

Harnessing the hidden potential of complex data

Andy Bromley

**Complex Data & Data Exchange
Specialist**

Agenda

- 1. Introduction to Complex Data**
- 2. Why You Should Care About Complex Data**
- 3. Examples of Complex Data Applications**
- 4. Recommendations**
- 5. Informatica & Complex Data**
- 6. Questions**

1. Introduction to Complex Data

Structured Data is just the tip of the iceberg



Over 80% of
enterprise data is in
unstructured formats

So why are so many
DI projects missing
it?

Complex Data Definitions

- **Structured Data**
 - Simple/ Tabular: RDBMS; Flat file; Table Dump
- **Unstructured Data**
 - Word Processing; Spreadsheets; Emails; PDF's; RDBMS Text Fields; RSS
- **Semi-Structured Data**
 - Simple Schema: Most industry standards
- **Complex Structured Data**
 - Complex Schema; Large File Sizes

Analytic Data Integration

- **Structured Data dominates Data Warehouse**
- **This trend reflects way DW typically organised**
- **Little to no complex data**
- **But Complex data represents ~75% of data in organisation**
- **“Single Version of Truth” should surely be “Complete version of the truth”?**

Operational Data Integration

- **Multi Enterprise Data Integration/ Data Exchange has existed for 30 years**
- **More important than ever as companies seek to collaborate more closely with**
 - Customers
 - Partners
 - Suppliers

2. Why You Should Care

Corporate Performance

- **Analysis of Complex Data can provide very different insight into corporate performance**
 - Eg Metrics & other aggregated data related to supply chain
 - Products; Suppliers; Distributors
- **Achieving a complete & accurate view of a business necessitates including Complex data sources as Data Warehouse Sources**

Industry Standards

Financial	Insurance	B2B Standards
SWIFT MT SWIFT MX NACHA FIX Telekurs FpML BAI – V2.0\Lockbox CREST DEX IFX TWIST UNIFI (ISO 20022) SEPA FIXML MISMO	DTCC-NSCC ACORD-AL3 ACORD XML	UN\EDIFACT EDI-X12 EDI ARR EDI UCS+WINS EDI VICS RosettaNet OAGI
	Healthcare	
	HL7 HL7 V3 HIPAA NCPDP CDISC	Other
		IATA-PADIS PLMXML NEIM

Cost Savings/ Efficiencies

- **Many companies have a legacy of siloed point to point solutions for B2B Data Exchange**
- **Our experience is that Hand coding is typical method employed for dealing with Complex Data**
- **Adopting industry standards promotes compatibility & standardisation**

Data Quality

- **DQ issues with Structured Data are gaining attention & visibility with many companies**
- **The Data Quality implications for Complex Data are identical**
 - Analytic Data Driven Decision Making
 - Operational processes using Complex Data

Stewardship/ Expertise

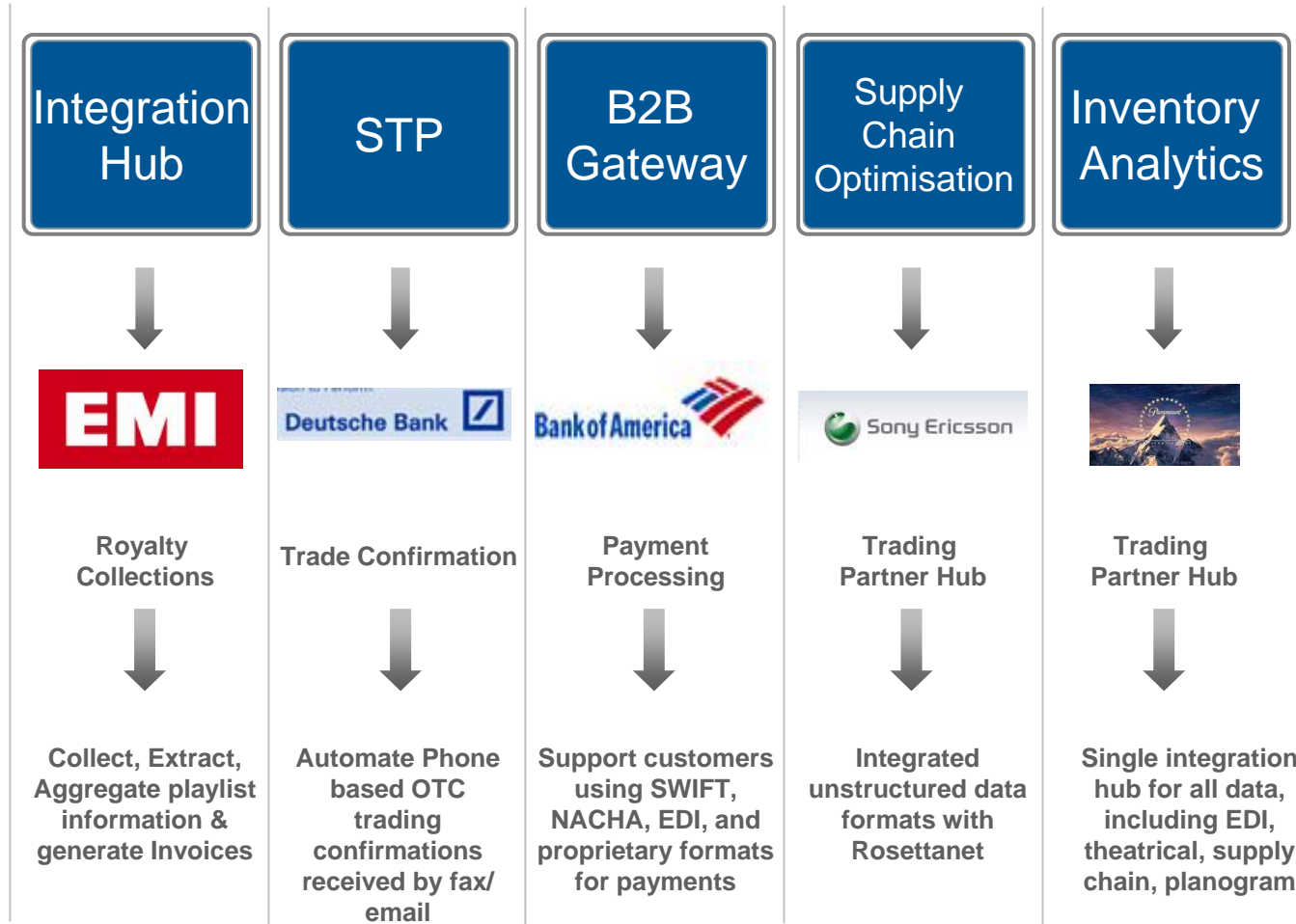
- **Complex Data challenges/ opportunities within each business require an ‘expert’**
- **Technically & Business savvy**
- **Can see ‘beyond’ the firewall**

New Business Opportunities

- **Multi Enterprise Data Integration offers many business opportunities**
 - Customer Centricity
 - Supply chain optimisation
 - White labelling
 - Infrastructure consolidation
 - BPO

3. Complex Data Case Studies

Complex Data Case Studies



4. Recommendations

Improve representation of Complex Data in DW

XML



Trade confirmation data in XML integrated into Data Warehouse to process new products quickly and increase volume of trades

Flat Files



Positional
Name = Value
^/>>Delimited<¥^



Royalty reports with songs usage data to data warehouse supporting usage analysis and payments

Industry Standards



Connect SEPA/SWIFT and other formats to payments systems and data warehouse

Documents



Extract information from companies financial reports in formats like PDF and Excel

Have 'Stewards' & provide tools

- **Appoint Complex Data Stewards for your organisation– either formally or informally**
- **Someone who will have visibility of your data inside & outside the firewall**
- **Provide Stewards with productivity tools for analysing Complex Data & dealing with exceptions & mappings**

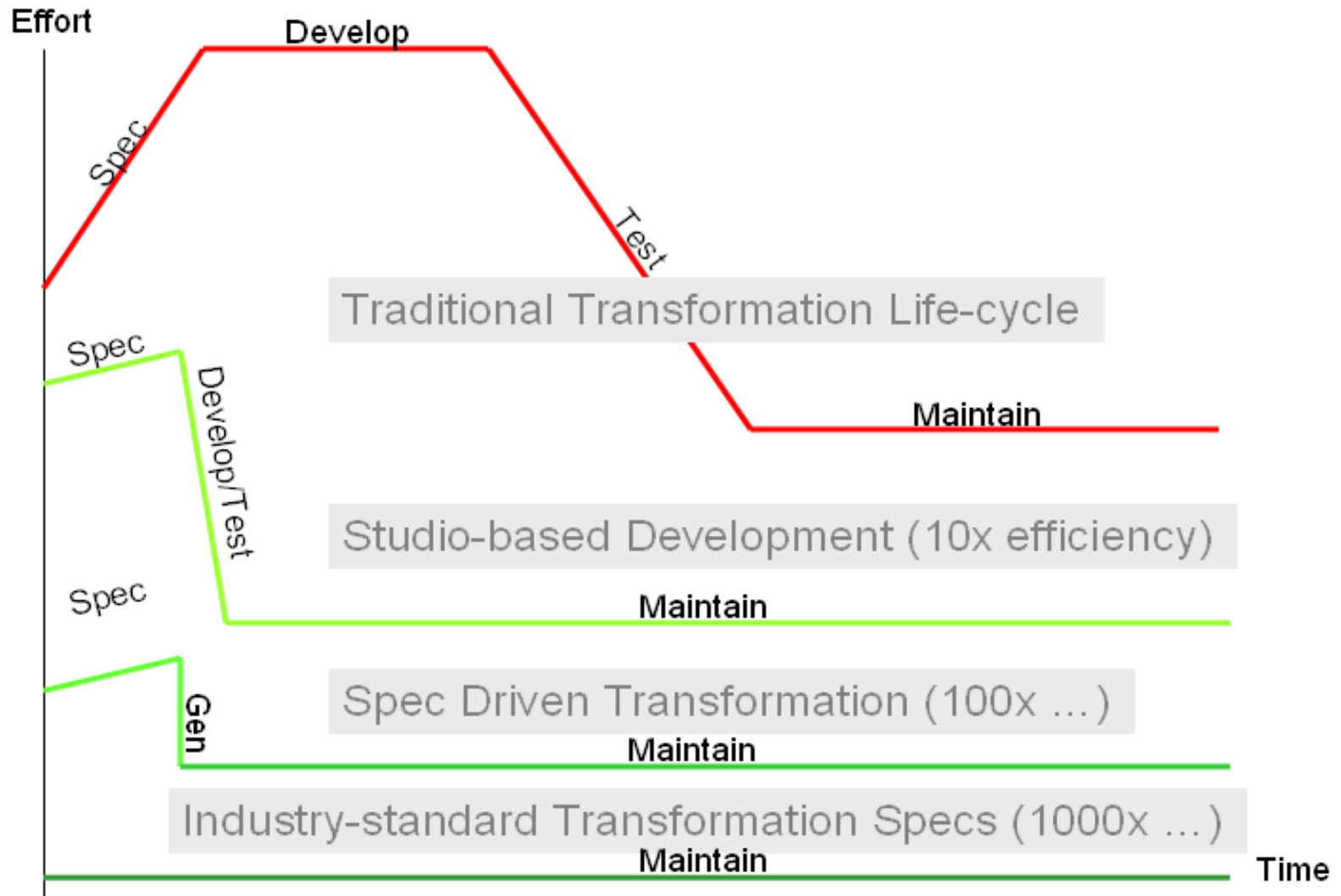
Extend DI infrastructure

- **Start looking at ways to extend your existing DI infrastructure to include complex data types**
- **What Industry Standards data types are relevant today & in the medium/ long term**
- **Operational applications of Complex data often require Real Time capabilities**
- **How does your existing DI infrastructure interoperate with your messaging infrastructure?**

Select tools that cover both

- **It is usually more efficient & cost effective to have Data management tools that combine functionality for both Structure & Complex data types**
- **Avoid separate siloes for ETL & Complex Data/ Batch & Real time**

Embrace Industry Standards



Look for Business opportunities

- **Complex Data applications can bring great business value**
 - Closer collaboration with customers & partners
 - Improved efficiencies & STP
 - Outsourcing opportunities
 - White Labelling
 - Legacy modernisation
 - Standards Adoption

5. The Informatica Platform

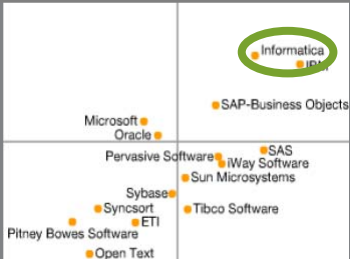
About Informatica

Cloud Data Integration

"Informatica's mission to integrate data from business to business... meets a need of many companies... Others in this market (including IBM, Microsoft and Oracle) have not yet pursued this progressive use of data integration"



2008 Data Integration Magic Quadrant



Gartner

Application ILM

Informatica was assessed as having the largest share, more than 33 percent, in packaged applications for database archiving based on number of customers for 2007 by Gartner, Inc. in a research report released October 13, 2008

Gartner

2009 Data Quality Magic Quadrant



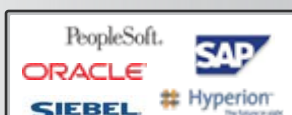
Gartner

B2B Data Exchange

"Informatica supports the requirements of cross-organizational data exchange, so users can apply familiar and trusted data integration tools and techniques to the relatively new and growing practice of B2B data integration."



Cloud Computing



Application



Database



Unstructured

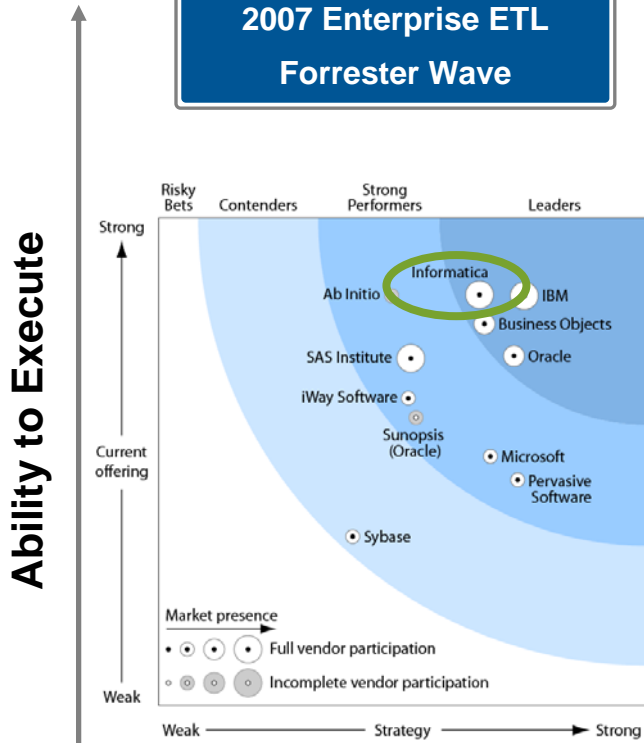


Partner Data

Why Informatica?

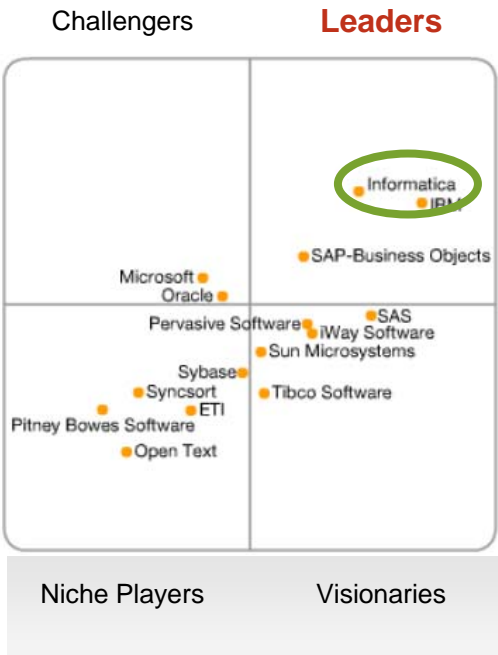
Proven Technology Leadership

2007 Enterprise ETL Forrester Wave



Source: Forrester Research 2007

2008 Data Integration Magic Quadrant



Source: Gartner Research (September 2008)

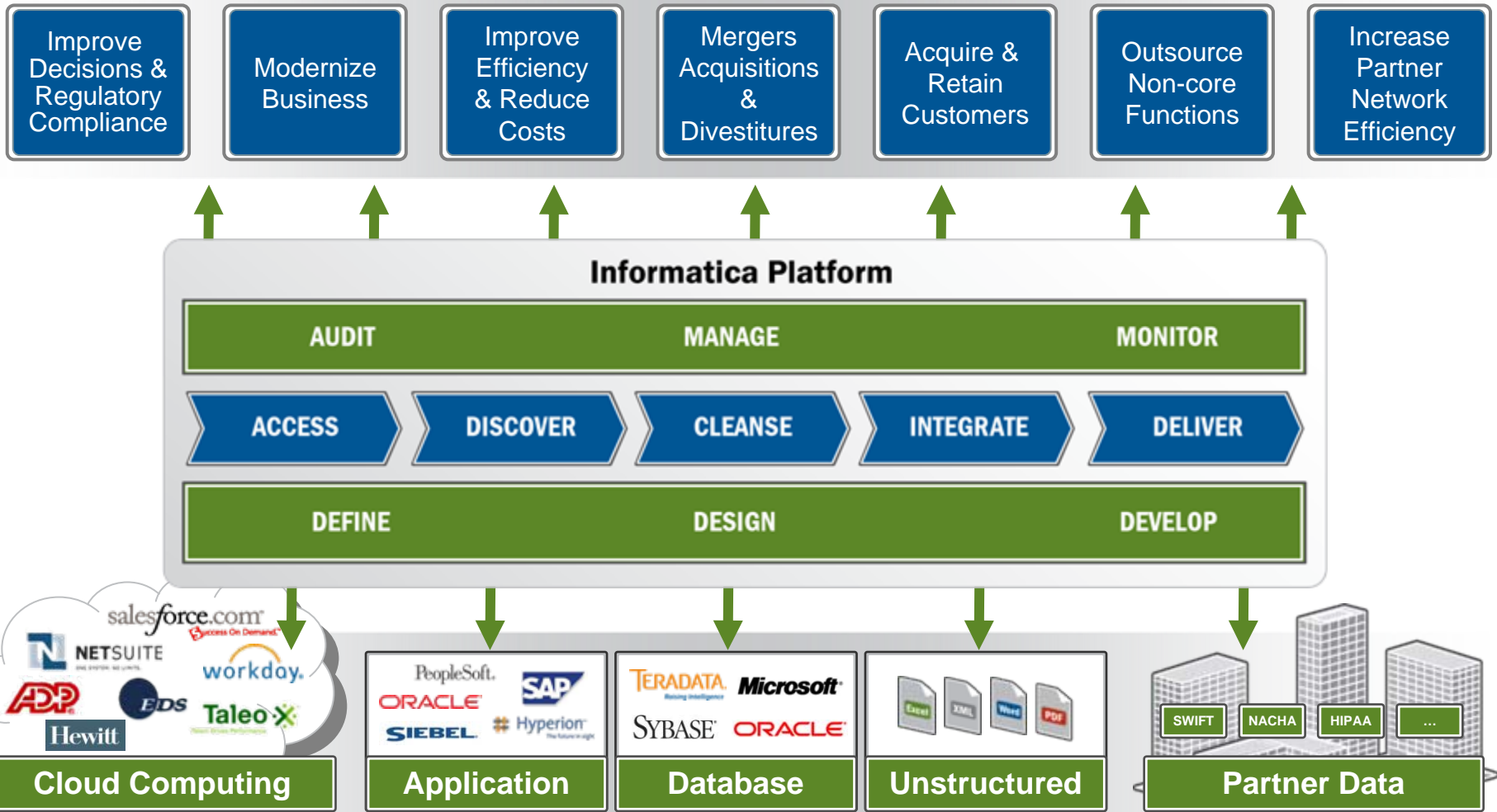
2008 Data Quality Magic Quadrant



Source: Gartner Research (June 2008)

Completeness of Vision

The Informatica Platform



Over 3,450 Companies Rely on Informatica

Financial Services and Insurance



Government and Public Sector



Healthcare and Life Sciences



Manufacturing



Retail and Services



Telecommunications



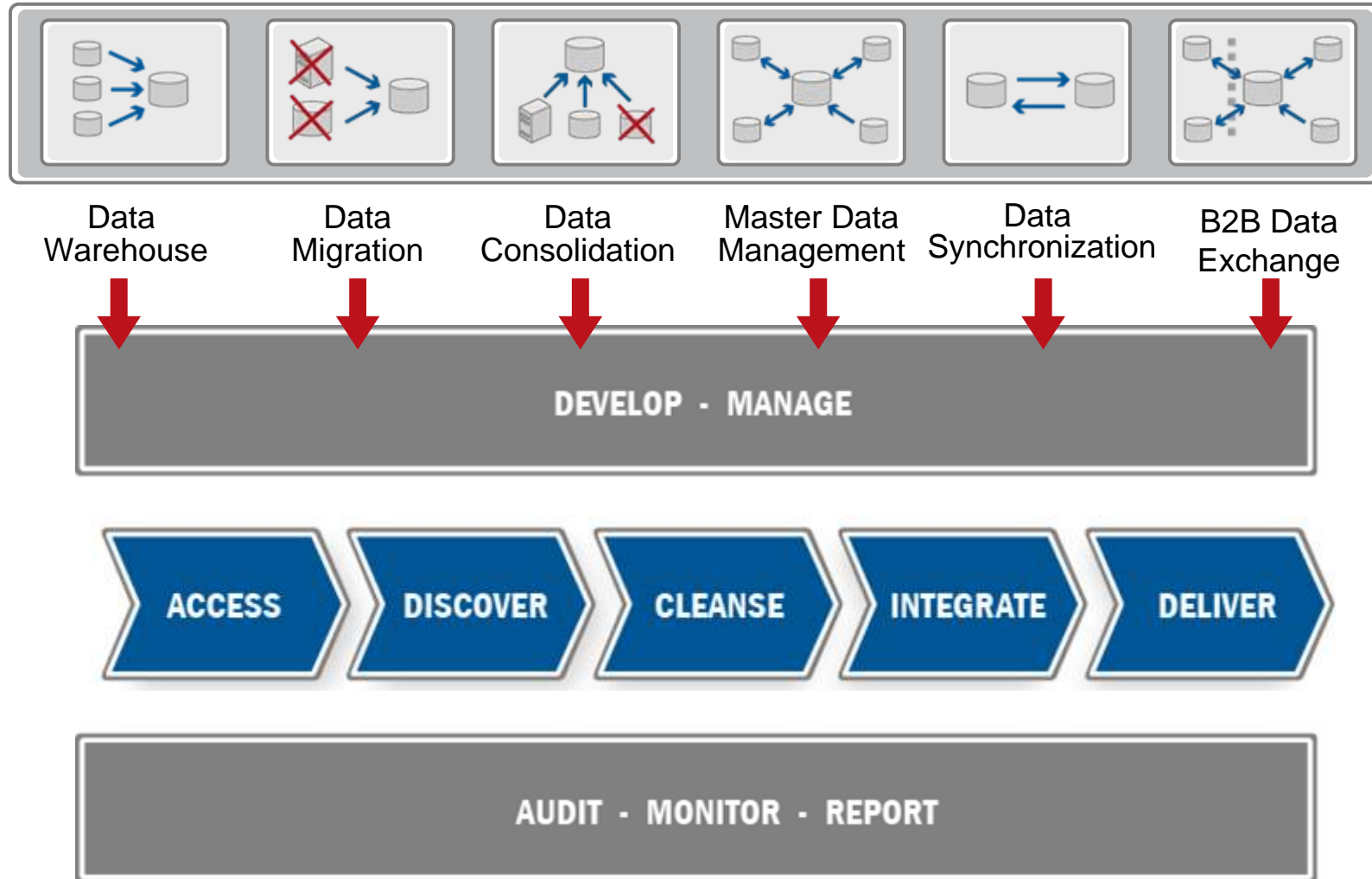
Transportation and Distribution



Utilities and Energy



The Informatica Platform



6. Q&A