

Informatica PowerCenter Data Profiling Option

Assessing the Initial and Ongoing Quality of All Enterprise Data

Dave Ramos

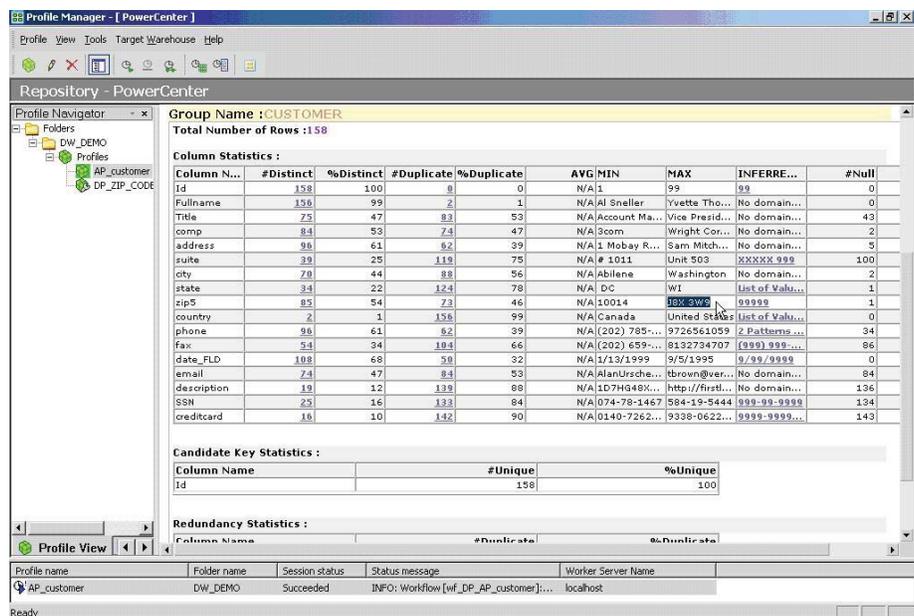
Dir. of Business Intelligence & Analytics
LinkShare

“We’re impressed with Informatica’s ability to continue to deliver a wide range of capabilities across their product line that address, and even anticipate, the demands placed on organizations and hence on their technology. Release 8.5 maintains that tradition. There is so much in this release that reveals that this is a company that listens to its customers, and then tries to deliver more.”

BENEFITS

- Reduce initial and ongoing data quality assessment time
- Lower development costs by automating data discovery
- Accelerate time to delivery by reducing risk and rework

The PowerCenter® Data Profiling Option extends PowerCenter’s ability to provide comprehensive, accurate information about the content, quality, and structure of data in virtually any operational system. Fully integrated into PowerCenter, these data profiling capabilities automatically leverage the platform’s unified development environment, performance, metadata capabilities, and universal data access. With this option, developers can automatically assess the initial and ongoing structure and quality of data regardless of its location or type.



Profile Manager - [PowerCenter]
Repository - PowerCenter
Group Name : CUSTOMER
Total Number of Rows : 158

| Column Name | #Distinct | %Distinct | #Duplicate | %Duplicate | AVG | MIN | MAX | INFERRE... | #Null |
|-------------|-----------|-----------|------------|------------|-----|---------------|-----------------|-----------------|-------|
| id | 158 | 100 | 0 | 0 | N/A | 1 | 99 | 99 | 0 |
| Fullname | 138 | 99 | 2 | 1 | N/A | Al Sneller | Yvette Tho... | No domain... | 0 |
| Title | 25 | 47 | 83 | 53 | N/A | Account Ma... | Vice Presid... | No domain... | 43 |
| comp | 84 | 53 | 74 | 47 | N/A | Scom | Wright Cor... | No domain... | 2 |
| address | 96 | 61 | 62 | 39 | N/A | Mobay R... | Sam Mitch... | No domain... | 5 |
| suite | 39 | 25 | 119 | 75 | N/A | # 1011 | Unit 503 | XXXXX 999 | 100 |
| city | 79 | 44 | 88 | 56 | N/A | Abilene | Washington | No domain... | 2 |
| state | 34 | 22 | 124 | 78 | N/A | DC | WI | List of Valu... | 1 |
| zip5 | 85 | 54 | 73 | 46 | N/A | 10014 | 86X 3WB | 99999 | 1 |
| country | 2 | 1 | 156 | 99 | N/A | Canada | United States | List of Valu... | 0 |
| phone | 96 | 61 | 62 | 39 | N/A | (202) 785... | 9726561059 | 2 Patterns... | 34 |
| Fax | 54 | 34 | 104 | 66 | N/A | (202) 859... | 6122794707 | (999) 999... | 86 |
| date_FLD | 188 | 68 | 58 | 32 | N/A | 1/13/1999 | 9/5/1995 | 9/88/999 | 0 |
| email | 74 | 47 | 84 | 53 | N/A | AlanUrsche... | brown@ver... | No domain... | 84 |
| description | 19 | 12 | 139 | 88 | N/A | ID7HG48X... | http://first... | No domain... | 136 |
| SSN | 25 | 16 | 133 | 84 | N/A | 074-78-1467 | S84-19-5444 | 999-99-9999 | 134 |
| creditcard | 16 | 10 | 142 | 90 | N/A | 0140-7262... | 9338-0622... | 9999-9999... | 143 |

Candidate Key Statistics :

| Column Name | #Unique | %Unique |
|-------------|---------|---------|
| id | 158 | 100 |

Redundancy Statistics :

| Column Name | #Duplicate | %Duplicate |
|-------------|------------|------------|
| id | 0 | 0 |

Profile View

| Profile name | Folder name | Session status | Status message | Worker Server Name |
|--------------|-------------|----------------|--|--------------------|
| AP_customer | DW_DEMO | Succeeded | INFO: Workflow [wf_DP_AP_customer]:... | localhost |

In virtually any operational system, the Data Profiling Option extends PowerCenter’s ability to furnish comprehensive, accurate information about the quality, content, and structure of data.

Data Profiling Option Key Features

Data Profiling

- Reduces the time it takes for developers to assess the structure, content, and quality of data by supplying a single interface for the entire profiling process
- Instills confidence in data by automatically and accurately profiling any data accessible to PowerCenter
- Provides the choice of scanning every single data record or using algorithmic sampling of source systems to deliver accurate assessments
- Enables developers to create custom profiling rules, which can be used to validate data and ensure that data is transformed to meet the rule condition on an ongoing basis
- Displays data profiling results immediately through an interactive profiling mode and provides ongoing data profiling and quality metrics through a batch mode

Full Integration Across Entire Data Integration Platform

- Leverages PowerCenter's performance, scalability, and comprehensive data access
- Automatically harnesses all of PowerCenter's sophisticated parallelization and grid capabilities
- Empowers developers to extend profiling capabilities by using PowerCenter mapplets to support more complex scenarios

Data Profiling Option Benefits

Reduce Initial and Ongoing Data Quality Assessment Time

With this option, developers can streamline data profiling efforts by using a wizard-driven, automated interface for conducting the initial profiling assessment, including data content, structure, and quality. Users can easily rerun the resulting data profile mappings to continually measure quality improvement and analyze changes to source data over time. IT departments can be more productive because they use the same user interface for the entire process.

Lower Development Costs by Automating Data Discovery

PowerCenter is unique because powerful data profiling capabilities are embedded within the same environment used for developing data integration solutions, reducing developer ramp-up time and development costs. By automating data discovery, IT teams can avoid the labor-intensive, manual work of creating one-off queries or scripts to assess source data structure or quality.

Accelerate Time to Delivery by Reducing Risk and Rework in Data Integration Projects

Organizations can automatically profile any data accessible to PowerCenter, eliminating upfront assumptions and guesswork by building detailed statistics on the state of the actual data. Data integration projects can be delivered to specification and on time because developers can scope out potential data quality issues faster and with much greater accuracy than with the traditional method of hand-coded queries.

Learn More

Let Informatica help you determine how the PowerCenter Data Profiling Option can maximize the business value of your data.

Visit us at www.informatica.com/powercenter or call 1.800.653.3871.

About Informatica

Informatica is a leading provider of enterprise data integration software and services. With Informatica, organizations can gain greater business value by integrating all their information assets from across the enterprise. Thousands of companies worldwide rely on Informatica to reduce the cost and expedite the time to address data integration needs of any complexity and scale.

INFORMATICA
The Data Integration Company™

Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 www.informatica.com

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Ireland · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

© 2008 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and PowerCenter are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

6684 (08/08/2008)