

Informatica

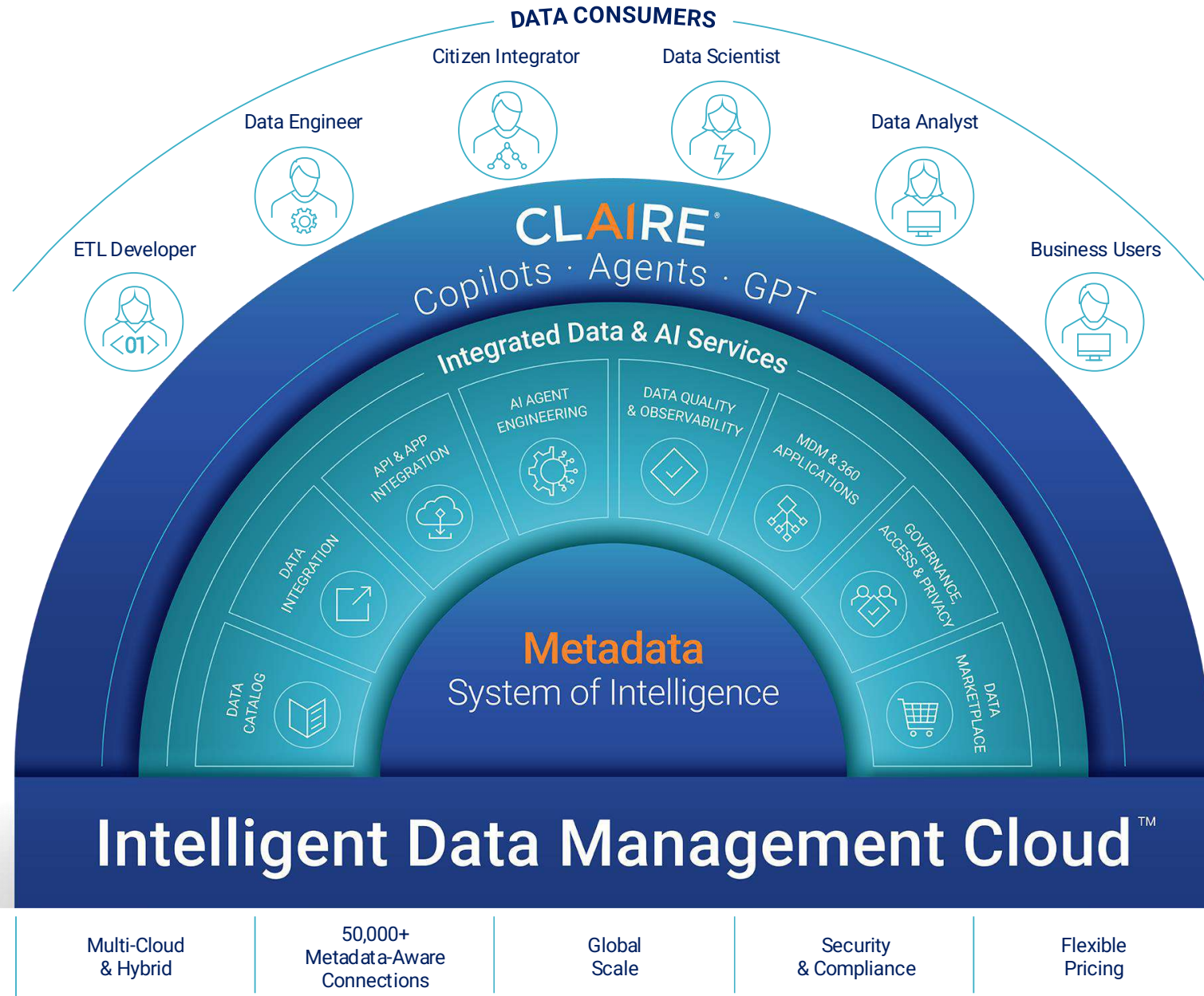
Data Management & Data Quality

2026-03-18

Sebastien Cognet – Data Governance strategic Engineer

**Where data
& AI come to** 

IDMC



Context - Why?

Quality

**I do not trust anymore
this report**

I want to ensure
specific reporting data sources
are exposing good data quality

- Profile
- Define Rule
- Apply

Quality Governance

**My domain is never on
data quality scope**

I want to generalize all data quality
best practices

- Profile "All"
- Standardize
- Monitor

Observability

**We are not reactive on
data issue**

I want to be alerted if something
unusual happen

- Regular check
- Deviation

AI Agent

**What if I want to
innovate?**

I want to let "anybody" benefit from
our data quality assets

- Suggest
- 100% no code
- Improve

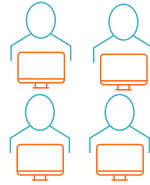
Solutions – How?

DEMO

Quality

- Profile
- Define Rule
- Apply

Users



AI



Quality Governance

- Profile "All"
- Standardize
- Monitor



Observability

- Regular check
- Deviation



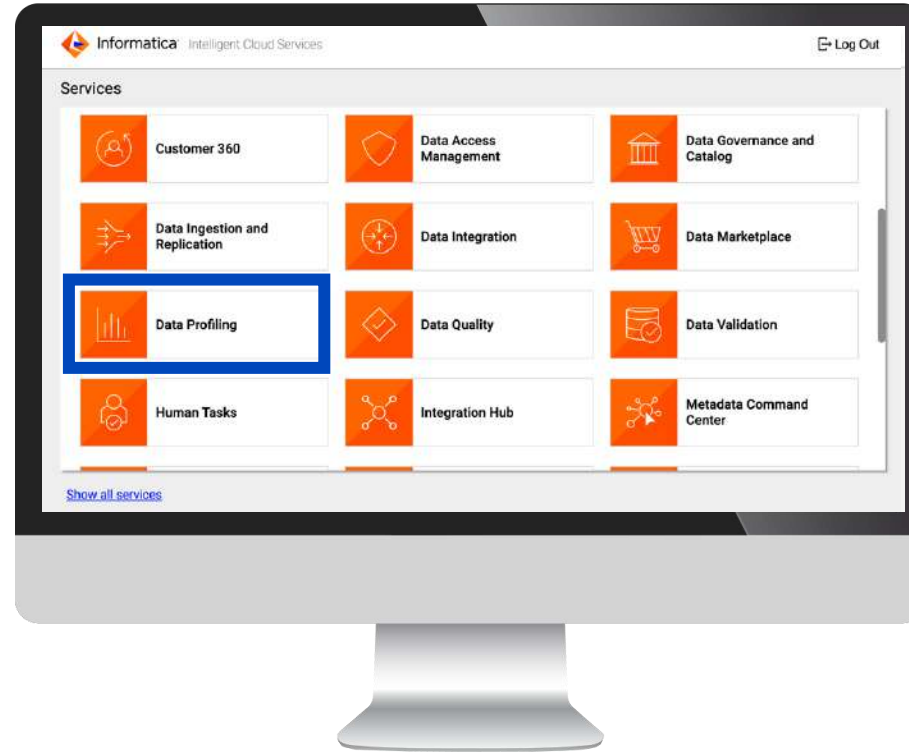
AI Agent

- Suggest
- 100% no code
- Improve



DEMO

Quality



Where data & AI come to 

LIVE DEMO

Profile_DataBricks_dim_customers - with scorecards

Run [More] [Close]

- Results
- Definition**
- Rules
- Metrics
- Schedule
- Insights

Asset Details

Name: * Profile_DataBricks_dim_customers - with scorecards

Description: Country Validity in CDGC - <https://cdgc.dm-us.informaticacloud.com/asset/b0d0510e-08e0-4cbc-a4b1-71c6dffc3554?type=RuleInstance&name=Validity%20Country>.

Location: [Cloud DQ Standard Demo - Oracle\@Profile](#)

Created by: unity-us-cust-retail-v2-admin

Last Updated: Oct 13, 2025, 3:45 PM

Used Storage (MB): 29.39

Source Details

Connection: DatabricksUnityEMEA (ODBC)

Source Object: databricks_dim_customers

Advanced Options

CLAIRE capabilities: Enabled ?

Profile Settings

Run profile on: All rows for complete analysis

Drill down: Yes

- Columns**
- Filters
- Data Preview

Select columns to run the profile on. (Selected 31 of 31)

Find [Search] [Sort] [Settings]

<input checked="" type="checkbox"/>	Columns	Type	Native Data Type	Precision	Scale	Nullable	Key
<input checked="" type="checkbox"/>	CustomerKey	HIPRECDECIMAL	decimal				
<input checked="" type="checkbox"/>	GeographyKey	HIPRECDECIMAL	decimal				
<input checked="" type="checkbox"/>	CustomerLabel	STRING	varchar				

Profile_DataBricks_dim_customers - with scorecards

Profile run 29 of 29 | 31 of 31 Columns | 14 of 14 Rules | 18871 Rows (All rows) | Oct 13, 2025, 5:29 PM

Run ... X

Results Definition Rules Metrics Schedule Insights

View: Columns and Rules with: All Statistics

Find [] [] [] [] []

Columns	Value Distribution	% Null	# Null	% Distinct	# Distinct	% Non-dist	# Non-dist	# Patterns	% of Top Pa	Minimum
LastName		0%	0	0.03%	5	99.97%	18866	2	97.94%	3
NameStyle		2.04%	385	43.74%	8254	54.22%	10232	2	97.96%	19
BirthDate		2.04%	385	0.02%	3	97.94%	18483	3	97.95%	1
MaritalStatus		99.97%	18865	0.02%	4	0.01%	2	5	99.97%	0
Gender		2.04%	385	0.04%	7	97.92%	18479	3	97.93%	1
EmailAddress		2.04%	385	97.95%	18485	0.01%	1	10	19.65%	22
YearlyIncome		0%	0	0.1%	18	99.9%	18853	3	86.31%	5
TotalChildren		2.04%	385	0.04%	7	97.92%	18479	3	97.95%	1

Details Rules

Trend

Type Trend % Cl

Null 0%

Distinct 0%

Non-distinct 0%

Data Types (2)

- Choose Profile Ru
- Compare Column
- Compare Profile R
- Scorecard Metrics
- Delete Profile Run
- Profile Jobs
- Export Profile Res
- Create Exception

Data Preview (10 Rows)

Show only profiled columns

Queries [] [] []

	CustomerKey	GeographyKey	CustomerLabel	Title	FirstName	MiddleName	LastName	NameStyle	BirthDate	MaritalStatus	Suffix
1	18431	595	29430	Miss	Donna	H	Goel	false	2021-11-04 00:00:00	M	Dr
2	17407	503	28406	Msieur	Gerald	Jr	Srini	true	1963-03-23 00:00:00	M	
3	18427	529	29426	Mdm	Neil		Gomez	false	1974-10-02 00:00:00	S	
4	17403	529	28402	Mdme	Kristin	R	Kumar	true	1965-05-26 00:00:00	S	
5	18423	564	29422		Ian	C	Adams	false	1976-08-19 00:00:00	S	

Profile_DataBricks_dim_customers - with scorecards

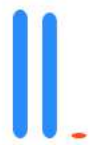
Profile run 29 of 29 | 31 of 31 Columns | 14 of 14 Rules | 18871 Rows (All rows) | Oct 13, 2025, 5:29 PM

Run ⋮ ✕

Results | Definition | Rules | Metrics | Schedule | Insights

Column: Gender

7 distinct values (3 non-unique values, 4 unique values)



Back to Summary



Sort By: Frequency

Export values to dictionary.

Details | Rules



#	Value	Frequency	Percentage	Length	
<input checked="" type="checkbox"/>	1	M	9349	49.54%	1
<input checked="" type="checkbox"/>	2	F	9131	48.39%	1
<input checked="" type="checkbox"/>	3	NULL	385	2.04%	
<input type="checkbox"/>	4	Femme	2	0.01%	5
<input type="checkbox"/>	5	Fem	1	0.01%	3

- Data Types (2) >
- Patterns (3) >
- Most Frequent Values (5) >

Data Preview (10 Rows)

Show only profiled columns

Queries

Filter: Gender: Value: M

Clear All Apply

Gender	EmailAddress	YearlyIncome	TotalChildren	NumberChildrenAtHorn	Education	Occupation	HouseOwnerFlag	NumberCarsOwned	AddressLine1	AddressLine2
Femme	donna17@adventu...	30000.0000	0	0	Partial College	Clerical	0	1	Galeriestr 6818	

- New...
- Explore
- My Jobs
- My Import/Export...
- Profile_DataBrick...
- Validate_Country**

Validate_Country

Convert to Advanced

Validate [More] [Close]

Definition **Configuration**

▶ Design [Icons]

Properties: Validate_Country

General

Rule logic definition

Inputs

		Input	Operator	Condition		Action
	if	Input_Country	is within	Dictionary(countries_infa)	then	Valid
Rule Logic	or	if	No rule statement is valid		then	INVALID

Test

- + New...
- 🔍 Explore
- 📁 My Jobs
- 📁 My Import/Export...
- 📊 Profile_DataBrick...
- 🔗 rs_nb_children_c...

rs_nb_children_consistency_advanced Advanced

[Edit](#) [Validate](#) [...](#) [X](#)

Definition [Configuration](#)

Rule Editor

[Manage Inputs and Outputs](#) ✓ 🔍 📄 📄 📄

```
1 F TotalChildren = null AND NumberChildrenAtHome = null THEN nb_children_consistency = 'VALID'  
2 ELSEIF TotalChildren = null AND NumberChildrenAtHome != null THEN nb_children_consistency = 'INVALID'  
3 ELSEIF TotalChildren < NumberChildrenAtHome THEN nb_children_consistency = 'INVALID'  
4 ELSE nb_children_consistency = 'VALID' ENDIF  
5
```

Test Data ⓘ

Runtime Environment: Choose an option: ▾

Test



Find



< Collapse

Profile_DataBricks_dim_customers - with scorecards

Run [More] [Close]

Results Definition Rules Metrics Schedule **Insights**

View: All Insight Types in: All Columns

Find [Search Icon] [Sort Icon] [Settings Icon]

<input type="checkbox"/>		Score	Insight Type	Columns	Status
<input type="checkbox"/>	Column includes one or more null, blank, or empty values or values that contain only z...	Low	Completeness Check	HouseOwnerFlag	
<input type="checkbox"/>	Column includes one or more null, blank, or empty values or values that contain only z...	Low	Completeness Check	LastName	
<input type="checkbox"/>	the column has a high standard deviation.	Medium	Column Length Deviation	LastName	
<input type="checkbox"/>	umn has a high standard deviation.	Low	Column Token Deviation	LastName	
<input type="checkbox"/>	aracters	Low	Special Characters	LastName	
<input type="checkbox"/>	Column includes one or more null, blank, or empty values or values that contain only z...	Low	Completeness Check	EmailAddress	
<input type="checkbox"/>	the column has a high standard deviation.	High	Column Length Deviation	EmailAddress	
<input type="checkbox"/>	Column includes one or more null, blank, or empty values or values that contain only z...	Low	Completeness Check	Education	
<input type="checkbox"/>	the column has a high standard deviation.	Low	Column Length Deviation	Education	
<input type="checkbox"/>	spelling.	High	Spelling Analysis	Education	
<input type="checkbox"/>	umn has a high standard deviation.	High	Column Token Deviation	Education	
<input type="checkbox"/>	the column has a high standard deviation.	Medium	Column Length Deviation	AddressLine2	
<input type="checkbox"/>	The number of tokens in the column has a high standard deviation.	Medium	Column Token Deviation	AddressLine2	
<input type="checkbox"/>	The data may contain special characters	High	Special Characters	AddressLine2	
<input type="checkbox"/>	More than 70% increase on the number of distinct values in the column between the current profile run and the previ...	Low	Distinct Variation	Suffix	
<input type="checkbox"/>	The number of tokens in the column has a high standard deviation.	High	Column Token Deviation	Suffix	
<input type="checkbox"/>	The data may contain special characters	High	Special Characters	Suffix	
<input type="checkbox"/>	The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	Title	
<input type="checkbox"/>	More than 70% increase on the number of distinct values in the column between the current profile run and the previ...	Low	Distinct Variation	Title	
<input type="checkbox"/>	The number of tokens in the column has a high standard deviation.	Low	Column Token Deviation	Title	
<input type="checkbox"/>	The data may contain special characters	High	Special Characters	Title	
<input type="checkbox"/>	Data appears incomplete. The column includes one or more null, blank, or empty values or values that contain only z...	Low	Completeness Check	BirthDate	
<input type="checkbox"/>	Data appears incomplete. The column includes one or more null, blank, or empty values or values that contain only z...	Low	Completeness Check	MaritalStatus	

- All Insight Types
- Completeness Check
- Column Token Deviation
- Spelling Analysis
- Column Length Deviation
- Number Value Distribution
- Special Characters
- Completeness Variation
- Distinct Variation
- Uniqueness Check

- New...
- Home
- Explore
- Bundles
- My Jobs
- Templates
- My Import/Export...
- M_Cleanse_DeDu...
- M_Cleanse_Valid...
- M_Parse_TV_Des...

M_Parse_TV_Descriptions - for data preview demo Valid

Switch to Advanced... Save Run

Design

- Source
- Target
- Access Policy



Properties Preview Parse_TV_Desc

Row

TV_BRAND	TYPE	SCREEN_RES	SCREEN_REFRESH	PROD_CODE	PRICE	Overflow	ToUpperCase	DESCRIPTION
RCA	LCD	1080P	60HZ	22LA45RQ		-- 22" CLASS - - - ...	RCA - 22LA45R...	RCA - 22LA45RQ -- 22" Class - LCD - 1080p - 60Hz - HDTV
LG	LED	720P	60HZ	22LV2500		-- 22" CLASS - - - ...	LG - 22LV2500 - ...	LG - 22LV2500 RCA - 22LA45RQ -- 22" Class - LCD - 1080p - 60Hz - HDTV
COBY	LCD	720P		TF-TV2225		-- TF-TV2225 22...	COBY - TF-TV22...	Coby - TF-TV2225 -- TF-TV2225 22" 720p LCD TV - 16:9 - HDTV
COBY		720P		TFDVD2295		-- TFDVD2295 2...	COBY - TFDVD2...	Coby - TFDVD2295 -- TFDVD2295 22" TV/DVD Combo - 16:9 - 1366 x 768 - 720p
COBY	LED	1080P	60HZ	LED-TV2326	\$259.99	KIT - 23" CLAS...	COBY - LED-TV2...	Coby - LED-TV2326 KIT - 23" Class - LED - 1080p - 60Hz - HDTV \$259.99
COBY	LCD	1080P	60HZ	TF-DVD2395	\$299.99	KIT - 23" CLAS...	COBY - TF-DVD...	Coby - TF-DVD2395 KIT - 23" Class - LCD - 1080p - 60Hz - HDTV DVD Combo \$299.99
COBY	LED	1080P	60HZ	LEDTV2326		-- 23" CLASS - - - ...	COBY - LEDTV2...	Coby - LEDTV2326 -- 23" Class - LED - 1080p - 60Hz - HDTV
RCA	LED	1080P	60HZ	LED24A45RQ	\$159.99	-- REFURBISHE...	RCA - LED24A4...	RCA - LED24A45RQ -- Refurbished 24" Class - LED - 1080p - 60Hz - HDTV \$159.99
INSIGNIA?	LCD	1080P	60HZ	NS-24L240A13	\$169.99	-- 24" CLASS / / ...	INSIGNIA? - NS...	Insignia? - NS-24L240A13 -- 24" Class / LCD / 1080p / 60Hz / HDTV \$169.99

1 - 100 of 100

< 1 of 1 >

Items per Page 100

- New...
- Home
- Explore
- My Jobs
- My Import/Export...

Home

Assets by Type and Status

Asset Type	Invalid/Valid	Invalid	Valid	Total
Rule Specification		6	627	633
Dictionary		0	282	282
Verifier		0	26	26
Cleanse		0	208	208
Deduplicate		0	24	24
Parse		1	97	98

Runtime Environments

- Informatica Cloud Hosted Ag... Up and Running
- Unity AWS Stopped
- Unity Local Cluster Stopped
- azlicsetl02d No Secure Agent
- cloudagent-sandbox-unityv2-1 No Secure Agent
- Data_Quality_Runtime_Enviroi No Secure Agent
- cdgcworkshop-base.infaclo... Stopped
- presales-managed-sa-group Up and Running

Recent Assets

[View All](#)

Name	Location	Type	Last Updated	Status
Validate_Country	Cloud DQ Standard Demo - Oracle\rules_kpis	Rule Specification	Mar 28, 2024, 11:01 AM	Valid
rs_COUNTRY_CompletenessCheck_2	Cloud DQ Standard Demo - DataBricks\@P...	Rule Specification	Feb 12, 2025, 10:44 AM	Valid
rs_check_input_is_not_null	Cloud DQ Standard Demo - DataBricks\Co...	Rule Specification	Feb 12, 2025, 10:44 AM	Valid
rs_completeness	Cloud DQ Standard Demo - Oracle\rules_kpis	Rule Specification	Mar 28, 2024, 10:52 AM	Valid
rs_completeness	Cloud Data Governance	Rule Specification	Feb 13, 2026, 9:38 AM	Valid
rs_gender_validity	Cloud DQ Standard Demo - Oracle\rules_kpis	Rule Specification	Mar 27, 2024, 12:00 PM	Valid
rs_COUNTRY_SpellingAnalysis_2	Cloud DQ Standard Demo - Oracle\@Profile	Rule Specification	Mar 27, 2024, 11:15 AM	Valid

< Collapse

- New...
- Home
- Browse
- Data Access
- Saved Searches
- Profile_DataBrick...
- Validity Country

Top Level

Validity Country

DATA QUALITY RULE OCCURRENCE

- Overview
- Score
- Stakeholders
- Tickets
- History

OVERALL RATING: ★★★★★(1) Add Rating

LIFECYCLE: PUBLISHED

LAST UPDATED: Oct 13, 2025, 8:33 AM

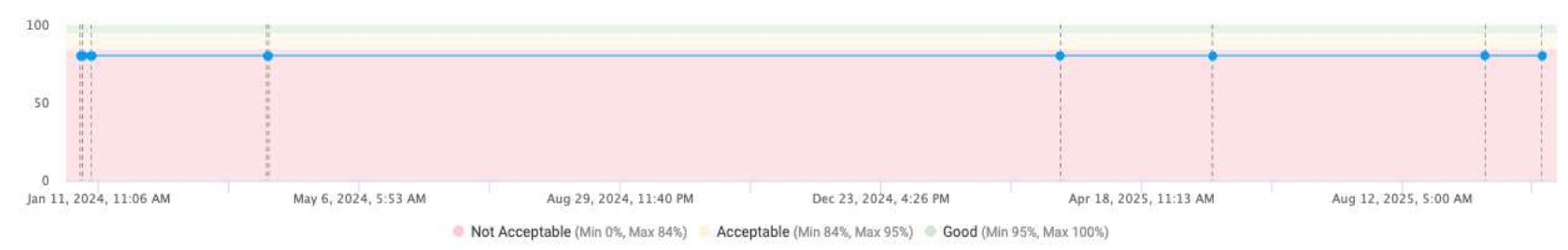
Latest Data Quality Score



No Change

Score Status: ● Not Acceptable
 Total Rows: 18871
 Failed Rows: 3766

Score Trend



Total Runs (19)

Data Preview

Date	Latest Score	Total Rows
Oct 13, 2025, 8:33 AM	● 80.04%	18871
Sep 18, 2025, 5:47 AM	● 80.04%	18871
May 20, 2025, 5:47 AM	● 80.05%	18871
May 20, 2025, 5:31 AM	● 80.05%	18871
Mar 13, 2025, 2:04 PM	● 80.04%	18871
Mar 27, 2024, 4:05 AM	● 80.06%	18870
Mar 27, 2024, 4:00 AM	● 80.06%	18870
Mar 27, 2024, 3:53 AM	● 80.06%	18870
Mar 27, 2024, 3:21 AM	● 80.06%	18870
Mar 26, 2024, 7:24 AM	● 80.06%	18870

Preview of Failed Rows

From: admin@informaticacloud.com
Sent on: Tuesday, September 10, 2024 8:23:12 PM
To: fboulbel@informatica.com
Subject: Exception task 'Except_DataBricks_dim_customers_score-2' completed successfully

Rename file



Organization: Unity CDGC Generic Customer Re...
Date: Sep 10, 2024, 1:23 PM

Hi,
This email summarizes an exception task job that ran in the above organization.

Job Summary: Exception task ran **successfully: Identified 2.0% exceptions**

Exception Task Job: [Except_DataBricks_dim_customers_score-2](#)
Start Date: Sep 10, 2024, 1:22 PM
End Date: Sep 10, 2024, 1:23 PM
Started By: unity-us-cust-retail-v2-admin
Status: **Success**

Source Rows: Indicates the number of rows in the data source that the exception task job read.	18870
Exception Rows: Indicates the number of rows in which the exception task job found data quality issues.	386
Number of Issues: Indicates the total number of data quality issues that the exception task job found in the exception rows.	771

Percentage of Exceptions Found:

2.0%

Download Exception Data: [Except_DataBricks_dim_customers_score-2](#)

Counts per priority:

 **Critical**
0

 **Major**
0

 **Minor**
771

Top 10 Issues

Shows the top 10 issues, each with a count of one or more

Description	Count
should not be empty	770

DQ Exception task email notification sample



al.

- New...
- Home
- Browse
- Data Access
- Saved Searches

Delta Lake AdventureWorks / unitycatalogospn / adworksforecast / databricks_dim_customers /

Country

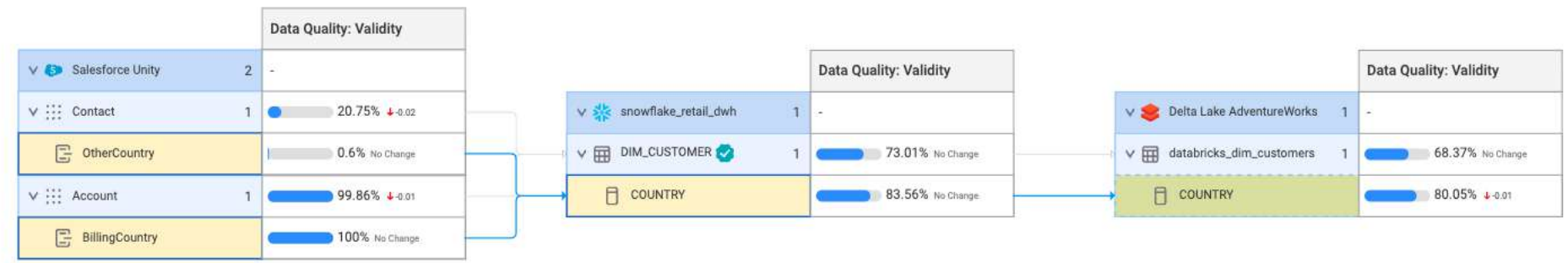
COLUMN / COUNTRY

OPEN TICKETS: 1 (1 task overdue!)
SENSITIVITY: PUBLIC
LIFECYCLE: PUBLISHED
LAST UPDATED: Sep 18, 2025, 6:27 AM

- Overview
- Lineage
- Relationships
- Data Quality
- Data Observability
- Reference Data
- Stakeholders
- Attributes
- Tickets
- History

View in the context of: databricks_dim_cust...
Overlay Selected: Certification, Data Qual

Find Assets



- + New...
- Home
- Explore
- Monitor
- Configure
- Customize
- Access Control

Delta Lake Adven...

Delta Lake AdventureWorks

< Back Next > Save Copy Run X

- 1 Registration
- 2 Configuration
- 3 Associations
- 4 Schedule

- ✓ Metadata Extraction
- ✓ Lineage Discovery
- ✓ Data Profiling and Quality
- Data Observability
- ✓ Data Classification
- ✓ Relationship Discovery
- ✓ Glossary Association

Data Profiling

Enable Data Profiling:

Connection and Runtime

Runtime Environment: unity-us-retail-v2-New X v

Filters

Specify data profiling filters: ? No Yes

> Show examples

Exclude Metadata v Catalog.Schema v Named v unitycatalogspn.adworksforcast.v_mo... x unitycatalogspn.adworksforcast.daily_... x Select + X

Parameters

Modes of Run: * ? Keep Signatures Only Keep Signatures and Values

Profiling Scope: * ? Incremental Full

Sampling Type: * ? Custom Query v

- + New...
- Home
- Explore
- Monitor
- Configure
- Customize
- Access Control

Delta Lake Adven...

Delta Lake AdventureWorks

< Back Next > Save Copy Run X

- 1 Registration
- 2 Configuration
- 3 Associations
- 4 Schedule

- ✓ Metadata Extraction
- ✓ Lineage Discovery
- ✓ Data Profiling and Quality
- Data Observability
- ✓ Data Classification
- ✓ Relationship Discovery
- ✓ Glossary Association

Data Profiling

Data Quality

Enable Data Quality:

Use Data Profiling Parameters

Connection and Runtime

Runtime Environment: unity-us-retail-v2-New

Parameters

Data Quality Rule Automation: * ? Yes No

Apply on all Data Elements

Cache Result: * ?

Agent Cache

Run Rule Occurrence Frequency: * ? Yes No

Results

Quality

*I do not trust
anymore this
report*

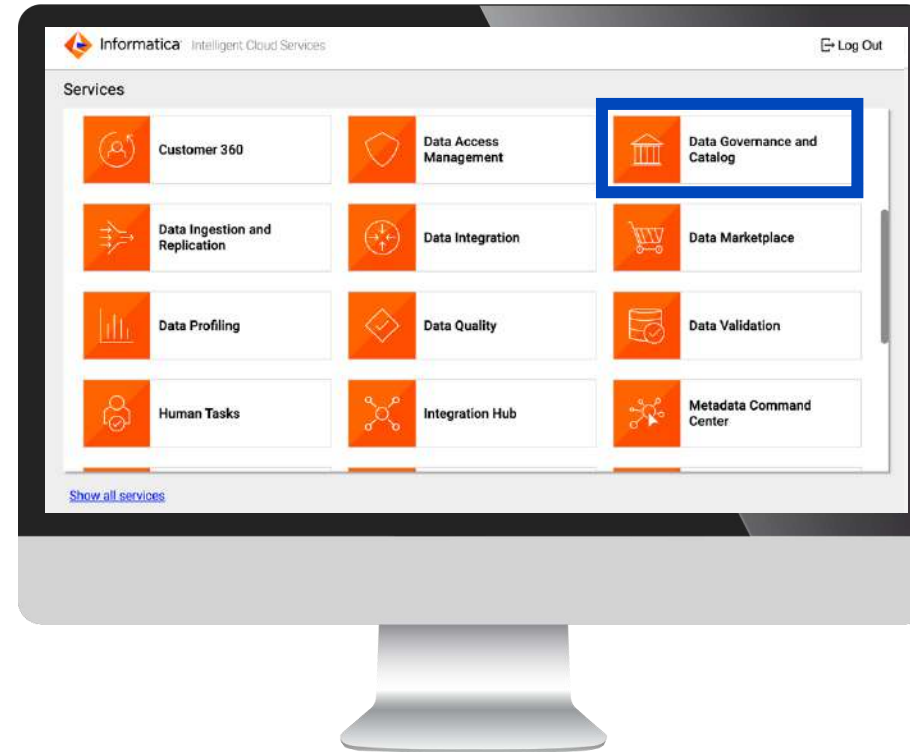
Consumers trust BI reports

Solutions

- Profiling
- Cleansing
- Matching & Parsing

DEMO

Quality Governance



2

Quality Governance

My domain is never on data quality scope

I want to generalize all data quality best practices

- Profile "All"
- Standardize
- Monitor

- Profile "All"
- Standardize
- Monitor



LIVE DEMO

- New...
- Home
- Browse
- Data Access
- Saved Searches
- Validity Country

Global Data Quality and Observability

Shared with: All Users

Governed Technical Assets

107.4K Column (Relational Database ...)	5.6K Table (Relational Database ...)
427 Catalog Source	424 Hierarchical File (Hierarchical files)
242 Schema (Relational Database ...)	77 Mapping task (Informatica Intellige...)

Governed Business Assets

304 Business Term	209 Policy
186 Process	80 Subdomain
56 AI Model	44 Metric
37 System	13 Domain

Global DQ Score

Group By: Threshold Result

HR DQ Score

Group By: Threshold Result

Data Lake Observability

▲ 1/6 ▼

File System Observability

▲ 1/3 ▼

Browse

- Glossary**
- Data Catalog
- System & Dataset
- Process
- Policy
- Project
- AI Model
- Business Area
- Legal Entity
- Geography
- Regulation
- AI System

Glossary (13)

Name ↑	Type	Description	Last Updated	Stakeholders
AI Model Performance Met... (8)	Domain	Comprehensive domain covering all AI ...	Nov 26, 2025, 4:4...	
Demo Description Generati... (6)	Domain	Demo Description Generation Using CL...	May 6, 2025, 1:28...	
Demo Dom Full Permission (1)	Domain	-	May 22, 2025, 12:...	-
Finance (9)	Domain	Financial operations domain encompass...	Nov 26, 2025, 4:4...	
Géographique et Statistiqu... (2)	Domain	Ce domaine se focalise sur les effets ré...	Apr 11, 2025, 8:2...	-
Human Resources Data (4)	Domain	-	Dec 6, 2024, 2:44 ...	-
Juridique et Médiation (2)	Domain	Ce domaine porte sur les options de rés...	Mar 21, 2025, 10:...	-
Litige et Gestion (2)	Domain	Ce domaine couvre la gestion des litige...	Mar 21, 2025, 10:...	-
Marketing (1)	Domain	Marketing governed assets	Dec 6, 2024, 3:57 ...	-
Party Data (7)	Domain	Data related to customers of the organi...	Mar 21, 2025, 10:...	
Product Data (9)	Domain	Data relted to products sold by the orga...	Dec 6, 2024, 12:3...	
Profil Demandeur (2)	Domain	Ce domaine met l'accent sur les inform...	Mar 21, 2025, 10:...	-
Sales Data (3)	Domain	Terms related to retail sales	Dec 6, 2024, 12:1	

AI Model Performa... DOMAIN

- Overview**
- Related Assets (9)

Description:
 Comprehensive domain covering all AI model performance evaluation metrics including classification accuracy, regression error measures, and predictive performance
[Show More](#)

PUBLISHED

Hierarchy: ?
 Top Level

Last Updated:
 Nov 26, 2025, 4:42 PM

Reference ID: ?
 GEN-DOM-AI-Model-Performance

Data Quality: ?
 -

Stakeholders: ?

Fouad B
 Governance Owner

Governance ...
 Governance Adm...

Related Assets: ?

- New...
- Home
- Browse**
- Data Access
- Saved Searches

Browse

- Glossary
- Data Catalog**
- System & Dataset
- Process
- Policy
- Project
- AI Model
- Business Area
- Legal Entity
- Geography
- Regulation
- AI System

Data Catalog (181)

- Select Types
- Alteryx (1)
 - Amazon Athena (1)
 - Amazon Redshift (1)
 - Amazon S3 (13)
 - AWS Glue (1)
 - Custom Lineage (5)
 - Databricks (15)
 - Databricks (15)

All Assets Find

Name	Location	Last Updated	Stakeholders
> ACME CRM (1)		Jul 24, 2025, 10:0...	
> ACME CRM BACKUP (1)		Jul 24, 2025, 10:0...	
> ACME CRM COPY (1)		Jul 24, 2025, 10:0...	
> ACME PIM (1)		May 6, 2025, 12:5...	-
> ADF Ad Works (5)	data Factory etl to move ad work...	May 6, 2025, 1:22...	-
> ADF Calling DataBricks (4)	ing databricks & passing source...	Sep 17, 2024, 6:5...	-
> adls 2 enablement HR Team	enablement HR Team	Feb 22, 2026, 9:3...	-
> adls marking enablement (1)	Microsoft Azure Data L... data lake storage for enablement purpo...	Feb 25, 2026, 9:3...	JA
> adls r&D enablement (1)	Microsoft Azure Data L... data lake storage for enablement purpo...	Feb 21, 2026, 9:3...	JA
> adls_partition_custom_YYY... (1)	Microsoft Azure Data L...	Jun 19, 2025, 4:4...	-
> adls_partition_with_multiple... (1)	Microsoft Azure Data L...	Jun 19, 2025, 5:0...	-

1 - 25 of 181 Items

ACME CRM ORACLE CATALOG SOURCE

Overview Related Assets (1)

Description:

PUBLISHED

Hierarchy: Top Level

Last Updated: Jul 24, 2025, 10:08 AM

Stakeholders:

- MAC Corpor... cdgc demo user
- Governance ... cdgc demo user

Show More

Related Assets:



- New...
- Home
- Browse
- Data Access
- Saved Searches
- Profile_DataBrick...
- Validity Country
- COUNTRY
- Delta Lake Adven...
- enriched_sales_f...

Delta Lake AdventureWorks / unitycatalogspn / adworksforecast /

Create Data Set

(Add Business Name)

OVERALL RATING: ★★★★★ (1) Add Rating

LIFECYCLE: PUBLISHED

LAST UPDATED: Jul 31, 2025, 11:18 PM

Overview **Contains** Lineage Relationships Data Quality Data Observability Reference Data Stakeholders Attributes Tickets History

Data Elements (14)

<input type="checkbox"/>	Business Name ↑	Name	Description	Glossaries	Sensitivity	Classifications	Classification Cate...	Null Distinct Non-Distinct %	CLA
<input type="checkbox"/>	Country	SalesTerritoryCou...	Name of the counr...	Country	PUBLIC	Country	Address Inform...	<div style="width: 100%; height: 10px; background-color: #2e8b57;"></div>	...
<input type="checkbox"/>	Customer Email Address	emailAddress	A customer email a...	Country	CONFIDENTIAL	Address +1	Address Inf...	+2 <div style="width: 80%; height: 10px; background-color: #ff4500;"></div>	
<input type="checkbox"/>	Email Address	EmailAddress	The email address t...	Email Add... +1	CONFIDENTIAL	Address +1	Address Inf...	+2 <div style="width: 80%; height: 10px; background-color: #ff4500;"></div>	
<input type="checkbox"/>	English Education	EnglishEducation	English Education r...		-			<div style="width: 80%; height: 10px; background-color: #ff4500;"></div>	
<input type="checkbox"/>	English Occupation	EnglishOccupation	English Occupation...		-			<div style="width: 80%; height: 10px; background-color: #ff4500;"></div>	
<input type="checkbox"/>	English Product Catego...	EnglishProductCa...	English Product Ca...		-			<div style="width: 100%; height: 10px; background-color: #2e8b57;"></div>	
<input type="checkbox"/>	English Product Name	EnglishProductNa...	English Product Na...		-			<div style="width: 100%; height: 10px; background-color: #2e8b57;"></div>	
<input type="checkbox"/>	English Product Subcat...	EnglishProductSu...	English Product Su...		-			<div style="width: 100%; height: 10px; background-color: #2e8b57;"></div>	
<input type="checkbox"/>	Gender	Gender	One's identity as fe...	Gender	-			<div style="width: 80%; height: 10px; background-color: #ff4500;"></div>	
<input type="checkbox"/>	Sales Amount	SalesAmount	total purchase amo...	Sales Am...	-			<div style="width: 80%; height: 10px; background-color: #2e8b57;"></div>	
<input type="checkbox"/>	Sales Category	sales_category	Sales Category ide...		-			<div style="width: 100%; height: 10px; background-color: #2e8b57;"></div>	
<input type="checkbox"/>	Sales Territory Region	SalesTerritoryRegi...	Sales Territory Regi...		-			<div style="width: 80%; height: 10px; background-color: #2e8b57;"></div>	
<input type="checkbox"/>	Ship Date	ShipDate	Ship Date represen...		-			<div style="width: 80%; height: 10px; background-color: #2e8b57;"></div>	
<input type="checkbox"/>	Transaction Date	OrderDate	The date a transact...	Transacti...	-			<div style="width: 80%; height: 10px; background-color: #2e8b57;"></div>	

Country
COLUMN SALESTERRITORYCOUNTRY

Overview Related Assets (20)

Hierarchy: Delta Lake AdventureWo... / ... / enriched_sales_forc...

Related Data Products

This asset is linked only to unpublished Data Products in Data Marketplace. For more information, contact the stakeholder of the asset.

Asset Groups: MAC Marketing, MAC Sales, MAC Corporate



Data Observability Events:



Top Values:
United States
Australia
United Kingdom

Data Types:
Documented: string
Inferred: String

- New...
- Home
- Browse
- Data Access
- Saved Searches
- Profile_DataBrick...
- Validity Country
- COUNTRY

Delta Lake AdventureWorks / unitycatalogspn / adworksforecast / databricks_dim_customers /

Country

COLUMN COUNTRY

- Overview
- Lineage
- Relationships
- Data Quality
- Data Observability
- Reference Data
- Stakeholders
- Attributes
- Tickets
- History

Visual representation of the data that leads into and out of the asset that is open.

Description

Name of the country, part of an address. The list of valid countries and their corresponding ISO Code is defined in [this reference table](#) which could be enriched and customized by dictionary owner.

iso_full_country_name_infa

Definition Configuration

Show more

Technical Description

-

Glossaries

Accepted (1) Country (100%)

CLAIRE Recommendations (0) No recommendations for now.

Declined (0) No declined Glossary assets.

Data Element Classifications

Accepted (1) Country

Declined (0) No declined Classification assets.

OPEN TICKETS 1 (Task overdue) ! SENSITIVITY PUBLIC LIFECYCLE PUBLISHED LAST UPDATED Sep 18, 2025, 6:27 AM

Stakeholders

- Regulatory Office Governance Administrator
- AI IT Team NIST - IT Team
- Customer Success Governance Administrator

View All

Asset Groups (3)

MAC Marketing MAC Sales MAC Corporate

Data Quality

86.7%

Rules Against Target

Good (1) Acceptable (2) Not Acceptable (1)

View Details

Open Data Observability Events

- New...
- Home
- Browse
- Data Access
- Saved Searches
- Profile_DataBrick...
- Validity Country
- COUNTRY
- Delta Lake Adven...
- enriched_sales_f...
- Country

Party Data / Party Data - Core Entities /

Country

BUSINESS TERM

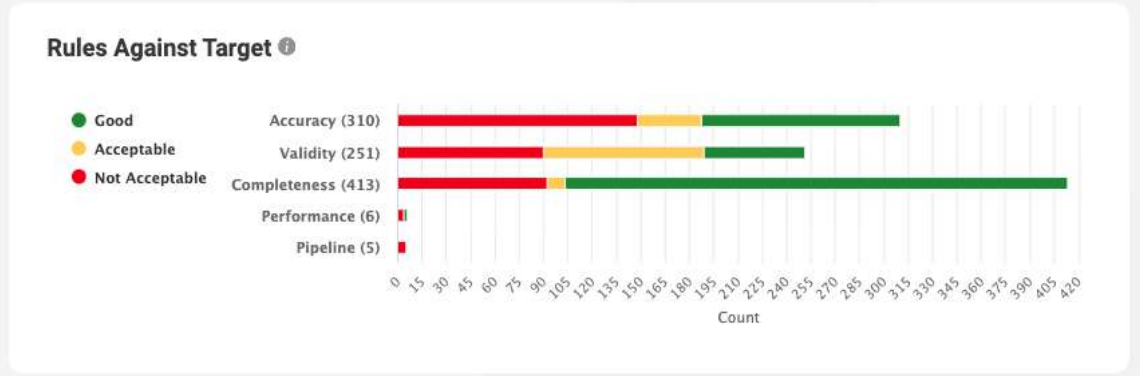
OVERALL RATING: ★★★★★ (1) Add Rating

CRITICAL DATA ELEMENT: NO

LIFECYCLE: PUBLISHED

LAST UPDATED: May 22, 2025, 12:20 AM

- Overview
- Hierarchy
- Relationships
- Data Quality**
- Stakeholders
- Tickets
- History



Rule Occurrences (985)

The rule occurrences that are visible to you are determined by the access policies assigned to you.

Filter: Status: Good, Not Acceptable, Acc... Add Filter

Date	Name	Dimension	Latest Score	Total Rows	Failed Rows	Primary Data Element	Primary Data Element Hierarchy	Secondary Data Element	Description
Oct 13, 2025, 8:3...	Validity Country	Validity	80.04%	18.9K	3.8K	COUNTRY	Delta Lake A... / ... / databricks_d...		Validat
Oct 13, 2025, 1:0...	Auto-MailingCountry-Performance-Ema...	Performance	98.9%	1K	11	MailingCountry	Salesforce Un... / SF-Un... / Cont...		it shou
Oct 2, 2025, 2:59 ...	Auto-OtherCountry-Completeness-Cou...	Completeness	0.6%	1K	994	OtherCountry	Salesforce Un... / SF-Un... / Cont...		Countr
Oct 2, 2025, 2:59 ...	Auto-MailingCountry-Validity-Country - ...	Validity	98.9%	1K	11	MailingCountry	Salesforce Un... / SF-Un... / Cont...		country
Oct 2, 2025, 2:59 ...	Auto-MailingCountry-Validity-Email Add...	Validity	0%	1K	1K	MailingCountry	Salesforce Un... / SF-Un... / Cont...		Email A
Oct 2, 2025, 2:59 ...	Auto-MailingCountry-Completeness-E...	Completeness	98.9%	1K	11	MailingCountry	Salesforce Un... / SF-Un... / Cont...		Email A
Oct 2, 2025, 2:59 ...	Auto-OtherCountry-Validity-Country - Va...	Validity	0.6%	1K	994	OtherCountry	Salesforce Un... / SF-Un... / Cont...		country

- New...
- Home
- Browse**
- Data Access
- Saved Searches
- Profile_DataBrick...
- Validity Country
- COUNTRY
- Delta Lake Adven...
- enriched_sales_f...
- Country

Browse

- Glossary**
- Data Catalog
- System
- Process
- Policy
- Project
- AI Model
- AI System
- Business Area
- Legal Entity
- Geography
- Regulation

Glossary (12)

Name ↑	Type	Description	Last Updated	S
> Litige et Gestion (2)	Domain	Ce domaine couvre la gestion des litige...	Mar 21, 2025, 2:2...	-
> Marketing (1)	Domain	Marketing governed assets	Dec 6, 2024, 6:57 ...	-
> Party Data (7)	Domain	Data related to customers of the organi...	Mar 21, 2025, 2:3...	F
▼ Product Data (9)	Domain	Data relted to products sold by the orga...	Dec 6, 2024, 3:38 ...	G
> Product Classifications (6)	Subdomain	Product Classifications are hierarchical ...	May 6, 2025, 2:21...	-
> Product Data - Core Entities (3)	Subdomain	Core Product related entities	May 31, 2022, 3:0...	-
> Product Data - Other (2)	Subdomain	Other product related data not part of a...	May 31, 2022, 3:0...	-
> Product Descriptors (8)	Subdomain	Data elements describing products	May 31, 2022, 3:0...	-
> Product Hierarchies (1)	Subdomain	Hierarchical product data	May 31, 2022, 3:0...	-
> Product Identifiers (2)	Subdomain	Data Identifying products	May 31, 2022, 3:0...	-
> Product Performance Data (15)	Subdomain	Performance related product data	May 31, 2022, 3:0...	-
Product Risk Data	Subdomain	Risk data related to products	May 31, 2022, 3:0...	-
System Date	Business Term	This is the system date derived from the...	Feb 18, 2025, 8:3...	-
> Profil Demandeur (2)	Domain	Ce domaine met l'accent sur les inform...	Mar 21, 2025, 2:2...	-
> Sales Data (3)	Domain	Terms related to retail sales	Dec 6, 2024, 3:15 ...	G

Product Data DOMAIN

Overview Related Assets (11)

Description:
Data relted to products sold by the organisation. V2

PUBLISHED

Hierarchy:
Top Level

Asset Groups:
MAC Corporate MAC Producter

Last Updated:
Dec 6, 2024, 3:38 AM

Reference ID:
DOM-4



No Change
Oct 1, 2025, 7:43 AM

Stakeholders:
Governance Administrators
Governance Administrator

Related Assets:



Results

Observability

*We are not
reactive on data
issue*

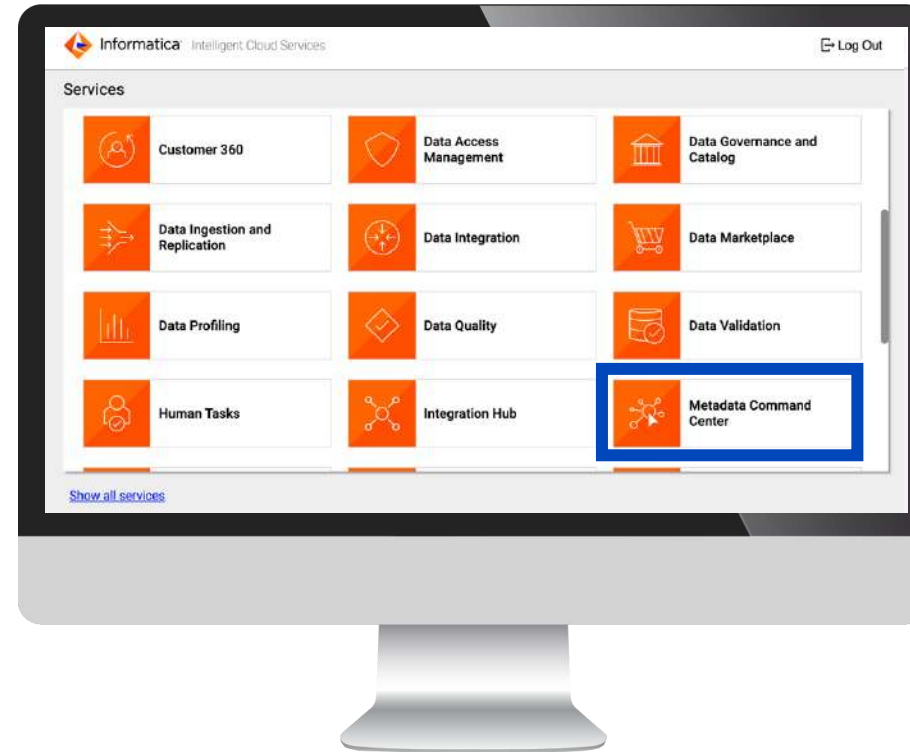
You react to change

Solutions

- Alerting
- Anomaly detection

DEMO

Observability



Where data & AI come to 

3

Observability

We are not reactive on data issue

I want to be alerted if something unusual happen

- Regular check
- Deviation

- Regular check
- Deviation



LIVE DEMO

- New...
- Home
- Browse
- Data Access
- Saved Searches
- Profile_DataBrick...
- Validity Country
- COUNTRY
- Delta Lake Adven...
- enriched_sales_f...
- Country
- Product Data

Global Data Quality and Observability

Group By: Metric

Data Discovery @Scale

34.5K Column (Relational Databas...)	5.2K Table (Relational Databas...)
738 View (Relational Databas...)	292 Flat File (Flat file)

Group By: Metric

DQ @Scale - DQ Automation

28.1K Data Quality Rule O...	95 Data Quality Rule Te...
---------------------------------	-------------------------------

Group By: Threshold Result

DataObs File Monitoring By Severity

High: 23
Normal: 14

Group By: Severity

Supervised File Monitoring

Not Acceptable: 6 (85.71%)
Acceptable: 0 (0%)
Good: 1 (14.29%)

Group By: Threshold Result

Shared by: JA

<h4>Completeness</h4> <p>+11.26%</p> <table border="1"><tr><td>Not Acceptable: 810 (18.74%)</td></tr><tr><td>Acceptable: 112 (2.59%)</td></tr><tr><td>Good: 3401 (78.67%)</td></tr></table>	Not Acceptable: 810 (18.74%)	Acceptable: 112 (2.59%)	Good: 3401 (78.67%)	<h4>Consistency</h4> <p>+140.99%</p> <table border="1"><tr><td>Not Acceptable: 219 (48.56%)</td></tr><tr><td>Acceptable: 82 (18.18%)</td></tr><tr><td>Good: 150 (33.26%)</td></tr></table>	Not Acceptable: 219 (48.56%)	Acceptable: 82 (18.18%)	Good: 150 (33.26%)
Not Acceptable: 810 (18.74%)							
Acceptable: 112 (2.59%)							
Good: 3401 (78.67%)							
Not Acceptable: 219 (48.56%)							
Acceptable: 82 (18.18%)							
Good: 150 (33.26%)							
<h4>Uniqueness</h4> <p>No Change</p> <table border="1"><tr><td>Not Acceptable: 0 (0%)</td></tr><tr><td>Acceptable: 2 (22.22%)</td></tr><tr><td>Good: 7 (77.78%)</td></tr></table>	Not Acceptable: 0 (0%)	Acceptable: 2 (22.22%)	Good: 7 (77.78%)	<h4>Timeliness</h4> <p>No Change</p> <table border="1"><tr><td>Not Acceptable: 17 (21.52%)</td></tr><tr><td>Acceptable: 0 (0%)</td></tr><tr><td>Good: 62 (78.48%)</td></tr></table>	Not Acceptable: 17 (21.52%)	Acceptable: 0 (0%)	Good: 62 (78.48%)
Not Acceptable: 0 (0%)							
Acceptable: 2 (22.22%)							
Good: 7 (77.78%)							
Not Acceptable: 17 (21.52%)							
Acceptable: 0 (0%)							
Good: 62 (78.48%)							
<h4>Performance</h4> <p>-0.38%</p> <table border="1"><tr><td>Not Acceptable: 83 (17.47%)</td></tr><tr><td>Acceptable: 32 (6.74%)</td></tr><tr><td>Good: 360 (75.79%)</td></tr></table>	Not Acceptable: 83 (17.47%)	Acceptable: 32 (6.74%)	Good: 360 (75.79%)	<h4>Pipeline</h4> <p>+0.3%</p> <table border="1"><tr><td>Not Acceptable: 161 (29.65%)</td></tr><tr><td>Acceptable: 46 (8.47%)</td></tr><tr><td>Good: 336 (61.88%)</td></tr></table>	Not Acceptable: 161 (29.65%)	Acceptable: 46 (8.47%)	Good: 336 (61.88%)
Not Acceptable: 83 (17.47%)							
Acceptable: 32 (6.74%)							
Good: 360 (75.79%)							
Not Acceptable: 161 (29.65%)							
Acceptable: 46 (8.47%)							
Good: 336 (61.88%)							
<h4>File Monitoring</h4>							

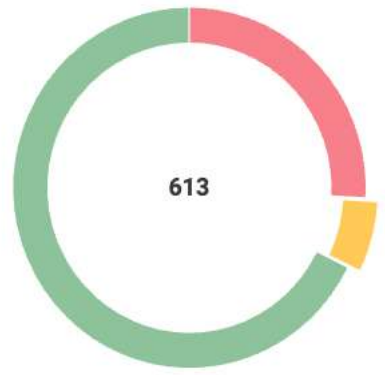
- New...
- Home
- Browse
- Data Access
- Saved Searches
- Profile_DataBrick...
- Validity Country
- COUNTRY
- Delta Lake Adven...
- enriched_sales_f...
- Country
- Product Data

[Back to Global Data Quality and Observability](#)

Global DQ Score ▼

About

Select a section from the donut chart to see details on the table.



● Not Acceptable: 2.5K (25.9%) ● Acceptable: 613 (6.38%) ● Good: 6.5K (67.72%)

Group By: Threshold Result

Data Quality Scores: Acceptable (613)

Filter Threshold Result: Acceptable

Name	Dimension	Latest Score	Total Rows	Failed Rows	Primary Da
RO Consistency TotalChildren Greater Thar	Consistency	92.42%	18.9K		...
Auto-EMAILADDRESS-Reliability-Re	RO Consistency TotalChildren Greater Than ChildrenAtHome		51.8K	7.2K	EMAILADD
Auto-email-Pipeline-Reliable email	Pipeline	75%	8	2	email
Auto-email-Pipeline-Reliable email	Pipeline	80%	10	2	email
Auto-email-Pipeline-Reliable email	Pipeline	83.33%	12	2	email
Auto-email-Completeness-Email Address -	Completeness	75%	8	2	email
Auto-POSTALCODE_DETAIL-Validity-Postal	Validity	71.12%	658	190	POSTALCO
Auto-POSTAL_CODE-Validity-Postal Code -	Validity	73.91%	92	24	POSTAL_CI
Auto-GENDER-Validity-Gender - Validity - NI	Validity	79.1%	1K	209	GENDER
Auto-BillingPostalCode-Validity-Postal Cod	Validity	89.14%	903	98	BillingPosti
Auto-POSTAL_CODE-Validity-Postal Code -	Validity	87.62%	202	25	POSTAL_CI
Auto-C_EMAIL-Performance-Email comple	Performance	84.61%	13	2	C_EMAIL
Auto-POSTALCODE-Validity-Postal Code - \	Validity	75.82%	91	22	POSTALCO

1 - 25 of 613 Items

Page 1 of 25

Items Per Page: 25

RO Consistency TotalChild... DATA QUALITY RULE OCCURRENCE

Overview Related Assets (0)

Description:
TotalChildren should be higher or equal to NumberChildrenAtHome

PUBLISHED

Hierarchy:
Top Level

Custom Attributes:
Dimension: Actual Columns:
Consistency -

Last Updated:
Oct 13, 2025, 8:33 AM

Reference ID:
03083fee-58c6-4839-8ef2-c77d8ec9e7d0

Stakeholders:

JA John Admin
Governance Administrator

Related Assets:

Customer Master

TABLE DIM_CUSTOMER

OVERALL RATING
★★★★★(4)
Add Rating

LIFECYCLE
PUBLISHED

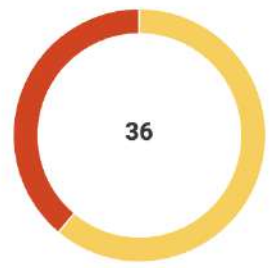
DATA MARKETPLACE
PUBLISHED

LAST UPDATED
Aug 25, 2025, 10:48 AM

- Overview
- Contains
- Lineage
- Relationships
- Data Quality
- Data Observability**
- Reference Data
- Stakeholders
- Attributes
- Tickets
- History

Open Events

Severity Metric



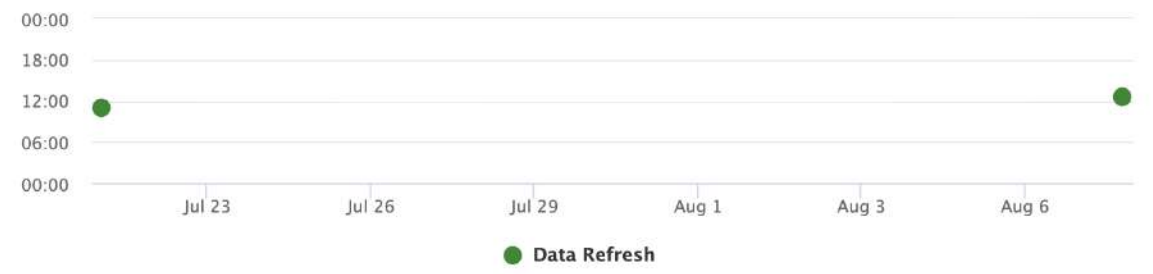
Normal: 22
High: 14

Group By: Severity

Freshness and Volume

Freshness Volume

Data Status: **ENABLED**



Open Events (36)

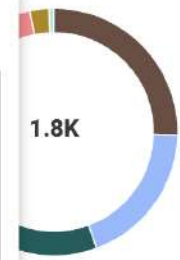
Filter, Sort, Settings icons

Name	Metric	Asset	Severity ↑	Last Detected	Status
Percentage of Duplicate Rows change in 'DateFirstPurchase'	Percentage of Duplicate Rows	DateFirstPurchase	High	Sep 17, 2025, 5:26 PM	OPEN
Percentage of Zero Rows change in 'NumberCarsOwned'	Percentage of Zero Rows	NumberCarsOwned	High	Sep 17, 2025, 5:26 PM	OPEN
Percentage of Distinct Rows change in 'FirstName'	Percentage of Distinct Rows	FirstName	High	Sep 17, 2025, 5:26 PM	OPEN

Open Events by Metric ⓘ

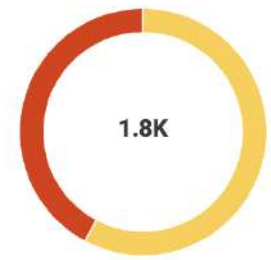
Find

- Percentage of Distinct Rows
- Percentage of Duplicate Rows
- Percentage of Blank Rows
- Percentage of Zero Rows
- Percentage of Null Rows
- Maximum Length
- Minimum Length
- Average



- Percentage of Distinct Rows: **457**
 - Percentage of Duplicate Rows: **341**
 - Average: **293**
 - Maximum Length: **235**
 - Minimum Length: **162**
- ▲ 1/3 ▼

Open Events by Severity ⓘ



- Normal: **1K**
- High: **759**

Group By: Severity



Filter Metric: ▾ × Add Filter

Name	Metric	Asset	Severity ↑	Last Detected	Status
Percentage of Duplicate Rows change in 'LastName'	Percentage of Duplicate Rows	☐ LastName	High	Sep 18, 2025, 10:37 AM	OPEN
Percentage of Distinct Rows change in 'AddressLine1'	Percentage of Distinct Rows	☐ AddressLine1	High	Sep 18, 2025, 10:37 AM	OPEN
Percentage of Distinct Rows change in 'BirthDate'	Percentage of Distinct Rows	☐ BirthDate	High	Sep 18, 2025, 10:37 AM	OPEN
Percentage of Duplicate Rows change in 'EmailAddress'	Percentage of Duplicate Rows	☐ EmailAddress	High	Sep 18, 2025, 10:37 AM	OPEN
Percentage of Distinct Rows change in 'MiddleName'	Percentage of Distinct Rows	☐ MiddleName	High	Sep 18, 2025, 10:37 AM	OPEN
Percentage of Duplicate Rows change in 'FirstName'	Percentage of Duplicate Rows	☐ FirstName	High	Sep 18, 2025, 10:37 AM	OPEN
Percentage of Duplicate Rows change in 'DateFirstPurchase'	Percentage of Duplicate Rows	☐ DateFirstPurchase	High	Sep 18, 2025, 10:37 AM	OPEN
Percentage of Distinct Rows change in 'Billing_Date'	Percentage of Distinct Rows	☐ Billing_Date	High	Sep 18, 2025, 10:37 AM	OPEN

- 1 Registration
- 2 Configuration**
- 3 Associations
- 4 Schedule

- ✓ Metadata Extraction
- ✓ Lineage Discovery
- ✓ Data Profiling and Quality
- ✓ Data Observability**
- ✓ Data Classification

To get accurate results for data observability, set the Profiling Scope option to Full in the Data Profiling and Quality configuration.

Enable Data Observability:

Parameters

Minimum Number of Data Points: * ?

Maximum Events to Generate: * ?

Filters

Metric Filters

- No filters ?
- Filter conditions ?



Uniqueness

Percentage of Dist...

Percentage of Dupli...

Completeness

Percentage of Null ...



- New...
- Home
- Browse
- Data Access
- Saved Searches
- Profile_DataBrick...

Global Data Quality and Observability

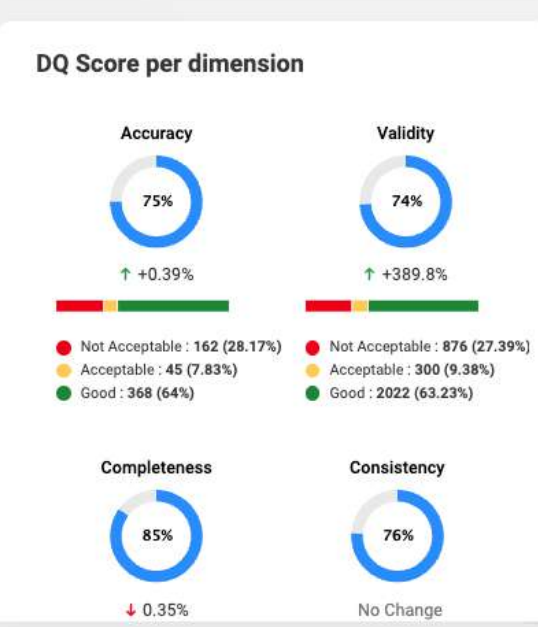
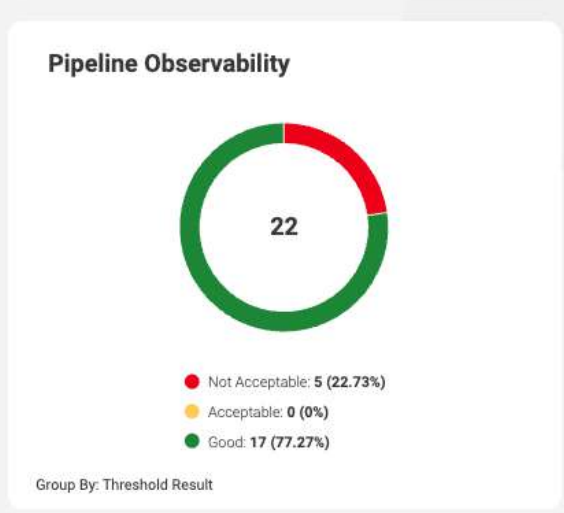
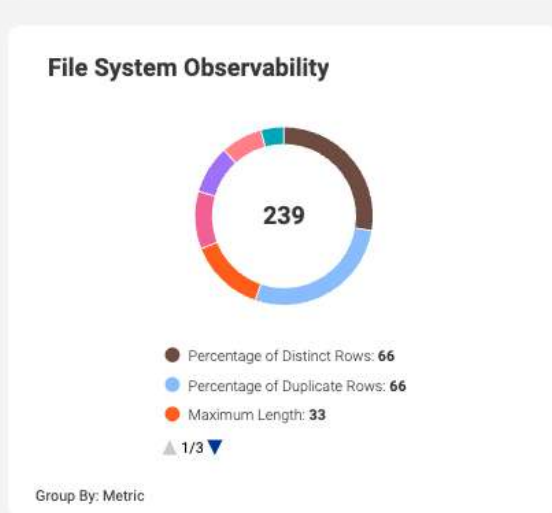
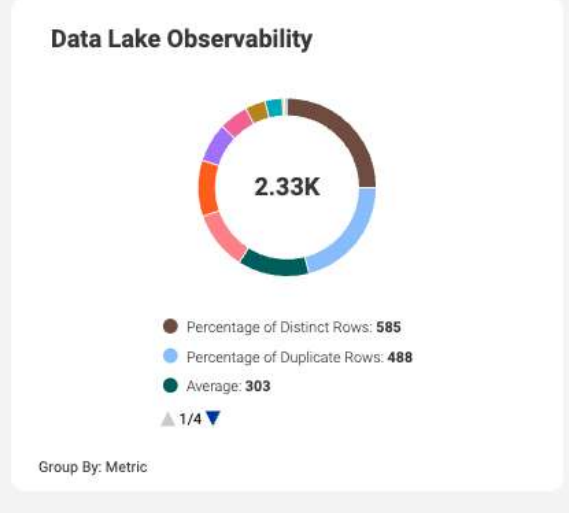
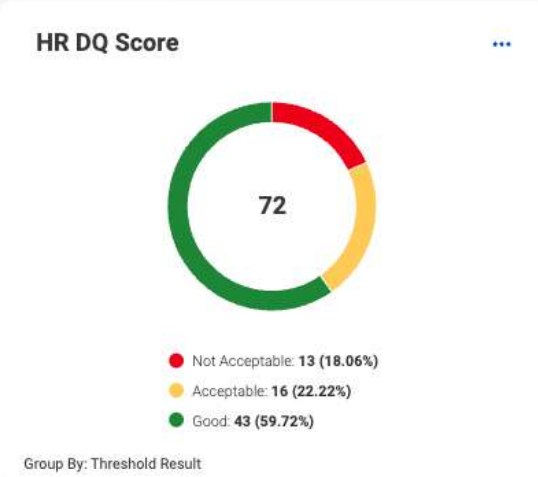
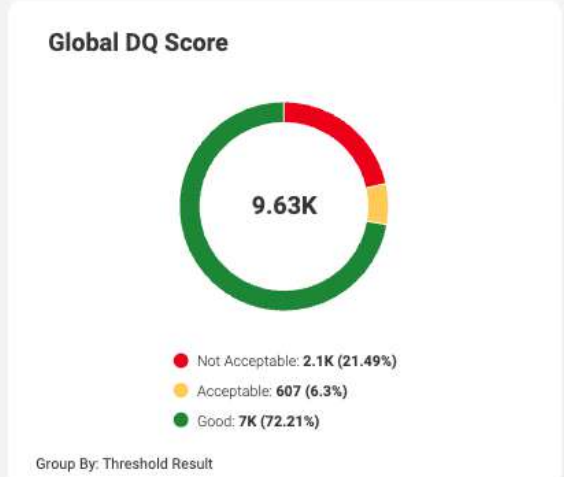
Shared with: All Users

Governed Technical Assets

46.2K Column (Relational Databas...)	3.1K Table (Relational Databas...)
397 Catalog Source	284 Hierarchical File (Hierarchical files)
236 Schema (Relational Databas...)	64 Mapping task (Informatica Intellig...)

Governed Business Assets

259 Business Term	208 Policy
186 Process	71 Subdomain
48 AI Model	33 Metric
32 System	12 Domain



Data Discovery @Scale

DQ @Scale - DQ Automation

DataObs File Monitoring By Severity

Percentage of Zero Rows change in 'DIMCUST_NumberCarsOwned_DWH'

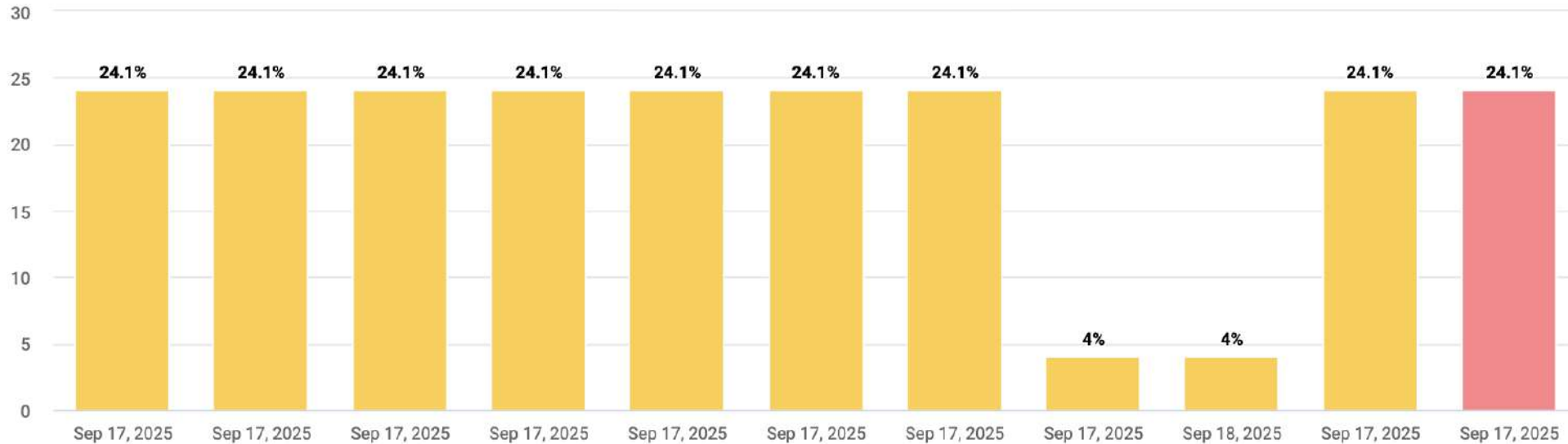
DATA OBSERVABILITY EVENT

Overview

Summary

The value usually remains constant but has suddenly changed.

Percentage of Zero Rows



YOY Growth Revenue Report

TABLEAU WORKSHEET YOY Growth Revenue

OVERALL RATING
★★★★★(1)
Add Rating

LIFECYCLE
PUBLISHED

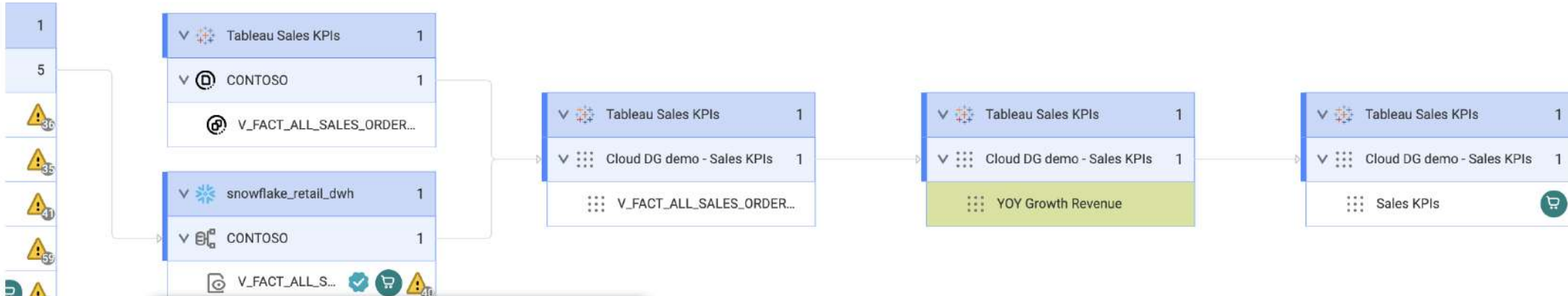
LAST UPDATED
May 22, 2025, 5:44 PM

- Overview
- Contains
- Lineage**
- Relationships
- Data Quality
- Reference Data
- Stakeholders
- Attributes
- Tickets
- History

Data Set Level

Overlay Selected: Certification, Data Obsei x

Find Assets



Highlight Path | Isolate Path

Data Observability Events (40)

Row Count 6 hours ago

STATUS **OPEN** | SEVERITY Normal

+ 39 more

Linked Data Products



Results

Quality Governance

*My domain is
never on data
quality scope*

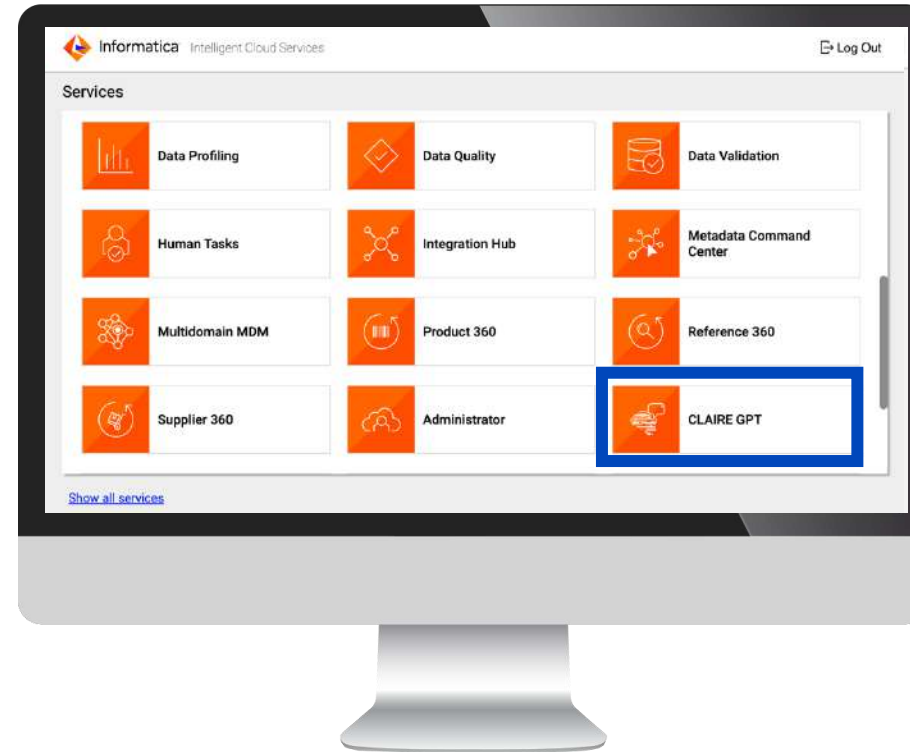
You scale at enterprise level

Solutions

- Metrics
- Data Lineage
- Verification

DEMO

AI Agent
Clean my data!



Where data & AI come to 

4

AI Agent

What if I want to innovate?

I want to let "anybody" benefit from our data quality assets

- Suggest
- 100% no code
- Improve

- Suggest
- 100% no code
- Improve



LIVE DEMO

Home Insert Draw Page Layout Formulas Data Review View Automate
Comments Share

Paste Aptos Narrow (Bod...) 12 A⁺ A⁻
General
Conditional Formatting Insert Format as Table Cell Styles
Delete Format
Σ Sort & Filter Find & Select
Sensitivity Add-ins Analyse Data

A1 fx ACCT_NO

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	ACCT_NO	CUST_ID	ACCT_TYP	CURR_BAL	AVAIL_BAL	OPEN_DT	LAST_TXN_DT	ACCT_STAT	BRNCH_CD	PROD_CD	INT_RT	CR_LMT	YTD_INT	MNTHLY_FEE	KYC_STAT	CUST_SEG	MTD_DPST	MTD_WTHDRW	
2	ACC10012345	C000123	DDA	15420.5	15420.5	15/03/2019	05/10/2025	ACT	BR001	CHK001	0.05			0	VERIFIED	RETAIL	2500	800	
3	ACC10012346	C000124	SAV	45000	45000	22/07/2020	07/10/2025	ACT	BR001	SAV002	2.5		1125	5	VERIFIED	MASS_AFF	5000	0	
4	ACC10012347	C000125	CCA	8500	3500	10/01/2021	30/09/2025	ACT	BR003	CRD003	18.99	15000			25	VERIFIED	RETAIL		
5	ACC10012348	C000126	DDA	250.75	250.75	20/05/2018	15/11/2025	ACT	BR002	CHK001	0.05			0	EXPIRED	RETAIL	0	150	
6	ACC10012345	C000123	DDA	15420.5	15420.5	15/03/2019	05/10/2025	ACTIVE	BR001	CHK001	0.05			0	VERIFIED	RETAIL	2500	800	
7	ACC10012349	C000127	SAV	120000	120000	30/11/2022	06/10/2025	ACT	BR005	SAV003	3.25		3900	0	VERIFIED	AFFLUENT	15000	2000	
8	ACC10012350	C000128	MMA	75000	75000	14/02/2023	01/10/2025	ACT	BR001	MMA001	4.1		2500	15	VERIFIED	MASS_AFF	0	5000	
9	ACC10012351	C000129	CCA		5000	08/09/2020	04/10/2025	ACT	BR004	CRD002	21.49	10000			35	VERIFIED	RETAIL		
10	ACC10012352	C000130	DDA	5840.25	5840.25	17/06/2021	07/10/2025	ACT	BR002		0.1			0	VERIFIED	RETAIL	1200	450	
11	ACC10012353	C000131	SAV	2500	2500	25/08/2019	10/05/2022	ACT	BR003	SAV002	2.5		62.5	5	PENDING	RETAIL	0	0	
12	ACC10012354	C000132	CCA	12000	0	19/03/2022	28/09/2025	ACT	BR001	CRD003	19.99	12000			25		RETAIL		
13	ACC10012355	C000133	DDA	-250	-250	05/12/2020	06/10/2025	ACT	BR005	CHK002	0			12.5	VERIFIED	RETAIL	500	1200	
14	ACC10012356	C000134	SAV	95000	95000	06/15/2021	03/10/2025	ACT	BR002	SAV004	3.75		3562.5	0	VERIFIED	AFFLUENT	10000	0	
15	ACC10012357	C000135	MMA	50000	50000	01/09/2023	07/10/2025	ACT	BR004	MMA001	4.1		1800	15	VERIFIED	MASS_AFF	8000	3000	
16	ACC10012358	C000136	DDA	1250.3	1250.3	20/11/2017	22/08/2021	DORMANT	BR003	CHK001	0.05			0	VERIFIED	RETAIL	0	0	
17	ACC10012359	C000137	CCA	7800	2200	14/07/2021	05/10/2025	ACT	BR001	CRD002	21.49	10000			35	VERIFIED	RETAIL		
18	ACC10012360	C000138	SAV	25000		10/05/2022	02/10/2025	ACT	BR005	SAV002	2.5		625	5	VERIFIED	MASS_AFF	3000	1000	
19	ACC10012361	C000139	DDA	8920.15	8920.15	28/02/2019	07/10/2025	ACT	BR002	CHK002	0.1			0	VERIFIED	RETAIL	2000	600	
20	ACC10012362	C000140	CCA	15000	10000	05/10/2020	29/09/2025	ACT	BR004	CRD004	17.99	25000			40	VERIFIED	AFFLUENT		
21	ACC10012349	C000127	SAV	120000	120000	30/11/2022	06/10/2025	ACT	BR005	SAV003	3.25		3900	0	VERIFIED	AFFLUENT	15000	2000	
22	ACC10012363	C000141	MMA	150000	150000	20/07/2023	04/10/2025	ACTIVE	BR001	MMA002	4.5		6750	10	VERIFIED	HNW	20000	0	
23	ACC10012364	C000142	DDA	420	420	12/09/2018	20/12/2024	ACT	BR003	CHK001	0.05			0	EXPIRED	RETAIL	100	50	
24	ACC10012365	C000143	SAV	0	0	30/04/2021	15/01/2025	ACT	BR002	SAV001	1.5		0	5	VERIFIED	RETAIL	0	0	
25	ACC10012366		CCA	5600	1400	18/08/2022	06/10/2025	ACT	BR005	CRD002	21.49	7000			35	VERIFIED	RETAIL		
26	ACC10012367	C000145	DDA	32500	32500	25/01/2020	07/10/2025	ACT	BR001	CHK003	0.25			0	VERIFIED	MASS_AFF	5000	1500	
27	ACC10012368	C000146	SAV	180000	180000	15/03/2023	05/10/2025	ACT	BR004	SAV005	4		7200	0	VERIFIED	HNW	25000	5000	
28	ACC10012369	C000147	CCA	9999.99	111.01	22/11/2021	25/09/2025	ACT	BR002	CRD003	18.99	10000			25	PENDING	RETAIL		

FB - Oct 14 - clean Bank Account Last modified: 10/14/2025, 09:39 AM

VT

could you clean this file



I will use the Data Quality Cleansing skill to clean your file bank_acct_dataset.xlsx.

> Reasoning

Data Preview



bank_acct_dataset Data Preview

TABULAR LIST

Data preview of 50 random records from the dataset

> Reasoning

Data Profile



Data Profile

TABULAR LIST

Profiling statistics of the dataset.



Ask CLAIRE...



FB - Oct 14 - clean Bank Account | Last modified: 10/14/2025, 09:39 AM

Data Profile

Data Profile

#	name	dataType	valueCount	nullCount	distinctCount	duplicateCount	average
1	ACCT_NO	String	30	-	28	2	NaN
2	CUST_ID	String	30	1	27	2	NaN
3	ACCT_TYP	String	30	-	4	4	NaN
4	CURR_BAL	Float	30	2	26	2	38524.02
5	AVAIL_BAL	Float	30	1	26	3	38102.88
6	OPEN_DT	Datetime	30	-	28	2	NaN
7	LAST_TXN_DT	Datetime	30	-	16	6	NaN
8	ACCT_STAT	String	30	-	3	2	NaN
9	BRNCH_CD	String	30	-	5	5	NaN
10	PROD_CD	String	30	1	13	7	NaN
11	INT_RT	Float	30	-	15	7	6.15
12	CR_LMT	Float	30	23	5	1	12714.29
13	YTD_INT	Float	30	17	12	1	2857.35
14	MNTHLY_FEE	Float	30	7	5	3	3.8
15	KYC_STAT	String	30	-	6	4	NaN
16	CUST_SEG	String	30	1	6	5	NaN
17	MTD_DPST	String	30	-	15	8	NaN
18	MTD_WTHDRW	Float	30	7	12	4	1145.65

Showing 18 of 18

Data Profile



Data Profile

TABULAR LIST

Profiling statistics of the dataset.

Data Quality Cleansing High-Level Plan

- **Infer Metadata Attributes****
Perform on-the-fly metadata inference for the uploaded file to determine key attributes such as data types, domains, and possible value ranges for each column.
- **Identify Critical Data Elements****
Based on the inferred metadata and the context of the dataset, identify the most important or key data elements that are essential for downstream processes or analyses.
- **Assess Functional Dependencies****
Analyze the relationships among the critical data elements to identify potential functional dependencies within the dataset, which helps in understanding data integrity and consistency.
- **Compute Statistical Data Quality Metrics and Recommend Fixes****

Ask CLAIRE...

FB - Oct 14 - clean Bank Account | Last modified: 10/14/2025, 09:39 AM

Bank Account Dataset - Column Level Enrichment

Column Level Metadata Enrichment

#	Name	Data Type	Data Kind	Entity	Sensitivity	Generated Description
1	ACCT_NO	String	categorical	Account	Non-sensitive	Unique alphanumeric identifier for each bank account. Used as the primary key for the dataset. Duplicates exist and should be cleaned for uniqueness.
2	CUST_ID	String	categorical	Customer	Sensitive	Identifies each customer. Links accounts to individual customers. Some account...
3	ACCT_TYP	String	categorical	Account Type	Non-sensitive	Type of bank account, such as DDA (Demand Deposit Account), CCA (Credit Card Account), SAV (Savings...
4	CURR_BAL	Float	numeric	Account Balance	Sensitive (financial)	Type of bank account, such as DDA (Demand Deposit Account), CCA (Credit Card Account), SAV (Savings Account), or MMA (Money Market Account).
5	AVAIL_BAL	Float	numeric	Account Balance	Sensitive (financial)	Available balance in the account, representing funds accessible for withdrawal or spending. May differ...
6	OPEN_DT	Datetime	datetime	Account Opening D...	Non-sensitive	Date when the account was opened. Format varies (ISO 8601 and MM/DD/YYYY). Used for account ag...
7	LAST_TXN_DT	Datetime	datetime	Transaction Date	Non-sensitive	Date of the most recent transaction on the account. Useful for activity and dormancy analysis.
8	ACCT_STAT	String	categorical	Account Status	Non-sensitive	Current status of the account, such as ACT (Active), ACTIVE, DORMANT, or EXPIRED. Indicates whether the...
9	BRNCH_CD	String	categorical	Branch	Non-sensitive	Code identifying the branch where the account is held. Useful for geographic or branch-level analysis.
10	PROD_CD	String	categorical	Product	Non-sensitive	Code representing the specific banking product associated with the

1. **Dataset Name:** Not explicitly provided
Description: This dataset contains detailed information about customer bank accounts, including account identifiers, customer identifiers, account types, balances, product and branch codes, interest rates, credit limits, accrued interest, monthly fees, KYC (Know Your Customer) status, customer segments, and monthly transaction summaries. Each row represents a specific account held by a customer, with columns capturing both static attributes (such as account type, branch code, and product code) and dynamic financial metrics (such as current and available balances, year-to-date interest, and monthly deposits/withdrawals). The dataset also tracks account status, opening and last transaction dates, and compliance-related information. The data is suitable for profiling customer segments, analyzing account activity, and supporting compliance and risk management processes. There are some duplicate and missing values, and the primary key is the account number (ACCT_NO), though some duplicates exist and should be cleaned.
Domain: Banking
Primary Key: ACCT_NO (duplicates exist and should be cleaned for uniqueness)

Bank Account Dataset - Column Level Enrichment
 TABULAR LIST

Inferred metadata attributes for each column in the dataset.

> Reasoning

Ask CLAIRE...

FB - Oct 14 - clean Bank Account | Last modified: 10/14/2025, 09:39 AM

Statistical Data Quality Metrics

Statistical Issues

#	Issue Name	Issue Description	Dimension	Fix Type	Samples Causing Issue	Reason To Fix
1	uniqueness_ACCT_...	Duplicate account numbers found: 'ACC10012345' and 'ACC10012349' each appear twice, with identical or nearly identical details. Example: {ACCT_NO: 'ACC10012345', CUST_ID: 'C000123', CURR_BAL: 15420.50}.	Uniqueness	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012345, CUST_ID: C000123 ACCT_NO: ACC10012349, CUST_ID: C000127 	Ensuring each account unique is critical for account-level analysis.
2	completeness_CU...	Missing customer ID in at least one record. Example: {ACCT_NO: 'ACC10012366', CUST_ID: None}.	Completeness	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012366, CUST_ID: None 	Customer ID is essential for accounts to customers maintaining referential integrity.
3	completeness_CU...	Missing current balance in some records. Example: {ACCT_NO: 'ACC10012371', CURR_BAL: None}.	Completeness	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012371, CURR_BAL: None ACCT_NO: ACC10012351, CURR_BAL: None 	Current balance is a critical metric. Missing values affect financial reporting and analysis.

> Reasoning

The identified statistical issues include:

- **Duplicate account numbers** (e.g., ACC10012345 and ACC10012349 appear twice)
- **Missing values** in key fields such as customer ID, current balance, available balance, product code, customer segment, monthly fee, month-to-date withdrawal, credit limit, and year-to-date interest
- **Unusually high values** (outliers) in current balance, available balance, interest rate, credit limit, monthly fee, and month-to-date withdrawals

These issues are evidenced by specific examples, such as missing customer IDs (e.g., ACC10012366), duplicate account numbers, and outlier values like a current balance of 180,000.0.

Statistical Data Quality Metrics
TABULAR LIST

List of statistical issues identified in the dataset along with their descriptions

> Reasoning

The identified standardization issues are:

- **format_OPEN_DT**: Inconsistent date formats in the OPEN_DT column

Ask CLAIRE...

FB - Oct 14 - clean Bank Account | Last modified: 10/14/2025, 09:39 AM

Candidates for Standardization

Standardization Issues

#	Issue Name	Issue Description	Dimension	Fix Type	Samples Causing Issue	Reason To Fix
1	format_OPEN_DT	Inconsistent date formats in OPEN_DT column: some values are in ISO 8601 format (e.g., '2019-06-...)	Consistency	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012356, OPEN_DT: 06/15/2021 ACCT_NO: ACC10012345, OPEN_DT: 2019-03-15T00:00:00 ACCT_NO: ACC10012370, OPEN_DT: 2019-06-30T00:00:00 	Standardizing date format is essential for reliable data analysis, sorting, filtering, and reporting for active accounts, and consistency should be enforced for others.
2	format_LAST_TXN...	LAST_TXN_DT values are in ISO 8601 format with time ('YYYY-MM-DDTHH:MM:SS'), which may be...	Consistency	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012370, LAST_TXN_DT: 2025-10-03T00:00:00 ACCT_NO: ACC10012356, LAST_TXN_DT: 2025-10-03T00:00:00 ACCT_NO: ACC10012345, LAST_TXN_DT: 2025-10-05T00:00:00 	Removing unnecessary components from date reduces storage, improves performance, and ensures consistency for others.
3	label_ACCT_STAT	Inconsistent account status labels in ACCT_STAT: both 'ACT' and 'ACTIVE' are used for active accounts, and...	Consistency	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012345, ACCT_STAT: ACT ACCT_NO: ACC10012345, ACCT_STAT: ACTIVE 	Standardizing status labels ensures accurate grouping, filtering, and reporting of account status for others.

Standardizing date format is essential for reliable data analysis, sorting, filtering, and reporting for active accounts, and consistency should be enforced for others.

- case_KYC_STAT:** KYC_STAT values are in uppercase, but case consistency should be enforced.
- case_CUST_SEG:** CUST_SEG values are in uppercase, but case consistency should be enforced.
- label_ACCT_TYP:** ACCT_TYP uses short codes; if full names are required, these should be standardized.

These issues are evidenced by the descriptions of inconsistent formats, label variations, and case usage in the relevant columns.

Candidates for Standardization
TABULAR LIST

List of standardization issues identified in the dataset along with their descriptions

> Reasoning

> Show Code

Cleaning code has been generated for all the issues.

> Reasoning



Ask CLAIRE...

FB - Oct 14 - clean Bank Account Last modified: 10/14/2025, 09:39 AM

Candidates for Standardization

Standardization Issues

#	Dimension	Fix Type	Samples Causing Issue	Reason To Fix	Suggested Fixes
			05100:00:00 • ACCT_NO: ACC10012345, LAST_TXN_DT: 2025-10-05T00:00:00		
3	Status labels in and 'ACTIVE' units, and...	Consistency	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012345, ACCT_STAT: ACT ACCT_NO: ACC10012363, ACCT_STAT: ACTIVE ACCT_NO: ACC10012358, ACCT_STAT: DORMANT 	Standardizing status labels ensures accurate grouping, filtering, and reporting of account states... <ul style="list-style-type: none"> Map all variants of active status ('ACT', 'ACTIVE') to a single standard label (e.g., 'ACTIVE'). Similarly, ensure all other statuses use a consistent set of labels. In Python, use pandas.Series.replace() with a mapping dictionary.
4	uppercase 'NG', consistency...	Consistency	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012353, KYC_STAT: PENDING ACCT_NO: ACC10012348, KYC_STAT: EXPIRED ACCT_NO: ACC10012370, KYC_STAT: VERIFIED 	Enforcing case consistency prevents issues with grouping, filtering, and downstream processing, especially... <ul style="list-style-type: none"> Convert all KYC_STAT values to uppercase using pandas.Series.str.upper().
5	uppercase 'IT', case...	Consistency	fixable	<ul style="list-style-type: none"> ACCT_NO: ACC10012370, CUST_SEG: RETAIL ACCT_NO: ACC10012371 	Consistent casing ensures reliable grouping and filtering, and prevents errors if new data is appended in... <ul style="list-style-type: none"> Convert all CUST_SEG values to uppercase using pandas.Series.str.upper().

Showing 12 of 12

for others.

- **case_KYC_STAT**: KYC_STAT values are in uppercase, but case consistency should be enforced.
- **case_CUST_SEG**: CUST_SEG values are in uppercase, but case consistency should be enforced.
- **label_ACCT_TYP**: ACCT_TYP uses short codes; if full names are required, these should be standardized.

These issues are evidenced by the descriptions of inconsistent formats, label variations, and case usage in the relevant columns.

Candidates for Standardization
TABULAR LIST

List of standardization issues identified in the dataset along with their descriptions

> Reasoning

> Show Code

Cleaning code has been generated for all the issues.

> Reasoning



Ask CLAIRE...

FB - Oct 14 - clean Bank Account | Last modified: 10/14/2025, 09:39 AM

Clean Data Sample bank_acct_dataset

Clean data sample

Reason for Fixes	AVAIL_BAL	OPEN_DT	LAST_TXN_DT	ACCT_STAT	BRNCH_CD	PI
Fill missing credit limit values with 0.0 for non-credit card accounts, convert all opening and last...	150000.0	2023-07-20	2025-10-04	ACTIVE	BR001	M
Fill missing credit limit values with 0.0 for non-credit card accounts, convert all opening and last...	150000.0	2017-11-20T00:00:00	2017-11-20	ACTIVE	BR001	M
Fill missing credit limit values with 0.0 for non-credit card accounts, fill missing year-to-date interest with...	1250.3	2017-11-20	2021-08-22	DORMANT	BR003	Cl
Fill missing credit limit values with 0.0 for non-credit card accounts, fill missing year-to-date interest with...	1250.3	2017-11-20	2021-08-22	DORMANT	BR003	Cl
Fill missing current balance with 0.0 for new or inactive accounts or flag for review, fill missing credit limit...	88000.0	2022-12-08	2025-10-07	ACTIVE	BR001	M
Fill missing current balance with 0.0 for new or inactive accounts or flag for review, fill missing credit limit...	88000.0	2022-12-08	2025-10-07	ACTIVE	BR001	M
Fill missing product codes with 'Unknown', fill missing credit limit values with 0.0 for non-credit card...	5840.25	2021-06-17	2025-10-07	ACTIVE	BR002	Ur
Fill missing product codes with 'Unknown', fill missing credit limit values with 0.0 for non-credit card...	5840.25	2021-06-17	2025-10-07	ACTIVE	BR002	Ur

Before and After Values

OPEN_DT:
2017-11-20T00:00:00 → 2017-11-20

Cleaning code has been generated for all the issues.

> Reasoning

Clean Data Sample bank_acct_dataset
TABULAR LIST

Data preview showing before and after values for records affected by cleansing rules

Agent Log of clean data for bank_acct_dataset
AGENT LOG

Data quality analysis dashboard generated successfully.

Data cleansing for your uploaded file bank_acct_dataset.xlsx has been completed successfully. What would you like to do next?

- Download the cleansed file
- Show a summary of fixes applied
- Analyze data quality of another file
- Recommend data quality rules for this dataset
- Show a before-and-after data preview

Ask CLAIRE...

FB - Oct 14 - clean Bank Account | Last modified: 10/14/2025, 09:39 AM

Agent Log of clean data for bank_acct_dataset

Here's a comprehensive clean data quality report

Rules Overview

500

Sample Records

22

Total Rules

16

Fixable Rules

6

Non-Fixable Rules

168

Total Violations Fixed

16.22% Quality Improvement

Data Quality Dimension Overview

#	Dimension	Fixable Issues	Non-Fixable Issues	Total Violations Fixed
1	Uniqueness	1		
2	Completeness	9		57
3	Accuracy		6	
4	Consistency	6		111

Rule Details 14

#	Rule Name	Violation Count	Description
1	format_OPEN_DT	28	Inconsistent date formats in OPEN_DT column:...
2	format_LAST_TXN_DT	28	LAST_TXN_DT values are in ISO 8601 format w...
3	label_ACCT_TYP	28	ACCT_TYP values use short codes ('DDA', 'CCA'...

Cleaning code has been generated for all the issues.

> Reasoning



Clean Data Sample bank_acct_dataset

TABULAR LIST

Data preview showing before and after values for records affected by cleansing rules



Agent Log of clean data for bank_acct_dataset

AGENT LOG

Data quality analysis dashboard generated successfully.

Data cleansing for your uploaded file bank_acct_dataset.xlsx has been completed successfully. What would you like to do next?

Download the cleansed file

Show a summary of fixes applied

Analyze data quality of another file

Recommend data quality rules for this dataset

Show a before-and-after data preview

Ask CLAIRe...



CLAIRe GPT might make mistakes and is not designed for high risk use cases. Verify all outputs before proceeding.

Can you please help me clean my citizen dataset Last modified: 03/17/2026, 01:31 AM

Service Request Data Cleaning Mapplet

Open in Data Integration [Icons]

The mapplet implements fixes for completeness, validity, and standardization issues.



Rules Code

```

PYTHON

# Pre-Partition Python Code
# Python version: 3.9
import pandas as pd
import numpy as np
import re

all_rows_data = []

def is_null(value):
    """
    Check if a value is null (NaN or None).
    """
    return pd.isnull(value)

def is_empty_string(value):
    """
    Check if a value is an empty or all-whitespace string.
    """
    return isinstance(value, str) and value.strip() == ""
  
```

Agent Log of clean data for CITIZEN_SERVICES_v2
REPORT
Data quality analysis dashboard generated successfully.

> Reasoning

Service Request Data Cleaning Mapplet
MAPPLET ARTIFACT
Cleans and standardizes service request data for completeness, validity, and consistency.

Citizen dataset cleansing and assessment completed successfully. Let me know what you would like to do next.

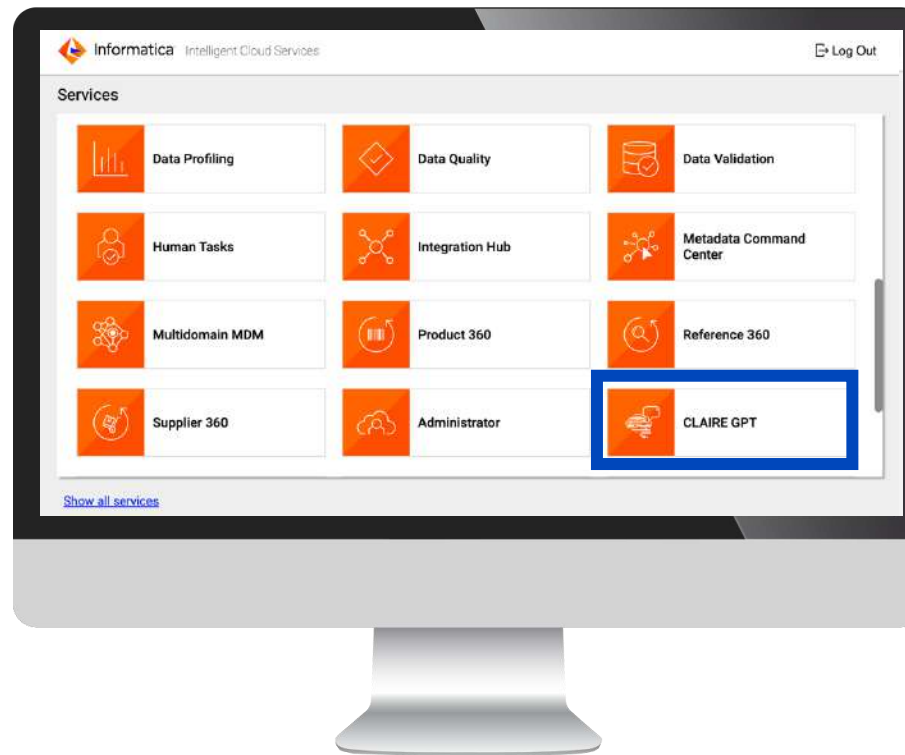
- Review the final data quality dashboard
- Download the generated mapplet for future cleansing
- Cleanse another dataset

👍 Cards (8)

Ask CLAIRE... [Icons]

DEMO

AI Agent
Enrich my Quality



Where data & AI come to 

create data quality rule using attached file Last modified: 02/11/2026, 07:45 PM

JA

create data quality rule using attached file

 dq_finance_business_rules_v1.xlsx



I will use the **Data Quality Rule Generation Skill** to help you create data quality rules using the attached file dq_finance_business_rules_v1.xlsx.

Starting the process now.

> Reasoning

45 rules are extracted from file dq_finance_business_rules_v1.xlsx

Here are some sample rules:

Completeness

- **Account_Must_Have_Customer**
Every bank account must be linked to a valid customer identifier. Check that Customer ID field is not empty or null.
- **Customer_Must_Have_Valid_Email**
Email address is required for digital communication. Check that Email Address field is not empty AND follows valid email format (contains @ and domain).

Validity

- **NoNegativeBalance_WithoutOverdraft**
Accounts without overdraft protection should not have negative balances. Check if Current Balance or Available Balance is less than zero AND account is not designated as overdraft-enabled.
- **Email_Must_Follow_Valid_Format**



Ask CLAIRE...



create data quality rule using attached file Last modified: 02/11/2026, 07:45 PM

Extracted Data Quality Rules (45)

Where needed, I've provided the extracted and generated rule names, fields and data types, dimensions as per the requirements of IDMC.

CLAIRE generated content is in gray.

#	Rule Name	Description	Dimension	Input Fields/Type	Dictionary
1	Account_Must_Have_Customer	Every bank account must be linked to a valid customer identifier. Here are more details on check to...	COMPLETENESS	• CUST_ID/string	
2	Account_Balance_Must_Be_Present	Current balance must be recorded for all accounts. Here are more details on check to perform :Check...	COMPLETENESS	• CURR_BAL/float	
3	Available_Balance_Must_Be_Present	Available balance must be recorded for all accounts to determine transaction limits. Here are more...	COMPLETENESS	• AVAIL_BAL/float	
4	Product_Code_Must_Be_Assigned	Every account must be associated with a valid product code. Here are more details on check to perform...	COMPLETENESS	• PROD_CD/string	
5	KYC_Status_Must_Be_Recorded	KYC status must be documented for all accounts. Here are more details on check to perform :Check that KY...	COMPLETENESS	• KYC_STAT/string	
6	Customer_Segment_Must_Be_Assigned	Customer segment classification must be assigned for analytics and targeted services. Here are more...	COMPLETENESS	• CUST_SEG/string	
7	NoNegativeBalance_WithoutOverdraft	Accounts without overdraft protection should not have negative balances. Here are more details on...	VALIDITY	• CURR_BAL/float • AVAIL_BAL/float	
8	ACTIVE_LastTxnDtWithin180Days	Accounts marked as active should have transactions within the last 180 days. Here are more details on...	ACCURACY	• LAST_TXN_DT/date/time	
9	DORMANT_NoRecentTxn180Days	Accounts marked as dormant should not have recent transaction activity. Here are more details on...	CONSISTENCY	• ACCT_STAT/string • LAST_TXN_DT/date/time	
10	ACCT_STAT_StandardizeACCT...	Account status must use...	CONSISTENCY	• ACCT_STAT/string	

Showing 45 of 45

Accuracy

- ACTIVE_LastTxnDtWithin180Days**
Accounts marked as active should have transactions within the last 180 days. Check if Last Transaction Date is more than 180 days in the past AND account status is ACTIVE.
- YTD_INT_MatchesBalanceAndRate**
Year-to-date interest should align with current balance and interest rate. Check if calculated interest (Balance × Rate) matches recorded YTD Interest within acceptable variance.

Uniqueness

- No_Duplicate_Customer_Records**
Each customer should have only one record in the system. Check that Customer ID appears only once in the table.
- CRD_NO_ACCT_NO_NoMultipleActiveCards**
Multiple active cards on one account may indicate duplicates or fraud. Check if Account Number has multiple cards with same characteristics and status ACTIVE.

Extracted Data Quality Rules (45)
SAMPLE DQ RULES

45 DQ rules identified from
dq_finance_business_rules_v1.xlsx

Ask **CLAIRE**...

CLAIRE GPT might make mistakes and is not designed for high risk use cases. Verify all outputs before proceeding.

create data quality rule using attached file Last modified: 02/11/2026, 07:45 PM

Data Quality Rules Generation Plan

- 1 Rules specifications validation** This step will filter out invalid rule specifications with respect to the rules repository.
- 2 Rules deduplication** This step will check if any rule specifications match already existing data quality rules to avoid generating duplicates.
- 3 Cluster rules specifications** This step will refactor semantically identical rule specifications into a single specification.
- 4 Generate rules script** This step will generate executable scripts for each rule specification.
- 5 Generate rules test data** This step will generate test data for each rule specification.
- 6 Test rules** This step will test each rule specification using generated scripts and test data.
- 7 Create a dashboard** This step will provide a summary report of all outputs.
- 8 Persist rules** This step will publish rule specifications into the CLAIRE project and CLAIRE folder.



Cards (1)

Do you accept the generated plan for converting your rule specifications into data quality rules, or would you like to modify any steps before proceeding?

Yes, proceed with the plan

No, modify the plan by:

JA

Yes, proceed with the plan



Ask CLAIRE...



create data quality rule using attached file Last modified: 02/11/2026, 07:45 PM

Generated DQ Rule Test Data (45)

Rules are generated based on business requirements and data patterns from uploaded file.

#	Rule Name	Description	Dimension	Input Fields/Type	Dictionary
1	Account_Must_Have_Customer	Every bank account must be linked to a valid customer identifier. Here are more details on check to...	COMPLETENESS	• CUST_ID/string	
2	Account_Balance_Must_Be_Present	Current balance must be recorded for all accounts. Here are more details on check to perform :Check...	COMPLETENESS	• CURR_BAL/float	
3	Available_Balance_Must_Be_Present	Available balance must be recorded for all accounts to determine transaction limits. Here are more...	COMPLETENESS	• AVAIL_BAL/float	
4	Product_Code_Must_Be_Assigned	Every account must be associated with a valid product code. Here are more details on check to perform...	COMPLETENESS	• PROD_CD/string	
5	KYC_Status_Must_Be_Recorded	KYC status must be documented for all accounts. Here are more details on check to perform :Check that KY...	COMPLETENESS	• KYC_STAT/string	
6	Customer_Segment_Must_Be_Assigned	Customer segment classification must be assigned for analytics and targeted services. Here are more...	COMPLETENESS	• CUST_SEG/string	
7	NoNegativeBalance_WithoutOverdraft	Accounts without overdraft protection should not have negative balances. Here are more details on...	VALIDITY	• CURR_BAL/float • AVAIL_BAL/float	
8	ACTIVE_LastTxnDtWithin180Days	Accounts marked as active should have transactions within the last 180 days. Here are more details on...	ACCURACY	• LAST_TXN_DT/date/time	
9	DORMANT_NoRecentTxn180Days	Accounts marked as dormant should not have recent transaction activity. Here are more details on...	CONSISTENCY	• ACCT_STAT/string • LAST_TXN_DT/date/time	
10	ACCT_STAT_StandardizedCodes	Account status must use standardized codes across all	CONSISTENCY	• ACCT_STAT/string	

Showing 45 of 45

> Reasoning

Number of rule scripts getting generated: 45



Generated DQ Rule Scripts (45)

DQ RULES ARTIFACTS

CLAIRE identified and generated 45 new DQ rules from uploaded file.

- Number of rules for which script generation succeeded: 45
- Number of rules for which script generation failed: 0

> Reasoning



Generated DQ Rule Test Data (45)

DQ RULES ARTIFACTS

CLAIRE identified and generated 45 new DQ rules from uploaded file.

- Number of rules for which test data generation succeeded: Test data was successfully generated for 45 rules.
- Number of rules for which test data generation failed: Test data...

Ask CLAIRE...



create data quality rule using attached file Last modified: 02/11/2026, 07:45 PM



DQ Rule Test Results (45)



Rules are generated based on business requirements and data patterns from uploaded file.

#	Rule Name	Description	Dimension	Input Fields/Type	Dictionary
1	Account_Must_Have_Customer	Every bank account must be linked to a valid customer identifier. Here are more details on check to...	COMPLETENESS	• CUST_ID/string	
2	Account_Balance_Must_Be_Present	Current balance must be recorded for all accounts. Here are more details on check to perform :Check...	COMPLETENESS	• CURR_BAL/float	
3	Available_Balance_Must_Be_Present	Available balance must be recorded for all accounts to determine transaction limits. Here are more...	COMPLETENESS	• AVAIL_BAL/float	
4	Product_Code_Must_Be_Assigned	Every account must be associated with a valid product code. Here are more details on check to perform...	COMPLETENESS	• PROD_CD/string	
5	KYC_Status_Must_Be_Recorded	KYC status must be documented for all accounts. Here are more details on check to perform :Check that KY...	COMPLETENESS	• KYC_STAT/string	
6	Customer_Segment_Must_Be_Assigned	Customer segment classification must be assigned for analytics and targeted services. Here are more...	COMPLETENESS	• CUST_SEG/string	
7	NoNegativeBalance_WithoutOverdraft	Accounts without overdraft protection should not have negative balances. Here are more details on...	VALIDITY	• CURR_BAL/float • AVAIL_BAL/float	
8	ACTIVE_LastTxnDtWithin180Days ✖ Partial generation	Accounts marked as active should have transactions within the last 180 days. Here are more details on...	ACCURACY	• LAST_TXN_DT/date/time	
9	DORMANT_NoRecentTxn180Days ✖ Partial generation	Accounts marked as dormant should not have recent transaction activity. Here are more details on...	CONSISTENCY	• ACCT_STAT/string • LAST_TXN_DT/date/time	
10	ACCT_STAT_StandardizedCodes	Account status must use standardized codes across all	CONSISTENCY	• ACCT_STAT/string	

Showing 45 of 45

Number of rules for which we are running tests: 45



DQ Rule Test Results (45)

DQ RULES ARTIFACTS

CLAIRE identified and generated 45 new DQ rules from uploaded file.



DQ Rule Test Results Summary

TEST SUMMARY

Results of executing tests

- **Number of rules for which tests passed:** 35 rules successfully met all test criteria.
- **Number of rules for which tests failed:** 10 rules did not meet the required test criteria.

> Reasoning



Data Quality Rules Agent Log

REPORT

Data Quality rules applied and test generated successful



Ask CLAIRE...



create data quality rule using attached file Last modified: 02/11/2026, 07:45 PM

Data Quality Rules Agent Log

Here's a log of what's been done for dq generation.

Summary of Generated Rules

45 Extracted Rules | 0 Existing Rules | 35 Rules Generated | 10 Rules Not Generated

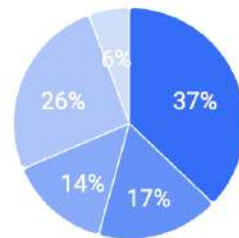
77.78% Rules Generated

Other Key Indicators

Clustering Candidates: 45
Clusters Representatives: 0

Rules Generated - Indicates the total number of valid, unique rules along with cluster representatives after removing duplicates and failures.

Distribution of Rules across Data Quality Dimensions



Data Quality Dimension	Number of Rules	Percentage
1 COMPLETENESS	13	37.14
2 VALIDITY	6	17.14
3 CONSISTENCY	5	14.29
4 ACCURACY	9	25.71
5 UNIQUENESS	2	5.71

Rules not generated

#	Rule Name	Description	Dimension	Input Fields/Type	Dictionary
1	ACTIVE_LastTxnDtWithin180Days	Accounts marked as active should ...	ACCURACY	• LAST_TXN_DT/date/time	

Data Quality Rules Agent Log
REPORT

Data Quality rules applied and test generated successfully.

77.78% Rules Generated

Cards (6)

Would you like to publish the 35 successfully tested rules to the CLAIRE project and CLAIRE folder, or do you want to review the failed rules before proceeding?

Publish Rules | Don't Publish Rules

JA

Publish Rules



Publishing the 35 successfully tested data quality rules to the CLAIRE project and CLAIRE folder as requested.

Ask CLAIRE...



CLAIRE GPT might make mistakes and is not designed for high risk use cases. Verify all outputs before proceeding.

Results

Quality

*I do not trust
anymore this
report*

Consumers trust BI reports

Quality Governance

*My domain is
never on data
quality scope*

You scale at enterprise level

Observability

*We are not
reactive on data
issue*

You react to change

AI Agent

*What if I want to
innovate?*

You can innovate everyday

Solutions

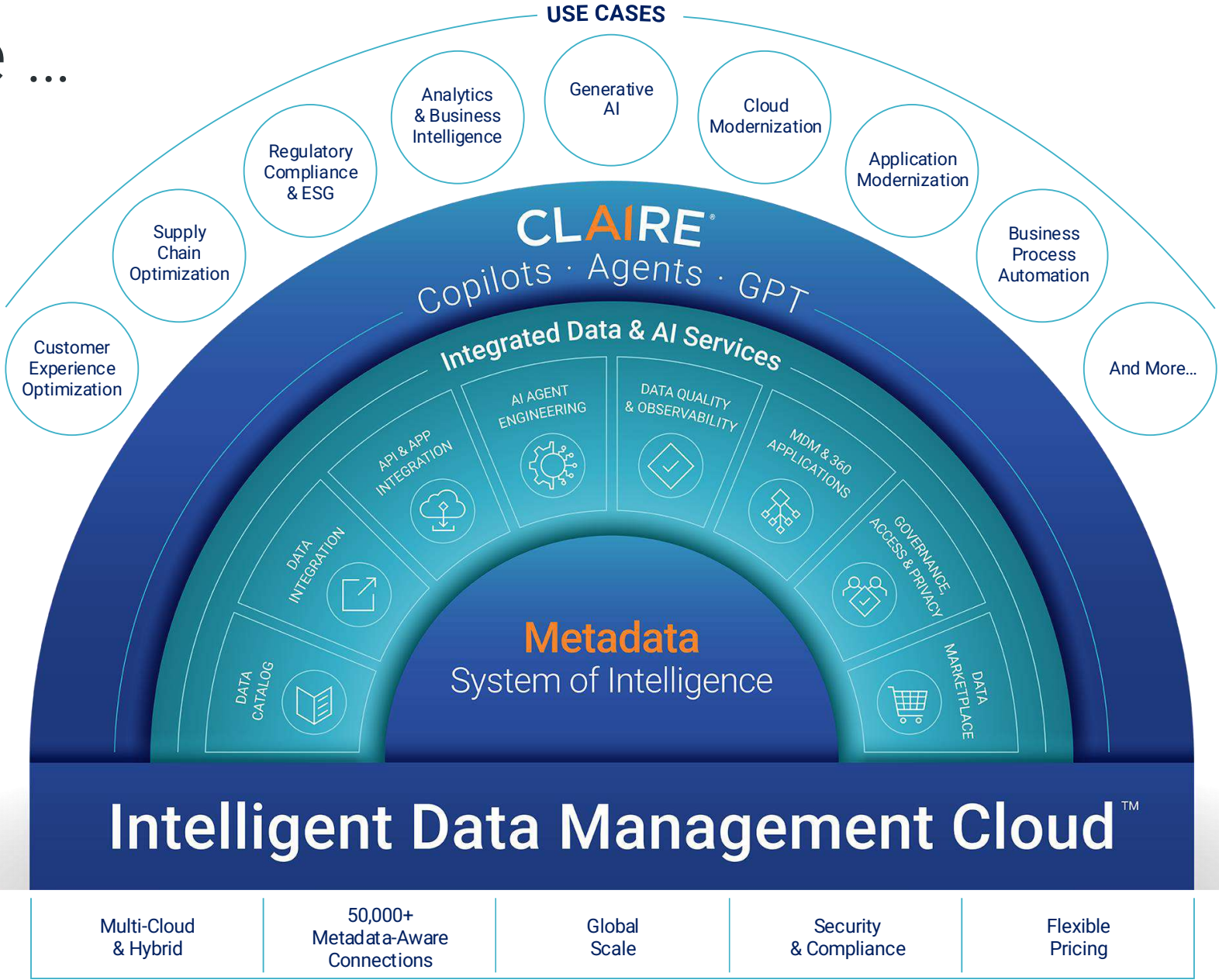
- Profiling
- Cleansing
- Matching & Parsing

- Metrics
- Data Lineage
- Verification

- Alerting
- Anomaly detection

- Auto Rule Generation

And more ...



Where data & AI come to

