



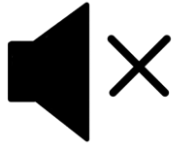
22 July, 2025

Leveraging Real-time Analytics with Ingestion, Replication and Change Data Capture

- Bharath Raghavendran, Principal Product Manager, 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.

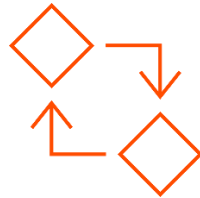
Feature Rich Success Portal



Bootstrap trial and
POC Customers



Enriched Customer
Onboarding
experience



Product Learning
Paths and Weekly
Expert Sessions

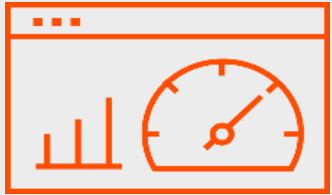


Informatica
Concierge



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

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

1 CDIR Overview

2 Connectors and CDC

3 Streaming sources and targets

4 Real time replication - Snowflake

5 Demo

6 Q&A

CDIR Overview

Cloud Data Ingestion and Replication

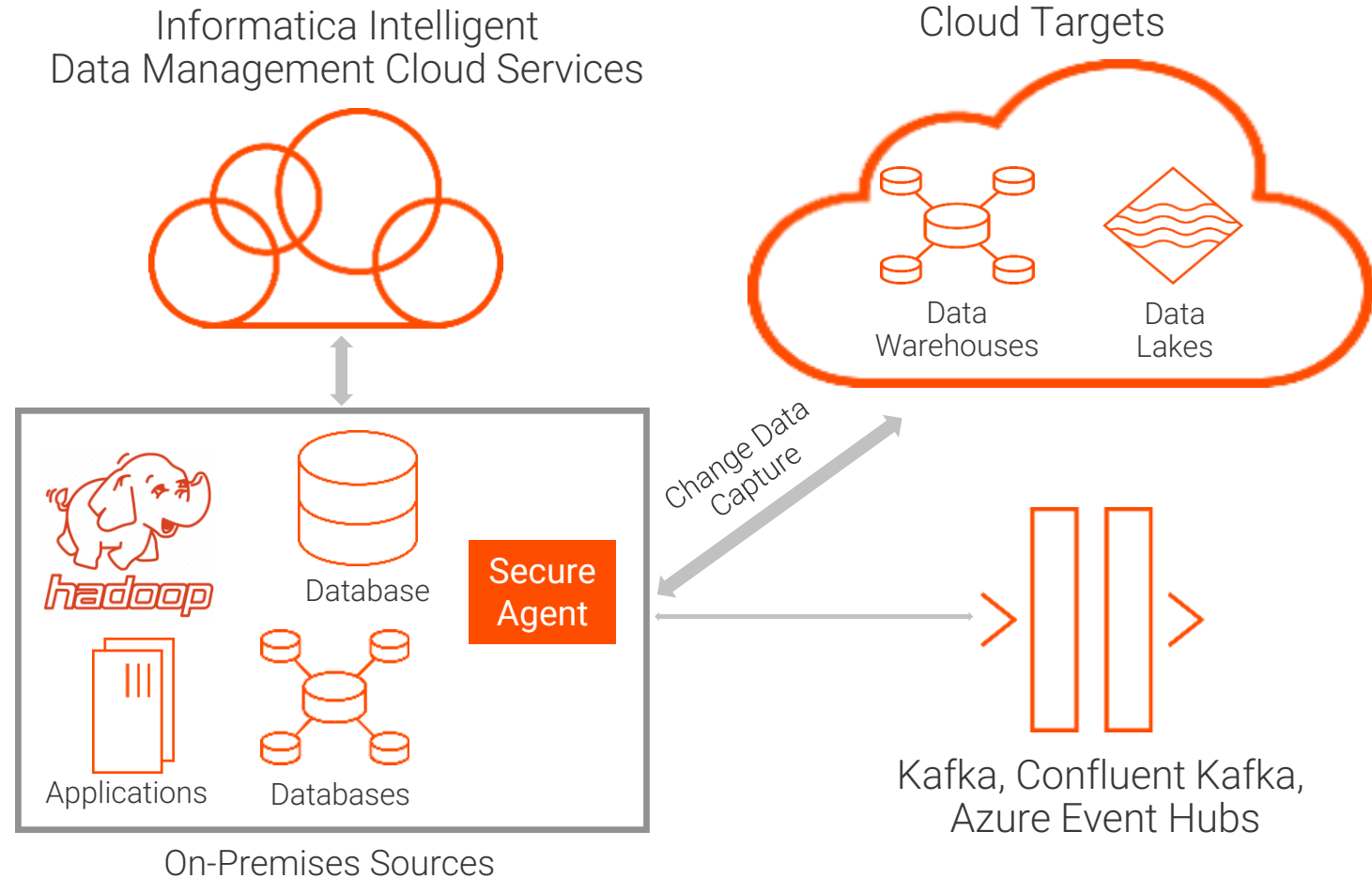
Apps, files, streaming, DBs and CDC

Provides wizard approach to quickly ingest data into the cloud (or to wherever the Secure Agent can reach)

Maintain synchronization between applications and sources as they continue to change

Real-time monitoring of ingestion jobs with lifecycle management and alerting in case of issues

Orchestrate data ingestion in **hybrid/cloud** as **managed** and **secure** service



CDIR – Low Code wizard

Why best

- The only **unified platform** for ingestion, synchronization, replication/CDC and transformation. With one single UI for different source patterns such as databases, applications, streaming, files, etc.
- Easily ingest/replicate enterprise data using batch, streaming, real-time and change data capture (CDC) into cloud data warehouses, lakes and messaging hubs.
- Automate job creation with a **CLAIRE-powered GenAI, prompt-based task generation**.
- Highly scalable with the ability to ingest billions of rows and millions of files within hours.

Solution capabilities

Easy;

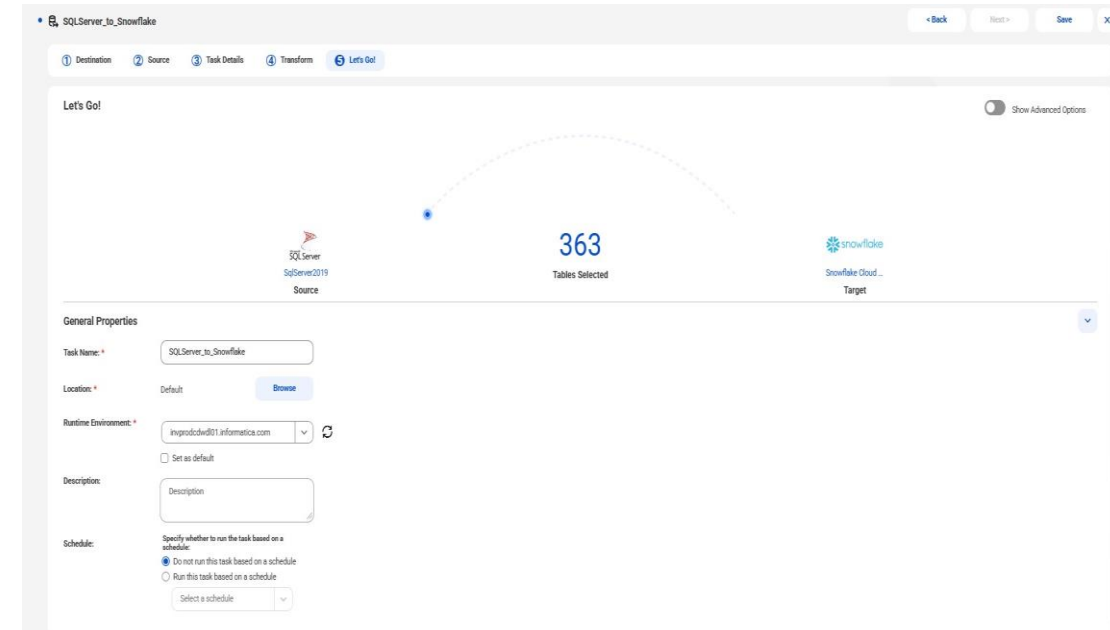
Four-step wizard for data engineers to ingest and replicate files, databases, applications, CDC and streaming data.

Efficient:

- Automatic change data capture (CDC) to ingest data into cloud data warehouses.
- Multiple replication mechanisms including log-based, query-based, trigger-based, API-based and consistency validation

Cost-effective:

- Ability to leverage CLAIRE copilot to automate and accelerate the design and development of ingestion/ replication logic.



Customer value

A Spanish insurance company improved developer productivity by 80% and operating costs by 50% with Informatica's Cloud Mass Ingestion.

Data was made available in the right format in a governed manner to data scientists, data analysts and line of business users for downstream consumption.

Connectors and CDC - CDIR

Data Ingestion and Replication - Support Matrix

Applications

- Salesforce
- SAP ECC, S/4
- ServiceNow
- Microsoft Dynamics 365
- NetSuite
- Oracle E-Business Suite
- Workday
- Adobe Analytics
- Google Analytics
- Zendesk
- Marketo
- Salesforce Marketing Cloud
- Oracle ERP and SCM

Databases

- Oracle w/ CDC
- SQL Server w/ CDC
- PostgreSQL w/ CDC
- Mainframe AS 400 and Db2 z/os w/ CDC
- Db2 LUW
- Teradata
- SAP Hana
- Azure SQL DB
- Netezza
- MySQL w/ CDC
- Mongo DB

Streaming

- Kinesis Streams
- Kafka
- AMQP
- Azure Event Hubs
- MQTT
- OPC UA
- REST
- JMS

Files

- Hadoop
- SFTP, FTP, FTPS
- Local Directory
- S3, GCS, ADLS

Targets

- MFST ADLS, Blob, Synapse
- AWS S3, Redshift
- Snowflake
- Google GCS, BigQuery
- Oracle, SQL Server, Postgres
- Databricks Delta
- Flat files
- Kafka/ Azure EventHubs

Change Data Capture (CDC): Overview

What is change data capture (CDC)?

CDC is the process of capturing changes made to a data source as and when they happen so that just the change can be passed on to consumers.

Methods for change data capture (CDC)

Timestamp-based CDC

Capture changes to the database using a timestamp column.

- **Pros**
 - Easy and simple way to capture changes
- **Cons**
 - Availability of timestamp column
 - Adverse impact on the database load
 - Doesn't include all the changes
 - Performance challenges

Trigger CDC

A database trigger is a special stored procedure that is run when specific actions occur within a database.

- **Pros**
 - The trigger is activated as soon as the table is modified
- **Cons**
 - Pauses the database update until the functionality of the trigger has been completed which causes latency

Log-based CDC

Extract changes from the database log files.






- **Pros**
 - Most efficient method of CDC
 - Non-intrusive method to capture changes
 - No load on the DB server
 - Scales very well for large volumes
- **Cons**
 - Requires access to the database logs

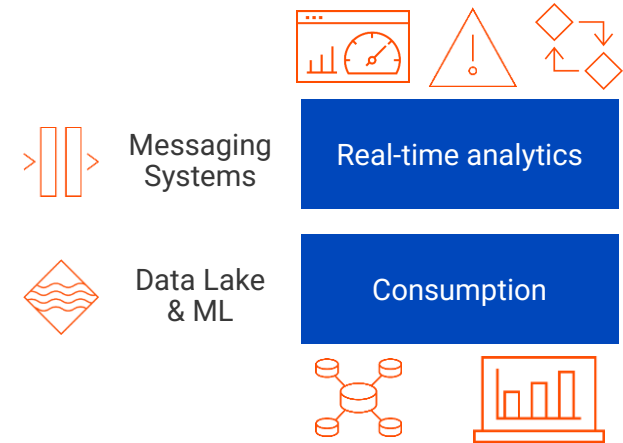
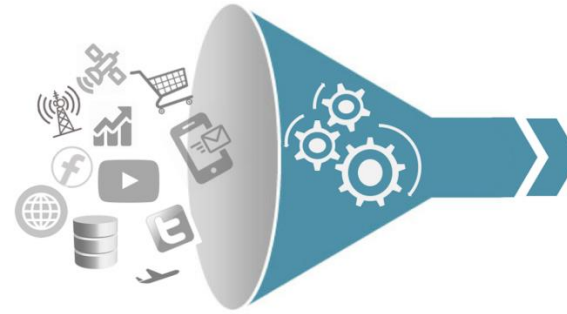
Informatica Preferred Approach and Used in Cloud Data Ingestion and Replication

Streaming Source and Targets - CDIR

Streaming Ingestion & Replication

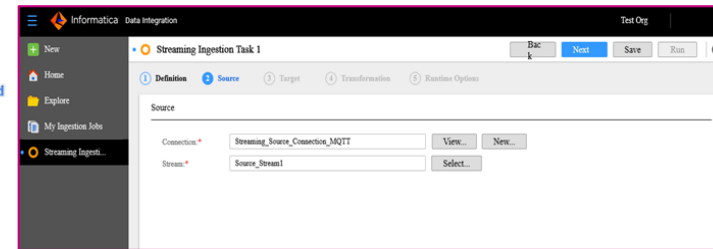
Provides streaming ingestion capabilities as part of IICS data ingestion service

-  Sensor Data
-  Machine Data/IOT
-  WebLogs
-  Social Media
-  Messaging Systems



Ingest streaming data: logs, clickstream, social media, Kafka Kinesis, S3, ADLS, Firehose, etc.

Real-time monitoring of ingestion jobs with lifecycle management and alerting in case of issues



Orchestrate application data ingestion in hybrid/cloud as a managed and secure service



Real time replication to Snowflake - CDIR

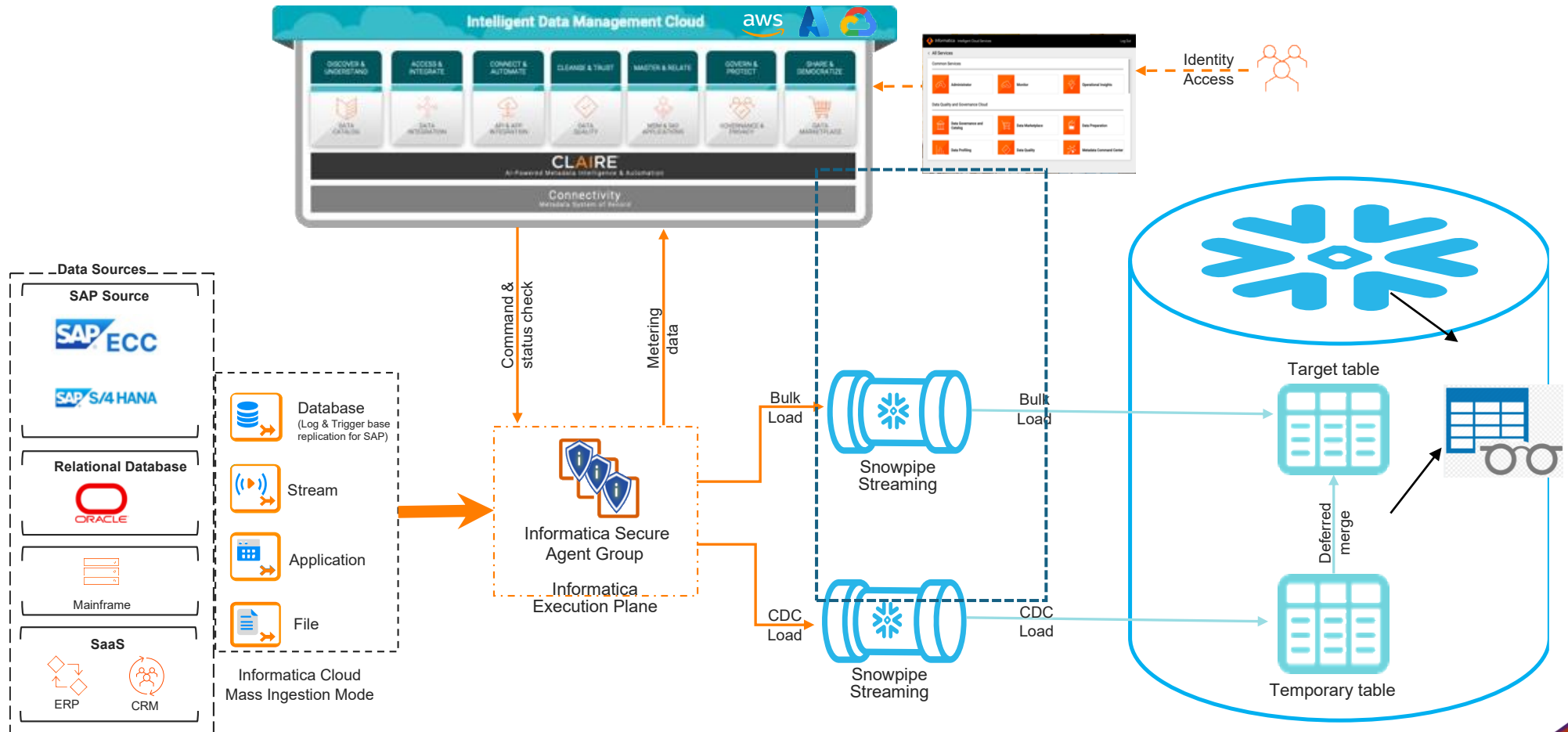
Replication using Superpipe

- Superpipe is a feature leveraging Snowflake's **Snowpipe streaming** and **Deferred Merge** for high-performance real-time ingestion into Snowflake Data Cloud.
- Real-time data view is always available regardless of CDC's deferred merge interval set.

- ✓ **Up to 3.5 times performance gains**
- ✓ **Up to 40% lesser Snowflake credits consumed**

Use case (Tech)	Optimization	Benefit
Real-time Ingestion	Use Snowpipe Streaming to bulk load and CDC on to target table and change table	Snowpipe streaming streams the changes. View created on the final target table and change table helps user have a near real time view always.
Deferred merge	Merge/Apply CDC changes on periodic basis instead of doing on transactional boundaries	Reduces TCO by doing merge in batches while providing real time replicated view of change data

Superpipe Architecture



Solution Approach

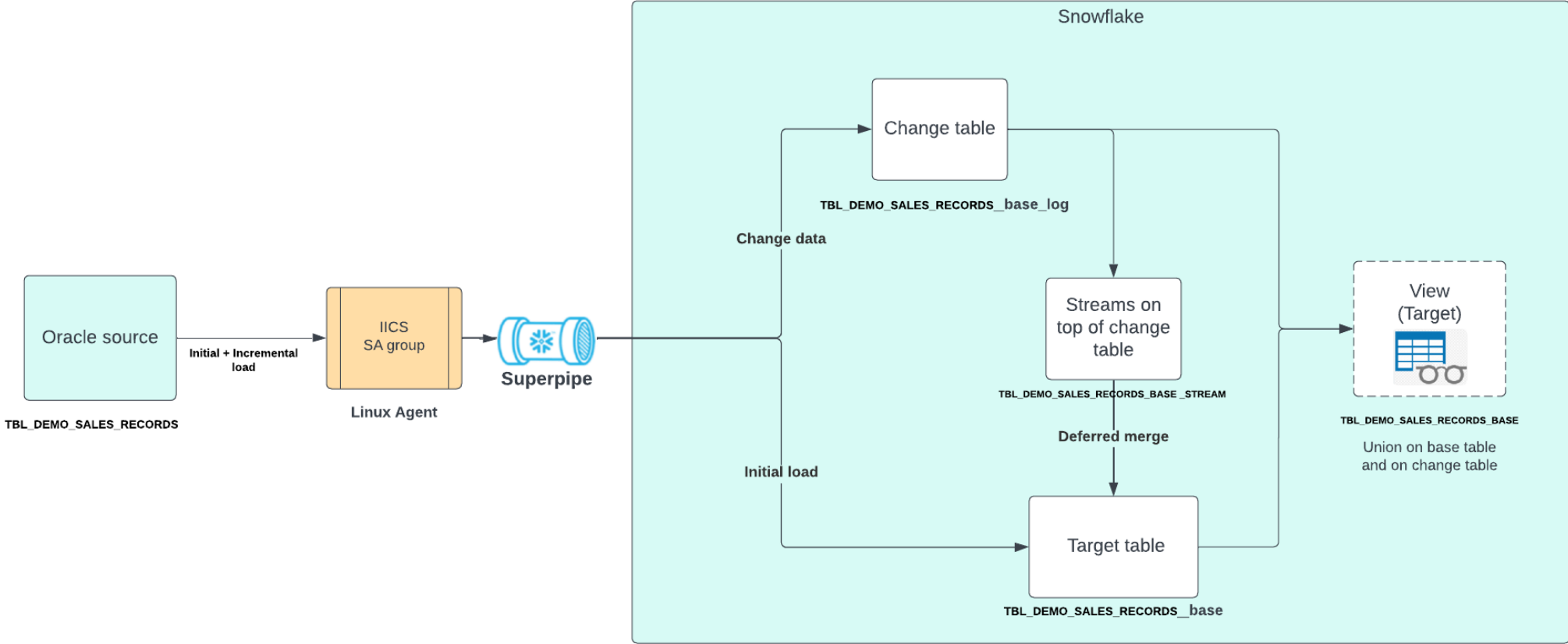
▼ **Advanced**

Add Last Replicated Time:

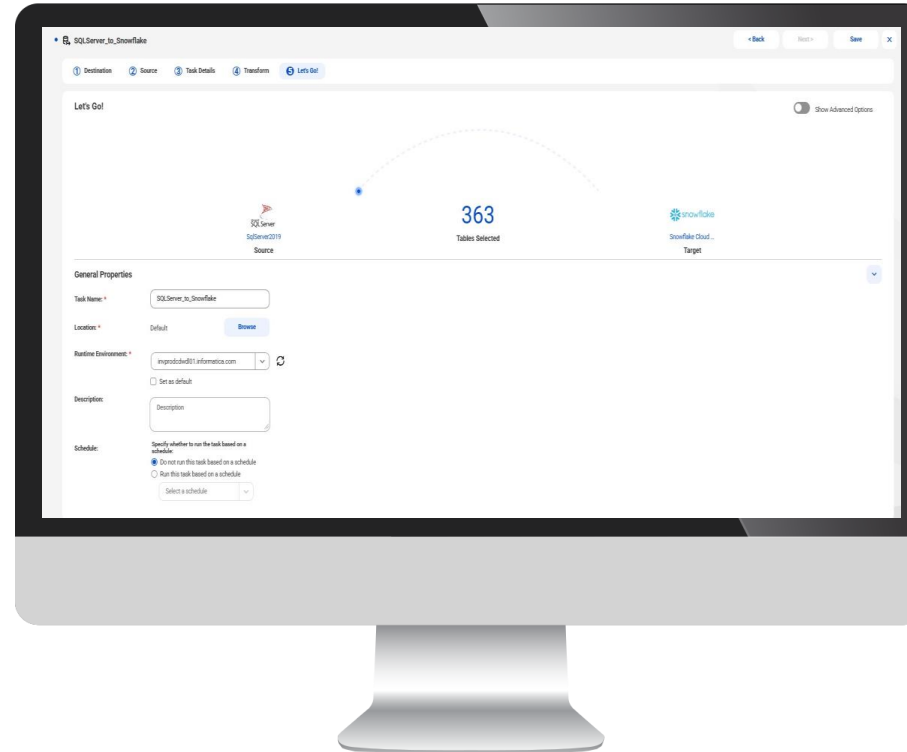
Prefix for Metadata Columns:

Superpipe:

Merge Frequency:



DEMO



Use Case: Harnessing Multi-Source Streaming Data for Smarter planning

Media Vision Broadcasting newly started streaming live cricket match in addition to movies and TV shows

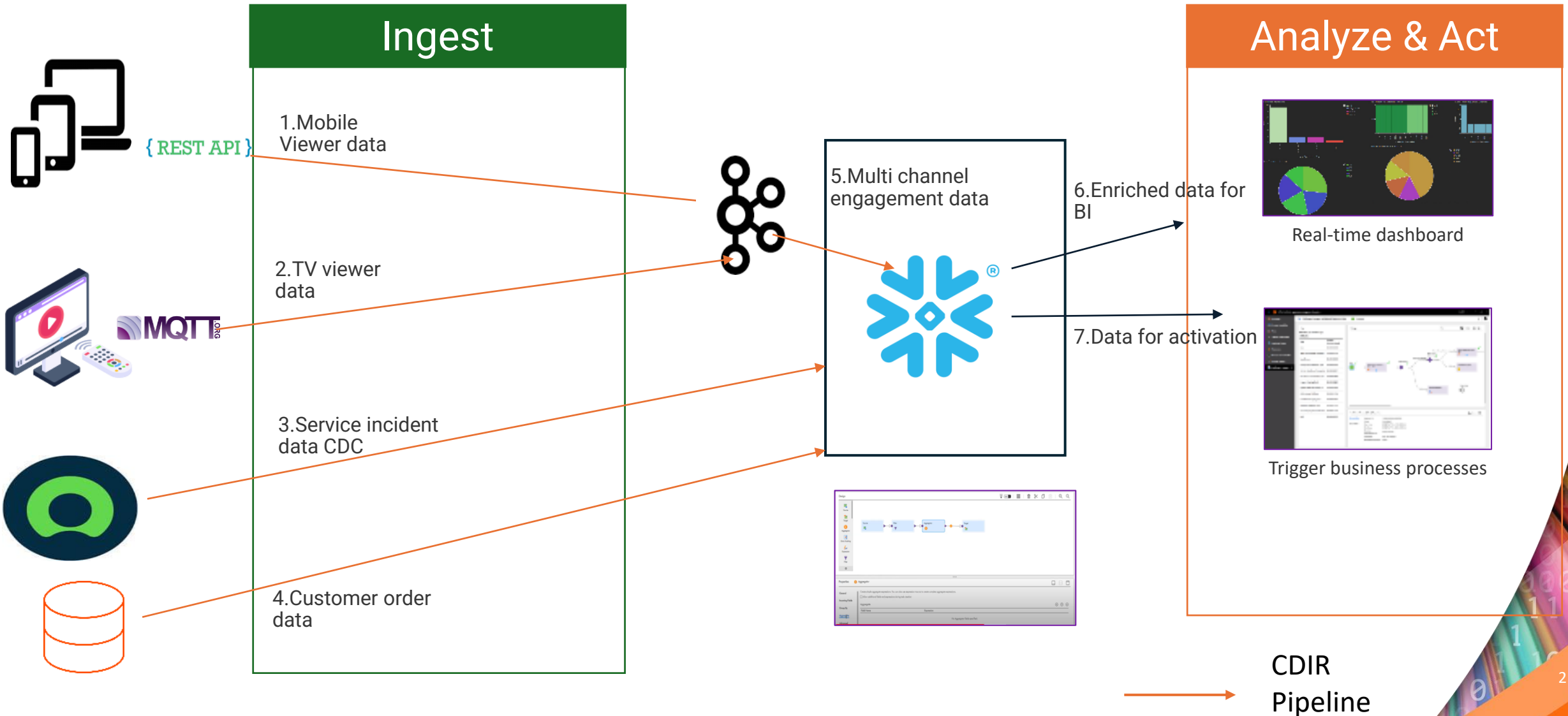
Business challenge

- Needs real-time viewer sentiment insights to validate the initiative
- Timely feedback is critical to avoid costly investments and optimize future marketing strategies.

What do they want to do?

- Collect real-time viewer data from mobile apps, smart TVs, and websites in one place.
- Combine this with incident reports and customer subscription info to see the full picture.
- Build a data pipeline to quickly turn this information into insights for better decisions and campaigns

Data Replication Flow



Register for trial experience





Q&A





Thank You

