

InfoSphere™

Information Platform & Solutions

Trends and Tactics for Improving Data Quality on the Mainframe



Information Management software

Patrick H. Connolly & Krishna Mamidipaka

InfoSphere Product Marketing

Information Centric Business Issues



- Too much information, not knowing what's important
 - Not using demand signals to drive supply chain
 - Not using customer analysis to tailor marketing and sales
 - Not leveraging valuable unstructured information



- Multiple versions of the truth
 - Problems managing customer, product and partner interactions
 - Regulatory compliance inhibited by poor transparency



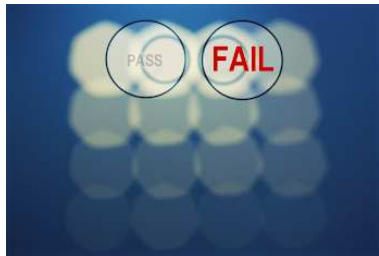
- Lack of trusted information
 - Incomplete, out-of-date, inaccurate, misinterpreted data
 - Difficult to understand or control how information is used



- Lack of agility
 - Inability to take advantage of opportunities for innovation
 - Escalating costs due to inflexible systems and changing needs

The Cost of Dirty Data

83% of data integration projects either overrun or fail



Scrap and rework
Increased costs

Lack of consumer confidence

Inaccurate or incomplete data is a leading cause of failure in business-intelligence and CRM projects



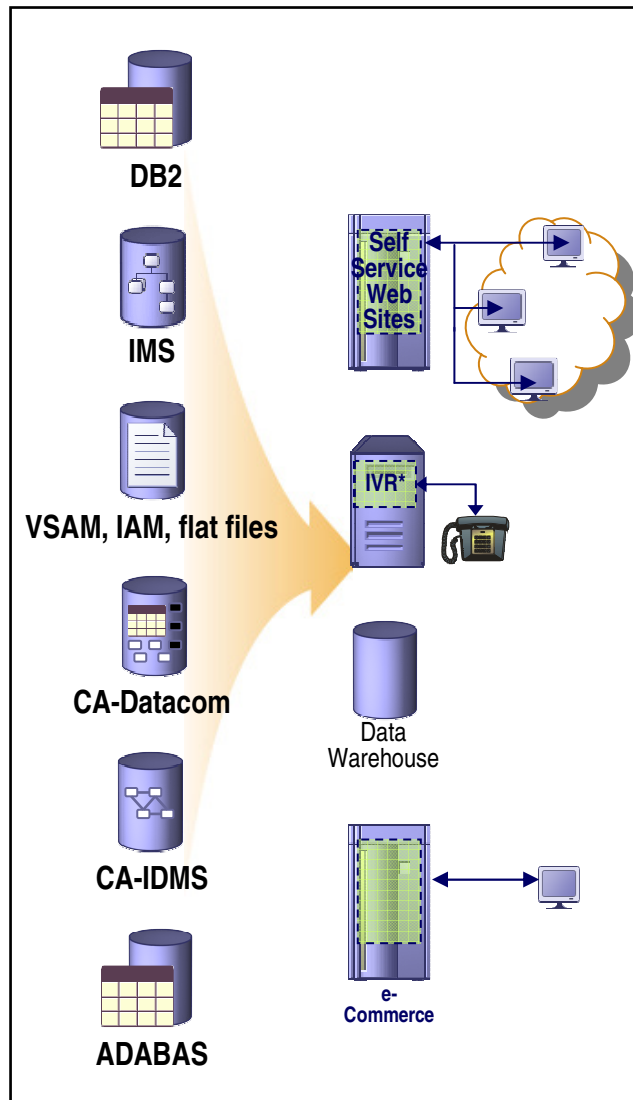
Lost opportunities

Low data quality costs companies \$611 billion annually

25% of time is spent clarifying bad data

Undetected defects will cost 10 to 100 times as much to fix upstream

Data residing on the Mainframe is mission critical



- **Applications of System z data?**
 - Self-service applications
 - Data Warehousing, MDM, BI applications etc
 - Web-based portals for e-commerce
- **Why is management of System z data critical?**
 - 60% of enterprise data is sourced from the mainframe
 - *Could be used directly or replicated to a distributed platform*
 - Increasing application consolidation activity on the mainframe
- **Why is management of System z data challenging?**
 - Proprietary databases – multiple, complex APIs
 - Dearth of skills – scarce and becoming more costly
 - Integration with modern initiatives not native to this data

Trusted data on the mainframe is necessary to ensure the success of various business and IT initiatives and hence be competitive in the market

What the Analysts are Saying...

- **“Data integration and data quality are no longer simply technology tools, they must enable an organizational competency responsible for delivering Trusted Data.”**
 - *Rob Karel, Forrester Research*
- **“If you look at...any business function in your company, you're going to find some direct cost there attributed to poor data quality.”**
 - *Ted Friedman, Gartner*
- **“For organizations in a competitive environment, data quality is a matter of survival and then of competitive advantage. For organizations in the public and not-for-profit sectors, data quality is a matter of survival and then of stewardship of stakeholder (taxpayer or contributor) resources.”**
 - *Larry English, Information Impact International*
- **“Poor quality data costs the typical organization 20% of revenue... if we can free up that 20%, we can create an economic boom that will make the 1990s look like a depression!”**
 - *Tom Redman, Navesink Consulting*

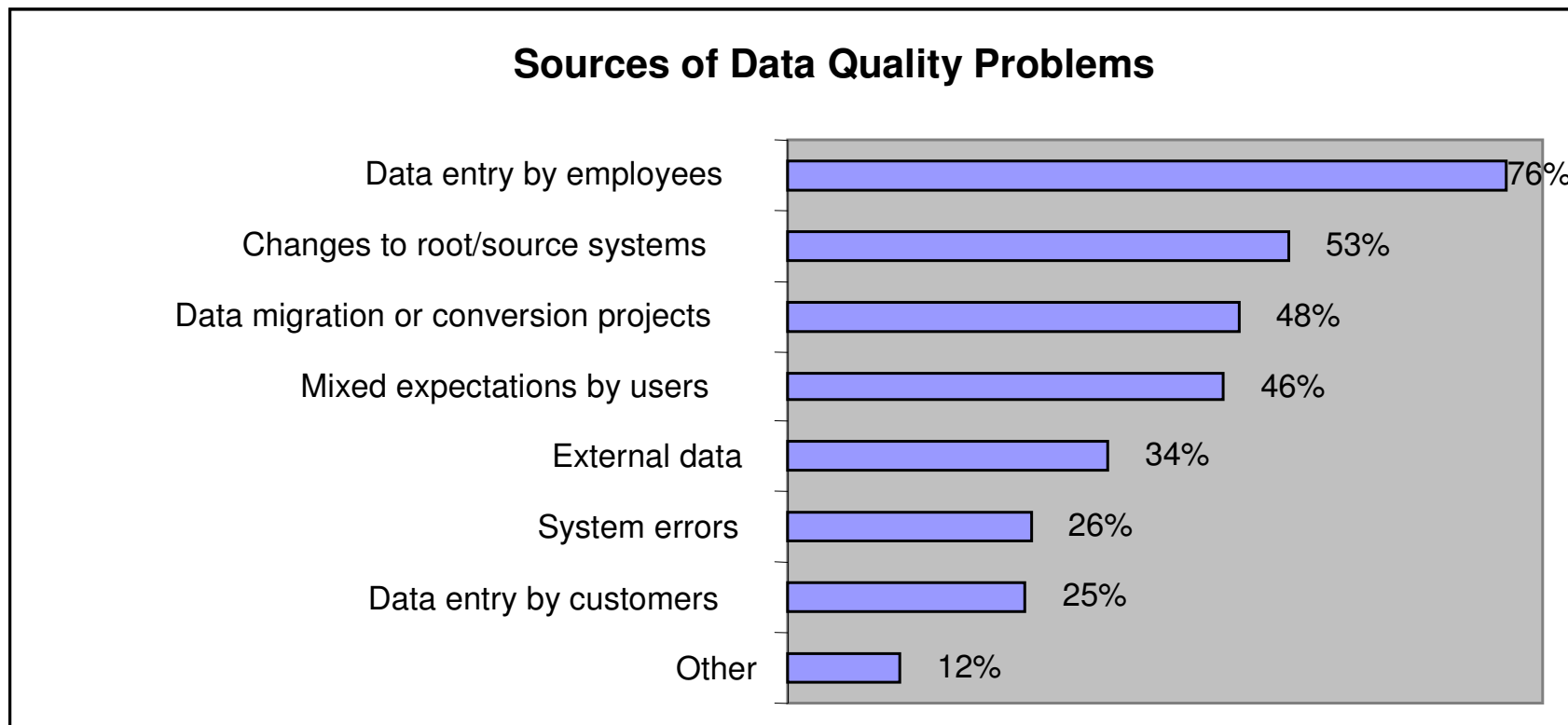
Can you Trust the Quality of Data driving your Business?

- **The challenge of data quality**
 - Good enough is good enough?
- **Multiple applications**
 - Some standards have emerged, but it's still not easy
 - Legacy, add-on, and web systems can compound the issues
- **Multiple instances**
- **Mandatory upgrades**
 - A golden opportunity?
- **Unmet expectations**
 - ROI, cost savings, operational efficiencies



Enterprise Solutions Have a Data Quality Challenge

- **Limited capabilities in monitoring & inability to easily improve the quality of information key systems**



Source: TDWI

Data Quality Approaches

- **Do the best you can for now: Opportunistic cleansing**
 - File cleansing, point-of-consumption cleansing
- **Next best: Cleanse and maintain**
 - Leverage best of breed tools
 - Prebuilt data quality modules
 - SOA-enabled Data Quality
- **Ideal case: Get to the root cause**
 - Implementations
 - Migrations
 - Upgrades
 - Instance Consolidations

Scenario 1: Database-Centric Data Quality

- **Situation**

- Company wants to do a marketing campaign – data lives in a number of different files. Quality is suspect at best (address data), duplicates within and across databases.

- **Solution**

- Service provider cleanses file, sends it back
- Standalone, point solutions

- **Outcomes**

- Usually based upon a specific pain point – crisis response
- Reactive, involving costly labor
- Inconsistent business rules
- After several iterations, it can become a routine process
 - ...but not necessarily addressing the root cause of the quality issue

Scenario 2: Business Intelligence

- **Situation**

- Doing some campaign management and need analytics of their customer and sales data. The data warehouse needs to be the source of truth. Leverage DW for an integrated view of the enterprise information assets and to drive strategic decisions (customer spend analysis, customer and product profitability, etc.
- Need to apply data quality in concert with the integration (ETL) process in order to clean up customer and product data.

- **Solution**

- Traditional ETL Processing combined with data cleansing (native or integrated)

- **Outcomes**

- Strategic decisions can be based upon accurate, consistent information – Information you can Trust

Scenario 3: Application Migration

- **Situation**

- Facing a migration from one or more legacy applications to a single application structure. May include ERP/CRM instance consolidation
- Superb opportunity to clean up data – both in content and structure as well as addressing the duplicates that may exist.

- **Solution**

- Aggressive data quality analysis and cleansing to clean up data in transition

- **Outcomes**

- Without doing this, you run the risk of creating a single, inconsistent version of the truth!
- This is a tremendous opportunity to clean up data – seize it!
- Risk – without ongoing data quality, what happens on Day One after go live?

Scenario 4: Real Time Data Quality

- **Situation**

- One or more application environments for CRM, ERP, or home grown applications. Need a manner for prevention – to proactively empower users and the applications they use to ensure the quality of data

- **Solution**

- One offs/custom solutions have been costly and difficult
- SOA helps a great deal with integration efforts
- Prebuilt modules for data quality provide low risk, low cost of ownership and quick time-to-value

- **Outcomes**

- Estimates vary, but a large number of data quality errors are the result of data entry anomalies
- Preventative measures like these make a big difference
- Create a firewall for data quality at the source!

Scenario 5: Master Data Integration

- **Situation**

- Your organization has decided to embark on an MDM (or CDI) initiative. This can include any and all of the previous scenarios in terms of data quality challenges – to gain a single version of the truth, you need clean, concise data.
- Doing MDM without a comprehensive approach to Master Data Integration (MDI) including data quality will never deliver consistent results.

- **Solution**

- Master Data Integration – a comprehensive approach to integrating, standardizing and cleansing data destined for a Master Data repository.

- **Outcomes**

- Successful MDM relies on MDI including Data Quality at multiple points in order to ensure that you're harmonizing accurate, consistent information

So, What Constitutes Data Quality?

- Data is standardized
- Data is fit for purpose (conforms to rules)
- Each record is unique
- View of information is complete
- Records are certified against authoritative sources
- Lineage is understood
- Data quality is measured over time



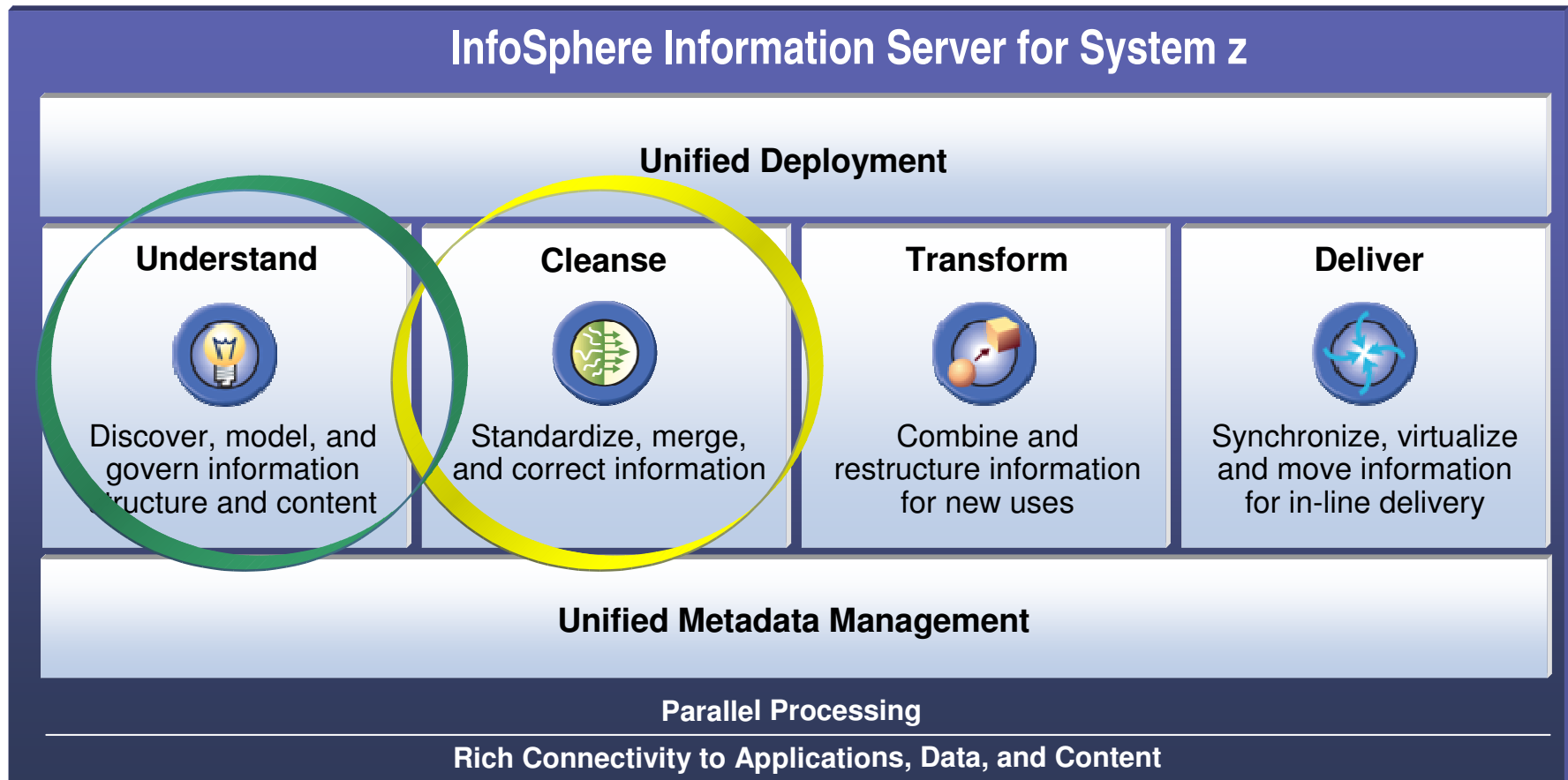
What Do You Need to Establish a Data Quality Program?

- **A foundation platform that centralizes quality rules and provides auditable data quality**
- **Business-driven, data-centric design environment for data quality rules**
- **An ongoing process for data quality**
- **A way to measure quality over time**
- **Universal deployment of quality rules across all points of entry**
- **Data quality ownership and data governance**
- **Management sponsorship and a corporate mandate for data quality improvement**



IBM InfoSphere Information Server

A Platform Enabling Enterprise Data Quality



Data Profiling & Analysis: InfoSphere Information Analyzer for System z

- **Native interface to multiple types of Mainframe data sources. Keep z data on z.**
- **Analyzes data sources to discover *structure, contents and quality* of information**
 - Finds missing, inaccurate and inconsistent data
- **Provides secure in-depth analysis of heterogeneous data sources**
 - Column Analysis
 - Table Analysis
 - Primary Key Analysis
 - Foreign Key Analysis
 - Duplicate Analysis
- **Enables ongoing measurement and baseline reporting of information quality**
- **Results instantly shared across IBM Information Server**



Subject Matter Experts



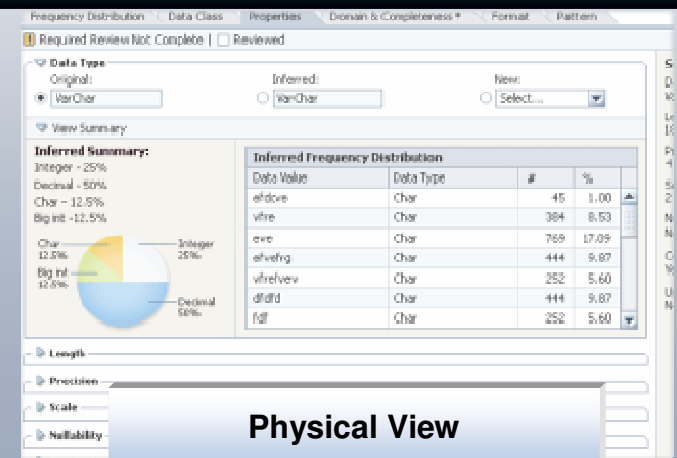
Data Analysts

Understand



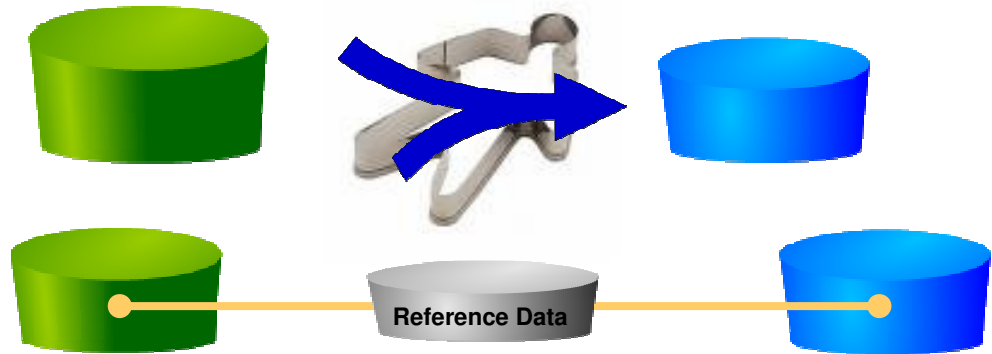
InfoSphere Information Analyzer for System z

Analyze source data structures, and monitor adherence to integration and quality rules



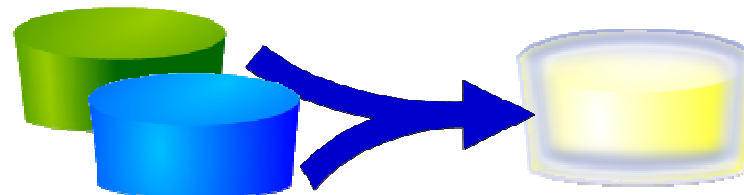
How Can We Cleanse Data?

- Standardization



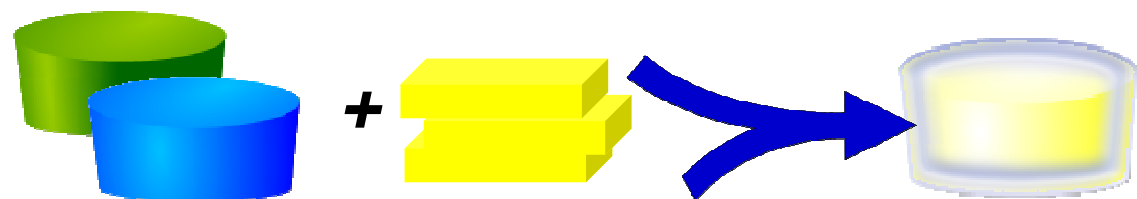
- Verification

- Identify Matches & Duplicates



- Manage Matches

- Enrich Data



Data Cleansing: InfoSphere QualityStage for System z

- **Native interface to multiple types of Mainframe data sources. Keep z data on z.**
- **Provides specialized data quality processing**
 - Ensures clean, standardized, de-duplicated information
 - Enables a single version of the truth
 - Supports global postal verification
- **Provides visual tools for designing quality rules and matching logic**
 - Seamlessly integrated with DataStage (one engine, one metamodel, one UI)
 - Precisely calibrates matching rules
- **Allows quality logic to be deployed seamlessly within ETL, or as shared services**



Subject Matter Experts



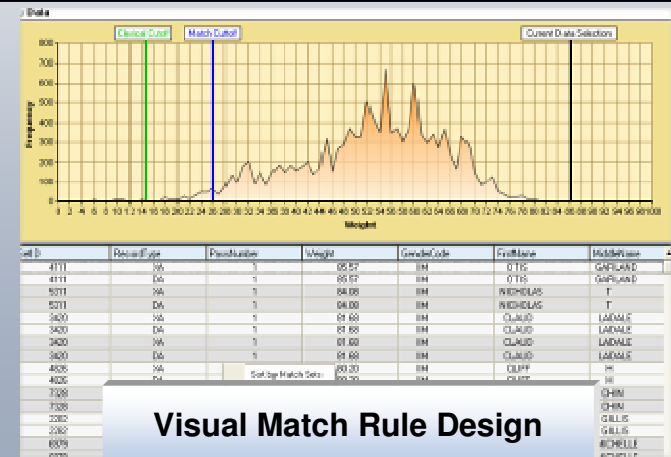
Data Analysts

Cleanse



InfoSphere QualityStage™ for System z

Standardize and correct source data fields, and match records together across sources to create a single view



Visual Match Rule Design

How Can IBM Help?

- **Comprehensive platform for data quality**
- **Experience and repeatable process for helping organizations set up data quality programs**
- **Domain and industry-specific expertise in establishing repeatable data quality services**
- **Data quality assessment offering to report on existing data quality and establish the business value of a data quality program**
- **Contact your IBM representative for more information**

The Data Quality Assessment Offer

- Do your key initiatives sometimes come in late, over budget, and then fail to deliver the information promised? A lack of complete, valid, accurate and relevant facts may be the root cause
- Let IBM shine the light on your data sources to determine if the content, quality, and structure of your data meets your business requirements
- IBM experts are ready to work with your organization to profile your data sources and quickly establish the quality of your data
- Components of the DQA
 - **Register** – Discuss your requirements to better understand the quality of your data
 - **Discovery Call** – Get into the details and define the scope for the DQA. Identify the number of sources, the number of records or rows, the attributes or fields of interest and how the data should be organized for the data quality assessment
 - **Kick Off** – Confirm the scope with the execution team, answer any outstanding questions and establish the timeline for the DQA
 - **Results & Recommendations** – IBM personnel present the results of the data quality assessment including the written DQA Report and appropriate recommendations

Resources

IBM InfoSphere Information Server for System z

http://www-306.ibm.com/software/data/integration/info_server_system_z/

IBM InfoSphere Information Analyzer for System z

http://www-306.ibm.com/software/data/integration/information_analyzer/

IBM InfoSphere QualityStage for System z

<http://www-306.ibm.com/software/data/integration/qualitystage/>

IBM InfoSphere Information Server for System z – Connectivity Software

<http://www-306.ibm.com/software/data/integration/connectivity/>



All you need to know about:

- > IBM Data Management > Information Integration > Master Data Management > Enterprise Content Management > Business Intelligence & Performance Management

For:

- > Business and IT Executives > Managers
- > IT Professionals > DBA's > Developers

IBM Information

>>> On Demand

2008

Register Now!

October 26-31, 2008

Mandalay Bay Las Vegas, Nevada

Expect more in 2008!

750+ business leadership and technical education sessions

200 Customer speakers

Free certification tests for all IBM SWG products

www.ibm.com/events/informationondemand

2007 Highlights:
www.ibm.com/software/data/conf/2007/

