Talend Data Quality Framework - Data Dictionary

Introduction

This document provides a description of all tables and views used in the Data Quality Framework (DQF).

Note: DQF is built on the Talend Data Quality Mart, which uses its own set of tables and views documented at help.talend.com and is not part of this document.

All DQF Tables are prefixed with DQF_ or MV_ and all Talend Data Quality Mart tables are prefixed with TDQ_.

All DQF views are prefixed with BI_ and Talend Data Quality Mart views are prefixed with TDQ_.

Contents

Tables

DQF tables

Ensure you deploy all DQF tables in the same database/schema as the Talend Data Quality Mart, with the same access permissions. The DQF tables enhance the data quality mart by providing additional capabilities beyond data profiling.

DQF_AUTO_ANALYSIS

Purpose	Reserved for future use			
Column Name	**PK**	**Description**	**Relationships**	**Example**
AUTO_ANALYSIS_PK				
SEMANTIC_TYPE_FK				
SEMANTIC_TYPE_NAME				
CLOUD_BASE_TYPE				
STUDIO_TYPE				
ANALYSIS_TYPE				
INDICATOR_CATEGORY				
INDICATOR				
IS_DEFAULT_RULE				
MIN_THRESH				
MAX_THRESH				
THRESHOLDS_ARE_PERCENTAGE				
DQ_DIMENSION_DEFAULT_MAPPING				

DQF_DATASET

Purpose	This table holds reference data about a dataset that will be monitored for data quality.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DATASET_PK		The primary key for this table		1
SOURCE_SYSTEM		The source system / application where the dataset was sourced from		Salesforce
SOURCE_OBJECT_NAME		The physical name of the dataset in the source system		Account
SOURCE_PK_DEFINITION		The primary key column for the source dataset		ID
BUSINESS_SYSTEM		The business name for the source system		CRM
BUSINESS_NAME		The business name for the dataset		Customer Account
OWNER_FK		Data owner for the dataset	FK to DQF_PERSON	1
STEWARD_FK		Data steward for the dataset	FK to DQF_PERSON	2
MDF_OBJECT_ID_FK		Reserved for future use		

DQF_DATASET_ATTRIBUTE

Purpose	This view holds additional custom attributes for a dataset using a key-value pair schema. It is not currently leveraged in the demo dashboard.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
ATTRIBUTE_PK		Primary key for this table		1
DATASET_FK		ID of the dataset	FK to DQF_DATASET	1
NAME		Name of the attribute		Department
VALUE		Value for the attribute		Sales

DQF_DATASET_RUN

Purpose	This table holds the history of every ruleset execution (or RUN) for each dataset. That is, for a given DQF_RUN, this table will contain 1..n datasets.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_DATASET_PK		Primary key for this table		1
RUN_FK		ID for the rule execution	FK to DQF_RUN	1
DATASET_FK		ID of the dataset	FK to DQF_DATASET	1
TDQ_REP_RUNTIME_SOURCE_FK		Rule execution date/time	FK to TDQ mart tables (not enforced via constraints)	2023-01-04 19:33:49.000
TDQ_REP_RUNTIME_STD_FK		Reserved for future use		
TDQ_REP_LABEL_FK		Name of the report from Talend Studio	FK to TDQ mart tables (not enforced via constraints)	DQF_DEMO_DATASET1_RPT
TOTAL_ROW_COUNT		Number of rows in scope from the dataset at the time of rule execution		999

DQF_DQ_RULE

Purpose	This table holds the metadata that defines each data quality rule for each dataset.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DQ_RULE_PK		Primary key for this table		1
DATASET_FK		ID of the dataset	FK to DQF_DATASET	2
DESCRIPTION		A business user friendly description of the data quality rule		
TDQ_REP_LABEL_FK		Report name from Talend Studio / TDQ mart. Needs to be exact to allow joins between the DQF metadata and the TDQ mart.	FK to TDQ mart tables (not enforced via constraints)	
TDQ_AN_LABEL_FK		Analysis name from Talend Studio / TDQ mart. Needs to be exact to allow joins between the DQF metadata and the TDQ mart.	FK to TDQ mart tables (not enforced via constraints)	
TDQ_ELT_TABLE_NAME_FK		Source table name from Talend Studio (as defined in the metadata repository) / TDQ mart. Needs to be exact to allow joins between the DQF metadata and the TDQ mart.	FK to TDQ mart tables (not enforced via constraints)	
TDQ_ELT_COLUMN_NAME_FK		Source column from Talend Studio (as defined in the metadata repository) / TDQ mart. Needs to be exact to allow joins between the DQF metadata and the TDQ mart.	FK to TDQ mart tables (not enforced via constraints)	
TDQ_IND_LABEL_FK		Indicator name from Talend Studio (as defined in the metadata repository) / TDQ mart. Needs to be exact to allow joins between the DQF metadata and the TDQ mart. Note: When using a “business rule” analysis, this will be the name of the business rule from Studio / TDQ mart.	FK to TDQ mart tables (not enforced via constraints)	
INDICATOR_VALUE_IS_PASS_COUNT		Determines if the Indicator value is a Pass or fail (*)		1
OWNER_FK		Data owner for the rule	FK to DQF_PERSON	1

Purpose	This table holds the metadata that defines each data quality rule for each dataset.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
STEWARD_FK		Data Steward for the rule	FK to DQF_PERSON	2
DQ_DIMENSION		DQ dimension. Choose from Uniqueness, Accuracy, Timeliness, Validity, Consistency, or Completeness.		Accuracy
RULE_PRIORITY		Rule priority – currently used in notification rules, for example, receive a notification for all failures with a priority four or higher		4
BUSINESS_COLUMN_NAME		Business name for the column		LASTNAME
IS_ACTIVE		Is the rule still active? Setting to zero will exclude this rule from execution when a DQF Run includes the associated dataset.		1
RULE_SCOPE		Determines if rule is executed on a single row (ROW) or entire dataset (DATASET)		ROW

(*) INDICATOR_VALUE_IS_PASS_COUNT is either zero (false), one (true), or minus one (special case for dataset scope rules). When Talend Studio executes an indicator, it returns the measurement value for that indicator. However, the interpretation of that value – whether the value indicates a success or failure - is dependent on how the user interprets it and what the indicator is measuring.

For example, an indicator could identify that one or more NULL values exist in a particular column, and it may be desirable to define a rule that specifies NULL values are not allowed for that column. The count stored in the TDQ mart will be the total number of nulls, which can be interpreted as the total number of failed checks. In this case, the INDICATOR_VALUE_IS_PASS_COUNT is set equal to one, which may appear to be counter intuitive, but there are reasons for this (Talend may choose to rename the column in a future release for clarity).

As a second example, the regular expression indicator records a count of rows that conform to the given regex in the TDQ mart, which means you need to do a calculation to get the number of failed rows by subtracting the pass count from the total in scope row count. In this instance the INDICATOR_VALUE_IS_PASS_COUNT should be set to zero.

There is a special case when defining a rule that has DATASET scope rather than ROW scope in the RULE_SCOPE field. An example of this would be a timeliness rule where there must be at least one record in the dataset that has been updated in the last week. This is calculated using a basic column analysis, a row count indicator, and a SQL filter for the analysis that contains the condition. If the analysis/row count indicator returns zero rows, this indicates that the dataset is failing the condition. To denote this special case, set the value minus one in the INDICATOR_VALUE_IS_PASS_COUNT field and DATASET in the field RULE_SCOPE. For more examples, see the **Talend Data Quality Framework - Dataset Onboarding Guide**.

DQF_DQ_RULE_ATTRIBUTE

Purpose	This table holds additional custom attributes for a DQ rule using a key-value pair schema. These attributes are used on the dataset summary tab of the demo dashboard to allow the user to slice and dice by attributes relevant to their business.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
ATTRIBUTE_PK		Primary key for this column		
DQ_RULE_FK		ID of the DQ rule	FK to DQF_DQ_RULE	5
NAME		Name of the attribute		Data Domain
VALUE		Value of the attribute		Customer

DQF_FAILED_QUERY_TEMPLATE

Purpose	This table holds SQL templates used for DQ rule execution. At runtime this template is dynamically parsed and then executed on the database as SQL pushdown to identify rows that do not meet a data quality rule.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
TEMPLATE_PK		Primary key for this table		
TEMPLATE		The actual SQL template		SELECT [PK_FIELD] as ROW_KEY, [CHECK_FIELD] as ROW_FAILED_VALUE FROM [TABLE] WHERE [CHECK_FIELD] IS NULL AND ([FILTER])
DESCRIPTION		Description of the template		SQL template for the Null Count indicator
TDQ_AN_TYPE_FK		Analysis type from Talend Studio: MULTIPLE_COLUMN | COLUMNS_COMPARISON | TABLE – is extensible, just needs to match value in the TDQ mart	FK to TDQ mart tables (not enforced via constraints)	MULTIPLE_COLUMN
TDQ_IND_LABEL_FK		Indicator name from Talend Studio – must be exact match to TDQ mart value	FK to TDQ mart tables (not enforced via constraints)	Null Count
DATA_TYPE		Reserved for future use		
DATABASE		Type of database this template supports – like Oracle, SQL server, and others – if “Default”, then any ANSI SQL database		MySQL
DATABASE_VERSION		Version of database this template supports, not mandatory		8

DQF_FAILED_QUERY_VARIABLE

Purpose	This table holds the values that are inserted into the SQL templates found in DQF_FAILED_QUERY_TEMPLATE. Selected values are also used to keep a history of the rows that were in scope for a given combination of rule, dataset, and DQF run.*			
Column Name	**PK**	**Description**	**Relationships**	**Example**
VARIABLE_PK		Primary key for the table		45
TEMPLATE_FK		ID of the template	FK to DQF_FAILED_QUERY_TEMPLATE	5
DQ_RULE_FK		ID of the Rule	FK to DQF_DQ_RULE	3
KEY		Template placeholder Key, must be surrounded by square backets. [PK_FIELD] and [FILTER] (or [FILTER_A]) are mandatory for every DQ Rule, even if those variables are not defined in the associated template.*		[PK_FIELD]
VALUE		Variable to replace with		COMPANYNUMBER

*For more details, see the **Talend Data Quality Framework - Dataset Onboarding Guide**.

DQF_FAILED_ROW

Purpose	This table holds a record of all failed rows that did not pass a data quality rule (**Note**: More than one rule could have been executed on a single row). Also, this table incorporates slowly changing dimensions; therefore, if the same break occurs on multiple runs, a new row will not be inserted.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
FAILED_ROW_PK		Primary key for this table		67
SCD_ID		Slowly Changing Dimension ID		6|13441837|Source
SCD_VERSION		SCD version		1

Purpose	This table holds a record of all failed rows that did not pass a data quality rule (**Note**: More than one rule could have been executed on a single row). Also, this table incorporates slowly changing dimensions; therefore, if the same break occurs on multiple runs, a new row will not be inserted.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
SCD_ACTIVE		Is the row still active? That is, the failure was detected in the last DQF Run that included the associated dataset.		1
SCD_START		When the break was first detected		2023-01-04 19:38:45.253
SCD_END		When the break stopped occurring, null if still ongoing		2023-01-05 19:35:47.222
DQF_RUN_ID_START_FK		Run ID when break was first detected		23
DQF_RUN_ID_END_FK		Run ID when break stopped occurring, null if still ongoing		24
DQ_RULE_FK		ID of the Rule	FK to DQF_DQ_RULE	7
DATASET_FK		ID of the dataset	FK to DQF_DATASET	2
SOURCE_ROW_KEY		Primary key for the dataset		SC691519
ROW_FAILED_VALUE		Value of the failed row/column		IP13329R
FAILURE_REASON		Reason for failure		Company Number must conform to a pattern
CONTEXT		Context name from Talend studio – currently hardcoded to “Source”. Will be used in future release.		
STEWARD_LINK		URL to failed row (if available)		

DQF_FAILED_ROW_TMP

Purpose	This table updates the Slowly Changing Dimensions table DQF_FAILED_ROW and is part of the materialized view refresh for MySQL. It is truncated on each DQF run.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
ROW_ID		Primary key for this table (autoincrement)		5
DQ_RULE_FK		ID of the Rule	FK to DQF_DQ_RULE	7
DATASET_FK		ID of the dataset	FK to DQF_DATASET	2
SOURCE_ROW_KEY		Primary key for the dataset		SC691519
ROW_FAILED_VALUE		Value of the failed row/column		IP13329R
FAILURE_REASON		Reason for failure		Company Number must conform to a pattern
CONTEXT		Context name from Talend studio – currently hardcoded to “Source”. Will be used in future release.		
DQF_RUN_ID_START_FK		The current DQF run		23
SCD_ID		Slowly Changing Dimension ID		6|13441837|Source
STEWARD_LINK		Not used		

DQF_NOT_CHANNEL

Purpose	All notifications are sent using a channel; this table holds all available channels.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
CHANNEL_PK		Primary key for this table		1
LABEL		Name of channel (email or Slack included in this release, but extensible)		Email
ADMIN_EMAIL		Email address of channel admin		myemail@company.com
IS_ACTIVE		Is channel active		1

DQF_NOT_CHANNEL_PERSON

Purpose	List of contact identifiers for people by channel			
Column Name	**PK**	**Description**	**Relationships**	**Example**
CHANNEL_PERSON_PK		Primary key for the table		5
PERSON_FK		ID of the person	FK to DQF_PERSON	2
CHANNEL_FK		ID of the channel	FK to DQF_NOT_CHANNEL	1
CHANNEL_IDENTIFIER		Email id of the person or Slack channel name (extensible to other systems)		myemail@company.com

DQF_NOT_DEFINITION

Purpose	Defines a notification			
Column Name	**PK**	**Description**	**Relationships**	**Example**
NOT_DEF_PK		Primary key for this table		4
DQ_RULE_FK		ID for the DQ Rule	FK to DQF_DQ_RULE	8
DATASET_FK		ID for the dataset	FK to DQF_DATASET	2
SERVICE_METRIC_FK		Reserved for future use	FK to DQF_SERVICE_METRIC	
NOTIFICATION_TYPE		When to send a notification (*)		
NOTIFICATION_SCOPE		Scope for the notification (**)		
NOTIFICATION_PER_ROW		Should a notification be sent for each failing row?		
THRESHOLD		Reserved for future use		
PRIORITY_LEVEL		Used only when NOTIFICATION_TYPE is RULE_PRIORITY. Defines the minimum rule priority level for sending a notification. For example, if this is set to 4, a notification is sent only if there are one or more failures of a priority 4 or priority 5 rule.		4
DQ_DIMENSION		DQ Dimension – for reference but not used in logic		Timeliness
DESCRIPTION		Description of the notification - for reference but not used in logic		Last file update notification

(*) NOTIFICATION_TYPE: Possible values are:

- FAILED_ROWS
- RULE_PRIORITY

(**) NOTIFICATION_SCOPE: Possible values are:

- RULE
- DATASET
- GLOBAL

When scope is DATASET, RULE_PRIORITY and DATASET_FK must be set. When scope is GLOBAL, RULE_PRIORITY must be set.

DQF_NOT_GROUP

Purpose	Reference data for Notification Groups			
Column Name	**PK**	**Description**	**Relationships**	**Example**
NOT_GROUP_PK		Primary key for the table		1
LABEL		Group name		SalesOps, group with no members
GROUP_CHANNEL_IDENTIFIER		Pipe delimited list of identifiers for the channel when the notification is to be sent just to the group identifier (such as an email distribution list or slack channel) vs sending to the individuals in a group		1:sales@company.com|2:C021JR7DS564J
IS_ACTIVE		Is the group active?		

DQF_NOT_GROUP_MEMBER

Purpose	Table that holds the list of members for a Notification Group			
Column Name	**PK**	**Description**	**Relationships**	**Example**
MEMBER_PK		Primary key for this table		1
NOT_GROUP_FK		ID of Notification Group	FK to DQF_NOT_GROUP	2
PERSON_FK		ID of the Person	FK to DQF_PERSON	2
IS_ACTIVE		Is this member active?		1

DQF_NOT_SUBSCRIPTION

Purpose	Define subscribers for notifications – that is, which people or groups will receive notifications and how			
Column Name	**PK**	**Description**	**Relationships**	**Example**
NOT_SUB_PK		Primary key for the table		1
NOT_DEF_FK		ID of the Notification	FK to DQF_NOT_DEFINITION	2
NOT_CHANNEL_FK		ID of the Channel	FK to DQF_NOT_CHANNEL	1
PERSON_FK		ID of the Person – set person OR group	FK to DQF_PERSON	2
GROUP_FK		ID of the Notification Group – set person OR group	FK to DQF_NOT_GROUP	4
THRESHOLD_OVERRIDE		Reserved for future use		
SEND_PER_RECURRENCE		Send notification on every DQF run, not just the first DQF run, in which the event occurred		1
IS_ACTIVE		Is subscription active?		1
INCLUDE_IN_AGGREGATE_MSG		Include notification in an aggregate message. An aggregate message allows subscribers to receive a single message when they are subscribed to multiple notifications.		1

DQF_NOTIFICATION

Purpose	List of all notifications generated to date			
Column Name	**PK**	**Description**	**Relationships**	**Example**
NOTIFICATION_PK		Primary key for this table		4
RUN_FK		ID of the Rule Execution	FK to DQF_RUN	3
SENT		Has the notification been sent to the channel?		1
CREATED		Date/time when notification created		2023-01-30 16:54:03.823
FAILED_ROW_FK		ID of the failed row (if applicable)	FK to DQF_FAILED_ROW	17624
NOT_SUB_FK		ID of the Subscription	FK to DQF_NOT_SUBSCRIPTION	3
DESCRIPTION		Notification description		The most recent ACCOUNTS_LASTMADEUPDATE in the dataset must be within the last two months
COUNT_FAILED_ROWS		Used when NOTIFICATION_PER_ROW is set to false on the table DQF_NOT_DEFINITION – holds the total number of failures reported by the notification		56

DQF_PERSON

Purpose	Reference data for named individual users of the DQF framework			
Column Name	**PK**	**Description**	**Relationships**	**Example**
PERSON_PK		Primary key for this table		1
FIRST		First name		
LAST		Last name		
EMAIL		Email address		
TITLE		Title		
IS_ACTIVE		Is person active		1

DQF_ROW_RULE_HISTORY

Purpose	This table holds the record of every row that was in scope (determined by the analysis filter), for every rule, and for each execution that includes the dataset (whether that execution resulted in a failure or not).			
Column Name	**PK**	**Description**	**Relationships**	**Example**
ROW_RULE_HISTORY_PK		Primary key for this table		4
SCD_ID		Slowly changing dimension ID		1|13441837|Source
SCD_VERSION		SCD version		1
SCD_ACTIVE		SCD active		1
SCD_START		SCD start		2023-01-04 19:40:46.124
SCD_END		SCD end		2023-01-08 19:38:41.222
DQF_RUN_ID_START_FK		Run ID when this row was first checked	FK to DQF_RUN	33
DQF_RUN_ID_END_FK		Run ID when this row no longer checked	FK to DQF_RUN	37

Purpose	This table holds the record of every row that was in scope (determined by the analysis filter), for every rule, and for each execution that includes the dataset (whether that execution resulted in a failure or not).			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DQ_RULE_FK		ID of the DQ Rule	FK to DQF_DQ_RULE	15
DATASET_FK		ID of the dataset	FK to DQF_DATASET	2
SOURCE_ROW_KEY		Primary key of the dataset		00913705
CONTEXT		Context name from Talend Studio – hardcoded to "Source" in this release		Source

DQF_ ROW_RULE_HISTORY_TMP

Purpose	This table updates the Slowly Changing Dimensions table DQF_ROW_RULE_HISTORY. It is truncated on each DQF run.			
Column Name	**PK**	**Description**	**Relationships**	**Example**
TMP_ROW_ID		Primary key for this table (autoincrement)		5
DQ_RULE_FK		ID of the Rule	FK to DQF_DQ_RULE	7
DATASET_FK		ID of the dataset	FK to DQF_DATASET	2
SOURCE_ROW_KEY		Primary key for the dataset		SC691519
CONTEXT		Context name from Talend Studio – currently hardcoded to "Source". Will be used in future release.		Source
DQF_RUN_ID_START_FK		The current DQF run		23
SCD_ID		Slowly Changing Dimension ID		6|13441837|Source

DQF_RUN

Purpose	List of all executions (DQF runs)			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		Primary key for this table		1
RUN_START		Date/time when run started		2023-01-04 19:33:14.200
RUN_END		Date/time when run completed		2023-01-04 19:38:16.876
STATUS		Status of the run		COMPLETE

DQF_SCORE_BY_DATASET_HIST

Purpose	Holds the aggregated Dataset DQ scores over time			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DATASET_FK		Dataset ID	FK to DQF_DATASET	1
DATASET_NAME		Duplicated from DQF_DATASET for readability / efficiency		Customer
RUN_FK		Run ID	FK to DQF_RUN	1
RUN_TIME		Duplicated from DQF_DATASET_RUN for readability / efficiency		2023-05-10 17:26:27
TOTAL_RULES_APPLIED		Total number of rules applied to this dataset on this run		12345
TOTAL_RULES_FAILED		Total number of rules that failed on this dataset on this run		999
DATASET_SCORE		The calculated DQ score for this dataset for this run		88.99

DQF_SEMANTIC_TYPE

Purpose	Reserved for future use			
Column Name	**PK**	**Description**	**Relationships**	**Example**
SEMANTIC_TYPE_PK				
NAME				
TYPE				
LAST_SYNC				
DICT_TABLE_NAME				
REGEX_PATTERN				

DQF_SEMANTIC_TYPE_COMPOUND

Purpose	Reserved for future use			
Column Name	**PK**	**Description**	**Relationships**	**Example**
ST_COMPOUND_ROW_PK				
PARENT_SEMANTIC_TYPE_FK				
CHILD_SEMANTIC_TYPE_FK				

DQF_SERVICE_METRIC

Purpose	Reserved for future use			
Column Name	**PK**	**Description**	**Relationships**	**Example**
SERVICE_METRIC_PK				
LABEL				

Materialized Views (MySQL only)

MV_ALL_ROWS

Purpose	Holds a row for every DQ rule applied to every source row in the current run, including status			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		Current Run ID		1
DQ_RULE_FK		DQ Rule ID		2
DATASET_FK		Dataset ID		3
SOURCE_ROW_KEY		Source PK		6788990
ROW_FAILED_VALUE		Failure value (if any)		United Kingdo
FAILURE_REASON		Failure reason (if any)		Invalid country
STEWARD_LINK		URL for steward to click on (not used in demo Power BI dashboard in this release)		https://myapp/6788990
RULE_NAME		Rule plain english description		Country name must be a valid ISO description
OWNER_FK		Data owner ID		3
DQ_RULE_PK		DQ Rule ID		2
DQ_DIMENSION		DQ rule mapped DQ dimension		Validity
RULE_PRIORITY		Rule priority		3
BUSINESS_COLUMN_NAME		Business friendly column label		Country
IS_ACTIVE		Is rule active?		1
RULE_SCOPE		Rule scope: ROW or DATASET		ROW

MV_DATASET_RULES

Purpose	Dataset level rule information for current run			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DATASET_FK		Dataset ID		1
SOURCE_ROW_KEY		Source row PK		2643
DQ_RULE_FK		DQ Rule ID		12
FAILED		Did the rule fail?		0

MV_DATASET_RULE_COUNTS

Purpose	Dataset level rule information for current run			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DATASET_FK		Dataset ID		1
SOURCE_ROW_KEY		Source row PK		2643
DATASET_RULE_IN_SCOPE_COUNT		How many dataset level rules applied to this row?		1
DATASET_RULE_FAILED_COUNT		How many dataset level rules failed for this row?		0

MV_DQF_DATASET_SUMMARY

Purpose	Dataset summary information for the latest run of each dataset			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DATASET_FK		Dataset ID		1
TOTAL_RULES_APPLIED		How many rules applied to this dataset?		3702
TOTAL_RULES_FAILED		How many rules failed for this dataset?		168
TOTAL_RULES_PASSED		How many rules passed for this dataset?		3534
DATASET_SCORE		Dataset score		95.46
TOTAL_ROW_COUNT		Number of in scope source rows		1234
GOOD_ROWS		Rows with zero DQ issues		1066
BAD_ROWS		Rows with one or more DQ issues		168
COMPLETENESS_SCORE		Completeness score		93.19
VALIDITY_SCORE		Validity score		null
ACCURACY_SCORE		Accuracy score		null
UNIQUENESS_SCORE		Uniqueness Score		null
TIMELINESS_SCORE		Timeliness Score		100
CONSISTENCY_SCORE		Consistency Score		null

MV_DQF_ROW_SCORE

Purpose	Row level information for the latest run			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		Last run ID		1
DATASET_FK		Dataset ID		2
SOURCE_ROW_KEY		Source row PK		179907
NUM_RULES_APPLIED		How many rules applied to this row?		10
NUM_RULES_FAILED		How many rules failed for this row?		1
ROW_SCORE		Row score		90

MV_DQF_ROW_SCORE_BY_DIMENSION

Purpose	Row level information for the latest run by dimension			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		Last run ID		1
DATASET_FK		Dataset ID		2
SOURCE_ROW_KEY		Source row PK		179907
DQ_DIMENSION		DQ Dimension		Accuracy
NUM_RULES_APPLIED		How many rules applied to this row?		3
NUM_RULES_FAILED		How many rules failed for this row?		1
DIMENSION_SCORE		Row dimension score		66.67

Views

Calculation of Data Quality scores

The views support two types of data quality score calculations:

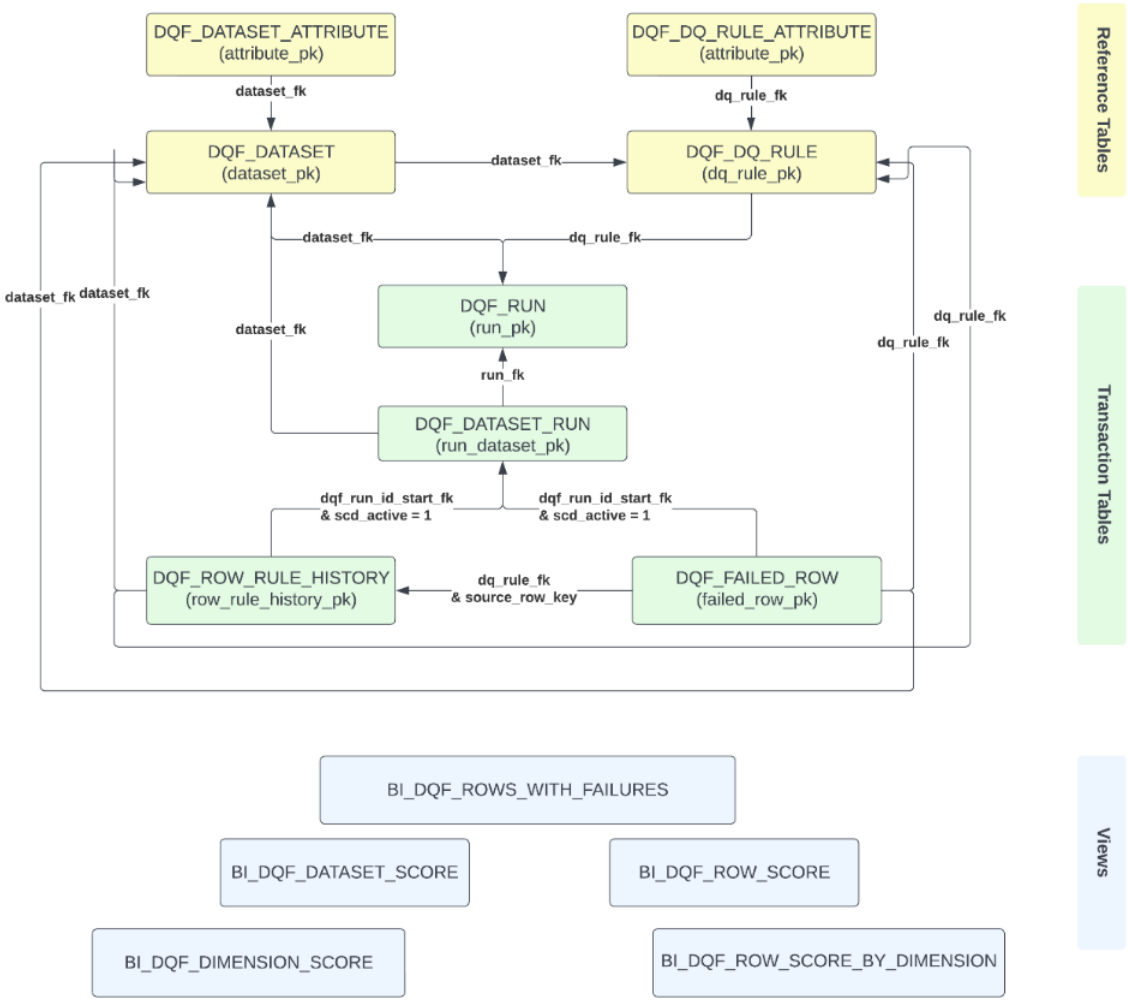
- **By DQ Rule** – (Deprecated) In this methodology, all pass/failures are aggregated by Rule. For example, the framework calculates how many rows failed a particular rule, multiply it by the rule priority, and do a weighted average across all the rules.
- **By Row & Dataset** – In this methodology, total pass/failures are aggregated by row and by dataset. For example, the framework might execute three rules on three different columns of a row. The total pass/fail score will be calculated for the entire row, rather than the individual column. These pass/fail rule counts are further used to derive the score for the entire dataset. This is the approach used within the latest dashboards.

$$\text{Data Quality Score} = \frac{\text{Total Rules Passed}}{\text{Total Rules Executed}} \times 100$$

Materialized views in MySQL

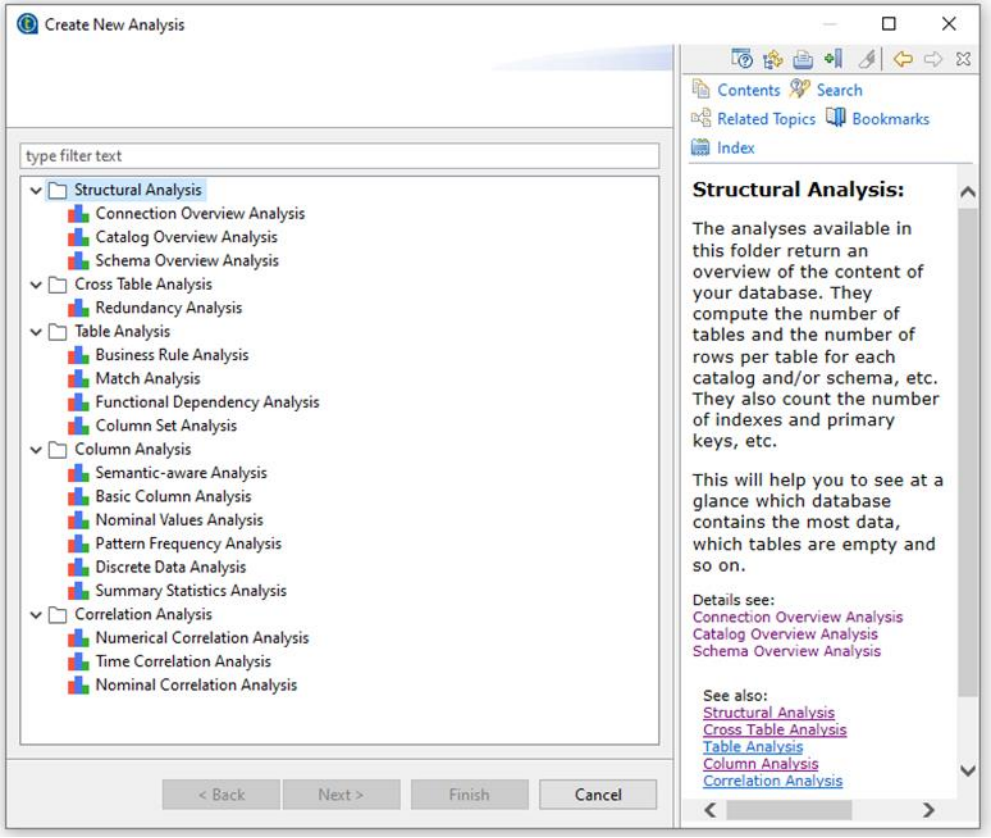
DQF currently supports both Snowflake and MySQL. However, MySQL is unable to deliver query performance similar to Snowflake, therefore causing the need to store pre-aggregated metrics to ensure that the dashboards are performant. Unfortunately, MySQL does not support materialized views, so Talend has created a set of MV_* tables that the Talend Jobs populate at the end of every run.

View hierarchy



Types of Data Profiles

Talend Studio supports 17 types of data profiles, called analyses.



The results of these analyses can be exported to the Data Quality Mart, except for the correlation analyses (numerical, time, and nominal).
Each analysis can be categorized into one of five areas:

- Column analysis — generally profiles single columns of data
- Overview analysis — analyzes data models and structures
- Column set analysis — performs an analysis across a set of columns (full records)
- Match analysis — compares a set of columns and creates groups of similar records
- Comparison analysis — compares two datasets that have the same columns, similar to a reconciliation

The following table maps the analysis type to the database view.

Types of data analysis stored in the Data Quality Mart		**Column analysis** bi_column_analysis	**Overview analysis** bi_overview_analysis	**Column set analysis**	**Match analysis**	**Comparison analysis** bi_comparison_analysis
Structural analysis						
1	Connection overview		X			
2	Catalog overview		X			
3	Schema overview		X			
Cross table analysis						
4	Redundancy					X
Table analysis						
5	Business rule	X				
6	Match				X	
7	Functional dependency					X
8	Column set			X		
Column analysis						
9	Semantic-aware	X				

Types of data analysis stored in the Data Quality Mart		**Column analysis** bi_column_analysis	**Overview analysis** bi_overview_analysis	**Column set analysis**	**Match analysis**	**Comparison analysis** bi_comparison_analysis
10	Basic column	X				
11	Nominal values	X				
12	Pattern frequency	X				
13	Discrete data	X				
14	Summary statistics	X				

BI_COLUMN_ANALYSIS

Purpose	Talend Studio supports many types of indicators that operate on a Column. The results of these types of analyses will be available through this view.			
Related views	BI_COMPARISON_ANALYSIS & BI_OVERVIEW_ANALYSIS			
Column Name	**PK**	**Description**	**Relationships**	**Example**
REPORT_NAME		The report name from the Studio profiler		
RUN_TIME		Execution of the profile		
AN_PK		ID of the Analysis	FK to TDQ_ANALYSIS	
ANALYSIS_NAME		Name of the Analysis		
CONNECTION_NAME		Connection name from Talend Studio		

Purpose	Talend Studio supports many types of indicators that operate on a Column. The results of these types of analyses will be available through this view.			
Related views	BI_COMPARISON_ANALYSIS & BI_OVERVIEW_ANALYSIS			
Column Name	**PK**	**Description**	**Relationships**	**Example**
CONTEXT_NAME		Context name from Talend Studio		
TABLE_NAME		Table name from the underlying data source		
COLUMN_NAME		Column name from the underlying data source		
INDICATOR_CATEGORY		Category of Indicator from Talend Studio		
INDICATOR_NAME		Name of Indicator from Talend Studio		
REGULAR_EXPRESSION				
ROWS_PROCESSED		No. of rows included in the Analysis		
INDICATOR_VALUE		Outcome of the execution of the Indicator		
INDICATOR_DATE_VALUE		Outcome of the execution of the Indicator if date data type		
VALUE_PK		Link to actual value of the row	FK to TDQ_VALUES	

BI_COMPARISON_ANALYSIS

Purpose	This view contains results of profiles (analysis) that compare two datasets that have the same columns – similar to a reconciliation.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
ANALYSIS_PK		ID of the Analysis from Talend Studio	FK to TDQ_ANALYSIS	
REPORT_NAME		Report name from Talend Studio		
REPORT_RUNTIME		Date/time from Talend Studio		

Purpose	This view contains results of profiles (analysis) that compare two datasets that have the same columns – similar to a reconciliation.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
ANALYSIS_NAME		Analysis name from Talend Studio		
INDICATOR_PK		ID of the Indicator		
INDICATOR_NAME		Indicator name from Talend Studio		
INDICATOR_CATEGORY		Indicator category from Talend Studio		
INDICATOR_DESCRIPTION		Indicator description from Talend Studio		
TABLE_A_PK		ID of the first table		
CONNECTION_NAME_A		Connection name for the first table		
TABLE_NAME_A		Name of the first table		
COLUMN_NAME_A		Colum name from the first table		
TABLE_B_PK		ID of the second table		
CONTEXT_NAME_A		Context for the first table		
CONNECTION_NAME_B		Connection name for the second table		
TABLE_NAME_B		Name of the second table		
COLUMN_NAME_B		Column name for the second table		
CONTEXT_NAME_B		Context for the second table		
NUMBER_OF_ROWS		Number of rows in scope for the match		
NUMBER_ROWS_MATCHED		Number of rows that matched		
MATCH_PERCENTAGE		Percentage match		

BI_OVERVIEW_ANALYSIS

Purpose	This view contains results of profiles (analysis) on data models and structures.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
REPORT_NAME		Report name from Talend Studio		
REPORT_RUN_TIME		Date/time when report was executed		
ANALYSIS_PK		ID of the Analysis from Talend Studio		
ANALYSIS_NAME		Analysis name from Talend Studio		
DATABASE_CONNECTION_NAME		Connection name from Talend Studio		
CATALOG_NAME		Catalog name from Talend Studio		
SCHEMA_NAME		Schema name from Talend Studio		
CONTEXT_NAME		Context name from Talend Studio		
TABLE_OR_VIEW_NAME		Table or view name from Talend Studio		
NUMBER_OF_ROWS		Number of rows (in table or view)		
NUMBER_OF_KEYS		Number of keys		
NUMBER_OF_INDEXES		Number of indexes		
NUMBER_OF_TRIGGERS		Number of triggers		

BI_DQ_FAILED_ROWS

Purpose	This view contains details of all the rows that failed a data quality rule. It holds the primary key of the row and the value that failed.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
FAILED_ROW_ID		Primary key for this table		
REPORT_PK		Primary key for the Talend Studio report		
DQ_RULE_PK		Primary key of the Data quality rule		
FIRST_DETECTED_RUN_ID		The RUN_ID when this break was first detected		
FIRST_DETECTED_ON		The date/time when this break was first detected		
BREAK_CLOSED_ON		The date/time when this break was closed		
ROW_KEY		The primary key of the row that failed		
ROW_FAILED_VALUE		The value of the row that caused the failure		
FAILURE_REASON		Description of the reason for failure		

BI_DQ_METRICS

Purpose	(Deprecated) This view brings together all relevant data across all views, so a dashboard can be built easily by using this view as the primary source. This is deprecated because this view aggregates all data at the data quality rule level, that is, total pass/fail by rule. The alternate approach is to measure data quality at the row (BI_DQF_ROW_SCORE) & dataset level (BI_DQF_DATASET_SCORE).			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		ID of the Rule execution	FK to DQF_RUN	
RUN_START		Date/time when rule executed		
RUN_END		Date/time when rule execution completed		
RUN_STATUS		Status of rule execution		
REPORT_PK		ID of the Report from Talend Studio		
REPORT_NAME		Report name from Talend Studio		
RUN_TIME		Rule execution date/time		
CONNECTION_NAME		Connection name from Talend Studio		
CONTEXT_NAME		Context name from Talend Studio		
ANALYSIS_TYPE		Analysis name from Talend Studio		
DQ_RULE_PK		ID of the DQ Rule	FK to DQF_DQ_RULE	
RULE_NAME		Name of the DQ Rule		
RULE_PRIORITY		Rule priority (*)		
DATASET_PK		ID of the Dataset	FK to DQF_DATASET	
DATASET_SOURCE_SYSTEM		Dataset Source system from Talend Studio		
DATASET_SOURCE_OBJECT		Dataset Source from Talend Studio		

Purpose	(Deprecated) This view brings together all relevant data across all views, so a dashboard can be built easily by using this view as the primary source. This is deprecated because this view aggregates all data at the data quality rule level, that is, total pass/fail by rule. The alternate approach is to measure data quality at the row (BI_DQF_ROW_SCORE) & dataset level (BI_DQF_DATASET_SCORE).
Related views	

Column Name	PK	Description	Relationships	Example
DATASET_BUSINESS_SYSTEM		Business name of source system		
DATASET_BUSINESS_NAME		Business name of the dataset		
DATASET_OWNER		Dataset owner name		
DATASET_STEWARD		Dataset steward name		
TDQ_TABLE_NAME		Physical table name from Talend Studio		
TABLE_NAME		Business table name		
TDQ_COLUMN_NAME		Physical column name from Talend Studio		
COLUMN_NAME		Business Column name		
INDICATOR_CATEGORY		Indicator category from Talend Studio		
INDICATOR_NAME		Indicator name from Talend Studio		
DQ_DIMENSION		DQ dimension		
DATA_OWNER		Rule data owner		
DATA_STEWARD		Rule data steward		
ROWS_PROCESSED		Number of rows processed by rule		
ROWS_FAILED		Number of rows that failed		
ROWS_PASSED		Number of rows that passed		
ROW_SCORE		Raw score = rows_passed / total_rows * 100		

Purpose	(Deprecated) This view brings together all relevant data across all views, so a dashboard can be built easily by using this view as the primary source. This is deprecated because this view aggregates all data at the data quality rule level, that is, total pass/fail by rule. The alternate approach is to measure data quality at the row (BI_DQF_ROW_SCORE) & dataset level (BI_DQF_DATASET_SCORE).			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
WT_SCORE		Weighted score = (rows_passed / total_rows * 100) * rule_priority		

(*) RULE_PRIORITY – is used to calculate the weighted score.

BI_DQ_RULE_ATTRIBUTE

Purpose	This view is a “flattened” view of the Rule attributes, which is useful to slice-and-dice DQ scores using custom data domains in the dashboard.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
DQ_RULE_PK		ID of the Rule	FK to DQF_DQ_RULE	
DATASET_FK		ID of the dataset	FK to DQF_DATASET	
RESPONSIBLE_TEAM		Custom attribute		
CONSUMING_TEAM		Custom attribute		
DATA_DOMAIN		Custom attribute		

BI_DQF_DATASET_SCORE

Purpose	(Deprecated) This view holds the aggregated DQ score for each dataset based on the score of the individual rows for each run (also known as rule execution).			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		The ID for the rule execution	FK to DQF_RUN	
DATASET_PK		The ID of the Dataset	FK to DQF_DATASET	
RUN_TIME		The date/time of the rule execution		
DATASET_SCORE		The aggregated data quality score for the entire dataset		
TOTAL_ROW_COUNT		The number of rows in the dataset at the time of execution.		

BI_DQF_DATASET_SUMMARY

Purpose	This view holds only the latest (that is, last executed run) details for each individual dataset. This view will NOT have any historical information.			
Related views	(In MySQL only) This view is now an alias to the materialized view – MV_DQF_DATASET_SUMMARY			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		The ID for the rule execution	FK to DQF_RUN	
DATASET_PK		The ID of the dataset	FK to DQF_DATASET	
RUN_TIME		The date/time of the rule execution		
TOTAL_RULES_APPLIED		The total rules applied		
TOTAL_RULES_FAILED		The number of rules that failed		
TOTAL_RULES_PASSED		The number of rules that passed		
DATASET_SCORE		The aggregated data quality score for the entire dataset		

Purpose	This view holds only the latest (that is, last executed run) details for each individual dataset. This view will NOT have any historical information.			
Related views	(In MySQL only) This view is now an alias to the materialized view – MV_DQF_DATASET_SUMMARY			
Column Name	**PK**	**Description**	**Relationships**	**Example**
TOTAL_ROW_COUNT		The number of rows in the dataset at the time of execution		
GOOD_ROWS		The number of rows where there wasn't a single data quality failure		
BAD_ROWS		The number of rows where there was at least one data quality failure		
COMPLETENESS_SCORE		The aggregated score for a DQ dimension		
VALIDITY_SCORE		The aggregated score for a DQ dimension		
ACCURACY_SCORE		The aggregated score for a DQ dimension		
UNIQUENESS_SCORE		The aggregated score for a DQ dimension		
TIMELINESS_SCORE		The aggregated score for a DQ dimension		
CONSISTENCY_SCORE		The aggregated score for a DQ dimension		

BI_DQF_DIMENSION_SCORE

Purpose	This view holds the aggregated DQ score by DQ dimension per dataset and per run.			
Related views	(In MySQL only) This view is an alias for the table – MV_DQF_ROW_SCORE_BY_DIMENSION			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		The ID for the rule execution	FK to DQF_RUN	
DATASET_FK		The ID of the dataset	FK to DQF_DATASET	
DQ_DIMENSION		The name of the DQ dimension		
DIMENSION_SCORE		The score for the DQ dimension		
TOTAL_ROW_COUNT		The number of rows that were in scope for this dimension		

BI_DQF_DQ_ATTRIBUTES

Purpose	(Deprecated) This view holds additional data at a rule level using a key-value pair structure. This is useful to slice-and-dice DQ scores using custom data domains.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
REPORT_NAME		The name of the report from Talend Studio		
RUN_PK		The ID for the rule execution	FK to DQF_RUN	
RUN_TIME		The date/time of the rule execution		
DATASET_PK		The ID of the dataset	FK to DQF_DATASET	
TABLE_NAME		The name of the dataset from Talend Studio		
DQ_ATTRIBUTE		The attribute key		
ATTR_DEPARTMENT		The attribute value		
WT_AVG_SCORE		The weighted average score by rule		

BI_DQF_DQ_DATASET

Purpose	This view holds reference data for each dataset.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
DATASET_PK		The ID for the dataset	FK to DQF_DATASET	
SOURCE_SYSTEM		The name of the application where the dataset is sourced from		
SOURCE_OBJECT_NAME		The technical / physical name of the dataset		
SOURCE_PK_DEFINITION		The technical/physical name of the primary key		

Purpose	This view holds reference data for each dataset.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
DSET_BUSINESS_SYSTEM		The department / group / acronym for the source system		
DSET_BUSINESS_NAME		The business name for the source application		
DSET_OWNER_FK		The ID for the data owner		
DSET_STEWARD_FK		The ID for the data steward		
MDF_OBJECT_ID_FK		(reserved for future use)		
DSET_DATA_OWNER_PK		The ID for the data owner		
DSET_DATA_OWNER_FIRST		The first name of the data owner		
DSET_DATA_OWNER_LAST		The last name of the data owner		
DSET_DATA_OWNER_EMAIL		The email address of the data owner		
DSET_DATA_OWNER_TITLE		The title of the data owner		
DSET_DATA_OWNER_IS_ACTIVE		Is the data owner still active		
DSET_DATA_STEWARD_PK		The ID for the data steward		
DSET_DATA_STEWARD_FIRST		The first name of the data steward		
DSET_DATA_STEWARD_LAST		The last name of the data steward		
DSET_DATA_STEWARD_EMAIL		The email address of the data steward		
DSET_DATA_STEWARD_TITLE		The title of the data steward		
DSET_DATA_STEWARD_IS_ACTIVE		Is the data owner still active		

BI_DQF_DQ_RULE

Purpose	The view holds reference data for the data quality rule.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RULE_PK		The Primary key for the Rule		
RULE_DATASET_PK		The ID of the dataset on which the rule will be applied	FK to DQF_DATASET	
RULE_DESCRIPTION		The description of the rule		
TDQ_REP_LABEL_FK		FK to the Report Label from Talend Studio		
TDQ_AN_LABEL_FK		FK to the Analysis Label from Talend Studio		
TDQ_ELT_TABLE_NAME_FK		FK to the physical table name from Talend Studio		
TDQ_ELT_COLUMN_NAME_FK		FK to the physical column name from Talend Studio		
TDQ_IND_LABEL_FK		FK to the Indicator name from Talend Studio		
INDICATOR_VALUE_IS_PASS_COUNT		Define if the value of the indicator is to be treated as a Pass count. (*)		
RULE_DATA_OWNER_FK		FK to the Data Owner		
RULE_DATA_STEWARD_FK		FK to the Data Steward		
RULE_DQ_DIMENSION		DQ dimension of the rule		
RULE_PRIORITY		Priority of the rule vs other rules		
BUSINESS_COLUMN_NAME		Business name for the column		
RULE_DATA_OWNER_PK		ID for the Data Owner (same as FK)		
RULE_DATA_OWNER_FIRST		Data Owner first name		
RULE_DATA_OWNER_LAST		Data Owner last name		
RULE_DATA_OWNER_EMAIL		Data Owner email		

Purpose	The view holds reference data for the data quality rule.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RULE_DATA_OWNER_TITLE		Data Owner title		
RULE_DATA_OWNER_IS_ACTIVE		Is the data owner still active (**)		
RULE_DATA_STEWARD_PK		ID for the Data Steward (same as FK)		
RULE_DATA_STEWARD_FIRST		Data Steward first name		
RULE_DATA_STEWARD_LAST		Data Steward last name		
RULE_DATA_STEWARD_EMAIL		Data Steward email		
RULE_DATA_STEWARD_TITLE		Data Steward title		
RULE_DATA_STEWARD_IS_ACTIVE		Is the Data Steward still active		

(*) INDICATOR_VALUE_IS_PASS_COUNT is either zero (false), one (true), or minus one (special case for dataset scope rules). When Talend Studio executes an indicator, it returns the measurement value for that indicator. However, the interpretation of that value – whether the value indicates a success or failure – is dependent on how the user interprets it and what the indicator is measuring.

For example, an indicator could identify that one or more NULL values exist in a particular column, and it may be desirable to define a rule that specifies NULL values are not allowed for that column. The count stored in the TDQ mart will be the total number of nulls, which is interpreted as the total number of failed checks. In this case set the INDICATOR_VALUE_IS_PASS_COUNT equal to one, which may appear to be counter intuitive, but there are reasons for this (Talend may choose to rename the column in a future release for clarity).

As a second example, the regular expression indicator records a count of rows that conform to the given regex in the TDQ mart, thus a calculation is needed to get the number of failed rows by subtracting the pass count from the total in scope row count. In this instance, the INDICATOR_VALUE_IS_PASS_COUNT should be set to zero.

There is a special case when defining a rule that has DATASET scope rather than ROW scope in the RULE_SCOPE field. An example of this would be a timeliness rule where there must be at least one record in the dataset that has been updated in the last week. This is calculated using a basic column analysis, a row count indicator, and a SQL filter for the analysis that contains the condition. If the analysis/row count indicator returns zero rows, this indicates that the dataset is failing the condition. To denote this special case, set the value minus one in the INDICATOR_VALUE_IS_PASS_COUNT field and DATASET in the field RULE_SCOPE. For more details, see the **Talend Data Quality Framework - Dataset Onboarding Guide**.

(**) RULE_DATA_OWNER_IS_ACTIVE – useful to identify if you have data owners who might no longer be with the organization but are still responsible for a rule.

BI_DQF_NOT_DEFINITION

Purpose	View that has the reference data for Notification definitions. Notifications are alerts sent to a user (usually by email) or to a system (such as Slack) to inform them of some event that occurs when executing the data quality rules.
Related views	

Column Name	PK	Description	Relationships	Example
NOT_DEF_PK		Primary key for the Notification definition		
DQ_RULE_FK		ID of the DQ rule	FK to DQF_DQ_RULE	
DATASET_FK		ID of the dataset	FK to DQF_DATASET	
SERVICE_METRIC_FK		ID of the service metric	FK to DQF_SERVICE_METRIC	
NOTIFICATION_TYPE		Type of Notification (*)		
NOTIFICATION_SCOPE		Scope of Notification (**)		
NOTIFICATION_PER_ROW		Should the notification be sent per row?		
THRESHOLD		Reserved for future use		
PRIORITY_LEVEL		Minimum priority level for notification generation		
DQ_DIMENSION		The DQ dimension if NOTIFICATION_SCOPRE is RULE		
DESCRIPTION		Notification description		

Purpose	View that has the reference data for Notification definitions. Notifications are alerts sent to a user (usually by email) or to a system (such as Slack) to inform them of some event that occurs when executing the data quality rules.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RULE_PK		ID for the Rule that triggers a notification	FK to DQF_DQ_RULE	
RULE_DATASET_PK		ID for the Rule Dataset	FK to DQF_DATASET	
RULE_DESCRIPTION		Description of the rule		
TDQ_REP_LABEL_FK		Physical Report Name from Talend Studio		
TDQ_AN_LABEL_FK		Physical Analysis Name from Talend Studio		
TDQ_ELT_TABLE_NAME_FK		Physical Table Name from Talend Studio		
TDQ_ELT_COLUMN_NAME_FK		Physical Column Name from Talend Studio		
TDQ_IND_LABEL_FK		Physical Indicator Name from Talend Studio		
INDICATOR_VALUE_IS_PASS_COUNT		Should the Indicator value be considered as a Pass?		
RULE_DATA_OWNER_FK		FK for the Rule Owner		
RULE_DATA_STEWARD_FK		FK for the Rule Steward		
RULE_DQ_DIMENSION		Rule DQ Dimension		
RULE_PRIORITY		Priority of the rule		
BUSINESS_COLUMN_NAME		User-friendly name for the column		
RULE_DATA_OWNER_PK		ID for Rule Owner (same as FK)		
RULE_DATA_OWNER_FIRST		Rule Data Owner First Name		
RULE_DATA_OWNER_LAST		Rule Data Owner Last Name		
RULE_DATA_OWNER_EMAIL		Rule Data Owner Email		

Purpose	View that has the reference data for Notification definitions. Notifications are alerts sent to a user (usually by email) or to a system (such as Slack) to inform them of some event that occurs when executing the data quality rules.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RULE_DATA_OWNER_TITLE		Rule Data Owner Title		
RULE_DATA_OWNER_IS_ACTIVE		Is Rule Data Owner active?		
RULE_DATA_STEWARD_PK		ID of Rule Data Steward		
RULE_DATA_STEWARD_FIRST		Rule Data Steward First Name		
RULE_DATA_STEWARD_LAST		Rule Data Steward Last Name		
RULE_DATA_STEWARD_EMAIL		Rule Data Steward email		
RULE_DATA_STEWARD_TITLE		Rule Data Steward title		
RULE_DATA_STEWARD_IS_ACTIVE		Is the Rule Data Steward still active?		
DATASET_PK		ID of the Dataset		
SOURCE_SYSTEM		The name of the application where the dataset is sourced from		
SOURCE_OBJECT_NAME		The technical / physical name of the dataset		
SOURCE_PK_DEFINITION		The technical / physical name of the primary key		
DSET_BUSINESS_SYSTEM		The department / group / acronym for the source system		
DSET_BUSINESS_NAME		The business name for the source application		
DSET_OWNER_FK		FK for the Dataset owner		
DSET_STEWARD_FK		FK for the Dataset Steward		
MDF_OBJECT_ID_FK				
DSET_DATA_OWNER_PK		ID for Dataset Owner (same as FK)		

Purpose	View that has the reference data for Notification definitions. Notifications are alerts sent to a user (usually by email) or to a system (such as Slack) to inform them of some event that occurs when executing the data quality rules.
Related views	

Column Name	PK	Description	Relationships	Example
DSET_DATA_OWNER_FIRST		Dataset Owner First Name		
DSET_DATA_OWNER_LAST		Dataset Owner Last Name		
DSET_DATA_OWNER_EMAIL		Dataset Owner Email		
DSET_DATA_OWNER_TITLE		Dataset Owner Title		
DSET_DATA_OWNER_IS_ACTIVE		Is Dataset Owner active?		
DSET_DATA_STEWARD_PK		ID of Dataset Steward		
DSET_DATA_STEWARD_FIRST		Dataset Steward First Name		
DSET_DATA_STEWARD_LAST		Dataset Steward Last Name		
DSET_DATA_STEWARD_EMAIL		Dataset Steward email		
DSET_DATA_STEWARD_TITLE		Dataset Steward title		
DSET_DATA_STEWARD_IS_ACTIVE		Is the Dataset Steward still active?		
SERVICE_METRIC_PK		ID of the service metric		
SERVICE_METRIC_LABEL		Name of the service metric		

(*) NOTIFICATION_TYPE: Possible values are:

- FAILED_ROWS
- RULE_PRIORITY

(**) NOTIFICATION_SCOPE: Possible values are:

- RULE
- DATASET

- GLOBAL

BI_DQF_NOT_GROUP_MEMBER

Purpose	All notifications are sent to a Group, and this view holds the list of members in a group. All members will receive the notification.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
MEMBER_PK		ID for the member		
NOT_GROUP_PK		ID of the Notification Group	FK to DQF_NOT_GROUP	
PERSON_PK		ID of the person	FK to DQF_PERSON	
GROUP_LABEL		Name of the Group		
GROUP_CHANNEL_IDENTIFIER		Name of the channel (usually for slack)		
MEMBER_TITLE				
MEMBER_FIRST_NAME				
MEMBER_LAST_NAME				
MEMBER_EMAIL				

BI_DQF_NOT_SUBSCRIPTION

Purpose	Users can subscribe to notifications, so they will only get what they are interested in. This view has the details of their subscriptions.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
NOT_SUB_PK		Primary key of the Subscription		
PERSON_PK		ID of the person	FK to DQF_PERSON	
SUBSCRIBER_NAME		Name of subscriber		
CHANNEL_LABEL		Type of communication channel - Slack or Email		
GROUP_FK		ID of the Group		
GROUP_NAME		Name of Group		
NOT_DEF_PK		See description for BI_DQF_NOT_DEFINITION		
DQ_RULE_FK				
DATASET_FK				
SERVICE_METRIC_FK				
NOTIFICATION_TYPE				
NOTIFICATION_SCOPE				
NOTIFICATION_PER_ROW				
THRESHOLD				
PRIORITY_LEVEL				
DQ_DIMENSION				
DESCRIPTION				
RULE_PK				

Purpose	Users can subscribe to notifications, so they will only get what they are interested in. This view has the details of their subscriptions.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RULE_DATASET_PK				
RULE_DESCRIPTION				
TDQ_REP_LABEL_FK				
TDQ_AN_LABEL_FK				
TDQ_ELT_TABLE_NAME_FK				
TDQ_ELT_COLUMN_NAME_FK				
TDQ_IND_LABEL_FK				
INDICATOR_VALUE_IS_PASS_COUNT				
RULE_DATA_OWNER_FK				
RULE_DATA_STEWARD_FK				
RULE_DQ_DIMENSION				
RULE_PRIORITY				
BUSINESS_COLUMN_NAME				
RULE_DATA_OWNER_PK				
RULE_DATA_OWNER_FIRST				
RULE_DATA_OWNER_LAST				
RULE_DATA_OWNER_EMAIL				
RULE_DATA_OWNER_TITLE				
RULE_DATA_OWNER_IS_ACTIVE				

Purpose	Users can subscribe to notifications, so they will only get what they are interested in. This view has the details of their subscriptions.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RULE_DATA_STEWARD_PK				
RULE_DATA_STEWARD_FIRST				
RULE_DATA_STEWARD_LAST				
RULE_DATA_STEWARD_EMAIL				
RULE_DATA_STEWARD_TITLE				
RULE_DATA_STEWARD_IS_ACTIVE				
DATASET_PK				
SOURCE_SYSTEM				
SOURCE_OBJECT_NAME				
SOURCE_PK_DEFINITION				
DSET_BUSINESS_SYSTEM				
DSET_BUSINESS_NAME				
DSET_OWNER_FK				
DSET_STEWARD_FK				
MDF_OBJECT_ID_FK				
DSET_DATA_OWNER_PK				
DSET_DATA_OWNER_FIRST				
DSET_DATA_OWNER_LAST				
DSET_DATA_OWNER_EMAIL				

Purpose	Users can subscribe to notifications, so they will only get what they are interested in. This view has the details of their subscriptions.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
DSET_DATA_OWNER_TITLE				
DSET_DATA_OWNER_IS_ACTIVE				
DSET_DATA_STEWARD_PK				
DSET_DATA_STEWARD_FIRST				
DSET_DATA_STEWARD_LAST				
DSET_DATA_STEWARD_EMAIL				
DSET_DATA_STEWARD_TITLE				
DSET_DATA_STEWARD_IS_ACTIVE				
SERVICE_METRIC_PK				
SERVICE_METRIC_LABEL				

BI_DQF_NOTIFICATION

Purpose	This view contains the history of all notifications sent to date.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
NOTIFICATION_PK		Primary key for this table		
RUN_FK		ID of the Rule execution		
SENT		Has the notification been sent?		
NOTIFICATION_CREATED_DT		When notification was created?		
FAILED_ROW_FK		FK for the row that failed (where relevant)		
NOTIFICATION_DESCRIPTION		Description of the notification		
SUBSCRIBER_OR_GROUP		Name of person or group		
NOT_SUB_PK		ID of subscription		
PERSON_PK		ID of the person	FK to DQF_PERSON	
SUBSCRIBER_NAME		See description for BI_DQF_NOT_SUBSCRIPTION		
CHANNEL_LABEL				
GROUP_FK				
GROUP_NAME				
NOT_DEF_PK				
DQ_RULE_FK				
DATASET_FK				
SERVICE_METRIC_FK				
NOTIFICATION_TYPE				
NOTIFICATION_SCOPE				

Purpose	This view contains the history of all notifications sent to date.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
NOTIFICATION_PER_ROW				
THRESHOLD				
PRIORITY_LEVEL				
DQ_DIMENSION				
DESCRIPTION				
RULE_PK				
RULE_DATASET_PK				
RULE_DESCRIPTION				
TDQ_REP_LABEL_FK				
TDQ_AN_LABEL_FK				
TDQ_ELT_TABLE_NAME_FK				
TDQ_ELT_COLUMN_NAME_FK				
TDQ_IND_LABEL_FK				
INDICATOR_VALUE_IS_PASS_COUNT				
RULE_DATA_OWNER_FK				
RULE_DATA_STEWARD_FK				
RULE_DQ_DIMENSION				
RULE_PRIORITY				
BUSINESS_COLUMN_NAME				
RULE_DATA_OWNER_PK				

Purpose	This view contains the history of all notifications sent to date.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
RULE_DATA_OWNER_FIRST				
RULE_DATA_OWNER_LAST				
RULE_DATA_OWNER_EMAIL				
RULE_DATA_OWNER_TITLE				
RULE_DATA_OWNER_IS_ACTIVE				
RULE_DATA_STEWARD_PK				
RULE_DATA_STEWARD_FIRST				
RULE_DATA_STEWARD_LAST				
RULE_DATA_STEWARD_EMAIL				
RULE_DATA_STEWARD_TITLE				
RULE_DATA_STEWARD_IS_ACTIVE				
DATASET_PK				
SOURCE_SYSTEM				
SOURCE_OBJECT_NAME				
SOURCE_PK_DEFINITION				
DSET_BUSINESS_SYSTEM				
DSET_BUSINESS_NAME				
DSET_OWNER_FK				
DSET_STEWARD_FK				
MDF_OBJECT_ID_FK				

Purpose	This view contains the history of all notifications sent to date.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
DSET_DATA_OWNER_PK				
DSET_DATA_OWNER_FIRST				
DSET_DATA_OWNER_LAST				
DSET_DATA_OWNER_EMAIL				
DSET_DATA_OWNER_TITLE				
DSET_DATA_OWNER_IS_ACTIVE				
DSET_DATA_STEWARD_PK				
DSET_DATA_STEWARD_FIRST				
DSET_DATA_STEWARD_LAST				
DSET_DATA_STEWARD_EMAIL				
DSET_DATA_STEWARD_TITLE				
DSET_DATA_STEWARD_IS_ACTIVE				
SERVICE_METRIC_PK				
SERVICE_METRIC_LABEL				

BI_DQF_PERSON

Purpose	This view holds the reference data for a person.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
PERSON_PK		Primary key for this person		
PERSON_FULL_NAME		Full name		
PERSON_FIRST_NAME		First name		
PERSON_LAST_NAME		Last name		
PERSON_EMAIL		Email address		
JOB_TITLE		Job title		

BI_DQF_ROW_SCORE

Purpose	This view holds the individual score for each row (not just the failed rows).			
Related views	BI_DQF_ROWS_WITH_FAILURES			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		ID of the Rule execution		
DATASET_FK		ID of the dataset		
SOURCE_ROW_KEY		Primary key of the row		
NUM_RULES_FAILED		No of rules that failed on this row		
NUM_RULES_APPLIED		No of rules that passed on this row		
ROW_SCORE		% score – rules_failed / rules_applied * 100		

BI_DQF_ROW_SCORE_BY_DIMENSION

Purpose	This view holds the aggregated row score by data quality dimension.			
Related views	(In MySQL only) This view is an alias for the table MV_DQF_ROW_SCORE_BY_DIMENSION			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		ID of the rule execution	FK to DQF_RUN	
DATASET_FK		ID of the dataset	FK to DQF_DATASET	
SOURCE_ROW_KEY		Primary key of the row		
DQ_DIMENSION		DQ dimension		
NUM_RULES_FAILED		No of rules that failed on this row for the given dimension		
NUM_RULES_APPLIED		No of rules that passed on this row for the given dimension		
ROW_SCORE		% score – rules_failed / rules_applied * 100		

BI_DQF_ROWS_WITH_FAILURES

Purpose	This rule counts the number of rules applied compared to rules failed for both types of rules, that is, ROW level rules and DATASET level rules. This is one of the most important views and is used as the base for other views.			
Related views	(In MySQL only) This view is an alias for the table MV_DQF_ROW_SCORE_BY_DIMENSION			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		ID of the rule execution	FK to DQF_RUN	
DATASET_FK		ID of the dataset	FK to DQF_DATASET	
SOURCE_ROW_KEY		Primary key of the row		
DQ_DIMENSION		DQ dimension		

Purpose	This rule counts the number of rules applied compared to rules failed for both types of rules, that is, ROW level rules and DATASET level rules. This is one of the most important views and is used as the base for other views.			
Related views	(In MySQL only) This view is an alias for the table MV_DQF_ROW_SCORE_BY_DIMENSION			
Column Name	**PK**	**Description**	**Relationships**	**Example**
NUM_RULES_APPLIED_ON		No of rules that were applied on this row for the given dimension		
NUM_RULES_FAILED		No of rules that failed on this row for the given dimension		

BI_DQF_WHAT_HAPPENED_TO_A_ROW

Purpose	This view is a useful view for debugging purposes when you want to investigate what happened to a particular row, such as how many rules were applied on that specific row and if there was a failure.			
Related views	(In MySQL only) This view is an alias to this table MV_ALL_ROWS			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_FK		ID of the rule execution	FK to DQF_RUN	
DATASET_FK		ID of the dataset	FK to DQF_DATASET	
SOURCE_ROW_KEY		Primary key for the row		
DQ_RULE_FK		ID of the rule	FK to DQF_DQ_RULE	
IS_ACTIVE		Is the rule active		
RULE_SCOPE		Rule scope – ROW level or DATASET level		
DESCRIPTION		Rule description		
DQ_DIMENSION		DQ Dimension		
ROW_FAILED_VALUE		Value of the row that failed		
FAILURE_REASON		Reason for failure		

BI_ROW_METRICS

Purpose	This view replaces the deprecated BI_DQ_METRICS. This view allows the ability to drill down into all failed rows.			
Related views	BI_DQF_WHAT_HAPPENED_TO_A_ROW			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_FK		ID of the rule execution	FK to DQF_RUN	
RUN_TIME		Rule execution date / time		
DATASET_TOTAL_ROW_COUNT		No. of rows that were in scope at the time of rule execution		
DATASET_FK		ID of the dataset	FK to DQF_DATASET	
DATASET_NAME		Name of the dataset		
DATA_OWNER_EMAIL		Data Owner name		
DATA_STEWARD_EMAIL		Data Steward name		
SOURCE_ROW_KEY		Primary key of the row		
DQ_RULE_FK		ID of the rule	FK to DQF_DQ_RULE	
RULE_IS_ACTIVE		Is the rule active		
DQ_DIMENSION		DQ dimension		
RULE_SCOPE		Scope of the rule – ROW or DATASET		
DESCRIPTION		Rule description		
DATA_ELEMENT		Name of the column rule was applied on		
RULE_PRIORITY		Rule priority		
ROW_FAILED_VALUE		Value of the column that failed		
FAILURE_REASON		Reason for failure		

BI_SCORE_BY_DATASET

Purpose	This view shows the score for the latest run across all datasets.			
Related views	BI_DQF_ROW_SCORE			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		ID of the rule execution	FK to DQF_RUN	
DATASET_PK		ID of the dataset	FK to DQF_DATASET	
RUN_TIME		Rule execution date/time		
TOTAL_RULES_APPLIED		Total rules applied		
TOTAL_RULES_FAILED		Total rules failed		
DATASET_SCORE		% score – sum of all row scores / total rows in dataset		
TOTAL_ROW_COUNT		Total rows in scope for the dataset at time of execution		
GOOD_ROWS		Total rows without a single rule failure		
BAD_ROWS		Total rows with at least one rule failure		

BI_SCORE_BY_DATASET_DIMENSION

Purpose	This view shows the score by dimension across all runs and datasets.			
Related views	BI_DQF_ROW_SCORE_BY_DIMENSION			
Column Name	**PK**	**Description**	**Relationships**	**Example**
RUN_PK		ID of the rule execution	FK to DQF_RUN	
RUN_TIME		Rule execution date / time		
DATASET_FK		ID of the dataset	FK to DQF_DATASET	
DQ_DIMENSION		DQ Dimension		

Purpose	This view shows the score by dimension across all runs and datasets.			
Related views	BI_DQF_ROW_SCORE_BY_DIMENSION			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DIMENSION_SCORE		Aggregated score for the DQ dimension		
TOTAL_ROW_COUNT		Total rows in scope at the time of rule execution		

BI_SCORE_BY_DATASET_HIST

Purpose	This view shows the score for all rule executions across all datasets (Same as BI_SCORE_BY_DATASET, but it covers all executions rather than just the latest one). **Note**: The Snowflake dashboard uses this view, but the MySQL dashboard uses the table DQF_SCORE_BY_DATASET_HIST for performance reasons.			
Related views	BI_DQF_ROW_SCORE			
Column Name	**PK**	**Description**	**Relationships**	**Example**
DATASET_PK		ID of the dataset	FK to DQF_DATASET	
DATASET_NAME		Dataset name		
RUN_PK		ID of the rule execution	FK to DQF_RUN	
RUN_TIME		Rule execution date / time		
TOTAL_RULES_APPLIED		Total rules applied		
TOTAL_RULES_FAILED		Total rules failed		
DATASET_SCORE		% score – sum of all row scores / total rows in dataset		
TOTAL_ROW_COUNT		Total rows in scope for the dataset at time of execution		
GOOD_ROWS		Total rows without a single rule failure		
BAD_ROWS		Total rows with at least one rule failure		

BI_V_IND_HISTO

Purpose	Core view that queries the system tables and views of the Talend Studio DQ Mart. This view is primarily used by BI_COLUMN_ANALYSIS, BI_COMPARISON_ANALYSIS, and BI_OVERVIEW_ANALYSIS.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
REP_UUID		For descriptions, see the Physical Data Model (PDM) page in Talend Help Center.		
AN_UUID				
REP_LABEL				
AN_LABEL				
AN_CREATION_DATE				
AN_AUTHOR				
AN_DATA_FILTER				
REP_CREATION_DATE				
REP_AUTHOR				
REP_STATUS				
ELT_UUID				
ELT_CONNECTION_UUID				
ELT_CONNECTION_NAME				
ELT_CATALOG_NAME				
ELT_SCHEMA_NAME				
ELT_TABLE_NAME				
ELT_COLUMN_NAME				
ELT_CTX_NAME				

Purpose	Core view that queries the system tables and views of the Talend Studio DQ Mart. This view is primarily used by BI_COLUMN_ANALYSIS, BI_COMPARISON_ANALYSIS, and BI_OVERVIEW_ANALYSIS.			
Related views				
Column Name	**PK**	**Description**	**Relationships**	**Example**
IND_UUID				
IND_LABEL				
IND_CATEG_UUID				
IND_CATEGORY				
INO_REGEXP				
CAL_DATE				
TIME_HOUR				
TIME_MINUTE				
RUN_TIME				
INDV_VALUE_TYPE_INDICATOR				
INDV_INT_VALUE				
INDV_REAL_VALUE				
INDV_DATE_VALUE				
INDV_ROW_COUNT				
VALUE_PK				