

IBM



Analytics Forum

Transforming Industries and Professions 2015

Data Integration - Game Changing Use Cases

Stuart Cranch

Information Governance Leader

IBM Analytics, Asia Pacific



Indian Space Program

Many individuals with myopic vision questioned the relevance of space activities in a newly independent nation, which was finding it difficult to feed its population. Their vision was clear if Indians were to play meaningful role in the community of nations, they must be second to none in the application of advanced technologies to their real-life problems. They had no intention of using it as a means of displaying our might.

Dr APJ Abdul Kalam - former Indian President

The same theory, in my opinion, applies to the adoption of Analytics



Targeted use of appropriate technology to drive timely valuable outcomes

A \$74 million Indian spacecraft entered orbit around Mars today after an almost year long voyage, and for 11 percent the cost of the U.S.'s Maven probe

“This mission costs less than it takes to make a Hollywood movie,”

Modi - 24 September 2014.



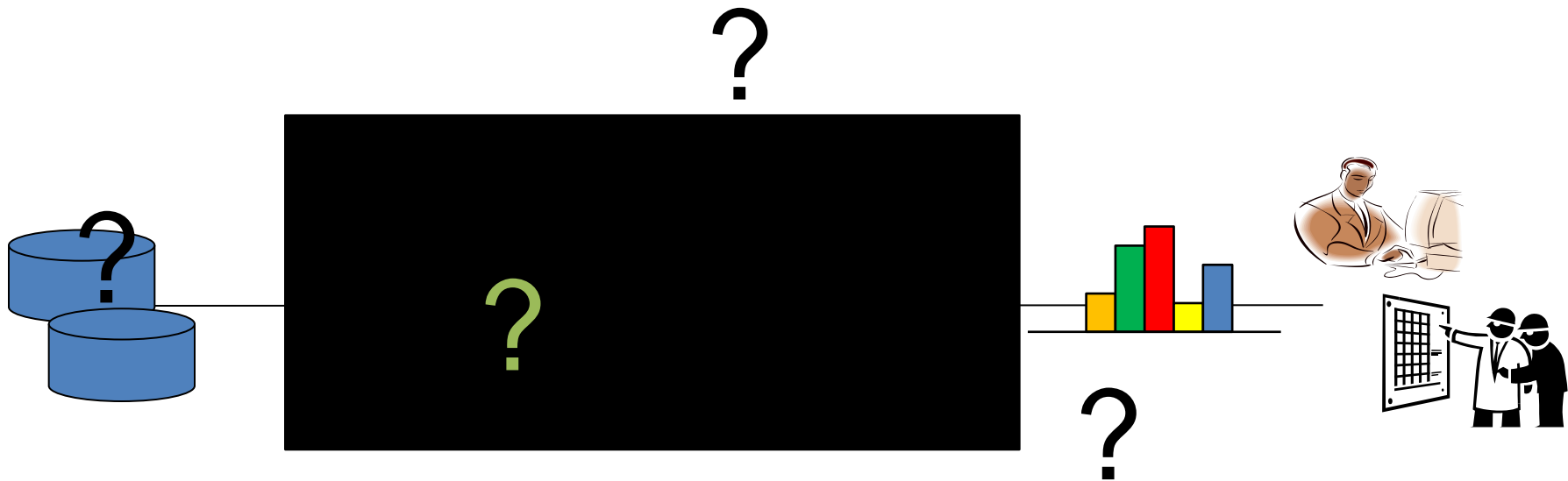
So what

does this mean to you?

Your global competitors (and local ones) are implementing Information Governance technologies to support Analytical outcomes to drive revenue and reduce costs



Information Landscape



Sources, targets, intermediate landing areas, transformations, processing platforms, web services, definitions, models, staging areas, external systems, and more...

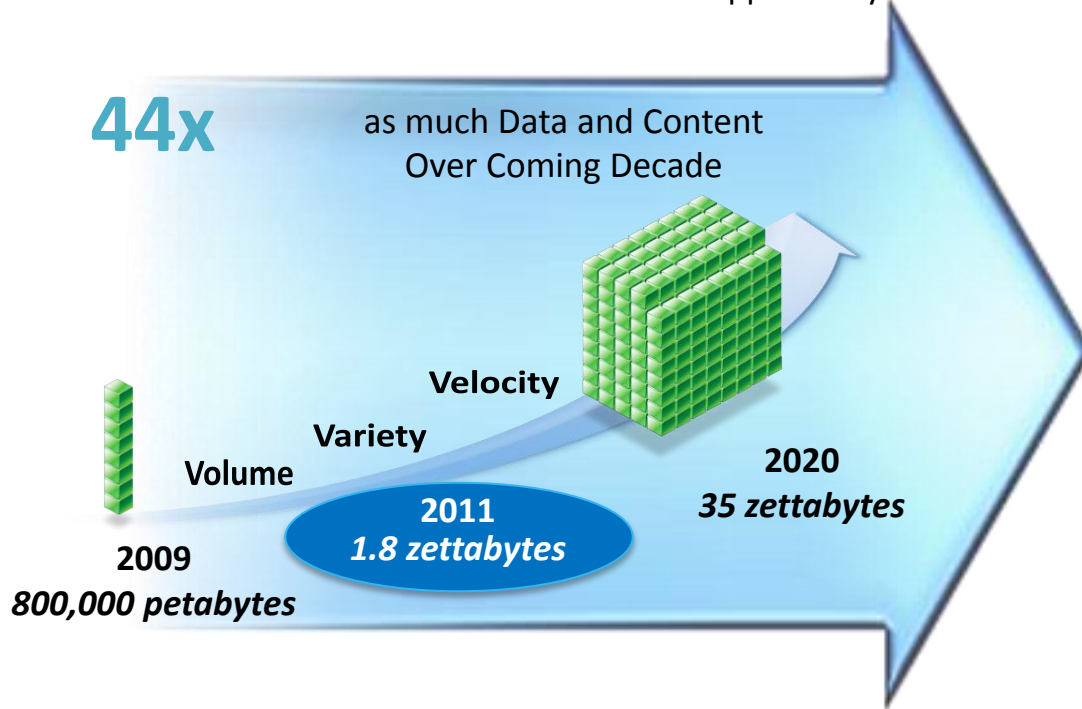
How big is your “black box”?



We All Know Data is Growing ...

... and the number of sources and consumers of that data is growing too

Information is at the Center Of a New Wave of Opportunity...



Number of data sources is growing

Number of consumers is growing

Total number of integration points is growing

Information Integration & Governance is more important than ever before

The Big Data Paradox: More Data, Less Confidence

Big data increases importance of delivering trusted information to your analytical platforms



1 in 3

Make decisions on untrusted information

1 in 2

Don't have necessary information

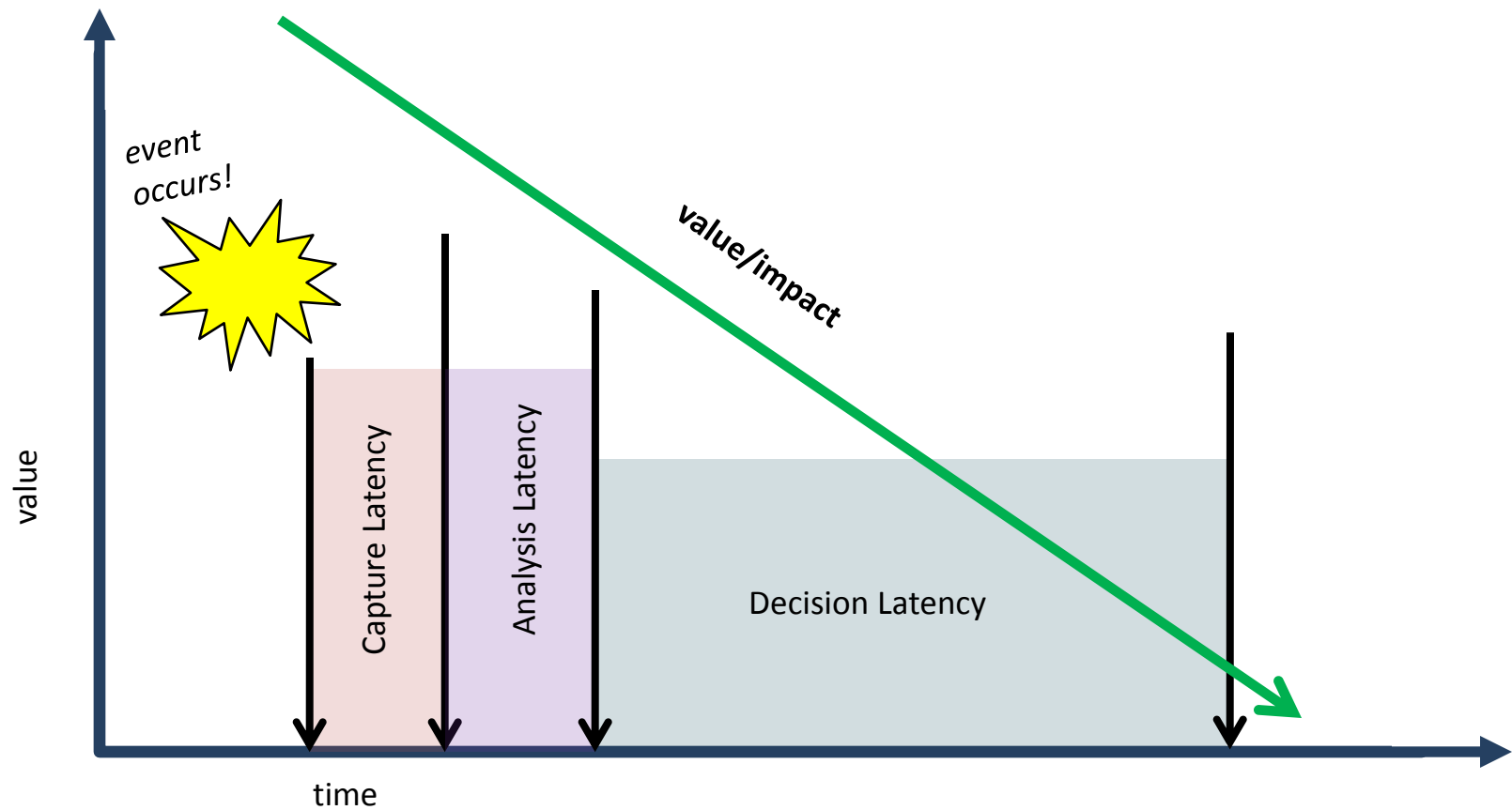
60%

Have more data than they can use

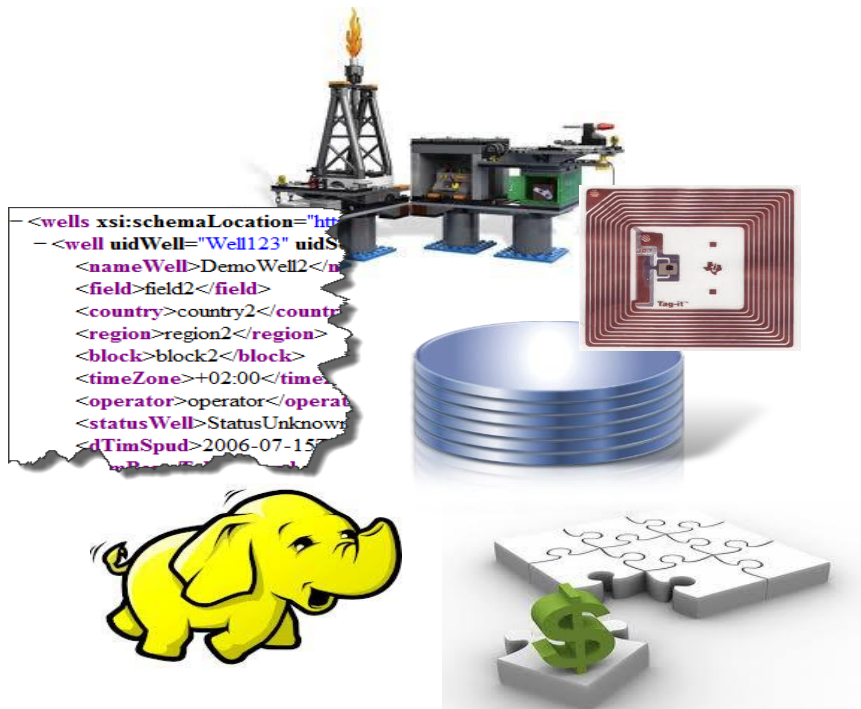
40%

Time spent on each big data project to understand information

Decision Latency



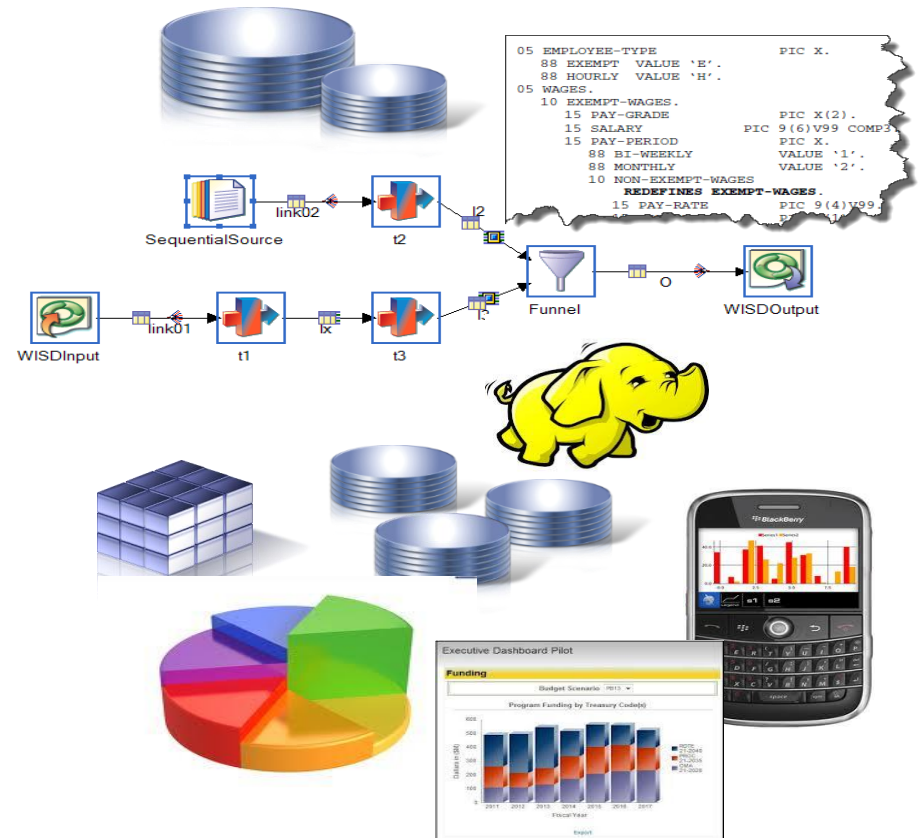
Variables Increasing Decision Latency: Capture



- What sources are critical to your organisation? y, finance, HR, operations...
- ...and also health and safety, trading systems, Customer Service, Real-time Analytics...
- Application systems such as SAP, JDE, Oracle Financials
- Protocols and Standards, XML, HL7, FATCA, EBICS, etc.
- Web Services, batch, real time messaging...
- ...and all of it in repositories, already in Big Data, or spread across the enterprise

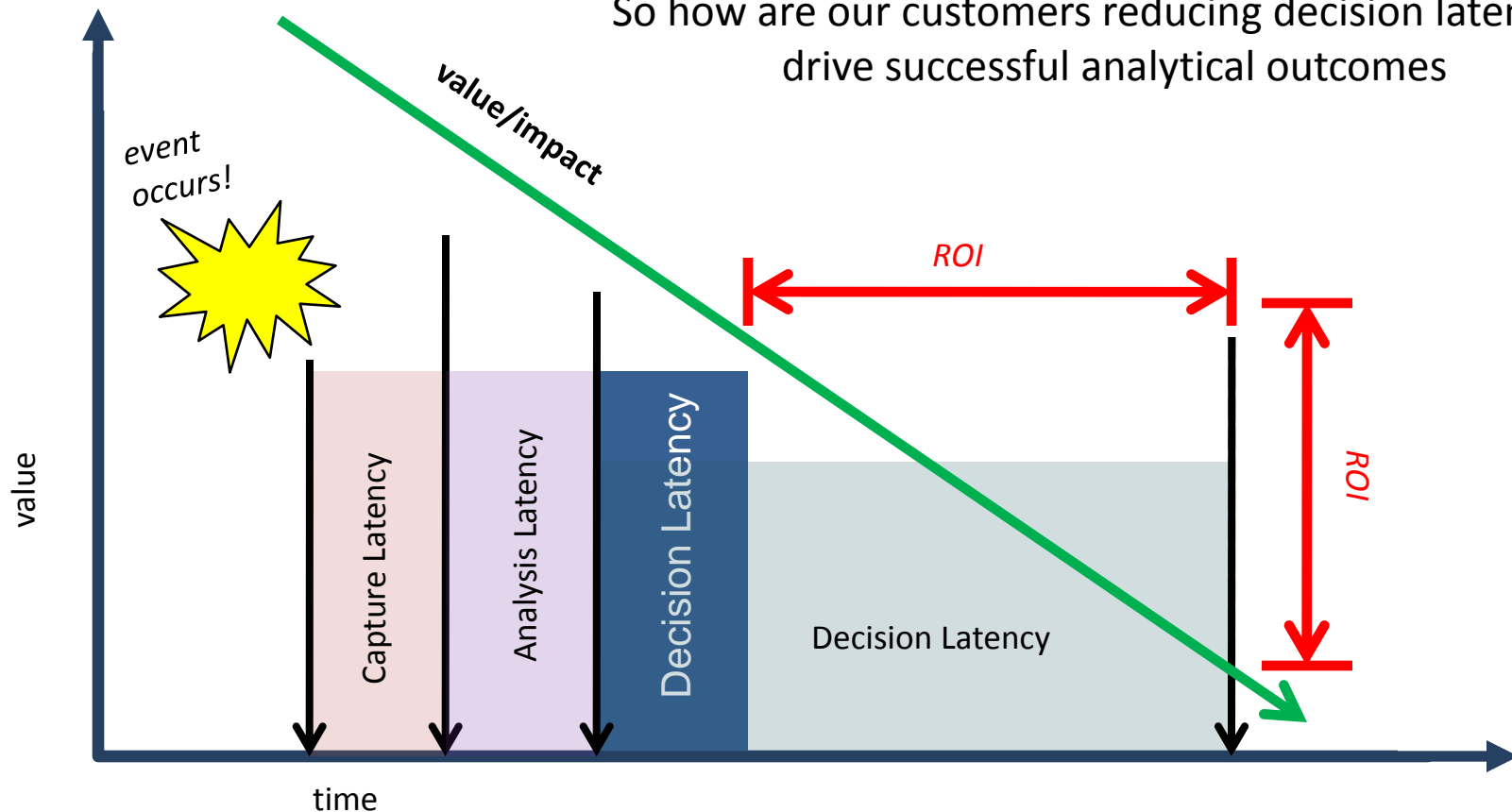
Variables Increasing Decision Latency: Analysis

- Complexity of how data is delivered: ETL, EAI, Big Data, Hadoop, custom Java, messaging, FTP scripts, stored procedures, COBOL
- Requirements to transform more data in less time!
- Structural transformation, aggregation, statistical analysis, data quality cleansing, etc.
- Distribution to data marts, spreadsheets, cubes, Excel, Portals, Smart phones, Reporting tools, social media, alerts, emails



Decision Latency

So how are our customers reducing decision latency to drive successful analytical outcomes



How do we reduce Decision Latency?

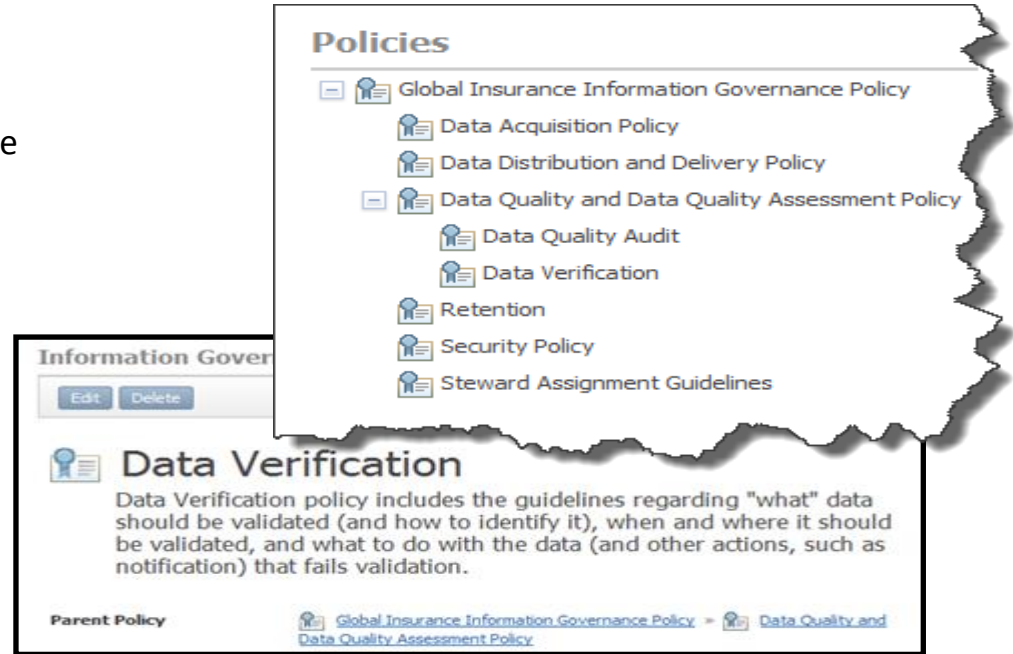
- Develop policies and rules for information governance
- Establish a common vocabulary
- Define and meet data quality objectives
- Exploit new architectures for massive parallelism and transform more data more quickly, enabling faster decision making for critical indicators
- Retire data securely that no longer required
- Find the data gold in your customers and prospects
- Enable lineage reporting that includes visibility to all aspects of the information supply chain
 - Link goals-objectives-definitions-IT artifacts-documents-people
 - Help decision makers find the right resources and people “in the know”
 - Increase confidence in the facts and trust in the data

Successful Customers have recognised this is a journey rather than a one off project.



Develop Policies and Rules for Information Governance

- Policies document a organisations information governance objectives
- Establish guidelines regarding
 - Use of data throughout the enterprise
 - Achievement and maintenance of data quality
 - Retirement of data
- Promote stewardship and define how stewards are selected.
- Outline security measures and privacy requirements



Policies

- [-] [icon] Global Insurance Information Governance Policy
 - [icon] Data Acquisition Policy
 - [icon] Data Distribution and Delivery Policy
- [-] [icon] Data Quality and Data Quality Assessment Policy
 - [icon] Data Quality Audit
 - [icon] Data Verification
- [icon] Retention
- [icon] Security Policy
- [icon] Steward Assignment Guidelines

Information Governance

Edit Delete

[icon] **Data Verification**

Data Verification policy includes the guidelines regarding "what" data should be validated (and how to identify it), when and where it should be validated, and what to do with the data (and other actions, such as notification) that fails validation.

Parent Policy [Global Insurance Information Governance Policy](#) > [Data Quality and Data Quality Assessment Policy](#)

This is an enterprise initiative, however successful organisations pilot and grow with business taking ownership.

Establish a Common Vocabulary

- Common understanding and agreement of terminology that is critical to the business; a glossary
- Driven by the business; available to the entire enterprise
- Increase collaboration between business and IT
- Terms are governed, vetted, approved at highest levels. Stewardship is a critical component

Location

Office Location

Drilling Location

Accident Location

...and how important is “accuracy” as a characteristic?

Equivalent Distillation Capacity

UEDC

EDC

eq_dst_cp

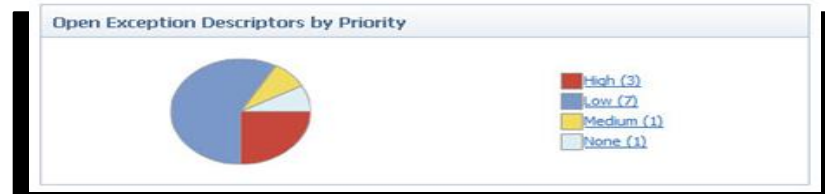
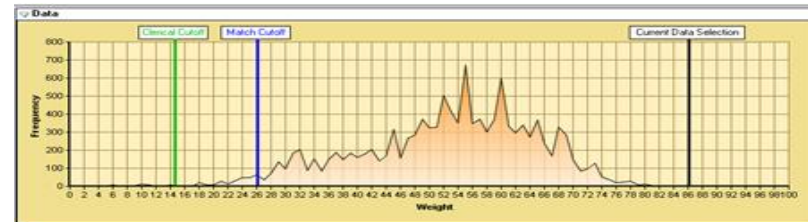
Established a Information Governance group to manage the gap between business and IT



Define and Meet Data Quality Objectives

- How good is our data? Exactly what is in it? How do we know?
- Who owns it?
- Who analyzed it? How recently?
- When was the data refreshed?
- Are we fixing our data quality problems? How are they being managed and monitored?
- Do we have metrics for measuring our data quality conditions?

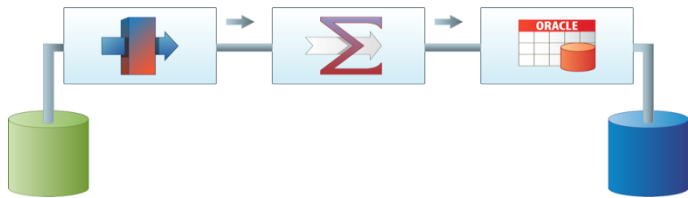
Column Frequency	
Most Frequent Formats	
Project :	GlobalCoIntegration
Report Name :	Most Frequent Formats - WorldCoShipTo StoreID
Date/ Time Executed :	01/07/2007 12:09:21
Generated By :	infosrvr admin
Customer Description :	WorldCo ShipTo StoreID Formats
Host Name :	IBM-9E7D87EBAF0
Data Store :	SQLServer01
Data Store Alias :	
Database Name :	dbo
Table Name :	WorldCoShipTo



Measurement is key, fixing it is optional and dependent of business drivers and outcomes.

Exploit New Architectures – Not all Data Integration is equal

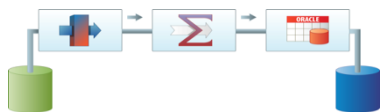
Application Assembly: One Dataflow Graph



Codeless creation of data integration logic and jobs, reusable across the enterprise

Unlimited data scalability wherever data integration processing occurs. Run in ETL, Database or Hadoop.

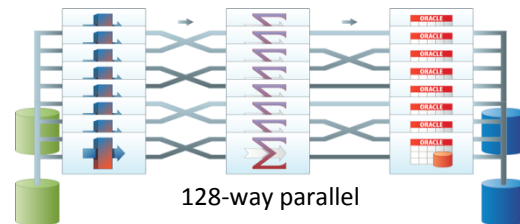
Application Execution: Sequential or Parallel



sequential

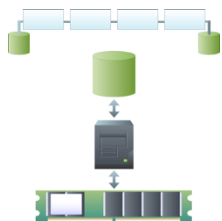


4-way parallel

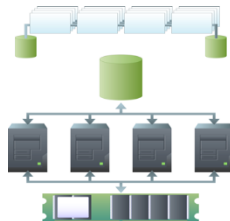


128-way parallel

Hardware Platform



Uni-processor



SMP



128 Processor Grid, Cluster, or MPP, Hadoop



Retire data securely that no longer required

Use Case

Identify relevant data

- Dormant vs. active data
- Related data / complete business object
- Sensitive data

Value

- ✓ Time to market / productivity
- ✓ Higher quality outcome

Use Case

- Application testing
- Outsourced dev & test
- DevOps, Mobile First

Value

- ✓ Time to market
- ✓ Higher quality
- ✓ Lower cost



- Retaining & disposing data for compliance
 - Application retirement / consolidation
 - Operational efficiency

Use Case

Value

- ✓ Reduce risk
- ✓ Reduce cost
- ✓ Improve performance

Use Case

Value

- Data privacy compliance
 - Outsourcing
 - Monetising data

- ✓ Reduce risk
- ✓ Capitalize new opportunity

Find the data gold in your customers and prospects

8%

... but only of their customers agree!

80%

of CEOs think they deliver a superior customer experience ...

75%

do not believe companies tell the truth in ads

78%

of consumers trust peer recommendations



86%

use multiple channels

4 in 10

Smart phone users search for an item in a store

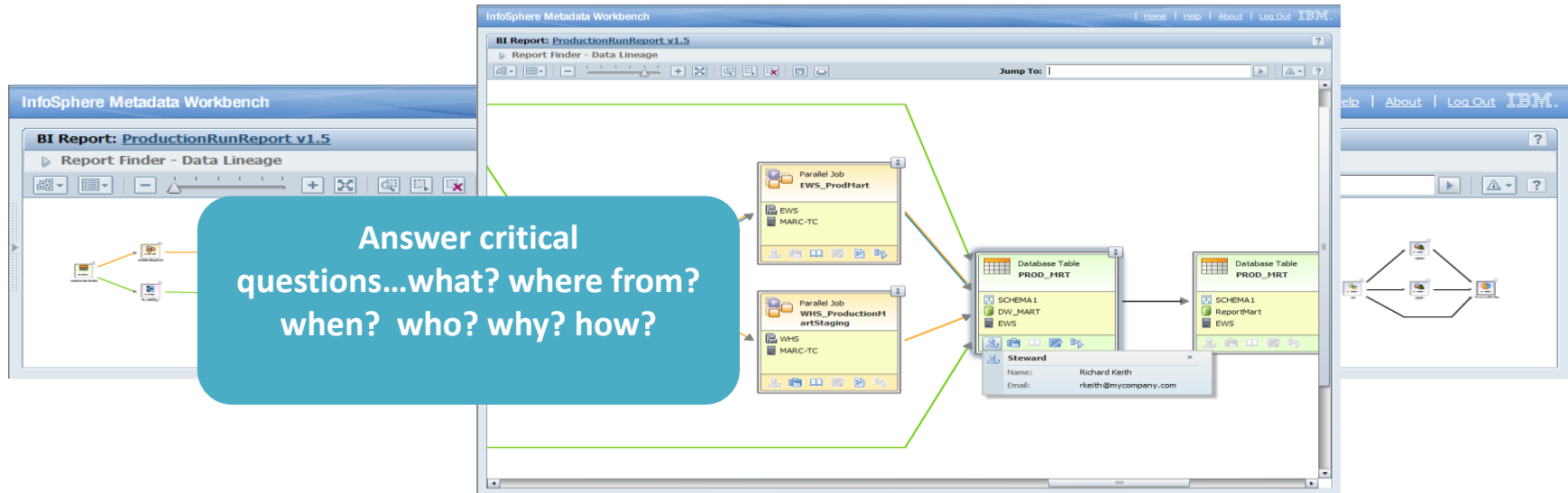
4-5x

more than average is spent by multi-channel buyers

- Know your customers relationships and value to your organisation
- Drive value from existing and new customers
- Our successful customers have embarked on a target approach to master data management

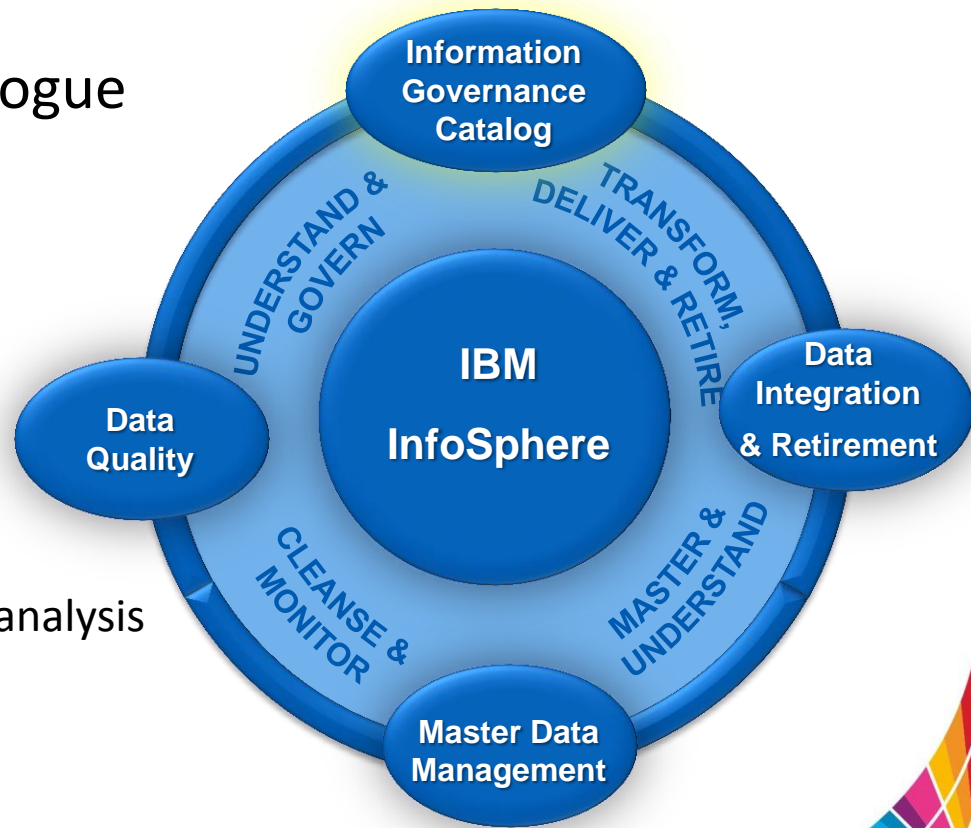
Enable Lineage Reporting

- Lineage reporting illustrates the relationships that tie everything together
- Data, processes, their calculations...and also the people, the statistics, the history, and the meaning, and more



Where to start?

- Information Governance Catalogue
 - Business Glossary sandpit
 - Lineage analysis
- Data Integration & Replication
 - Big Data Latency Assessment
- Data Quality
 - Quality Assessment
- Master Data Management
 - Single view of customer or product analysis
- Data retirement
 - Discovery and analysis



Suncorp-Metway Ltd is integrating data for smarter marketing

Need

- Create a single, integrated view of customers from multiple, previously siloed sources for smarter marketing that limits internal conflict among brands and cuts costs
- Suncorp Insurance customer base is 5 million across 26 brands
- Low growth and margin, with more competitors entering the market
- Providing niche offerings is the key to success
- Customer cannibalization – 1 in 3 new customers is a transfer between brands, costing up to \$100 per transfer

Benefits

- Reduced 23 million source records to roughly 9 million unique accounts that contain all customer data in one place
- Saved approximately US \$10 million annually on integration and associated costs
- Significantly increased market share without increasing marketing spend
- Closed loop marketing leveraging accurate trusted customer data delivery to analytical marketing systems

The Indian state Odisha to reduce fraud and identify the real citizens in need of social welfare benefits through the implementation of a Single View of Citizen Resident Data Hub

Need

- The Department of Food Supplies and Consumer Welfare has been charged under the National Food Security Act to identify the citizens who are the REAL BENEFICIARIES (Below Poverty Line) and the ones who should be awarded Ration cards.
- Requires the State to get a Single view of the Resident from multiple data bases including the National Population Register, voter card, e-Vahan, motorboat, electricity connections, LPG connections, pension and human resource management system Motor vehicle, National Population register, Electricity connection
- Cost of services increasing, welfare not reaching the correct recipients

Benefits

- Improved decision making
- Visibility of beneficiaries and the Socio-economic Improvements of the State
- Scalability for new workloads, data volumes and future data sources types
- Gain a rapid understanding of citizens, applicants and providers across multiple programs and cases
- Provide the right service to eligible citizens with improved access

News Corporation changing business environment requires a change to business as usual with the implementation of smart management of customer information

Need

- The Publisher-Consumer Landscape is changing Along with consumer Brand Loyalty and propensity to pay
- The Publisher-Advertiser Landscape has also changed, Need for targeted ads.
- The need for more information on our audience
- Have basic information, however this is fragmented across business units and channels
- Data volumes are growing, need for a closer interaction, know me, by respect me.
- A whole of News Corp Strategy for processes to manage and monitor

Strategy

- Data Quality and Linkage: Name parsing and 'correctness', Address parsing / PAF match, Phone & Email parsing, Salacious/Profanity & Quality flags, audience connection
- Linkage: Data Services: Master Data Management, Privacy compliance, Governance & Access, Trends and Metrics
- Analytics: Insights and Execution, Formatted for use, Ad hoc data analysis, Modelling & Segmentation

Benefits

- 10% More newsletter article clicks, 10% More articles read per session, Strong Ad revenue growth
- 20% Increase on Churn retention rate (no discounting required)
- Increased engagement, Decreased risk, Increased revenue, Less 'gut feel'



News Corporation changing business environment requires a change to business as usual with the implementation of smart management of customer information

Need

- The Publisher-Consumer Landscape is changing Along with consumer Brand Loyalty and propensity to pay
- The Publisher-Advertiser Landscape has also changed, Need for targeted ads.
- The need for more information on our audience
- Have basic information, however this is fragmented across business units and channels
- Data volumes are growing, need for a closer interaction, know me, by respect me.
- A whole of News Corp Strategy for processes to manage and monitor

Strategy

- Data Quality and Linkage: Name parsing and 'correctness', Address parsing / PAF match, Phone & Email parsing, Salacious/Profanity & Quality flags, audience connection
- Linkage: Data Services: Master Data Management, Privacy compliance, Governance & Access, Trends and Metrics
- Analytics: Insights and Execution, Formatted for use, Ad hoc data analysis, Modelling & Segmentation

Benefits

- 10% More newsletter article clicks, 10% More articles read per session, Strong Ad revenue growth
- 20% Increase on Churn retention rate (no discounting required)
- Increased engagement, Decreased risk, Increased revenue, Less 'gut feel'

Thank you

Your India Information Governance Contact is

Medha Nair

nmedha@in.ibm.com

My Details: Stuart@au1.ibm.com

