

Data Quality Assessment Methodology Course

Highlights

- **Increased scoping accuracy and development of realistic implementation timelines**
- **Faster ETL job development**
- **Improved ETL job accuracy**
- **Higher levels of data quality**
- **More accurate reporting**

Unlock the value of enterprise data

Organizations face an information challenge. Where is it? How do I get it when I need it in the form I need? What does it mean? What insight can I gain from it? Can I trust it? How do I control it? The list goes on, and the challenges grow unceasingly if organizations cannot ensure they have access to authoritative, consistent, timely and complete information.

Services from IBM Information Integration Solutions complement the IBM WebSphere® Information Integration platform, which integrates and transforms any data and content to deliver information clients can trust for their critical business initiatives. Designed to ensure customer success with IBM solutions, IBM information integration services range from technical support to education to professional services.

Maximize your software investment with IBM's proven implementation methodology

You've made an investment in the industry-leading enterprise data integration software suite. Now, the question is: Do you know how to leverage its implementation for the greatest impact? IBM's Data Quality Assessment Methodology course teaches your implementation team our proven method of leveraging our products for maximum impact.

Learning objectives

At the end of this course, your implementation team will be able to:

- *Describe the components of a data quality assessment (DQA)*
- *Identify necessary subject matter experts and describe their roles and responsibilities in a DQA*
- *Identify the appropriate IBM WebSphere tools for performing a DQA*
- *Identify the appropriate tool for measuring a particular facet of data quality*
- *Describe the options for storing, communicating and acting on the analysis results of a DQA*

Business value

By applying the methodology learned in this course, your implementation team will be able to deliver:

- *Realistic project timelines based on your unique data quality issues.*
- *Accurate project scoping that encompasses only the required and best data sources.*
- *The highest levels of data quality required to fully realize your enterprise applications' true potential and attain their business objectives.*

The Data Quality

Methodology in action

To better understand the course and its benefits, let's take a look at the DQA methodology being applied to a data element.

In the example on the right, through the application of the methodology, we've identified that social security number (SSN) will be used as a unique identifier in the new target.

We'll now apply the methodology to the data source to identify potential data quality issues surrounding the SSN.

What customers are saying

- *"Excellent presentation of the methodology and function of the product. One of the best mixes of methodology and application of any software class I have attended"*
- *"The class was very educational, informative and very professionally taught"*
- *"It's always more effective to learn from someone who's actually executed (the lesson), and [the instructor] has a wealth of experiences that he was able to leverage as examples for our class"*

IBM WebSphere Information Integration Products covered in this course

- *IBM WebSphere ProfileStage™*
- *IBM WebSphere QualityStage™*
- *IBM WebSphere AuditStage™*

IBM Information Integration Solutions Professional Services

To help you achieve success in meeting your business objectives through information integration solutions, and to accelerate your solution design and implementation, the IBM Information Integration Solutions team offers a broad range of planning, education, design and engineering, and implementation services. Whatever industry your company serves, our Professional Services team can help you implement the right IBM WebSphere Information Integration solution.

For more information

For more information about this or any other IBM Information Integration Solutions Professional Services offerings, please contact your IBM marketing representative or visit <http://www-306.ibm.com/software/data/integration>

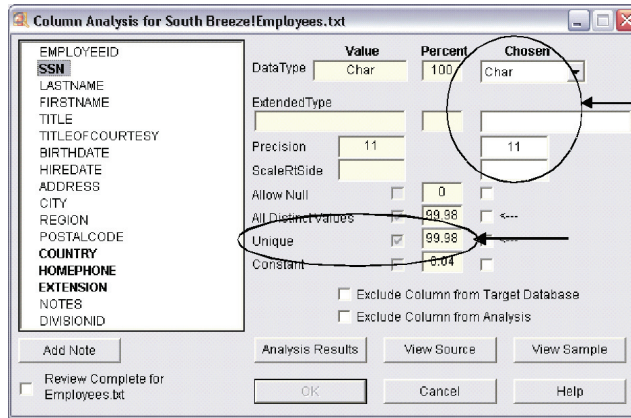
The Data Quality Methodology in Action

Step One:

Identify data domains requiring analysis

Step Two:

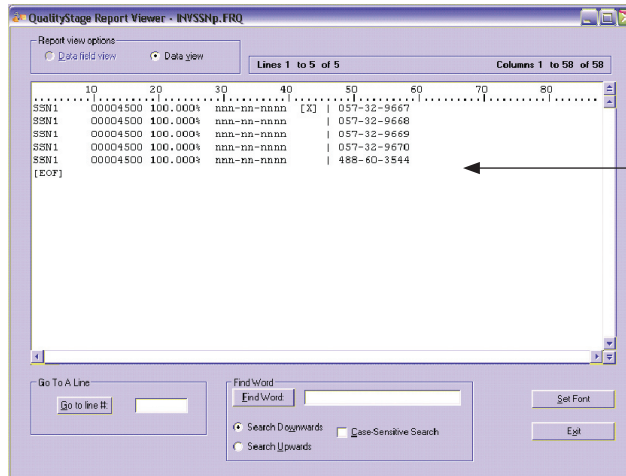
Measure the domain and structural quality facets and compare against target requirements



Column Analysis has revealed that the data is currently being stored as an 11-byte character. Our gap analysis has revealed that the new target will be storing SSN as a 9-byte integer. In step three, we'll identify how SSN is currently being formatted and provide that crucial information to our ETL programmers. Column analysis has also revealed that SSN is not currently a unique value.

Step Three:

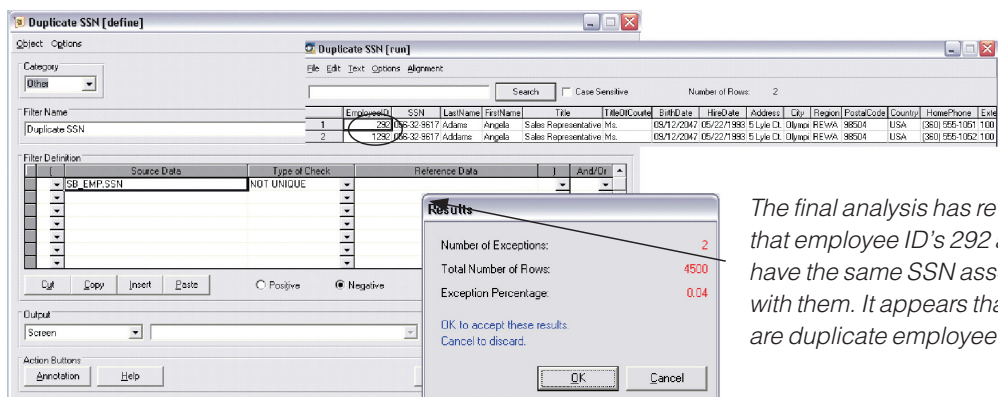
Based on the Column Analysis results, identify the format of the domain



Further analysis has revealed that the field is shown to be always formatted as a string of 3 numerics, hyphen (-), 2 numerics, hyphen (-), 4 numerics.

Step Four:

Identify exceptions to business rules for further analysis and normalization



The final analysis has revealed that employee ID's 292 and 1292 have the same SSN associated with them. It appears that these are duplicate employee records.



© Copyright IBM Corporation 2005

IBM Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
12-05
All Rights Reserved

AuditStage, IBM, the IBM logo, ProfileStage, QualityStage and WebSphere are trademarks of International Business Machines Corporation in the United States, other countries or both.

Other company, product or service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice.

The IBM home page on the Internet can be found at **ibm.com**