

Power your Cloud Data Warehouse/Data Lake with IICS

Shiv Patel

Principal Customer Success Technologist



Agenda

1

Cloud Data
Warehouse/Data
Lake Blueprint

2

IICS Services

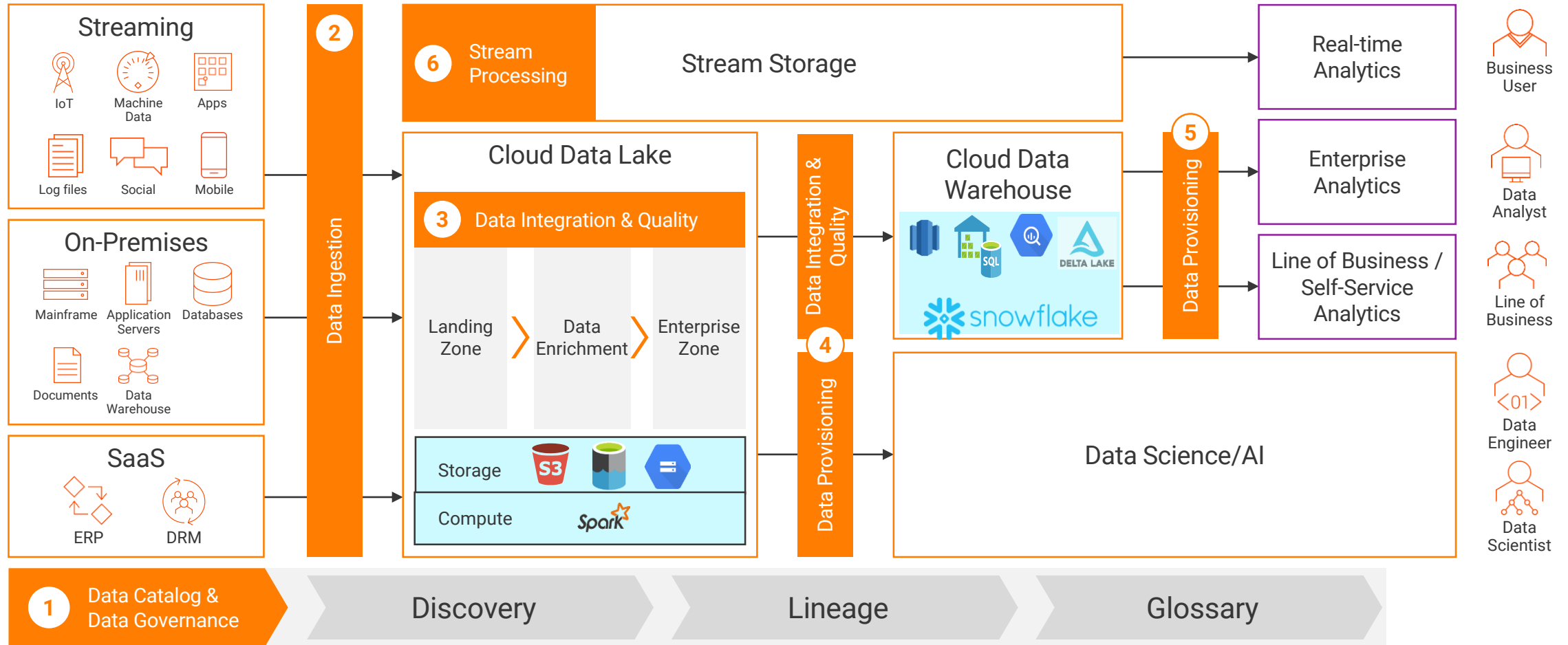
3

Cloud Data
Integration

4

Demo

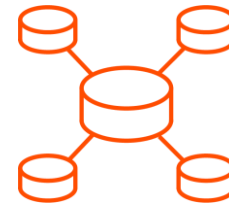
Blueprint for Cloud Data Warehouse/Data Lake



Cloud Data Warehouses and Data Lakes to Enable Analytics and AI in the Cloud



Data Lakes



Data Warehouse



Cloud Lakehouse Data Management Pillars

DATA INTEGRATION



- Codeless integration
- Mass ingestion for files, DB and streaming
- Push down optimization
- Serverless and elastic scaling
- Spark-based processing in the cloud
- Broad connectivity
- Stream processing
- MLOps

DATA QUALITY



- Data profiling
- Data quality rules
- Dictionaries to manage values lists
- Cleansing, standardization, parsing, verification and deduplication/consolidation processes
- Integrated into data integration
- Data quality analytics

METADATA MANAGEMENT



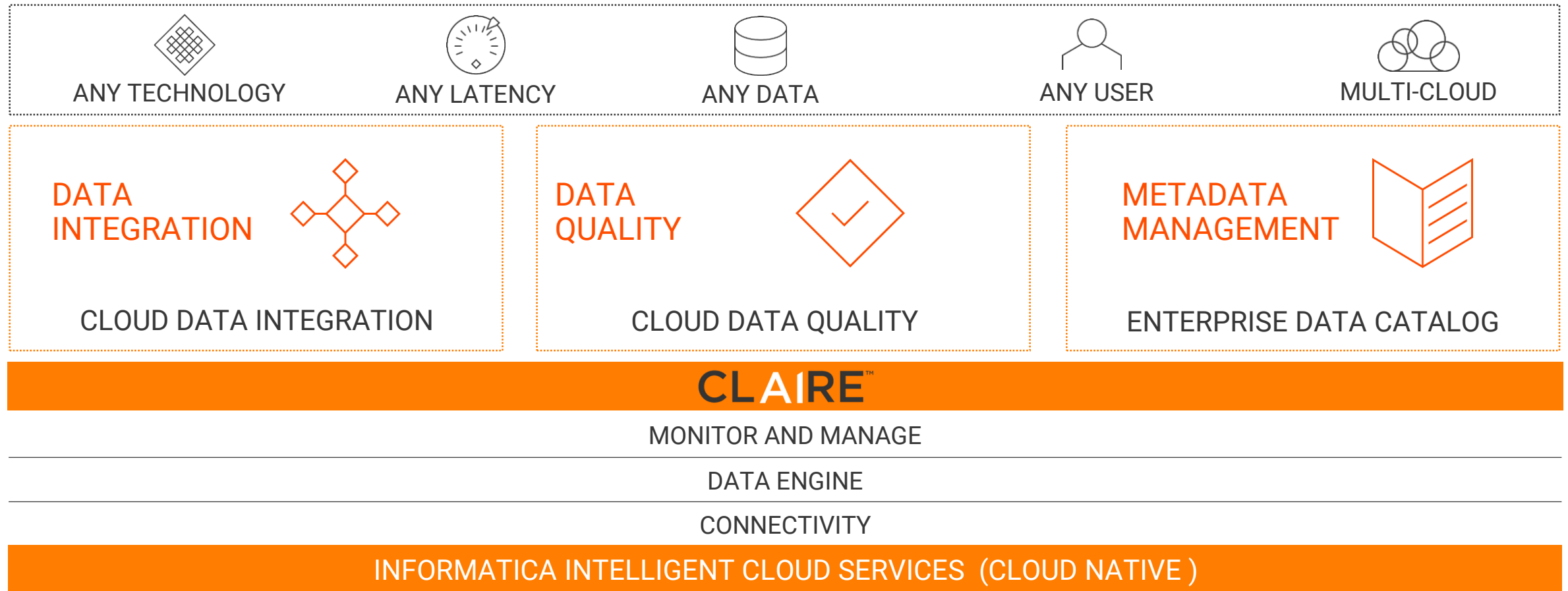
- Data discovery
- End-to-end lineage
- Metadata – technical, business, operational, usage
- Connect and scan metadata – databases (DW, DL), apps, ETL, BI tools and others
- Common metadata foundation

AI-DRIVEN AUTOMATION

CLOUD NATIVE

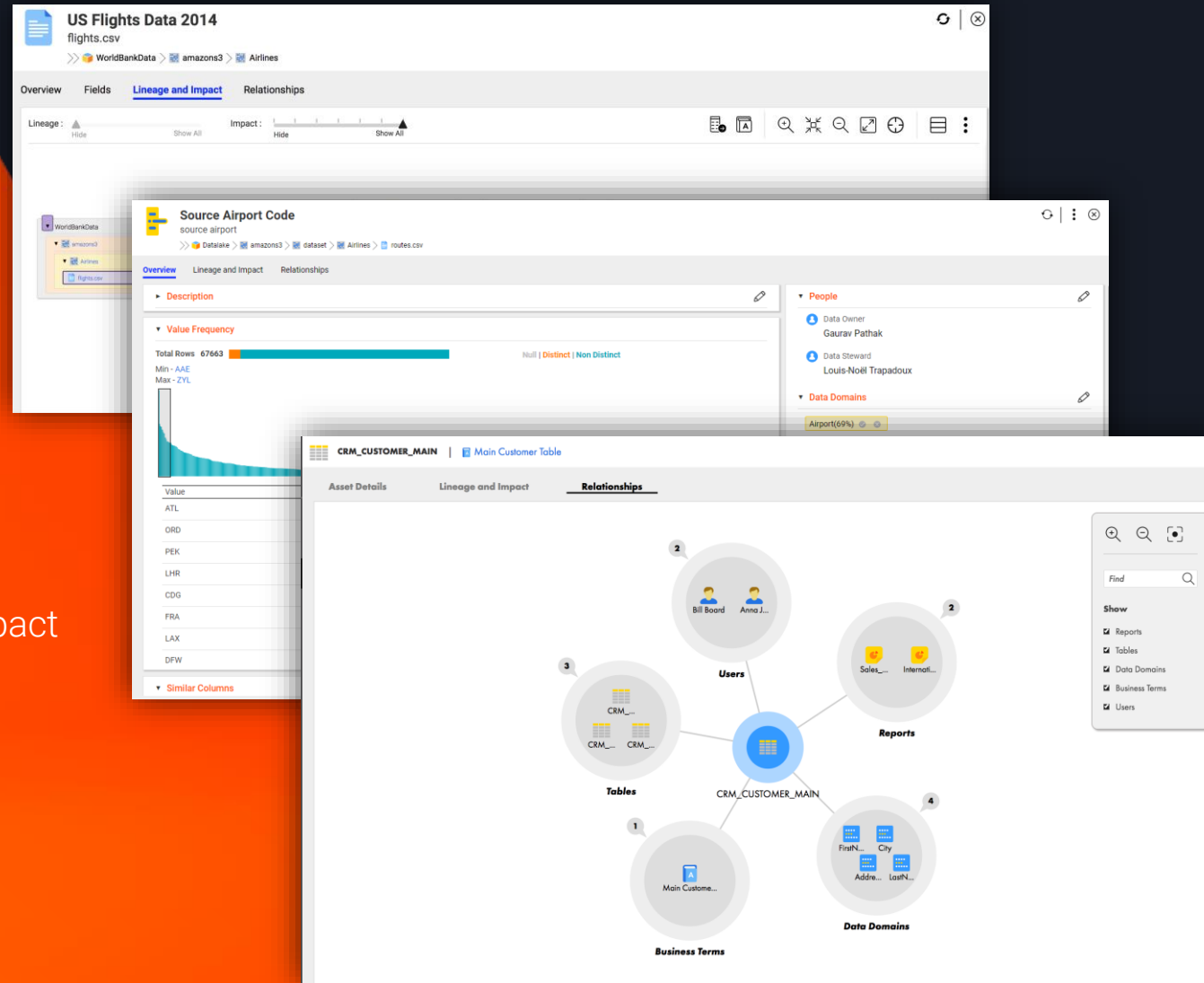
Multi cloud • API driven • Microservices • Containerization • Serverless architecture • Minimal install and setup • Auto-upgrades
Usage-based pricing • Trust certifications

Informatica Intelligent Cloud Services



Intelligent Metadata Management

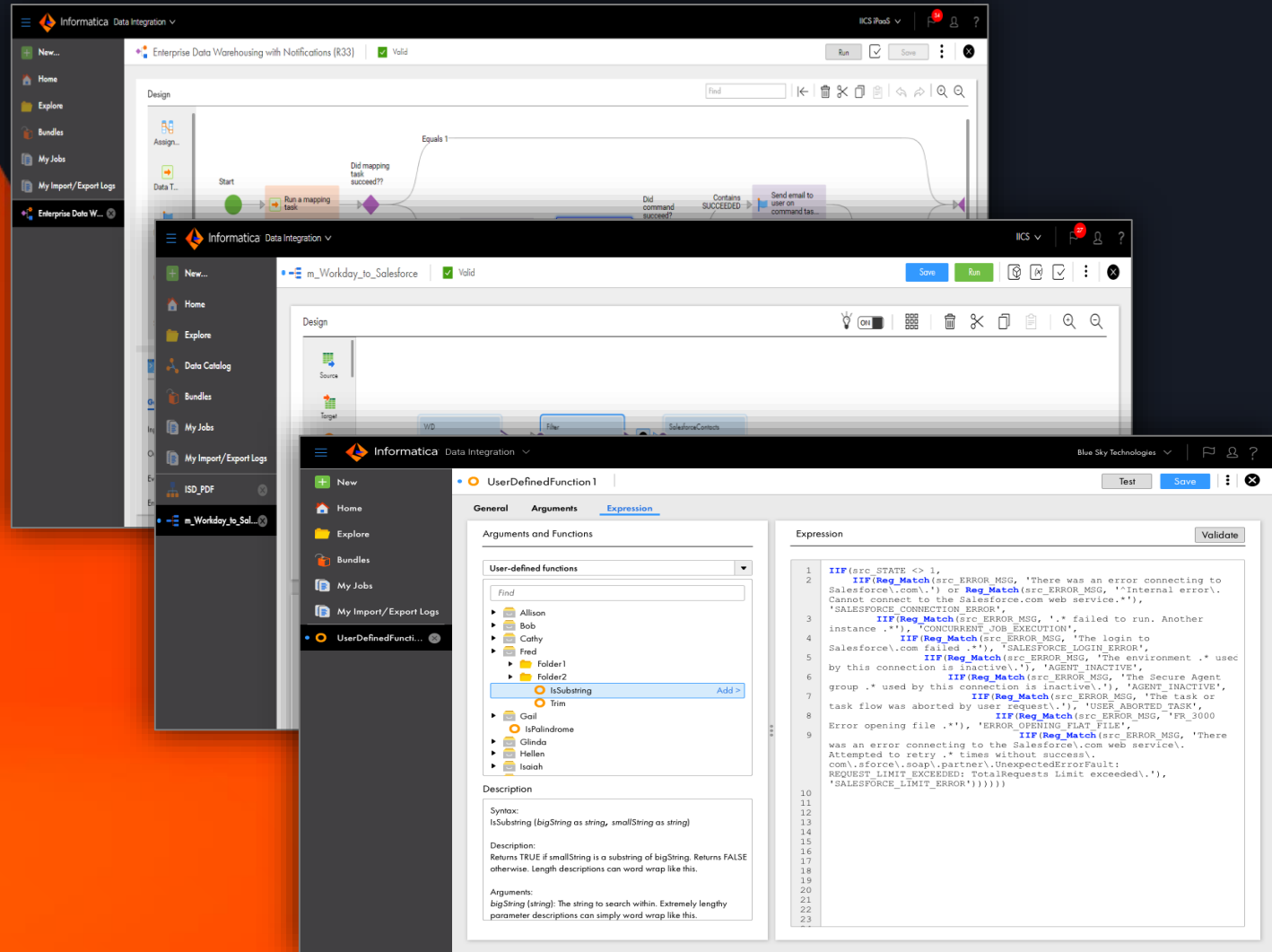
- Catalog all of your data across multi-cloud, on-premises, ETL, BI tools, and applications
- Quickly find and understand your data using AI/ML
- End to end data lineage with change impact
- Collaborate and tap into shared data knowledge across the enterprise
- Support all business use cases with integrated solution
- Scale to support tens of millions of data objects
- Intelligently link business terms to the technical data estate
- Automatically understand lineage of data across the organization



Enterprise Data Catalog

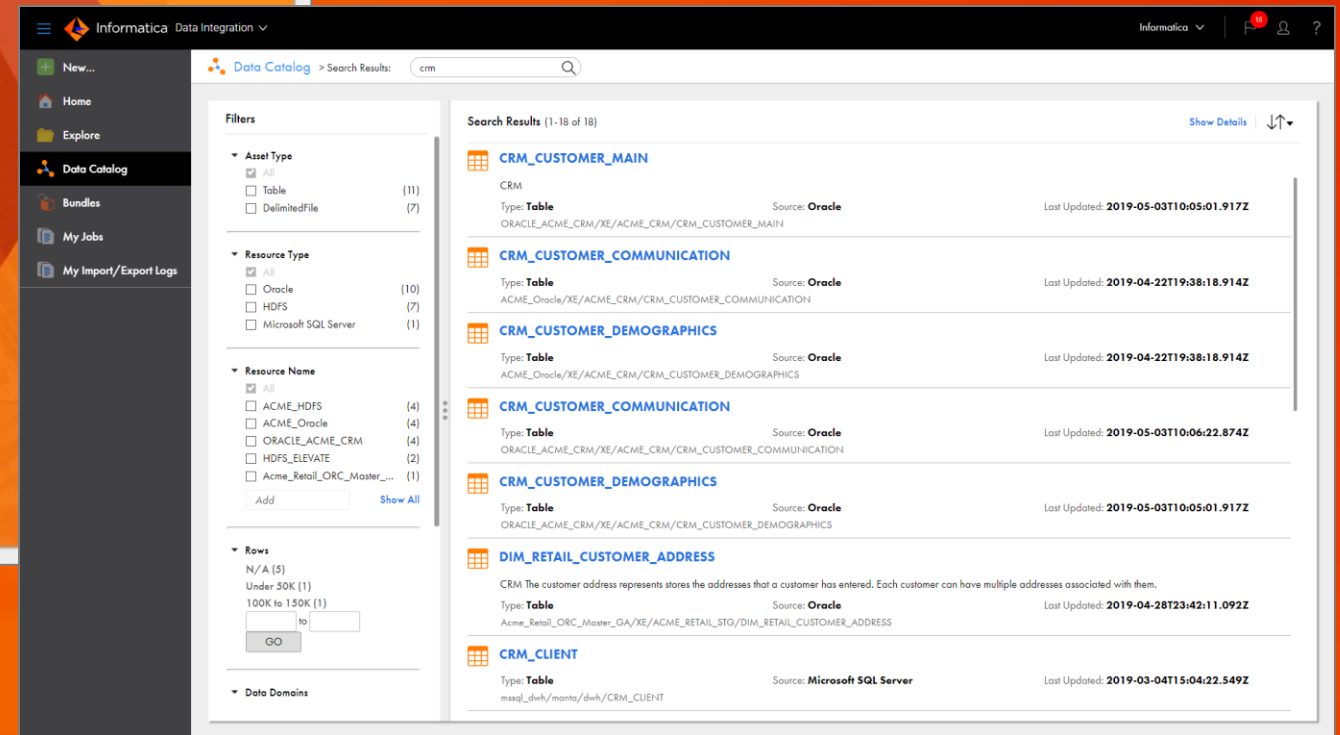
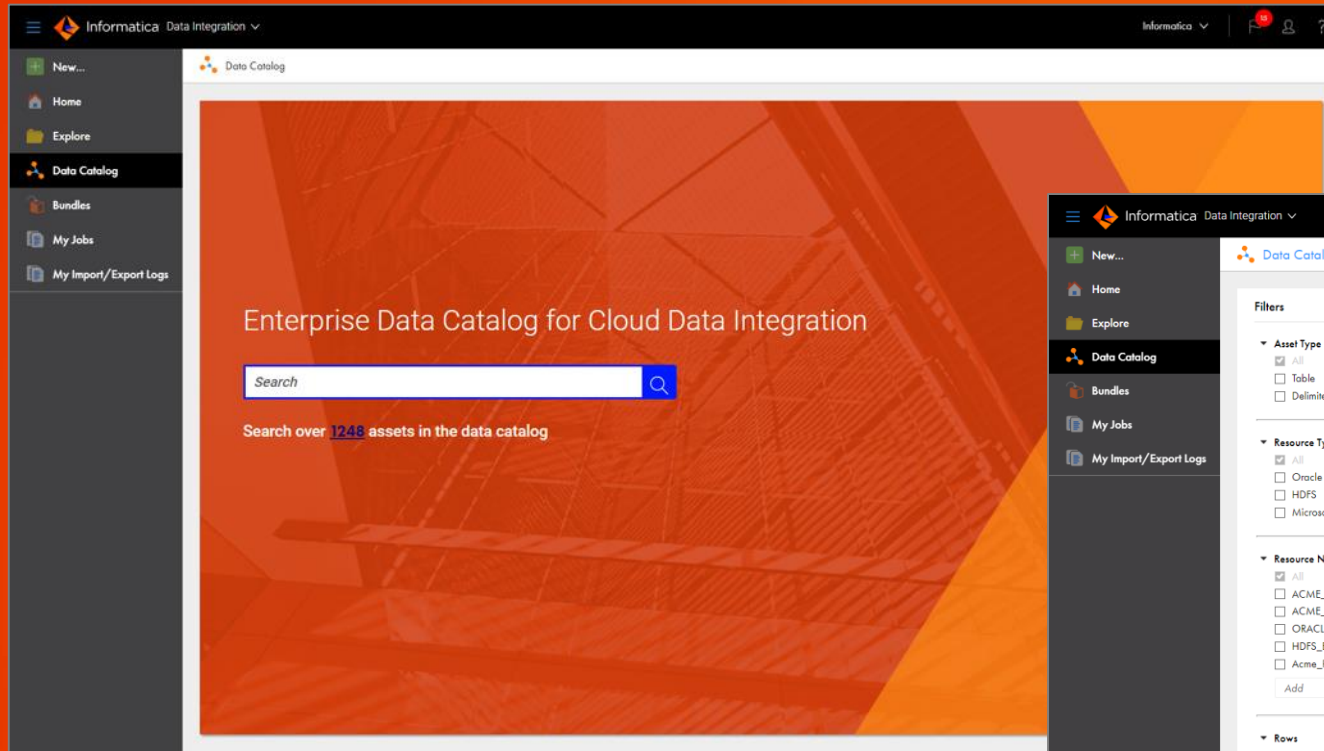
Data Integration

- Build simple to complex data integration loads using a mapping designer with out-of-the-box advanced data integration transformations.
- Connector support for any data type or any pattern (ETL or ELT)
- Support of Mass ingestion of any formats- files, databases, CDC or streaming
- Automatically discover any data type
- Intelligent transformation recommendation
- Serverless execution mode
- Auto tuning, auto scaling of DI jobs for greater performance & cost saving
- Heat map view of the jobs to identify critical times and peak hours for better resource planning.



Cloud Data Integration

Integrated Data Discovery



Cloud Data Integration

Data Quality

- Ensure trusted data for CDW/DL
- Empower self-service and business ownership
- Identify and prioritize data issues
- Intelligent discovery and classification of various domains including sensitive data, across structured and unstructured sources
- Intelligent rule recommendations
- Build once and re-use everywhere across cloud and on-premises
- Natural Language Processing (NLP) to auto generate data quality rules
- Automate data quality assessment and reporting across all sources
- Embed DQ processes with Cloud Data Integration

The image displays three overlapping screenshots of Informatica software interfaces. The top screenshot shows the 'Subject Registry' with 'Subject Details' for a subject named 'Antony Martial'. The middle screenshot shows the 'Data Profiling' interface for 'ProfileCustomerData', displaying a table of statistics for various columns. The bottom screenshot shows the 'Intelligent Rules Aut...' interface, listing rules for validation, with one rule highlighted: 'Validates the 'street' line of an address (Recommended by CLAIRE)'.

Informatica Secure@Source

Subject Registry

Subject Details

6 Data Stores, 8 Data Domains, 155 Fields / Files

Group by: None

FullName: Antony Martial, LastName: -, CompanyName: Manchester United, Address: Kormangala, City: Manchester, PhoneNumber: 8040201002, Salary: -

Data Stores

- FILES_NFS
- LOCAL_FILES
- One_Drive_Files
- SR_CUSTOMER_DATA
- SR_GOLDEN_DATA
- Unstructured_Files

Informatica Data Profiling

ProfileCustomerData, Profile run 3 of 3, 11 of 14 Columns, 5 of 3 Rules, 637 Rows (All rows), Oct 18, 2019, 7:52 AM

Columns	Value Distribution	% Null	# Null	% Distinct	# Distinct	% Non-distinct	# Non-distinct	# Patterns
ADDRESS		0.78%	5	90.42%	579	8.8%	56	2
ADDRESS2		55.57%	354	23.55%	150	20.88%	133	3
CITY		2.31%	16	53.53%	341	43.94%	280	7
COUNTRY		0%	0	2.67%	17	97.33%	620	6
CUSTOMER_ID		0%	0	80.35%	328	17.43%	111	4
CUSTOMER_NAME		0%	0	80.38%	312	19.62%	125	1
LAST_ORDER_DATE		43.33%	276	1.68%	107	39.87%	254	5
PHONE		6.12%	39	84.14%	536	9.74%	62	6
POSTALCODE		3.3%	21	81.48%	519	15.22%	97	6
STATE		41.29%	263	15.07%	96	43.64%	278	6
STATUS		1.88%	12	0.78%	5	97.34%	620	5
CheckCompleteness	Input Columns: STATUS	0%	0	0.31%	2	99.69%	635	2
CheckCompleteness	Input Columns: COUNTRY	0%	0	0.31%	2	99.69%	635	2

Intelligent Rules Aut...

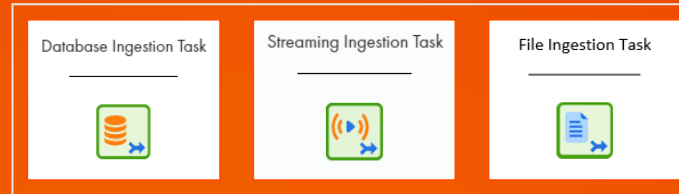
Definition, Rules, Schedule

Rules (4)

Name	Description	Type
<input type="checkbox"/> Validate_Street_Line	Validates the 'street' line of an address (Reco...	Rule Specification
<input type="checkbox"/> Validate_Street_Line	Validates the 'street' line of an address (Reco...	Rule Specification
<input type="checkbox"/> Validate_Street_Line	Validates the 'street' line of an address (Recommended by CLAIRE)	Rule Specification
<input type="checkbox"/> Validate_URL	Validates if the given input is a URL or not (Re...	Rule Specification

Cloud Data Quality

Ingest Data from Database, Streaming, Files



DATABASE INGESTION

The screenshot displays the Informatica Data Integration console for a task named 'demo_dbms_msc2m_load_1'. The task is of type 'Database Ingestion Task' and is currently in a 'Completed' state. The configuration details show the source as 'Oracle Staging DB - Patient Data' and the target as 'PATIENT_STAGING'. The task has processed 10,000 records. The 'Table Selection Rules' section shows the 'Include' operator with a condition of 'Supplemental logging for all columns'. The 'Advanced' section shows the 'Restart Point for Incremental Load' set to 'Default'.

STREAMING INGESTION

The screenshot shows the 'IngestionDemo' task configuration in the Informatica Data Integration console. The task is of type 'Streaming Ingestion Task' and is currently in a 'Completed' state. The configuration details show the source as 'Kafka' and the target as 'DemoTopic1'. The task has processed 10,000 records. The 'Advanced Properties' section shows the 'Producer Configuration Properties' set to 'key1=value1,key2=value2', the 'Metadata Fetch Timeout in milliseconds' set to '5000', and the 'Batch Flush Size in bytes' set to '1048576'.

FILE INGESTION

The screenshot displays the Informatica Data Integration console for a task named 'ADSSG012_PatientData_json_2019'. The task is of type 'File Ingestion Task' and is currently in a 'Completed' state. The configuration details show the source as 'FTP in_SensorData' and the target as 'ADSSG012_PatientData'. The task has processed 10,000 records. The 'Source Connection Details' section shows the 'Connection Type' set to 'Advanced FTP V2' and the 'Source Directory' set to 'C:\Users\Administrator\Documents\ADSSG012_PatientData'. The 'Table Selection Rules' section shows the 'Include' operator with a condition of 'Supplemental logging for all columns'.

Cloud Mass Ingestion

Change Data Capture (CDC) from Databases & DW's

The screenshot shows the Informatica Cloud Mass Ingestion interface for a task named "Oracle-to-Snowflake-Schema-Drit". The interface has a sidebar on the left with icons for various tasks. The main panel has tabs for "Definition", "Source", "Target", and "Runtime Options". The "Definition" tab is active, showing fields for Name, Location, Runtime Environment, Description, and Load Type. The Name field is "Oracle-to-Snowflake-Schema-Drit", Location is "Default", Runtime Environment is "EMDBMIDOCKRRH01", and Description is "This task ingest initial and incremental CDC data from Oracle to Snowflake and also handles schema drift". The Load Type dropdown is open, showing options: "Initial Load", "Incremental Load", and "Initial and Incremental Loads" (which is selected).

Informatica Mass Ingestion

Oracle-to-Snowflake-Schema-Drit

1 Definition 2 Source 3 Target 4 Runtime Options

▼ Definition

Name: Oracle-to-Snowflake-Schema-Drit

Location: Default Browse

Runtime Environment: EMDBMIDOCKRRH01

Description: This task ingest initial and incremental CDC data from Oracle to Snowflake and also handles schema drift

Load Type: Select a load type

- Initial Load
- Incremental Load
- Initial and Incremental Loads

The screenshot shows the Informatica Cloud Mass Ingestion interface for a task named "Ora2SnCombinedTestMay11". The interface has tabs for "Definition", "Source", "Target", and "Runtime Options". The "Runtime Options" tab is active, showing "Schema Drift Options" and "Advanced" sections. The "Schema Drift Options" section has dropdowns for "Add Column:", "Modify Column:", "Drop Column:", and "Rename Column:". The "Add Column:" dropdown is open, showing options: "Ignore", "Replicate", "Stop Job", and "Stop Table" (which is selected). The "Advanced" section has a field for "Number of Rows in Output File:".

Ora2SnCombinedTestMay11

1 Definition 2 Source 3 Target 4 Runtime Options

▼ Schema Drift Options

Add Column: Replicate

Modify Column: Replicate

Drop Column: Ignore

Rename Column: Ignore

▼ Advanced

Number of Rows in Output File:

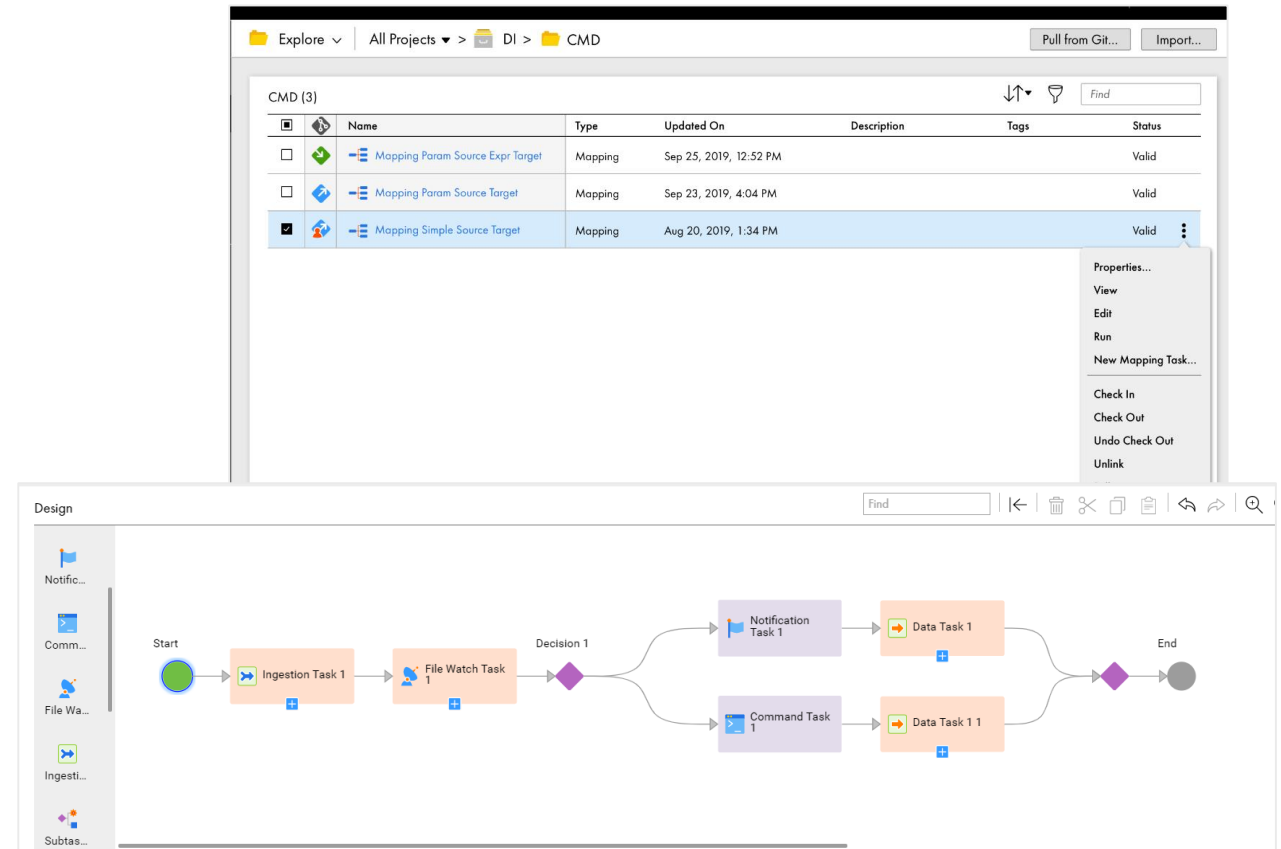
Cloud Mass Ingestion

Cloud Data Integration



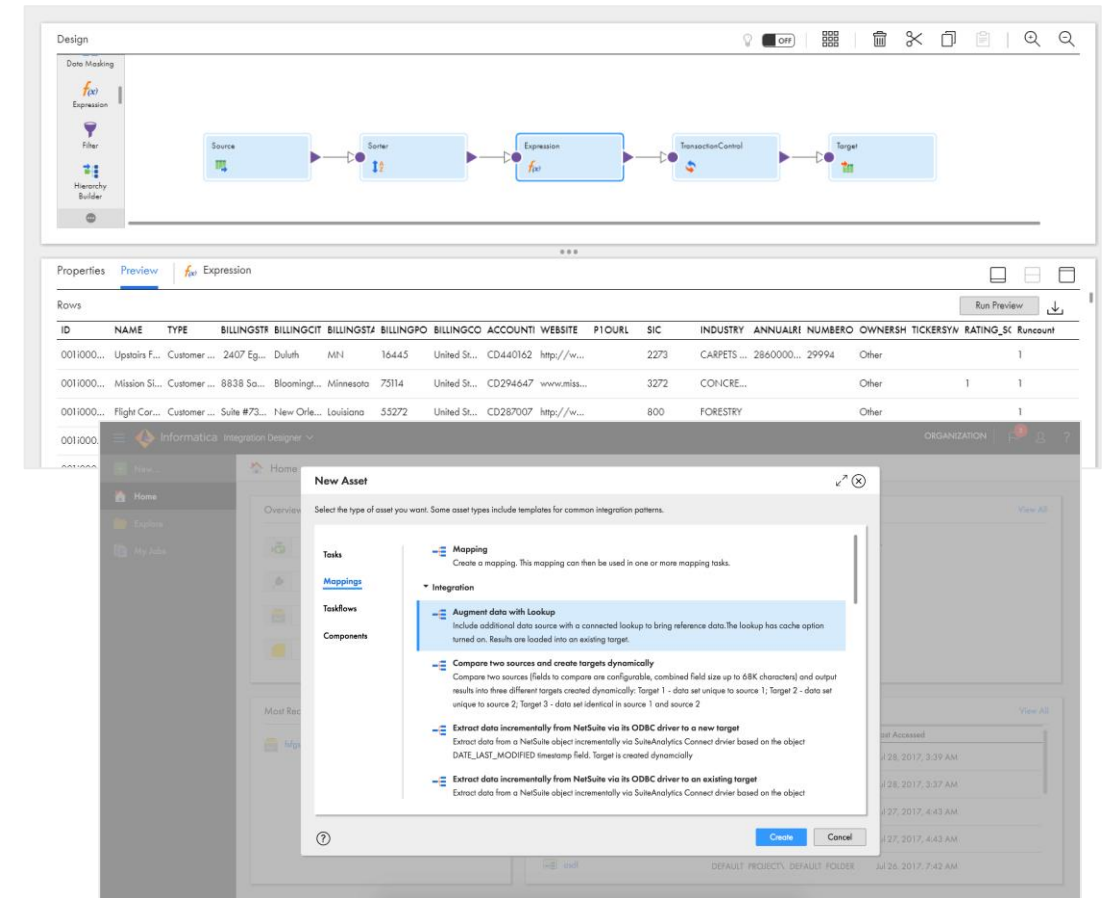
Enterprise Class

- One-click integration with GitHub for all source control needs
- Extensive transformation capabilities without any need for hand-coding
- The ability to automate and orchestrate all your integration tasks easily using the taskflow designer
- Options to choose the best runtime for your jobs with support for on-premises and cloud runtimes



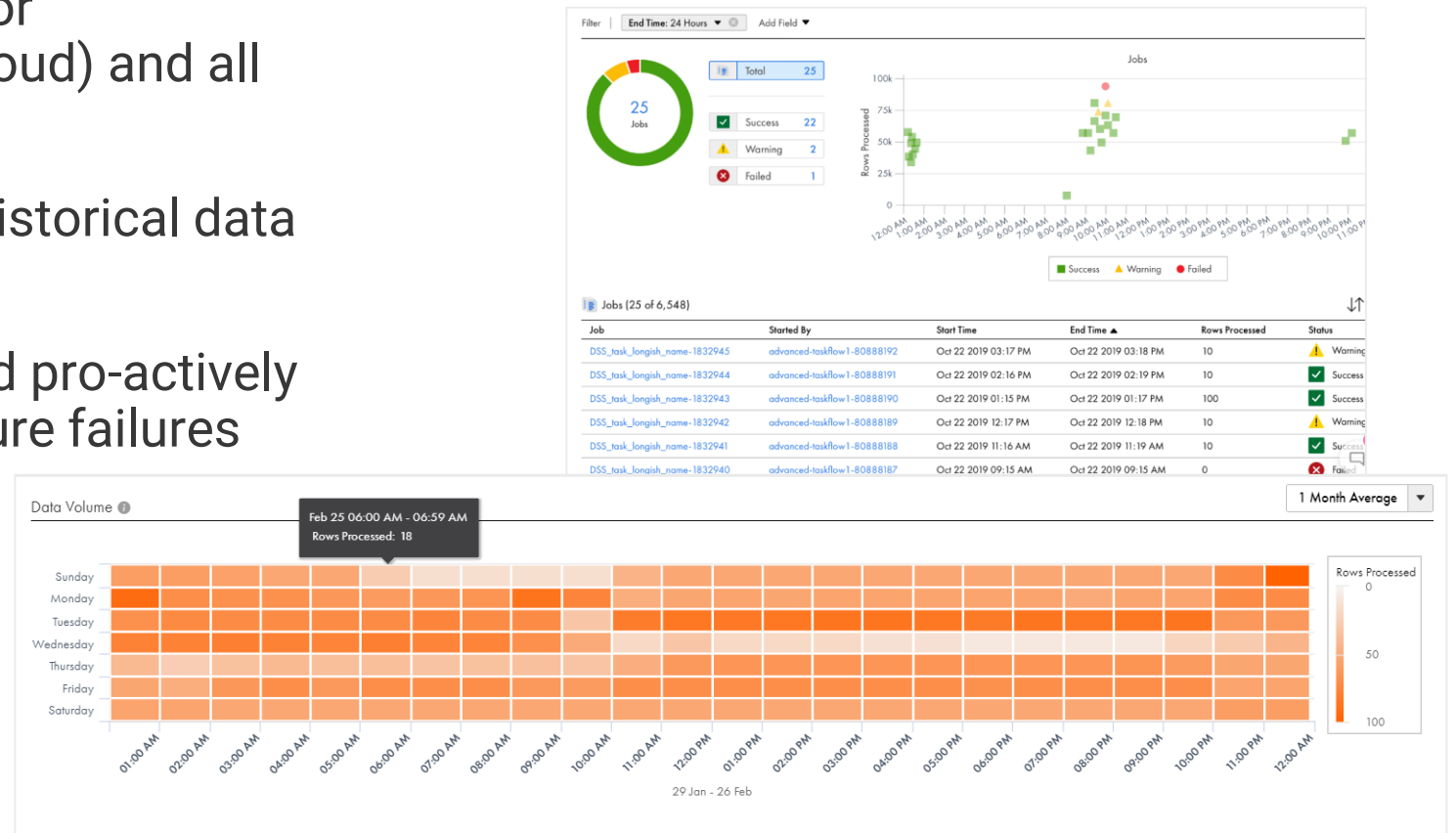
Developer Productivity

- Template and Wizard-Based Development
- Validate your mapping logic as you build with in-line data preview
- Build once and reuse business logic with the support for mapplets which are user-defined functions
- Jump start your integration projects with out-of-the-box templates available for common data warehousing patterns
- Turn your mapping into a template with parameters and promote re-use and consistency across business units



Operations

- Consolidated service to monitor infrastructure (on-premises/cloud) and all your jobs and connections
- Identify and spot trends with historical data and visual analytics
- Configure alerts to monitor and pro-actively respond to job and infrastructure failures



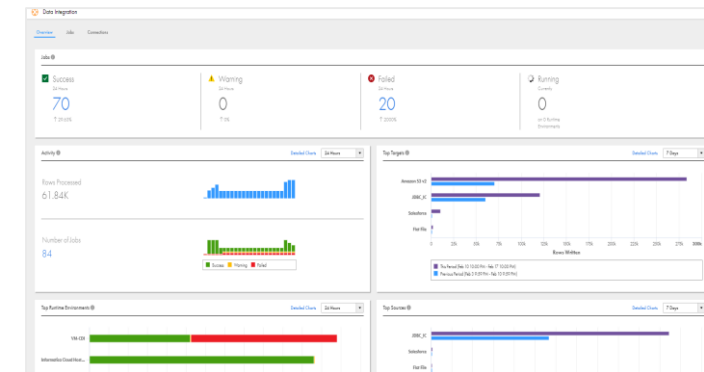
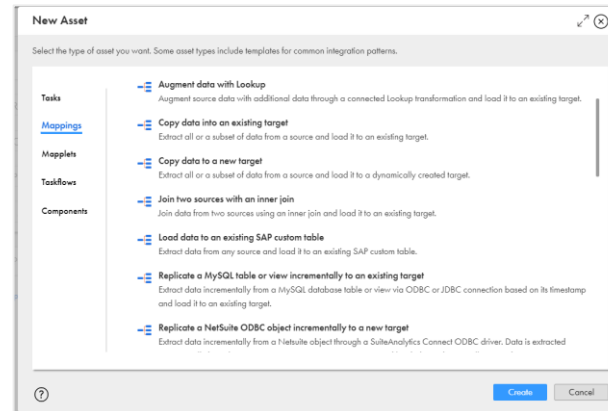
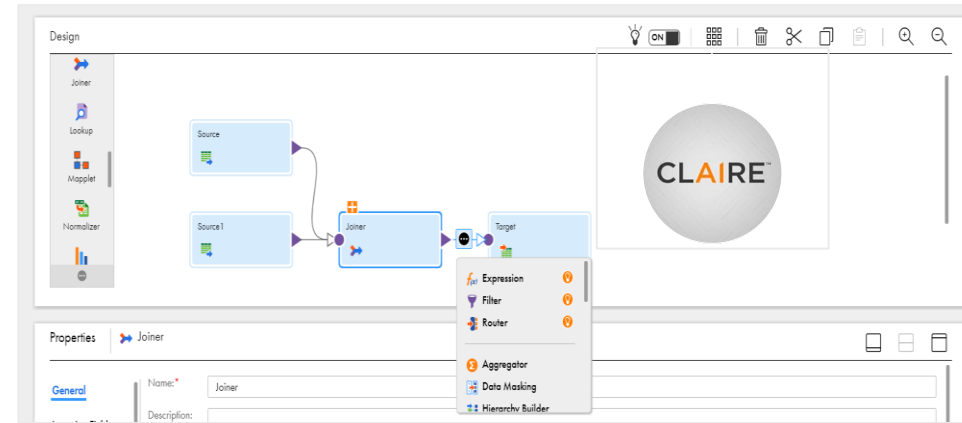
Cloud Data Integration

Step up performance of your data integration and data warehouse

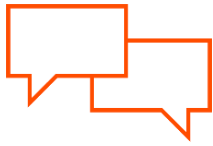
Build business processes with flow wizards, integration templates and AI-driven recommendations

High-performance loading with advanced partitioning and pushdown optimizations

Visual dashboards to monitor and manage your integrations and infrastructure seamlessly



Plug-and-Play Connectivity to **Any Data Type**



Social Data



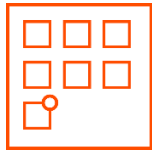
Machine Data



Data Lake



IoT Data



Application Data



SaaS Applications



Big Data



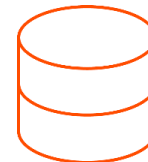
Web Services



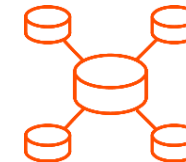
Mainframe



Local Files

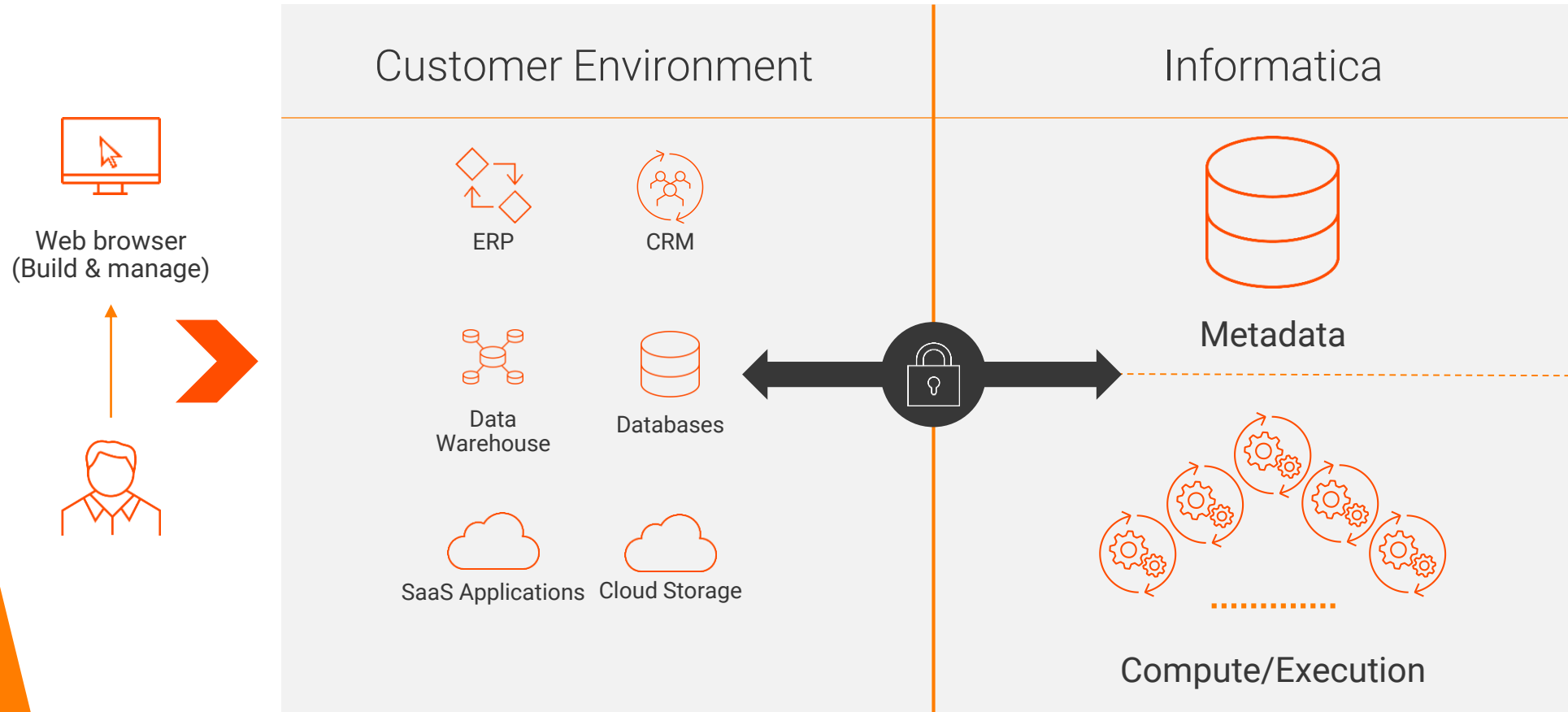


Databases



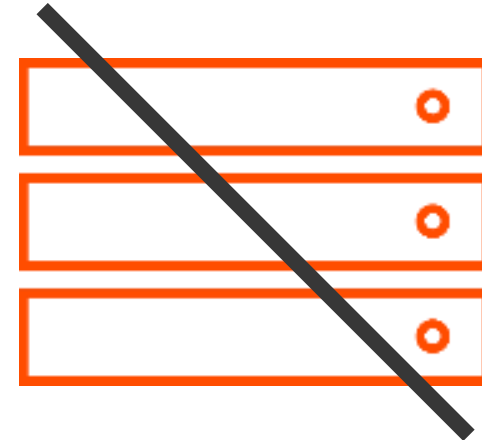
Data Warehouses

Serverless Integration (New!)

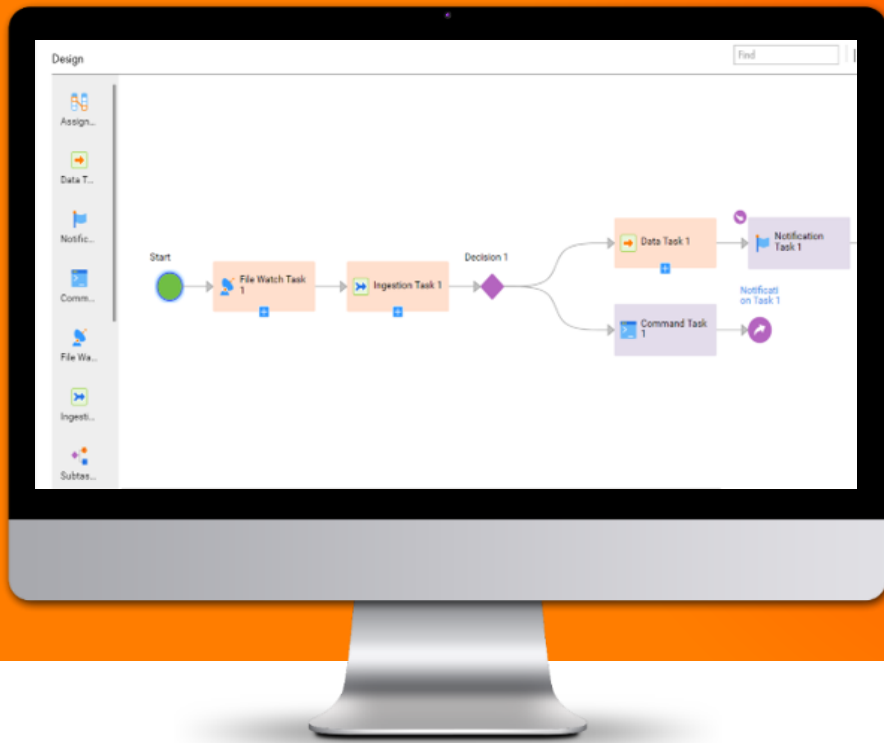


Serverless Integration (New!)

- No servers to manage
- Continuous scaling
- Built-in resiliency and highly available
- Elastic
- Consumption-based pricing



Demo



Summary

- Ability to easily integrate with any data
- Built in data quality rules
- Support for multiple patterns
- Integrated cataloging
- Flexible deployment
- CLAIRE™-powered recommendations

Please email your CSM or Account Manager if you are interested in trying out these products



Thank You