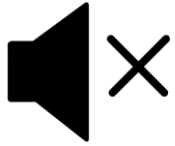Informatica®

23 April, 2024

Unlock the Power of Informatica Cloud: FAQs, Use-cases, and Best Practices

- Anita Ayyagari, Senior Solutions Architect, CSA

- Ananya Venkat, Solutions Architect, CSA

- Chinnari Damerla, Solutions Architect, CSA

- Rashmi P, Solutions Architect, CSA

- Akshada Sable, Solutions Architect, CSA

Where data & AI come to LIFE

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 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.
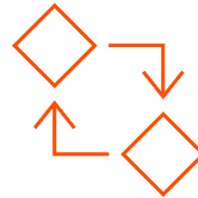
Feature Rich Success Portal
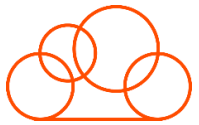
Bootstrap trial and POC Customers

Enriched Customer Onboarding experience

Product Learning Paths and Weekly Expert Sessions

Informatica Concierge

Tailored training and content recommendations

More Information

Success Portal
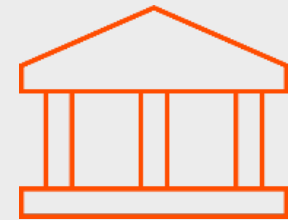
https://success.informatica.com

Communities & Support

https://network.informatica.com

Documentation

https://docs.informatica.com

University

https://www.informatica.com/in/services-and-training/informatica-university.html
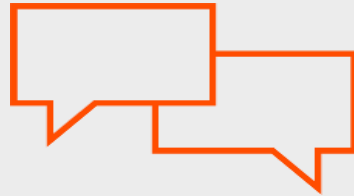
Safe Harbor

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.

Agenda

- Overview of IDMC services

- FAQ's and Best Practices on

    ✓ Secure Agent and Security

    ✓ Cloud Data Integration(CDI)

    ✓ Cloud Mass Ingestion(CMI)

    ✓ Cloud Data Governance and Catalog & Cloud Data Quality(CDGC and CDQ)

- Q&A

Informatica

Anita Ayyagari

PART 1 – FAQ- IDMC Services and Security

Where data & AI come to LIFE
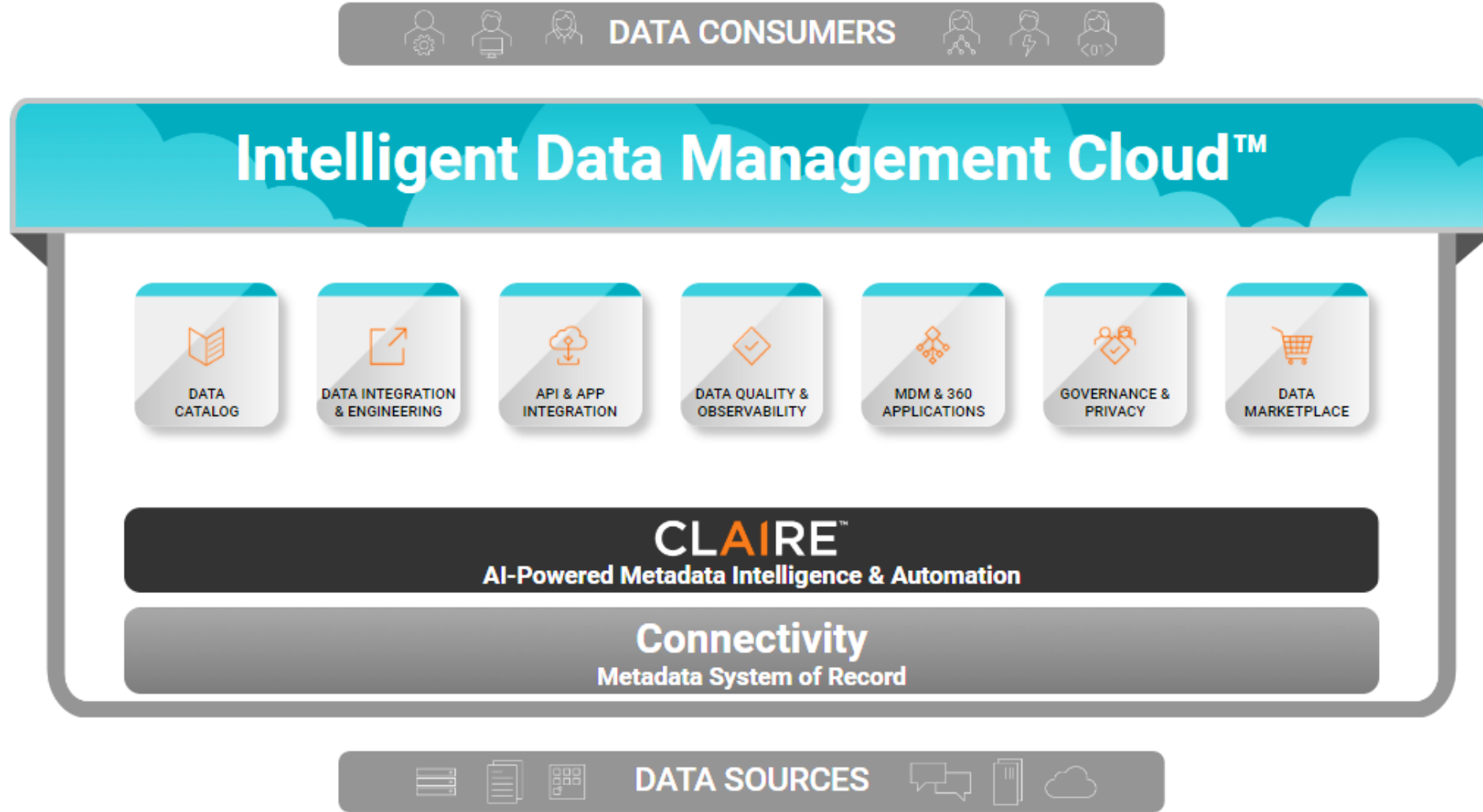
FAQ - IDMC Services

Question:

I believe Informatica was more of ETL tool does it even have other data management services ?

Overview of IDMC Services

IDMC Services

1. Data Integration

2. Cloud Application Integration

3. Mass Ingestion

4. Data quality and observability

5. Data Governance and Marketplace

6. Infacore

7. Modelserve

Informatica

FAQ- IDMC Services

That's nice to know of various services offered by Informatica

Question:

What is this CLAIRE Engine and how does it help improve our experience with the product?

Power of CLAIRE

Discover	......	Domain Identification
Model	......	Attribute & Hierarchy Recommendations
Acquire	......	Schema Mapping
Cleanse	......	Rule Recommendation
Merge	......	Identity Matching
Govern	......	Policy Classification
Search	......	Dataset Recommendation

FAQ- IDMC Services

Question:

What is secure agent and what does it do?

Informatica®

Informatica Cloud

Administration

Administrator

Application Integration Console

API Manager

API Portal

SAML Setup | User Access Control | Runtime Environments Config | Data Integration Connections & Schedules

Monitor

Monitor

Discovery IQ

Operational Insights

Running Jobs, Session Logs | Analytics, Diagnostics, Best Practices | Integrated operational monitoring / Predictive analytics

Design & Develop

Data Integration

Integration Hub

Application Integration

Tasks, Mappings, Task flows | Orchestration Process, Connections, Service Connector | Topics, Publications, Subscriptions

---

Cloud Hosted Process Engine

Metadata Repository Services
(Designs, Connections, Schedulers & Other Configurations)

Load Balancer

🔒 Secure Channel

Secure Agent Group

Secure Agent

Agent Manager

Agent Core

Tomcat

Channel Client, Metadata, Catalog | Data Integration Server

Process Server | CIH Processor

rDTM

Info on Agent

- Process Engine Logs
- Agent Core Logs
- Session Logs
- User Parameters
- Connections, Schedules

Secure Agent

Agent Manager

Agent Core

Tomcat

Channel Client, Metadata, Catalog | Data Integration Server

Process Server | CIH Processor

rDTM

Agent Manager	Java UI client used to manage (start, stop, proxy config) secure agent
Agent Core	Resilient ever-running single-threaded process checks for upgrades and manages Tomcat process
Channel Client	Provides secure, resilient communication to IICS channel servers
Catalog & rDTM	Catalog for Database DDL support and rDTM for Data Transformation
Metadata	Used to fetch metadata for files, databases, ODBC endpoints, IICS toolkit endpoints

Access Points & Connectors

File FTP SFTP

JMS

Special DB & App Connectors

JDBC/ODBC

HTTPS, HTTP REST/SOAP
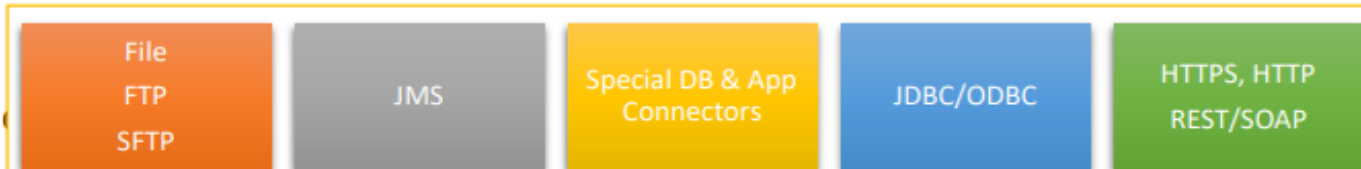
Informatica

FAQ- IDMC Services

That does explain about secure agent and security aspects to an extent but....

Question:

What are the different levels of security that are incorporated when data is processed by Informatica cloud?
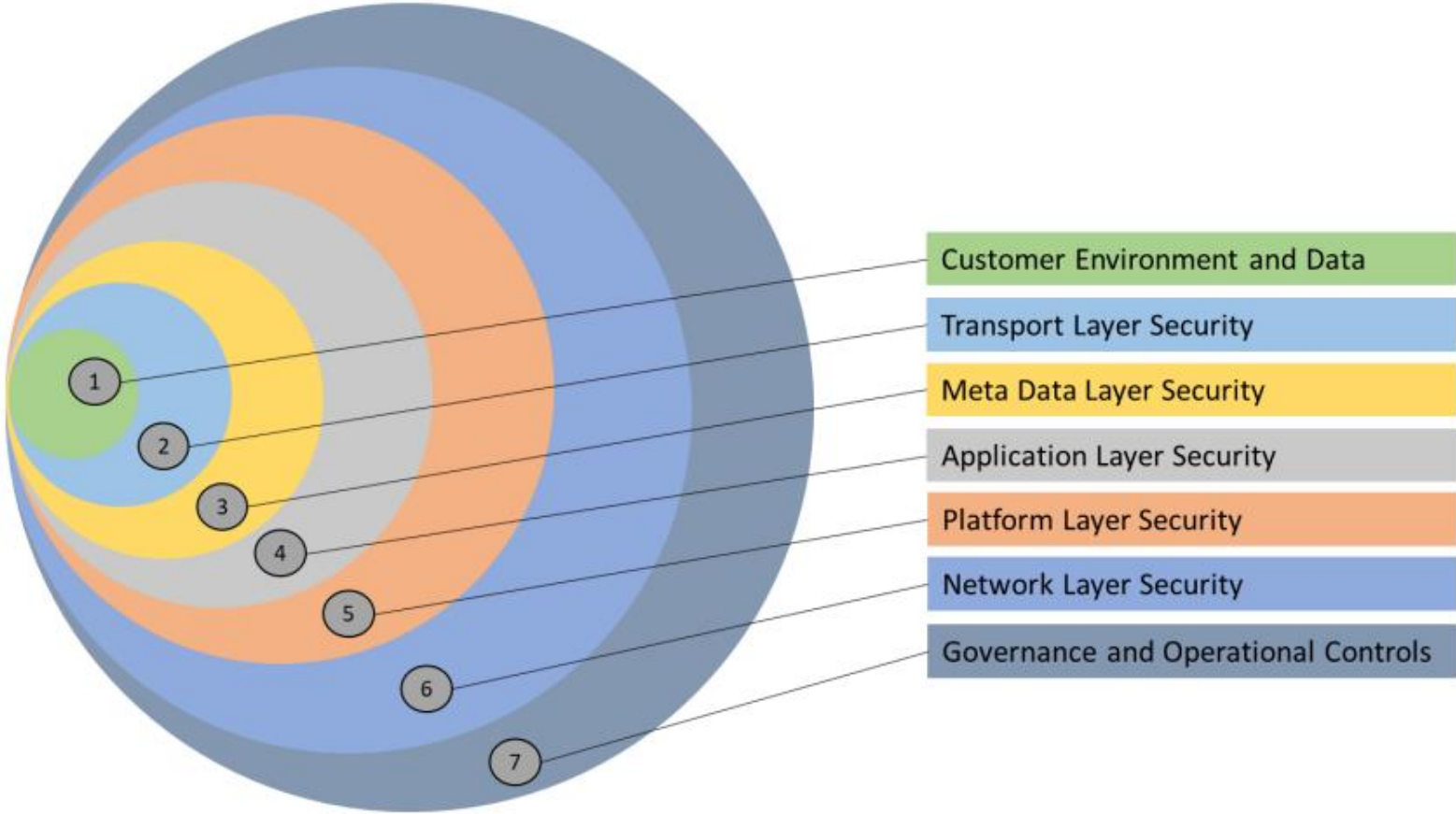
FAQ- IDMC Services

DATA IN CLOUD???

IS OUR DATA
SECURE???

Informatica

Levels of Security

1	Customer Environment and Data
2	Transport Layer Security
3	Meta Data Layer Security
4	Application Layer Security
5	Platform Layer Security
6	Network Layer Security
7	Governance and Operational Controls

Chinnari Damerla

PART 2 – FAQ- Cloud Data Integration (CDI)

Where data
& AI come to LIFE

FAQ- CDI

Question:

What is CDI ? When it is used?

FAQ- CDI
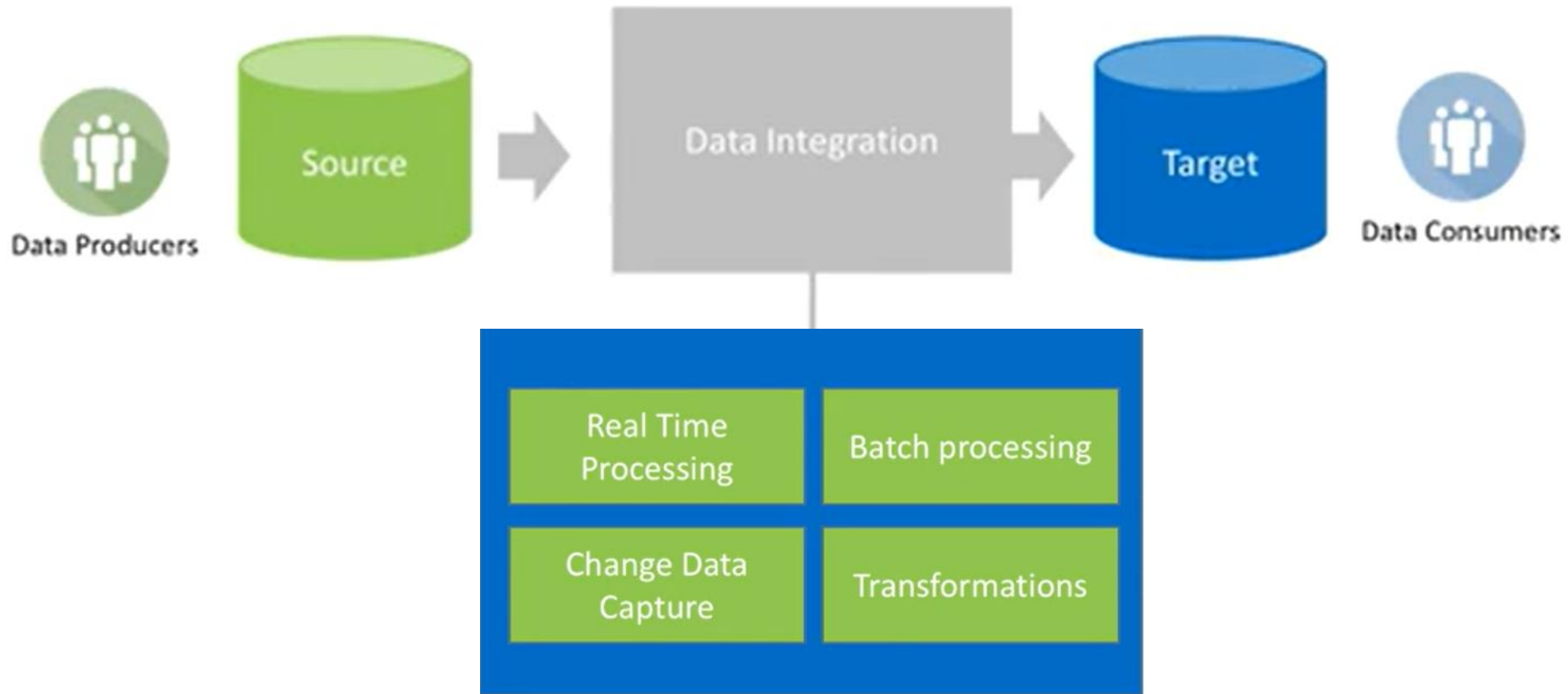
FAQ- CDI

Question:

What is DIS and what are different parameters for tuning the Data Integration Server?

FAQ- CDI

Data Integration Server is the Secure Agent service that runs data Integration jobs such as mappings, tasks and task flow instances

We can optimize performance of the data Integration Server by configuring some of its service properties. For example , we can change the network resiliency settings or the connection timeout period for the Secure Agent.

We can set the maxDTMProcesses custom property to improve performance

We can configure the INFA_MEMORY property and JVM options to achieve optimal performance and avoid Java heap and memory-related errors.

FAQ- CDI

maxDTMProcesses :

The recommended value for this property varies based on the connection types of the jobs that the agent runs:
•If the jobs use file-based or ODBC connections, the recommended value for the Secure Agent machine is 0.75 * # logical CPUs
•If the jobs use cloud data warehouse or cloud data lake connectors, the recommended value for the Secure Agent machine is
0.33 * # logical CPUs

Note :
This property should not be set higher than the number of CPUs on the Secure Agent machine because it will cause performance bottlenecks in task execution time.
When you set this property, make sure that you do not exceed the terms of your license agreement.

Custom Configuration Details

Service	Type	Sub-type	Name	Value	Sensitive
Data Integration Server	Tomcat		maxDTMProcesses	16	☐

Informatica

FAQ- CDI

JVM Options:

- Each mapping task on the Data Integration Server executes, runs as a unique process on the Secure Agent machine called the DTM process (pmdtm).

- The properties are

1. **-XX:ParallelGCThreads**

    - This property determines the number of garbage collection threads for the JVM. By default, it is computed as 3 + 5N/8 where N is the number of CPUs on the agent machine. The recommended value for this property is 2.

    - Tuning this property helps reduce the overall CPU utilization of the Secure Agent machine under concurrent loads by reducing unwanted context switching of the JVM threads.

2. **-XX:CICompilerCount**

    - This property determines the number of compiler threads for the JVM for byte code optimization. The default value is computed based on the number of CPUs of the agent machine. The recommended value for this property is 2.

    - Tuning this property helps reduce the overall CPU utilization of the Secure Agent machine under concurrent loads by reducing unwanted context switching of the JVM threads.

FAQ- CDI

3. **-Xmx**

- This property determines the heap size of the JVM for the DTM. The default value is 64MB.

- For mappings that process large data volumes with multiple partitions, the default value can cause mapping failure because of insufficient Java Heap Space. The recommended value for this property is 2024MB

System Configuration Details   Reset All

Service: Data Integration Server

Type: DTM

Type	Name	Value	Sensitive	
DTM	JVMClassPath	'pmserversdk.jar'	☐	✏ 00
DTM	JVMOption1	-XX:ParallelGCThreads=2	☐	✏ 00
DTM	JVMOption2	-XX:CICompilerCount=2	☐	✏ 00
DTM	JVMOption3	-Xmx2024m	☐	✏ 00
DTM	JVMOption4		☐	✏ 00

FAQ- CDI

INFA_MEMORY

- Configured to achieve optimal performance and avoid Java heap and memory-related errors

▼ **System Configuration Details**

Service:	Data Integration Server ⌄

Type:	Tomcat JRE ⌄

Type	Name	Value	Sensitiv
Tomcat JRE	INFA_SSL		☐
Tomcat JRE	INFA_MEMORY	'-Xms32m -Xmx512m'	☐
Tomcat JRE	JRE_OPTS	'-Xrs'	☐
Tomcat JRE	JAVA_LIBS		☐

FAQ- CDI

Question:

What is DTM and DTM process? Why it is important?

FAQ- CDI

What is DTM?

The component of the Data Integration Service that extracts, transforms, and loads data to complete a data transformation job. The DTM also runs profiles and data previews.

What is DTM process ?

The Data Integration Service starts the DTM process. The DTM is the process associated with the session of a task.

The DTM process performs the following tasks:

- Retrieves and validates session information from the repository.

- Performs pushdown optimization when the session is configured for push down optimization.

- Adds partitions to the session when the session is configured for dynamic partitioning

FAQ- CDI

- Forms partition groups when the session is configured to run on a grid.

- Expands the service process variables, session parameters, and mapping variables and parameters.

- Creates the session log.

- Validates source and target code pages.

- Verifies connection object permissions.

- Runs pre-session shell commands, stored procedures, and SQL.

- Sends a request to start worker DTM processes on other nodes when the session is configured to run on a grid.

- Creates and runs mapping, reader, writer, and transformation threads to extract, transform, and load data.

- Runs post-session stored procedures, SQL, and shell commands.

- Sends post-session email.

Question:

What are the DTM Performance Properties?

FAQ- CDI

DTM properties in mapping tasks are:

- **DTM Buffer Size**: The DTM buffer size specifies the amount of buffer memory used when the DTM processes a session.

- **Default Buffer Block Size:** The buffer block size specifies the amount of buffer memory used to move a block of data from the source to the target.

- These two parameters of the task can be changed only in the Mapping Task of a Mapping. By default, it is set to Auto in all the type of task. To change the same,

① **General**　　② **Runtime Options**

Advanced Session Properties

Session Property Name	Session Property Value
DTM Buffer Size	1GB
Default buffer block size	1MB

FAQ- CDI

The DTM buffer size and Block Size are calculated as follows:

- **DTM Buffer Block Size** = 2*X, where X is the row size in KB (It can be set approximately based on the data which you have in the source). Or Can count the number of fields in the sources and targets and multiply it with their precision * 2.

    For example: 5 fields in source with 255 precision and 5 fields in target with 255 precision, then 255*10*2. 2 is for multibyte characters as in agent by default, DataMovementMode is UNICODE

- **DTM Buffer Size** = (DTM Buffer Block Size) * 10* N, where N is the number of partitions Or 10 times the block size.

Note:

- When a task contains large amounts of character data, increase the DTM buffer size to 24 MB

- When a task contains 'n' partitions, increase the DTM buffer size to at least 'n' times the value for the task with one partition

- When a source contains a large binary object with a precision larger than the allocated DTM buffer size, increase the DTM buffer size so that the task does not fail.

FAQ- CDI

Question:

What are the possible reasons for high CPU usage and how to mitigate?

FAQ- CDI

When there is a high CPU usage, alerts are sent from the secure agent server.

What needs to be checked:

Task manager should be verified for 'pmdtm.exe' and 'zulu platform' which are using too much of CPU. On top of that, 'pmdtm.exe' appears to be running in multiple instances, each of them consuming CPU.

pmdtm.exe - For every task triggered, a pmdtm process is spawned on the secure agent server to carry over and perform the task requested. Ram used by this process is handled by the data processed by that task. CPU usage is determined by the intensity of the task, like number of transformations, string operations, conversion etc which requires processing power.

Zulu platform - Zulu platform is the JDK provider which is used by IICS. The zulu instances in the process list are there to help the secure agent and the services (Data Integration Server, Process Server etc) keep on running

If the **maxDTMProcesses** number is higher and there are many CPU intensive jobs that are running at the same time, then many pmdtm.exe processes will run for each of those tasks, causing a high CPU spike.

FAQ- CDI

How to Mitigate:

- Verify the jobs which are running at the same time and judge if those are heavy intensive jobs.

- Verify the CPU usage of those jobs and multiply the average value with maxDTMProcesses value. In this manner, a precedent can be set and according to that, CPU can be increased in secure agent server or maxDTMProcesses shall be reduced.

- Also, if any old process kept on running, causing heavy CPU usage, the secure agent should be stopped, and all the java processes should be manually killed before firing up secure agent again.

Question:

How to identify the low performing components in a mapping?

FAQ- CDI

When the execution of the task is completed, in the session log of a completed process, you can find a section called "LOAD SUMMARY".

In that you will have the subsection "***** RUN INFO FOR TGT LOAD ORDER GROUP [1], CONCURRENT SET [1] *****" under which you can see three threads. Each of these threads represents the Reader, Transformation and Writer respectively. These sections also contain information about three factors:

"Total Run Time" = time this thread was loaded

"Total Idle Time" = time this thread was loaded but waiting for other components

"Busy Percentage" = percentage of time this thread was actually working

From this information, you can decide if the bottleneck is either the source (Reader), the transformations or the target (Writer). Judging on the Busy Percentage you can also estimate if the process is waiting on an external endpoint (low Busy Percentage).

FAQ- CDI

Question:

When should we change the Java Heap Size for CDI Jobs?

FAQ- CDI

While executing the task, if we see any of the below errors

java.lang.OutOfMemoryError

java.lang.OutOfMemoryError:Java heap space

FATAL ERROR : Failed to allocate memory. Out of virtual memory.

This means, there is insufficient memory for the Java Runtime Environment to continue.\n# Native memory allocation(malloc) failed to allocate

How to calculate the the JVM heap size

Formula:

-Xmx<size>m -> size=SYSTEM_RAM/maxDTMProcesses

Ananya Venkat

PART 3 – FAQ- Cloud Mass Ingestion

Where data
& AI come to LIFE

FAQ- Cloud Mass Ingestion

Question:

What does CMI do and Why CMI ?

Informatica

Data Ingestion From Various Sources Is Critical For Cloud Modernization

Files

Mainframe
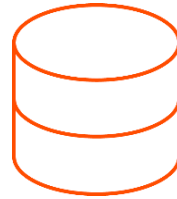
Database
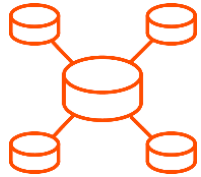
On-prem Data Warehouse

Machine Data

IoT Data

Streaming

Documents
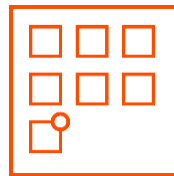
Apps

Data Ingestion

Cloud Data Lake Ingestion

Data Warehouse Modernization

Database Migration/ Synchronization

Real-Time Analytics

FAQ- CMI

Question:

What are the type of solutions provided by CMI?

Informatica

Cloud Mass Ingestion - Use case Patterns

Cloud Data Lake Ingestion

- Mass ingestion of files into cloud and on-premises data lakes

- Streaming and IoT data ingestion into a data lake

- Mass ingestion of on-premises apps, database content into a cloud or on-premises data lake

DWH Modernization Database Migration /DB Synchronization

- Mass ingestion of on-premises database, data warehouse, apps and mainframe content into a cloud data warehouse (Snowflake, Synapse)

- Synchronize ingested data with Change Data Capture (CDC) and applying

Kafka for Real-Time Analytics

- Log files and clickstream ingestion

- CDC ingestion

- IoT data ingestion

- App change data ingestion

FAQ- CMI

Question:

We keep hearing the term "schema drift" while we talk about CMI? What is it ?

Automatically addresses schema drift in source
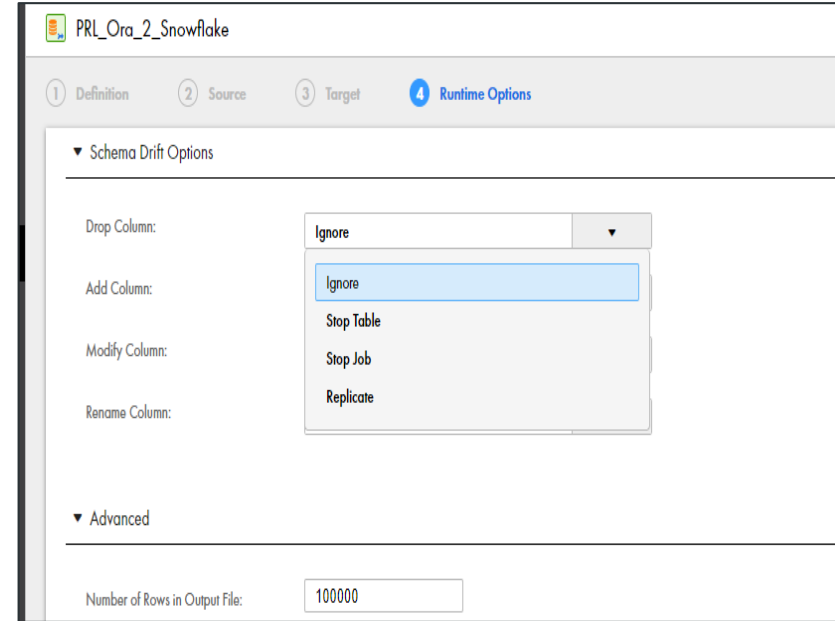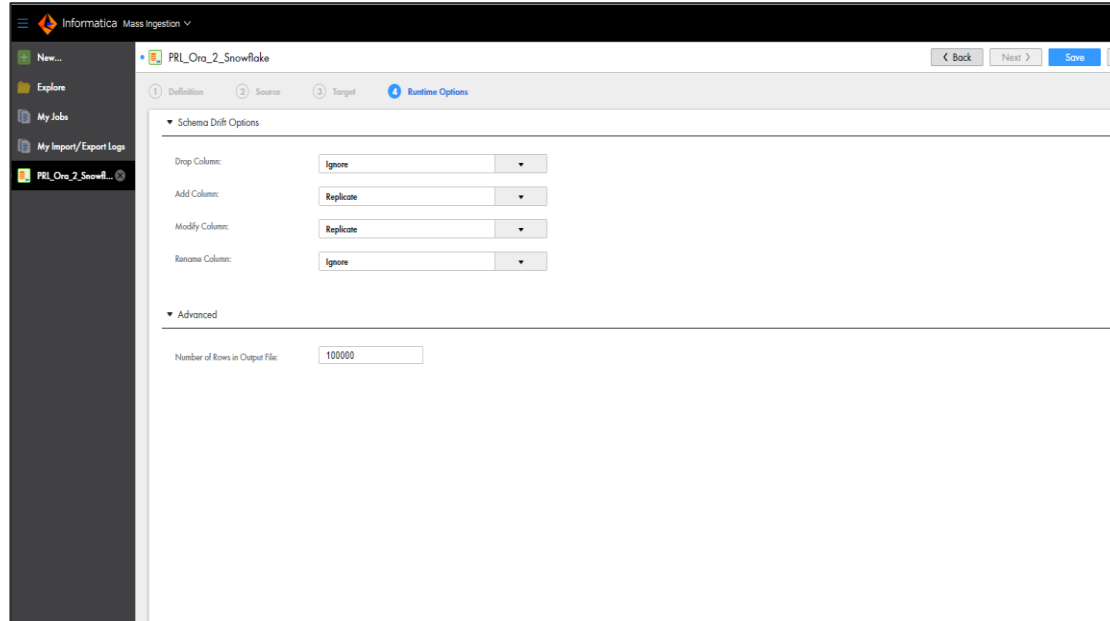
Recognizes Changes on the source schema
and Automatically Processes & Applies the Changes

FAQ- CMI

Question:

I do have a scenario wherein I need to perform Advanced Analytics on Application Data , can I explore CMI?

Informatica

Mass Ingestion Applications - Common Use Cases

- ✓ Analytics offload from SaaS application & DW consolidation
- ✓ Advanced Analytics on Application Data
  - ✓ Initial & change data ingestion, Apply changes, Schema drift, near real time ingestion

- ✓ Real time process automation & action operationalization (for App modernization)
  - ✓ CDC data ingestion onto Messaging Systems
  - ✓ Integration with CAI and CIH for downstream consumption

Informatica Intelligent Cloud Services

Cloud Targets

Data Warehouses
Data Lakes

Data offload and Advanced Analytics

Real time Process Automation and Operationalization

On-premise Apps

Secure Agent

Analytics

SaaS Apps

Messaging Storage
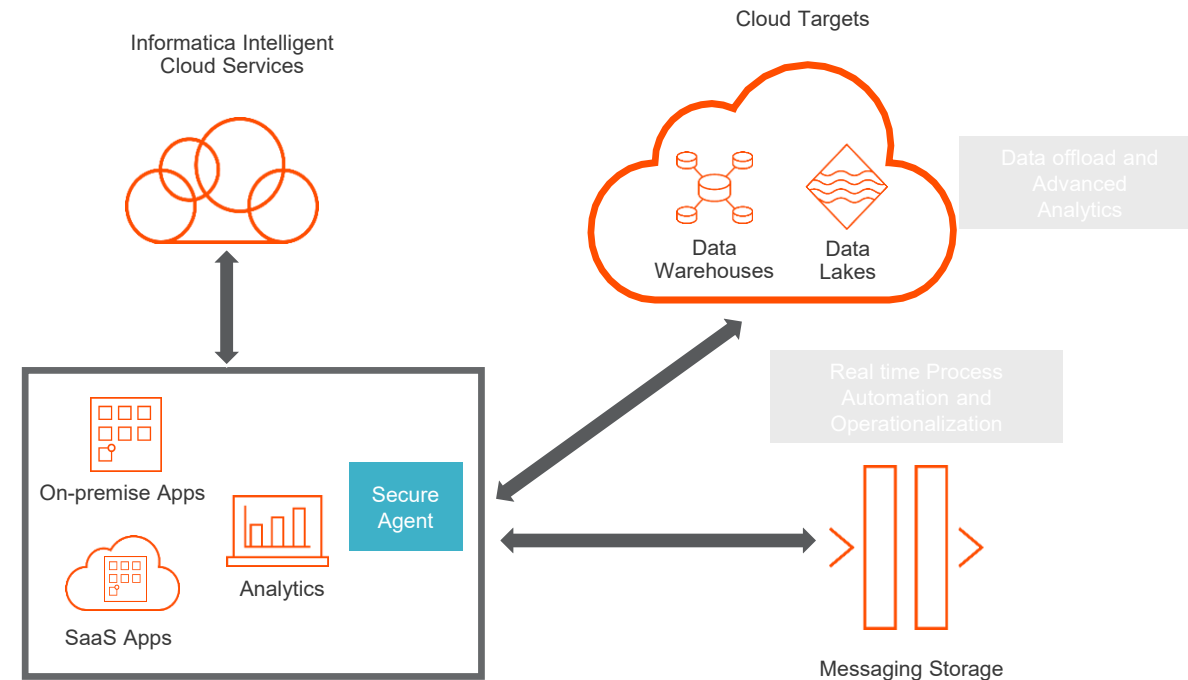
FAQ- CMI

Question:

We are a large bank, and we want to alert customers if there is any fraudulent transaction in the system, as soon as possible.

Informatica

Cloud Mass Ingestion
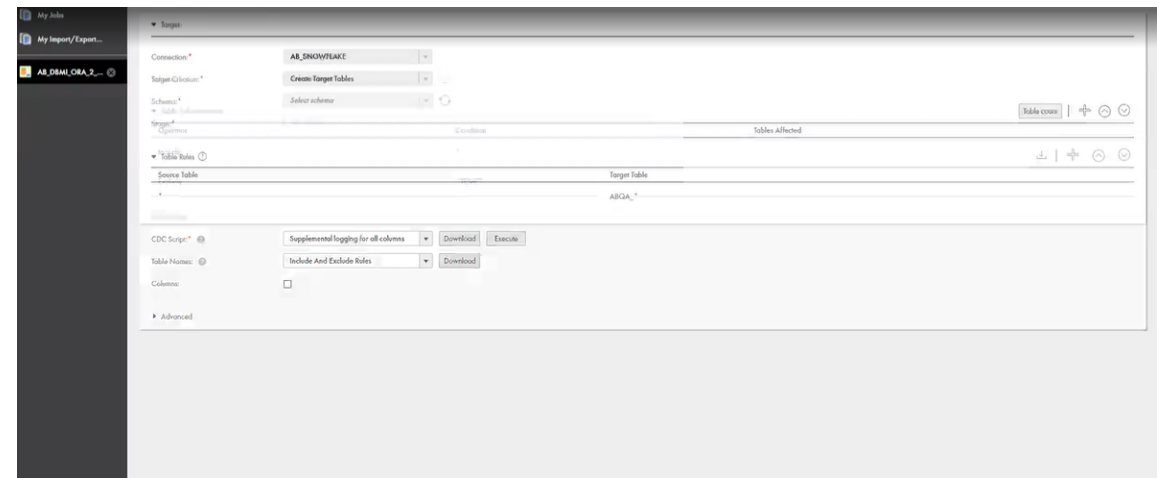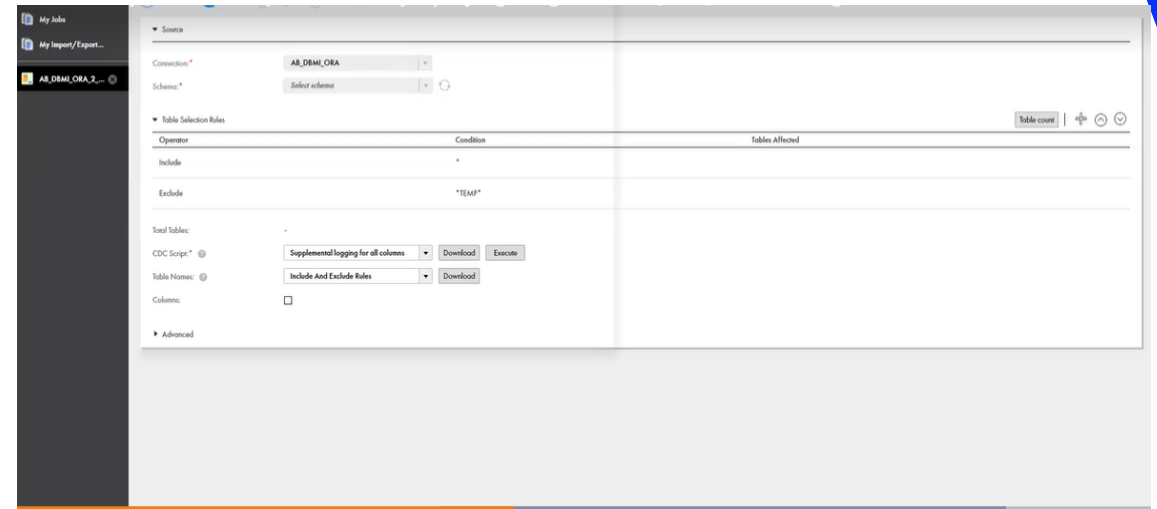
Real time monitoring & lifecycle management

Informatica

FAQ- CMI

Question:

Does Mass Ingestion Streaming support a Secure Agent group with more than one Secure Agent?

Akshada and Rashmi

Cloud Data Governance and Catalog & Cloud Data Quality

Where data & AI come to LIFE

How to effectively break down silos within our organization and promote cross-functional collaboration?

Implement Data Governance

The successful data-driven enterprise requires effective data governance. As data and its value scales and grows with the business, it's important that data be secure, and be aligned with business policies and priorities.

Ability to Implement, Automate, Enforce

➡ Link People, Policies & Processes and Data together

➡ Maximizing the Use of Data to Make Decisions

➡ The Imperative for all Data Projects

➡ The Value of Data Governance can be seen in Data Projects, where Data Governance is Absent

Where data & AI come to LIFE

CDGC

Centralized Platform

Business Assets

191 Business Term	**88** Policy
61 Process	**21** Data Set
12 Metric	**12** System
5 Domain	**4** AI Model

Technical Assets

3.4K Column	**817** View Column
393 Calculation	**249** Table
68 Catalog Source	**59** View
33 Procedure Instance	**17** Schema
12 Tableau Worksheet	
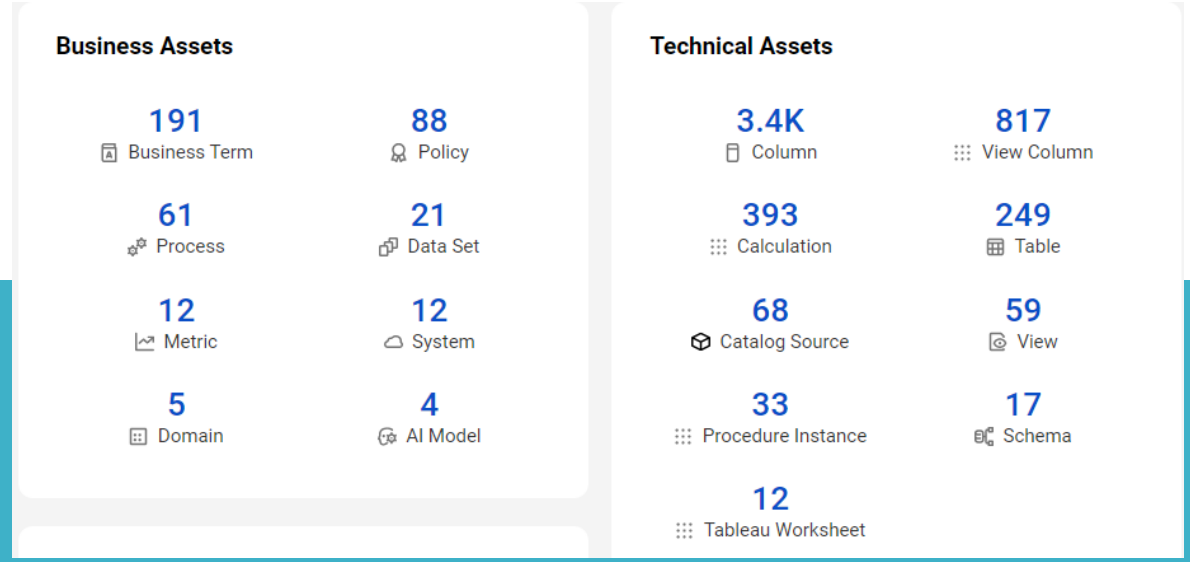
Glossary

Items (4)

Name	Type	Description
Party Data	Domain	Data related to customers of the organisation - both personal and corporate.
Corporate Data	Sub Domain	Corporate Data covers the data elements of corporate entities
Customer	Business Term	Customers are corporate entities or individuals that have an economic relations...
Party Data - Classifications	Business Term	Classifications related to a corporate entity or individual that has an economic re...
KYC/KYB Status	Business Term	Status of party as per KYC/KYB regulation.
Party Data - Core Entities	Sub Domain	This category defines core elements which can be related to a party that is eithe...
	Business Term	The party ID is the numeric representation used to identify the party in question.
	Sub Domain	Personal Data covers the data elements of persons / individuals
	Domain	Data related to products sold by the organisation.
	Domain	Retail Measures and Metrics
	Domain	Terms related to retail sales

➡ CDGC, a service of the Informatica Intelligent Data Management Cloud™ (IDMC), combines data governance, data catalog and data quality capabilities into a singular tool for automating data intelligence insights.

➡ CDGC offers broad and deep metadata connectivity that spans multi-cloud and on-premises environments

Data Governance

Informatica™

Collaboration

Common Data Standards and Taxonomies
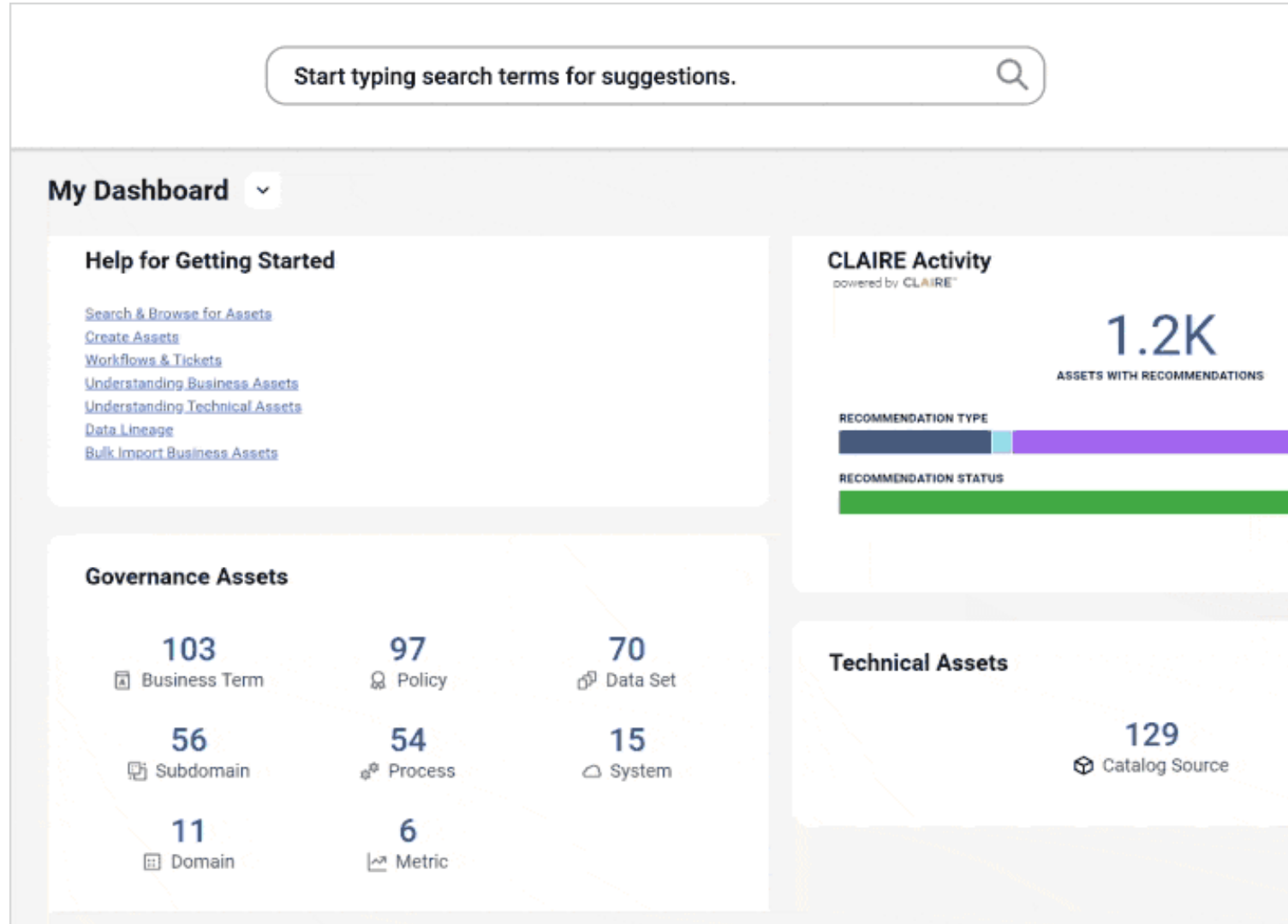
Data Quality Monitoring and Improvement

Chief Data Officer
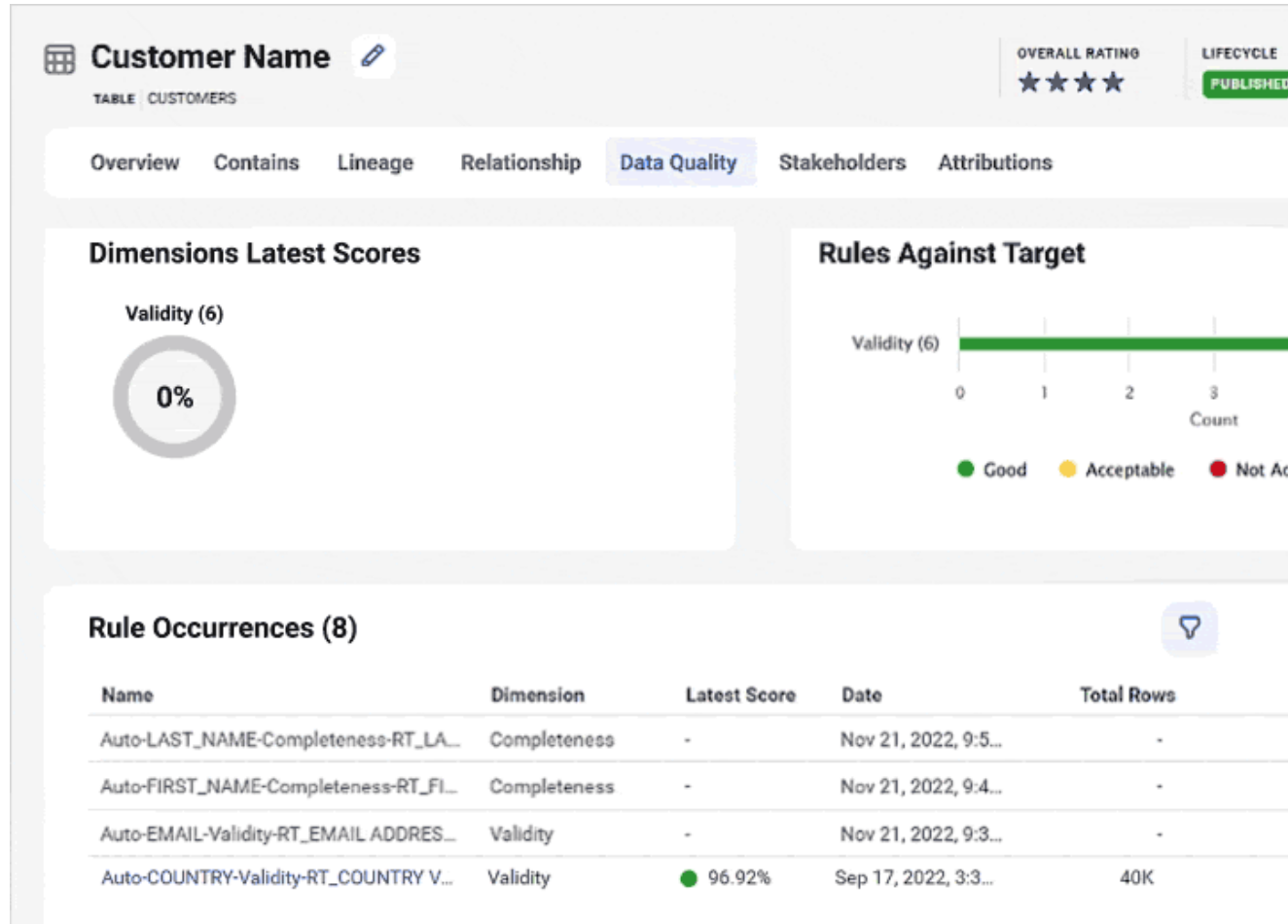
What process should be followed to compile a comprehensive report detailing all Critical Data Elements (CDEs) across our organization, ensuring accuracy, relevance, and alignment with strategic goals?"

Sensitive Data & Assessment of Privacy Risks

Data Privacy
Widgets to get insight on sensitive data assets

Data Privacy Quick Links

- 360 View Of ALL PII Data
- Snowflake PII Data - mask sensitive data in Snowflake when not authorized
- Email 360 view
- PII Policy Definition & usage

PII Business Terms

Name	Description
CSC / CVV	Card Security Code (CSC)/...
Trade Union Membership	Any membership with a tra...
Next of Kin / Emergency C...	The next of kin to the indivi...
First Name	First name of a person.
Country of Birth	The country that an individ...
Full Name	An individual's full name.
Credit/Payment Valid to date	Credit/Payment Valid to date
Postal Code	The postal code used to m...
Political Affiliations and Op...	The policitcal views held by...
Gender	One's identity as female or ...
Account Number	The number used to identi...

1 - 25 of 34 Items   View All

Data Classifications (338)

Name ↑	Type	Description	Sensitivity	Updated On
Albania National ID	Data Element	This matches with ...	HIGH	Apr 11, 2023, 7:22 ...
Amex	Data Element	This identifies Am...	HIGH	Apr 11, 2023, 7:20 ...
Argentina National ID (DNI)	Data Element	This matches with ...	HIGH	Apr 11, 2023, 7:22 ...
Australia Bank Account Number	Data Element	Classifies metadat...	HIGH	Apr 11, 2023, 10:0...
Australia Passport	Data Element	Classifies metadat...	HIGH	Apr 11, 2023, 7:20 ...

➡ Enable stewards to quickly identify datasets and data sharing lineage that may indicate potential privacy risks in order to gain more insight into policy compliance for sensitive information.

➡ With increased transparency, your data protection planning can help limit customer information exposure and avert risks from abuses and data loss

➡ Automatically tag and classify sensitive PII data and monitor privacy risks in customizable dashboards. Users can set a sensitivity level for pre-defined OOTB classifications.

Informatica

Business Analyst

As a Business analyst ,What steps should I take to ensure that the data I have retrieved is compliant with business requirements and suitable for analysis

Cloud Data Governance and Catalog:
Reduced expected exposure from compliance

Data Compliance: *The process of following various regulations and standards to maintain the integrity and availability of the sensitive/regulated data.*

© Informatica. Proprietary and Confidential.

Data Analyst

How do I leverage discovery capabilities to identify and shortlist data assets with high-quality and relevant data?

Data Governance Quality

Thank You

Where data & AI come to LIFE

Informatica