Informatica
from Salesforce

03 Mar, 2026

Maximize Data Velocity: Data Replication Evolved for the AI Era

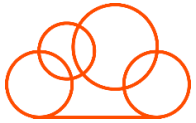
- Deepika R, Senior Technical Support Engineer, GCS
- Faisal Ishaq, Product Manager Senior Specialist, R&D

Where data & AI come to LIFE

Housekeeping Tips

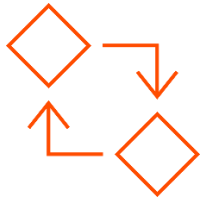
- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via 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

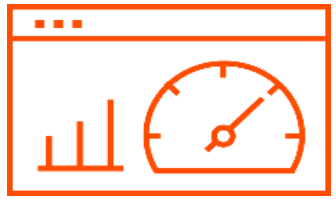


Product Learning Paths and Weekly Expert Sessions

Informatica Concierge

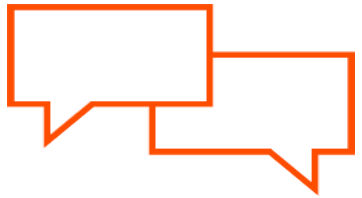
Tailored training and content recommendations

More Information



Success Portal

https://success.informatica.com

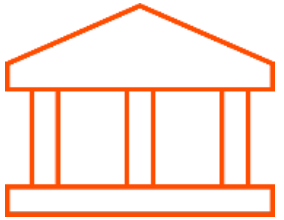


Communities & Support

https://network.informatica.com

Documentation

https://docs.informatica.com



University

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

Safe Harbor

Disclaimer: The information being provided herein is for informational purposes only. The development, release and timing of any Informatica product, service or functionality described herein remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made herein are based on information currently available, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products, services or functionality in the future.

Agenda

1. CDIR Overview & Trends
2. CDC Data Staging
3. Row Filters
4. Serverless Runtime
5. Orchestration-Data to Insights
6. Open Table
7. Q&A

Informatica Cloud Data Ingestion and Replication

Features, Capabilities, and Value Proposition

1

Simple: Easy to Use

- **Easy** 4-step wizard-based approach for building ingestion and replication tasks
- **Automatically** create data ingestion jobs using natural language prompts and **CLAIRE-assisted** guided experience
- **Monitor** ingestion jobs in real-time with lifecycle management and alerting in case of issues

2

Seamless: Unified and Comprehensive

- **Unified solution** for all ingestion and replication sources – databases, Files, streaming, mainframe, applications
- **Broad connectivity** with pre-built connectors for on-prem and cloud sources and targets
- **Post-replication transformation** to support end-to-end EL&T

3

Support: All Ingestion Patterns

- **Ingest and replicate** any data patterns like - batch, streaming, CDC into cloud data warehouses, lakes, message hubs
- **Multiple CDC methods** like – timestamp, trigger and log-based and **replication patterns**- Soft delete, Audit
- **Automatic schema drift** to ensure uninterrupted data replication

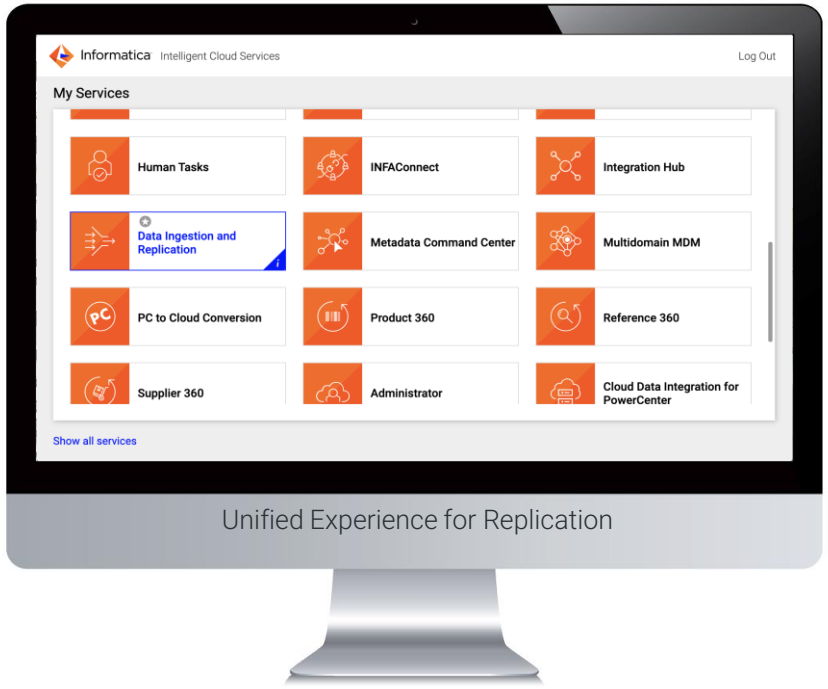
4

Scalable: Enterprise Grade Scalability

- **Ingest billions of rows** and millions of files in hours
- **Cost and performance optimized** for Snowflake target – Stream data 3.5x faster and 40% cheaper
- **CDC Data Staging** – efficient 1-to-many
- **Elastic performance** – Automatically scales up or down compute to handle data volume spikes

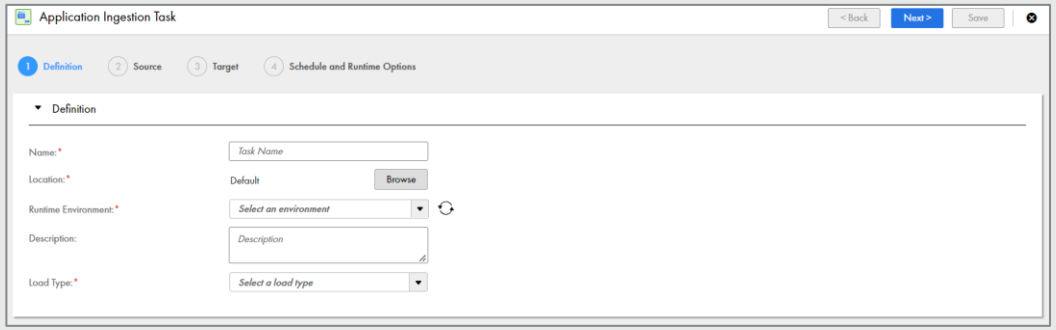
Cloud Data Ingestion and Replication

All ingestion patterns for files, databases, applications and streaming sources



Unified Experience for Replication

✓ Step-by-step wizard for designing and creating an ingestion and replication task



✓ Deployment, scheduling, real-time monitoring and lifecycle management

- Ingest in Real-Time
- Real-Time Monitoring
- Automated Schema Drift Handling

✓ Versatile out-of-the-box connectivity to sources and targets

- Databases & CDC
- Streaming Sources
- Files
- Applications

CDC Staging

What is CDC Staging ?

-Read once Write Multiple times

•It stages CDC data into a centralized cloud storage layer (using Apache Iceberg tables on S3, ADLS, or GCS).
•Decouples CDC extraction from application, improving modularity and scalability.

Challenge

- Multiple CDC tasks read from the same source, causing repeated log scans and high load.
- Each task uses a separate connection, increasing overhead and reducing performance.

Ask

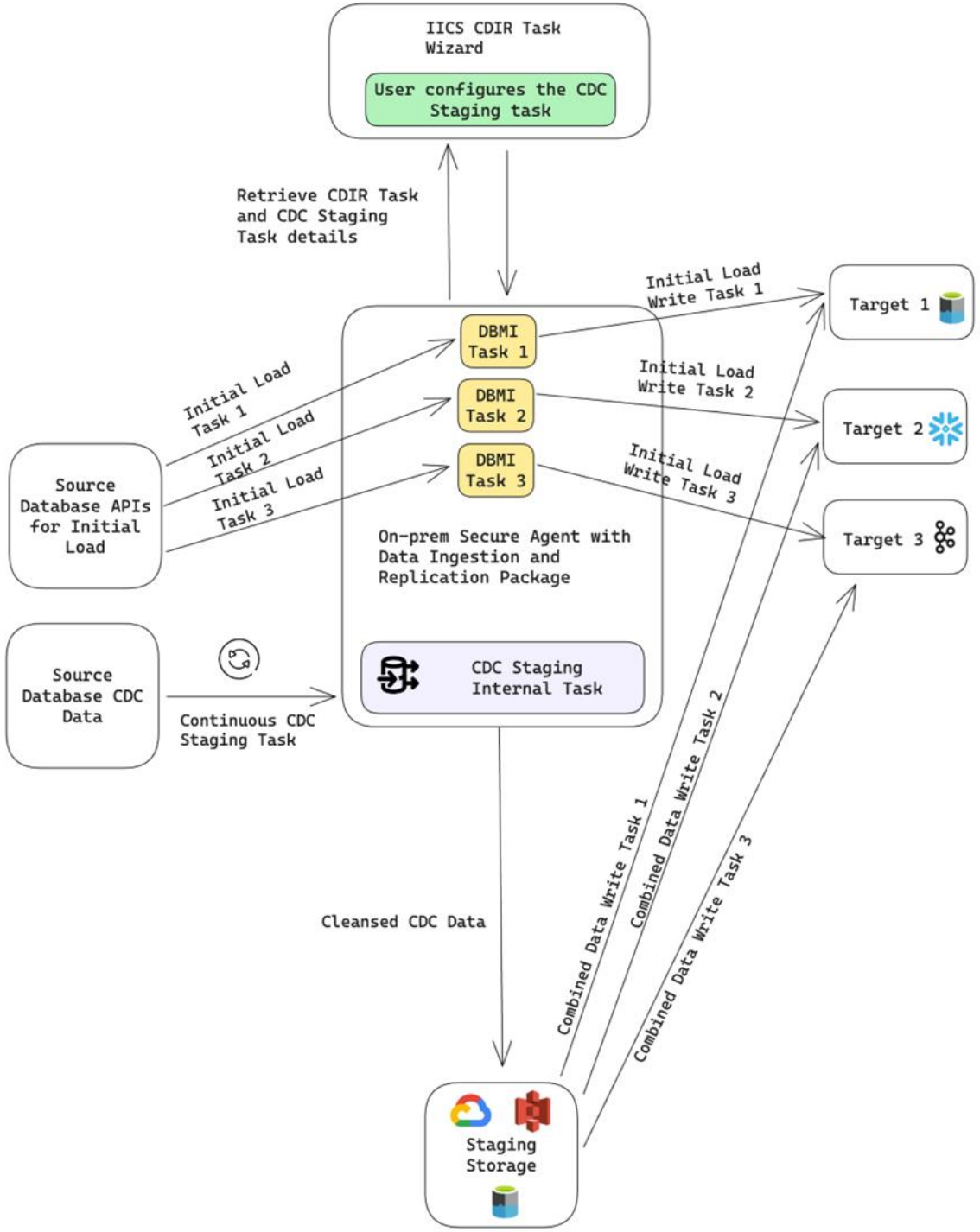
- Read CDC data once and share it across multiple tasks.
- Reduce the load on the source by avoiding multiple connections and repeated reads.

How did we Solve It

- Introduced a staging layer using Apache Iceberg tables on cloud storage to store CDC data after a single read.
- Used a CDC Group Job to read multiple schemas in one connection and let multiple Apply Jobs consume the staged data independently.

CDC Staging Architecture & Data Flow

- **Capture (CDC Group Job):** Ingests changes across multiple schemas via a single connection.
- **Stage (Iceberg on cloud data lake):** Writes all events to structured Iceberg tables (e.g., AllEventTable, SchemaHistoryTable) in S3/ADLS/GCS; retention and log-start are user-configured; **markers** (e.g., CDC_STAGING_FLUSH) ensure consistency and handle fuzzy-unload edge cases.
- **Apply (Downstream consumers):** Apply Jobs read staged data and apply ordered changes to target systems.



Why CDC Staging? Business Justification & Benefits

- **Reduce Source Load** - One connection instead of many reduces overhead on the production database.
- **Multi-Schema Support** - No separate tasks per schema enable simplified governance and monitoring.
- **Improved Performance** - Data is extracted and staged once in the cloud, enabling faster downstream reads from cloud storage and parallel processing without contention.
- **Scalability** - Multiple Apply Jobs consume the same staged data, allowing new pipelines to be added without re-extracting, which scales efficiently with business growth.
- **Maintainability** - Eliminate redundant log scans
- **Cloud Flexibility** - Supports S3, ADLS Gen2, and GCS
- **Customizable Properties** - Log retention (e.g., 14 or 30 days), start points, and flush frequency are tailored to meet specific environment needs.
- **Enterprise-Grade Storage** - Apache Iceberg ensures ACID compliance with snapshot isolation for consistency, along with data versioning and reliability.

Row Filters

Row Filters

Problem

- Large data transfers from source to target might create bottlenecks.
- Target systems receive ALL data, including unnecessary records.
- Unwanted data causes extra processing, overhead, and performance degradation.
- Slower replication and degraded system performance.

Ask

- Need ability to filter rows at source before replication.
- Specify which records actually need to go to target.
- Reduce data volume sent downstream.

Solution

- Filter data at ingestion point based on defined criteria.
- Only relevant rows replicated to target.
- Unnecessary records dropped at source.

- Filters source table rows based on user-defined conditions before applying to target.
- Evaluates each record individually to determine eligibility.
- Writes only records that meet criteria to the target, dropping all others.

Row Filters Architecture & Data Flow

Filter Configuration (CDIR Wizard)

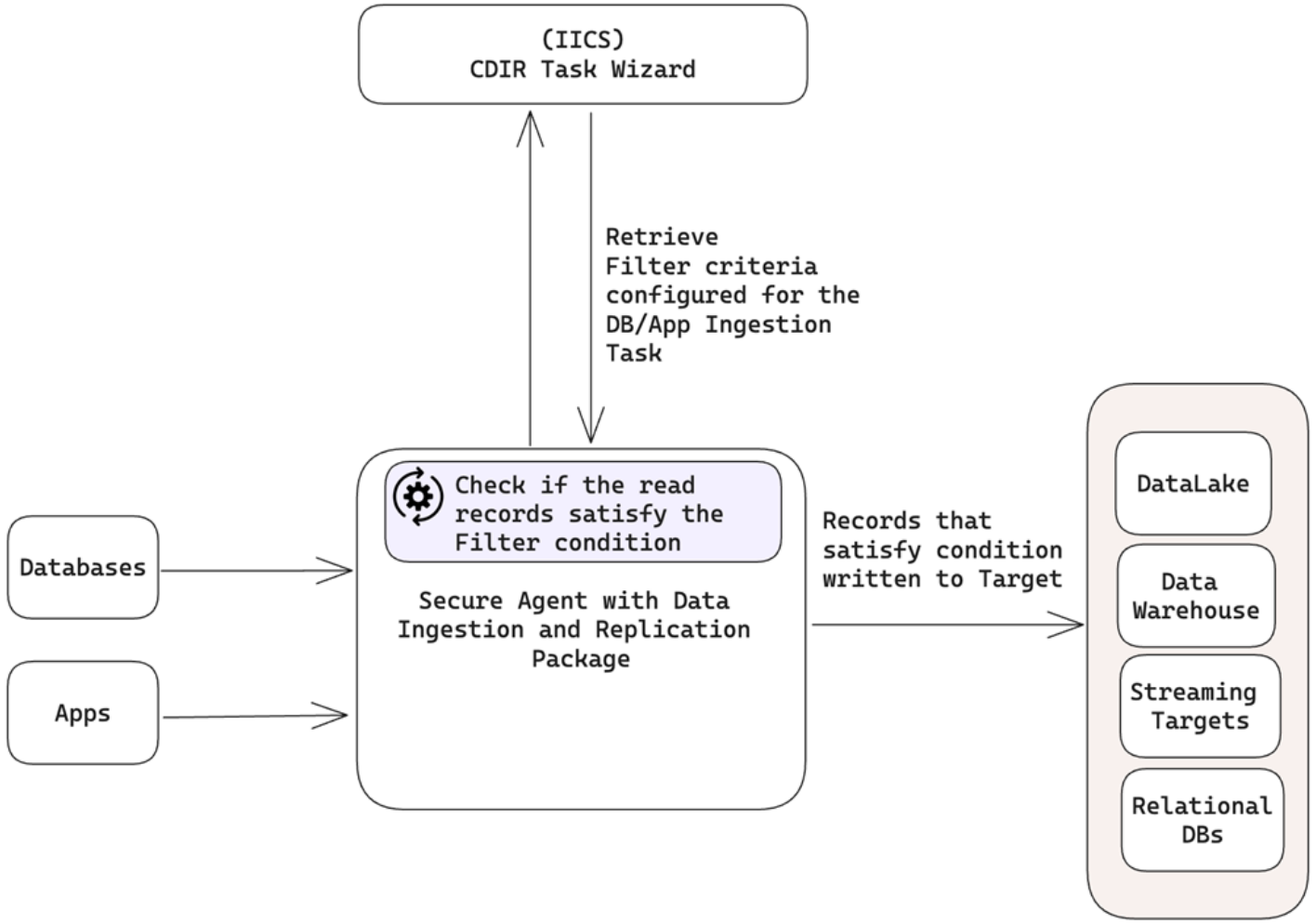
- Define filter condition per table or object
- Select columns to apply the filter on
- Choose filter syntax:
  - Simple
  - Advanced

Execution Flow

- For each record read from the source:
- Evaluate column values against filter criteria
  - If criteria met:
    - Write record to target
  - If criteria not met:
    - Drop record

Applicable Load Types

- Initial loads
- Incremental loads
- Combined loads

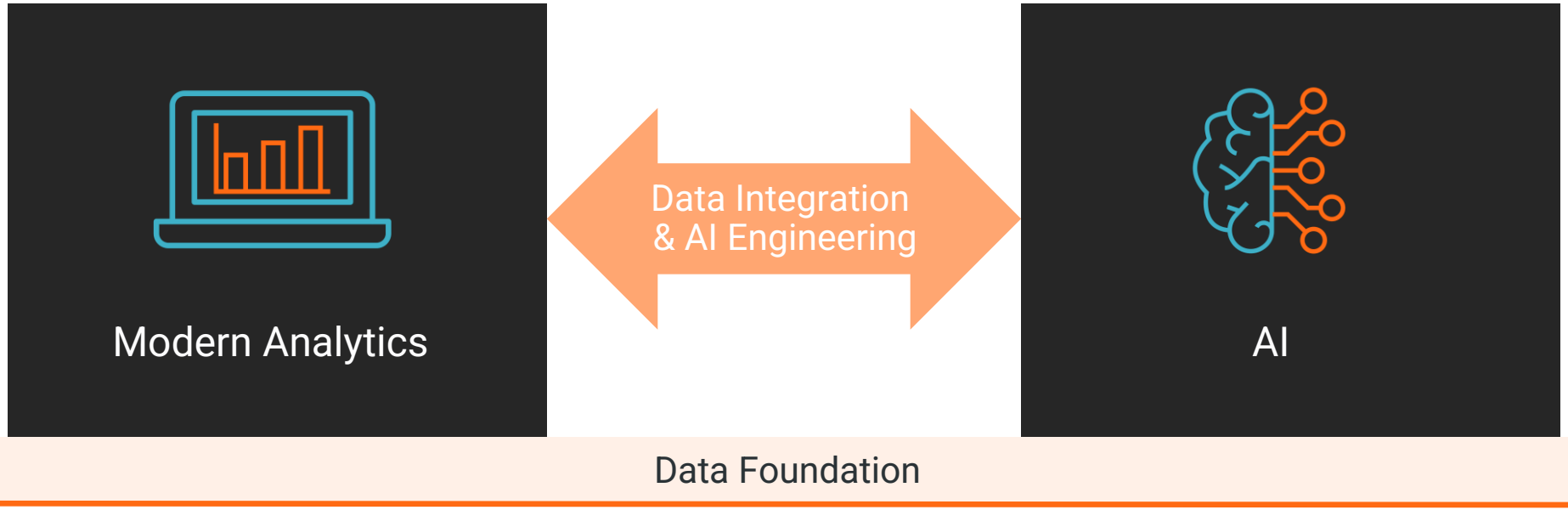


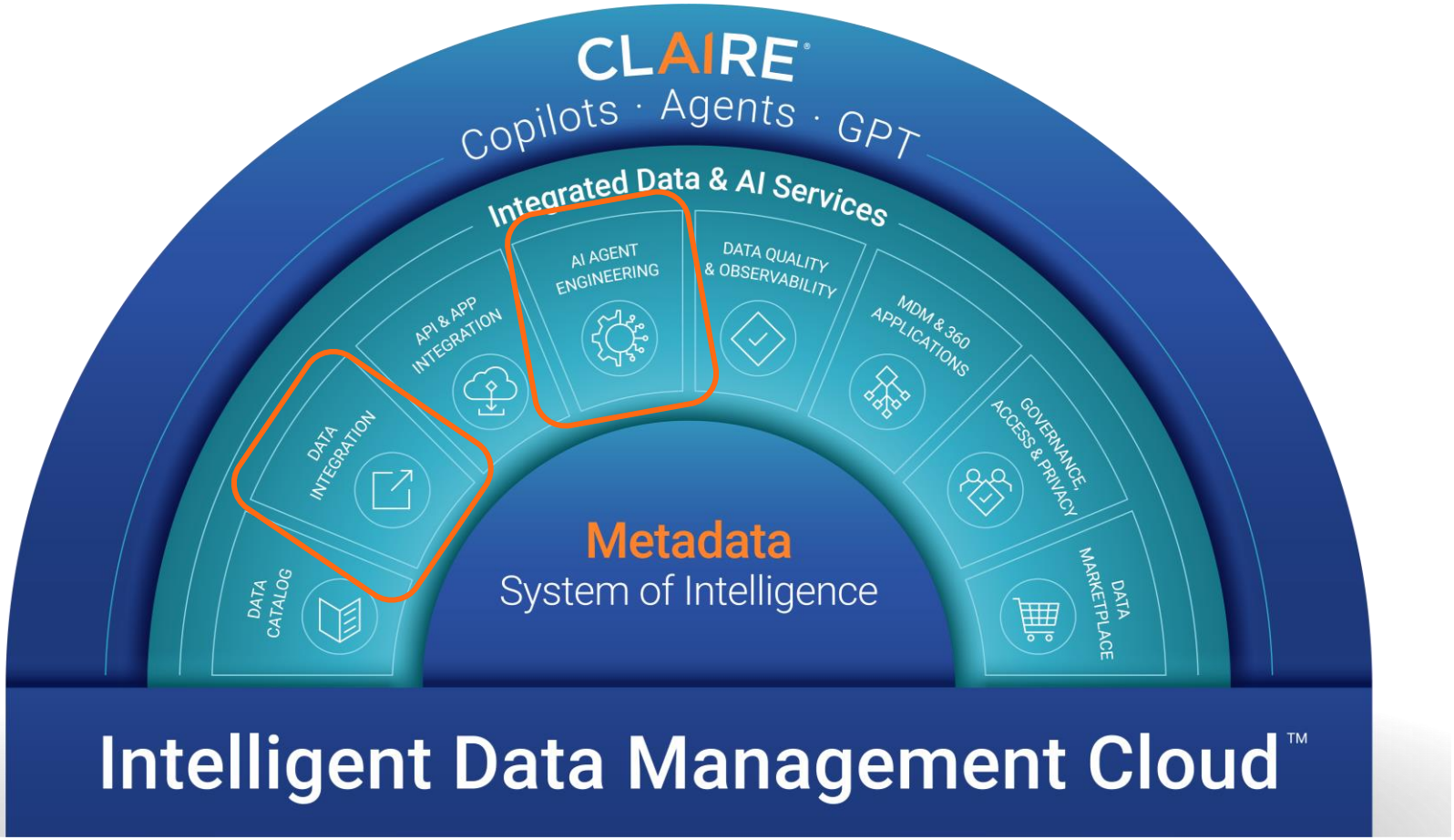
Why Row Filters? Business Justification & Benefits

- **Performance Enhancement**
  - Faster data replication by filtering at source
  - Reduced processing time on target systems
- **Security & Compliance**
  - Control what gets replicated based on policy
  - Meet regulatory requirements, filter out sensitive or unauthorized data
- **Time Savings**
  - Replicate only needed data
  - Faster downstream availability for analytics/reporting
  - Improved productivity

Trends & CDIR Overview

"Modern Analytics and AI are transforming enterprises into insight-driven, adaptive organizations"





Multi-Cloud & Hybrid	Metadata-Aware Conn	Global Scale	Security & Compliance	Flexible Pricing

	Modern Analytics	AI
New Products and Services	Insights into market trends Competitive intelligence	Customer behavior analysis Generate new product ideas
Automating/Optimizing Business Processes	Analytics to study past performance	Predict and forecast outcomes Virtual Assistants AI agents for autonomous actions
Enhance Customer Experience	Improved user experience	Behavior analysis, anticipate needs, and personalize offerings
Operational Efficiencies	Real time performance monitoring	Dynamic resource allocation based on forecasts/ predictions
Risk Management	Detect fraud and asset risks	Fraud prevention Simulate risk scenarios and mitigate potential risks
Employee Empowerment and Productivity	Data democratization enabling collaboration and faster actions and decisions	Augmented AI to increase efficiency and productivity

94%

of leaders expect their Data & Analytics functions to play an important role in enhancing decision making, gaining competitive advantage & delivering business value

92%

of companies plan to increase their AI investments over the next three years to enhance operational efficiency, drive innovation & boost employee productivity

Sources:

1. Data and Analytics Priorities and Challenges: 2024 Trends - Gartner
2. AI in the workplace: A report for 2025 – McKinsey & Company

Informatica

Data Complexity

Applications

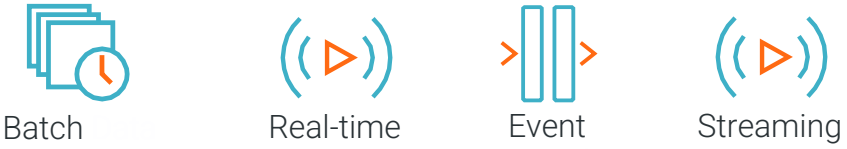
1000s of applications

Data Stores

149 zettabytes of data growing over 25% YoY

Data Structures

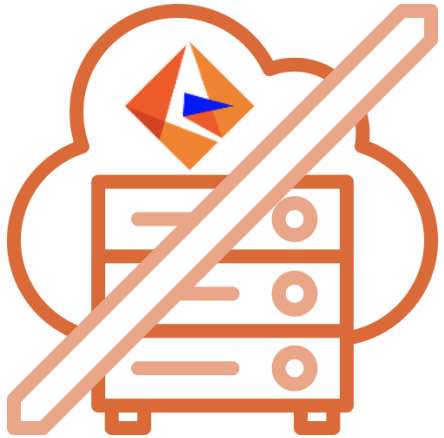
Data Latencies



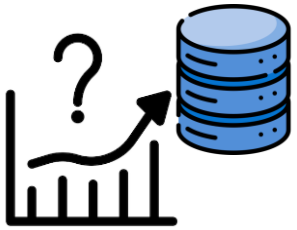
Serverless RTE

Why Serverless?

I don't want to spend time doing infrastructure maintenance and administration.

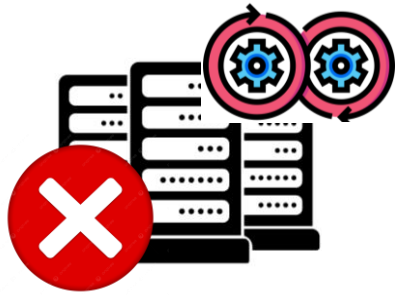


I want my data management staff to be more productive and focus on building solutions for business-needs, rather than investing time managing ingestion pipelines



I have dynamic or seasonally changing infrastructure needs. I want a runtime that can scale up or down on demand

I want to process massively concurrent workloads at scale and with high throughput performance



I want to kickstart my ingestion journey without going through the hassle of infra procurement, provisioning and patching

What does Informatica Serverless RTE for CDIR provide?

Highly Secure

- Informatica's Azure Serverless uses NIC V2 technology for cross-tenant private connectivity
- 7 layers of Enterprise Grade Security protection to protect tenant data.
- Serverless compute in DMZ is part of INFA's SOC2 certification boundary
- Trusted secure link used to link to tenant's network and controlled using tenant's security group and policies

High Performance and configurability

- Configure compute, memory and heap size requirements
- Set up multiple RTEs in INFA provisioned environment
- Connect to on-premise instances using private connectivity
- Support for Database and Application Ingestion sources/targets

How to set up a Serverless RTE?

Create an Informatica Azure Native ISV Resource

Basics | Tags | Review + create

Informatica IDMC is fully managed, cloud-native Data Management Platform

By interacting with the product creation, you consent to share your account information with the product provider for product support and other transactional purposes. For more information, please see the product provider's privacy T&C and Microsoft's Privacy T&C.

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * : Azure R&D CTG_DataEngine_2

Resource group * : DBMIQA_SERVERLESS_AZUREUSW

Create new

Azure resource details

The azure resource representing your Informatica organization, to be used for managing the organization from Azure.

Resource name * : dbmi_qa_resource

Region * : Central US EUAP

Informatica organization details

Create a new Informatica IDMC organization

Informatica Region * : West US 2 (USW1-1)

Organization :
- (•) Create a new organization
- ( ) Link to an existing organization (continue with existing Informatica Billing)

Review + create | < Previous | Next: Tags >

Configuring the resource group which provides isolation to the generated VMs, application and network security groups. The NSGs control the traffic to the SRTE in INFA tenant

This is a region to create a resource to manage the metadata of an INFA org that the customer either creates afresh or adds an already existing one to this Azure account. This resource manages the metadata of the INFA org and could be placed in ***any*** region.

Create an Informatica Azure Native ISV Resource

Resource name * dbmi_qa_resource

Region * Central US EUAP

Informatica organization details

Create a new Informatica IDMC organization

Informatica Region * West US 2 (USW1-1)

Organization
- (•) Create a new organization
- ( ) Link to an existing organization (continue with existing Informatica Billing)

Organization name: Organization name is generated based on Azure resource name

Plan: **CDI Free (Private Preview)**
Change plan

Billing term: (•) 1-Year

Price + Payment options: US$0.00/month

Subtotal: US$0.00 for 1 year

Experience the industry's only free, AI-powered solution to easily load, transform and integrate data

This is the region for an Informatica org. This could either be a new org which the user can create here or an existing org that the customer can link to this Azure resource. This region determines the corresponding regions where INFA can host the SRTE agent to run customer workloads

Create an Informatica Azure Native ISV Resource

Basics | Tags | Review + create

Tags are name/value pairs that enable you to categorize Azure resources and view consolidated billing by applying the same tag to multiple resources and resource groups. Learn More

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name	Value	Resource
		Informatica

Review + create | < Previous | Next: Review + create >

Create an Informatica Azure Native ISV Resource

Validating

Basics Tags Review + create

Basics

Subscription	Azure R&D CTG_DataEngine_2
Resource Group	DBMIQA_SERVERLESS_AZUREUSW
Resource name	dbmi_qa_resource
Region	Central US EUAP
Plan	CDI Free (Private Preview)
Billing term	1-Year
Price + Payment options	US$0.00/month
Subtotal	US$0.00 for 1 year

Terms

Terms of use | Privacy Policy

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with each Marketplace offering above, (b) authorize Microsoft to charge or bill my current payment method for the fees associated with my use of the offering(s), including applicable taxes, with the same billing frequency as my Azure subscription, until I discontinue use of the offering(s), and (c) agree that Microsoft may share my contact information and transaction details (including usage volume associated with the offering) with the seller(s) of the offering(s). Microsoft does not provide rights for third-party products or services. See the Azure Marketplace Terms for additional terms.

By clicking "Create", you give Microsoft permission to use or share your account information so that the provider or Microsoft can contact you regarding this product and related products.

Name	cdir cdir
Preferred e-mail address	cdir@InformaticaLLC.onmicrosoft.com

Create < Previous Next

DBMI_QA_SERVERLESS

Informatica Organization

Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Resource visualizer
Settings
Locks
Properties
Informatica
Serverless Runtime Environment
Help
Automation

Refresh Delete

Essentials

JSON View

Resource group (move)	: DBMIQA_SERVERLESS_AZUREUSW	SSO Login	: Login to IDMC organization
Location	: centraluseuap	Status	: Succeeded
Subscription (move)	: Azure R&D CTG_DataEngine_2	Plan	: CDI Free - Private Preview
Subscription ID	: 3aca207a-ac1f-4098-b412-ec4edfbbc4cb	Module	: Data Integration

Tags (edit) : businessunit : INTCLOUD | businessentity : MASSINGESTION | owneremail : alkumar@informatica.com | owner : RSPENCER

Get started with Informatica Intelligent Data Management Cloud

Launch IDMC

The getting started guide gets you up and running with Informatica on Azure as a Service.

IDMC Account Login

Getting Started

The getting started guide gets you up and running with Informatica on Azure as a Service.

View guide

New Feature - Data Ingestion/Replication

The Serverless runtime environment now supports creating data ingestion and replication tasks as well.

Learn more

DBMI_QA_SERVERLESS | Serverless Runtime Environment

Informatica Organization

Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Resource visualizer
Settings
Locks
Properties
Informatica
Serverless Runtime Environment
Help
Automation

Create Serverless Runtime Environment | Refresh

Search | Type : All | Name : All | Status : All

Name	Description	Type	Status	Status Message	Updated Time	Created Time
DBMI		Data Integration and...	Running	Environment is running.	2026-02-19T07:43:55.000Z	2025-05-13T06:39:20.000Z
DBMI_3		Data Integration and...	Running	Environment is running.	2026-02-19T09:35:31.000Z	2025-06-11T06:57:41.000Z
DBMI_OP		Data Integration and...	Failed		2026-02-17T13:20:13.000Z	2025-07-04T11:34:19.000Z
DBMI_REL_WEST_US		Data Integration and...	Failed	Contact Informatica Global Cust...	2026-02-10T08:19:28.000Z	2025-08-20T08:40:49.000Z

Create Serverless Runtime Environment

Create a serverless runtime environment for your informatica organization

Name * : Gartner_Serverless_RTE

Description : Enter Description

Task Type * : Data Integration and Data Ingestion/Replication

Maximum Compute Units Per Task * :

Task Timeout (Minutes) * : 2880

Select Data Integration and Data Ingestion/Replication

This does not apply to CDIR as the 'max compute units per task' calculation is done internally

Clone a Serverless Runtime Environment

Basics | **Platform Detail** | Runtime Configuration | Review+Create

Region * centraluseuap

Virtual network DBMI_VN (CDIEQA_AVS_EWE_RG)
Edit virtual network

Subnet * DBMI_SUBNET
Edit subnet 10.0.1.0 - 10.0.1.255 (256 addresses)

Managed Identity Select Managed Identity

Supplementary file location abfs://dbmiserverless@dbmistorageaccount.dfs.core.windows.net/dbmi

Region in which to set up the Serverless RTE. It is a limited list governed by the region of the INFA Azure pod which hosts this organization

Data Disks

Type	Server host/IP	Source mount	Target mount	Mount options
NFS	Enter details	Enter details	Enter details	Enter details

Custom properties

Key	Value
ServicePrincipalId	3e9a333c-21de-4a7f-afa8-009d3fb88a67
ServicePrincipalSecret	********

< Previous | Next: Runtime Configuration > | Review+Create

Clone a Serverless Runtime Environment

Basics | Platform Detail | **Runtime Configuration** | Review+Create

Service : **Database_Ingestion** | Type : **All** | Reset All | Add Property

Service	Type	Name	Value
Database_Ingest...	DBMI_AGEN...	serviceLogRetentionPe...	7
Database_Ingest...	DBMI_AGEN...	taskLogRetentionPeriod	7
Database_Ingest...	DBMI_AGEN...	serviceUrl	
Database_Ingest...	DBMI_AGEN...	logLevel	'TRACE'
Database_Ingest...	DBMI_AGEN...	useProxy	false
Database_Ingest...	DBMI_AGEN...	supportedLoadTypes	'INITIAL,INCREMENTAL...
Database_Ingest...	EXTERNAL_T...	maxTaskUnits	150
Database_Ingest...	EXTERNAL_T...	clusterStartupTimeout...	30
Database_Ingest...	EXTERNAL_T...	clusterStartupPollingInt...	5
Database_Ingest...	EXTERNAL_T...	clusterAdvancedSched...	true
Database_Ingest...	EXTERNAL_T...	taskLogSubFolder	'logs/infaJobLogs/dbmi'
Database_Ingest...	EXTERNAL_T...	taskStartTimeoutSecon...	900
Database_Ingest...	EXTERNAL_T...	taskResourceRequests...	'3000m'
Database_Ingest...	EXTERNAL_T...	taskResourceRequests...	'4Gi'
Database_Ingest...	DBMI_AGEN...	testProperty	'testValue'

Here you could specify the CPU and Memory requests and limits for individual tables

< Previous | Next: Review+Create > | Review+Create

Clone a Serverless Runtime Environment

Validation Passed

Basics | Platform Detail | Runtime Configuration | **Review+Create**

Basics

Name	Gartner_Serverless_RTE
Description	
Task Type	CDI
Max Compute Unit	1
TimeOut	2880

Platform Details

Region	centraluseuap
Virtual Network	/subscriptions/3aca207a-ac1f-4098-b412-ec4edfbbc4cb/resourceGroups/CDIEQA_AVS_EWE_RG/providers/Microsoft.Network/virtualNetworks/DBMI_VN
Subnet	/subscriptions/3aca207a-ac1f-4098-b412-ec4edfbbc4cb/resourceGroups/CDIEQA_AVS_EWE_RG/providers/Microsoft.Network/virtualNetworks/DBMI_VN/subnets/DBMI_SUBNET
Supplementary File Location	abfs://dbmiserverless@dbmistorageaccount.dfs.core.windows.net/dbmi
Managed Identity	

Data Disks

Custom Properties

ServicePrincipalId	3e9a333c-21de-4a7f-afa8-009d3fb88a67
ServicePrincipalSecret	********

< Previous | Next > | Clone

DBMI_QA_SERVERLESS

Informatica Organization

Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Resource visualizer
Settings
Locks
Properties
Informatica
Serverless Runtime Environment
Help
Automation

Refresh Delete

Essentials

JSON View

Resource group (move)	: DBMIQA_SERVERLESS_AZUREUSW	SSO Login	: Login to IDMC organization
Location	: centraluseuap	Status	: Succeeded
Subscription (move)	: Azure R&D CTG_DataEngine_2	Plan	: CDI Free - Private Preview
Subscription ID	: 3aca207a-ac1f-4098-b412-ec4edfbbc4cb	Module	: Data Integration

Tags (edit) : businessunit : INTCLOUD | businessentity : MASSINGESTION | owneremail : alkumar@informatica.com | owner : RSPENCER

Get started with Informatica Intelligent Data Management Cloud

Launch IDMC

The getting started guide gets you up and running with Informatica on Azure as a Service.

IDMC Account Login

Getting Started

The getting started guide gets you up and running with Informatica on Azure as a Service.

View guide

New Feature - Data Ingestion/Replication

The Serverless runtime environment now supports creating data ingestion and replication tasks as well.

Learn more

Ecosystem SSO
Organization
Licenses
SAML Setup
Metering
Settings
Users
User Groups
User Roles
Runtime Environ...
Serverless Enviro...
Connections
Add-On Connecto...
Schedules
Add-On Bundles
Swagger Files
Logs
Advanced Clusters
Collapse

Serverless Environments

New Serverless Runtime Environment

Manage the serverless runtime environments that can connect to your private cloud.

Serverless Runtime Environments

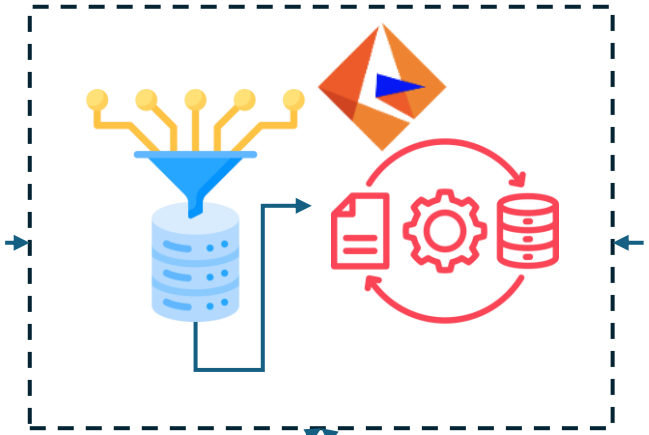
Name	Description	Type	Cloud Platform	Status	Status Message	Updated Time	Created Time
DBMI		Data Integration	AZURE	Up and Running	Environment is runn...	Feb 18, 2026, 11:43 ...	May 12, 2025, 11:...
DBMI_3		Data Integration	AZURE	Up and Running	Environment is runn...	Feb 19, 2026, 1:35 ...	Jun 10, 2025, 11:5...
DBMI_OP		Data Integration	AZURE	Failed		Feb 17, 2026, 5:20 ...	Jul 4, 2025, 4:34 A...
DBMI_REL_WEST_US		Data Integration	AZURE	Failed	Contact Informatica...	Feb 10, 2026, 12:19 ...	Aug 20, 2025, 1:40...

Orchestration-Data to Insights

CDIR integration with Taskflow

Why Taskflow Integration?

I want to continuously check data quality as my data is ingested every cycle



I want to standardize my KPI fields on the incremental data

I want the business logic data enrichment to be orchestrated seamlessly on event basis

I want to notify my data consumers once after each refresh cycle

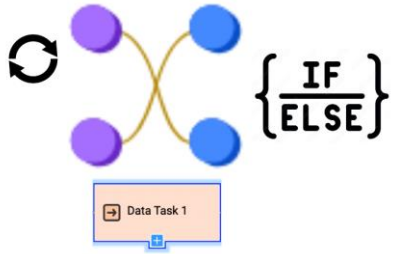
What does Taskflow Integration provide for CDIR customers

- Automatically triggers a Taskflow workflow post ingestion and replication

- Adds cycle IDs in the payload to let downstream processes identify changed records in a cycle

item	order_ date	return_ date	cycle_id
ball	Dec 21	NA	2
bat	Dec 22	NA	2
stump	Jan 1	Jan 4	3

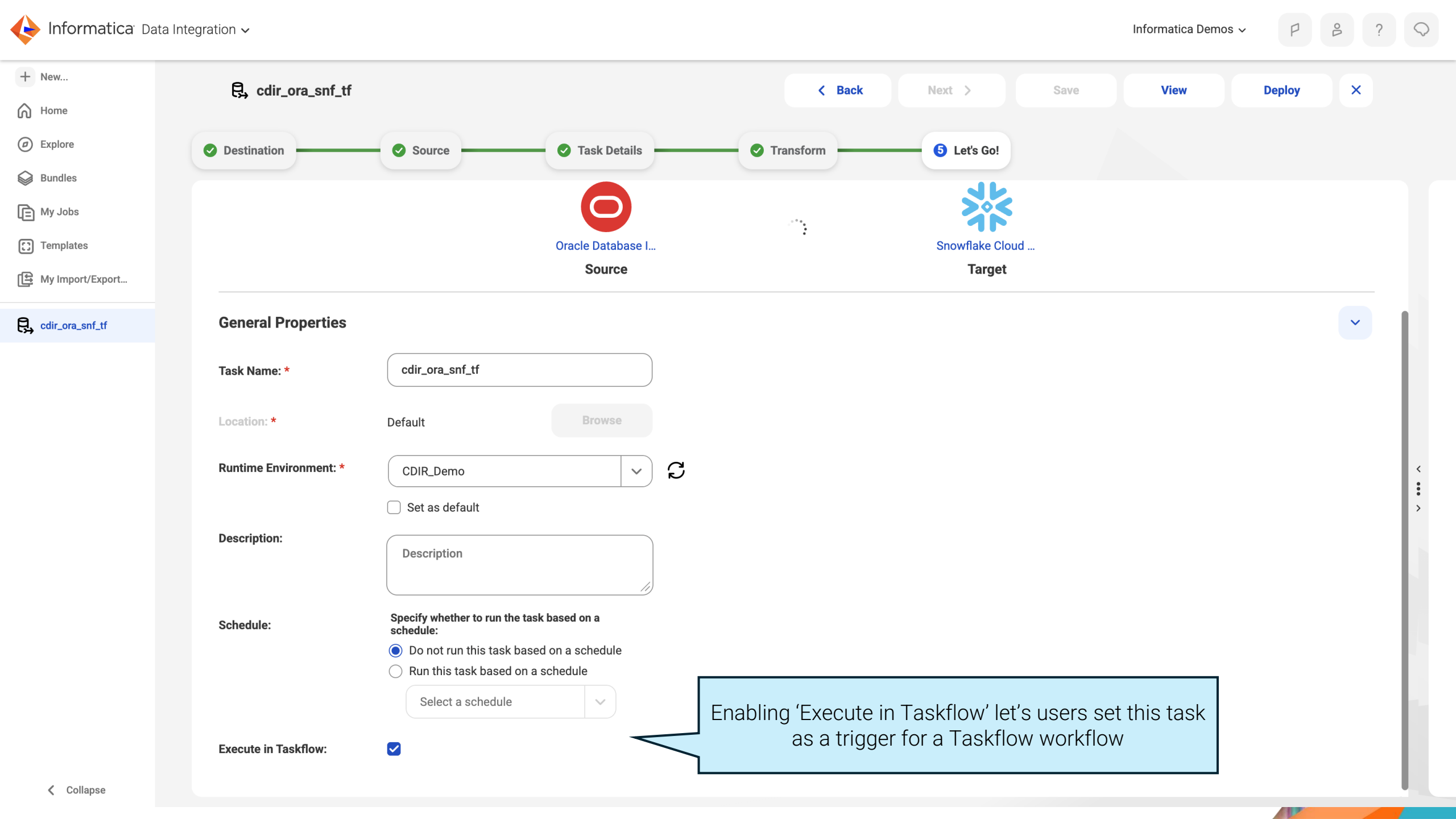
- Enables complex transformation logic using loops, conditions, expressions and mappings to be applied on ingested data

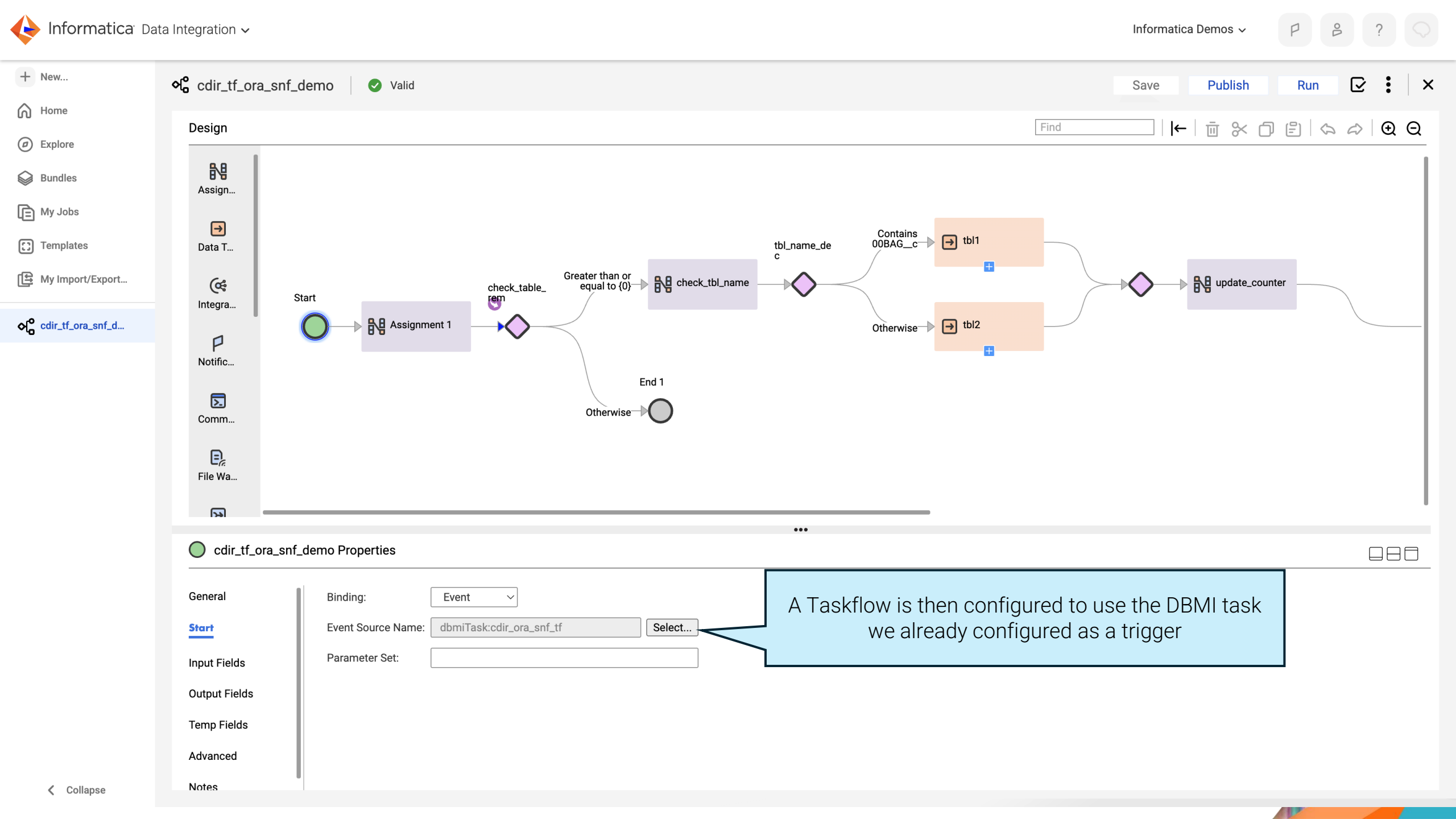


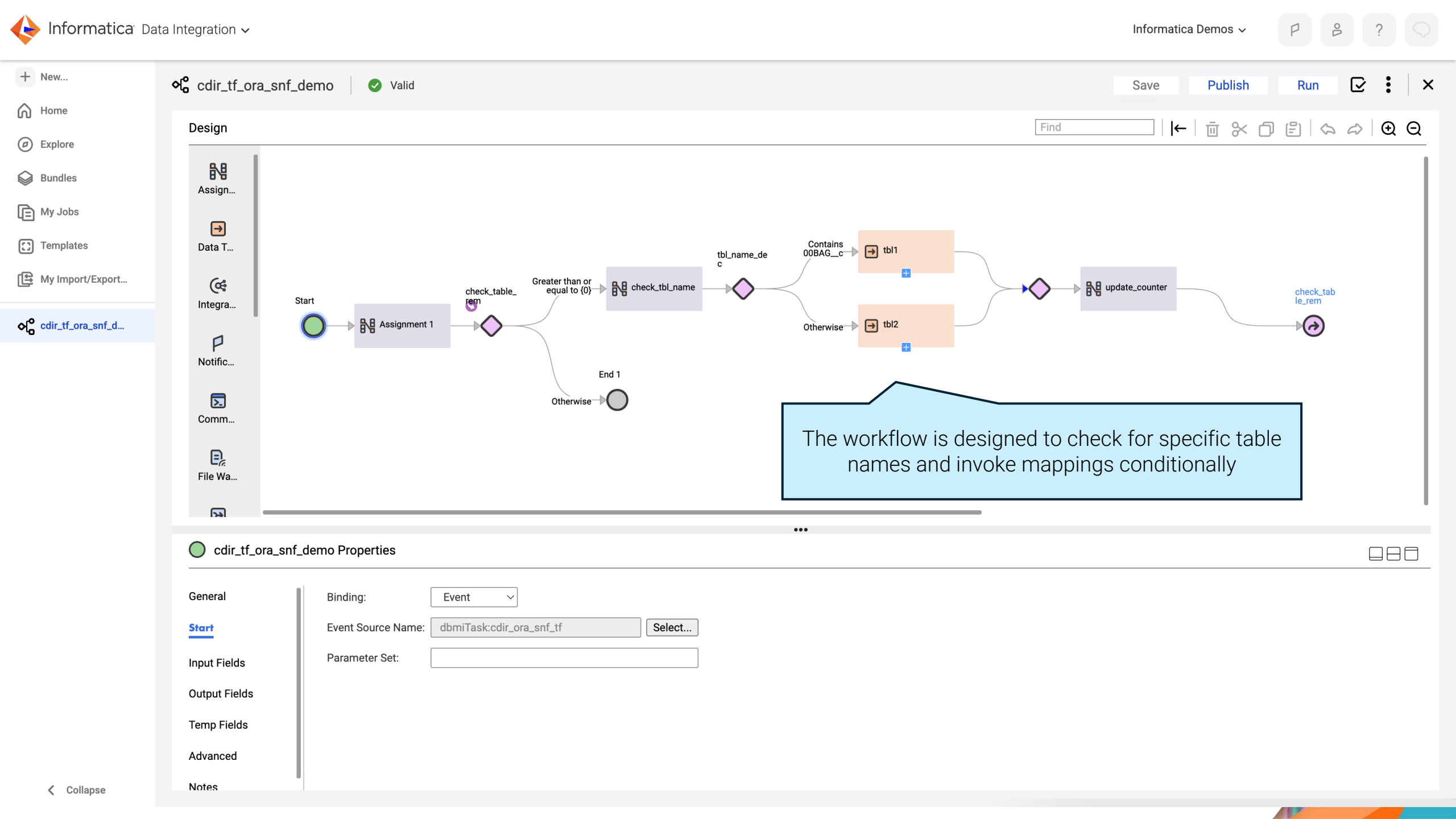
- Lets downstream tasks identify updated or resynced tables, processed cycle IDs and if a schema has changed on a table

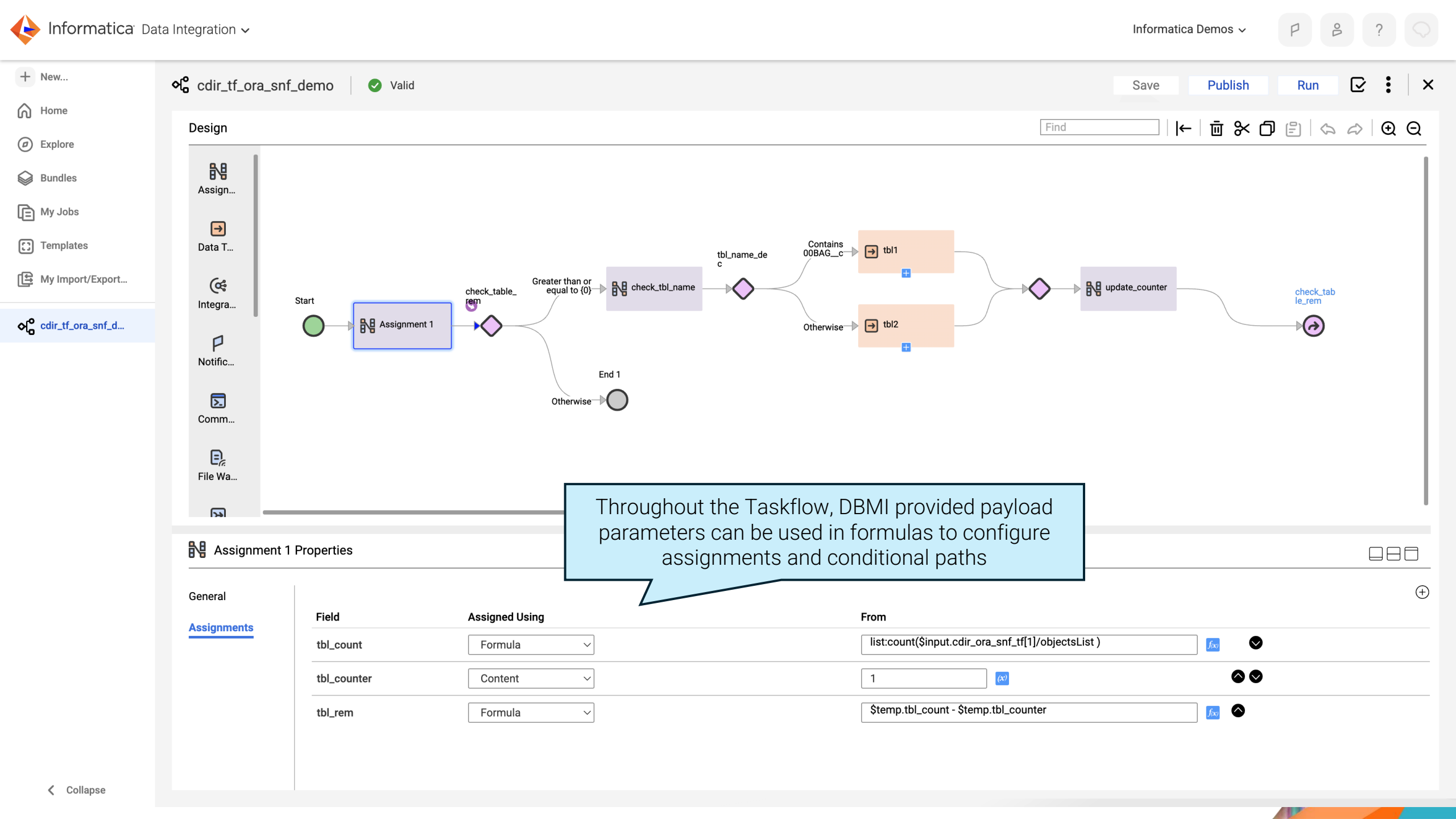
- **Retail_customers_to_Snowflake**
  - **Max Cycle Id**
  - **Min Cycle Id**
  - **Schema Name**
  - **Object List**
    - **Name**
    - **Resynced Cycle Id**

A quick run down


Informatica Data Integration
Informatica Demos
New...
Home
Explore
Bundles
My Jobs
Templates
My Import/Export...
cdir_ora_snf_tf
Collapse
cdir_ora_snf_tf
Back
Next
Save
View
Deploy
Destination
Source
Task Details
Transform
5 Let's Go!
Oracle Database I...
Source
Snowflake Cloud ...
Target
General Properties
Task Name: *
cdir_ora_snf_tf
Location: *
Default
Browse
Runtime Environment: *
CDIR_Demo
Set as default
Description:
Description
Schedule:
Specify whether to run the task based on a schedule:
Do not run this task based on a schedule
Run this task based on a schedule
Select a schedule
Execute in Taskflow:
Enabling 'Execute in Taskflow' let's users set this task as a trigger for a Taskflow workflow

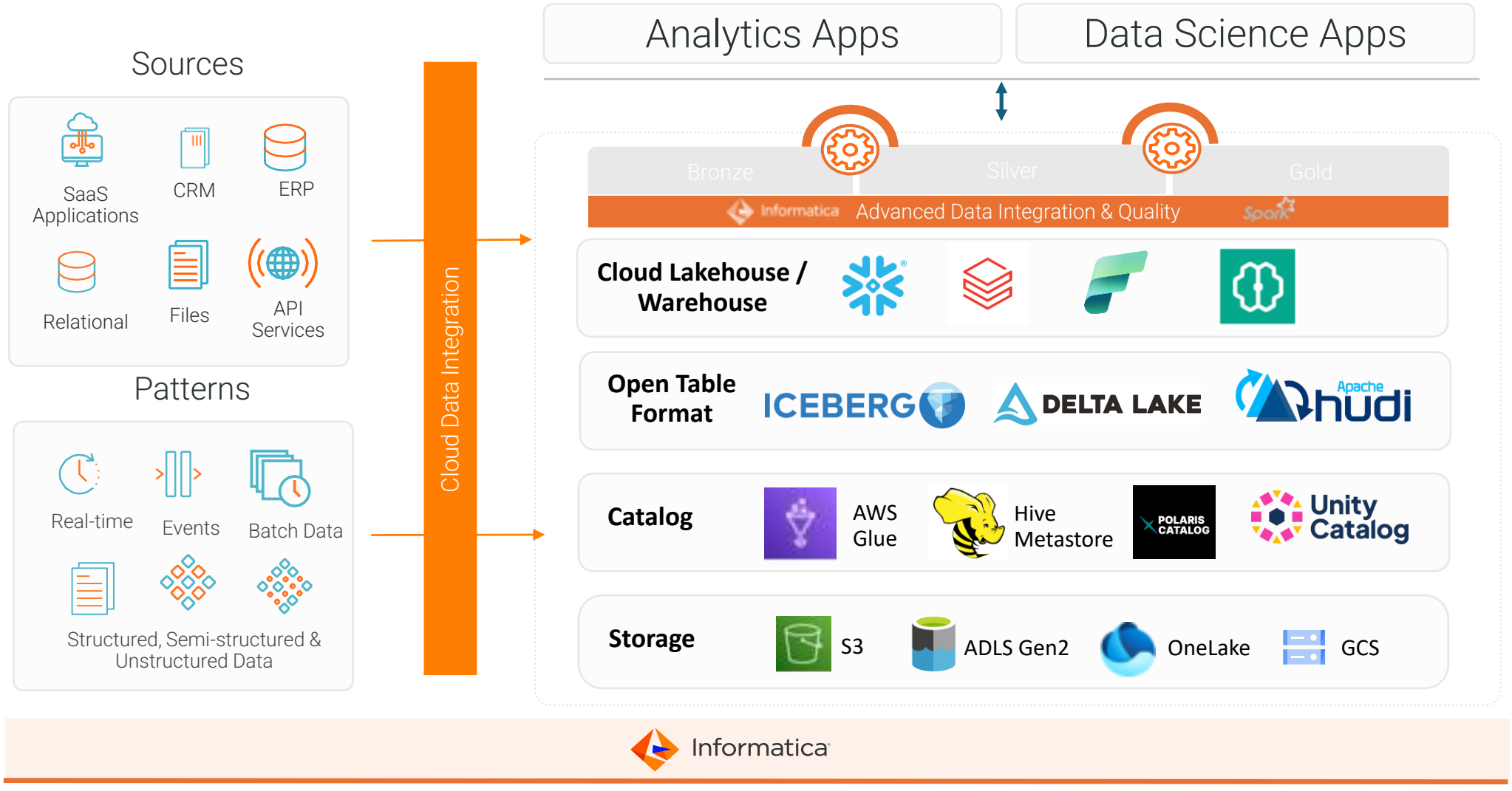

Informatica Data Integration
Informatica Demos
New...
Home
Explore
Bundles
My Jobs
Templates
My Import/Export...
cdir_tf_ora_snf_d...
Collapse
cdir_tf_ora_snf_demo
Valid
Save
Publish
Run
Design
Find
Assign...
Data T...
Integra...
Notific...
Comm...
File Wa...
Start
Assignment 1
check_table_
rem
Greater than or
equal to {0}
check_tbl_name
tbl_name_de
c
Contains
00BAG__c
tbl1
Otherwise
tbl2
update_counter
End 1
Otherwise
cdir_tf_ora_snf_demo Properties
General
Start
Input Fields
Output Fields
Temp Fields
Advanced
Notes
Binding:
Event
Event Source Name:
dbmiTask:cdir_ora_snf_tf
Select...
Parameter Set:
A Taskflow is then configured to use the DBMI task
we already configured as a trigger


Informatica Data Integration
Informatica Demos
New...
Home
Explore
Bundles
My Jobs
Templates
My Import/Export...
cdir_tf_ora_snf_d...
cdir_tf_ora_snf_demo
Valid
Save
Publish
Run
Design
Find
Assign...
Data T...
Integra...
Notific...
Comm...
File Wa...
Start
Assignment 1
check_table_rem
Greater than or equal to (0)
check_tbl_name
tbl_name_dec
Contains 00BAG__c
tbl1
Otherwise
tbl2
update_counter
check_table_rem
End 1
Otherwise
The workflow is designed to check for specific table names and invoke mappings conditionally
cdir_tf_ora_snf_demo Properties
General
Start
Input Fields
Output Fields
Temp Fields
Advanced
Notes
Binding:
Event
Event Source Name:
dbmiTask:cdir_ora_snf_tf
Select...
Parameter Set:
Collapse


Informatica Data Integration
Informatica Demos
New...
Home
Explore
Bundles
My Jobs
Templates
My Import/Export...
cdir_tf_ora_snf_d...
Collapse
cdir_tf_ora_snf_demo
Valid
Save
Publish
Run
Design
Find
Assign...
Data T...
Integra...
Notific...
Comm...
File Wa...
Start
Assignment 1
check_table_rem
Greater than or equal to (0)
check_tbl_name
tbl_name_dec
Contains 00BAG__c
tbl1
Otherwise
tbl2
update_counter
check_table_rem
End 1
Otherwise
Throughout the Taskflow, DBMI provided payload parameters can be used in formulas to configure assignments and conditional paths
Assignment 1 Properties
General
Assignments
Field
Assigned Using
From
tbl_count
Formula
list:count($input.cdir_ora_snf_tf[1]/objectsList)
tbl_counter
Content
1
tbl_rem
Formula
$temp.tbl_count - $temp.tbl_counter

Open Table

Modern Lakehouse Architecture with Open Table Formats



CDIR S3 Iceberg Support

Business Challenge:

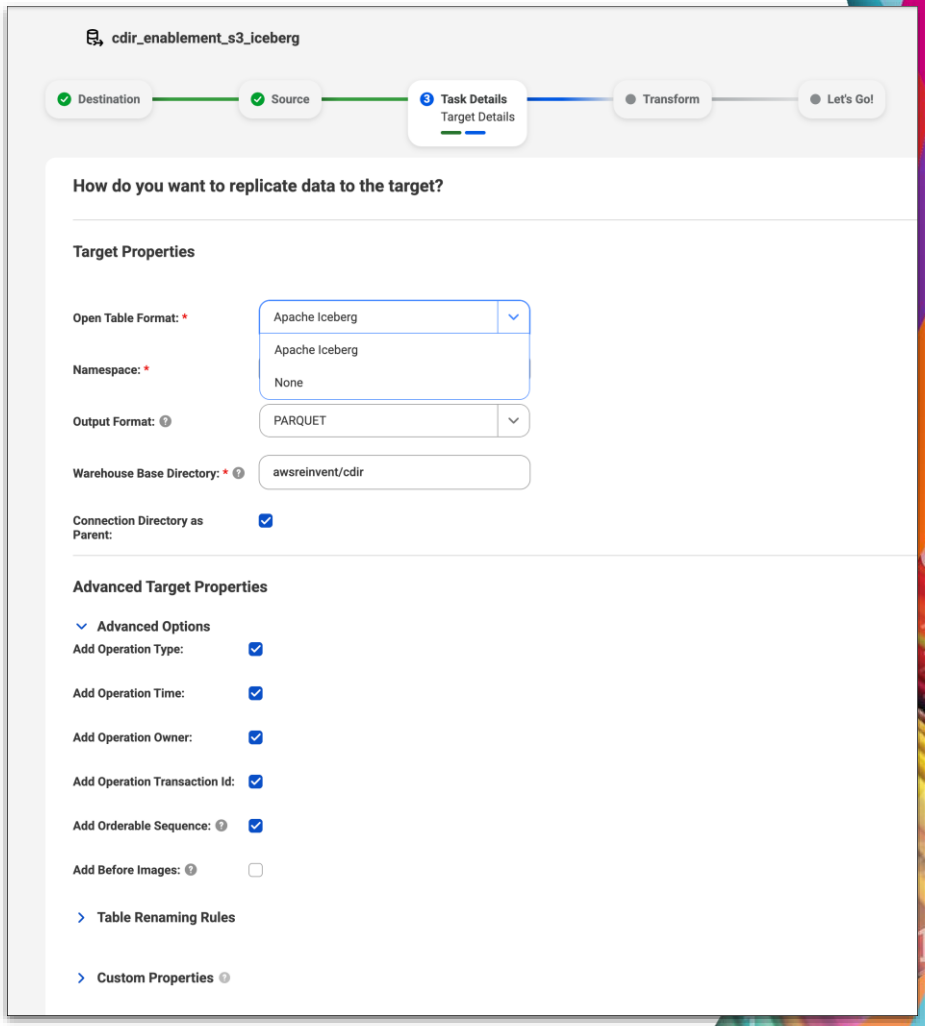
While Amazon S3 is widely supported as a target for database ingestion and replication tasks with flexible output formats such as CSV, Avro, and Parquet, many organizations are increasingly adopting the Iceberg table format to implement modern lake house architectures. However, the lack of native support for Iceberg in current ingestion workflows limits the ability to leverage its advanced features—such as ACID compliance, schema and partition evolution, and time travel—which are critical for managing complex data lakes with warehouse-like capabilities. This gap hampers data teams' agility, slows innovation, and prevents businesses from fully unlocking the value of their data lake investments.

Capability:

We now support initial and incremental ingestion to Amazon S3 using the Apache Iceberg open table format with Cloud Data Ingestion and Replication, providing seamless integration with Glue Catalog and optimized handling of large-scale data lakes.

Benefits to Informatica Customers:

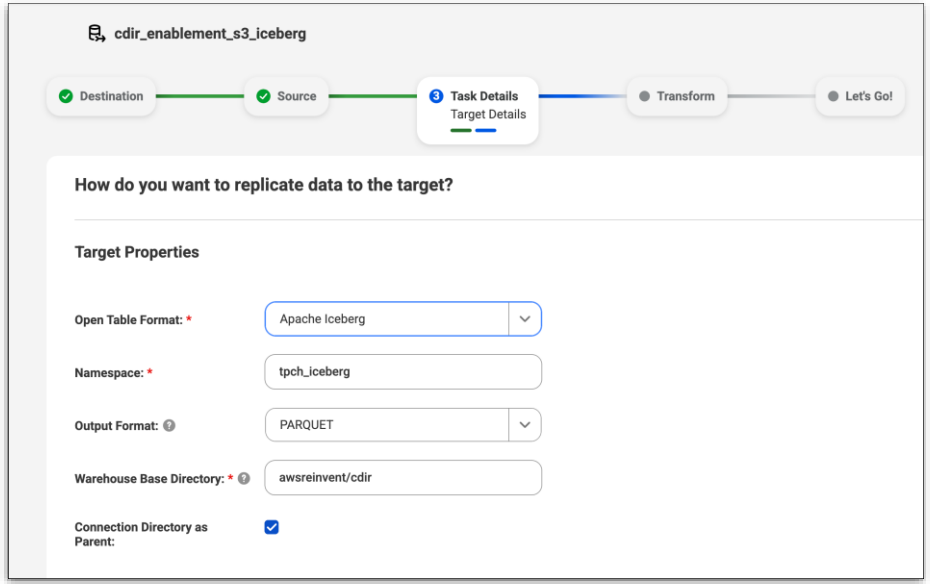
- Unlock modern lake house capabilities with Iceberg table format support for Amazon S3 targets.
- Seamlessly integrate advanced features like ACID compliance and schema evolution into existing ingestion workflows.
- Speed up data ingestion with seamless integration into AWS Glue Catalog.
- Expand connector options to boost agility and efficiency across your cloud data architecture.

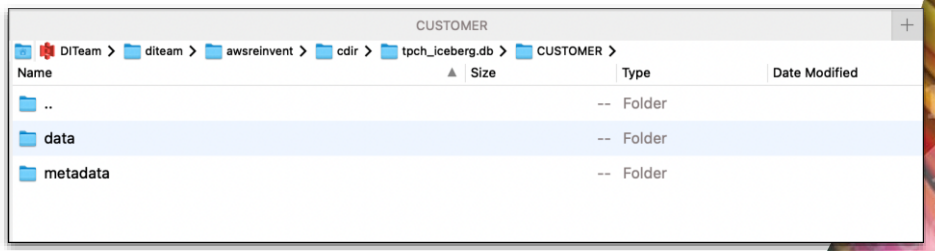


CDIR S3 Iceberg Support

Key Features

- New UI fields are added to support Iceberg Open Table format for S3 target.
  - Supports Initial Load (by default) for October release with out custom org flag
  - "Open Table Format" - Default value is "NONE"- works as traditional file formats
  - "Output Format" - Default value is PARQUET (Currently it is the only value supported)
  - "Namespace" - Name of the AWS Glue database tables will reside under
  - "Warehouse Base Directory" - This field holds text as value and, it's the folder path in S3
  - "Connection Directory as Parent" - Default value is "true" (checkbox)
- By default, iceberg table format option will only show for initial load task to S3 target.
- For selecting iceberg table format for S3 target for incremental and combined load, the below org level feature flag should be set to true.
  - cdir.s3.iceberg.cdc.combined.mode.enabled true/false





Informatica
from Salesforce

Thank You!

Where data & AI come to LIFE