

Data Citizens Circles

How to leverage more value from Collibra Data Quality

September 2023

Welcome!

Zoom housekeeping and navigation

If you can hear us now, you're all set.

At this time, you are **muted by default** and should only be able to hear the presenters speaking.

To **enable audio**, raise your hand 🙋!

If you are having issues joining audio via the web, using the dial in numbers provided in the Chat.

Please enter any questions you have into the **Q&A** section on the right side of your screen.

Technical issues? Use 'Chat' feature
To submit questions use 'Q&A' feature
Audio issues? Check 'Audio settings'



Disclaimer

Any unreleased services or features referenced in this or other presentations, press releases or public statements are only intended to outline our general product direction. They are intended for information purposes only, and may not be incorporated into any contract. This is not a commitment to deliver any material, code, or functionality (which may not be released on time or at all) and customers should not rely upon this presentation or any such statements to make purchasing decisions. The development, release, and timing of any features or functionality described for Collibra's products remains at the sole discretion of Collibra

Agenda

01 Collibra Data Quality API integration with Data Intelligence Cloud

02 Pushdown with Snowflake

03 New reporting capabilities

04 New UI/UX

05 Q/A

A glimpse into the future: DQ to DIC Integration



George Luft
Data Quality Specialist, Collibra

Realized business value:

Increased trust, efficiency and data product adoption



Empower business users with data quality

Foster data literacy and adoption by enriching data context with quality metrics.



Ensure data quality in every initiative

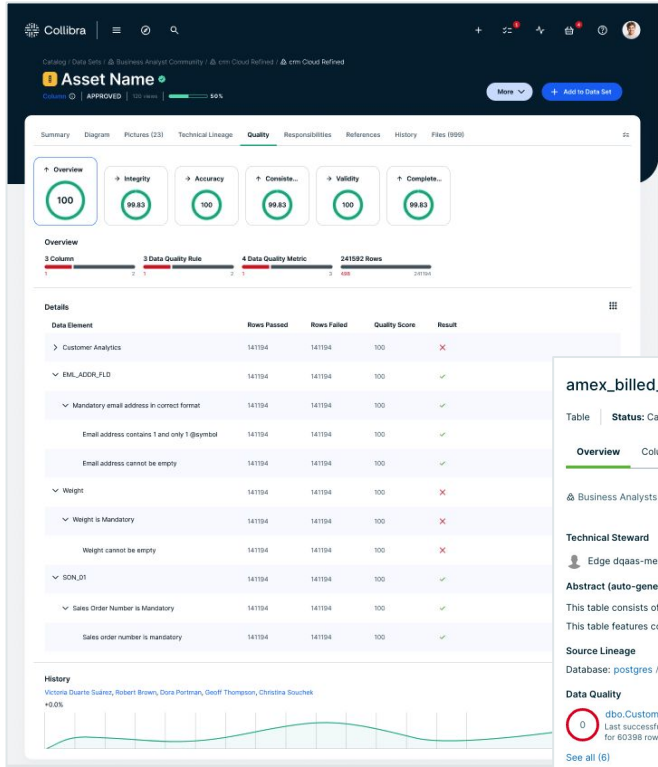
Increase trust in the business critical data tied to enterprise initiatives.



Govern data quality right from CDIC

Increase operational efficiency by governing and democratising data quality in Collibra Data Intelligence Cloud.

Empower teams with real-time Data Quality insights



Quality scores available in Data Marketplace & on Physical Data Assets

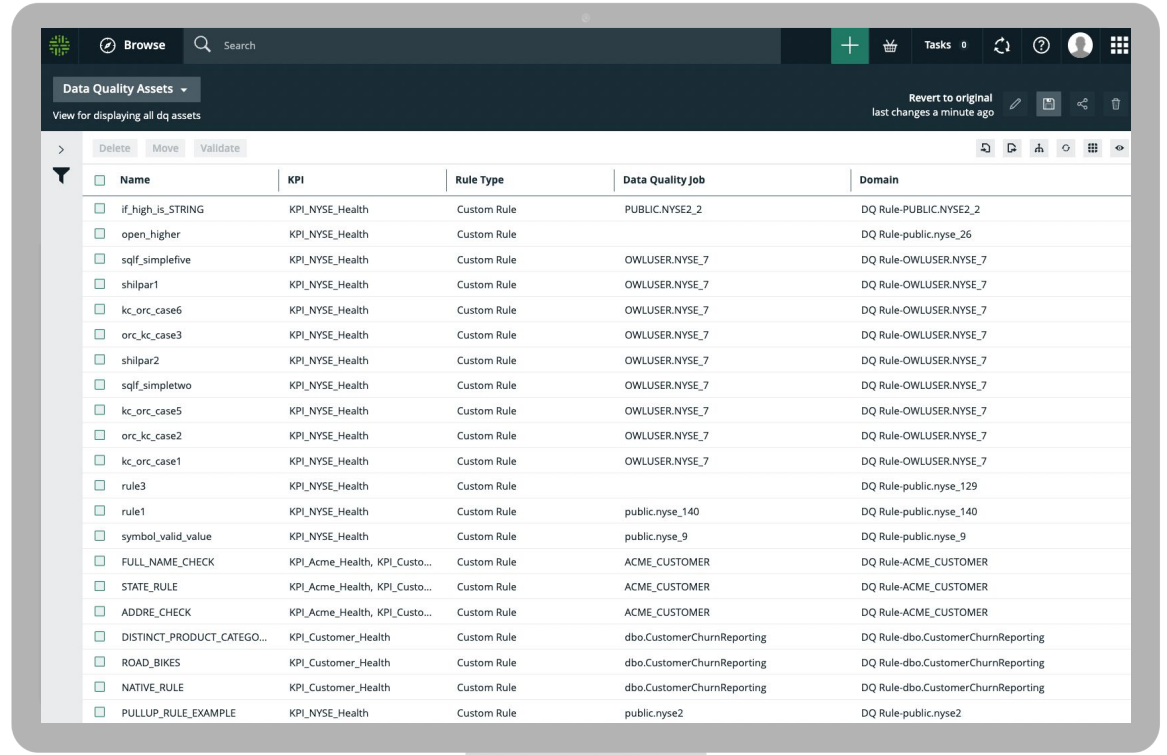
The screenshot shows the details for the 'amex_billed_transactions' table. It includes a table with columns for 'Status' and 'Candidate'. Below the table, there's an 'Overview' section with tabs for 'Columns (16)', 'Details', 'Responsibilities', and 'Actions (0)'. The 'Technical Steward' is listed as 'Edge dqas-meshon-two-edge1'. The 'Source Lineage' section shows the database as 'postgres / Schema: public'. The 'Data Quality' section displays three circular gauges: '0' for 'dbo.CustomerChurnReporting' (last successful run on 14 July 2023 for 60398 rows), '0' for 'default.bloomberg_2' (last successful run on 24 July 2023 for 500 rows), and '100' for 'default.nyse_38' (last successful run on 19 July 2023 for 102817 rows).



Drive literacy & adoption by minimizing clicks to find high quality data

Track DQ health across initiatives

Align Rules to KPIs to monitor DQ across data sources or lines of business to build trust in critical data



The screenshot displays a web-based interface for managing Data Quality Assets. At the top, there is a navigation bar with a 'Browse' button, a search field, and a 'Tasks' indicator. Below the navigation bar, a dropdown menu is set to 'Data Quality Assets', and a status message reads 'View for displaying all dq assets'. On the right side of the interface, there are options to 'Revert to original' and 'last changes a minute ago', along with icons for refresh, help, and user profile.

The main content area features a table with the following columns: Name, KPI, Rule Type, Data Quality Job, and Domain. The table lists 20 different data quality rules, each associated with a specific KPI and a data quality job.

Name	KPI	Rule Type	Data Quality Job	Domain
if_high_is_STRING	KPI_NYSE_Health	Custom Rule	PUBLIC.NYSE2_2	DQ Rule-PUBLIC.NYSE2_2
open_higher	KPI_NYSE_Health	Custom Rule		DQ Rule-public.nyse_26
sqf_simplefive	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
shilpar1	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
kc_orc_case6	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
orc_kc_case3	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
shilpar2	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
sqf_simpletwo	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
kc_orc_case5	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
orc_kc_case2	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
kc_orc_case1	KPI_NYSE_Health	Custom Rule	OWLUSER.NYSE_7	DQ Rule-OWLUSER.NYSE_7
rule3	KPI_NYSE_Health	Custom Rule		DQ Rule-public.nyse_129
rule1	KPI_NYSE_Health	Custom Rule	public.nyse_140	DQ Rule-public.nyse_140
symbol_valid_value	KPI_NYSE_Health	Custom Rule	public.nyse_9	DQ Rule-public.nyse_9
FULL_NAME_CHECK	KPI_Acme_Health, KPI_Custo...	Custom Rule	ACME_CUSTOMER	DQ Rule-ACME_CUSTOMER
STATE_RULE	KPI_Acme_Health, KPI_Custo...	Custom Rule	ACME_CUSTOMER	DQ Rule-ACME_CUSTOMER
ADDRE_CHECK	KPI_Acme_Health, KPI_Custo...	Custom Rule	ACME_CUSTOMER	DQ Rule-ACME_CUSTOMER
DISTINCT_PRODUCT_CATEGO...	KPI_Customer_Health	Custom Rule	dbo.CustomerChurnReporting	DQ Rule-dbo.CustomerChurnReporting
ROAD_BIKES	KPI_Customer_Health	Custom Rule	dbo.CustomerChurnReporting	DQ Rule-dbo.CustomerChurnReporting
NATIVE_RULE	KPI_Customer_Health	Custom Rule	dbo.CustomerChurnReporting	DQ Rule-dbo.CustomerChurnReporting
PULLUP_RULE_EXAMPLE	KPI_NYSE_Health	Custom Rule	public.nyse2	DQ Rule-public.nyse2

Assess pipeline Data Quality health for remediation

Data Element	Rows Passed	Rows Failed	Qual
Customer Analytics	1994	4	98.00
EMAIL	997	2	98.00
Shape Detections.Sales.Snowflake	997	2	98.00
Snowflake.Sales.Shapes	997	2	98.00
cost_code	997	2	98.00
Shape Detections.Sales.Snowflake	997	2	98.00
Snowflake.Sales.Shapes	997	2	98.00



Actionable Data Quality insights to drive Data Governance

Customer Analytics

Data Set Implemented (0) 100%

DQ Lineage

Columns flow through Data Attributes to Report Attributes. Columns have context of Tables & Schemas & Databases, and Business Terms. Report Attributes have context of Reports, Table...

Hierarchy left - right Complete Depth 50

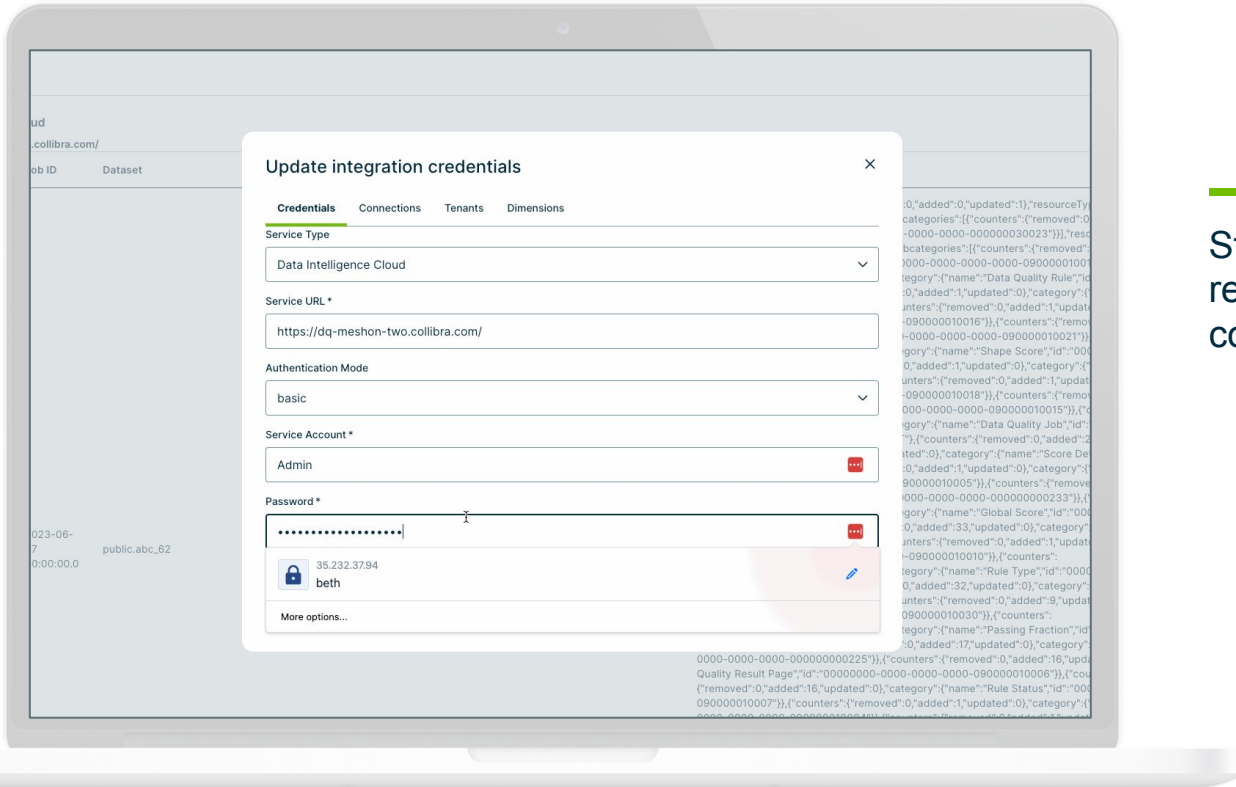
Diagram

Transaction Data (99%)

Customer Data (98%)

Customer Analytics (98%)

API-based integration with simple and easy mapping



Step by step guide for mapping relations to physical data assets, communities, and dimensions

2024 and beyond



Bi-Directional Integration

DQ Rule creation from Catalog by business users

Shared data classification between platforms

Alignment to business rules, code sets, and data stewards from Job/Rule Creation



MetaData Lineage Integration

Post **CDQ Scores** to Lineage

User **leverage DQ to assess what metadata to load** into DIC

Saturate catalog with high quality data from the start



Quality 360 Dashboard

Real-time dashboarding for enterprise level DQ in DIC

Highlight coverage of **ownership & assessment** for assets and their data domains

Create, maintain, and monitor DQ goals and KPIs



Data Quality in Lineage

Simple visual indicators to DQ thresholds & complex systems

Engage more stakeholders with DQ content

Workflow integration to trigger remediation

Estimated timeline

(subject to change)

Today

- APIs for integration available in Swagger
- New operating model in DIC including new DQ Job Asset

Q4

- Donut charts & aggregation in DIC (Banyan)
- DQ Scores in Data Marketplace (Banyan)

H2 24

- Create a rule in DIC
- Run a job from DIC

Q3

- UI based configuration in DQ
- Systemic table/column relations

H1 24

- DQ Scores on Asset Pages
- Improved Quality Tab & Reporting on DQ in DIC

Integration steps

Easier set-up from DQ

Update integration credentials ✕

Credentials | Connections | Tenant | Dimensions

Service Type
Data Intelligence Cloud ▾

Service URL *
https://dq-meshon-two.collibra.com

Authentication Mode
basic ▾

Service Account *
dq-user ⋮

Password *
 ⋮

Password Cannot Be Empty

Cancel **Next**

One-time task to connect your DIC

Configuration Control (from DQ)

Step by step mapping guide for relations, attributes, domains and communities

Credentials **Connections** Tenant Dimensions

Select Connection Enter DGC Database UUID. *

Select Connection Enter DGC Database UUID. *

Select Connection Enter DGC Database UUID. *

Credentials Connections **Tenant** Dimensions

You have established an integration between this DQ tenant and DIC. Mapped tenant: JP_TEST_...

Do you want to align your tenant to an existing DIC community?
 Yes No

Please select the name of the community it should be linked to.

Credentials Connections Tenant **Dimensions**

ACCURACY	Select Dimension <input type="text" value="Accuracy"/>
COMPLETENESS	Select Dimension <input type="text" value="Completeness"/>
CONSISTENCY	Select Dimension <input type="text" value="Completeness"/>
TIMELINESS	Select Dimension <input type="text" value="Accuracy"/>
UNIQUENESS	Select Dimension <input type="text" value="Duplication"/>

Ad-Hoc Sync or Active Schedule (from DQ)

Event-driven when job completes so data is always up-to-date

Collibra Search...

Integrations

Add and manage integrations between Collibra Data Quality & Observability and other applications

Data Intelligence Cloud
<https://dq-meshon-two.collibra.com>

Dataset	Last Update	Last Integration Status	Integration Status	Created By	Active
samples.botanic	2023-08-01T15:17:34.04164	COMPLETED	COMPLETED	brian	<input checked="" type="checkbox"/>
public.ah_currency	2023-08-14T23:59:18.605191	COMPLETED	COMPLETED	beth	<input checked="" type="checkbox"/>
default.attributes_2	2023-08-10T23:59:22.840066	COMPLETED	COMPLETED	beth	<input checked="" type="checkbox"/>
samples.census_tracts_alabama_4	2023-08-10T20:27:38.397376	COMPLETED	COMPLETED	DQ USER	<input checked="" type="checkbox"/>
samples.nyse_job_load_Rule_Delete_Test	2023-09-12T18:12:36.579518	COMPLETED	COMPLETED	beth	<input checked="" type="checkbox"/>
public.nyse2	2023-09-20T00:07:31.163058	COMPLETED	COMPLETED	mcampab	<input checked="" type="checkbox"/>
public.abc_20skedule2	2023-09-20T10:22:29.518179	COMPLETED	COMPLETED	brian	<input checked="" type="checkbox"/>
public.aclaims_master_9	2023-07-27T16:59:29.190221	COMPLETED	COMPLETED	susmitha	<input checked="" type="checkbox"/>
public.nyse_163_utc_23	2023-09-19T23:03:32.333215	COMPLETED	COMPLETED	beth	<input checked="" type="checkbox"/>
public.nyse_139	2023-09-20T04:03:38.324428	COMPLETED	COMPLETED	susmitha	<input checked="" type="checkbox"/>
OWLUSER.NYSE_7	2023-07-27T16:58:39.870138	COMPLETED	COMPLETED	zeyu2017	<input checked="" type="checkbox"/>
PUBLIC.NYSE2_2	2023-07-27T17:02:51.02137	COMPLETED	COMPLETED	susmitha	<input checked="" type="checkbox"/>

Integration Details

Frequency: Daily

Scheduled Time	Date Format	Mon	Tue	Wed	Thu	Fri	Sat	Sun	TimeZone	Frequency
-09-20 17:50:00	yyyy-MM-dd								UTC	DAILY
-09-20 17:30:00	yyyy-MM-dd								UTC	DAILY
-09-20 17:10:00	yyyy-MM-dd								UTC	DAILY
-09-20 00:00:00	yyyy-MM-dd								UTC	DAILY
-09-20 07:50:00	yyyy-MM-dd								US/Eastern	DAILY
-09-20 11:15:00	yyyy-MM-dd								UTC	DAILY
OWLDB2.NYSE_STOCKS_18	2023-09-19 07:20:23								UTC	DAILY
OWLUSER.NYSE_38_sh_rundt	2023-09-18 22:20:40								UTC	DAILY

Enable Integration

The screenshot displays the Collibra interface with two views of the 'public.accounts' dataset. The top view shows the dataset metadata bar with an 'Enable Integration' button highlighted by a green box and a green arrow. The bottom view shows the Dataset Manager table with an action menu open for the 'public.nyse_5' dataset, where the 'Enable Integration' option is also highlighted by a green box and a green arrow.

Dataset Manager Table:

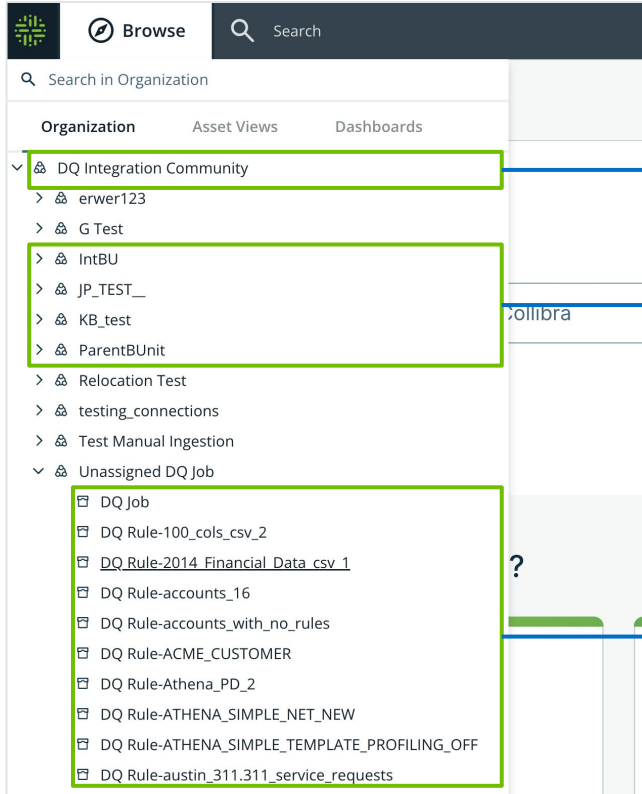
Dataset	Source Type	Schema	Table	Meta Tags	Server	Action
public.event	Redshift	public	event		redshift-cluster-1.cctgcaphfd.us-east-2.redshift.amazonaws.com	Acti...
DWLD82.NYSE_STOCKS_3	Db2	DWLD82	NYSE_STOCKS_3		10.150.0.3	
public.loan_customer	Postgres	public	loan_customer		35.190.143.55	
hl_rules_test	Oracle	OWLUSER	NYSE		jdbc	
public.nyse_5	Postgres	public	nyse		35.190.143.55	
OWLUSER.NYSE_rules	Oracle	OWLUSER	NYSE		jdbc	

Dataset Metadata Bar (public.accounts):

- Run Time: 00:00:19
- Rows: 50
- Columns: 4
- Alerts: 1
- Postgres
- Profile
- Rules: (2)
- Schedule: (2)

- Two ways to enable integration
 - Dataset Manager Action Menu
 - Dataset Metadata bar
- REACT_MUI must be TRUE

View Results in Data Intelligence Cloud - Organization



Top Level Community

- Can be created new for integration or aligned to an existing community

Sub-Communities

- Created to reflect business units in DQ
- If no business unit is aligned to a dataset, assets will land in the “Unassigned DQ Job Community”

Integration Domains

- DQ Job → Houses DQ job assets for all integrated datasets
- DQ Rule → Houses all DQ rule assets for a single integrated dataset
- DQ Score → Houses all DQ metric (score) assets for a single integrated dataset

Table Results in Data Intelligence Cloud - DQ Job Assets

- Summary of the DQ dataset
- Global Score attribute reflects most recent job score out of DQ
- Predicate & Job Row Filter indicate tables, columns, and row constraints added to the job
- Systemic relations to the table it impacts and data quality rules executed by the job will be visible
- Timestamps for the last run and successful run indicate how recent the insights are on the asset

The screenshot displays the Data Intelligence Cloud interface for a Data Quality Job Asset named 'default.austin_crime_2'. The asset is of type 'Data Quality Job' and has a status of 'Candidate'. The interface includes a navigation menu on the left with options like Overview, Tags, Comments, Diagram, Pictures, Quality, Responsibilities, References, History, and Files. The main content area shows the following details:

- Global Score:** 97
- Last Run Status:** Success
- Predicate:** select * from default.austin_crime limit 1000
- Job Row Filter:** 1,000
- represented by Business Asset:** A table with columns Name, Domain, and Definition. The row shows 'KPI_Athena_Health' in the Name column, 'dq-health-kpis' in the Domain column, and a trash icon in the Definition column.
- represents Table:** A table with columns Name, Domain, Description, and represented by. The row shows 'austin_crime' in the Name column, 'APPROVED_ATHENA_UP > Aws...' in the Domain column, and a trash icon in the represented by column.
- Last Job Run:** Mon Jul 24 09:40:03 UTC 2023
- Last Successful Job Run:** Mon Jul 24 09:40:03 UTC 2023

Column Results in Data Intelligence Cloud - DQ Rule Assets

- Assets created for all 9 areas of data quality analysis in DQ
- Description, rule status and rule type define the type of rule
- Measurement is the total possible score
- Predicate is the SQL statement for the rule
- Systemic relations to DQ Dimensions, DQ Score (metrics), columns, and DQ job available
- Hyperlink to findings page in DQ for easy cross tool navigation

The screenshot displays the 'EXCH:Data Type Range' rule asset in the Data Intelligence Cloud. The interface includes a navigation sidebar on the left with options like Overview, Tags, Comments, Diagram, Pictures, Quality, Responsibilities, References, History, and Files. The main content area shows the rule's description, status (Candidate), and type (Adaptive Rule). It also displays the measurement score (100) and the predicate SQL statement: `EXCH.$mixedTypePercent NOT BETWEEN 0.0 AND 0.0`.

The 'classified by Data' section contains a table with the following data:

Name	Domain	Description
EXCH:Data Type Range Score	DQ Score-kb_cxn_db2	AI enabled rules to assess beh...

The 'governs Data Quality Job' section contains a table with the following data:

Name	Domain
kb_cxn_db2	DQ Job

The 'Link to Data Quality Result Page' section includes a link to 'View Results in CDQ'.

The 'governs Column' section contains a table with the following data:

Name	Domain	Description
EXCH	APPROVED_DB2_UP > owlb2 ...	

View results in Data Intelligence Cloud - DQ Score Assets

- Quantitative output of rule execution
- Loaded rows indicate how many rows were analyzed
- Measurement is the total possible score and passing fraction is the actual score of the job run of the rule
- Score details will highlight a summary of the findings for shapes, duplicates, outliers, source, records, and patterns
- Systemic relation to the data quality rule allows for easy transitions between assets

The screenshot displays the 'EXCH:Data Type Range Score' asset in the Data Intelligence Cloud interface. The asset is of type 'Adaptive Rule Score' and is currently in 'Candidate' status. Key metrics shown include 102,815 loaded rows, a measurement of 100, and a passing fraction of 100. The score status is 'Observation'. A table at the bottom details the rule execution, showing the rule name 'EXCH:Data Type Range', domain 'DQ Rule-kb_cxn_db2', and description 'AdaptiveRule: AdaptiveRules a...'. The interface also features a navigation menu on the left with options like Overview, Diagram, Pictures, Responsibilities, References, History, and Files.

DQ Pushdown



Nita Dembla
Director, DQ Software Engineering



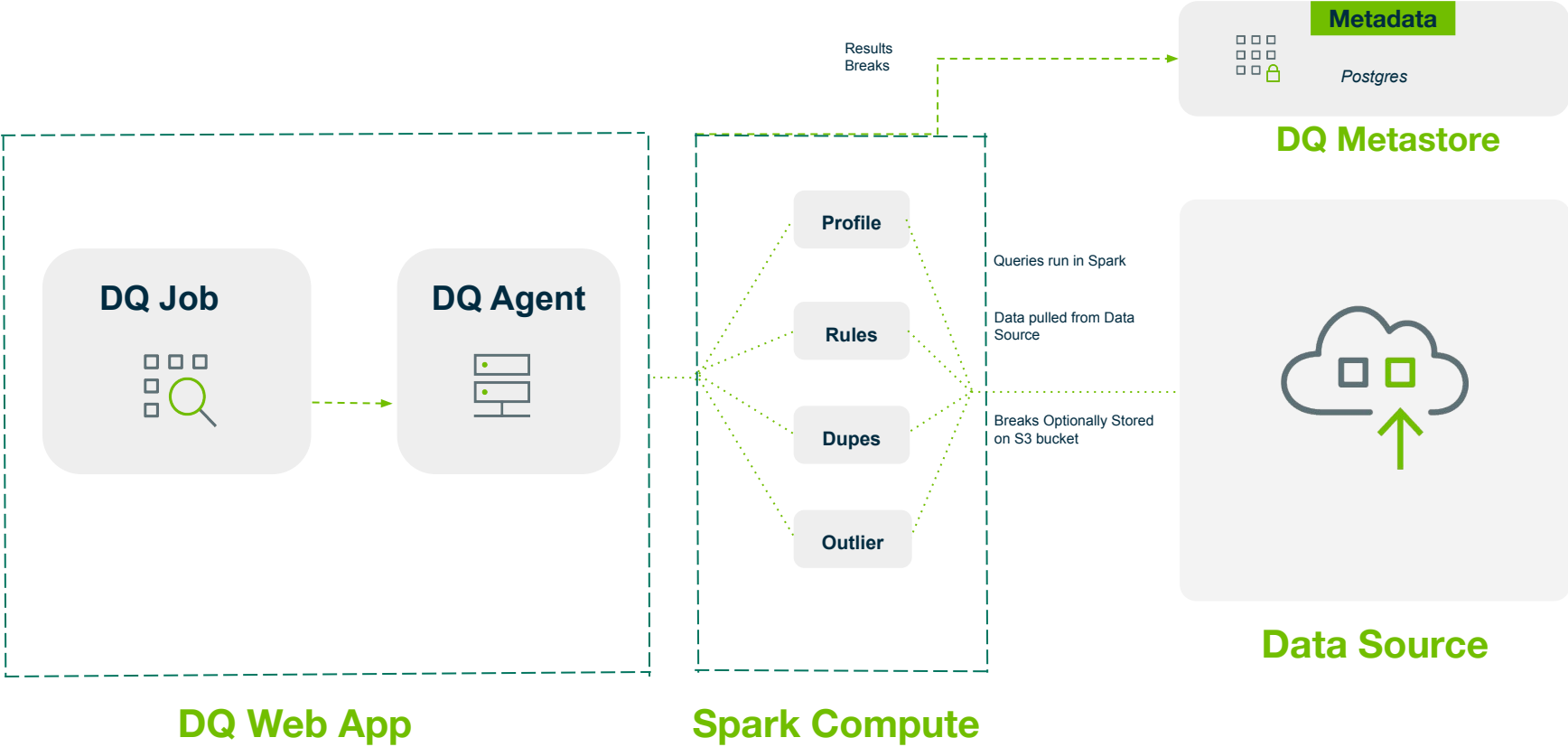
Laurent Weichberger
Sr. Manager DQ Customer Success

Pull-up

What is a **Pull-Up** (default) DQ process:

- Pull-up pull data from the customer data source into an Apache Spark DataFrame.
- Profile logic runs against the **DataFrame** in memory
 - Adaptive/Automatic Rules
 - Custom Rules
 - Advanced functionality (eg. Outliers, Patterns, Dupes, Validate Source, etc.)
- Break records, and DQ metrics, are stored in the customer Collibra DQ Metastore

Pull-up processing

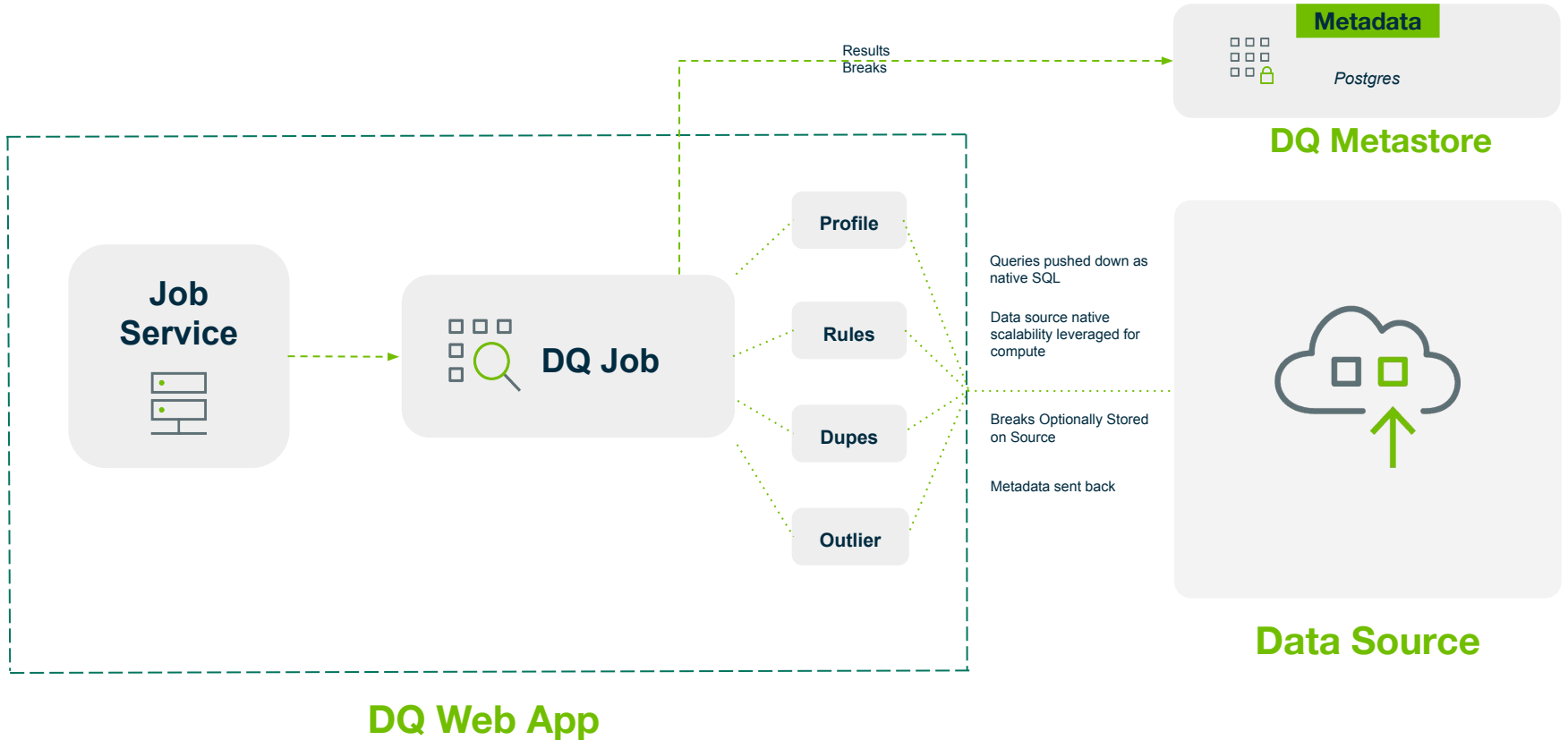


Pushdown

With **Pushdown** we have a totally different DQ process:

- Data **does not move** from the customer data source.
- Profile logic is packaged by DQ and sent to the data source.
- Logic runs against the target **Pushdown** platform (eg. Snowflake):
 - **We push:** Adaptive/Automatic Rules
 - **We push:** All rules - Custom, Freeform
 - **We push:** Advanced functionality (eg. Outliers, Dupes, etc.)
- Break records can be stored in the target **Pushdown** platform **Cloud Warehouse**

Pushdown processing



* Parallel execution of job activities queuing SQL queries across open connections.
* Asynchronous job execution in its own thread opens configurable number of jdbc connections.

Benefits of pushdown offering



Reduce costs, improved security

Eliminate the egress charges of reading all the data out to do processing



Get faster time to value

Eliminate dependencies on Spark compute and increase process speeds



Increase scalability & agility

Achieve unlimited scale by scaling your computational needs based on demand

Credits for 100 columnna

PUBLIC.DB_REPORTING2_3 PD

Run Time: 00:01:08 snowflake Profile Rules (35) fx Schedule

Rows: **1.00M**

Columns: 100

Alerts: 0

0.26 credits/run

1 M rows

PUBLIC.DB_REPORTING2_3 PD

Run Time: 00:03:01 snowflake Profile Rules (35) fx Schedule

Rows: **10.00M**

Columns: 100

Alerts: 0

0.80 credits/run

10 M rows

**100
Jobs**

80 credits

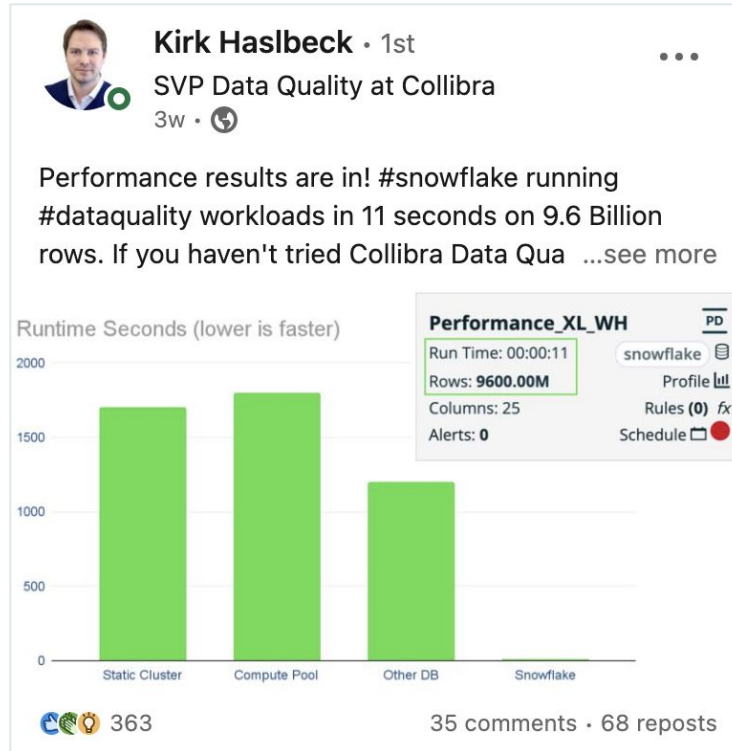
10 M rows

Collibra DQ Pushdown: Setup the Connection

- Permissions setup for - user, roles, Virtual Warehouse, databases, schema, etc.
 - ⇒ A template script for Snowflake is [here](#)
- Designate separate source `database . schema` to write break records for a connection, to which service account user will need additional privileges - `WRITE` and `CREATE TABLE`.
- Each DQ activity will write to an output table the results of the run.
- A separate table will list out the query to run to fetch the break records per activity of a DQ run.
- Timeline based pruning being considered to keep table sizes under check.

Data Quality Pushdown:

Get to value faster and with lower cost



Collibra DQ Pushdown Platforms

DQ Pushdown Platform Name	General Availability
Snowflake	GA Now
Databricks	GA Now
Google BigQuery	Beta Now
AWS Athena	Beta September 2023



New Reporting Capabilities



Christopher Schmidt
Data Quality Specialist

Dataset Dimension Report - Filter (1/2)



Dataset Dimension

An overview to view dataset dimensions with different filters to make more informed decisions and increase the value of my data.

Dataset

All

Business Unit

All

1

2

3



Accuracy



Consistency



Completeness



Uniqueness




Validity



Timeliness

Dataset Dimension Report - Filter (2/2)



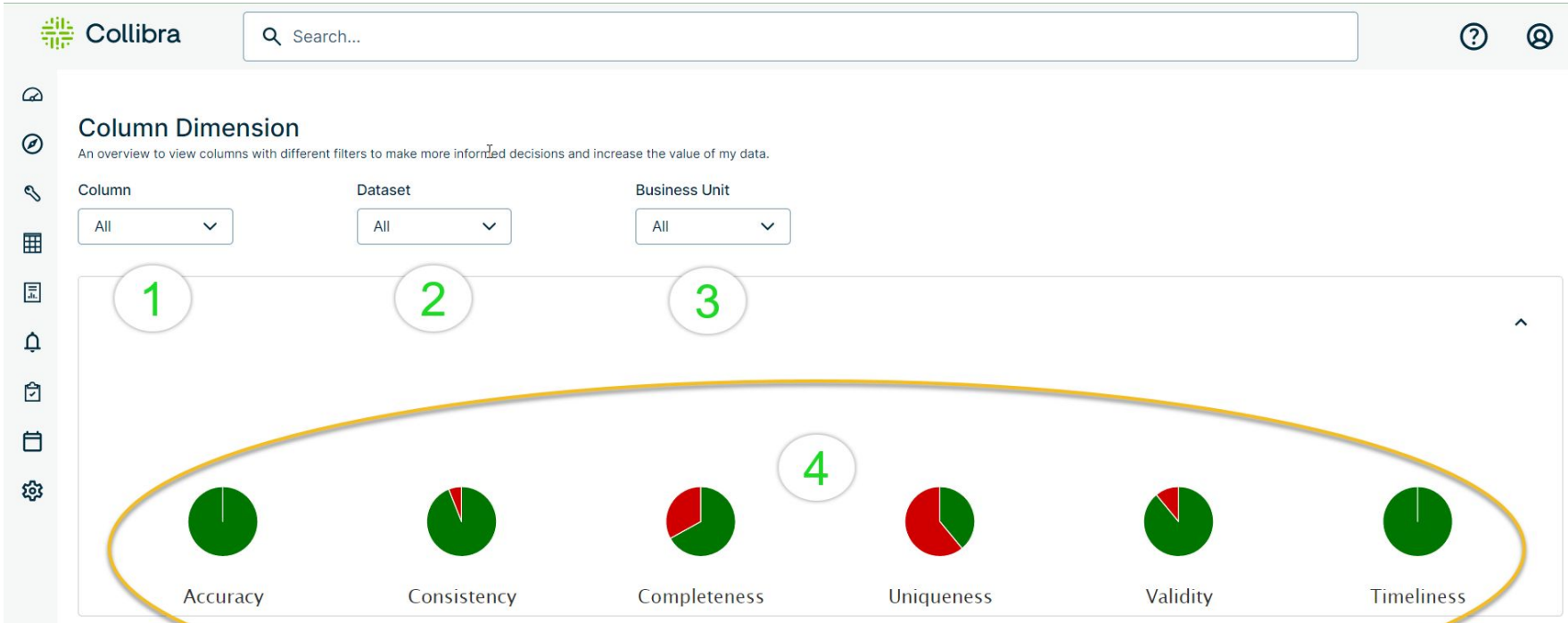
Intervals

Search... Current ^

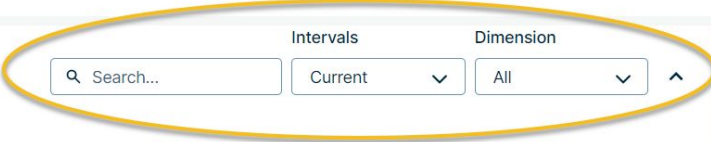
Dataset	Business Unit	Accuracy	Consistency	Completeness	Uniqueness	Validity	Timeliness
berka.accounts_11_alerts_rules_schedule	JP_TEST__	100	100	0	0	100	100
berka.accounts_12	PulseTest		100				
berka.accounts_13	PulseTest		100				
berka.transactions_postgresgcp	sdffsdfs		0		79		
KB_OWLUSER.CUSTOMER.ME				100			
public.abc_17			100				
public.customers_4_dev17410	sdffsdfs		100			98	
public.nyse2						100	
TPCH_SF001.CUSTOMER				100		0	

1-9 of 9 < 1 >

Column Dimension Dashboard (1/2)

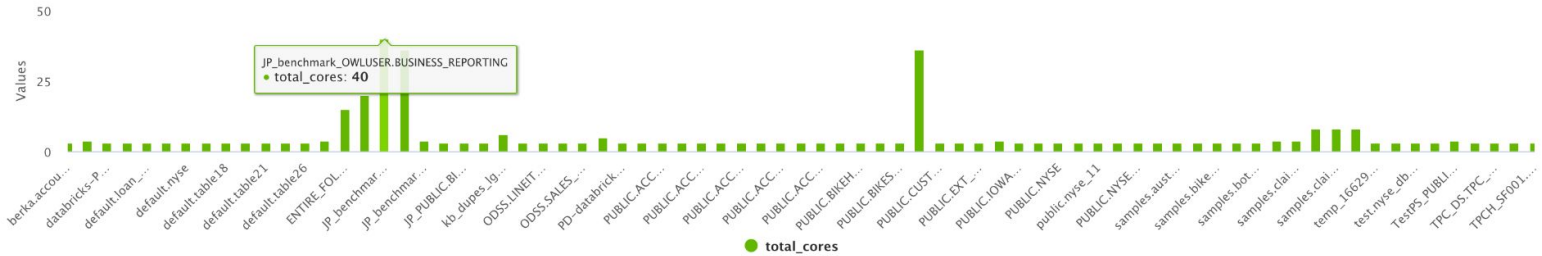


Column Dimension Dashboard (2/2)



Column	Rule Name	Dataset	Dimension	Business Unit	DQ Score
account_id ▾	if_account_id_is_valid_AMEX	berka.accounts_11_alerts_rules_schedule	UNIQUENESS	JP_TEST_	0
account_id ▾	kc_ff1	berka.accounts_13	CONSISTENCY	PulseTest	100
ADDR_LINE_2 ▾	if_ADDR_LINE_2_is_EMPTY	KB_OWLUSER.CUSTOMER.ME	COMPLETENESS		100
apples ▾	kc1	public.abc_17	CONSISTENCY		100
auto_make ▾	auto_make_uniquePercent	public.customers_4_dev17410	CONSISTENCY	sdfsfdfs	100
auto_year ▾	auto_year_valid_value	public.customers_4_dev17410	CONSISTENCY	sdfsfdfs	100
C_CUSTKEY ▾	if_C_CUSTKEY_is_STRING	TPCH_SF001.CUSTOMER	VALIDITY		0
date ▾	if_date_is_DATE_STR_TYPE	berka.accounts_11_alerts_rules_schedule	VALIDITY	JP_TEST_	100

Hardware Usage Report

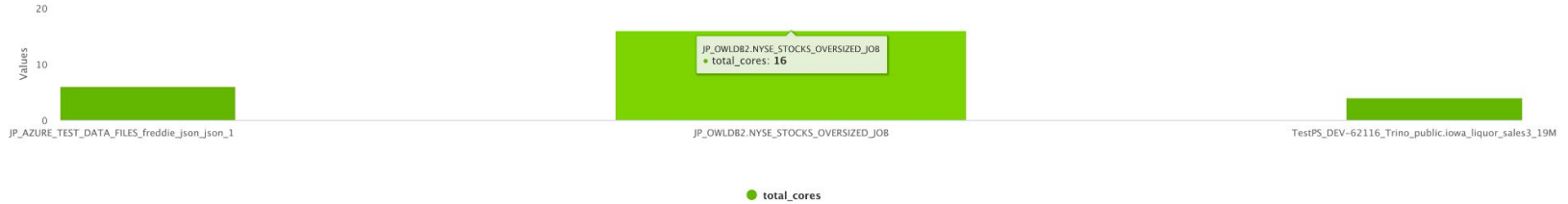


dataset	num_executors	executor_cores	executor_memory	total_cores	total_memory	avg_row_count ↓
ENTIRE_FOLDER	15	1	10g	15	150	234888087
samples.claim_full	8	1	24g	8	192	100716530
samples.claim_5	8	1	24g	8	192	100716530
samples.claim_3	4	1	24g	4	96	100716530
samples.claim_4	4	1	24g	4	96	100716530
JP_benchmark_samples.claim	2	2	8g	4	16	100716530

Oversized Job Report

Oversized Job Report

JOBS THAT HAVE MORE HARDWARE ASSIGNED THAN REQUIRED TO RUN



dataset	reason	recommended_total_memory	recommended_total_cores	recommended_num_executors	recommended_executor_memory	num_executors	executor_cores
JP_OWLDB2.NYSE_STOCKS_OVERSIZED_JOB	Too Much Memory & Cores	1	1	1	1	4	4
JP_AZURE_TEST_DATA_FILES_freddie_json_json_1	Too Much Memory & Cores	1	1	1	1	3	2
TestPS_DEV-62116_Trino_public.iowa_liquor_sales3_19M	Too Much Memory	23	3	1	23	2	2

Search...

1-3 of 3 < 1 >

User Activity Report

User Activity

THIS REPORT SHOWS CLICK ACTIVITY IN THE APPLICATION BY USER

Filter Type

lookback

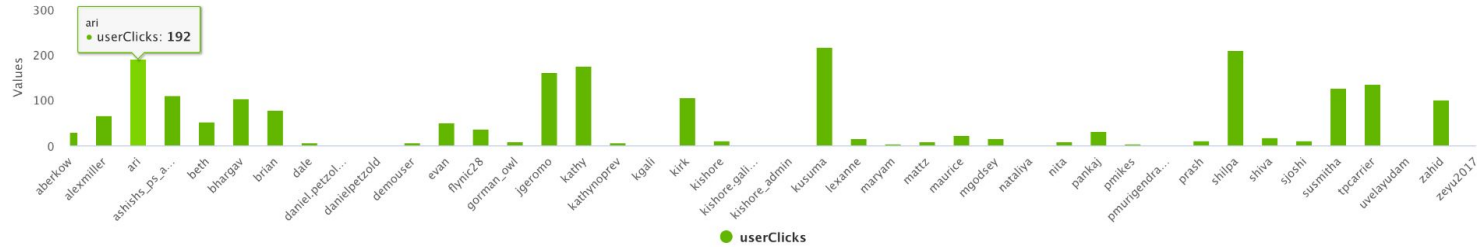
Last x Days

7

Limit

0

Go



Search...

username	userClicks ↓
kusuma	217
shilpa	211
ari	192
kathy	175
jgeromo	162
tpcarrier	136

Reporting capabilities coming in 2023.09

- Column filter is powered by the dataset filter = user selects a dataset and only the columns of that dataset will show
- Last X time filter powers the entire dashboard (based off job runs)
- Time Aggregation was simplified to a much more understandable view on the data (no filtering happening)
- Export LinkIDs functionality to download a csv of breaking records for a given rule (like it exists on the findings page)
- Sample SQL pulls for reports are provided in the documentation

New UI/UX



Eric Gerstner
Principal Solutions Engineer

Consolidated Flows

More data under quality management with fewer clicks

You're currently using Explorer beta. You can still [revert](#) to the classic Explorer at any time.

Job Creator Beta

SELECTED TABLE: ESG_DATA_OCTOBER

[Data Set Overview](#)

What's next?

Select a scanning method to create a DQ job.

Full Scan
Create a DQ Job that scans all rows of your table.
This may take a few minutes for larger tables.

Select

Partial Scan
Create a DQ Job that scans a section of your table.

Select

SQL Query
Compile SQL query manually for an advanced scan of your table.

Select

Agent: owl-dq-owl-agent-owl-dq-dev-0 (Online)

Not what you're looking for? [Go to Jobs Page](#) [Go to DQ Homepage](#)

Dataset Rules
Definitions for dataset level rules.

Select Dataset For Rule To Get Started

Search by dataset

App Frequent Runs: berka.accounts_13, berka.accounts_12, OWLUSER.NYSE_7

App Recent Runs

Table: austin_crime	Runs: 6	databricks	Ran 7 minutes ago
PUBLIC.AUSTIN_CRIME_16			
Table: AUSTIN_CRIME	Runs: 1	snowflake	Ran 12 minutes ago
PUBLIC.AUSTIN_CRIME_15			
Table: AUSTIN_CRIME	Runs: 6	snowflake	Ran 14 minutes ago
2014_Financial_Data_csv_1			
Table: 2014_Financial_Data.csv	Runs: 20	File:s3	Ran 15 minutes ago
PUBLIC.AUSTIN_CRIME_14			
Table: AUSTIN_CRIME	Runs: 6	snowflake	Ran 19 minutes ago

Show More

Rule
Workbench

Interactive
Results

Iterative
Refinement

Rule Workbench

Interactive Results

Iterative Refinement

The screenshot displays the Collibra Rule Workbench interface. At the top left is the Collibra logo and a search bar. A sidebar on the left contains navigation icons. The main area shows a 'Back to Rule List' link and a 'Type' dropdown set to 'Freeform SQL'. Below this, a SQL query is partially visible: '1 SELECT * FROM'. A modal window titled 'Rule Workbench' is open, showing a dropdown menu with the text 'using this collibra governed dataset'. The selected item is 'TPCDS_SF10TCL.STORE_RETURNS'. The dropdown list includes the following items: 'TPCDS_SF10TCL.STORE_RETURNS', '2014_Financial_Data_csv_1', 'dbo.nyse2_t1lookuprule', 'public.amex_client_profile_2', 'dbo.CustomerProductSales_5', 'OWLUSER.CUSTOMER_missing_data_chk', 'public.nyse_92', and 'OWLUSER.NYSE_7'.

Bulks Actions: Advanced ways to manage rules

The screenshot shows the Collibra interface with a search bar at the top. Below it, a section titled "Catalog by Columns" contains a table of datasets. A search dropdown menu is open over the "Run Mode" column, showing a list of options: react, cdata, connect, rules, admin, and promo. The table has columns for Dataset, Field Name, Data Class, Sensitive Label, Data Category, Business Unit, Source, Run Mode, and an unlabeled column. The first two rows have "AMEX" in the Data Class column.

Dataset	Field Name	Data Class	Sensitive Label	Data Category	Business Unit	Source	Run Mode	
100_cols_csv_1	condition	AMEX			JP_TEST	File:s3	draft	
100_cols_csv_1	although	AMEX			JP_TEST	File:s3	draft	
100_cols_csv_1	harder				JP_TEST	File:s3	draft	String
100_cols_csv_1	successful				JP_TEST	File:s3	draft	String
100_cols_csv_1	alike				JP_TEST	File:s3	draft	String
100_cols_csv_1	muscle				JP_TEST	File:s3	draft	String
100_cols_csv_1	test				JP_TEST	File:s3	draft	String
100_cols_csv_1	national				JP_TEST	File:s3	draft	String
100_cols_csv_1	talk				JP_TEST	File:s3	draft	String
100_cols_csv_1	forty				JP_TEST	File:s3	draft	String
100_cols_csv_1	whispered				JP_TEST	File:s3	draft	String
100_cols_csv_1	sand				JP_TEST	File:s3	draft	String

Better Alerting: Dataset Watching, Slack Integration, 1-click alert types

Collibra | Data Quality

Alerts

Set email alerts to send notifications when alert conditions are met.

Type ↑	Resource	Condition	Recipient(s)
Score Failed	Public.nyse	Score<75	data_quality@slack.com thomas@collibra.com brian@collibra.com
Job Failed	Public.nyse	FAILED	data_quality@slack.com kirk@collibra.com brian@collibra.com
Rule Failed	Public.nyse	accountRule > 0	thomas@collibra.com
Schedule Failed	Public.nyse	JOB DID NOT RUN	data_quality@slack.com kirk@collibra.com brian@collibra.com thomas@collibra.com ...

+ Create new alert Watch alert

- Score Failed
- Job Failed
- Rule Failed
- Schedule Failed

2022-03-22 Alert On

2022-03-22 2022-04-22 Alert On

Alert Off

Q/A

Wrap-up

- Thank you again for your participation!
- Your feedback is important to us
- Continue the collaboration on the Data Citizens Community [here](#)

Thank you

Integration Q&A

Data Intelligence Cloud Credentials

- Data Intelligence Cloud Base URL
- Basic Auth → Expanding to OAuth early next year
- Service Account
 - consumes author license
 - sysadmin & data steward roles

Update integration credentials ×

Credentials | Connections | Tenants | Dimensions

Service Type
Data Intelligence Cloud ▾

Service URL *
https://dq-meshon-two.collibra.com/

Authentication Mode
basic ▾

Service Account *
Admin ⌵

Password *
⌵

Password Cannot Be Empty

Cancel Next

Connections

- Auto-mapping based on name for schema, tables, and columns
 - can be manually overridden
 - "not configured" for no match
 - Views coming soon
 - Limited to JDBC Connections
→ File Connections on the 2024 roadmap
- Select Save and Continue

Update integration credentials

Credentials **Connections** Tenant Dimensions

Select Connection: Approved_Athena_UP | Enter DGC Database UUID: * 45b6b633-8ff7-4f02-bbaa-30e63e0c | Run

Selected DGC Database: APPROVED_ATHENA_UP>AwsDataCatalog

DQ Name	DGC Name
collibra_catalog_a530bbf2-bb30-446a- > a475-319ae7edce10_5e2e160b-5df7- 40b0-abd0-8305510e90ee_a1f7u2	Not Configured
> default	default
> a_rule_breaks	a_rule_breaks
> a_rule_multiple_breaks	a_rule_multiple_breaks
> a_surbhi_test	a_surbhi_test
> accounts	accounts


Cancel | Save And Continue

Tenants

- Available for Multi-Tenant DQ Customers → Single Tenant in backlog (2023.10)
- Will create a new community or use existing
- Defaults to “DQ Integration Community” (or tenant name)

Update integration credentials ✕

Credentials Connections **Tenants** Dimensions

 You have established an integration between this DQ tenant and DIC. Mapped tenant: DQ Integration Testing

Do you want to align your tenant to an existing DIC community?
 Yes No

Please select the name of the community it should be linked to.

DQ Integration Testing ▾

Cancel Next

Dimensions

- Custom Dimensions in DQ will be supported soon (2023.10)
- Will match on names
 - Can be overridden to match DIC

Update integration credentials ×

Credentials Connections Tenants **Dimensions**

DQ Dimensions	DIC Dimensions
ACCURACY	Select Dimension Accuracy
COMPLETENESS	Select Dimension Completeness
CONSISTENCY	Select Dimension Consistency
TIMELINESS	Select Dimension Integrity
UNIQUENESS	Select Dimension

Cancel Save

View Status in Data Intelligence Cloud

The screenshot shows the Data Intelligence Cloud interface. On the left, the 'ACTIVITIES' panel lists several synchronization tasks, all marked as 'Completed'. The main area displays an 'Import' dialog box with a 'Summary' section. The summary indicates that the import was completed without errors. Below the summary is a table showing the results of the import for various resources.

Resource	Added	Updated	Deleted
Communities	0	0	0
▶ Domains	2	0	0
▶ Assets	89	0	0
▶ Attributes	584	0	0
▶ Relations	205	0	0

- See Activities for synchronization status and number of assets inserted or updated
- If error occurs, detailed error messages are available via the import summary

View Quality Tab on Asset Page

The screenshot shows the Collibra Settings page. The 'Operating model' section is active, and the 'Data quality rules' tab is highlighted in a red box. A 'Create data quality rule' dialog is open, showing fields for Name, Description, Aggregation Path, and Categorisation, along with a table of metrics.

Create data quality rule

Name*
New DQ Rule

Description

Aggregation Path

Path*

Schema contains Table @ Table is represented by Data Asset @ Data Quality job is governed by Data Quality Rule @
Data Quality Rule executed by Data Quality Metric @

Construct the path through which you want to calculate your aggregated scores.

Categorisation*

classified by Data Quality Dimension

We can categorize all Rule elements based on a relation type. You will have the ability to see subcores on the asset types based on these categories.

Metrics

Metric	Description	Operation
Passing Fraction	The N of rows or entities that have passed the rule.	Average
Result	The result.	Logical AND
Rows Failed	The amount of rows that failed the rule.	Total
Rows Passed	The amount of rows that passed the rule.	Total

Add metric

Cancel Create

1. Add Data Quality Rules via the operating model section of your DIC Settings page
2. Create new data quality rules by defining aggregation paths relating the intended asset back to the DQ metric
3. To see the quality tab on the intended asset type, add the data quality rule to the global assignment of the asset type

Resources

DIC Integration Resources

Documentation site

- [Collibra DQ DIC Integration](#)
- [Configuring a new integration](#)

Professional Services and Coaching

- Customers are not expected to do this on their own
- Can use Coaching or PS
- Especially for migrating from existing DQ Connector
- See internal documentation of process

Sunset of existing DQ Connector

- 12-18 months

Thank you