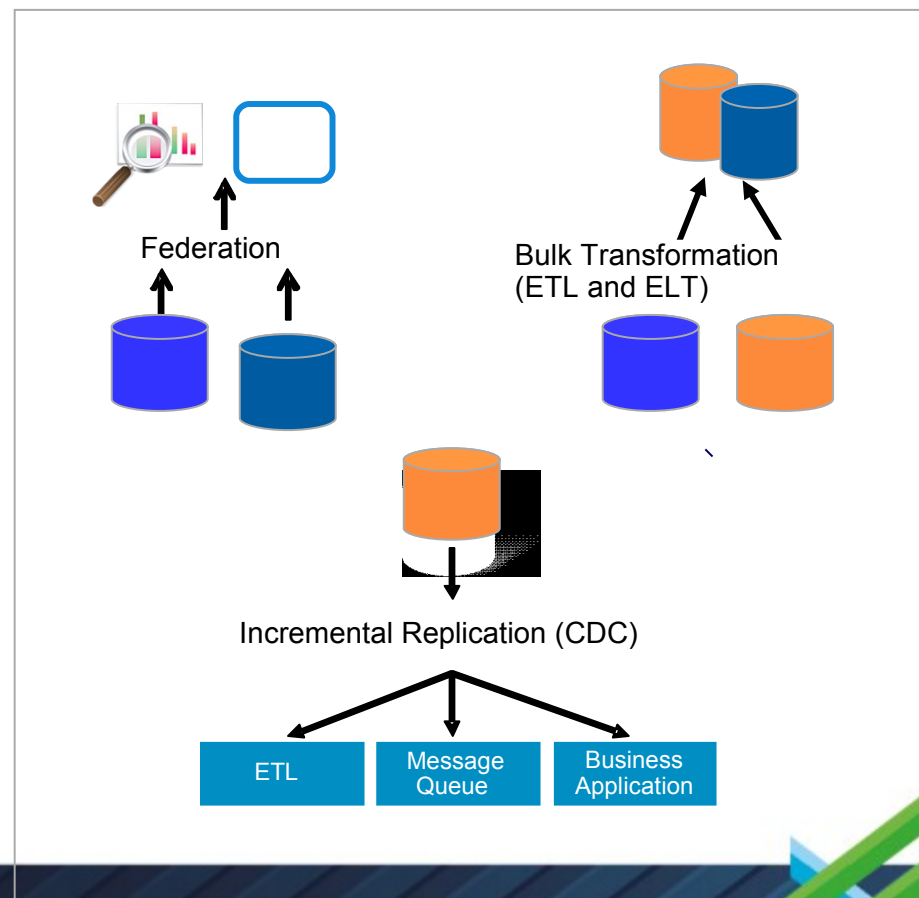
Deliver data efficiently according to your business requirements



Deliver

Leverage unique capabilities to:

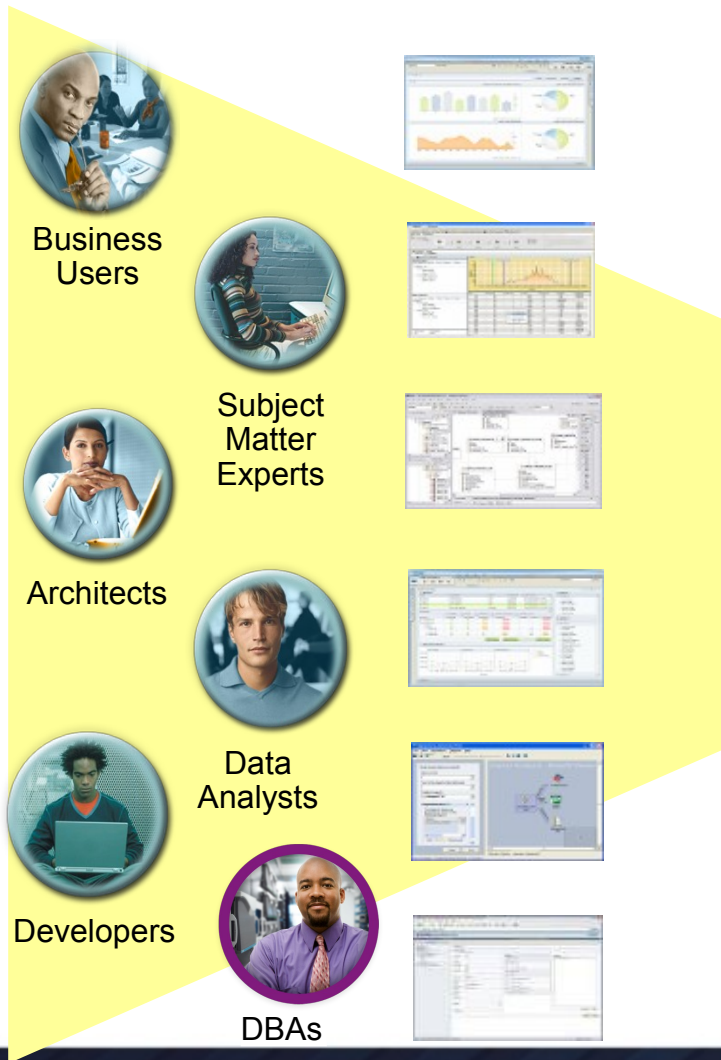
- Read from low-impact DBMS logs
- Apply any level of complex data integration/data quality logic
- Guarantee delivery through two-phased commit to one or more DBMS or MQ targets
- Support the most complex & challenging real-time integration requirements



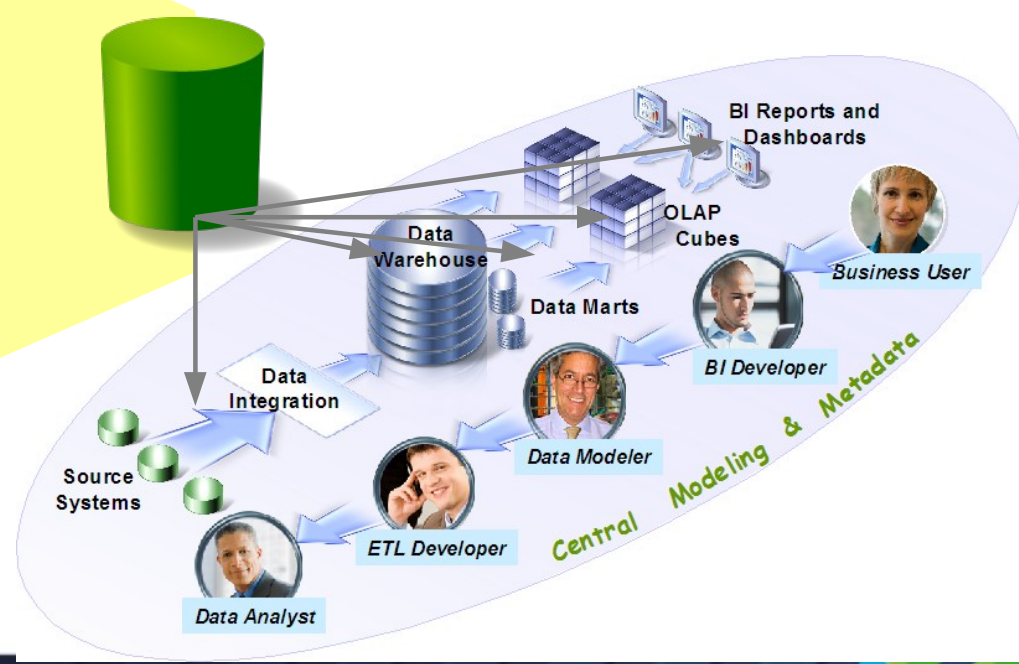
Common Metadata Layer - Why It's Important



Different roles create and consume different metadata



Common Metadata



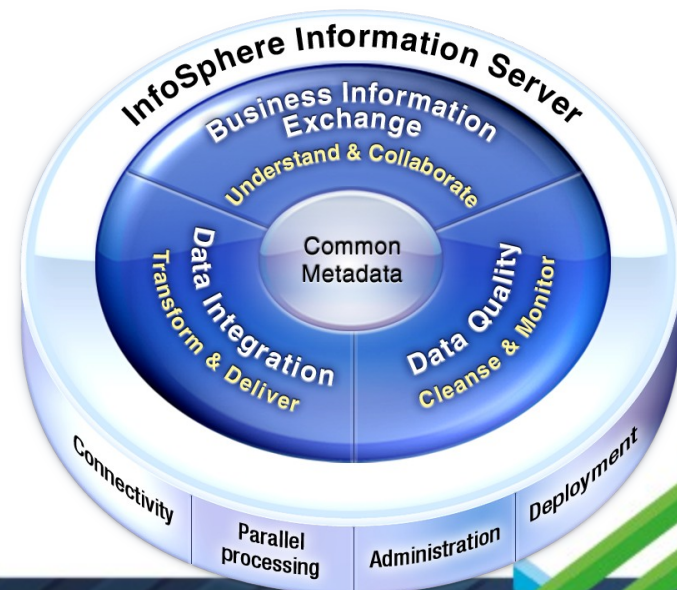
- Simplifies integration
- Facilitates change management and reuse
- Increases compliance with standards
- Increases trust and confidence in information

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



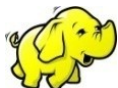
Unstructured Stage support

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?



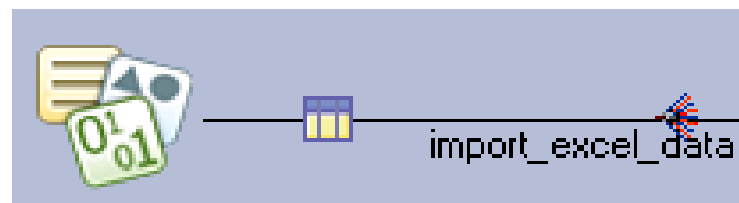
Assign terms to business processes

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

Native Excel Access

New Unstructured Stage provides....

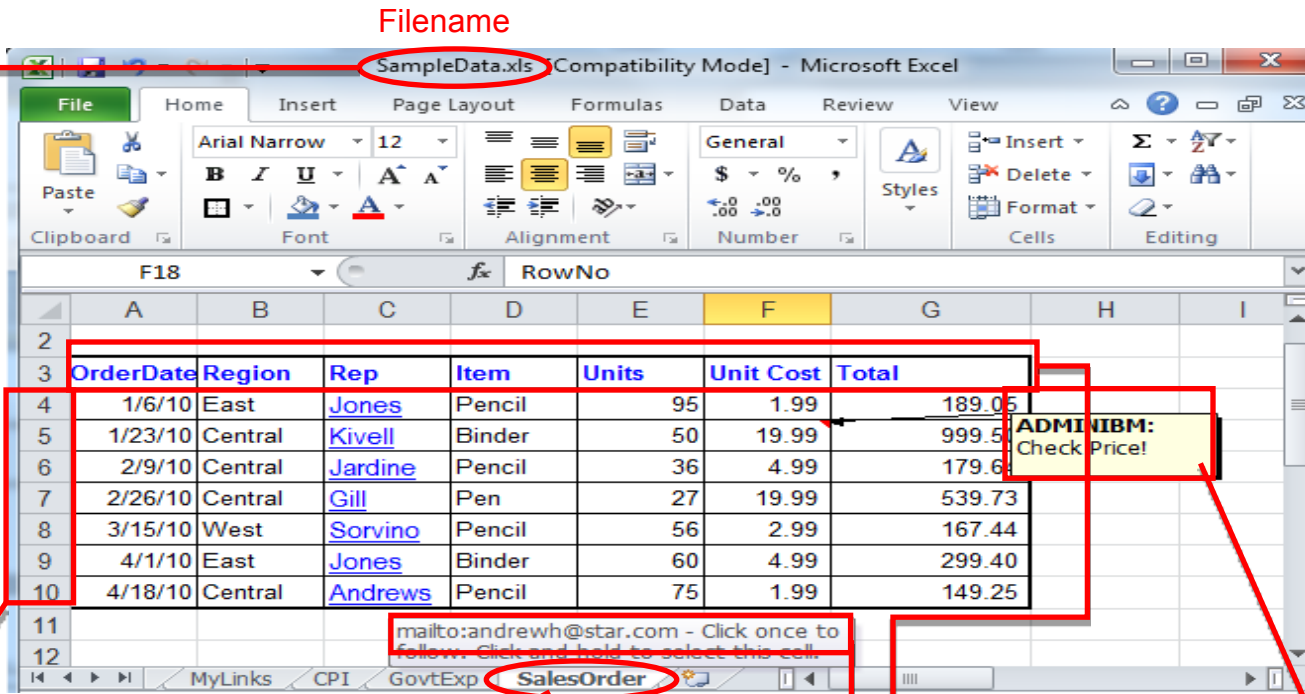
- **3-Step Configuration** – enter file name, select data range and setup column name mapping
- **Native Excel File Support** – reads data from any supported OS (Windows, Linux, Unix)
- **XLS Version Variants** - includes both Excel 97-2003 OLE2 (.xls) and 2007 OOXML (.xlsx) file format
- **Simplified 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



Row Number

Column Header

Sheet name

Hyperlink

Comment

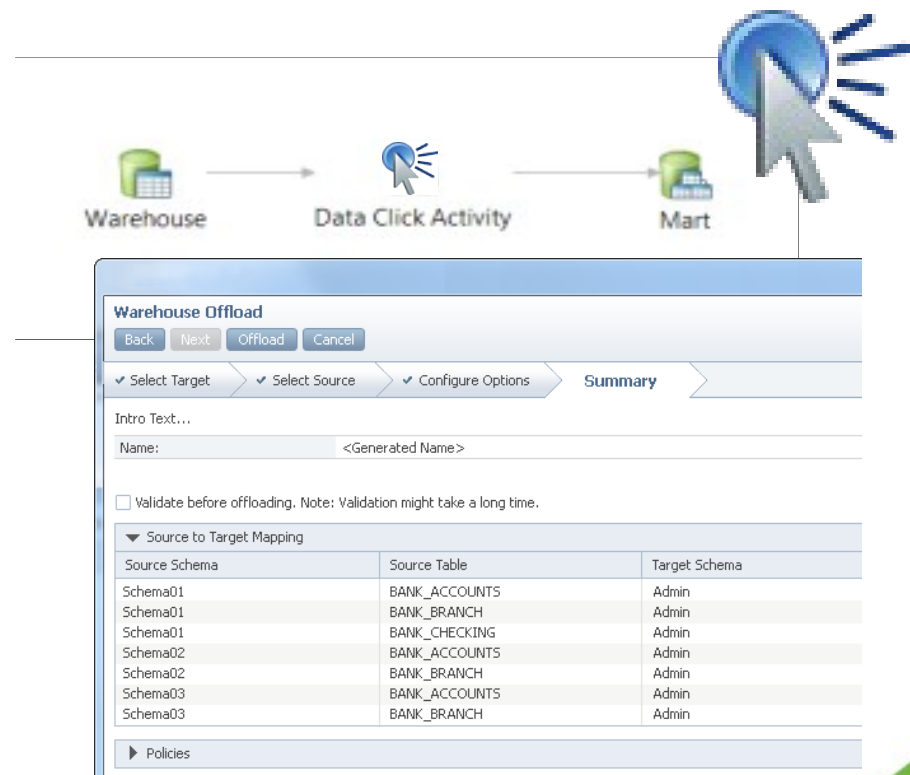
DataStage

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

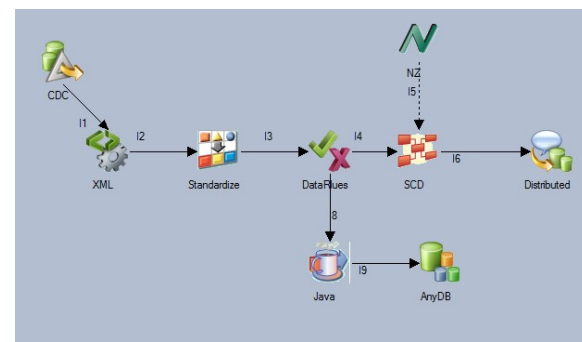


Faster Time to Value

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

The screenshot shows a web interface for 'Information Governance Policy Details'. The main heading is 'Customer Information'. Below the heading, there is a description: 'Ensure [customer](#) information complies with corporate [information governance standards](#) with special emphasis on data quality, data security, and data privacy. This policy applies to [North American](#) and [European](#) operations.' Below the description, there are sections for 'Labels (4)' with links to 'Data Privacy', 'Data Quality', 'Data Security', and 'PII', and 'Steward' with a profile icon for 'Marcellus Wallace'. A list of rules is shown under the 'Rules (6)' section, including: 'Customer Contact Address Formatting', 'Mask Customer PII in Dev and Test Environments', 'Required Fields For Customer Profile', 'Restrict Customer Financial Data to Accounts Under Level 7G', 'Store Customer Records For Ten (10) Years', and 'Validate Customer Identifier Uniqueness'. The interface also includes a search bar, navigation tabs (Search, Create, Terms, Categories, Rules, Policies, Labels, More), and a 'Quick Term Finder' search box.



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

The screenshot displays the 'InfoSphere Business Glossary' interface. At the top, there is a navigation bar with 'Search', 'Create', 'Terms', 'Categories', 'Rules', 'Policies', 'Labels', and 'More'. A 'Quick Term Finder' search box is also present. The main content area is titled 'Information Governance Rule Details' and features a shield icon next to the rule name 'Required Fields For Customer Profile'. Below the title, a description states: 'All customer profiles must contain populated values across key fields: [Name](#), [Age](#), [Gender](#), [E-Mail Address](#), [Purchase Frequency](#), [Last Purchase Date](#), [Home Address](#).' The interface lists 'Referencing Policies (1)' as 'Customer Information' and 'Labels (2)' as 'Data Quality, PII'. It includes expandable sections for 'General Information', 'Related Rules (2)', 'Implemented By (1)', 'Assigned Assets (1)' (showing 'Required Fields For Customer Profile'), 'Governs (2)', and 'Assigned Assets (2)' (showing 'CUST_PROFILE / DB2serv-456-34b > CUST_WH > SCHEMA-6' and 'CUST_HIST / DB2serv-456-34b > CUST_WH > SCHEMA-6'). A 'Notes' section is also visible at the bottom.



Reduced Risk

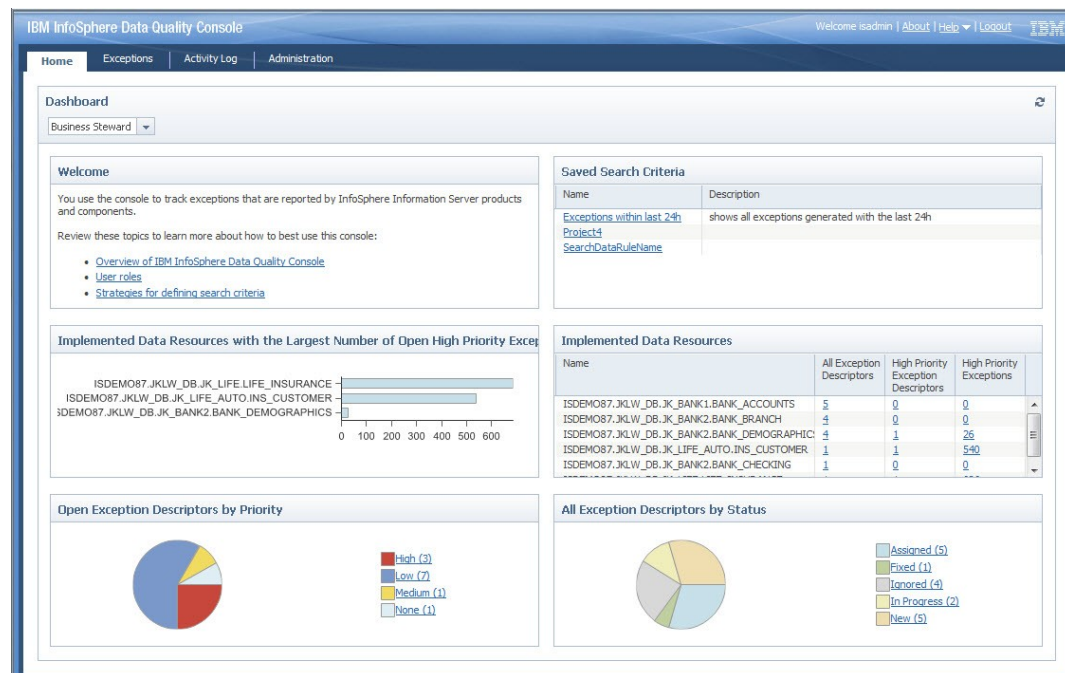
Sustainable Quality Capabilities - InfoSphere Data Quality Console Provides At-a-Glance Views



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

Policy Details
Customer Data Integrity
 Ensure that customer data be accurate and secure. To assure data integrity, collectors must take reasonable steps, such as using cross-referencing data against multiple sources, providing customer access to data, and destroying unneeded data or converting it to both managerial and technical measures to protect against loss and the unauthorized access, destruction, use, or disclosure of the data.

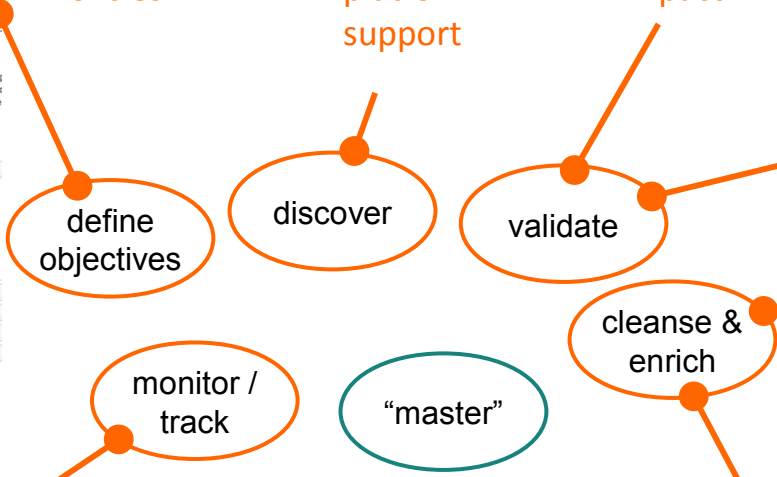
Labels (2): Project Falcon, Finance
 Steward: Ms. Jackie Smith

Business Rules (6)

- Destroy customer data after 3 years of non-use
- Obfuscate data for development
- Securely transmit key sensitive data
- Secure human data access
- Standardize customer address*
- Validate customer age

Related Reports

- Related Terms*
- Related Blueprints
- History



New Data Quality Console

IBM InfoSphere Exception Management Console

Dashboard: Business Stewardship

Welcome

You use the console to track exceptions that are reported by InfoSphere Information Server and InfoSphere Foundation tools products and components.

Business Stewardship Topics

- User roles
- Views

Business IT Assets with Most High Priority Exceptions

Asset Name	High Priority	Low Priority
PUMA-APPS-SouthBreach.CustomerAccounts	10	8
TIGER-APPS-Northwind.Employee.HR	2	2
PUMA-APPS-SouthBreach.Customer.Active	8	6
PUMA-APPS-SouthBreach.Customer.Inactive	2	5
FisherABC-FisherDEF-Fish123	5	3

Exceptions by Priority

Priority	Count
High Priority	754
Medium Priority	205
Low Priority	504
None	504

Exceptions by Status

Status	Count
New	754
Assigned	205
In Progress	504
Resolved	504
Known	504

Define Rule > Add Conditions > View Details

Drag the values from the example record to the appropriate output columns

Example Record

120Z FIZZY CORP CHEESE CAN

FIZZY 12 OUNCES

BRANDNAME UNITQUANTITY UNITMEASURE

CAN CHEESE

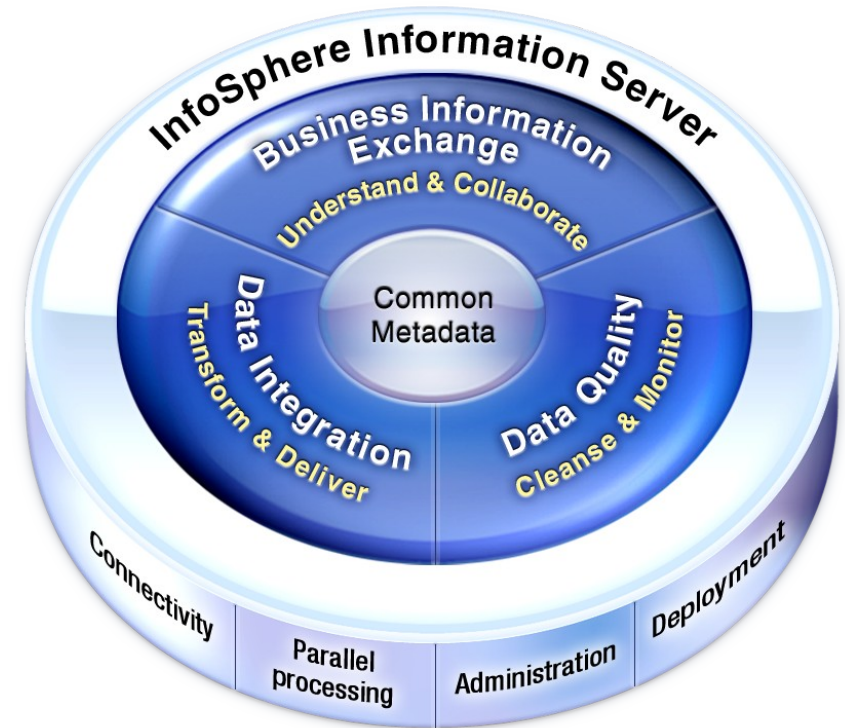
CONTAINERTYPE DESCRIPTION COUNTQUANTITY

Reset

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?

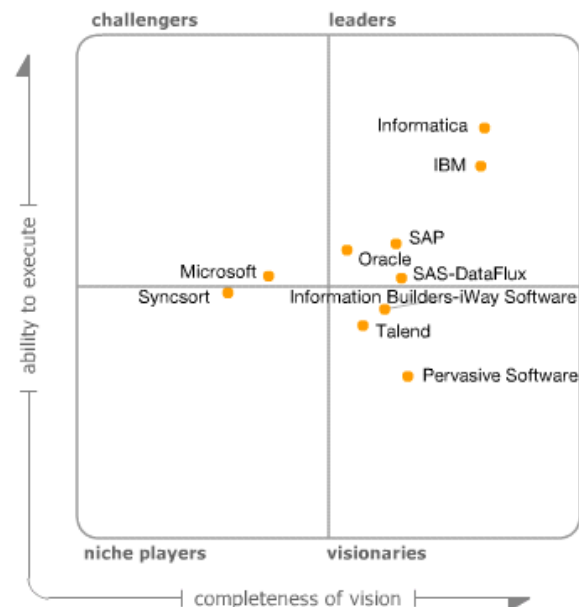


TABLE 1

Worldwide Data Integration and Access Software Revenue 2009-2011 (\$M)

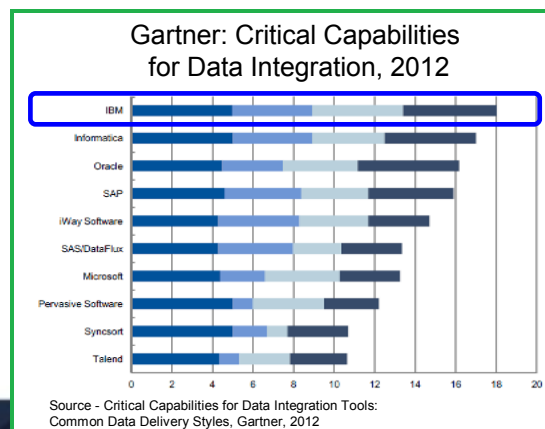
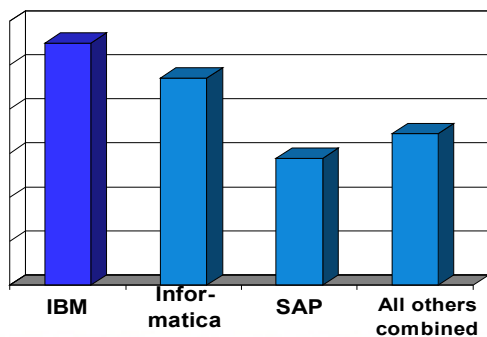
	2009	2010	2011
IBM	599.1	720.0	797.1
Informatica	400.8	455.0	549.6
SAS	433.6	455.7	510.8
SAP	338.9	407.2	493.8

Data Integration



As of October 2012

Gartner: IBM Leads Data Integration Tools Market*



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

Product Rating Chart

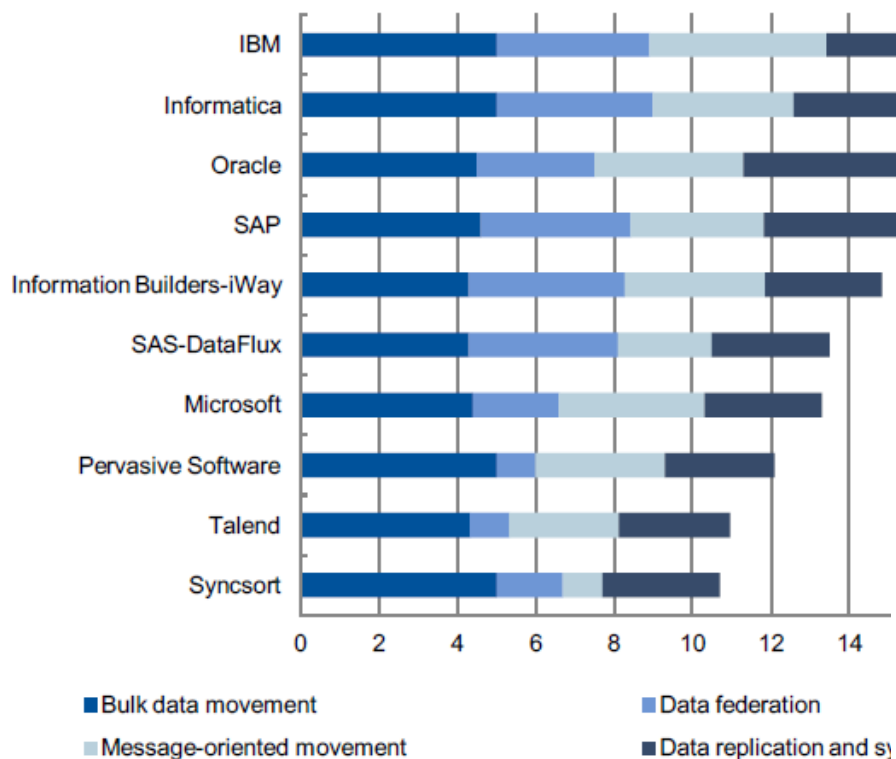


Figure 4. Product Score in Use Cases

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)

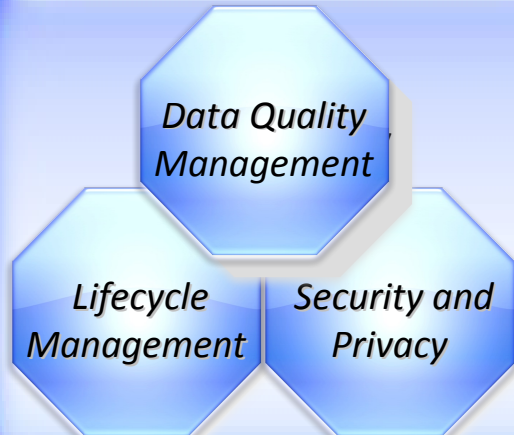
Senior Exec Expectation



Enabling Environment



Core Capabilities

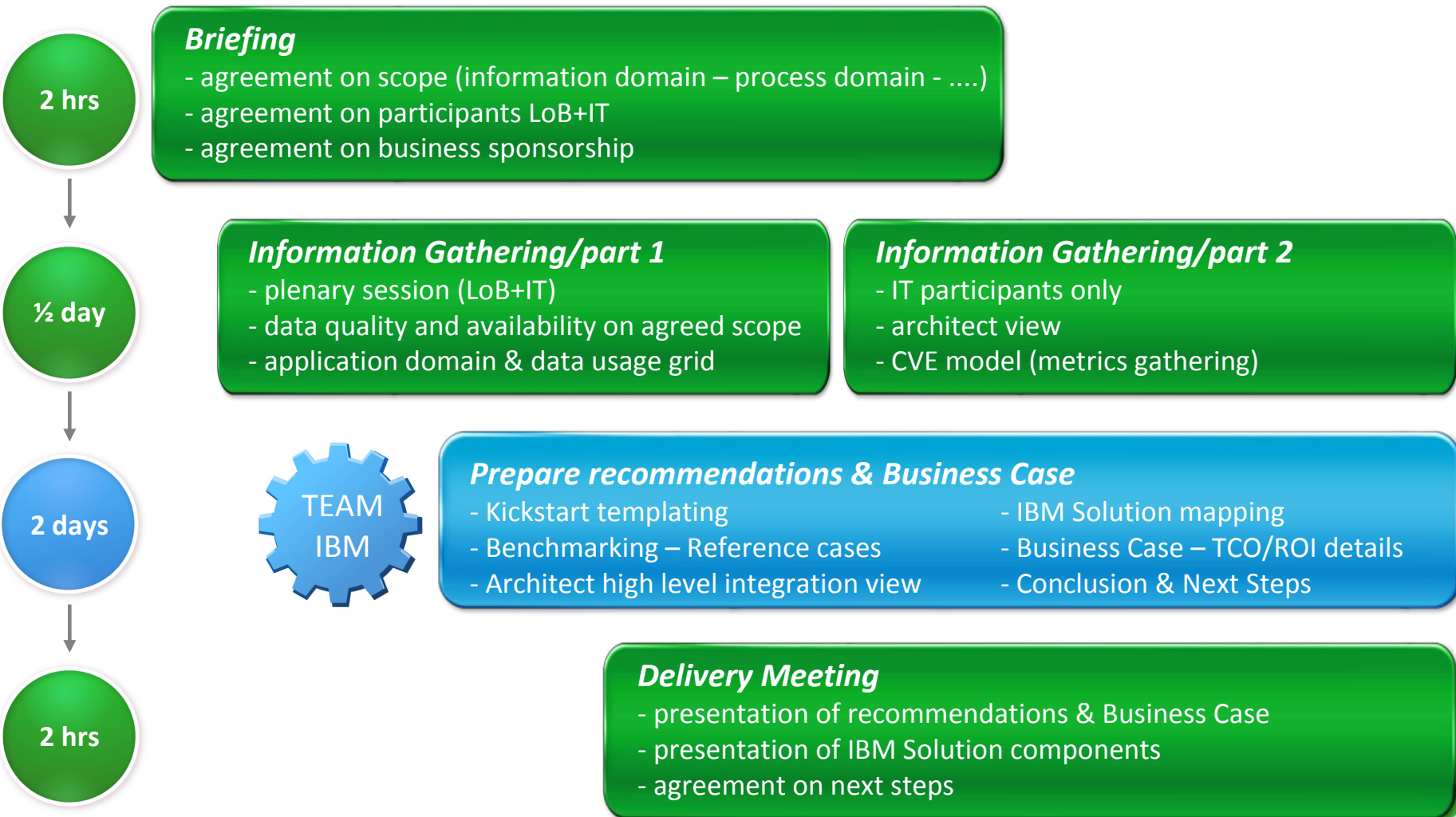


Supporting Disciplines



Flow and timing of typical Starter Kit engagement:

First Find the Pain, then Ask About Consequences



Data Profiling Lab



- Do you dare to profile YOUR data sample ?
- 10 participants from this audience
 - End users only !
 - First come – First serve base
 - Business card!
- Data Profiling Lab
 - Hands-on : bring your own data on the USB stick provided here
 - will take place in IBM Client Center
 - 25 September 2013 – Brussels
 - 02 October 2013 - Amsterdam



Your Information Governance Team in the Benelux



Andre De Loch

andre.de.locht@be.ibm.com

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

Brussels, BE IBM Sales & Distribution, Software Sales

Information Integration Sales Lead - Europe, report to Stefano Mino

Assistant: Christine Debecker



Toon Latinne

toon_latinne@be.ibm.com

32-2-339-3255 M: 32-477-84 10 54

Brussels, BE IBM Sales & Distribution, Software Sales

IT Specialist Information Integration



Frans Nieuwerth

frans_nieuwerth@nl.ibm.com

31-20-513-3653 M: 31-6 22925490

Amsterdam, NL IBM Sales & Distribution, Software Sales

Information Management Architect and Technical Specialist InfoSphere Master Data Managem



Robert Voermans [more info](#)

robert.voermans@nl.ibm.com

31-6-46257277 M: 31-6-46257277

Amsterdam, NL IBM Sales & Distribution, Software Sales

Client Technical Professional for Information Management



Lacko Darazsdi [more info](#)

lacko.darazsdi@nl.ibm.com

31-20-513-4704 M: 31-6 2111 2136

Amsterdam, NL IBM Sales & Distribution, Software Sales

Technical Sales: Information Management - Information Server/Identity Insight/Glob