

Ataccama DQ Analyzer



Ataccama DQ Analyzer is a unique and truly complete tool for data quality analysis. DQ Analyzer can rapidly reveal basic quality attributes of your data and easily proceed to advanced features for deeper and more accurate analysis. As its core, it uses Ataccama Data Quality Center, an advanced and proven tool for Data Quality Management which is used by a number of multi-national organizations for solving data quality problems.

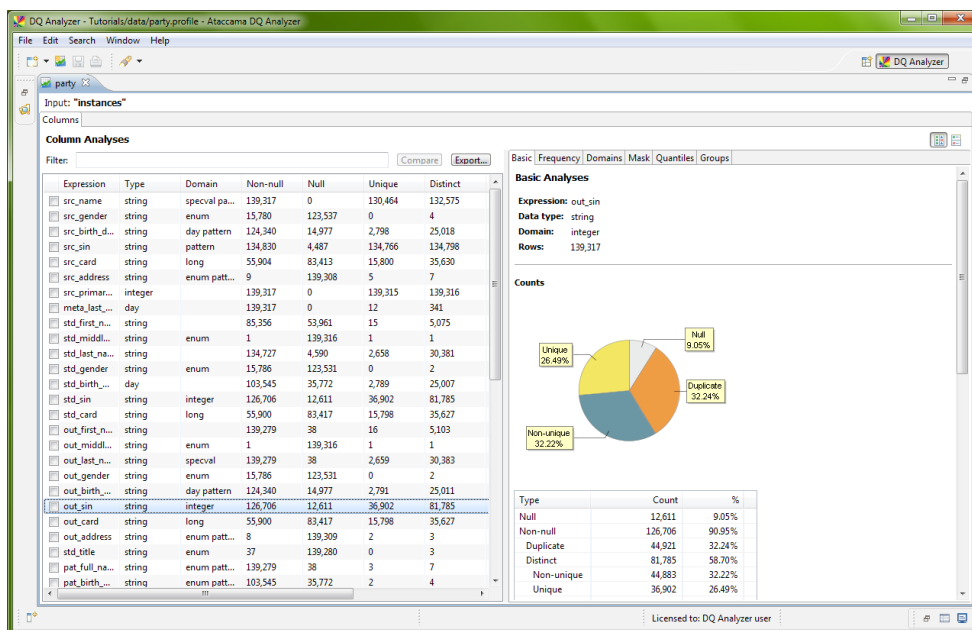
Problems with Data

Problems with data quality are pervasive and cause organizations numerous problems and incur many costs. They can result in correspondence being returned due to poor quality addresses, duplicate records in databases due to incorrect registration of names of clients or companies or the inability to contact a client due to errors in telephone numbers. These problems are common; however, organizations are frequently not able to accurately describe and quantify these errors.

financial resources and promote their implementation. In addition, data quality projects are very costly (in particular, the purchase of data cleansing tools), without the ability to foresee the benefits of such a project and the tools required for cleaning the data. In all these cases it is possible to perform a Data Quality Assessment, which enables very precise and rapid detection and quantification of errors in the client data and at the same time specifies the benefits of data quality improvement technology like Ataccama Data Quality Center (DQC) in the analyzed data.

Business Benefits

- Provides an objective picture of data quality - instantly.
- Gives an estimate of the benefits of data cleansing projects.
- Facilitates and accelerates data quality initiatives.
- Prevents making risky decisions based on inaccurate data.
- Completely free for commercial use.



Technical Advantages

- Basic analysis on data in a database or CSV file performed in a few minutes.
- DQ Analyzer is built on the Eclipse platform and its optimized user interface is designed to be intuitive and navigable.
- Simple problem solving. Plans created in DQ Analyzer can be used directly in Data Quality Center and Master Data Center without any modification.
- Large volumes of data. DQ Analyzer can handle tables with hundreds of millions of records and more.

Consequently, it's not possible to specify not only the effects of poor data, but also the cause, making it impossible to take effective measures to improve the quality of the data in corporate information systems. Also, the data quality projects which should solve the problem with data in organizations encounter difficulties while quantifying the impact of insufficient data quality and interpreting the financial effects. For these reasons, it is often difficult to get these projects the required

This makes it possible to quantify the financial impact of insufficient data quality and make a strong case for initiating a data quality project in the organization. At the same time it facilitates the subsequent project to clean the data since the data analysis uses the same procedures as for data cleansing. This accelerates the speed of the project and reduces its cost.

Easy Upgrade to DQC

With DQ Analyzer, you can easily upgrade to Ataccama Data Quality Center which has all the functionality you need to cleanse, standardize, de-duplicate, code, fix and otherwise enrich your data - and since it uses the same platform, language, and advanced user interfaces, it will feel like home to you. Also, if you are into SOA, note that every Data Quality Center process can be published as a service.

DQ Analyzer Algorithms



Text File Reader

Reads data from a delimited text file, such as CSV.



Condition

Splits the data stream based on a condition.



Column Assigner

A powerful and robust business rules engine that allows the use of complex expressions.



Fixed Width File Reader

Reads data from a text file with fixed width columns.



Union Same

Combines data streams that have the same format.



Trash

Ends the data flow.



JDBC Reader

Reads data directly from a database such as Oracle, MS SQL, Sybase, etc.



Regex Matching

Allows flexible parsing and matching using regular expressions.



Profiling

Flexible, fully configurable data profiler.

The screenshot shows the DQ Analyzer interface with a workflow diagram on the left and a data table on the right. The workflow includes steps like 'Text File Reader', 'Match emails in src_email', 'Extract valid emails', 'Match emails in src_address', 'Extract valid emails 2', 'Join the flows', 'Unparsed data Profiling', and 'Correct emails Profiling'. The data table shows columns: src_primary_key, src_name, src_email, and src_address.

src_primary_key	src_name	src_email	src_address
1	90	Tomasas Tingvald	SYSTEM MIGRATION
2	94	Beatrice Kornhauser	SYSTEM MIGRATION
3	97	Vogl Cristobal	SYSTEM MIGRATION
4	98	Debby Greenen	SYSTEM MIGRATION
5	101	Malett Jacqueline	SYSTEM MIGRATION
6	108	Starko Cand	SYSTEM MIGRATION
7	110	Waszkiewicz Charise	SYSTEM MIGRATION
8	111	Louwie Ardith	SYSTEM MIGRATION
9	112	Carline McPheron	SYSTEM MIGRATION
10	132	Toepke Erika	SYSTEM MIGRATION
11	133	Lina Turdo	SYSTEM MIGRATION
12	136	Pollom Lyman	SYSTEM MIGRATION
13	146	Wimbrow Mee	SYSTEM MIGRATION
14	147	Nick Teets	SYSTEM MIGRATION
15	155	Aleta Alicandro	SYSTEM MIGRATION
16	158	Yong Mokkerow	SYSTEM MIGRATION
17	161	Eiland Genaro	SYSTEM MIGRATION
18	168	Osbon Kendal	SYSTEM MIGRATION
19	172	Dibenedict Eileen	SYSTEM MIGRATION
20	176	Feduccia Milagros	SYSTEM MIGRATION
21	183	Rechart Carlotta	SYSTEM MIGRATION
22	191	Anaya He	SYSTEM MIGRATION
23	195	Corda Pansy	SYSTEM MIGRATION
24	203	Kinnbrew Niesha	SYSTEM MIGRATION
25	214	Wilda Moniak	SYSTEM MIGRATION
26	223	Knick Kathryn	SYSTEM MIGRATION
27	233	Una World	SYSTEM MIGRATION
28	242	Kira Beagley	SYSTEM MIGRATION
29	244	Joette Smolic	SYSTEM MIGRATION
30	257	Trevor Purdue	SYSTEM MIGRATION
31	273	Fyffe Lulu	SYSTEM MIGRATION
32	276	Malchom Madine	SYSTEM MIGRATION

Supported Analyses

Statistical Analysis

An overview of summary values, such as extremes, distribution and frequency analysis.

Domain Analysis

A configurable analysis of data types.

Mask and Group Analysis

An overview of value formats, groups and dimensions.

Business Rules

An analysis of the results of user-defined business rules.

Foreign Key and Dependency Analyses

An inside look into complex connections in the data.

Drill Through

The option to display individual records that correspond to aggregated results.

The screenshot shows the DQ Analyzer interface with a 'Foreign Keys' analysis window. It displays a table of foreign keys and a Venn diagram showing the relationship between 'person' and 'address' tables. The Venn diagram shows 48438 (52.84%) matched records, 43227 (47.16%) not matched records, and 71808 (64.43%) matched records between the two tables. Below the Venn diagram, there are two bar charts showing the distribution of records for 'Parent' and 'Child' tables.

Name	Left	Right
person to ...	person	address

Foreign Keys Analysis Results:

- Matched: 48438 (52.84%)
- Not matched: 43227 (47.16%)
- Matched: 71808 (64.43%)
- Not matched: 2420 (2.17%)
- Parent: 91665 records
- Child: 111456 records
- Matched: 48438 (52.84%)
- Not matched: 43227 (47.16%)
- Matched: 71808 (64.43%)
- Not matched: 2420 (2.17%)
- Null: 37228 (33.40%)

Ataccama Corporation is an international software company that prides itself in delivering cutting edge technology used for Data Quality and Master Data Management. Leading financial, commercial and government organizations choose Ataccama technologies for its proven ability to execute and deliver real business benefits.



ataccama

Ataccama Corporation
Le Parc Office Tower
8500 Leslie Street, Suite 200
Markham, ON L3T 7M8, Canada
Tel: + 1 (905) 881-7946
www.ataccama.com