Informatica

06 May, 2025

SQL ELT with Cloud Data Ingestion, Replication, and Integration

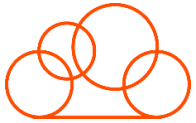
- Karthikeyan Mani, Senior Principal Product Manager, R&D
- Bharath Raghavendran, Principal Product Manager, R&D

Where data & AI come to LIFE

Housekeeping Tips

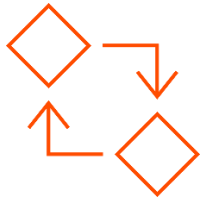
- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via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

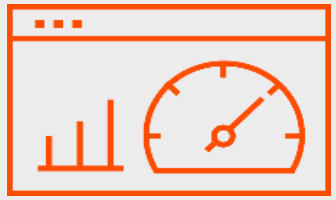


Product Learning Paths and Weekly Expert Sessions

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.

Informatica

SQL ELT with Cloud Data Ingestion, Replication and Integration

6 May, 2025

Karthikeyan Mani

Sr. Principal Product Manager

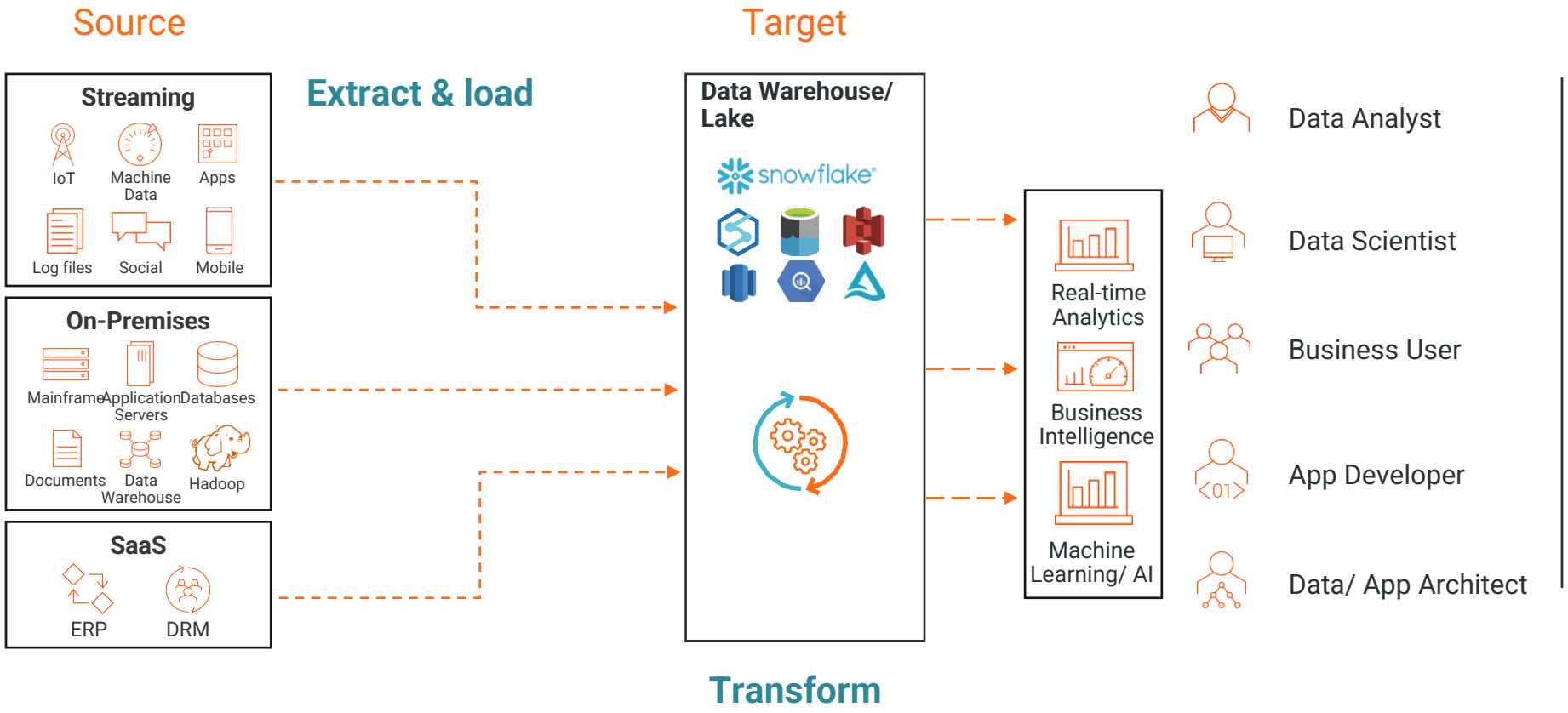
Bharath Raghavendran

Principal Product Manager

Agenda

1. Introduction to ELT
2. Replication Challenges and CDIR Overview
3. Overview and Use Case of SQL ELT
4. Reference Architecture
5. Demo
6. Q&A

Introduction to Extract, Load and Transform (ELT)



Get better results efficiently

- ☑ Combine data from different sources and formats
- ☑ Transform your data faster
- ☑ Manage data at petabyte scale
- ☑ Save time and money

Replication Patterns & Use Cases

1

Real-time analytics

- Enhanced decision making
- Immediate insights
- Operational Efficiency

2

Data Science and AI

- For Model training
- Trend analysis
- Next-gen operational app

3

Batch Analytics

- Analytics/ reporting
- Self-service data science (data as product)

- Customers replace hand coding with low-code ingestion/integration.
- Informatica data integration tools support ingestion, preparation (integration, filtering, enrichment), lineage and delivery of real-time ***and*** reference data at the needed latency, security, reliability and scale.

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.

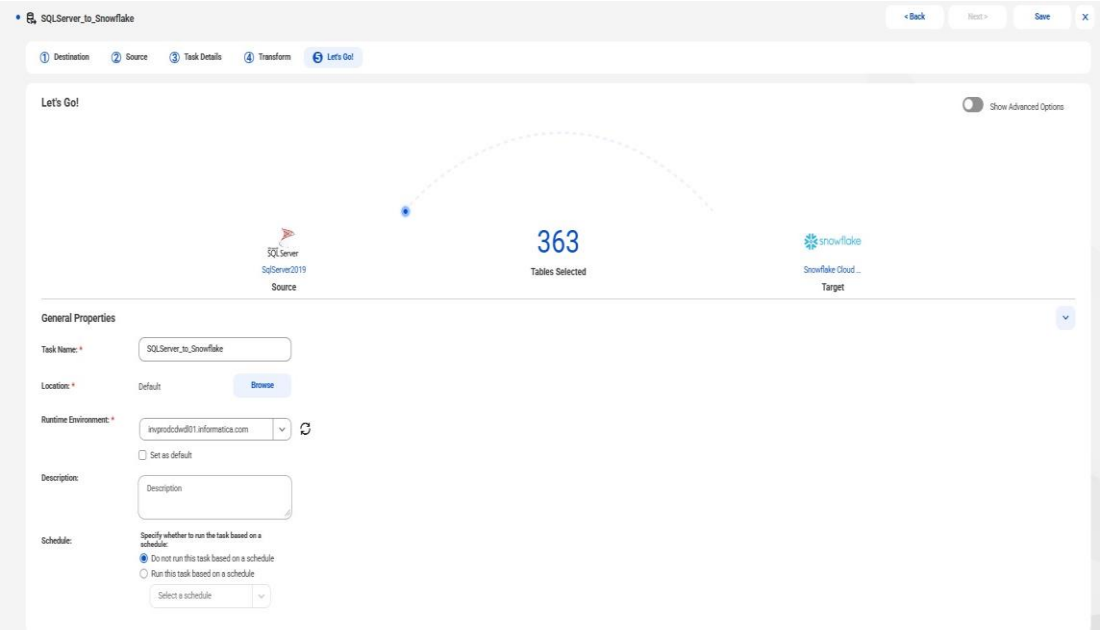
- Pros
  - Most efficient method of CDC
  - Non-intrusive method to capture changes
  - No load on the DB server
  - Scales very well for large volumes
- Cons
  - Requires access to the database logs

Informatica Preferred Approach and Used in Cloud Data Ingestion and Replication

Data Ingestion and Replication

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.

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 w/ CDC

Streaming

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

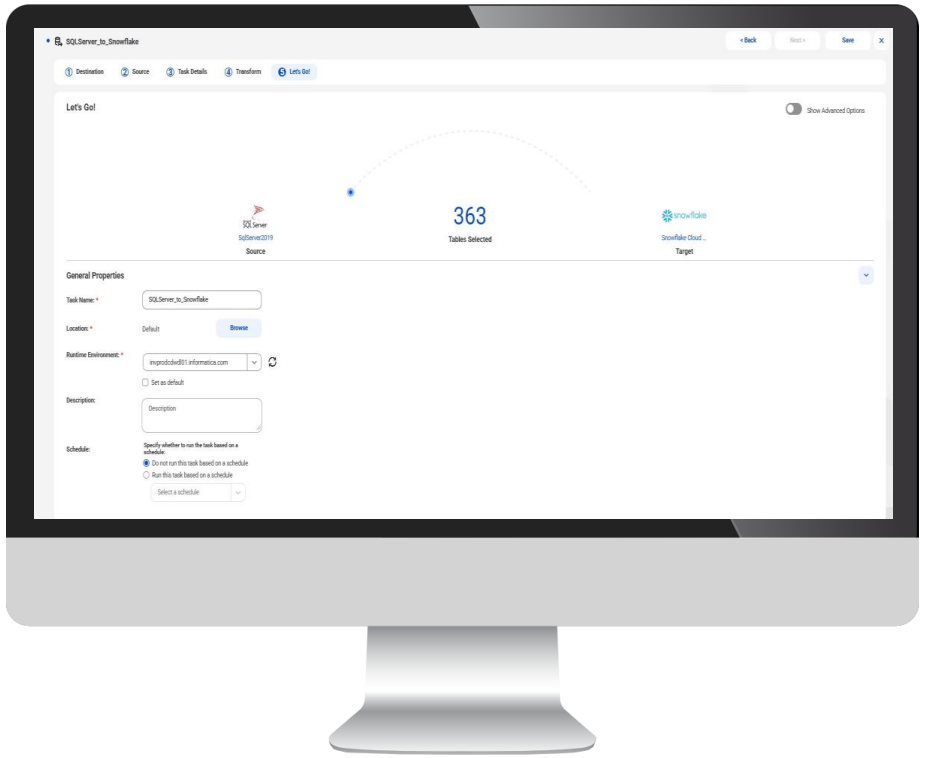
Files

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

Targets

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

DEMO

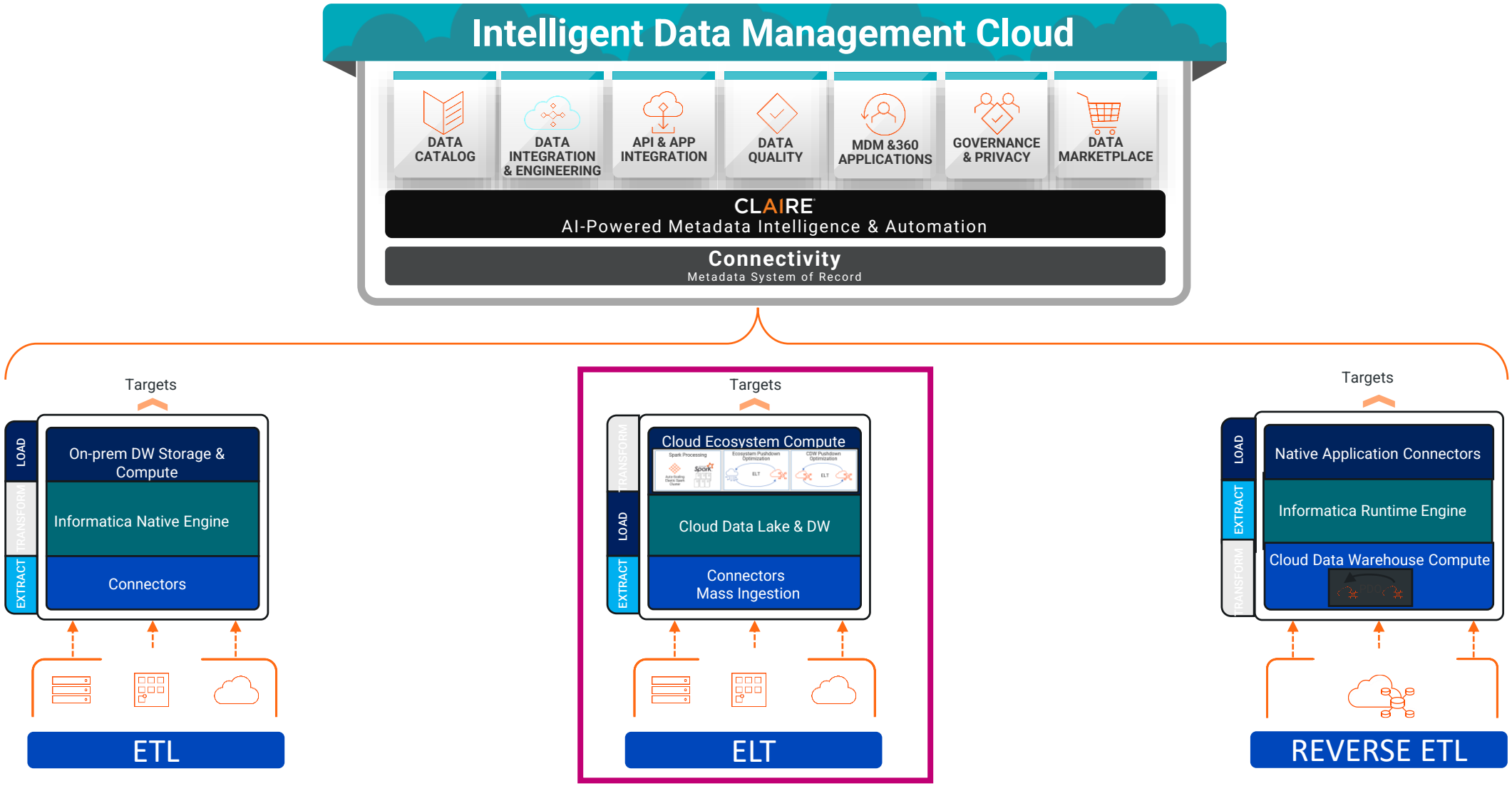

SQLServer_to_Snowflake
Back
Next
Save
Destination
Source
Task Details
Transform
Let's Go!
Let's Go!
Show Advanced Options
SQL Server
SqlServer2019
Source
363
Tables Selected
snowflake
Snowflake Cloud...
Target
General Properties
Task Name: *
SQLServer_to_Snowflake
Location: *
Default
Browse
Runtime Environment: *
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

Demo: Use Cases

Cloud Data Ingestion and Replication with Intelligent Data Management Cloud

In this demo, we will create an Ingestion and replication task for seamless data transfer using a wizard driven process. This will boost analytical capabilities, speed up insights, reduce costs, and optimize data flow. The initiative enhances productivity and sets a foundation for sustainable growth and competitive advantage in analytics.

Best in Class Support for all Data Integration & Engineering



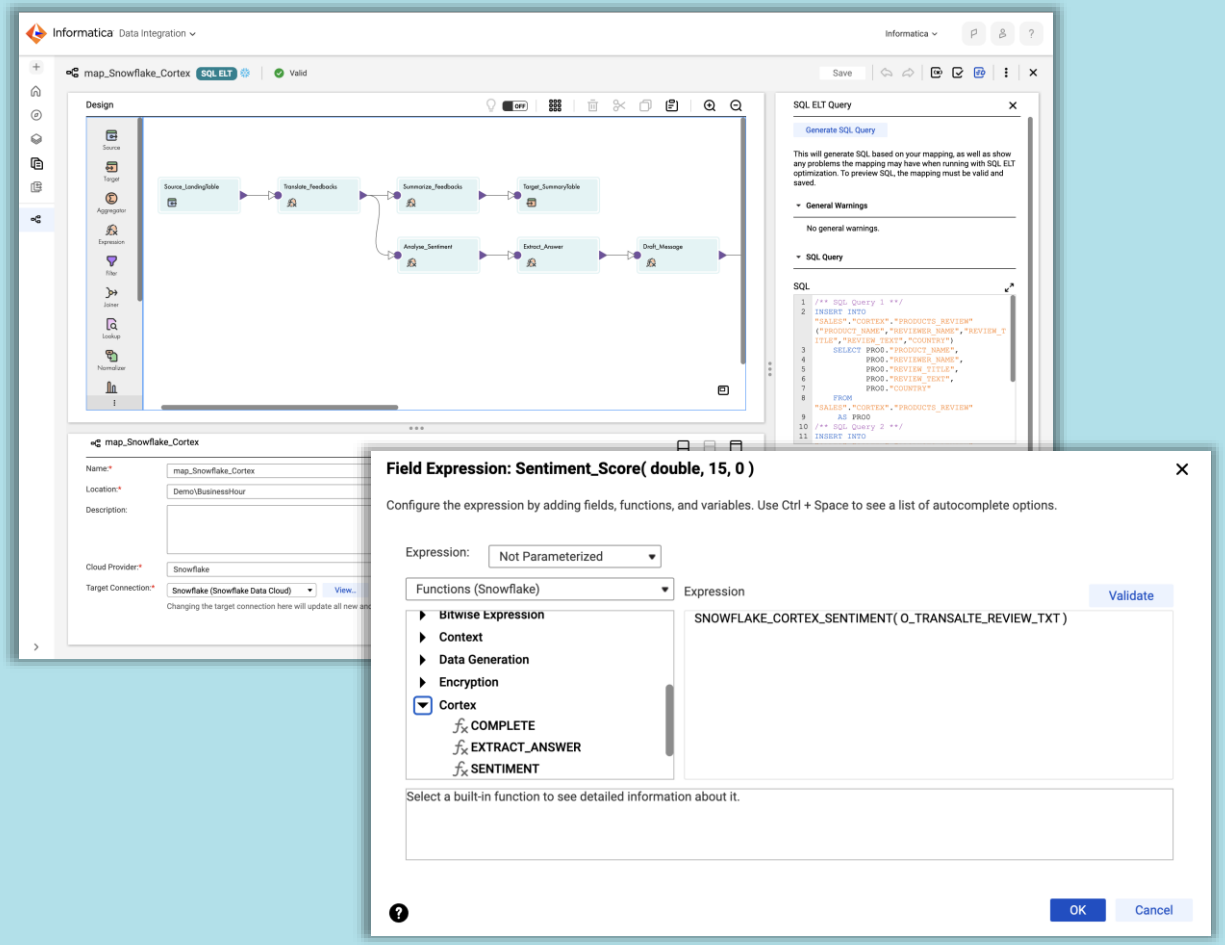
Informatica Ecosystem SQL ELT

Key highlights

- Support Snowflake, Databricks, Google Big Query, AWS Redshift and Fabric.
- Performance optimization for handling large data volumes efficiently with zero egress charge.
- Low code/ no-Code offers a user-friendly, drag-and-drop interface.
- OOTB supports for 100+ native functions.
- Support for Snowflake & Databricks Gen AI capabilities (Cortex & AI functions).

Benefits

- Guaranteed execution using the Cloud Data Warehouse (CDW) compute.
- Familiar semantics for quick onboarding of CDW users.
- SQL and data preview: Accelerated development of ELT mappings.



When to use Ecosystem SQL ELT

Data Science & analytics

Efficiently manage and analyze large datasets by reading data within the ecosystem and external data lakes (homogeneous/ heterogeneous).

Data warehouse modeling

Create a unified data repository to support business intelligence, reports and data analysis. Requires efficient data storage and quick query performance.

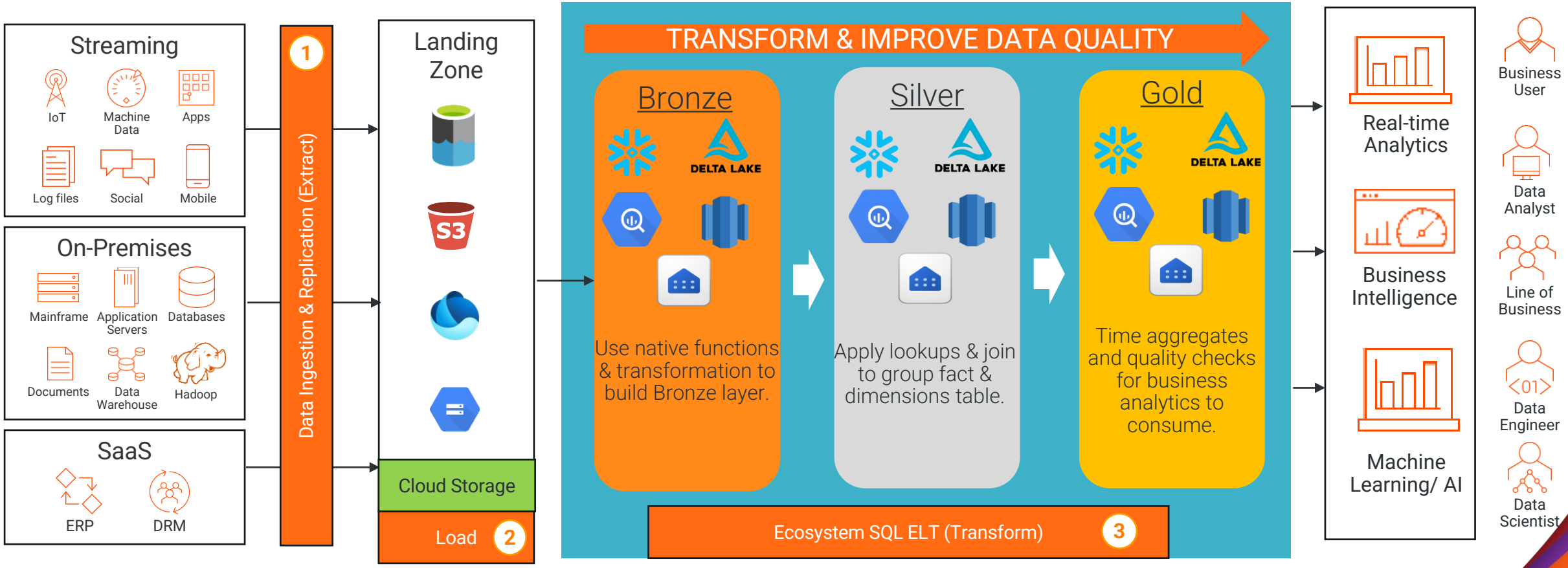
DI within same ecosystem

Extract data within the ecosystem, load it into a central warehouse and transform it for unified analysis.

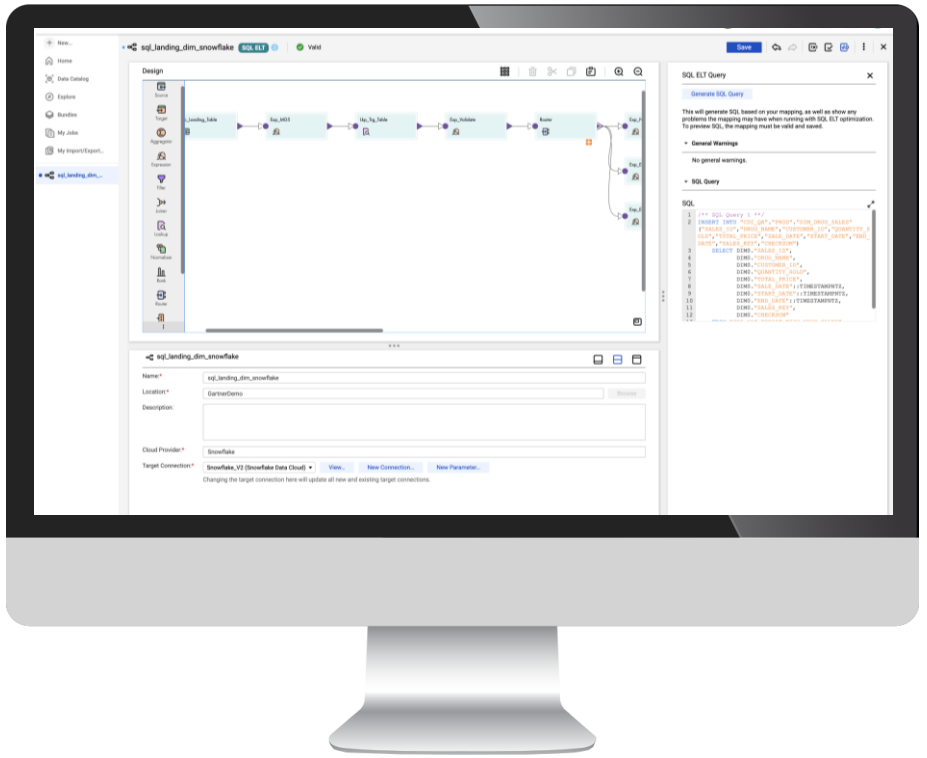
ML & model training

Large language model (LLM) functions available within the ecosystem enhance various applications by providing advanced natural language processing capabilities, leading to improved efficiency and automated tasks.

Medallion Architecture (Extract, Load, Transform)



DEMO


sql_landing_dim_snowflake
SQL ELT
Valid
Save
Design
SQL ELT Query
Generate SQL Query
This will generate SQL based on your mapping, as well as show any problems the mapping may have when running with SQL ELT optimization. To preview SQL, the mapping must be valid and saved.
General Warnings
No general warnings.
SQL Query
SQL
sql_landing_dim_snowflake
Name*
sql_landing_dim_snowflake
Location*
GartnerDemo
Description
Cloud Provider*
Snowflake
Target Connection*
Snowflake_V2 (Snowflake Data Cloud)
View...
New Connection...
New Parameter...
Changing the target connection here will update all new and existing target connections.

Demo: Use Cases

Cloud Data Integration with Intelligent Data Management Cloud

In this Demo, we will use Cortex AI to optimize data analytics in Snowflake Data Warehouses, benefiting business analysts. Reviews are translated into English to ensure consistency, while sentiment functions gauge sentiment scores. These are then categorized by sentiment and loaded into a Snowflake target table using SQL ELT mode. This enhances processing speed and efficiency, outperforming traditional methods, and equips analysts with accurate, high-quality data for informed decision-making and deeper consumer insights in retail.

Free Trial: Ecosystem SQL ELT

https://marketplace.informatica.com/forms/free-cloud-data-integration-modal.html

Snowflake

Databricks

Redshift

Fabric

GBQ

Q&A

Thank You