

Turbo Charging Data Governance Platform with Data Trust Score

Key take away

Trust and Data Quality are keys to making the most efficient use of data and data governance platforms. It is vital to measure and communicate the quality of data to ensure that stakeholders are making decisions based on good information. DataBuck enables Alation users to evaluate data quality with a trust score for data assets as part of the Alation Data Catalog.

Introduction

DataBuck, with its out-of-the-box Machine Learning (ML) capabilities, calculates a Data Trust Score for each data asset - a score that all stakeholders can universally understand. More importantly, DataBuck will map and update the data trust score to the Alation data catalog without human intervention or complex integration efforts.

The screenshot shows the Alation Data Catalog interface. At the top, there is the Alation logo and a search bar. Below the search bar, the 'Description' section is visible, followed by the 'FE Quality Score' section. The 'FE Quality Score' section displays the following information:

- Databuck validation result page**
- Validation Check: **2000_T1_ALT_Validation**
- Last Execution Date: **2021-11-20**
- Data Trust Score: **66.35 (Same as previous run)**

Below this information is a table with the following columns: DQ Check Name, Column Names, DQI, KeyMetric, and Measurement.

DQ Check Name	Column Names	DQI	KeyMetric	Measurement
Record Count Reasonability		100.00	Record Count	62585
Distribution Check	TransactionAmount, CommissionAmount	100.00	Columns Failed	0
Record Anomaly	TransactionAmount, CommissionAmount	99.99	Records Failed	11
Data Completeness	TranID, AgentID, Country, Instrument, SettlementCode, TransactionAmount, CommissionAmount	99.99	Nulls	34
String Value Drift (Drift and Orphan)	Country, Instrument, SettlementCode	100.00	Unique Values Changed	0
Length Check (Conformity)	SettlementCode	99.99	Records Failed	4

With DataBuck, now you can start the transition from a traditional manual model to a trust-based data-driven approach to data quality.

How It works

- **Scan:** DataBuck scans each data asset registered in your Data Governance Platform.

- **Auto Discover Metrics**- DataBuck autonomously creates data health metrics specific for each of the data assets. The well-accepted and standardized DQ tests are customized for each data set individually, leveraging AI/ML algorithms.
- **Monitor**- Health metrics are then translated to a data trust score. Health metrics are computed based on quality dimensions for each column in the data asset.
- **Alert**: DataBuck continuously monitors the health check metrics and trust score and alerts users when the trust score becomes unacceptable.

The deviation of the trust score displayed in the summary of analysis results shows how the quality score changed between the last two analyses. Every violation discovered can be double-clicked for further information:

- At the data asset level, expand the dimension to see which columns are affected. Click a column name to see the dimension details for that column.
- At the column level, click the dimension name for further details.

Users can then decide for each Data Quality violation to be either ignored or

evaluated during the analysis. Users can make this decision for the entire data asset or for individual columns.

The screenshot shows the Alation interface for a data asset named 'Commission Amount'. The breadcrumb path is 'FE / fe.dbo . t1 . commissionamount'. The asset name is 'Commission Amount' with a sub-label 'commissionamount' and three colored circles (red, orange, green). Below this, there are tabs for 'Overview', 'Queries 0', and 'Expressions 0'. The 'Description' section shows 'No description'. The 'FE Quality Score' section indicates a 'Validation Check: 2000_T1_ALT_Validation' and a 'Last Execution Date: 2021-11-20'. A table below lists the DQ checks and their results:

DQ Check Name	Records Failed	Fail Percentage
Null Check	0	0.00
Record Anomaly Check	7	0.01
Numerical Statistics Check	0	0.00
Custom Rule	8527	13.62

Summary

Trust and Data Quality are, but should not be, a popularity contest in Data Catalogs. Users vote their preference between different datasets. These subjective judgments on 1,000's of datasets were tolerated not because they are accurate but because they are slightly better than having no inkling of the trustworthiness of a dataset. DataBuck has taken the guesswork out and establishes trust with objective Data Quality criteria. With a few clicks, DataBuck populates the Data Catalog with a trust score and increases the usefulness of the catalog and the data tremendously.