

AI-Powered Data Excellence

with Informatica & Salesforce

Jonathon Bowring, Strategic Solution Engineer

jbowring@salesforce.com



Forward-looking statements



This presentation contains forward-looking statements about, among other things, trend analyses and statements regarding future events, anticipated growth and industry prospects, and our strategies, expectation or plans regarding product releases and enhancements. The achievement or success of the matters covered by such forward-looking statements involves risks, uncertainties and assumptions. If any such risks or uncertainties materialize or if any of the assumptions prove incorrect, results or outcomes could differ materially from those expressed or implied by these forward-looking statements. The risks and uncertainties referred to above include those factors discussed in Salesforce's reports filed from time to time with the Securities and Exchange Commission, including, but not limited to our ability to meet the expectations of our customers; uncertainties regarding AI technologies and their integration into our product offerings; the effect of evolving domestic and foreign government regulations; regulatory developments and regulatory investigations involving us or affecting our industry; our ability to successfully introduce new services and product features, including related to AI and Agentforce; our ability to execute our business plans; the pace of change and innovation and our ability to compete in the markets in which we participate; and our ability to maintain and enhance our brands.

This is Me

salesforce



Jonathon Bowring

Strategic Solution Engineer

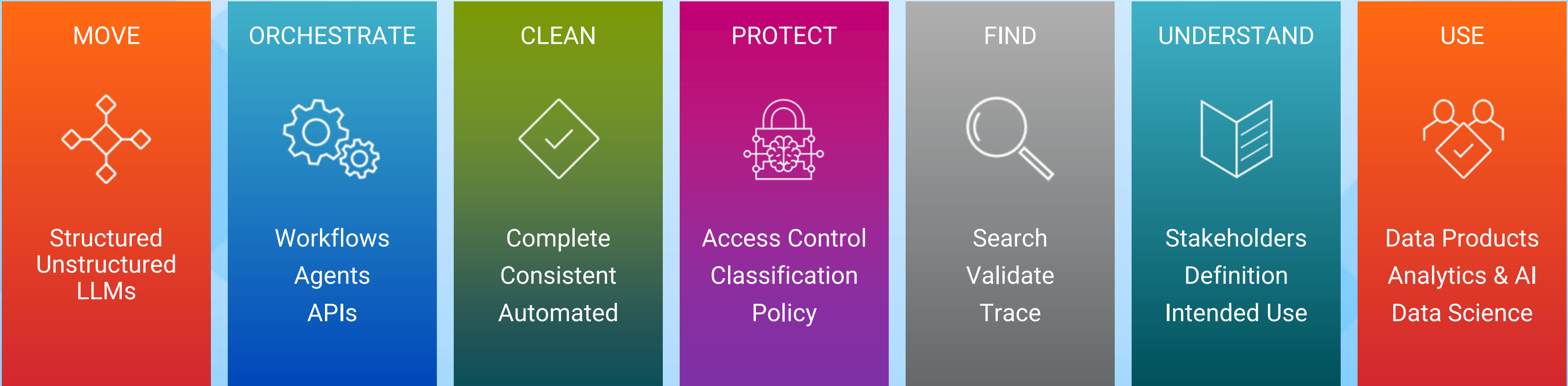
Using the attached image please generate an “This is Me” image that I can use for my presentation introduction.....



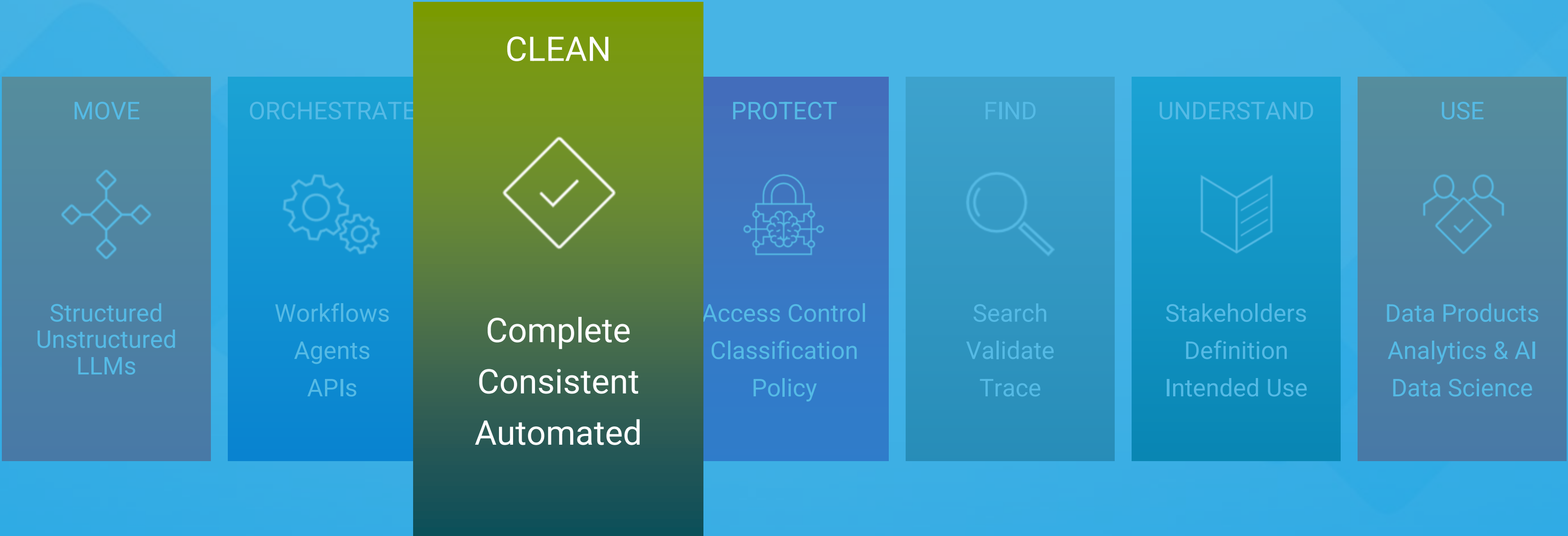
“Nearly eight in ten companies report using gen AI—yet just as many report no significant bottom-line impact. Think of it as ‘The Gen AI Paradox.’”

McKinsey Technology

The Data Value Chain



The Data Value Chain





SAP

SAP

SAP

SAP

SAP

**Trusted, high-quality
data is key
to ensuring ERP
projects are on-time
and within budget**





	A	B	C	D
1	TABLE_ENTITY	CRITICAL DATA ELEMENTS	CRITICAL DATA NAMES	BUSINESS_RULE_DESCRIPTION
2	EMPLOYEE_MASTER	COST_CTR	Cost Centre	
3	EMPLOYEE_MASTER	DOB	Date of Birth	
4	EMPLOYEE_MASTER	DEPT_CD	Department Code	
5	EMPLOYEE_MASTER	EMAIL	Email	
6	EMPLOYEE_MASTER	EMP_ID	Employee ID	
7	EMPLOYEE_MASTER	EMPL_STAT	Employee Status	
8	EMPLOYEE_MASTER	FST_NM	First Name	
9	EMPLOYEE_MASTER	FTE	Full Time Employee?	
10	EMPLOYEE_MASTER	HIR_DT	Hire Date	
11	EMPLOYEE_MASTER	JOB_CD	Job Code	
12	EMPLOYEE_MASTER	LST_NM	Last Name	
13	EMPLOYEE_MASTER	LOC_CD	Location Code	
14	EMPLOYEE_MASTER	LST_UPDT	LST_UPDT	
15	EMPLOYEE_MASTER	MGR_EMP_ID	Manager Employee ID	
16	EMPLOYEE_MASTER	PAY_GRD	Pay Grade	
17	EMPLOYEE_MASTER	SSN	Social Security Number	
18	EMPLOYEE_MASTER	TRM_DT	Termination Date	
19	PAYROLL_RECORDS	GROSS_PAY	Gross Pay	
20	PAYROLL_RECORDS	HRLY_RT	Hourly Rate	
21	PAYROLL_RECORDS	HRS_WRKD	Hours Workday	
22	PAYROLL_RECORDS	PAY_PRD_DT	Payroll Date	
23	PERFORMANCE_REVIEWS	GOALS_MET	Goals	
24	PERFORMANCE_REVIEWS	PERF_RTG	Performance Rating	
25	PERFORMANCE_REVIEWS	PERF_SCR	Performance Score	
26	PERFORMANCE_REVIEWS	REV_CMPLT	Review Complete	
27	PERFORMANCE_REVIEWS	REV_DT	Review Date	
28	PERFORMANCE_REVIEWS	RVWR_EMP_ID	Reviewer Employee ID	

???

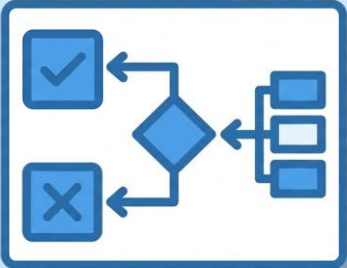




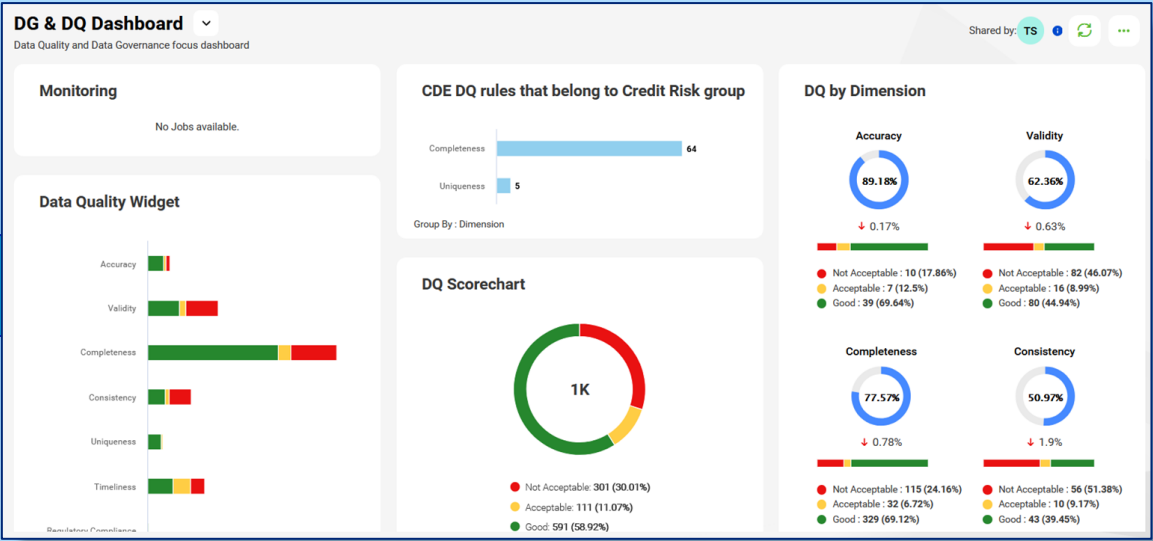
	A	B	C	D
1	TABLE ENTITY	CRITICAL DATA ELEMENTS	CRITICAL DATA NAMES	BUSINESS RULE DESCRIPTION
2	EMPLOYEE_MASTER	COST_CTR	Cost Centre	
3	EMPLOYEE_MASTER	DOB	Date of Birth	
4	EMPLOYEE_MASTER	DEPT_CD	Department Code	
5	EMPLOYEE_MASTER	EMAIL	Email	
6	EMPLOYEE_MASTER	EMPL_ID	Employee ID	
7	EMPLOYEE_MASTER	EMPL_STAT	Employee Status	
8	EMPLOYEE_MASTER	FST_NM	First Name	
9	EMPLOYEE_MASTER	FTE	Full Time Employee?	
10	EMPLOYEE_MASTER	HIR_DT	Hire Date	
11	EMPLOYEE_MASTER	JOB_CD	Job Code	
12	EMPLOYEE_MASTER	LST_NM	Last Name	
13	EMPLOYEE_MASTER	LOC_CD	Location Code	
14	EMPLOYEE_MASTER	LST_UPDT	LST_UPDT	
15	EMPLOYEE_MASTER	MGR_EMP_ID	Manager Employee ID	
16	EMPLOYEE_MASTER	PAY_GRD	Pay Grade	
17	EMPLOYEE_MASTER	SSN	Social Security Number	
18	EMPLOYEE_MASTER	TRM_DT	Termination Date	
19	PAYROLL_RECORDS	GROSS_PAY	Gross Pay	
20	PAYROLL_RECORDS	HRLY_RT	Hourly Rate	
21	PAYROLL_RECORDS	HRS_WRKD	Hours Workday	
22	PAYROLL_RECORDS	PAY_PRD_DT	Payroll Date	
23	PERFORMANCE_REVIEWS	GOALS_MET	Goals	
24	PERFORMANCE_REVIEWS	PERF_RTG	Performance Rating	
25	PERFORMANCE_REVIEWS	PERF_SCR	Performance Score	
26	PERFORMANCE_REVIEWS	REV_CMPLT	Review Complete	
27	PERFORMANCE_REVIEWS	REV_DT	Review Date	
28	PERFORMANCE_REVIEWS	RVWR_EMP_ID	Reviewer Employee ID	



IF KNA1.KTOKD == "0001" AND
KNKK.KLIMK < 10,000 AND
KNVV.ZTERM <> "D30"
THEN
 "Invalid"
ELSE
 "Valid"
END IF



"A sold-to customer with a
credit limit of up to \$10K
must use 30 day payment
terms"



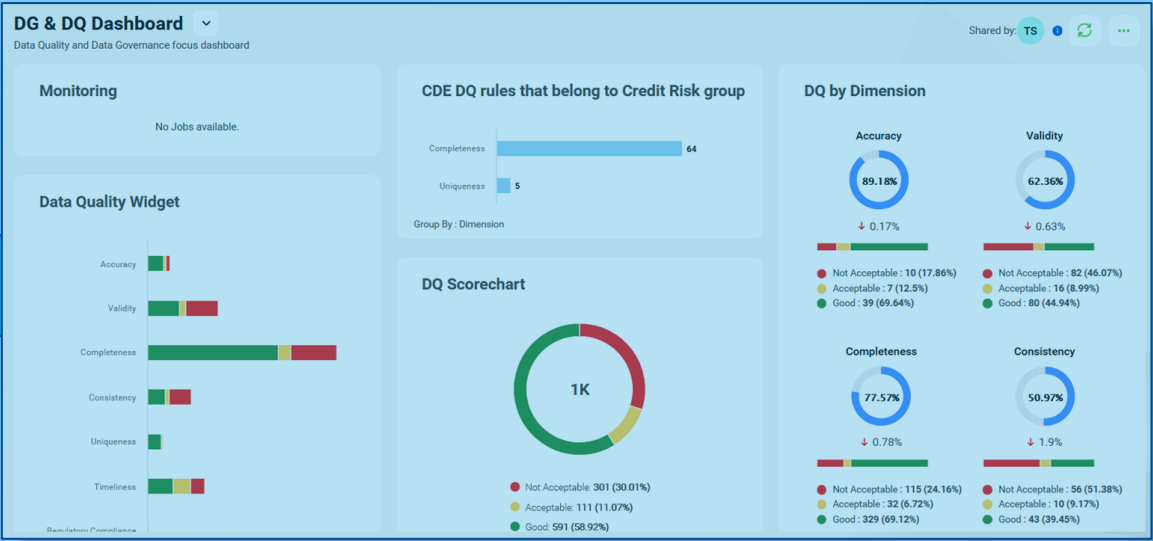


	A	B	C	D
1	TABLE ENTITY	CRITICAL DATA ELEMENTS	CRITICAL DATA NAMES	BUSINESS RULE DESCRIPTION
2	EMPLOYEE_MASTER	COST_CTR	Cost Centre	
3	EMPLOYEE_MASTER	DOB	Date of Birth	
4	EMPLOYEE_MASTER	DEPT_CD	Department Code	
5	EMPLOYEE_MASTER	EMAIL	Email	
6	EMPLOYEE_MASTER	EMPL_ID	Employee ID	
7	EMPLOYEE_MASTER	EMPL_STAT	Employee Status	
8	EMPLOYEE_MASTER	FST_NM	First Name	
9	EMPLOYEE_MASTER	FTE	Full Time Employee?	
10	EMPLOYEE_MASTER	HIR_DT	Hire Date	
11	EMPLOYEE_MASTER	JOB_CD	Job Code	
12	EMPLOYEE_MASTER	LST_NM	Last Name	
13	EMPLOYEE_MASTER	LOC_CD	Location Code	
14	EMPLOYEE_MASTER	LST_UPDT	LST_UPDT	
15	EMPLOYEE_MASTER	MGR_EMP_ID	Manager Employee ID	
16	EMPLOYEE_MASTER	PAY_GRD	Pay Grade	
17	EMPLOYEE_MASTER	SSN	Social Security Number	
18	EMPLOYEE_MASTER	TRM_DT	Termination Date	
19	PAYROLL_RECORDS	GROSS_PAY	Gross Pay	
20	PAYROLL_RECORDS	HRLY_RT	Hourly Rate	
21	PAYROLL_RECORDS	HRS_WRKD	Hours Workday	
22	PAYROLL_RECORDS	PAY_PRD_DT	Payroll Date	
23	PERFORMANCE_REVIEWS	GOALS_MET	Goals	
24	PERFORMANCE_REVIEWS	PERF_RTG	Performance Rating	
25	PERFORMANCE_REVIEWS	PERF_SCR	Performance Score	
26	PERFORMANCE_REVIEWS	REV_CMPLT	Review Complete	
27	PERFORMANCE_REVIEWS	REV_DT	Review Date	
28	PERFORMANCE_REVIEWS	RVWR_EMP_ID	Reviewer Employee ID	

IF KNA1.KTOKD == "0001" AND
KNKK.KLIMK < 10,000 AND
KNVV.ZTERM <> "D30"
THEN
 "Invalid"
ELSE
 "Valid"
END IF

A sold to customer with a
credit limit of up to \$10K
must use 30 day payment

8 Weeks to Establish a Baseline...





~22K

Invalid
Customers

Asset Value Up to
~\$52.8M

2-4

Cylinders Each

\$600

Per Cylinder

The Result!

salesforce



AI Data Quality Agent

Demonstration





Does CLAIRE GPT share data with other organizations?

Ask CLAIRE...



 Master Data Management



Stay updated with the latest capabilities of data management with AI.



Learn about the power of CLAIRE. Join our community now and begin learning from our experts and your peers!



Discover helpful tips and information to quickly perform tasks in CLAIRE GPT.





CLAIRE™ GPT

Agentic Data Management

Help me understand the security standards followed by I...

Show an overview of the <BE_name> business entity.

Show all <BE_name> business entity records that belong ...

Give insights on the type of sources and targets used in ...

Analyze the critical columns of the <table name> table to...

Does CLAIRE GPT share data with other organizations?

Create data quality rules using the attached business rule specifications



contact_dq_business_rules....



CLAIRE Security and Privacy



Data Analysis



Data Quality



Master Data Management



What's New in CLAIRE GPT

Stay updated with the latest capabilities of data management with AI.



Join CLAIRE Community

Learn about the power of CLAIRE. Join our community now and begin learning from our experts and your peers!



Quick Reference

Discover helpful tips and information to quickly perform tasks in CLAIRE GPT.



Informatica

CLAIRE GPT

Jon Bowring Suborg

Create data quality rules using the attached business rule specifications

Last modified: 12/07/2025, 06:53 PM

Extracted Data Quality Rules (29)

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	Contact_Must_Have_Unique_ID	Each contact must have a unique contact identifier	UNIQUENESS	• CONTACT_ID/string	
2	Contact_Must_Have_First_Name	All contacts must have a first name recorded	COMPLETENESS	• FIRST_NAME/string	
3	Contact_Must_Have_Last_Name	All contacts must have a last name recorded	COMPLETENESS	• LAST_NAME/string	
4	Contact_Must_Have_Valid_Email	Each contact must have a valid email address	ACCURACY	• EMAIL/string	
5	Contact_Must_Have_Phone_Number	All contacts should have a phone number recorded	COMPLETENESS	• PHONE_NUMBER/string	
6	Contact_Email_Must_Be_Unique	Each contact must have a unique email address	UNIQUENESS	• EMAIL/string	
7	Timely_LAST_UPDATED	Contact records should be updated timely	ACCURACY	• LAST_UPDATED/date/time	
8	Contact_Must_Have_Valid_Contact_Status	Contact must have a valid status indicating active/inactive	COMPLETENESS	• CONTACT_STATUS/string	
9	LAST_CONTACT_DT_If_Inactive	Inactive contacts must have a documented last contact date	CONSISTENCY	• LAST_CONTACT_DT/date/time	
10	Contact_Must_Have_Unique_Email	Each contact must have a unique email address	UNIQUENESS	• EMAIL/string	
11	Contact_Must_Have_Valid_Phone_Number	Phone number must be in valid national or international format	VALIDITY	• PHONE_NUMBER/string	
12	Contact_Must_Have_Non_Empty_Address	Address must have primary line, city, and postal code populated	COMPLETENESS	• ADDRESS_LINE1/string • CITY/string • ZIP_CODE/string	

Showing 29 of 29

contact identifier (CONTACT_ID).

Contact_Email_Must_Be_Unique:

Each contact must have a unique email address (EMAIL).

Completeness

Contact_Must_Have_First_Name:

All contacts must have a first name recorded (FIRST_NAME).

Contact_Must_Have_Non_Empty_Address:

Address must have primary line, city, and postal code populated (ADDRESS_LINE1, CITY, ZIP_CODE).

Accuracy

Contact_Must_Have_Valid_Email:

Each contact must have a valid email address (EMAIL).

Contact_Email_Bounce_Rate_Must_Be_Low:

Contacts with repeated email bounces should be flagged for review (EMAIL_BOUNCE_FLAG).

Validity

Contact_Must_Have_Valid_Phone_Number:

Phone number must be in valid national or international format (PHONE_NUMBER).

Contact_Postal_Code_Must_Be_Valid:

Postal code must match format and valid codes from postal reference (ZIP_CODE).

Consistency

LAST_CONTACT_DT_If_Inactive:

Inactive contacts must have a documented last contact date (LAST_CONTACT_DT).

Contact_Date_Created_Must_Be_In_Past:

Contact created date cannot be in the future (CREATED_DATE).

Extracted Data Quality Rules (29)

SAMPLE DQ RULES

29 DQ rules identified from contact_dq_business_rules.xlsx

Ask CLAIRE...

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

Informatica

CLAIRE GPT

Jon Bowring Suborg

Create data quality rules using the attached business rule specifications

Last modified: 12/07/2025, 06:53 PM

Extracted Data Quality Rules (29)

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	Contact_Must_Have_Unique_ID	Each contact must have a unique contact identifier	UNIQUENESS	• CONTACT_ID/string	
2	Contact_Must_Have_First_Name	All contacts must have a first name recorded	COMPLETENESS	• FIRST_NAME/string	
3	Contact_Must_Have_Last_Name	All contacts must have a last name recorded	COMPLETENESS	• LAST_NAME/string	
4	Contact_Must_Have_Valid_Email	Each contact must have a valid email address	ACCURACY	• EMAIL/string	
5	Contact_Must_Have_Phone_Number	All contacts should have a phone number recorded	COMPLETENESS	• PHONE_NUMBER/string	
6	Contact_Email_Must_Be_Unique	Each contact must have a unique email address	UNIQUENESS	• EMAIL/string	
7	Timely_LAST_UPDATED	Contact records should be updated timely	ACCURACY	• LAST_UPDATED/date/time	
8	Contact_Must_Have_Valid_Contact_Status	Contact must have a valid status indicating active/inactive	COMPLETENESS	• CONTACT_STATUS/string	
9	LAST_CONTACT_DT_If_Inactive	Inactive contacts must have a documented last contact date	CONSISTENCY	• LAST_CONTACT_DT/date/time	
10	Contact_Must_Have_Unique_Email	Each contact must have a unique email address	UNIQUENESS	• EMAIL/string	
11	Contact_Must_Have_Valid_Phone_Number	Phone number must be in valid national or international format	VALIDITY	• PHONE_NUMBER/string	
12	Contact_Must_Have_Non_Empty_Address	Address must have primary line, city, and postal code populated	COMPLETENESS	• ADDRESS_LINE1/string • CITY/string • ZIP_CODE/string	

Showing 29 of 29

matches already existing data quality rules to avoid generation of duplicate rules.

3

Cluster rules specifications

The step will check if some rules specification have identical logic and can be grouped into clusters.

4

Generate rules script

The step will generate a script for each rule specification.

5

Generate rules test data

The step will generate test data for each rule specification.

6

Test rules

The step will execute tests for each rule specification.

7

Create a dashboard

The step will provide summary of all outputs.

8

Persist rules

The step will publish rule specifications into CLAIRE project, CLAIRE folder.

👍👎

Cards (1)

Do you want to proceed with the following plan for generating data quality rules from your specifications? The steps are: 1) Rules specifications validation, 2) Rules deduplication, 3) Cluster rules specifications, 4) Generate rules script, 5) Generate rules test data, 6) Test rules, 7) Create a dashboard, 8) Persist rules. Please confirm or let me know if you want to include/exclude any steps.

Yes, proceed with the plan

No, modify the plan by:

Ask CLAIRE...

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

Informatica

CLAIRE GPT

Jon Bowring Suborg

Create data quality rules using the attached business rule specifications

Last modified: 12/07/2025, 06:55 PM

27 number of DQ rules clustered

Reasoning

Generated DQ Rule Scripts (27)

DQ RULES ARTIFACTS

CLAIRE identified and generated 27 new DQ rules from uploaded file.

Number of rules for which we are successfully able to generate script: 27 rules had scripts generated successfully.

Number of rules for which we failed to generate script: 0 rules failed script generation.

Reasoning

Generated DQ Rule Test Data (27)

DQ RULES ARTIFACTS

CLAIRE identified and generated 27 new DQ rules from uploaded file.

Generating

14 of 27 Completed

Ask CLAIRE...

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

Generated DQ Rule Test Data (27)

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

Rule Names

1

input_must_be_unique...

2

Contact_Must_Have_F...

3

Contact_Must_Have_L...

4

Contact_Must_Have_V...

5

Contact_Must_Have_P...

6

Timely_LAST_UPDATED

7

Contact_Must_Have_V...

8

LAST_CONTACT_DT_I...

9

Contact_Must_Have_V...

10

Contact_Must_Have_...

11

Contact Postal Cod...

Generating new rules

Contact_Must_Have_Valid_Phone_Number

Name:

Contact_Must_Have_Valid_Phone_Number

Description:

Phone number must be in valid national or international format

Dimension:

VALIDITY

Rule Code

IF REG_MATCH(PHONE_NUMBER, '^+?[0-9]{10,15}\$') THEN Output = 'VALID' ELSE Output = 'INVALID' ENDIF

Test Data

PHONE_NUMBER	Expected Output
	INVALID
1234567890	VALID
123456789012345	VALID
+123456789012345	VALID
123456789	INVALID
1234567890123456	INVALID
+12345abc901	INVALID
123-456-7890	INVALID
+12345678901	VALID

AI Data Cleansing Agent

Demonstration





CLAIRE™ GPT

Agentic Data Management

How do I create a page in MDM SaaS?

Show a data preview of the last 10 records from the <tab>

List the datasets that have poor data quality scores from...

List the assets with data quality issues by tracing the line...

Show all <BE_name> business entity records that don't b...

How does CLAIRE GPT handle hallucinations and bias?

Ask CLAIRE...

I



CLAIRE Security and Privacy



Data Governance



Data Analysis



Master Data Management



What's New in CLAIRE GPT

Stay updated with the latest capabilities of data management with AI.



Join CLAIRE Community

Learn about the power of CLAIRE. Join our community now and begin learning from our experts and your peers!



Quick Reference

Discover helpful tips and information to quickly perform tasks in CLAIRE GPT.



CLAIRE™ GPT

Agentic Data Management

How do I create a page in MDM SaaS?

Show a data preview of the last 10 records from the <tab>...

List the datasets that have poor data quality scores from...

List the assets with data quality issues by tracing the line...

Show all <BE_name> business entity records that don't b...

How does CLAIRE GPT handle hallucinations and bias?

Analyse, cleanse, standardise and remove duplicates in the attached file



contact_master....



CLAIRE Security and Privacy



Data Governance



Data Analysis



Master Data Management



What's New in CLAIRE GPT

Stay updated with the latest capabilities of data management with AI.



Join CLAIRE Community

Learn about the power of CLAIRE. Join our community now and begin learning from our experts and your peers!



Quick Reference

Discover helpful tips and information to quickly perform tasks in CLAIRE GPT.



Informatica

CLAIRE GPT

Jon Bowring Suborg

Analyse, cleanse, standardise and remove duplicates in the attached file

Last modified: 12/07/2025, 07:54 PM

contact_master

Data Preview

Privacy guidelines prevent us from storing sample data, but you can download the data for your reference.

	A CON_ID	A NAT_ID	A FST_NM	A LST_NM	A DOB	A CRTD_DT	A DEL_DT	A DE
1	CON48546	XXX-XX-2035	None	Miller	23/09/1966	6/05/2022	None	R&D
2	CON80711	XXX-XX-4706	Richard	Harris	6/03/1999	2/06/2000	22/06/2020	R&D
3	CON30664	XXX-XX-8740	Christopher	Lee	28/02/1995	14/06/2025	None	None
4	CON20418	XXX-XX-5802	Sarah	Williams	26/09/1993	17/06/2018	None	OPS
5	CON46531	XXX-XX-3344	Barbara	Williams	24/12/1966	18/09/2019	None	LGL
6	CON88539	XXX-XX-3755	Thomas	Johnson	24/07/1992	14/08/2014	11/09/2023	SLS
7	CON74040	XXX-XX-9967	Daniel	None	11/12/1983	20/10/2023	None	MKT
8	CON16059	None	Patricia	Martin	1/11/2003	18/03/2010	None	OPS
9	CON55844	XXX-XX-7474	Karen	Brown	23/02/1970	14/03/2015	None	OPS
10	CON18914	None	Linda	Davis	22/11/1987	29/01/2017	None	LGL
11	CON66578	XXX-XX-5978	Sarah	Martin	24/08/2004	None	None	SUP
12	CON17399	XXX-XX-6361	Nancy	Taylor	9/01/1968	11/05/2007	None	LGL
13	CON17399	XXX-XX-6361	Nancy	Taylor	9/01/1968	11/05/2007	None	LGL
14	CON73660	XXX-XX-4552	Elizabeth	Lee	5/02/1970	13/11/2015	None	SUP
15	CON31590	XXX-XX-5138	Daniel	Lee	17/04/1963	1/01/2025	None	HR
16	CON86843	XXX-XX-4111	Susan	Miller	12/12/1984	25/05/2013	3/04/2020	FIN
17	CON82959	XXX-XX-3829	Michael	Jones	2/03/1985	23/03/2024	None	MKT
18	CON83480	XXX-XX-6206	Richard	Davis	20/03/1994	9/04/2020	None	MKT
19	CON55534	XXX-XX-3407	Patricia	Martin	15/12/1968	4/02/2022	None	SLS
20	CON72919	XXX-XX-2323	None	Brown	22/08/1985	15/03/2006	None	IT
21	CON36232	XXX-XX-6590	David	Brown	29/03/1981	30/07/2018	None	HR
22	CON40500	XXX-XX-4014	Quinn	James	15/02/1990	10/02/2020	24/02/2025	OPS

Showing 270 of 270 Data Types (18): 17 String, 1 Other

Data preview of 270 records from the dataset

> Reasoning

The data profile has been successfully generated for the dataset. This analysis is based on a sample of 270 records.

Key Insights

- High null counts in CRTD_DT (188), DEL_DT (237), and LST_UPDT (180) suggest incomplete temporal tracking for many records.
- CON_ID shows high uniqueness (258 distinct values) indicating it likely serves as a primary identifier for contacts.
- The LVL column has only 3 distinct values with an average of 0.83, suggesting a simple hierarchical classification system.
- LOC_CD has only 6 distinct values with SEA location having the highest frequency (45), indicating geographic concentration.
- CON_STAT shows low variety (3 values) with most records (177) having "ACT" status, suggesting predominantly active contacts.
- EMAIL field has moderate completeness (83 nulls) with high uniqueness (155 distinct values), indicating personalized contact information.

Data Profile

TABULAR LIST

Profiling statistics of the dataset.

Calling the planner to create a step by step plan

Ask CLAIRE...

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



Data preview of Z/U records from the dataset

➤ Reasoning

Key Insights

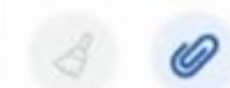
- High null counts in CRTD_DT (188), DEL_DT (237), and LST_UPDT (180) suggest incomplete temporal tracking for many records.
- CON_ID shows high uniqueness (258 distinct values) indicating it likely serves as a primary identifier for contacts.
- The LVL column has only 3 distinct values with an average of 0.83, suggesting a simple hierarchical classification system.
- LOC_CD has only 6 distinct values with SEA location having the highest frequency (45), indicating geographic concentration.
- CON_STAT shows low variety (3 values) with most records (177) having "ACT" status, suggesting predominantly active contacts.
- EMAIL field has moderate completeness (83 nulls) with high uniqueness (155 distinct values), indicating personalized contact information.

TABULAR LIST

Profiling statistics of the dataset.

- Calling the planner to create a step by step plan

Ask CLAIRE...



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

Informatica

CLAIRE GPT

Jon Bowring Suborg

Analyse, cleanse, standardise and remove duplicates in the attached file

Last modified: 12/07/2025, 07:55 PM

Column Level Enrichment for HR Personnel Dataset

#	Name	Data Type	Data Kind	Entity	Sensitivity	Generated Description
1	CON_ID	String	categorical	Contact/Employee Id...	Low	A unique alphanumeric identifier assigned to each individual in the dataset. Used as the primary key to...
2	NAT_ID	String	categorical	National Identifier	High	A partially masked national identification number (e.g., social security number or equivalent). Used f...
3	FST_NM	String	categorical	First Name	Medium	The given (first) name of the individual. Used for identification and communication. Some records have...
4	LST_NM	String	categorical	Last Name	Medium	The family (last) name of the individual. Used for identification and communication. Some records have...
5	DOB	Date	datetime	Date of Birth	High	The birth date of the individual, used for age calculation, eligibility, and verification. Sensitive personal...
6	CRTD_DT	Date	datetime	Record Creation Date	Low	The date when the individual's record was created in the system. Useful for tracking onboarding or data entry...
7	DEL_DT	Date	datetime	Record Deletion Date	Low	The date when the individual's record was marked as deleted or inactive. Indicates offboarding or contract...
8	DEPT_CD	String	categorical	Department Code	Low	A code representing the department or functional area (e.g., HR, IT, FIN, R&D) to which the individual belongs. Useful fo...
9	CON_CD	String	categorical	Contract Code	Low	A code indicating the type or category of contract or employment (e.g., SP1, MG2, EX1). Used for HR and payroll...
10	CON_STAT	String	categorical	Contract Status	Low	The current status of the individual's contract (e.g., ACT for active, TRM for terminated, LOA for leave of absence)...
11	CON_GRD	String	categorical	Contract Grade	Low	A code representing the grade or level of the individual's contract (e.g., G1...

Showing 18 of 18

Reasoning

1. Dataset Information

- Dataset Name: Not explicitly provided
- Description: This dataset contains detailed information about individuals, likely employees or contractors, within an organization. Each record represents a unique individual and includes identifiers (CON_ID, NAT_ID), personal information (first and last names, date of birth), employment details (department code, contract code, contract status, contract grade, level, reporting contact ID), organizational data (cost center, location code, country), and administrative metadata (creation date, deletion date, last update date, email). The dataset appears to track both current and former personnel, as indicated by contract status and deletion dates. Some columns contain missing values, and there are a small number of duplicate CON_IDs, suggesting that the dataset may include historical or updated records for the same individual. The data is suitable for HR analytics, workforce planning, and organizational reporting.
- Domain: Human Resources
- Primary Key: CON_ID

2.

Column Level Enrichment for HR Personnel Dataset

TABULAR LIST

Inferred metadata attributes for each column in the dataset.

Ask CLAIRE...

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

Informatica

CLAIRE GPT

Jon Bowring Suborg

Analyse, cleanse, standardise and remove duplicates in the attached file

Last modified: 12/07/2025, 07:55 PM

Critical Columns

#	Name	CLAIRE's Reasoning	Generated Description
1	CON_ID	Serves as the unique identifier for each individual in the dataset, enabling record distinction and linkage across HR processes, as highlighted in the enriched metadata and...	A unique alphanumeric identifier assigned to each individual in the dataset. Used as the primary key to distinguish records. While mostly unique, a few duplicates exist, possibly due to...
2	NAT_ID	Represents the national identification number, which is essential for uniquely identifying individuals at a national level and is critical for compliance, verification, and HR record...	A partially masked national identification number (e.g., social security number or equivalent). Used for uniquely identifying individuals at a national level. Contains sensitive personal...
3	FST_NM	First name is a core personal identifier used for communication, reporting, and individual recognition within HR and organizational processes.	The given (first) name of the individual. Used for identification and communication. Some records have missing values.
4	LST_NM	Last name is a fundamental personal identifier necessary for accurate identification, communication, and reporting in HR contexts.	The family (last) name of the individual. Used for identification and communication. Some records have missing values.
5	DOB	Date of birth is critical for age calculation, eligibility checks, and verification processes, supporting workforce planning and compliance.	The birth date of the individual, used for age calculation, eligibility, and verification. Sensitive personal information. Some records have missing values.
6	CRTD_DT	Indicates when a record was created, supporting onboarding tracking, data lineage, and audit trails in HR analytics.	The date when the individual's record was created in the system. Useful for tracking onboarding or data entry timelines. Many records have missing values, possibly for legacy data.
7	DEL_DT	Captures when a record is marked as deleted or inactive, which is essential for tracking offboarding, contract termination, and maintaining accurate workforce status.	The date when the individual's record was marked as deleted or inactive. Indicates offboarding or contract termination. Mostly null, suggesting most individuals are active or deletio...
8	DEPT_CD	Department code is vital for organizational structure analysis, workforce segmentation, and department-level reporting.	A code representing the department or functional area (e.g., HR, IT, FIN, R&D) to which the individual belongs. Useful for organizational structure analysis. Some records have missin...
9	CON_CD	Contract code classifies the type of employment or contract, which is necessary for HR, payroll, and workforce analytics.	A code indicating the type or category of contract or employment (e.g., SP1, MG2, EX1). Used for HR and payroll classification. Some records have missing values.
10	CON_STAT	Contract status provides the current employment state (active, terminated, leave of absence), which is fundamental for workforce status reporting and analytics.	The current status of the individual's contract (e.g., ACT for active, TRM for terminated, LOA for leave of absence). No missing values. Useful for workforce status reporting.
11	CON_GRD	Contract grade is used for compensation, hierarchy, and role analysis, supporting HR analytics and organizational planning.	A code representing the grade or level of the individual's contract (e.g., G1, G10). Used for compensation and hierarchy

Showing 18 of 18

• CON_GRD: Contract grade or level, used for compensation and hierarchy analysis.

• LVL: Employment level or FTE, indicates full-time equivalency.

• REP_CON_ID: Manager or reporting contact's identifier, establishes reporting lines.

• COST_CTR: Cost center code, used for financial tracking and budgeting.

• LOC_CD: Work location code, identifies primary work location.

• COUNTRY: Country of employment, includes both real and test data.

• EMAIL: Work email address, used for communication and is sensitive.

• LST_UPDT: Last update date, important for data maintenance and audit.

These columns are considered critical due to their roles in uniquely identifying individuals, managing sensitive information, and supporting key business processes.

Critical Columns

TABULAR LIST

Identified Critical Data Elements in the dataset.

Reasoning

Now that the critical data elements have been identified, I will assess the functional dependencies among these elements. This will help ensure data integrity and uncover any potential issues related to uniqueness or attribute relationships in the dataset.

Identifying the functional dependencies among the CDEs

Ask CLAIRE...

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

Informatica

CLAIRE GPT

Jon Bowring Suborg

Analyse, cleanse, standardise and remove duplicates in the attached file

Last modified: 12/07/2025, 07:55 PM

Functional Dependencies

#	Cluster	CLAIRE's Reasoning
1	- CON_ID - NAT_ID - EMAIL ...	CON_ID is the primary key and is intended to uniquely identify each individual. In practice, NAT_ID (national ID), EMAIL, and the combination of FST_NM, LST_NM, and DOB are also strong quasi-identifiers for a person. These columns are...
2	- CON_ID - CRTD_DT - DEL_DT ...	These columns are all record-level metadata that track the lifecycle of an individual's record: creation, deletion/inactivation, and last update. They are functionally dependent on the existence of a CON_ID and should be consistent for each unique...
3	- CON_ID - DEPT_CD - CON_CD ...	These columns describe the employment and organizational attributes of an individual at a given point in time. For a given CON_ID, these fields collectively define the employment context (department, contract type, status, grade, FTE level,...
4	- DEPT_CD - COST_CTR - LOC_CD ...	Department code, cost center, location code, and country are strongly interrelated in organizational structures. Typically, a cost center is assigned to a department and location, and both are situated within a country. This cluster is useful for validating...
5	- CON_CD - CON_GRD - CON_STAT ...	Contract code, contract grade, contract status, and employment level are all attributes of the employment contract. Certain contract codes only allow specific grades or statuses, and employment level (FTE) may be restricted by contract type...
6	REP_CON_ID	The reporting contact ID (manager's CON_ID) is a special field that establishes reporting lines. While it is functionally dependent on the individual's CON_ID, it also links to another CON_ID in the same dataset. This field is critical for validating...

Showing 6 of 6

Reasoning

The functional dependency clusters identified are:

- CON_ID, NAT_ID, EMAIL, FST_NM, LST_NM, DOB
- CON_ID, CRTD_DT, DEL_DT, LST_UPDT
- CON_ID, DEPT_CD, CON_CD, CON_STAT, CON_GRD, LVL, REP_CON_ID, COST_CTR, LOC_CD, COUNTRY
- DEPT_CD, COST_CTR, LOC_CD, COUNTRY
- CON_CD, CON_GRD, CON_STAT, LVL
- REP_CON_ID

These clusters represent groups of attributes that are functionally dependent on each other, as evidenced by the listed groupings.

Functional Dependencies

TABULAR LIST

Identified functional dependencies in the dataset.

Reasoning

Next, I will compute statistical data quality metrics for the dataset to identify issues such as missing values, duplicates, inconsistencies, and outliers. I will also propose specific fixes for each identified issue to improve the overall data quality.

- Identifying statistical data quality issues of the dataset

Ask CLAIRE...

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

Informatica

CLAIRE GPT

Jon Bowring Suborg

Analyse, cleanse, standardise and remove duplicates in the attached file

Last modified: 12/07/2025, 07:55 PM

Candidates for Standardization

#	Issue Name	Issue Description	Dimension	Fix Type	Samples Causing Issue	Reason To Fix
1	date_format_incon...	The DOB (date of birth) column contains multiple date formats, such as '23/09/1966', '6/03/1999', and...	Consistency	fixable	- CON_ID: CON48546, DOB: 23/09/1966 - CON_ID: CON80711, DOB: 6/03/1999 - CON_ID: CON16059, DOB: 1/11/2003	Consistent date formats are essential for accurate age calculation, eligibility checks,
2	date_format_incon...	The CRTD_DT (record creation date) column contains inconsistent date formats, e.g., '6/05/2022',...	Consistency	fixable	- CON_ID: CON48546, CRTD_DT: 6/05/2022 - CON_ID: CON30664, CRTD_DT: 14/06/2025 - CON_ID: CON80711, CRTD_DT: 2/06/2000	Standardizing date formats er reliable filtering, sorting, and t based analytics. Using a singl
3	date_format_incon...	The DEL_DT (deletion date) column shows inconsistent date formats, such as '22/06/2020', '11/09/2023',...	Consistency	fixable	- CON_ID: CON80711, DEL_DT: 22/06/2020 - CON_ID: CON88539, DEL_DT: 11/09/2023 - CON_ID: CON35971, DEL_DT: 1/01/2023	Consistent date formats are necessary for accurate offboz analysis and compliance repo
4	date_format_incon...	The LST_UPDT (last update date) column contains inconsistent date formats, such as '3/10/2024'...	Consistency	fixable	CON_ID: CON48546, LST_UPDT: 3/10/2024	Standardizing last update dat critical for audit trails and ens accurate record maintenance.

Showing 16 of 16

fictional country names, inconsistent abbreviations, and code variants or typos.

These issues are evidenced by the presence of multiple formats, inconsistent casing, and non-standardized labels in the sample values and descriptions provided for each column.

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

I will now compile a comprehensive final report summarizing all findings, quality assessments, proposed fixes, and the overall data cleansing strategy for your contact_master.csv file.

Preparing final report

Ask CLAIRE...

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

Analyse, cleanse, standardise and remove duplicates in the attached file Last modified: 12/07/2025, 07:55 PM

Agent Log of clean data for contact_master

Here's a comprehensive clean data quality report

Rules Overview

258

Sample Records

28

Total Rules

28

Fixable Rules

0

Non-Fixable Rules

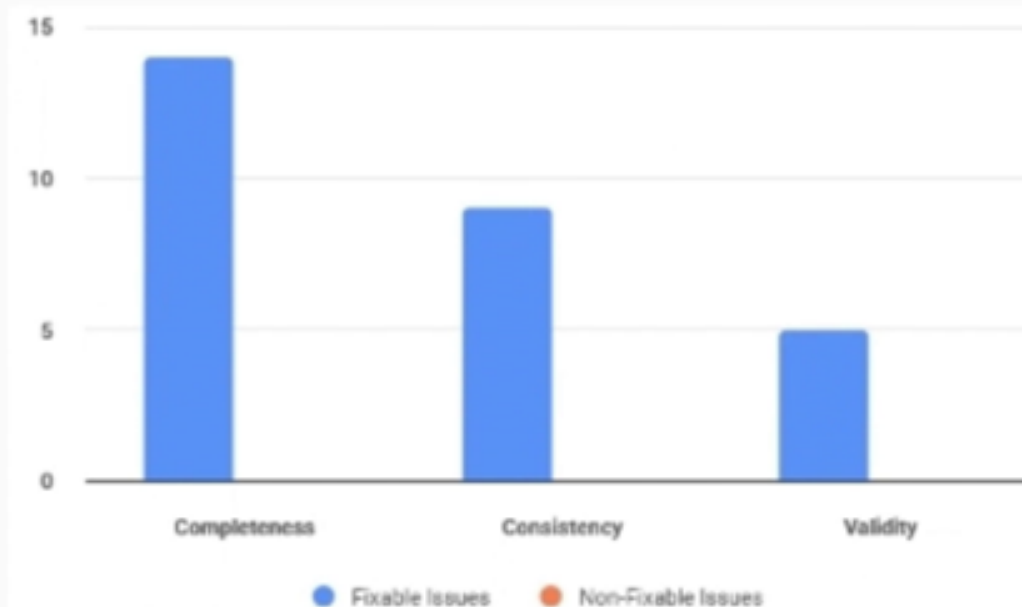
1659

Total Violations Fixed

35.72% Quality Improvement

Data Quality Dimension Overview

Data Quality Issues by Dimension



Data Quality Improvement by Dimension



Clean Data Sample contact_master

TABULAR LIST

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



Agent Log of clean data for contact_master

REPORT

Data quality analysis dashboard generated successfully.

> Reasoning



HR Contract Data Cleaning Maplet

MAPPLET ARTIFACT

Cleans and standardizes HR contract data for completeness, consistency, and validity.

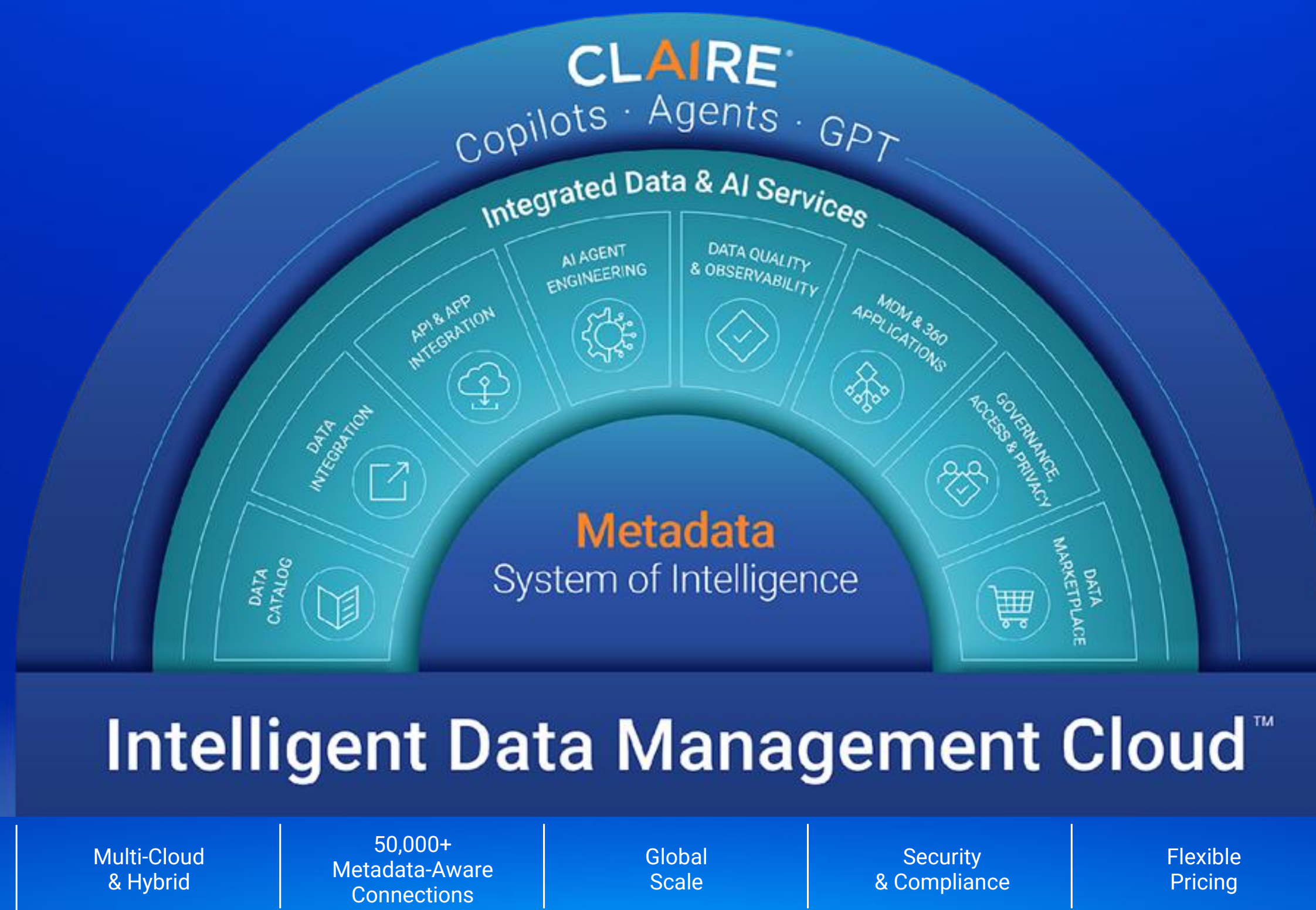
Analysis, cleansing, standardization, and deduplication of your contact_master.csv file are complete. A detailed report and a reusable maplet for future cleansing are now available.

Download the cleansed contact_master.csv file

View the data quality report for contact_master.csv

Ask CLAIRE...





Extend Your Agentic Enterprise



Data Cloud

Commerce Cloud

Analytics

Tableau

Slack

Marketing Cloud

Salesforce Chatter

Salesforce Pardot

Athena

Aurora

Aurora MySQL

Aurora PostgreSQL

DynamoDB

Redshift

SNS

SQS

Kinesis Firehose

Kinesis Streams

RDS MySQL (Community)

RDS Oracle

RDS PostgreSQL

RDS SQL Server

S3

Azure Blob Storage

Azure Data Factory

ADLS Gen1

ADLS Gen2

CDM Folders

Azure Synapse Script

Azure Synapse SQL

Fabric OneLake

Fabric DW

Azure EventHub

Cloud Storage

BigQuery

Google BigTable

Google Cloud Pub/Sub

Google Looker

Google Analytics

Google Sheets

Google Drive

Google APIs

Object Storage

Autonomous Database

Oracle Database

Oracle DB Cloud Service

Oracle Cloud Applications

snowflake

SAP

databricks

Anaplan

Xactly

SAP Concur

coupa

ORACLE
Configure, Price,
and Quote Cloud

workday.

veeva

OpenAI

ATLASSIAN

Amplitude

Adobe

cvent

kapow
SOFTWARE

servicenow.

JIRA

stripe

mri
REAL ESTATE SOFTWARE

ActiveCampaign >

ORACLE
Eloqua Marketing
Cloud Service

SAP Ariba

CallidusCloud

DOMO

erwin
by Quest

klaviyo

Marketo™
An Adobe Company

Magento

SAP SuccessFactors

Adaptive
Insights

qb
intuit quickbooks.

elasticsearch

f

Dropbox

Microsoft
Dynamics 365

zuora

zendesk

SIEBEL

A

File Processor

NETSUITE

ORACLE
CRM ON DEMAND

xero

SurveyMonkey

INTUIT
mailchimp

Satmetrix

Power BI

X

SAP

MicroStrategy

P

SWIFT

NETEZZA

shopify

OPC UA

ODBC

PostgreSQL

FHIR

JDBC

ASCI2

MLLP

LDAP

APPDYNAMICS

AMQP

Apache Atlas

Qlik

kafka

talend

JMS

Yellowbrick

IERADATA

ORACLE
JD EDWARDS
ENTERPRISEONE

RabbitMQ

IBM

ORACLE
JD EDWARDS
WORLD

TIBCO

ODATA

ORACLE
E-BUSINESS SUITE

ORACLE
PEOPLESOFT

software

box

WebSphere
MQ

mixpanel

300+

Unique Data Sources

50k+

Metadata-Aware Connections

CLAIRE AI-powered Data Management

Simplifying how organisations manage and consume data



CLAIRE[®] Copilot

In-Context Intelligence to generate data management artefacts like data pipelines, data quality rules, and glossary terms. Also includes ML-powered automations like next transformation recommendations and entity matching.

CLAIRE[®] GPT

Talk to your data across a complex enterprise ecosystem. Discover datasets, Explore using natural language questions, establish trust with metadata, and generate data management artefacts.

CLAIRE[®] Agents

Automate end-to-end data management goals with intelligent autonomy. Harness advanced AI reasoning and planning models to deliver a new level of productivity, data accuracy and scale.

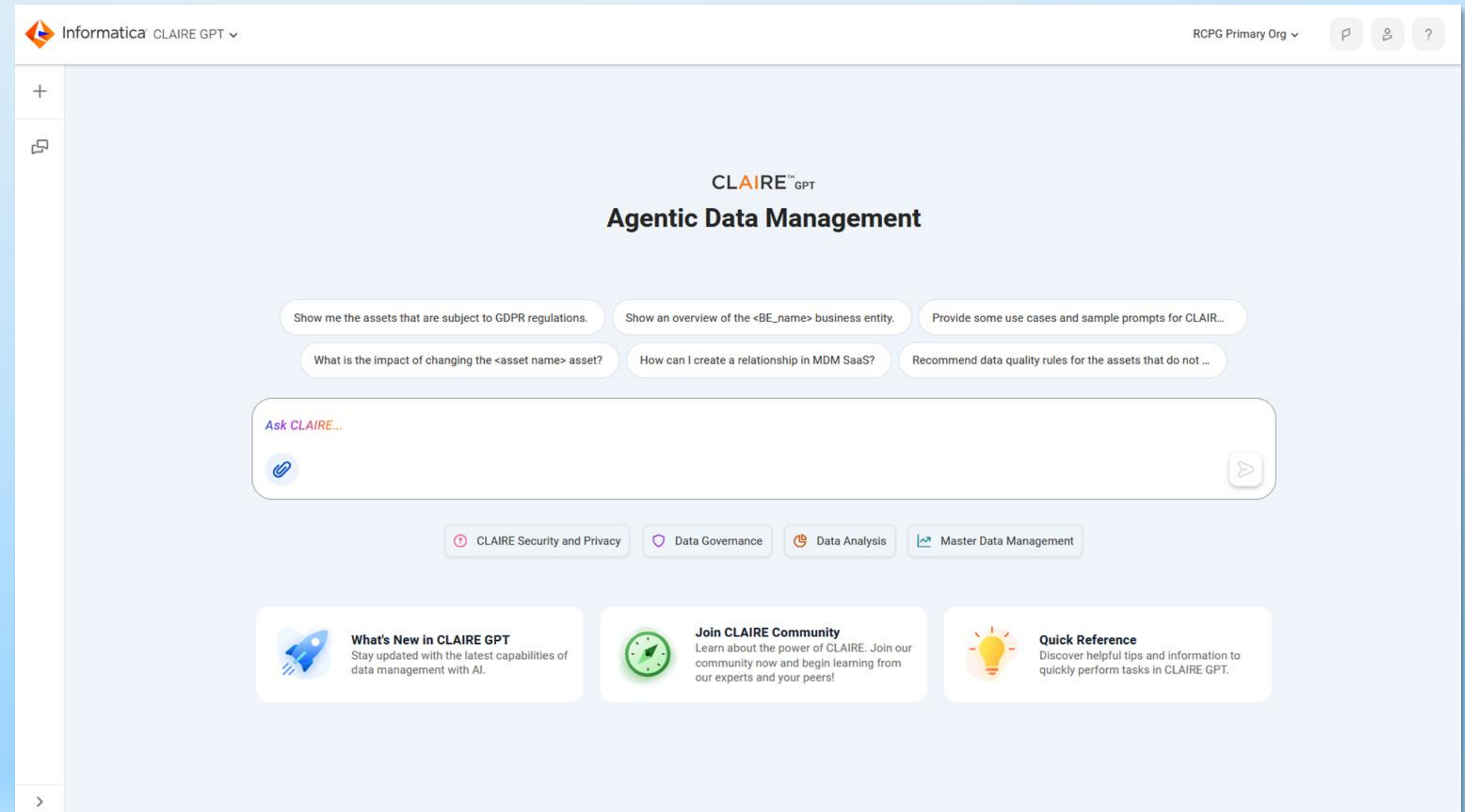


GenAI-Powered Data Management through Natural Language



GPT

Talk to your data across a complex enterprise ecosystem. Discover datasets, Explore using natural language questions, establish trust with metadata, and generate data management artefacts



CLAIRE[®]

AI Data Management Assistant



Copilot

In-Context Intelligence to generate data management artefacts like data pipelines, data quality rules, and glossary terms. Also includes ML powered automations like next transformation recommendations and entity matching

Informatica

Data Integration

Informatica CDP APJ

New...

Home

Explore

Bundles

My Jobs

Templates

My Import/Export...

Welcome Jonathon, what would you like to do today?

Do you use a cloud data warehouse as your primary destination?
If yes, we can set you up and optimize your processing so that it is faster and cheaper!

Yes, let's go

Remind me later

Ingest

Create an ingestion and replication task to replicate data from databases, files,...

Transform

Create a mapping in ETL or ELT mode to load and transform your data.

Orchestrate

Create a taskflow to define the control flow logic for your data integration and ingestio...

Recent assets

Status	Name	Location
Valid	dmt_Guidewire_to_Pos...	Customers'
Valid	mp_Guidewire_to_Post...	Customers'
Valid	tf_Guidewire_to_Postgr...	Customers'

View all

Recent jobs

Updated 2:40:16 PM PST

Status	Instance Name	Start Time
Success	dmt_Guidewire_to_Pos...	Feb 9, 2026, 10:40 PM
Success	dmt_Guidewire_to_Pos...	Feb 9, 2026, 10:37 PM
Stopped	mi_GitHub_to_Snowfla...	Feb 13, 2024, 5:42 PM

View all

Add another user

Go to Administrator to add a new user.

Let's get connected!

Add a connection to a source or destination.

Use our data transf...

Use a wizard to read and transform data from up to two sources (ETL).

Get insights!

Use Operational Insights to view the overall usage and health of your...

CLAIRE Copilot

Hi Jonathon, how can I help you with Data Integration today?
Here are some examples to help you get started.

Suggestions

How do I give feedback on my experience with ...

How do I summarize assets using CLAIRE Copi...

How do I create assets using CLAIRE Copilot?

Ask CLAIRE...

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

*CLAIRE Agents will start to be available for preview by the end of 2025

Data Quality Agent

Key Highlights

- Create, evaluate and operationalise DQ rules from business defined standard
- Deliver clean data for data consumption for analytical and operational use cases
- Proactive Data Quality for regulatory and analytical use cases

Benefits

- Increased productivity of data professionals with standardised rules creation and operationalisation
- Increased trust and satisfaction between data teams and business teams
- Better enforcement of standards for proactive data quality and compliance reporting

RT

We need to ensure the quality of medication administration data in our Epic EHR system. I'm uploading our hospital's medication administration policy document, which contains the data quality standards. I also have a spreadsheet, 'Medication_Formulary.xlsx', which lists all approved medications and their allowed routes of administration. We need to make sure that the data in our EHR conforms to these standards.

Medication_Formulary.xlsx

Medication_Administration_Policy.pdf

I understand that we need to ensure the quality of medication administration data. I have received the 'Medication Administration Policy.pdf' document and the 'Medication_Formulary.xlsx' spreadsheet.

Here's my initial interpretation of the key rule specifications:

- Rule Name: Patient ID Present; Rule Description: The Patient ID field must be present in the medication administration record.
- Rule Name: Medication Name Valid; Rule Description: The Medication Name must be a valid drug name from the hospital formulary.
- Rule Name: Dosage Valid; Rule Description: The Dosage field must be a valid dosage.
- Rule Name: Route of Administration Valid; Rule Description: The Route of Administration must be a valid route from the hospital formulary.
- Rule Name: Administration Time Valid; Rule Description: The Administration Time must be a valid time.
- Rule Name: Administering Nurse Valid; Rule Description: The Administering Nurse must be a valid nurse from the hospital staff.

Is my understanding of these rules correct?

Here's the plan:

1.

Check for existing data quality rules in the 'Medication Administration Quality.pdf' project within the 'Clinical Data' domain in the IDMC environment.

- Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque in justo erat. Ut rhoncus risus vitae ex tempus, at fringilla magna pellentesque.
- Nam felis ante, tincidunt vel efficitur in, facilisis vel leo.
- Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque in justo erat. Ut rhoncus risus vitae ex tempus, at fringilla magna pellentesque.
- Nam felis ante, tincidunt vel efficitur in, facilisis vel leo.

It will be done by the Data Quality Agent.

Medication_Administration_Policy.pdf

Medication_Formulary.xlsx

Data Quality Agent

2.

Generate New Rule Specifications (or mapplets, if necessary) for any rules not found in the existing rules.

- Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque in justo erat. Ut rhoncus risus vitae ex tempus, at fringilla magna pellentesque.
- Nam felis ante, tincidunt vel efficitur in, facilisis vel leo.

It will be done by the Data Quality Agent.

Data Quality Agent

Informatica CLAIRE GPT

General Hospital

Quality of Medication Administration Data

1920 x 72

Rule Name

Data Rule Details

Name * Rule Name

Description: Lorem ipsum dolor sit amet, consectetur adipiscing elit. In eu condimentum Mauris eget efficitur nulla, sit amet pellentesque lectus. Cras pretium con volutpat placerat. Ut sodales eros vitae orci consectetur, et blandit nulla m

Location: * Rule Creation Project VRule

Tags

Created on: 03/18//2025, 11:10 AM

Created by: CLAIRE Data Quality Agent

Last updated: 03/18//2025, 11:10 AM

Last updated by: 03/19//2025, 12:30 PM

Dimension: Validity

Rule Editor

```
1 IF (LENGTH(UPI)>20) THEN
2   Status="Invalid"
3 ELSEIF (REG_MATCH(UPI,"([A-Z0-9]{4})")=FALSE) THEN
4   Status = "Invalid"
5 ELSEIF ((REG_MATCH(System,"urnoid:[0-9]{1,3}-[0-9]{1,3}-[0-9]{1,3}") = FALSE) THEN
6   Status = "Invalid"
7 ELSEIF (LENGTH(Assigner)>0) THEN
8   Status = "Invalid"
9 ELSE
10  Status = "Valid"
11 ENDIF
12
```

Artefacts

Existing Rule (3)

Generated Rule (3)

Generated Rule Name 1

Generating and associating rules

Step 1: Checking for existing data quality rules in the 'Medication Administration Quality' project... Found 3 existing rules: 'Patient ID Present', 'Administration Time Valid', and 'Administering Nurse Valid'.

Patient ID Present

Quality of Medication Administration Data

Medication Administration Data Quality Dashboard

Data Quality Report

Generated on April 16, 2025 for Epic EHR - Medication Administration Records

Rules Generated: 3 Total Rules: 6

Summary of Generated Rules

Report 1

3 Generated Rules

3 Existing Rules

6 Total Rules

Types of Rules: 1

Completeness: 1

Validity: 2

Rules Generation Complete

Rules Association Complete

Dashboard Creation Complete

Operationalizing Rules (In Progress)

Distribution of Rules Across Data Quality Dimensions

Report 4

0 33%

1 50%

2 17%

Data Quality Dimension	Number of Rules	Percentage
Completeness	2	33%
Validity	3	50%
Relational Integrity	1	17%

Sample of Generated Data Quality Rules

Report 2

Rule Name	Rule Description	Rule Type	Status
Medication Name Valid	The Medication Name field must contain a valid drug name (either generic or brand name) from the hospital formulary.	Validity	New

*CLAIRE Agents will start to be available for preview by the end of 2025

Data Lineage Agent

Key Highlights

- **AI assisted** lineage - automatically parses files (scripts, spreadsheets,...) to extract data flow & transformation logic
- Handles **complex data transformation** (multi-scripting language)
- Automate catalogue source creation - seamlessly integrates with existing data catalogues to enrich metadata

Benefits

- Increase lineage source coverage - **Improved Troubleshooting** and Impact Analysis
- Allow documentation of complex lineage use cases - Support for Data Governance and Compliance
- Reduce manual effort and mistakes during custom lineage creation - **Enhanced Accuracy and Completeness** of Lineage

The screenshot displays the Informatica CLAIRE GPT interface. At the top, a user query is entered: "I want to verify the lineage for a dataset in the catalog. Please show me the lineage of the table Athlete_NOC". Below the query, a lineage diagram is shown, illustrating the flow from 'localSystems' to 'Data Lake' and then to 'OlympicGames_Analytics'. The diagram includes nodes for 'date', 'Athlete_NOC.csv', 'LTL_Athletes_NOC', and 'ATHLETES_NOC'.

Below the diagram, a chat interface shows the following conversation:

- GP:** This isn't the complete lineage. How can I find the rest of it?
- CLAIRE:** Please upload any ETL scripts used for these sources, and I can analyze them to find missing lineage.
- GP:** Here is my file
Athlete_NOC.py
- CLAIRE:** OK, I will analyze the ETL script and find lineage for the sources. Here is my plan for doing that.
Here's the plan:
 - Analyze script for entities and attributes**
I will analyze the file you provided **Athletes_NOC.py** and create a list of referenced entities and attributes.
 - Determine connection assignments (if needed)**
If I discover that some entities are reference assets in the catalog, I will determine the best connection assignments to use for merging with the complete asset definition.

Below the chat, a "Catalog Source Details" panel is displayed. It shows the following information:

- Name:** Custom Lineage 0425
- Description:** Custom Lineage for table Athletes_NOC
- Artifacts:** Entities and Attributes, Custom Lineage, Catalog Source Details
- Files Processed:** 1
- Lines of Code:** 28
- Entities Found:** 3
- Attributes Found:** 36
- Entity Links Generated:** 3
- Attribute Links Generated:** 77
- Artifacts Generated:** 2

On the right side, a "categorized_athletes.csv and Athlete_NOC.csv" panel shows the following findings:

- Found 18 lineage links between **noc_regions.csv** and **Athlete_NOC.csv**
- Found 27 lineage links between **bios.csv** and **categorized_athletes.csv**

Below this, a "Step 4. I will generate the catalog source details file." section shows the following statistics:

- Files processed: 1
- Lines of code processed: 28
- Entities found: 3
- Attributes found: 36
- Entity links generated: 3
- Attribute links generated: 77
- Artifacts: Lineage links, Catalog source details

At the bottom, a "Catalog Source Details File" section shows the file name "Custom lineage for table Athlete_Noc". A green checkmark indicates "All tasks have been completed successfully." Below this, there are buttons for "Download Artifacts" and "Publish to Catalog".

At the very bottom, a "Data Lineage Agent is done" message is displayed.

*CLAIRE Agents will start to be available for preview by the end of 2025

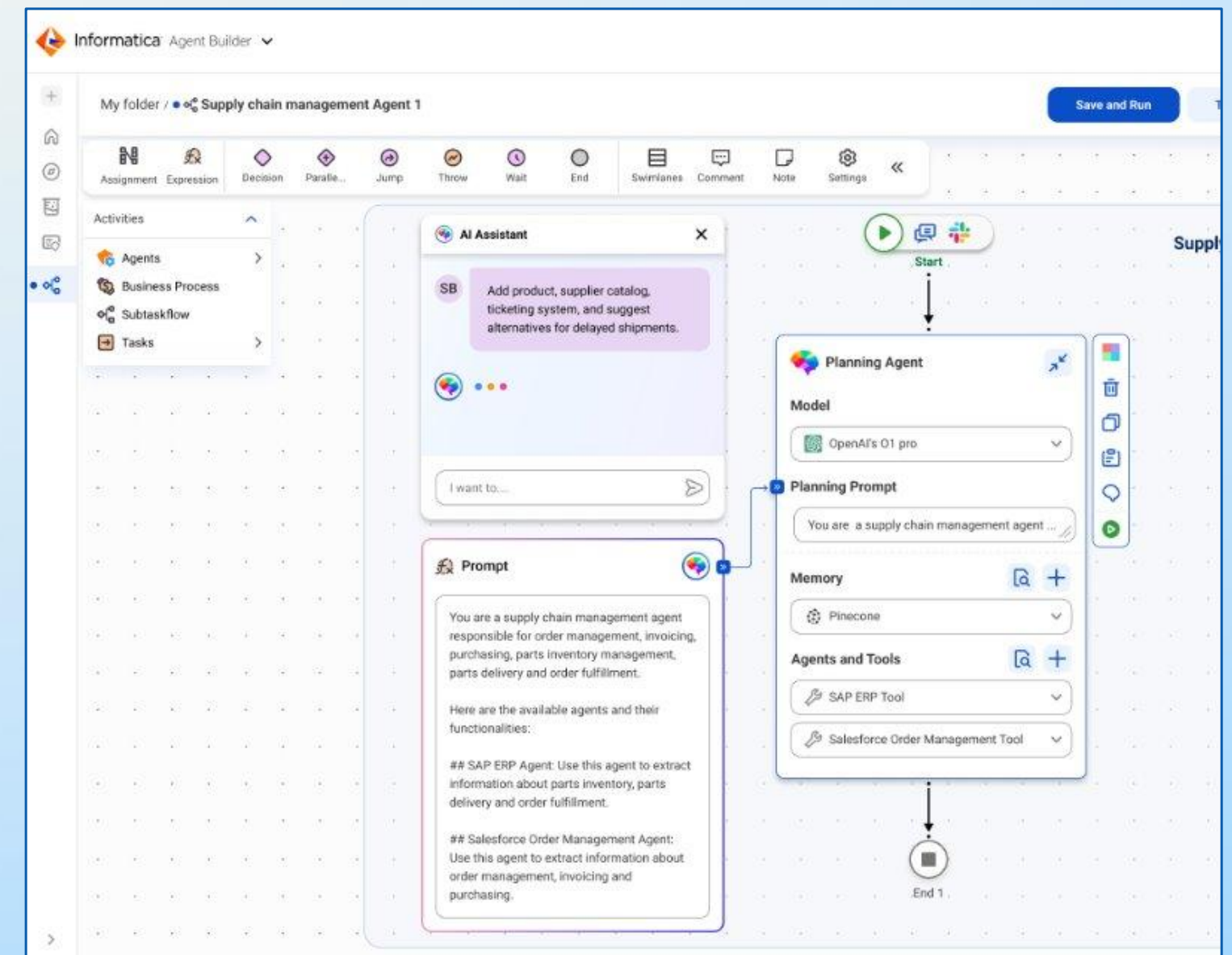
AI Agent Builder

Key Highlights

- Low-code/no-code "drag & drop" approach to build AI agents or a multi-agent framework
- Simplified AI-native user experience for building AI agents
- Built-in support for prompt engineering, testing, eval frameworks, and debugging
- Built-in support for creating an AI agent with an AI assistant

Benefits

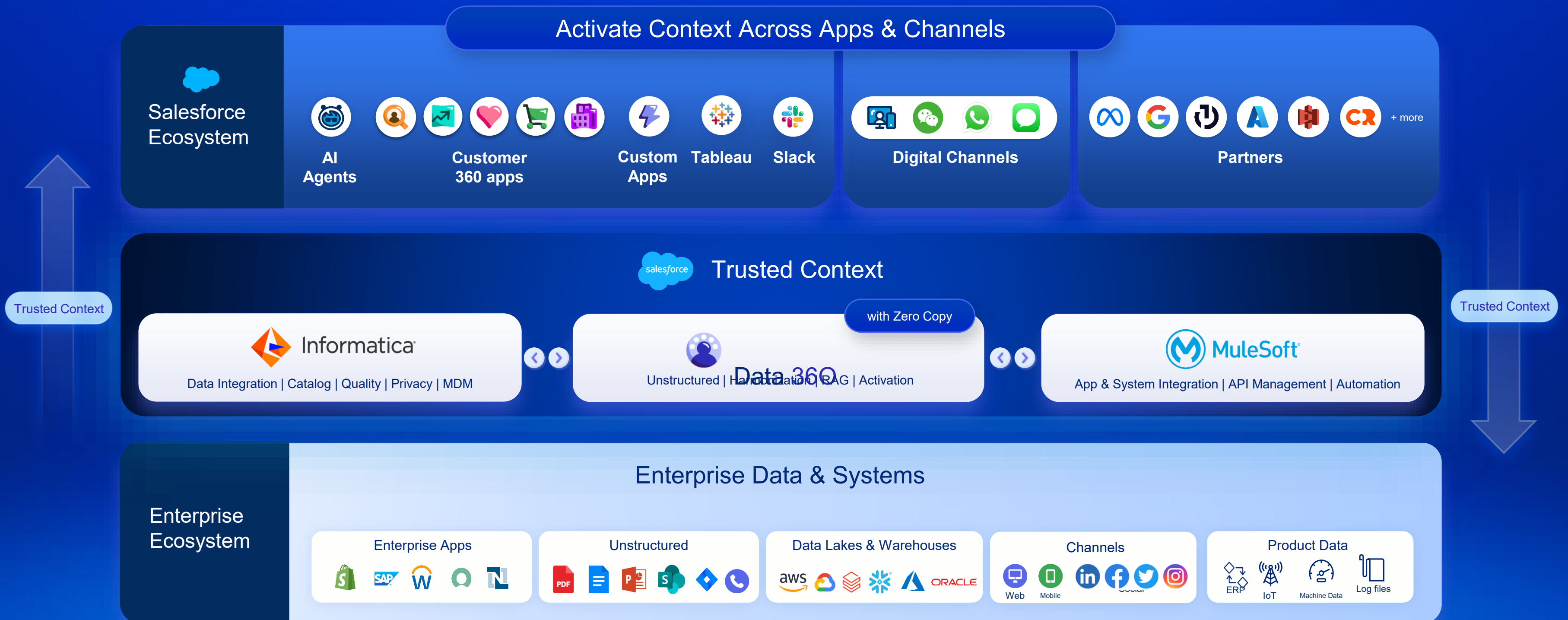
- No-code velocity with visual, drag-and-drop design lets business & IT teams create agents in minutes, not months
- Build accurate agents by reusing existing skills such as business processes, workflows, data integration pipelines, data quality rules to ground Agents with enterprise trusted data making agents deterministic and flexible
- Full SDLC support – Built in testing, evaluation, versioning, and deployment pipelines



COMING
SOON*

How it works - trusted context across Salesforce and the broader Enterprise

salesforce



We're Beyond the Hype with AI!

salesforce



**Continue the
Conversation**

@ the Informatica from
Salesforce Booth



**Data Agents Free
for Use**

Until end of July 2026



**Book a Demo
Today**

Speak to your Salesforce
or Informatica
representatives for more
info!

Thank you

