QC Use Cases

Created by Mike Timchenko
Last updated: Dec 10, 2021 • 3 min read • 12 people viewed

Conformance Tests verify that new data conforms to a set of rules defined for key tables. Rules may be applied to raw or processed tables. They may test tables in isolation, tables in comparison to other, related tables or the last-generated version of the same table. Rules can be defined by subject-matter experts or data engineers. Tests are run automatically on schedules, and results are reported on the wiki. Exceptions and failures may also be shared via email or Slack alerts.

Test Type	Description
Allowed Increment	Compares number of rows in table vs last test run. Error if row count is less than expected increment.
Uniqueness	Tests for absence of duplicate values from a column (or combination of columns) in the table. Null values are ignored. Error if duplicate values are found.
Primary Key	Tests for absence of duplicate values from a column (or combination of columns) in the table. Null values are not permitted. Error if duplicate or null values are found.
Prior Match	Tests for the presence of the same distinct column values (or combination of values) in a new build as are present in the same table of a prior build schema (or different schema). Error if prior values are not all included in latest data, or optionally if distinct values change.
Window Match	Test for the presence of the same column values (or combination of values) in the most recent time-window of a transactional table vs the prior time window of the same table in the same schema. Error if distinct prior values are not the same in both windows.
Value Match	Tests for presence (or absence) of column values (or combinations of values) vs static defined values or combinations of values defined directly in parameters. Error if values don't match static set.
Data Match	Tests for presence of values (or combination sof values) from current table vs another table. Null values are included in check. Error if values aren't found in second table.
Aggregate Match	Tests for matches in aggregated values distributed per one column value (or a combination of columns), for example the sum of sales_dollars per product per day vs the same values in another table. Error if aggregated values per columns don't match in both tables.
Not Null	Tests for absence of Null values in a column or combination of columns. (This test is also used by the Unique Key test.) Error if null values are found.

Condition Check	Tests for logical condition, typically within a record, though condition can also include a subquery. Error if condition is false.
Custom Query	**WARNING: all other tests check correct data, custom SQL suppose to query wrong data.** Runs a custom SQL query as a test. This allows atypical testing and exception reporting to be connected in to the Conformance engine and reporting infrastructure. A query should include only one return column with a name "error_data". Error if a query returns one or more rows.

Implementing Conformance Tests

1. Related tests are defined as part of a single test group;
2. Initially, all test parameters may be collected in an excel spreadsheet, which functions as a template;
3. Then test parameters should be copied into QC tables **<type_of_test>_test_set** in the defined QC schema;
4. Finally, a data engineer or an advanced user should include a QC ingredient into the build variation or create a new variation that triggers the Conformance QC Engine to perform the tests.