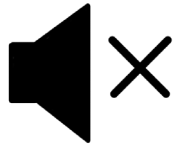


July 25, 2023

# Data Observability Framework

- Mayank Kumar Shrivastava, Principal Consultant, IPS

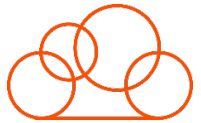
# Housekeeping Tips



- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available on our **Success Portal** - where you can also download the **slide deck** for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

# Feature Rich Success Portal

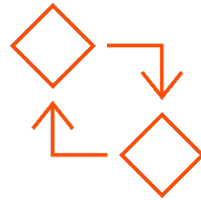
---



Bootstrap trial and  
POC Customers



Enriched Customer  
Onboarding  
experience



Product Learning  
Paths and Weekly  
Expert Sessions



Informatica  
Concierge



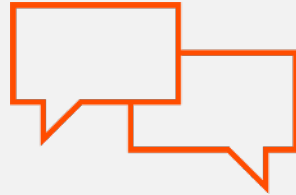
Tailored training and  
content  
recommendations

# More Information



## Success Portal

<https://success.informatica.com>



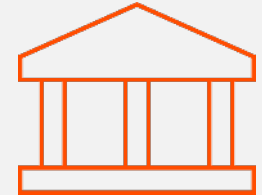
## Communities & Support

<https://network.informatica.com>



## Documentation

<https://docs.informatica.com>



## University

<https://www.informatica.com/in/services-and-training/informatica-university.html>

# Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

July 25, 2023

# Data Observability Framework

- Mayank Kumar Shrivastava, Principal Consultant, IPS

# Agenda

1

**Solution Overview**

2

**Key Design Principles**

3

**Technical  
Architecture**

4

**Solution High Level  
Steps**

5

**Sample Output and  
DQ Dashboard**

6

**Demo**

# Solution Overview

- This solution requires Informatica Intelligent Data Management Cloud (IDMC) platform.
- Here we are leveraging IDMC platform features like Mapplets, Mappings, Task Flow, Guide and others to create reusable, parameterized, meta driven and scalable solution to generate Data Quality results and perform REST API call to push the Data Quality scores to Informatica Cloud Data Governance & Catalog (CDGC).
- Once the solution is published, it can be scheduled to run on periodic basis like Daily, Monthly, etc or Ad Hoc as needed.
- Generated Data Quality results can then be exposed to 3<sup>rd</sup> party Business Intelligence tools like Tableau, Power BI, Qlik etc. which can help with creating customized dashboards and reports.



# Key Design Principles

**Metadata Driven**

**Reusability & Ease of  
Maintenance**

**Scalable**

**Parallelism**

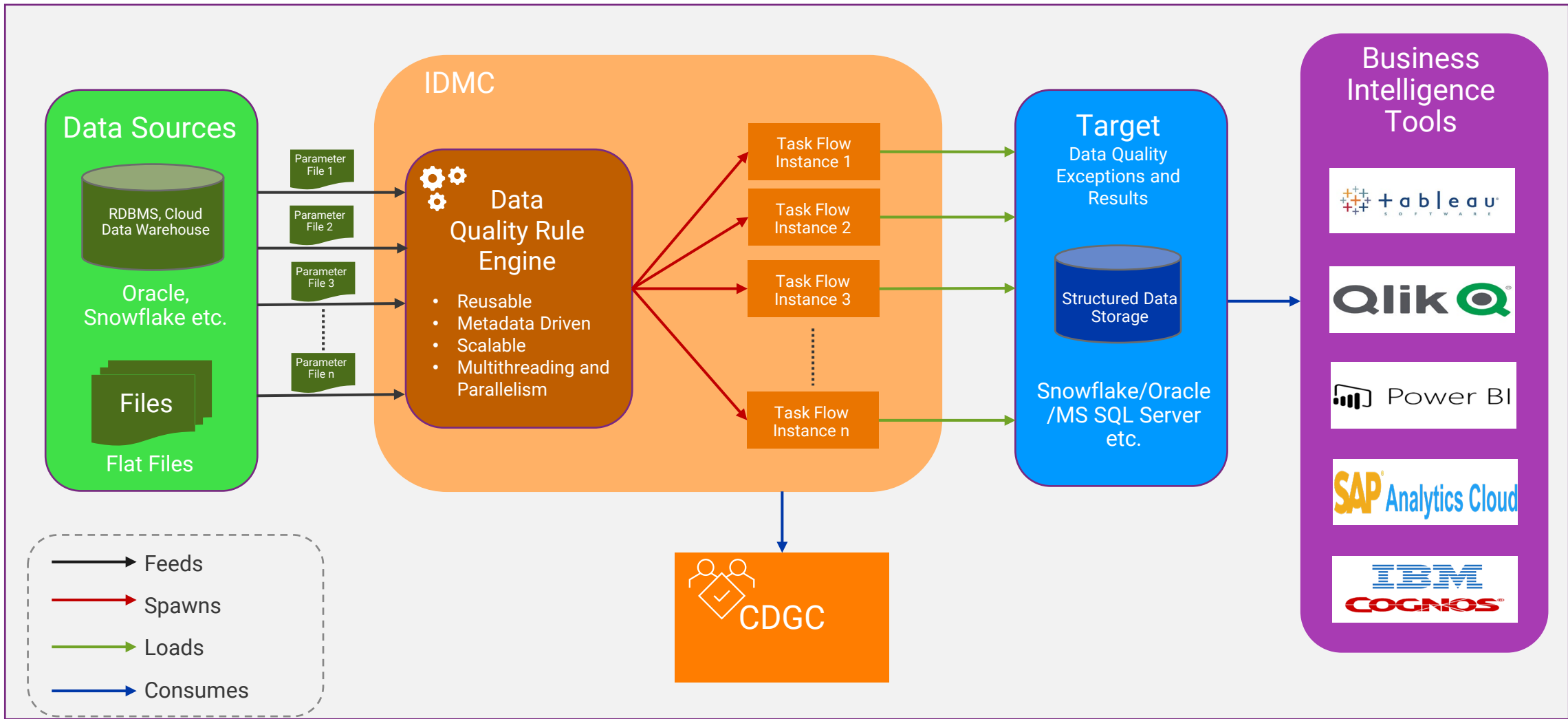
**Performance**

**Archival & Purging**

**Auditing**

**Failure Alerting**

# Technical Architecture



# Solution High Level Implementation Steps

- **Step-1:** Capture Data Quality rules and its related business assets like System, Data Set, Data Element, Glossary etc. in Informatica Cloud Data Governance & Catalog (CDGC).
- **Step-2:** Create reusable Data Quality Rules as Maplets in Informatica Cloud Data Integration.
- **Step-3:** Create parameterized Mapping which read data from source table/file and generates Data Quality result table in Informatica Cloud Data Integration.
- **Step-4:** Create a mapping that reads Data Quality results table data and generate JSON file containing required details like Execution Time, DQ Score, Exception Count, Total Count, Identity etc. to push DQ Results to Informatica CDGC.

# Solution High Level Implementation Steps

- **Step-5:** Create the logic to read JSON file data and make an API call to Informatica CDGC to push the DQ Results.
- **Step-6:** Orchestrate all created Mappings and other logics like Notifications in Informatica Cloud Data Integration and publish it.
- **Step-7:** Create APP Connection and Service Connector in Informatica Cloud Application Integration.
- **Step-8:** Create Guide in Informatica Cloud Application Integration to provide user interface to seed or execute.

# Sample Execution Output

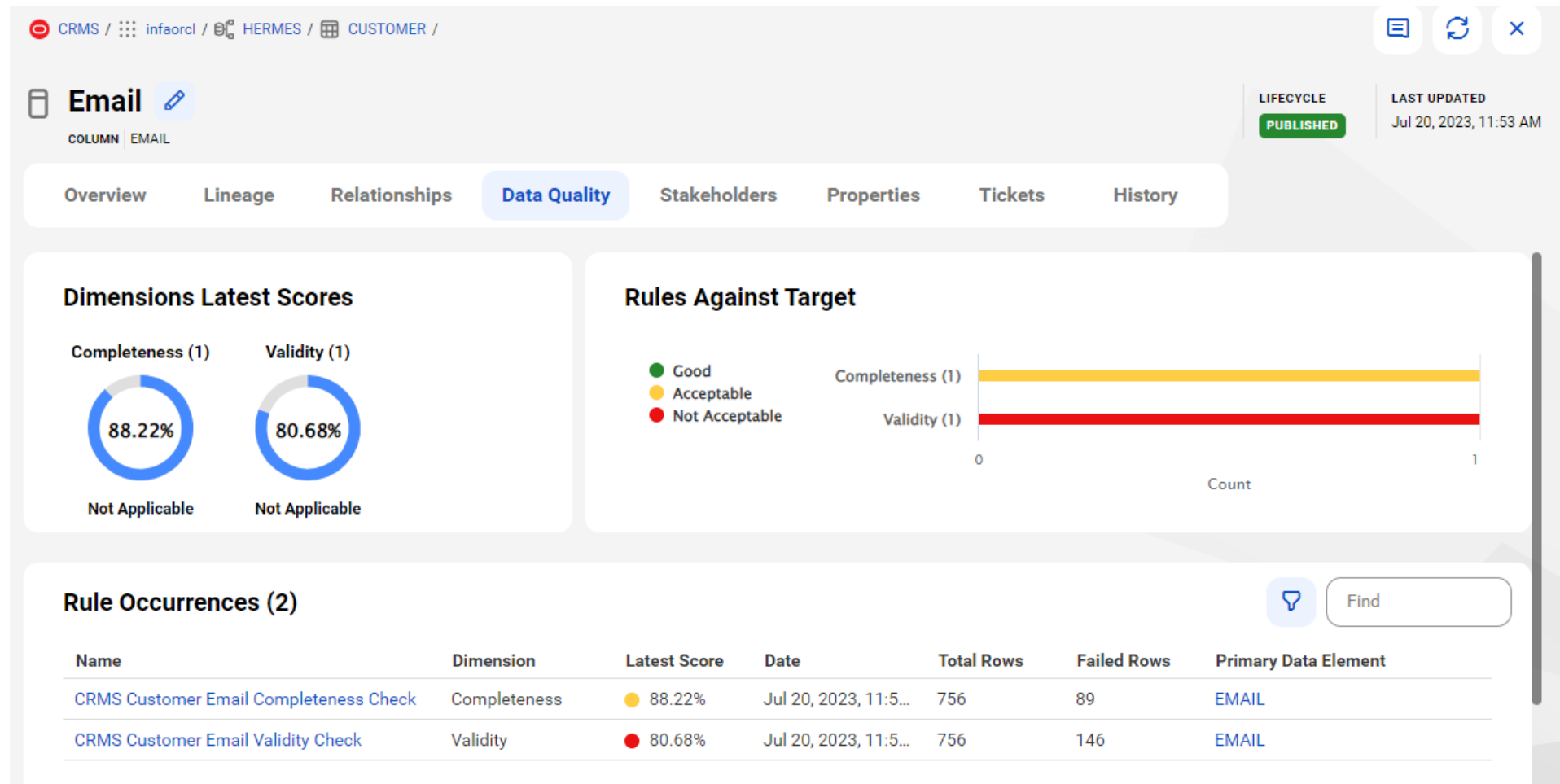
## Sample Data Quality Result Table

ABC SYSTEM_NAME	ABC DATA_SET_NAME	ABC ATTRIBUTE_NAME	ABC DQ_RULE_NAME	ABC DQ_DIMENSION	ABC INVALID_COUNT	ABC VALID_COUNT	ABC TOTAL_COUNT	ABC DQ_SCORE	ABC EXECUTION_DATE
CRMS	CUSTOMER	EMAIL	Customer_Email_Completeness_Check_1	Completeness	89	667	756	88.2275	2023_05_05
CRMS	CUSTOMER	EMAIL	Customer_Email_Validity_Check_1	Validity	146	610	756	80.6878	2023_05_05
CRMS	CUSTOMER	GENDER	Customer_Gender_Completeness_Check_1	Completeness	742	14	756	1.8519	2023_05_05
CRMS	CUSTOMER	PARTY_ID	Customer_ID_Completeness_Check_1	Completeness	2	754	756	99.7354	2023_05_05
CRMS	CUSTOMER	PARTY_ID	Customer_ID_Validity_Check_1	Validity	0	754	754	100	2023_05_05

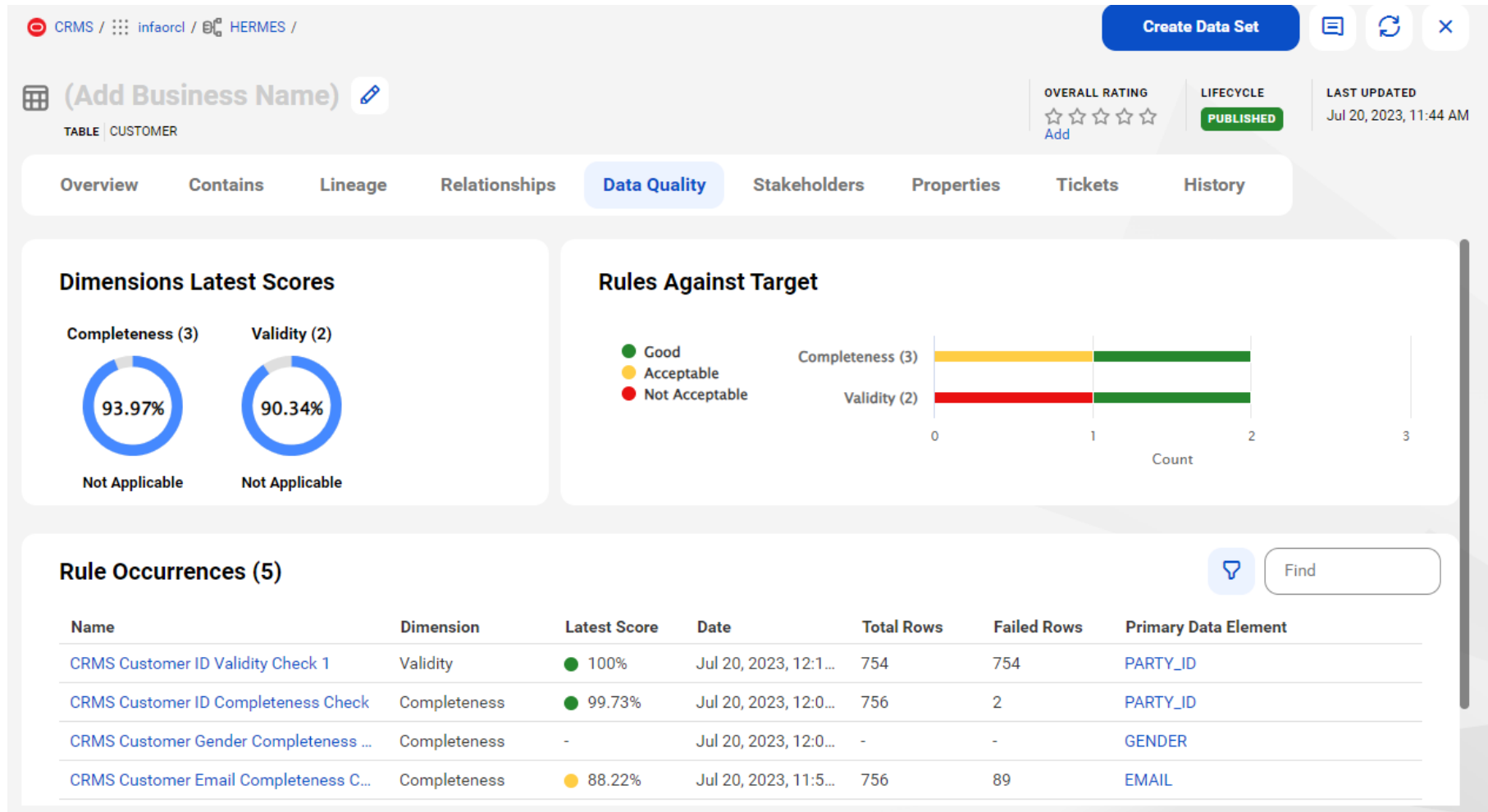
## Sample Data Quality Exception Table

	ABC SYSTEM_NAME	ABC DATA_SET_NAME	ABC ATTRIBUTE_NAME	ABC ATTRIBUTE_VALUE	ABC DQ_RULE_NAME	ABC DQ_DIMENSION	ABC EXECUTION_DATE
1	CRMS	CUSTOMER	GENDER	[NULL]	Customer_Gender_Completeness_Check_1	Completeness	2023_05_05
2	CRMS	CUSTOMER	EMAIL	[NULL]	Customer_Email_Completeness_Check_1	Completeness	2023_05_05
3	CRMS	CUSTOMER	EMAIL	AnnikaLarsson@ehptcom	Customer_Email_Validity_Check_1	Validity	2023_05_05
4	CRMS	CUSTOMER	EMAIL	BradPoorman@sts.com 915/677.1741	Customer_Email_Validity_Check_1	Validity	2023_05_05
5	CRMS	CUSTOMER	EMAIL	EugenEastJackson@worldafricancom	Customer_Email_Validity_Check_1	Validity	2023_05_05
6	CRMS	CUSTOMER	EMAIL	GregYancey@stsc.com	Customer_Email_Validity_Check_1	Validity	2023_05_05
7	CRMS	CUSTOMER	EMAIL	a&p@aptea.com	Customer_Email_Validity_Check_1	Validity	2023_05_05
8	CRMS	CUSTOMER	EMAIL	an@brookshire brothers.com	Customer_Email_Validity_Check_1	Validity	2023_05_05
9	CRMS	CUSTOMER	EMAIL	brookshire brothers@brookshire-brothers.com	Customer_Email_Validity_Check_1	Validity	2023_05_05
10	CRMS	CUSTOMER	EMAIL	c town@ctownsupermarkets.com	Customer_Email_Validity_Check_1	Validity	2023_05_05

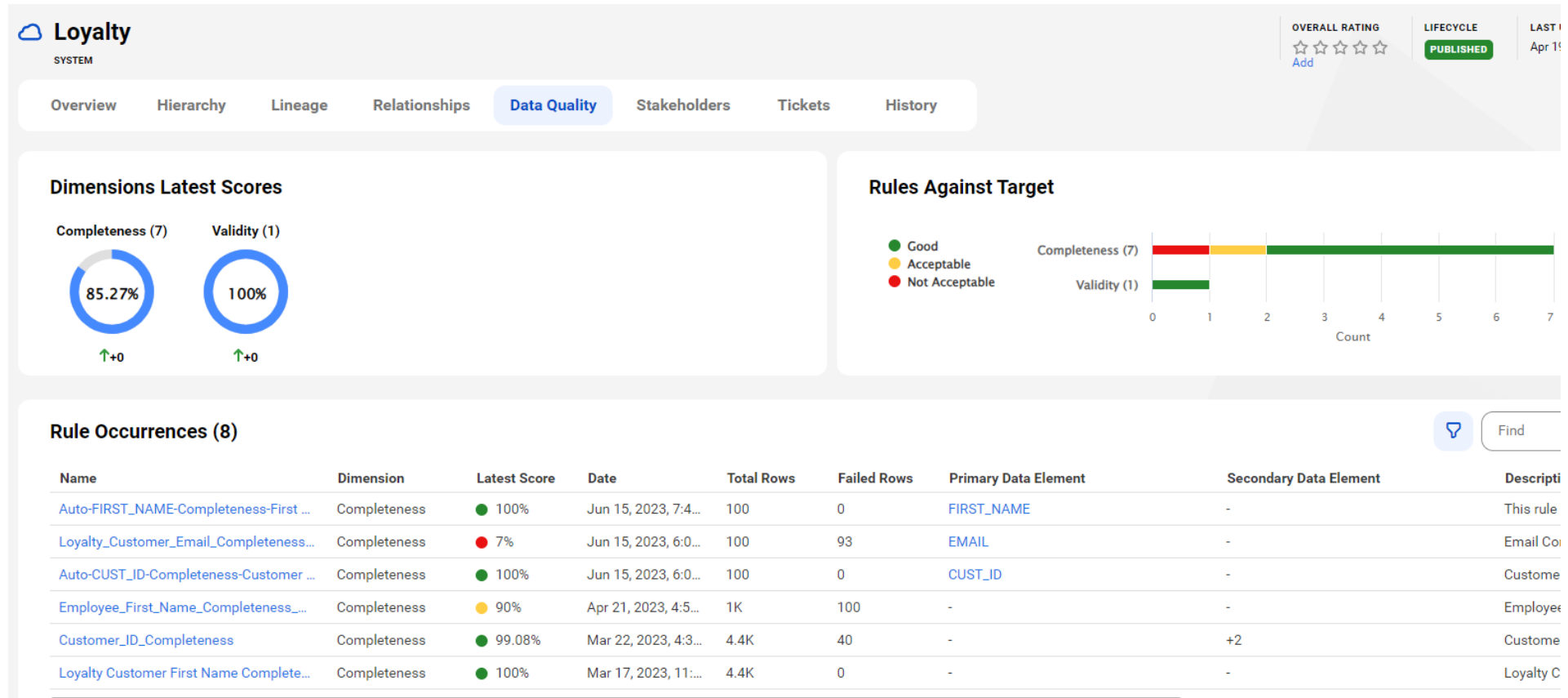
# Sample CDGC DQ Dashboard – Data Element



# Sample CDGC DQ Dashboard – Table

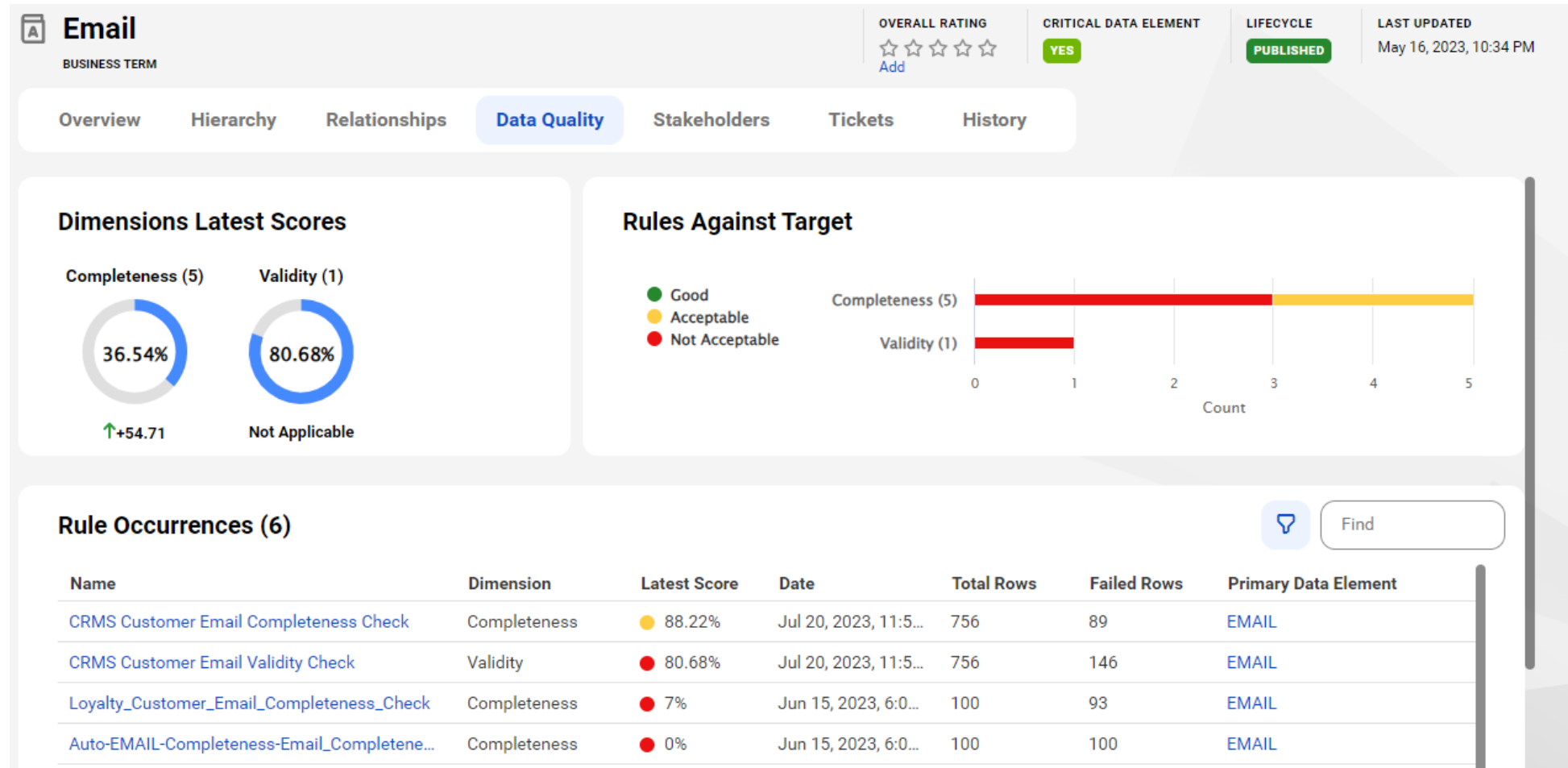


# Sample CDGC DQ Dashboard – System





# Sample CDGC DQ Dashboard – Glossary



# Demo

Thank you