

26th – 28th April 2022

Informatica Cloud Data Integration Bootcamp

Global Technical Alliances (GTA) and PTS Team

Agenda – Day1

9:30 AM–11:30 AM BST | 10:30 AM CEST – 12:30 PM CEST | 2:00 PM–4:00 PM IST

1 Welcome Note
and Program
Overview

2 Introduction to
IDMC

3 Cloud DW/DL
Modernization

4 Cloud Mass
Ingestion

5 Cloud Data
Integration

6 Cloud Data
Integration–E

Today's Presenters



Anil Datar

VP GTA



Salvatore Moretto

Director PTS



**Archana
Sundarakrishna**

Tech. Alliance &
Program Manager GTA



Pascal Hurel

Sen. Solutions Architect



Kilian Ingelfinger

Sen. Solutions Consultant

CDI Bootcamp : Associate

Recommended Prerequisites

1. CDW/DL Foundation Cert
2. Cloud On-boarding



Weekly Checkpoint calls –



In-Person

1. Three Day SME-led Bootcamp



Self-paced

1. Cloud Data Integration Services
Informatica University Course

1. Special Guest session

1. Closing Notes, INFA Guest Speakers

2. Bootcamp Exam & Associate Badge

1. Cloud Data Integration Services
Informatica University Course

1. CAI Services (IU Training Course)

1. CAI Services (IU Training Course)
2. IICS Professional Cert**

All trainings in this timeline are free of charge

** IU Certifications involve costs. Select Top Performers are entitled to free certification vouchers

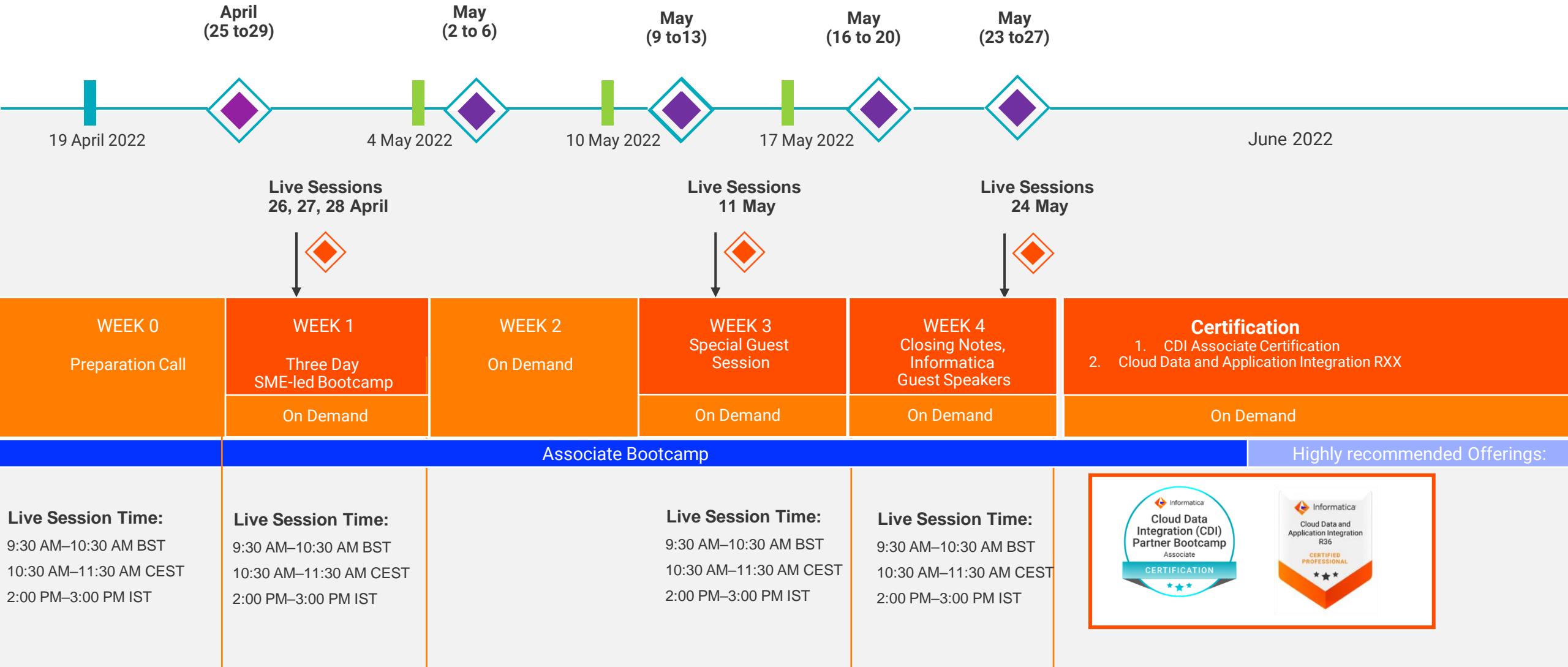
Weekly Cadence Call

9:30 AM–10:30 AM BST
2:00 PM–3:00 PM IST

Preparation Call

9:30 AM–10:30 AM BST
2:00 PM–3:00 PM IST

Key Initiative Timeline





IDMC

Introduction

IDMC vs. IIKS

Informatica's

Intelligent Data Management Cloud

... is a cloud native, comprehensive, by AI/ML-capabilities enhanced and automated **Cloud Data Management Platform** that is comprised of modular services like...

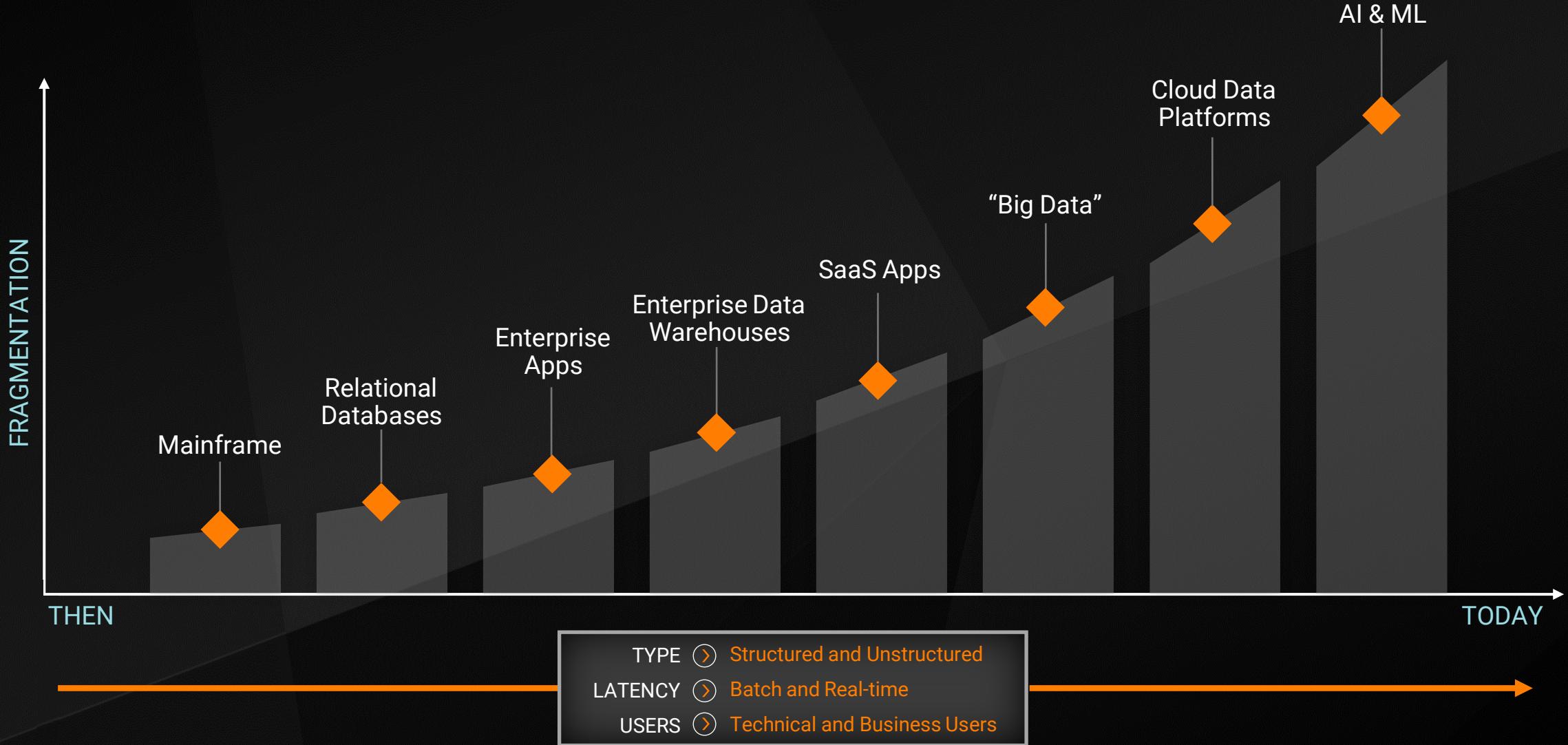
... Cloud Data Integration, Cloud Data Quality, Cloud Data Governance and Catalog...

Informatica Intelligent Cloud Services

... are exactly these modular **Services** the IDMC is comprised of, like...

... CDI, CDI-e, CDQ, Data Profiling, CDGC, CDMP, CMI, Monitor, Administrator, C360 SaaS, API Manager, API Portal, B2B Gateway, Application Integration, Integration Hub, Operational Insights...

Data Complexity Increasing Exponentially



Companies are Challenged by Data Complexity

68%

of organizations have not been able to operationalize AI across the organizations

65%

of organizations can't effectively manage customer experience in a multichannel journey

88%

of organizations are ill prepared for future supply chain disruptions

75%

of employees don't feel fully prepared to use data effectively and compliantly

75%

of organizations are actively migrating data to the cloud

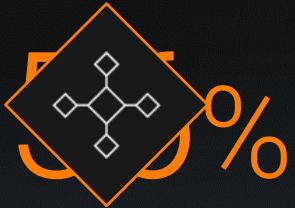
90%

of organizations, by 2022, will utilize multiple CSPs and will require significant augmented data management and integration

Companies are Investing in These Initiatives for Digital Transformation



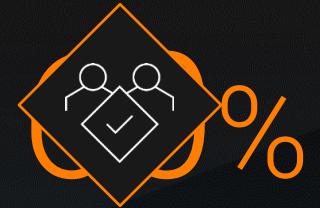
Approximately 40% of companies negative
affordability of AI/ML
practices that undervalue data



Approximately 40% of companies bought
integrated and
beyond integration



More than 30% of
managers say
applications improve
their master and
reference data functions



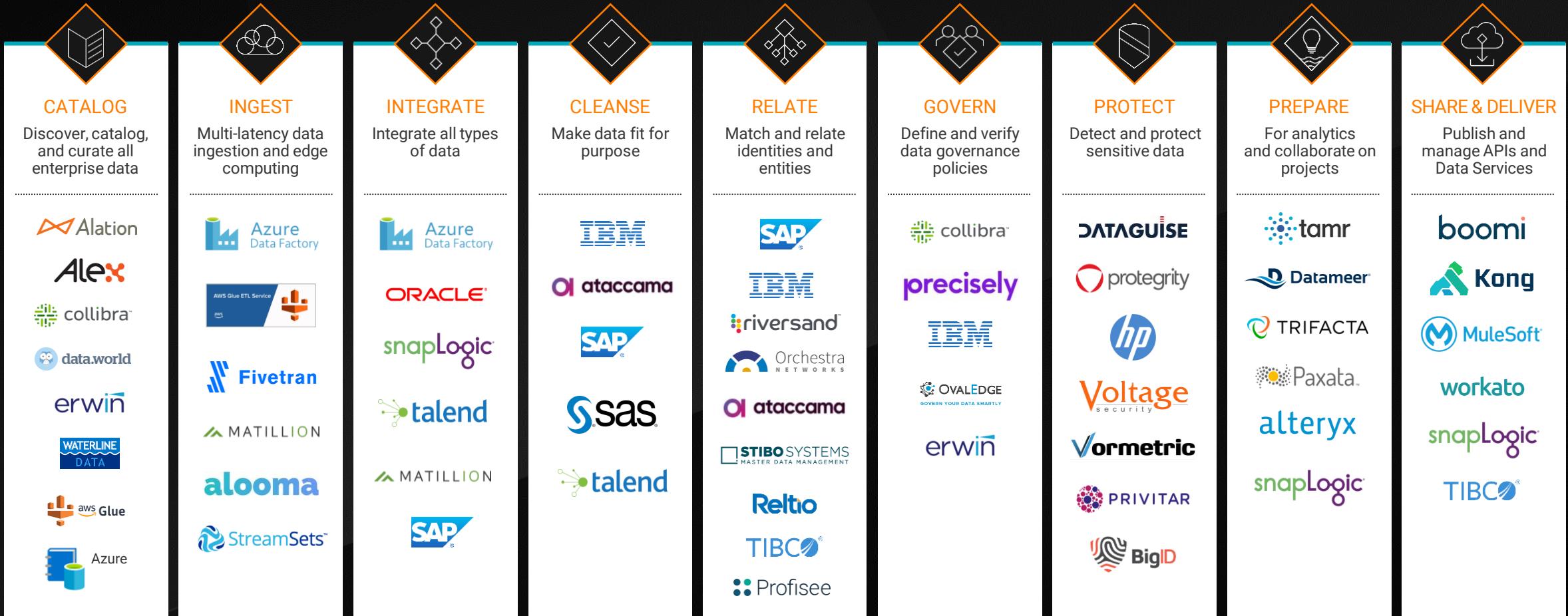
Data governance are
challenged by data
quality and complexity



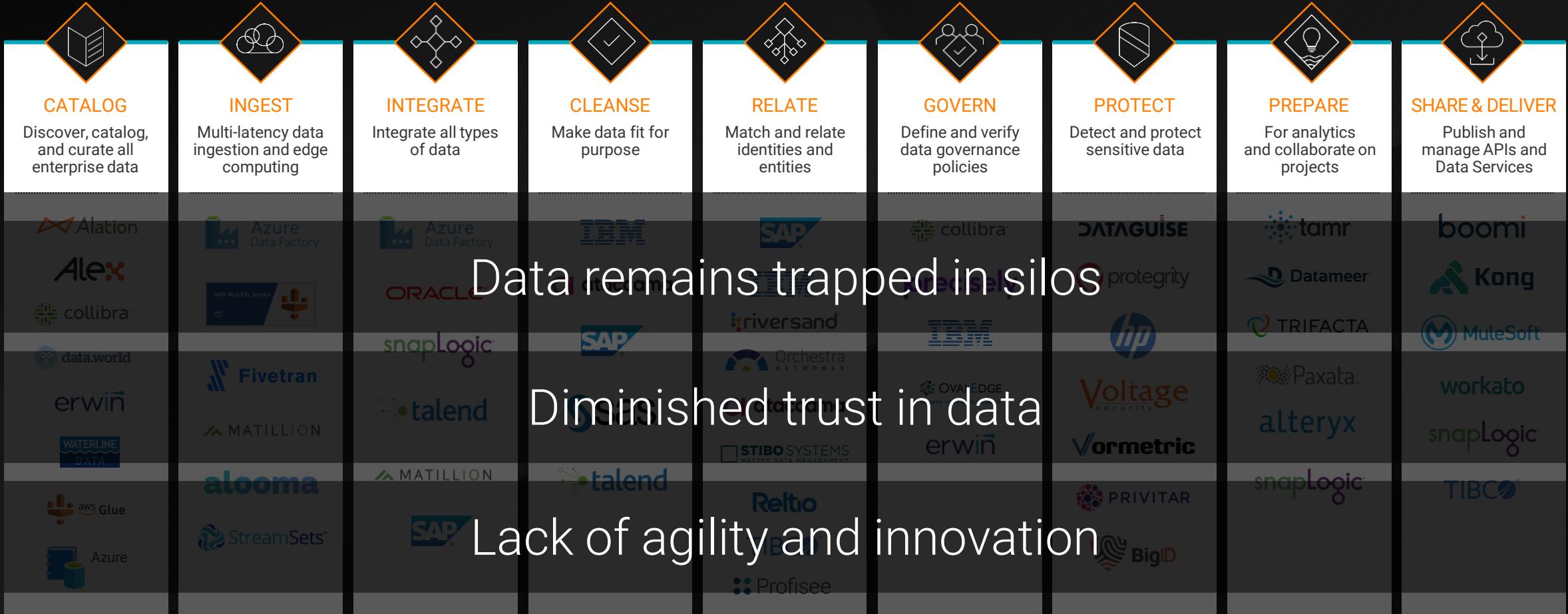
Cloud Modernization



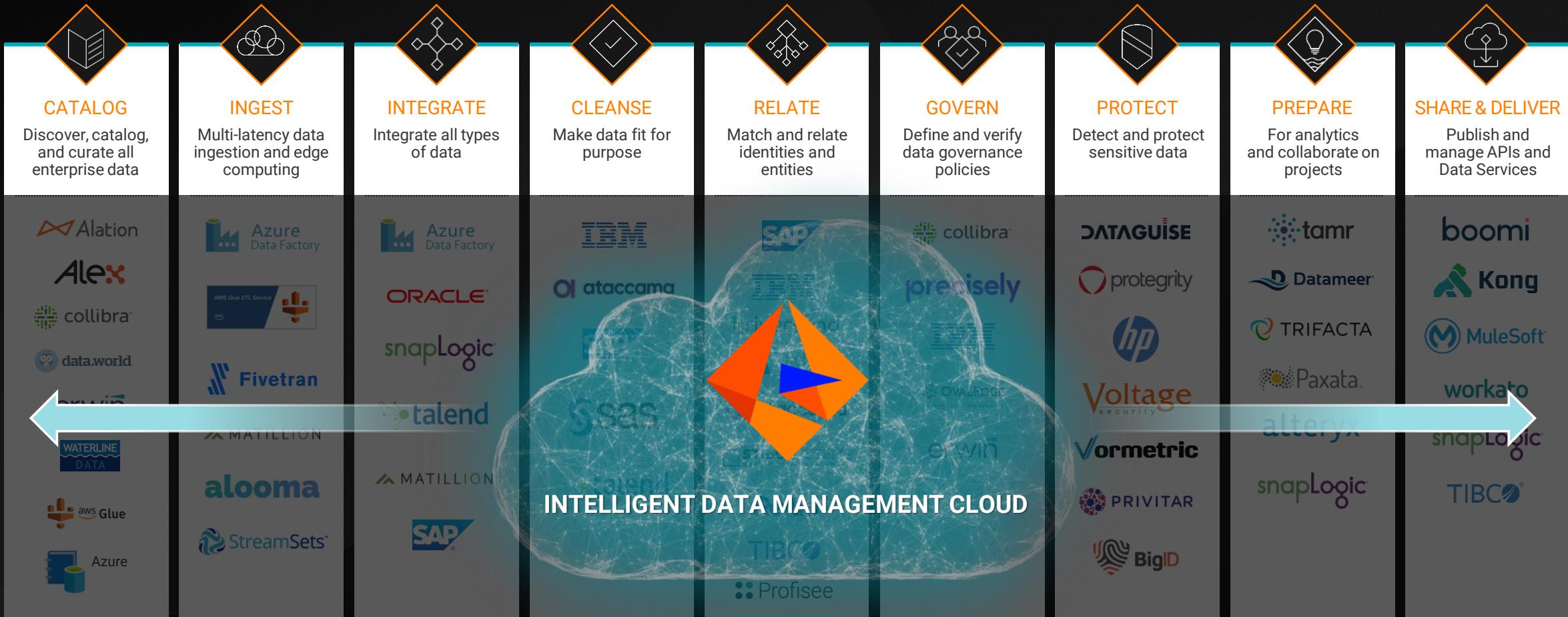
Data Management Landscape is Fragmented



Data Management Landscape is Fragmented



IDMC Delivers Best-of-Breed Products in a Single Platform

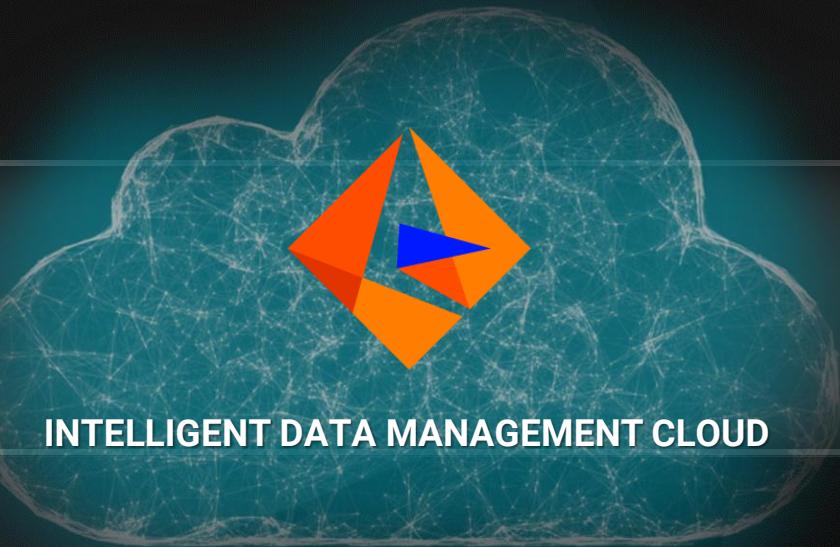


Pioneering the Intelligent Data Management Cloud

ANY
DATA



MULTI
CLOUD



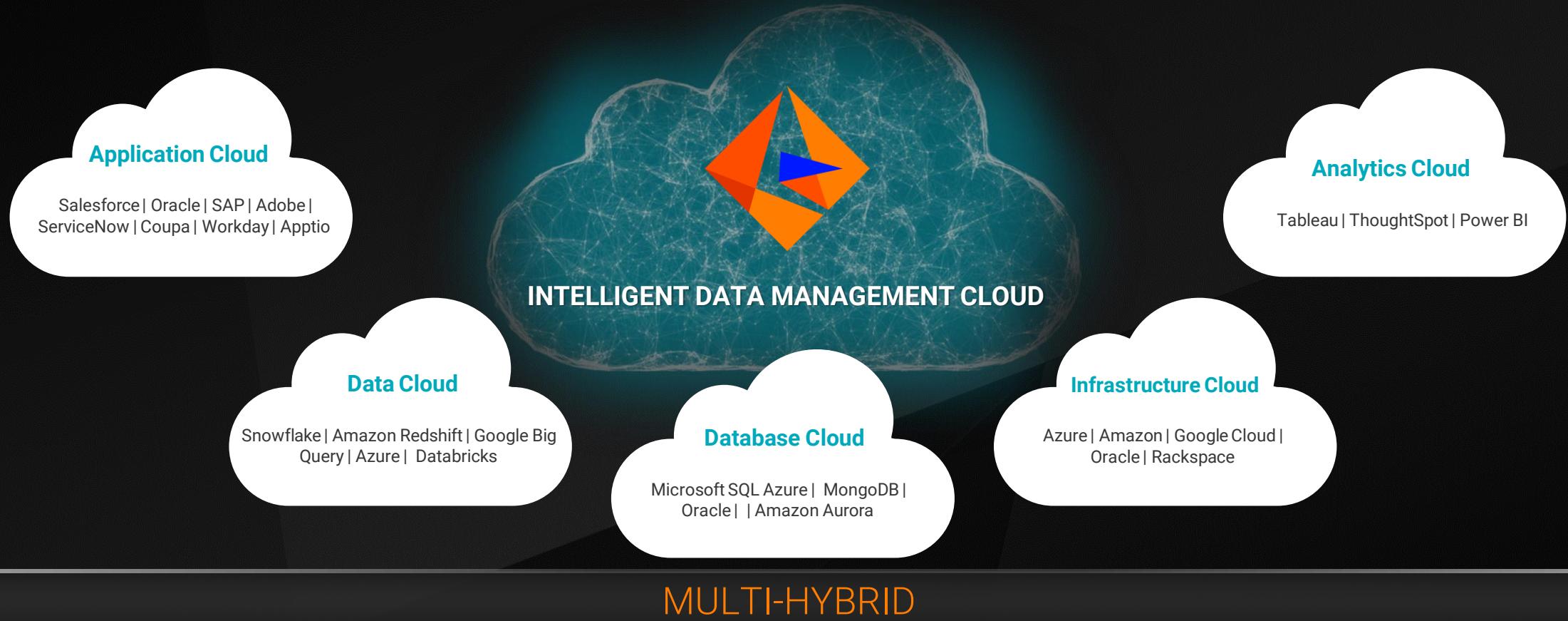
BATCH, REAL-TIME
& STREAMING



ANY
DATA USER



Pioneering a New Layer in the Enterprise Data Stack



INFRASTRUCTURE



DATABASES



DATA WAREHOUSE



BUSINESS
INTELLIGENCE TOOLS

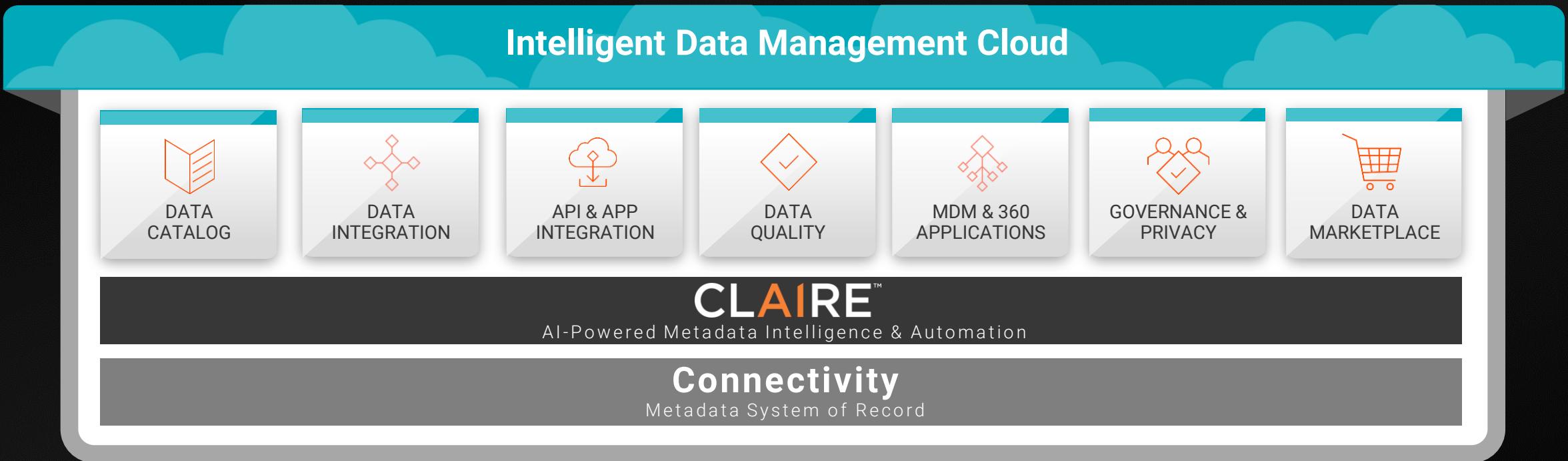


CUSTOM APPS



PACKAGED APPS

Best-of-Breed Product Suite



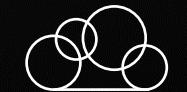
SaaS



Self-Managed



On-premises



Enterprise Cloud

IDMC Built for the Next Generation of Data Management Needs

Modern Technology Architecture



Powering Modern Capabilities



Cloud-native



Microservices-based



API-driven



Multi-cloud, hybrid

- ✓ Google-like search and discovery
- ✓ Single source of truth
- ✓ Low-code / no-code
- ✓ Single pane of glass
- ✓ Data democratization

28

Trillion

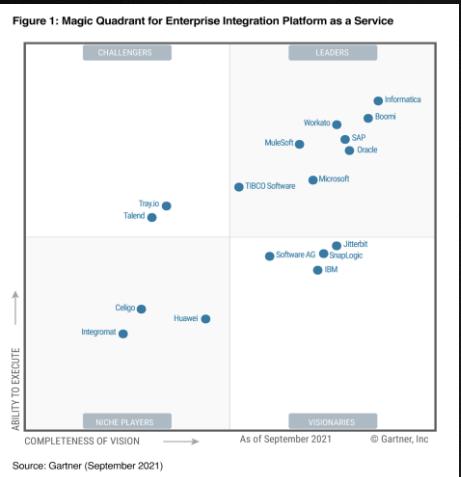
Transactions per month



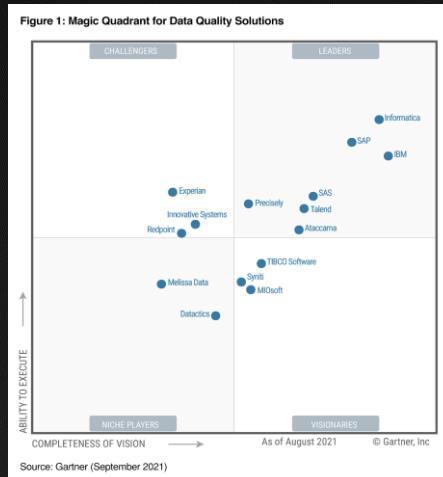
Best-of-Breed Products—A Leader in Four Gartner® Magic Quadrant™ Reports

Informatica is placed highest in ability to execute in **ALL** of these Magic Quadrant Reports

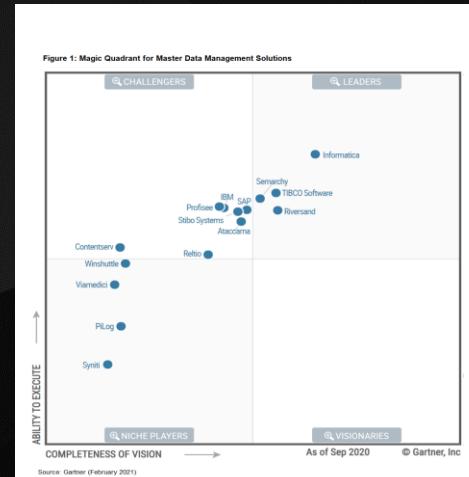
Enterprise Integration Platform as a Service



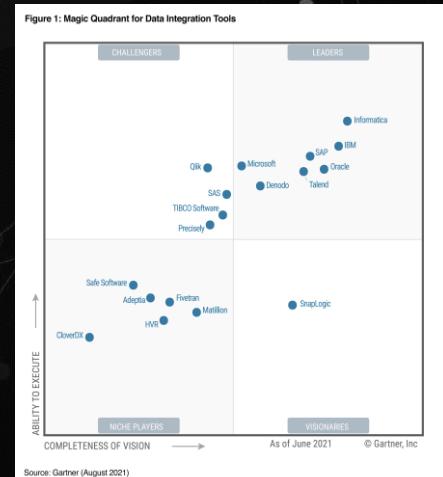
Data Quality Solutions



Master Data Management Solutions



Data Integration Tools



Sep 2021
Eric Thoo, et al.,
29 Sep 2021

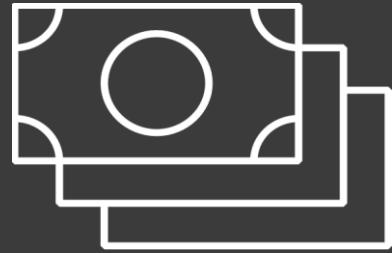
Sep 2021
Melody Chien, et al.,
29 Sep 2021

Jan 2021
Simon Walker, et al.,
27 Jan 2021

Aug 2021
Ehtisham Zaidi, et al.,
25 Aug 2021

These graphics were published by Gartner, Inc. as part of larger research documents and should be evaluated in the context of the entire document. The Gartner documents are available upon request from Informatica. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose. GARTNER and Magic Quadrant are registered trademarks and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and are used herein with permission. All rights reserved.

How you can benefit from IDMC



Higher Services
Revenue



One Platform
One Skill



Higher CSAT for
Repeat Business



Cloud Modernization

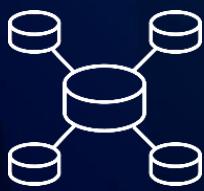
Data Warehouse, Lakes and App Modernization

Cloud Modernization

Data Warehouse, Lakes and Application Modernization Journey



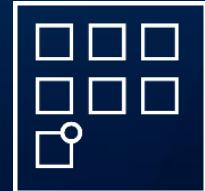
Data Engineering



Data Warehouse



Data Lakes



Applications



Data Science

Integration is the Common Theme



Google Cloud



databricks



snowflake®



Source:

1 – Harvard Business Review Services Survey, 2019



Informatica®

DATA CONSUMERS



ETL Developer



Data Engineer



Citizen Integrator



Data Scientist



Data Analyst



Business Users

Intelligent Data Management Cloud

DISCOVER & UNDERSTAND



DATA CATALOG

ACCESS & INTEGRATE



DATA INTEGRATION

CONNECT & AUTOMATE



API & APP
INTEGRATION

CLEANSE & TRUST



DATA
QUALITY

MASTER & RELATE



MDM & 360
APPLICATIONS

GOVERN & PROTECT



GOVERNANCE &
PRIVACY

SHARE & DEMOCRATIZE



DATA
MARKETPLACE

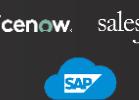
CLAIRE™

AI-Powered Metadata Intelligence & Automation

Connectivity

Metadata System of Record

DATA SOURCES



salesforce

SaaS Apps
Sources

+



On-premises
Sources

+



Real-time /
Streaming
Sources

Cloud Data Management Design Principles



Radical Simplicity

Simple and Easy Data Access,
Processing and Consumption
for All Data Practitioners



Productivity for ALL Users

10X to 100X Productivity
for Data Practitioners with
AI-Powered Automation



Elastic Scale

Cloud First, Cloud Native
Data Management at
Enterprise Scale

Microservices-based

API driven

Multi-tenant

Multi-cloud



Informatica®



Simplicity

Address Any Cloud Integration Need

Cloud Data Integration

Codeless, and optimized data integration

Cloud Data Quality

Identify, fix, and monitor data quality problems

Cloud Data Integration - Elastic

Dynamic scaling and automation for streamlined cloud-based processing on serverless Spark engine

Cloud Integration Hub

High-performance, governed pub/sub data hub

Advanced Serverless

Managed environment with no cloud administration, software, servers or clusters to manage.

Cloud B2B

Self-service partner onboarding and management

Cloud Mass Ingestion

Ingest Database, Change Data Capture, IoT, Streaming

Cloud Application Integration & API

No-code, no-build UX to build, monitor & maintain integrations and API

Simplicity Promotes Self-Service

Modern UX

No code/no build

5 step wizard

Pre-built templates



Architect



Operations



Application Developers



IT Specialist



Data Engineer



Data Scientist



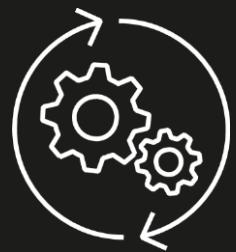
Business Analyst

Data Integration

Application & API Integration

Business Process Integration

Self-service Experience for All Users



Productivity

Only AI @ Scale Delivering the Enterprise System of Record for Metadata



Most Comprehensive Active Metadata Across the Enterprise

Broadest Deployment of AI/ML Algorithms

Market Leading AI for Data Management

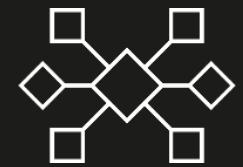
11 Petabytes of active metadata powering intelligent and automated data management

- ✓ Google-like search
- ✓ Amazon-like recommendations
- ✓ Social-graph of data
- ✓ Democratization of data
- ✓ NLP to create rules
- ✓ Automated match, merge and consolidation

Productivity is Driven by Automation

CLAIRE™

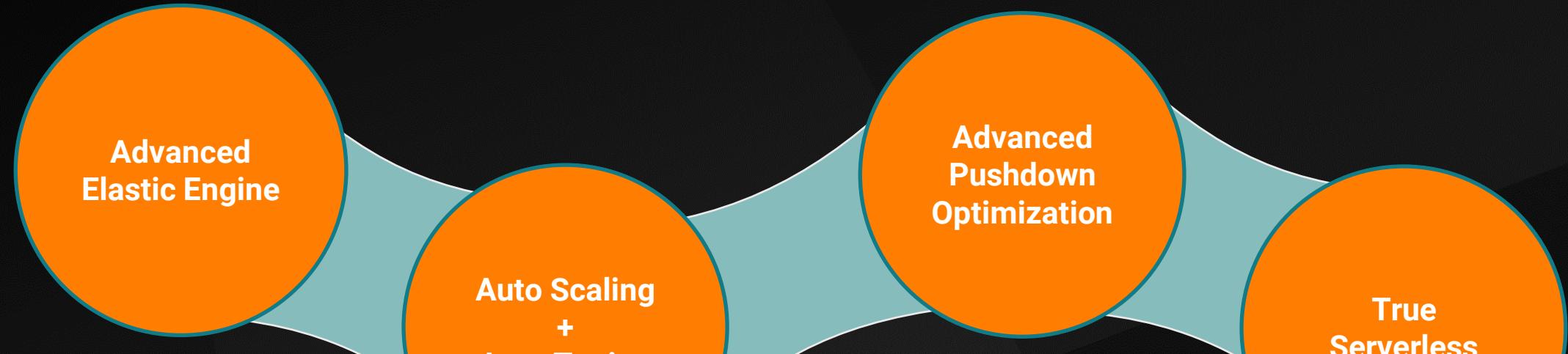


A white icon depicting a network structure, consisting of a central diamond shape connected to four smaller diamond shapes, which are further connected to eight even smaller square shapes, forming a cross-like pattern.

Scale

SCALE: Performant & cost-effective data management engine

CODELESS AND SERVERLESS



3TB

data processed
under 2 hours

- Router Performance
- Connectivity
- Performance such as S3 writes
- UDF Performance

65%

cheaper than commercial spark
vendor*

- CLAIRE based autotuning
- Local storage Scaling
- Workload aware auto scale
- Mapping runtime recommendations
- Spot instances

50X

more performant than
ETL

- Zero data egress charges
- Connectivity to all CDW/DL ecosystems
- Best-in-class performance
- Switch to PDO with One click

**True
Serverless**

60%

Lower TCO

- Multi-tenant compute cluster in our cloud
- No instances to manage for customers
- Easy trials
- Single bill for customers



The Informatica Difference

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



Social Data



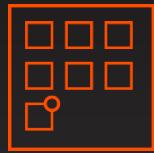
Machine Data



Data Lake



IoT Data



Application Data



SaaS Applications



Big Data



Web Services



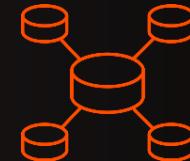
Mainframe



Local Files



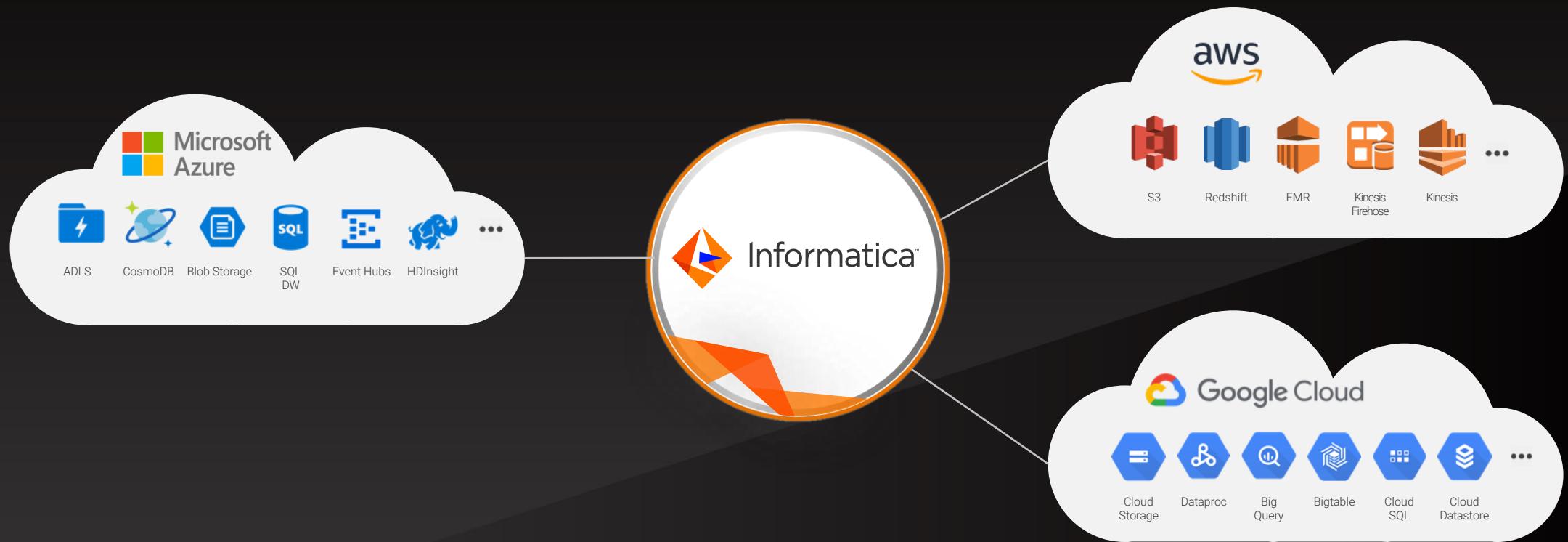
Databases



Data Warehouses

Multi Cloud Support

Informatica supports your Azure, AWS, Google Cloud, and on-prem environments



Achieve the core goal of delivering trusted, actionable data when and how the business needs it

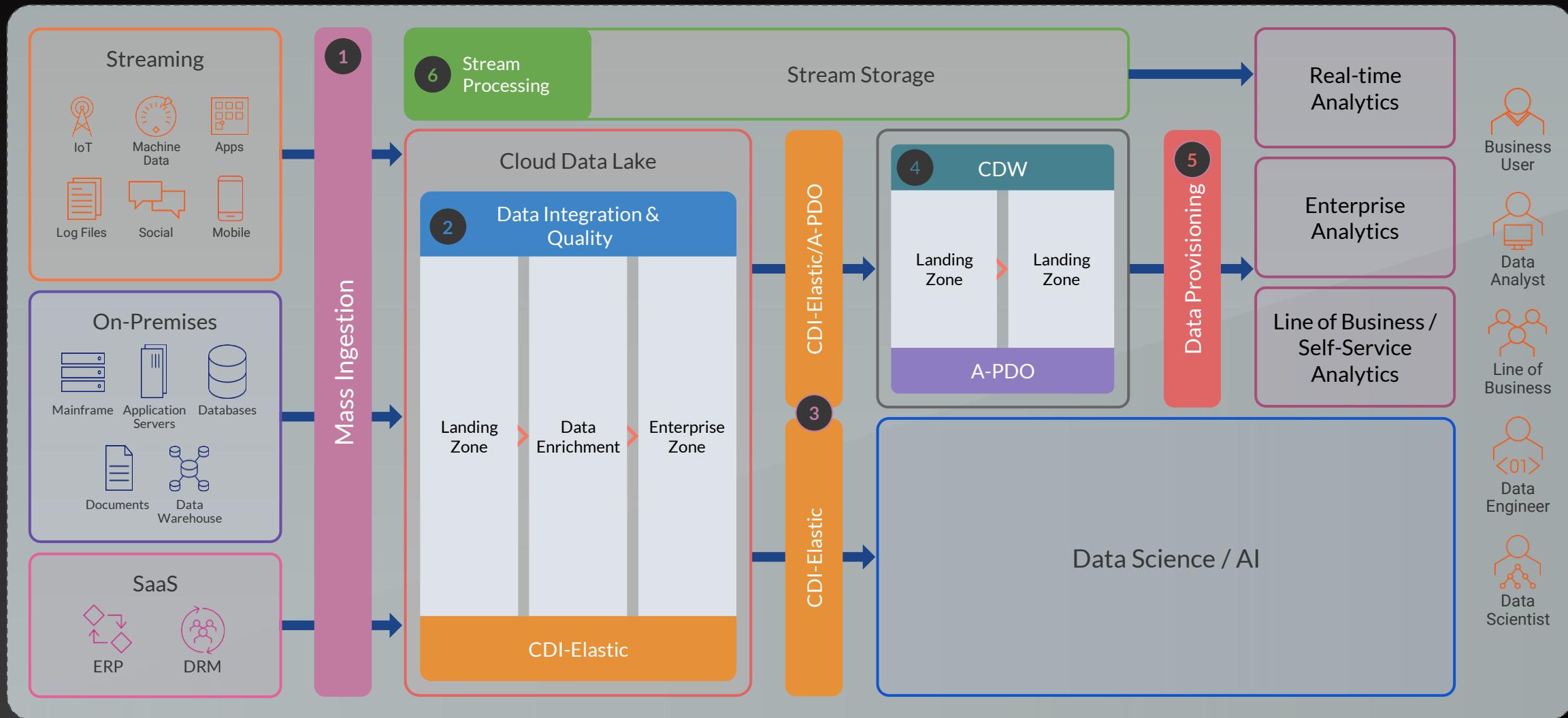
Most Secure and Trusted Cloud Data Management Provider

| | Informatica | aws | Microsoft | Google Cloud | salesforce | VENDOR A | VENDOR B | VENDOR C | VENDOR E | VENDOR D | VENDOR F | VENDOR G | VENDOR H |
|--|-------------|-----|-----------|--------------|------------|----------|----------|----------|----------|----------|----------|----------|----------|
| | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ | ✗ |
| | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ |
| | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ |



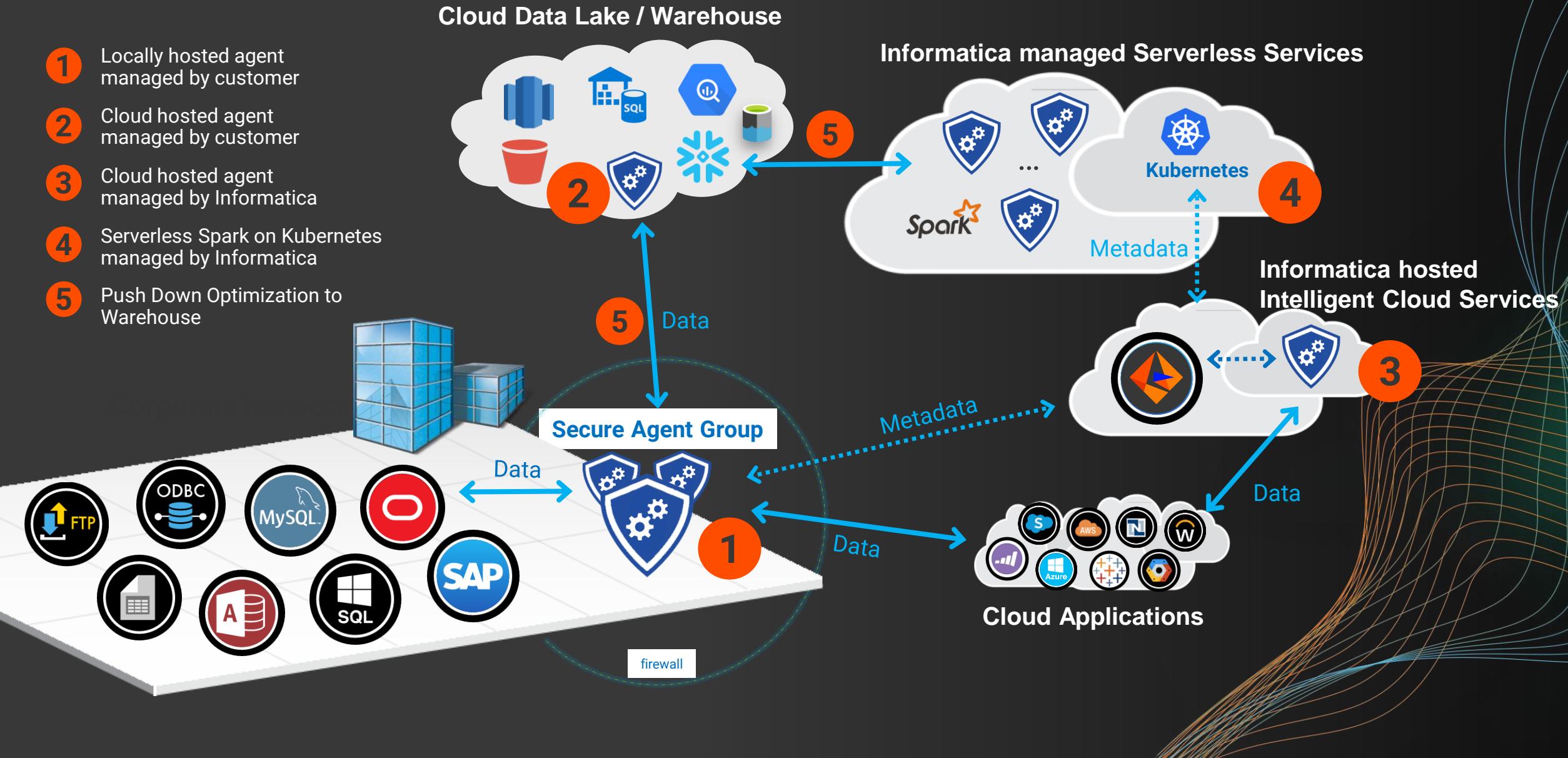
Informatica Data Warehouse and DataLake Architecture

Informatica Data Warehouse and DataLake Architecture



Informatica Cloud Integration Reference Architecture

- 1 Locally hosted agent managed by customer
- 2 Cloud hosted agent managed by customer
- 3 Cloud hosted agent managed by Informatica
- 4 Serverless Spark on Kubernetes managed by Informatica
- 5 Push Down Optimization to Warehouse



DEMO

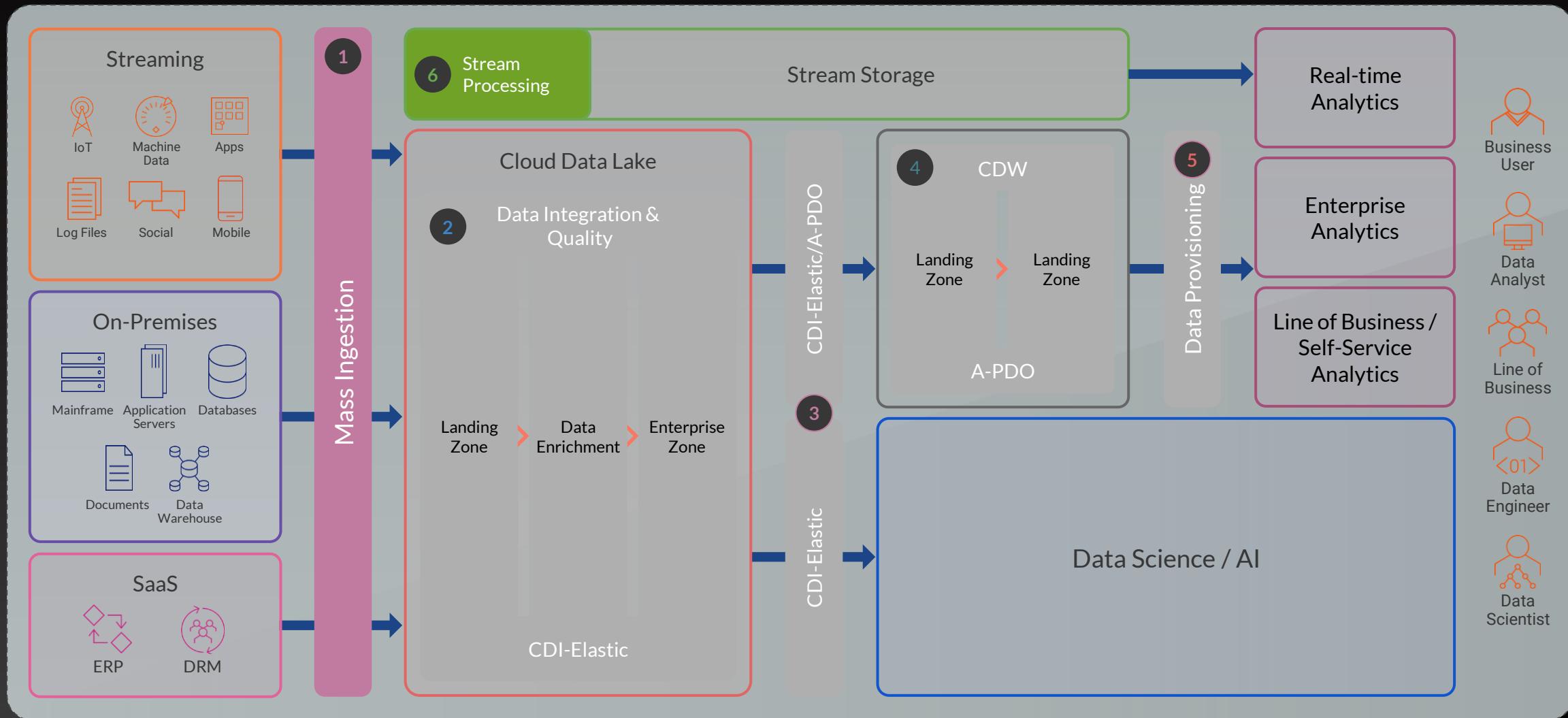
Admin Service



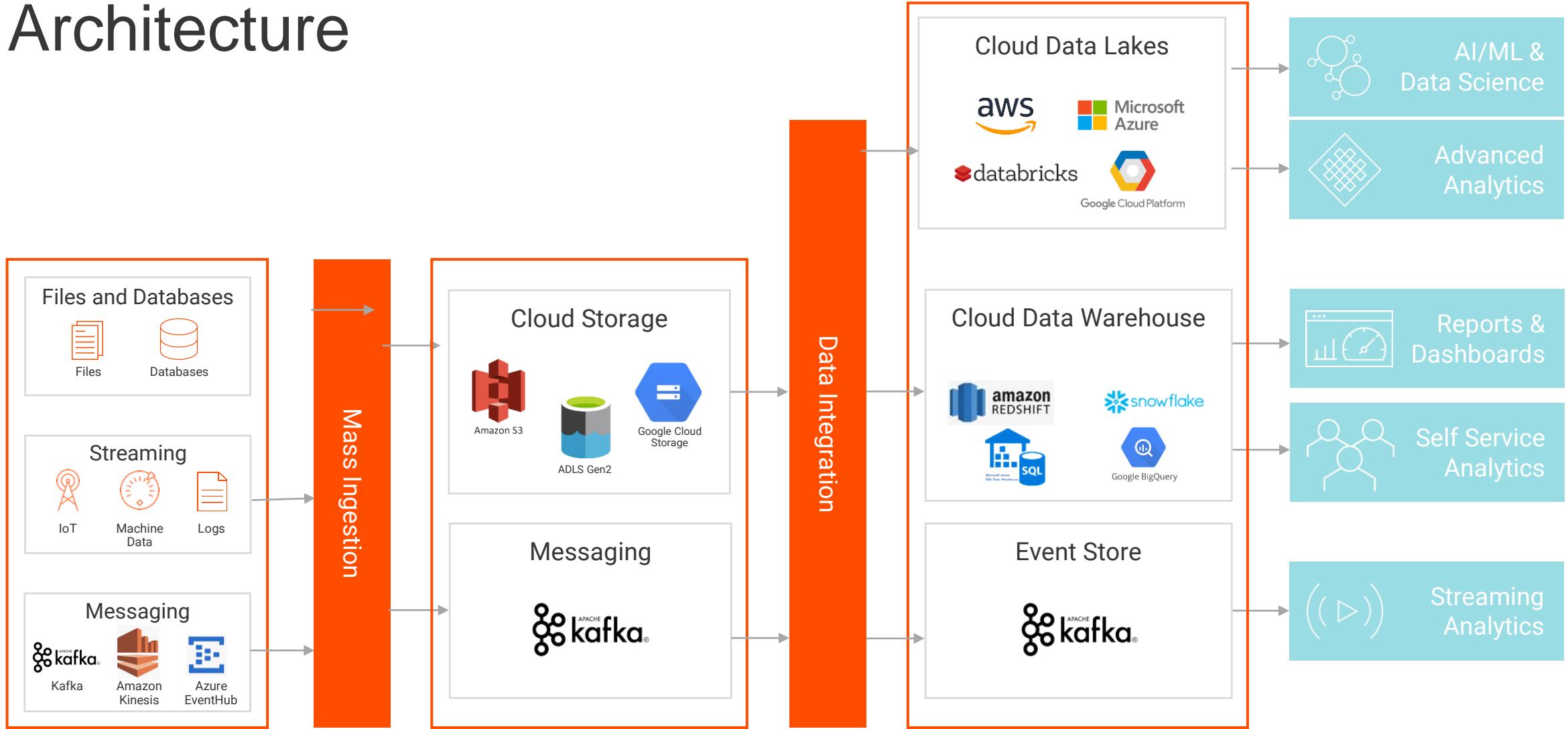


Cloud Mass Ingestion (CMI)

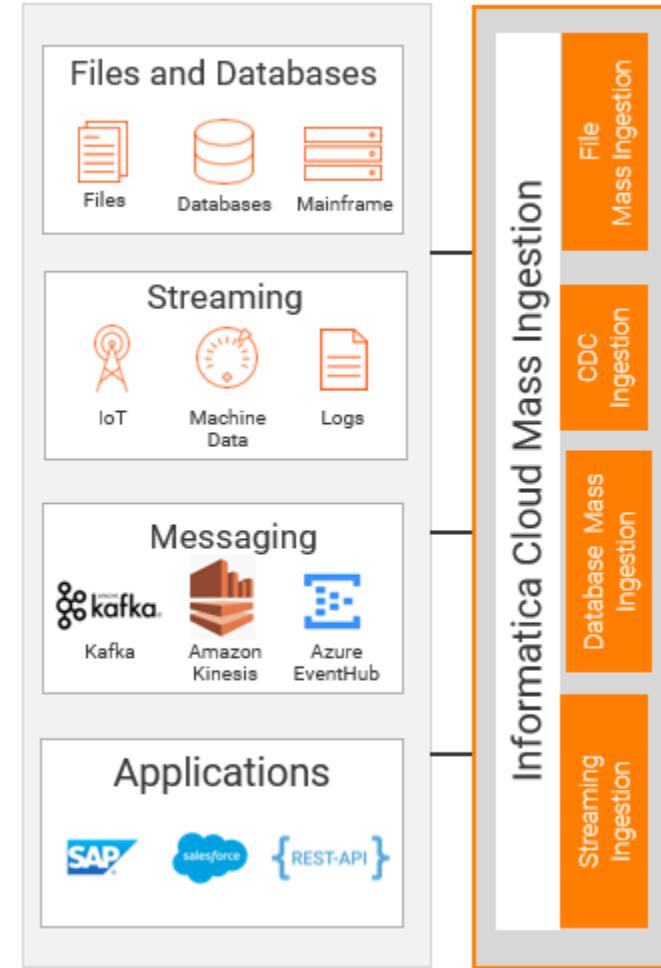
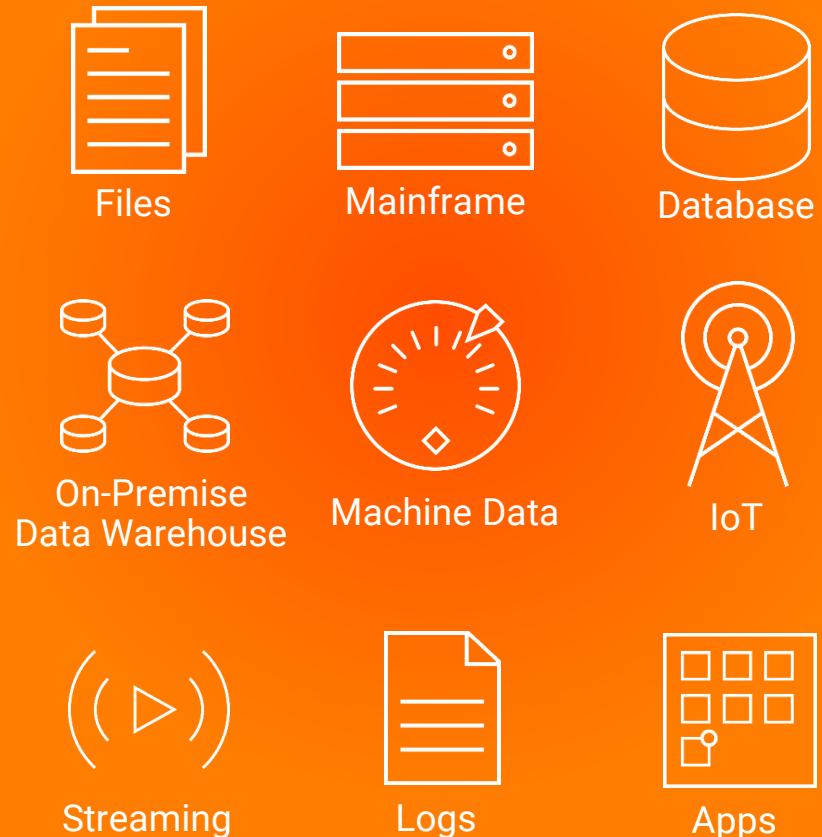
Informatica Data Warehouse and DataLake Architecture



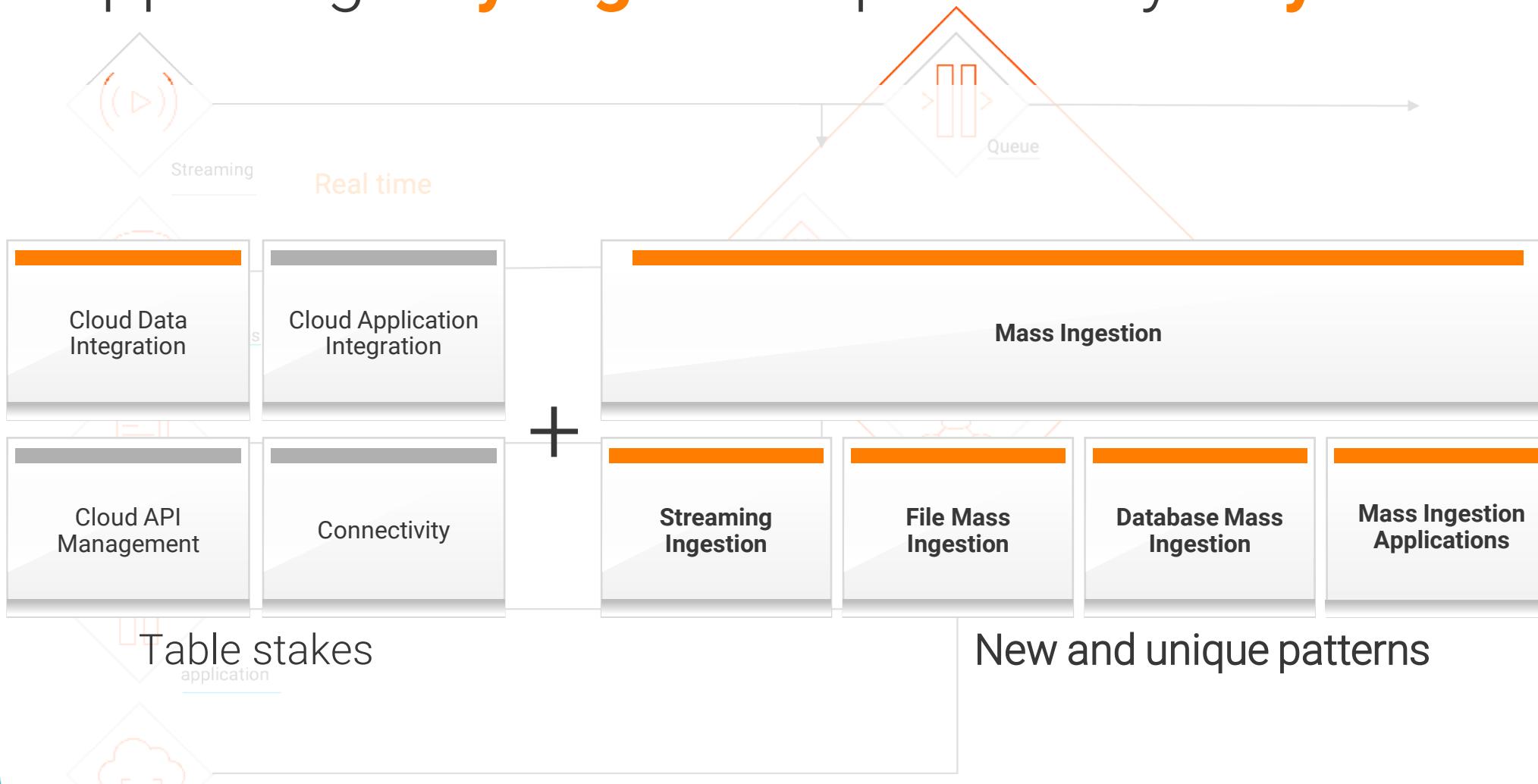
Enterprise Cloud Data Management – Reference Architecture



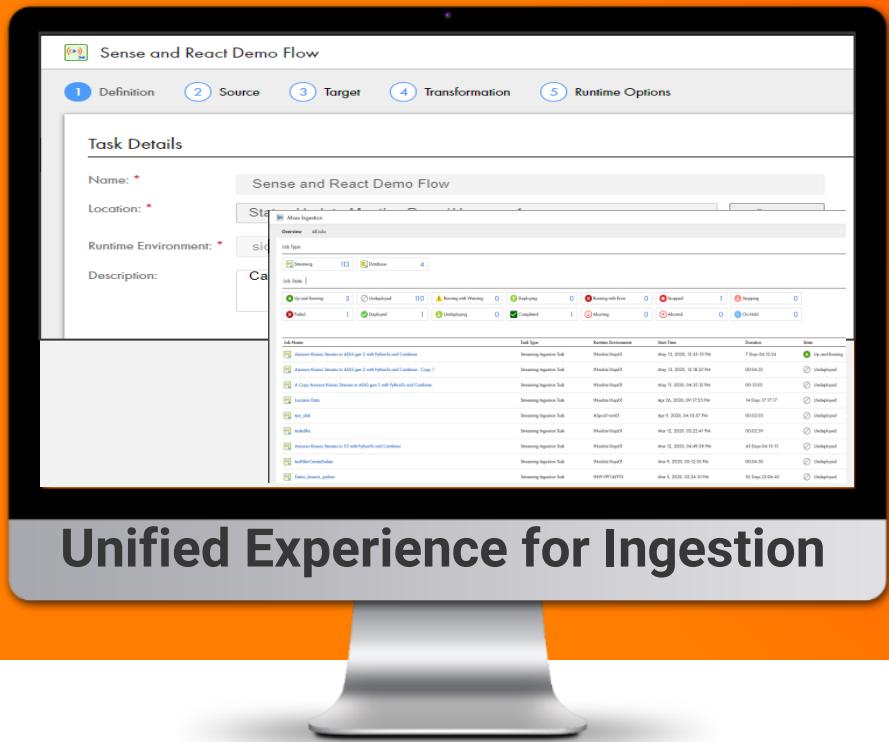
Ingesting large volume of data from a variety of sources into a CDWDL is a challenge



Supporting **Any Ingestion** pattern by **Any User**



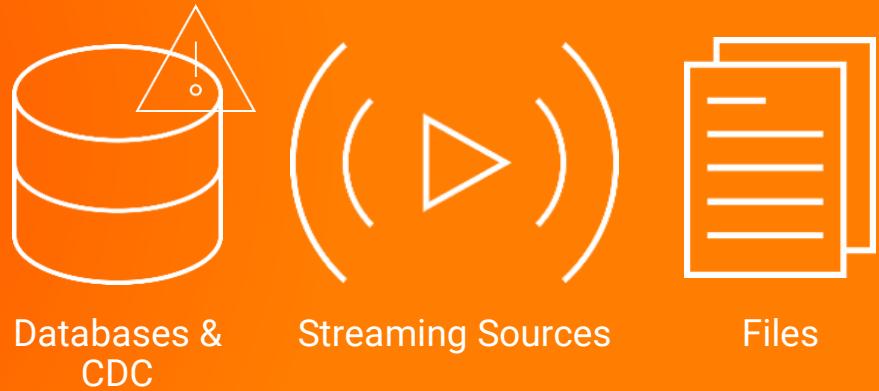
Cloud Mass Ingestion Service –Overview



Unified Experience for Ingestion



Real time monitoring



Databases &
CDC

Streaming Sources

Files

- Step by Step wizard for designing & creating an ingestion task
- Deployment, Scheduling, Real time Monitoring & Lifecycle Management
- Versatile Out of the box connectivity to sources & targets



Informatica®



Mass Ingestion Files

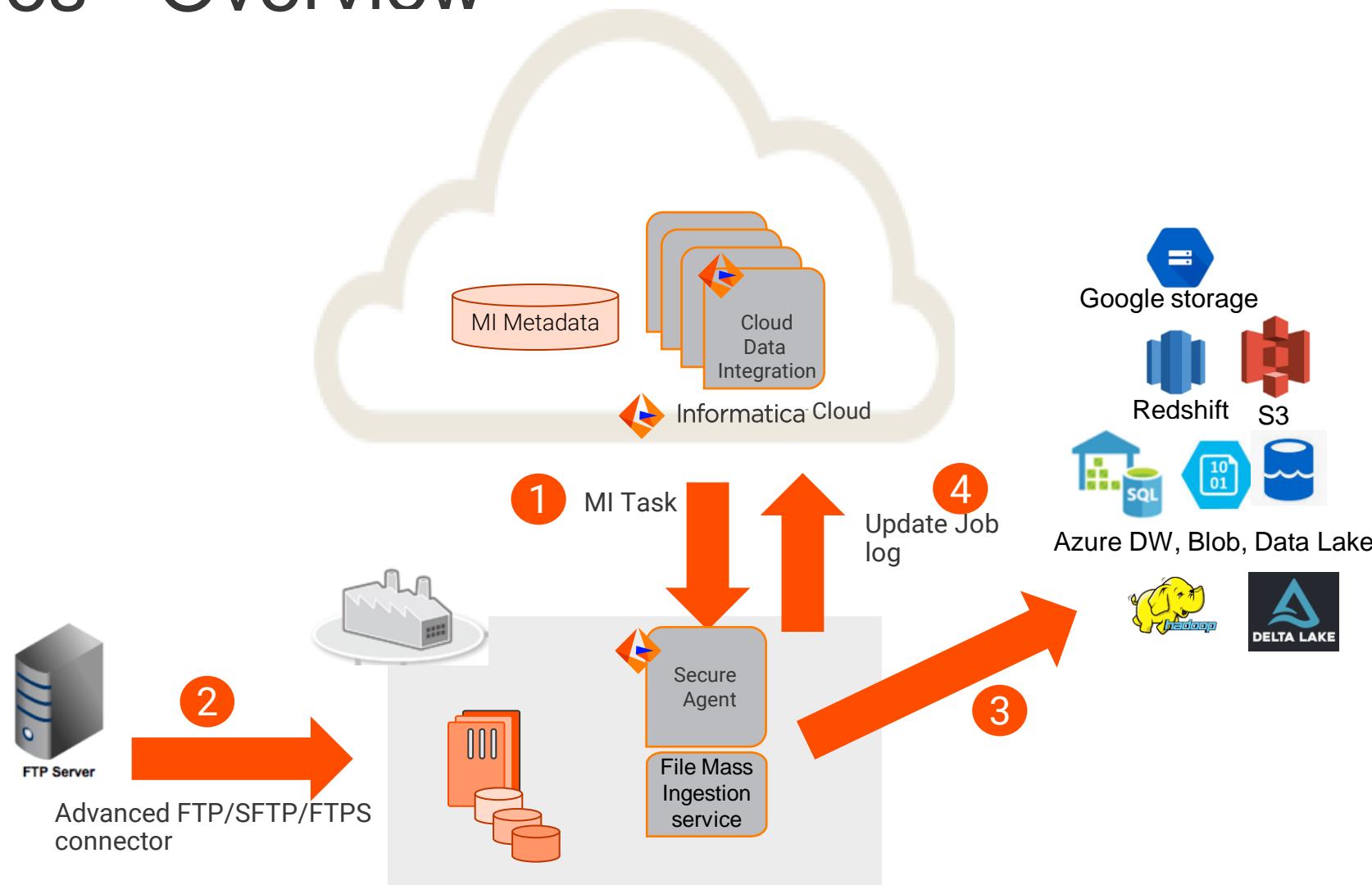
Mass Ingestion Files - Overview

Provides **file transfer capabilities** for exchanging files between on-premise and Cloud repositories, using standard protocols

Transfer **any file type** with a **high performance and scalability**

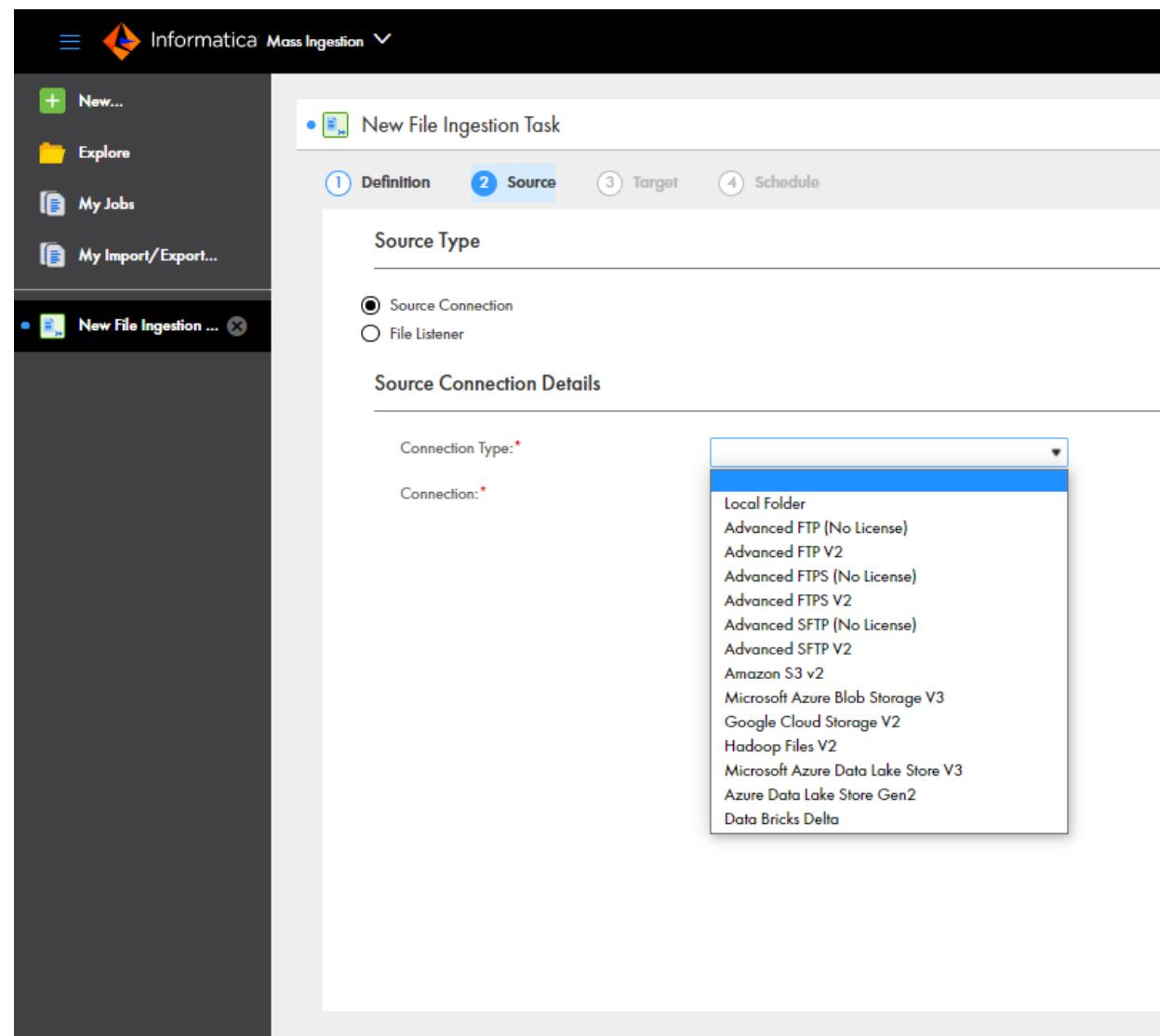
Job and file level **tracking and monitoring**

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



Main capabilities

- Unified user experience for all ingestion types (Streaming, Database, File)
- Simple, wizard-based task definition
- Wide list of supported sources/targets
- Advanced, highly scalable connectors for handling FTP/SFTP/FTPs
- Filter files by file name pattern, file size, file date



Main capabilities

- API, schedule or file event triggered
- File actions :
 - Compress/decompress (Zip, Gzip ,Tar)
 - Encrypt/decrypt (PGP)
- Highly scalable, any file type
- Unified monitoring and tracking experience
 - Tracking and monitoring - Job and file level

The screenshot displays two main windows of the Informatica Mass Ingestion application.

Top Window: A modal dialog titled "New File Ingestion Task" is open, showing the "Schedule" tab. It includes a "Schedule Details" section with three radio button options:

- Do not run this task on schedule
- Run this task on schedule
- Run this task by file listener

Bottom Window: The "Mass Ingestion" monitor dashboard. The left sidebar shows navigation links: "Running Jobs", "All Jobs", "Mass Ingestion" (which is selected), "Import/Export Logs", "File Transfer Logs", and "Source Control Logs". The main area displays monitoring statistics:

- Job Type: Streaming 0, Database 0, File 22758
- Job State: Failed 6710, Completed 16012, Aborted 36
- A table of recent job instances:

| Instance Name | Task Type | Runtime Environment | Start Time |
|---------------------------|---------------------|---------------------|-------------------------|
| adls_to_adls_SunOct0... | File Ingestion Task | AParentOrg | Oct 4, 2020, 8:06:13 AM |
| azureblob_to_redshift_... | File Ingestion Task | AParentOrg | Oct 4, 2020, 8:02:57 AM |
| s3_to_redshift_SunOct... | File Ingestion Task | AParentOrg | Oct 4, 2020, 7:59:52 AM |
| gcs_to_ftps_1SunOct0... | File Ingestion Task | AParentOrg | Oct 4, 2020, 7:59:22 AM |

File listener

Benefits

- A platform level asset that provides file listener capabilities that can be used by different services
- User can define/manage file listeners and different apps/services can register/invoke file listeners (via UI or API)
- Usage:
 - File Mass ingestion as a scheduling option-move files when they land in a specific folder
 - Taskflow:
 - Trigger taskflow when file event occurs
 - File watch inside a taskflow process
 - B2B Gateway - as a scheduling option- process files when they land in a specific folder

New Asset

Select the type of asset you want. Some asset types include templates for common integration patterns.

Tasks
Mappings
Maplets
Taskflows
Components

- Mierarchical Schema
Upload an XML schema or an XML or JSON sample file to use with a Hierarchy transformation.
- Intelligent Structure Model
After you create an intelligent structure model, you can associate it with a Structure Parser transformation and use the transformation in a mapping.
- Fixed-Width File Format
Configure reusable formats for fixed-width flat files to use in mappings and mapping tasks.
- Visio Template
Upload a Visio template from the Visio Template Designer to use the template's transformation logic in a Mapping task.
- User-Defined Function
Create a reusable function to use in transformation expressions.
- Shared Sequence
Create a reusable sequence to use in multiple Sequence Generator transformations.
- File Listener**
Create a listener that listens to files that arrive at a specific location to be used in mass ingestion tasks and other assets.

Informatica Data Integration

New... Home Explore Bundles My Jobs My Import/Export Logs New File Listener AFileListenerTestMI6642

AFileListenerTestMI6642

File Listener Details

| | |
|-----------------------------------|-------------------------|
| File Listener Name: [*] | AFileListenerTestMI6642 |
| Location: [*] | MI demo |
| Description: | Test File Listener |
| Runtime Environment: [*] | INVR7B2B10 |
| Source Type: [*] | Connector |
| Status: [*] | Enabled |
| Connection Type: [*] | Amazon S3 v2 |
| Connection: | S3 |

View

DEMO

CMI-Files





Mass Ingestion Databases

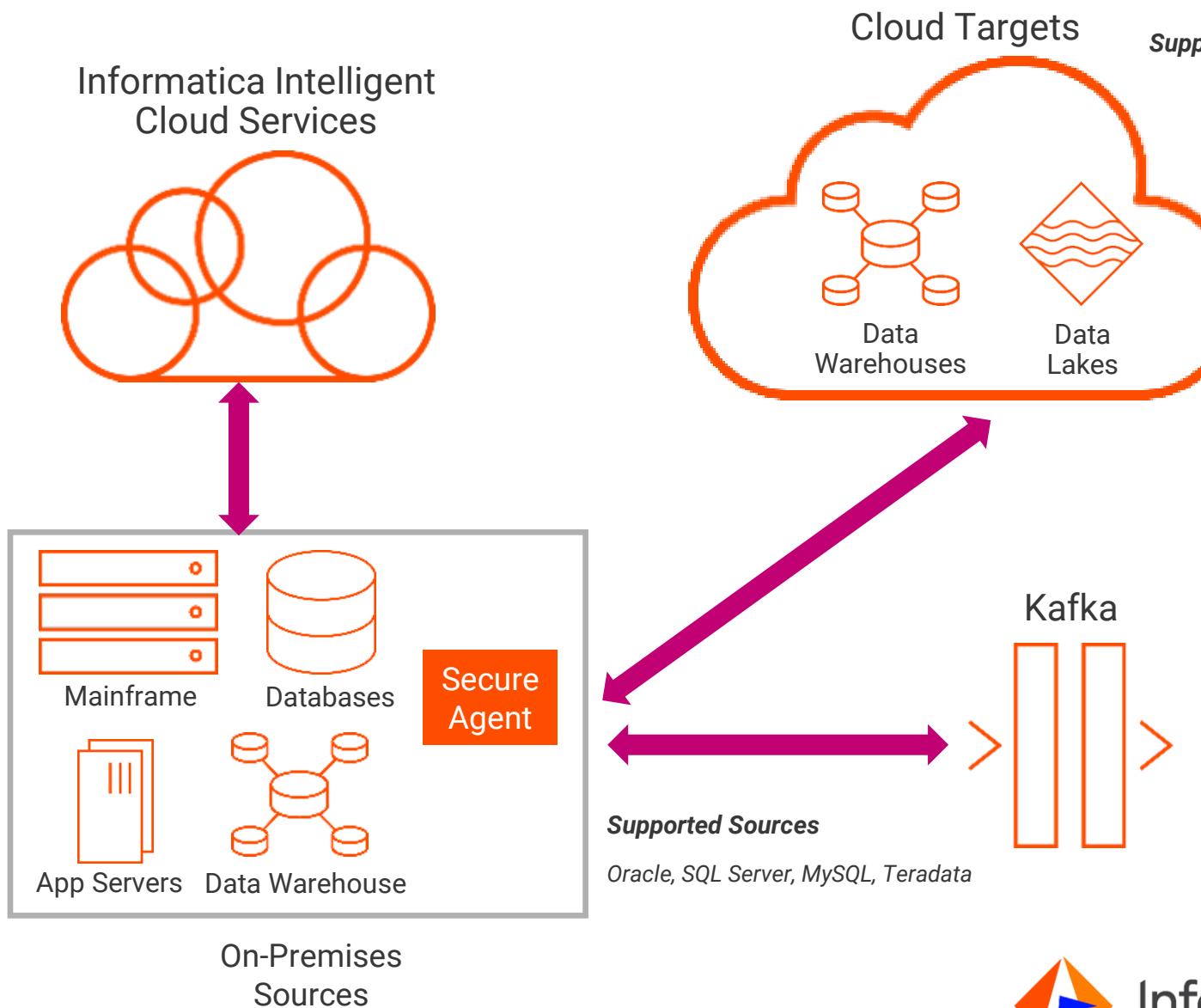
Cloud Mass Ingestion Databases

Provides Database ingestion capabilities as part of IICS Mass Ingestion service

Ingest relational database data from Oracle, SQL-Server & MySQL. Also supporting Schema Drift on CDC supported Databases

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

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



Benefits of Mass Ingestion Databases

1 Supports both data synchronization & real time analytics use cases

Faster decision making

2 Wizard driven experience for ingestion

Increase business agility

3 Efficiently ingest CDC data from 1000's of tables

No expensive maintenance

4 Automatic schema drift addressing

Increased trust in data assets

5 OOTB Connectivity to CDC sources, Data Lake & DWH targets

No need to hand code

6 Real time monitoring and alerting

Faster troubleshooting



Mass Ingestion Streaming

Mass Ingestion Streaming - Overview

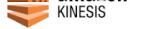
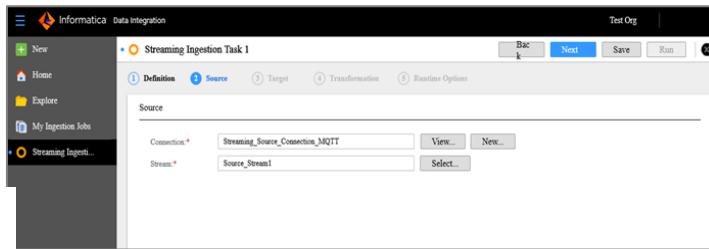
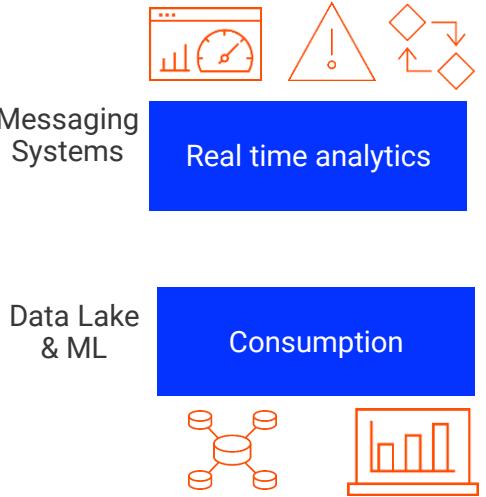
Provides streaming ingestion capabilities as part of IICS Data Ingestion service

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

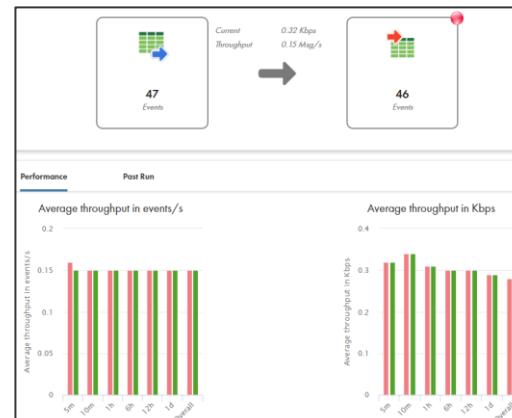
Real-time monitoring of ingestion jobs with lifecycle management and alerting in case of issues

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

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



{ REST API }



Informatica®

Benefits

1 Single ingestion solution for all patterns

Save time and money

2 Wizard driven experience for ingestion

Increase business agility

3 Enable business to ingest streaming data for their usage

Faster decision making

4 Edge transformations for cleansing data

Increased trust in data assets

5 Connectivity to streaming sources & targets

No need to hand code

6 Real time monitoring and alerting

Faster troubleshooting



Mass Ingestion Applications

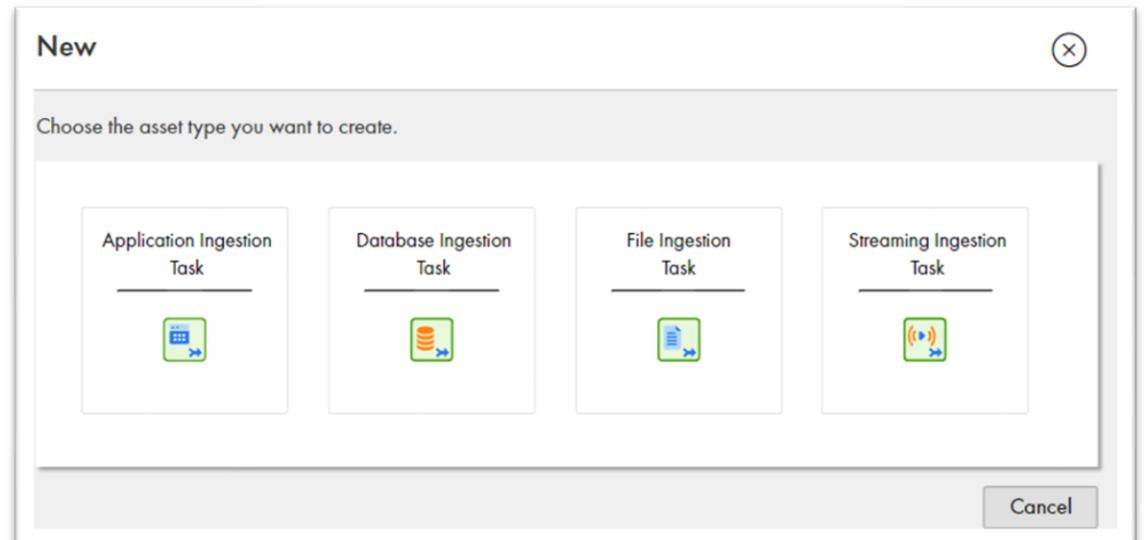
Mass Ingestion Applications

Benefits

- MIA can transfer data from Software-as-a-Service (SaaS) and on-premise applications to cloud-based data warehouses.
- The SaaS and on-premises applications used in your business or organization store large amounts of business-critical data on a daily basis. You can use MIA to transfer the data stored by your applications to cloud-based targets that can handle large volumes of data.
- After you transfer the data to the target, you can consolidate the data and use it for various purposes, such as advanced data analytics and data warehousing.

MIA can perform the following types of load operations:

- *Initial load*
 - Loads source data read at a single point in time to a target.
- *Incremental load*
 - Loads data changes continuously or until the ingestion job is stopped or ends.
- *Initial and Incremental load*
 - Performs an initial load of point-in-time data to the target and then automatically switches to propagating incremental data changes made to the same source objects on a continuous basis

A screenshot of the 'Application Ingestion Task' configuration screen. The screen has tabs at the top: 'Definition' (selected), 'Source', 'Target', and 'Schedule and Runtime Options'. The 'Definition' tab contains fields for 'Name' (Task Name), 'Location' (Default, with a 'Browse' button), 'Runtime Environment' (Select an environment), 'Description' (Description), and 'Load Type' (Select a load type). There are also expandable sections for 'Definition' and 'Advanced Settings'.

DEMO

CMI-Applications



Summary



Cloud native ingestion

- Unified service for ingestion from various sources
- Orchestration for ingestion from variety of patterns

Connectivity

- On-prem Database & CDC
- On-prem & cloud files
- IoT & Streaming
- Cloud data lakes, Datawarehouse and messaging hub

Wizard Driven Design

- Simple easy to use wizard
- Edge transformations
- Intent driven ingestion

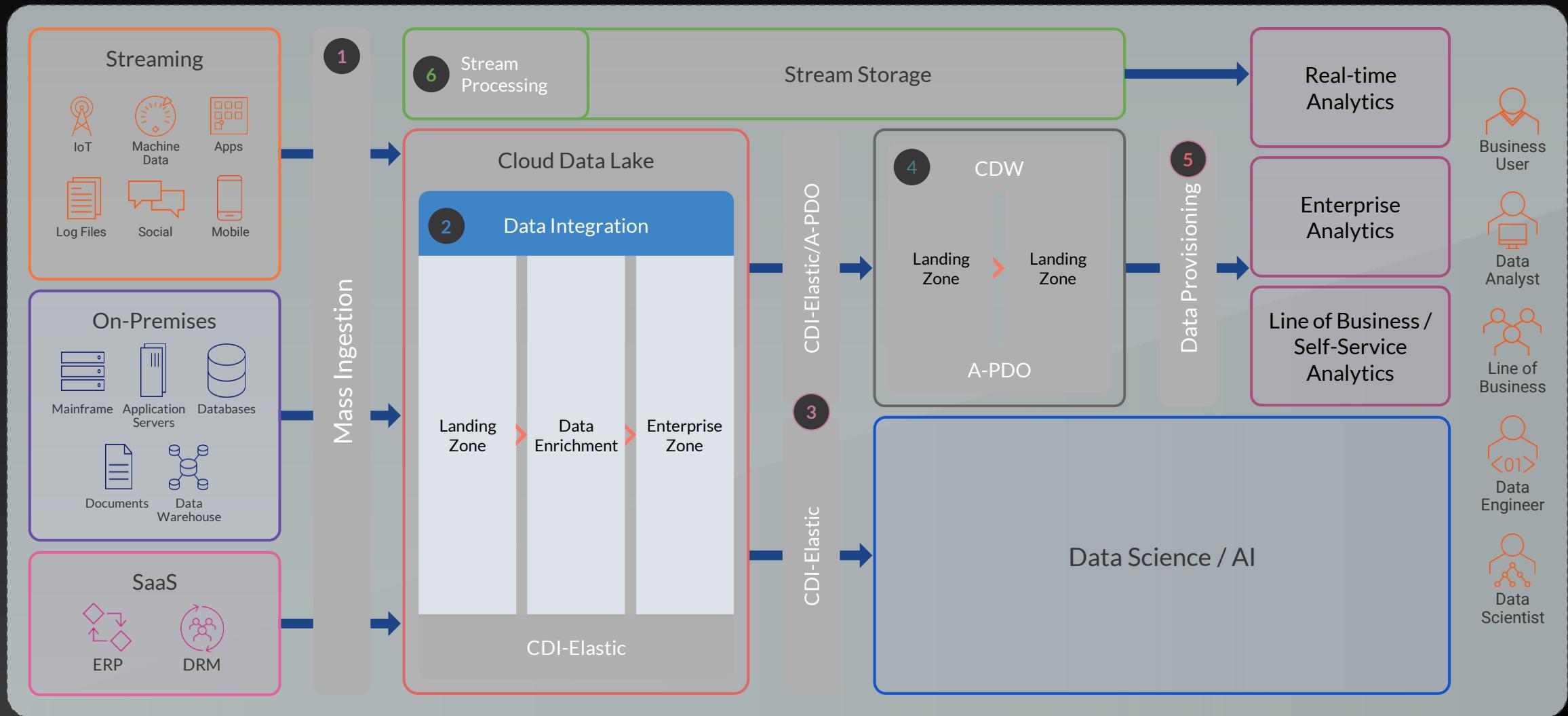
Real-time Monitoring

- Pictorial view of the ingestion job
- Real time flow visualization
- Lifecycle management

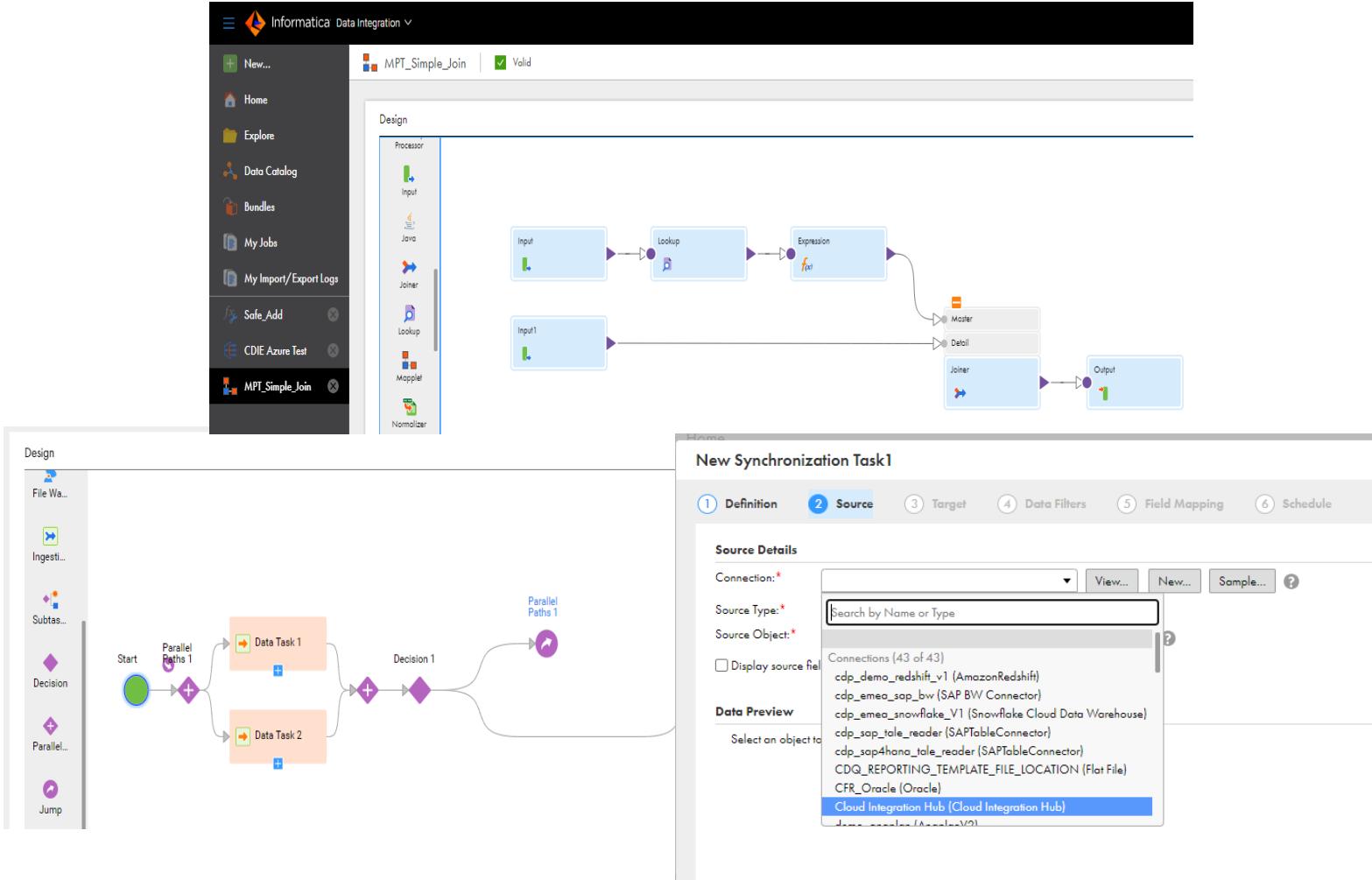


Cloud Data Integration (CDI)

Informatica Data Warehouse and DataLake Architecture



Multi Cloud Integrations using CDI



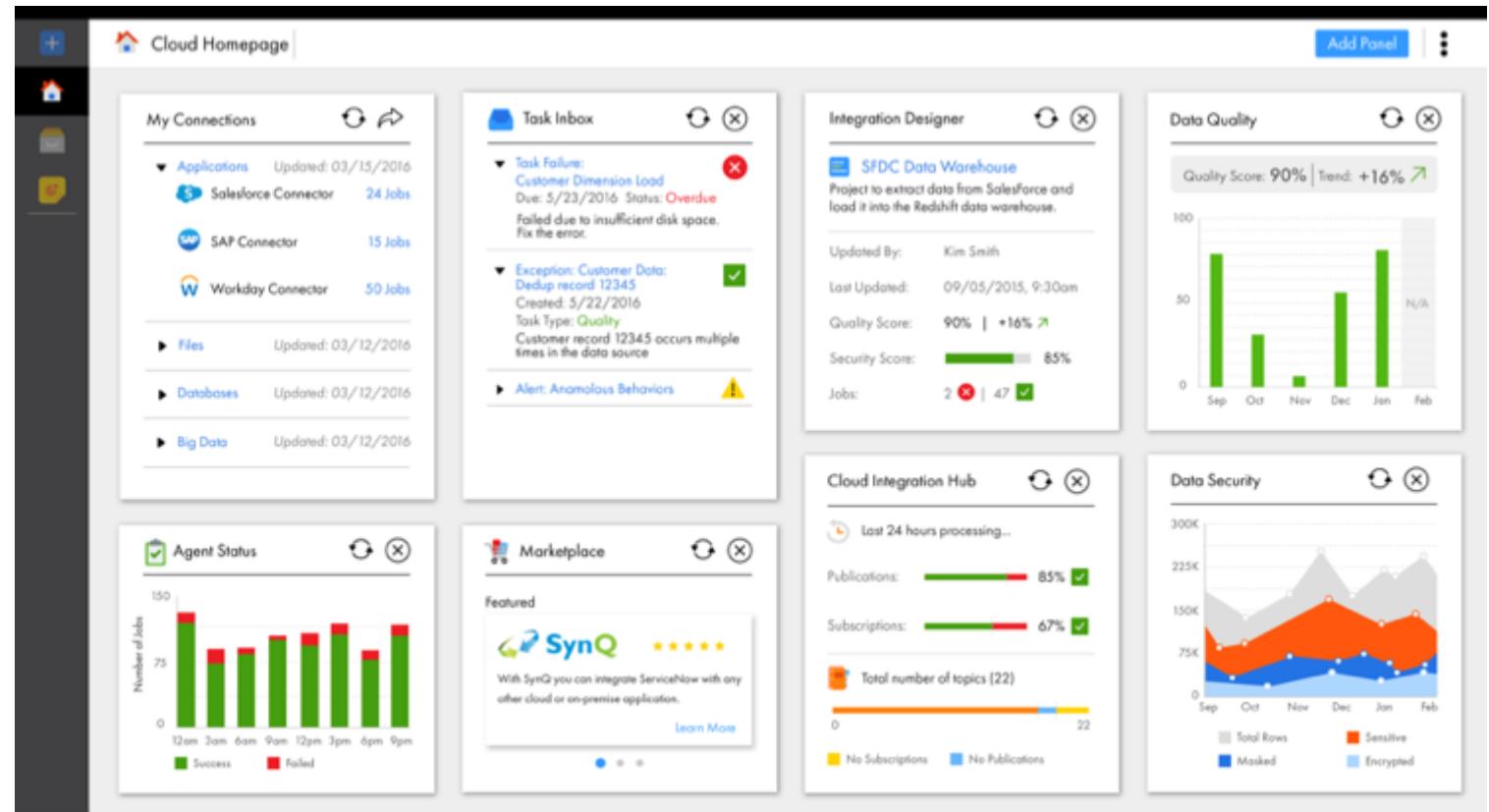
Key Platform Capabilities

- Ease of Use
- Templates and Wizards
- Micro-service Architecture
- Reusability
- Broad Hybrid and Multi-Cloud Connectivity
- No coding across the platform
- Performance optimizations like CDC, parallel processing, pushdown optimization, Mass Ingestion, etc

Hybrid, Multi-Cloud integrations using CDI Transformations and Patterns

With modern role-based unified experience

- Uniform front-end for cloud services
- Role-based, easy access, individualized “Home Page”
- Integrated access to Marketplace, Community and guided tutorials



Unified experience across all cloud services

Integration Task Wizards for Citizen Integrators

Edit DSS_sfdc_ora

① Definition ② Source ③ Target ④ Data Filters ⑤ Field Mapping ⑥ Schedule

Add Maplet... Refresh Fields

| Source: Account | | Target: EJS_ACCOUNTS | |
|-----------------|--------------------------|----------------------|-------------------|
| Status | Name | Status | Name |
| ✓ | Account ID | ✓ | ID |
| | Deleted | | NAME |
| | Master Record ID | | TYPE |
| ✓ | Account Name | ... | BILLINGSTREET |
| ✓ | Account Type | ... | BILLINGCITY |
| | Parent Account ID | ... | BILLINGSTATE |
| ✓ | Billing Street | ... | BILLINGPOSTALCODE |
| ✓ | Billing City | ✓ | BILLINGCOUNTRY |
| ✓ | Billing State/Province | ✓ | PHONE |
| ✓ | Billing Zip/Postal Code | ✓ | FAX |
| ✓ | Billing Country | ✓ | ACCOUNTNUMBER |
| | Billing Latitude | | |
| | Billing Longitude | | |
| | Billing Geocode Accuracy | | |

Actions Expression/Lookup

fx Id

fx Name

fx Type

fx BillingStreet

fx BillingCity

fx BillingState

fx BillingPostalCode

fx BillingCountry

fx Phone

fx Fax

fx concat(AccountNumber, Id)

?

Save

< Back

Next >

Finish

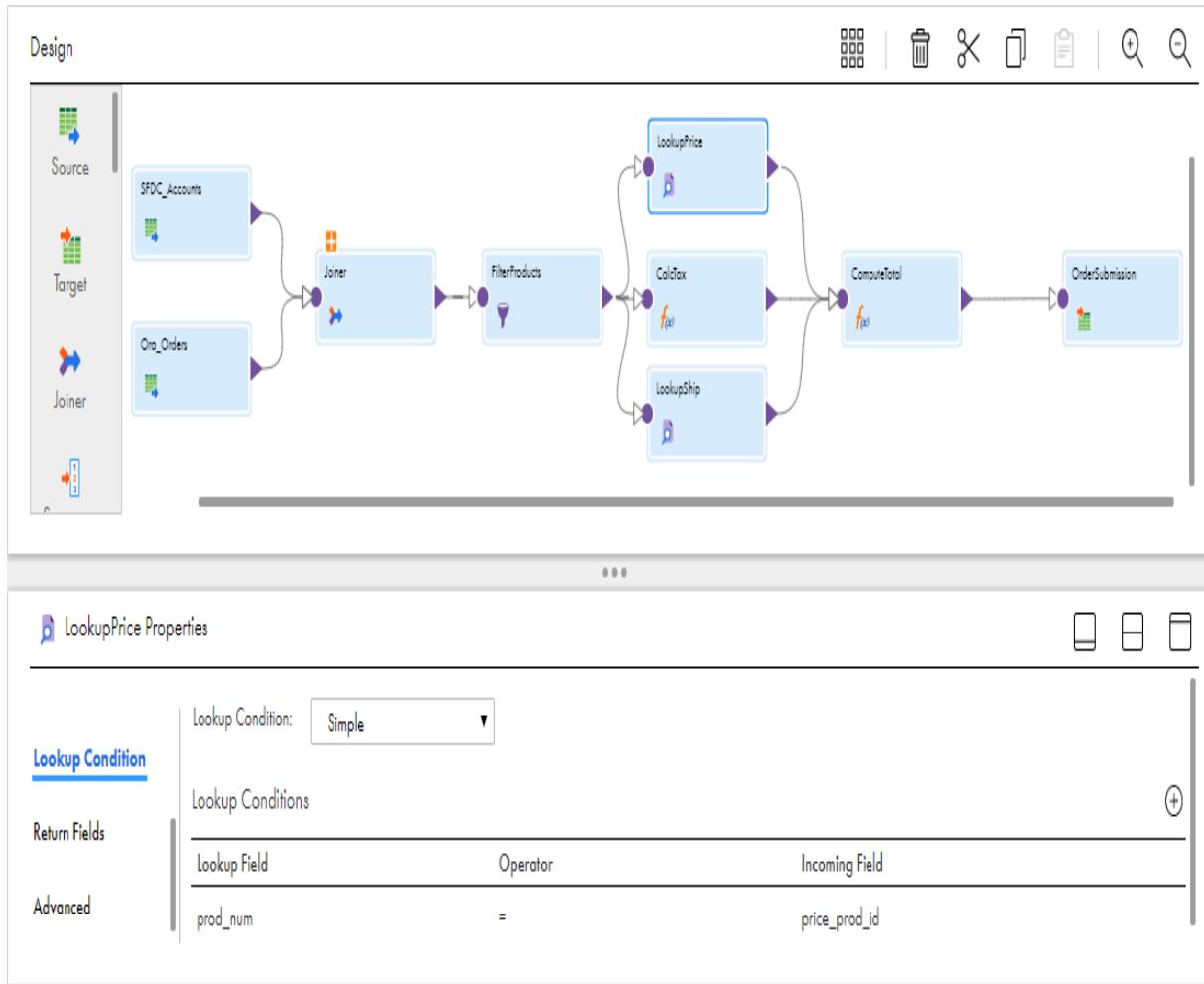
Cancel

DEMO

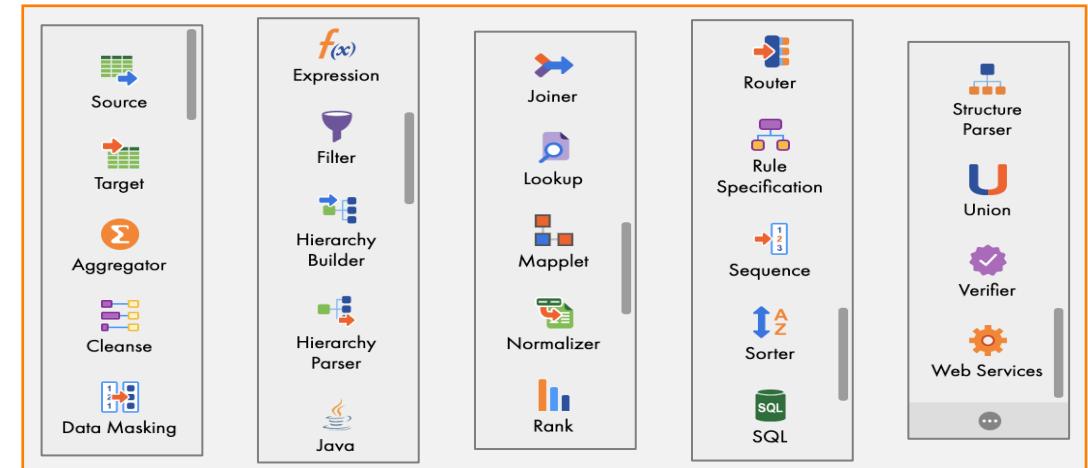
Synchronization



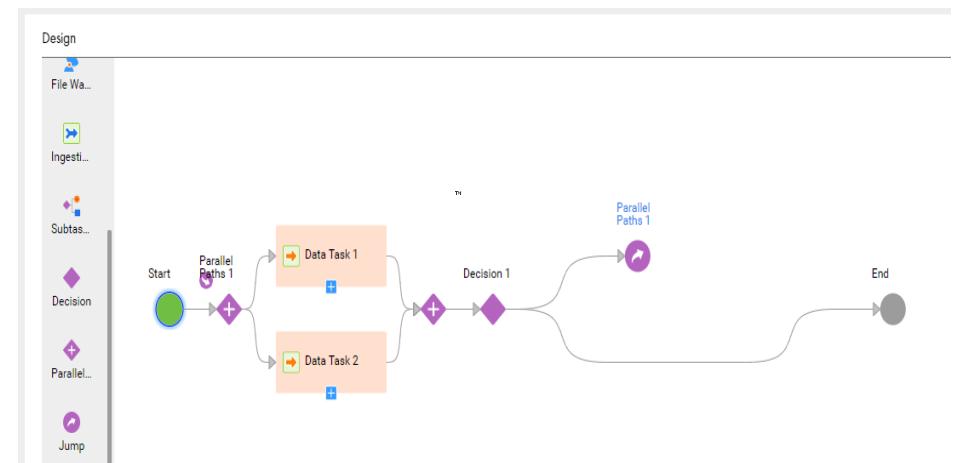
Cloud Mapping Designer for Integration Experts



Transformations



Task Flows



Intelligent Structure Discovery

The screenshot shows the Informatica Data Integration interface with the following details:

- Schema/Sample File:** multisheet_exampl.xlsx
- Discover Structure** button is highlighted.
- Display:** Orders
- Visual Model:** A hierarchical diagram showing the structure of the Excel file. It includes nodes for **Excel**, **Orders**, **Attributes**, **Table**, **unparsed1**, **unparsed2**, **sheet_index**, **sheet_name**, **organization1**, **Person D...**, **Products**, **table**, **element2**, **sheet_index**, **sheet_name**, **organization2**, **element3**, **Address_value1**, **Address_value2**, and **Orders_value1**, **Orders_value2**.
- Relational Output:** A table showing the mapping of discovered structures to relational output groups A and B.

| Display by: | Output groups |
|---------------|---------------|
| organization1 | A |
| ACME | B |
| ACME_llc | Address_value |
| Orders_value1 | Orders_value2 |
| organization2 | |

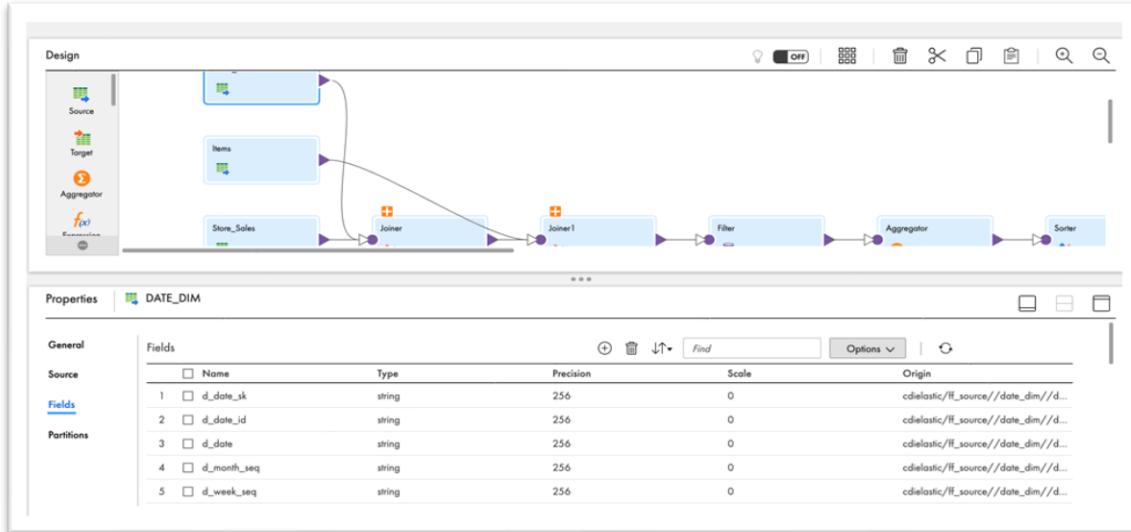
The **Orders_value1** row contains JSON data:

```
[{"O_ORDERKEY": 4416740, "O_CU...": {"L_orderkey": 4416740, "L_partkey": ...}, {"O_ORDERKEY": 4416741, "O_CU...": {"L_orderkey": 4416741, "L_partkey": ...}, {"O_ORDERKEY": 4416742, "O_CU...": {"L_orderkey": 4416742, "L_partkey": ...}, {"O_ORDERKEY": 4416743, "O_CU...": {"L_orderkey": 4416743, "L_partkey": ...}, {"O_ORDERKEY": 4416744, "O_CU...": {"L_orderkey": 4416744, "L_partkey": ...}}
```



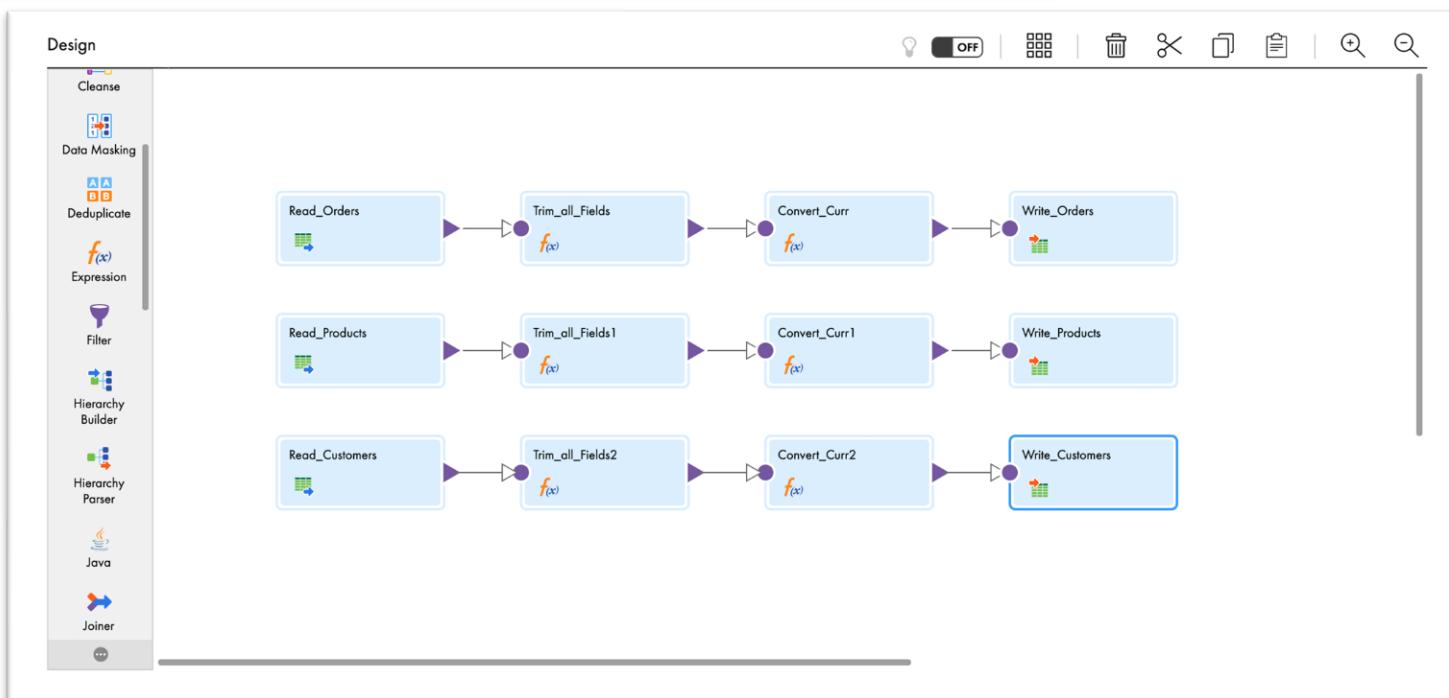
Dynamic Mappings in II CS

Mapping Challenges



Mappings are tightly bound to schemas

Change in metadata (data type, column, etc.) may involve manual changes to 100s of transformations and mappings



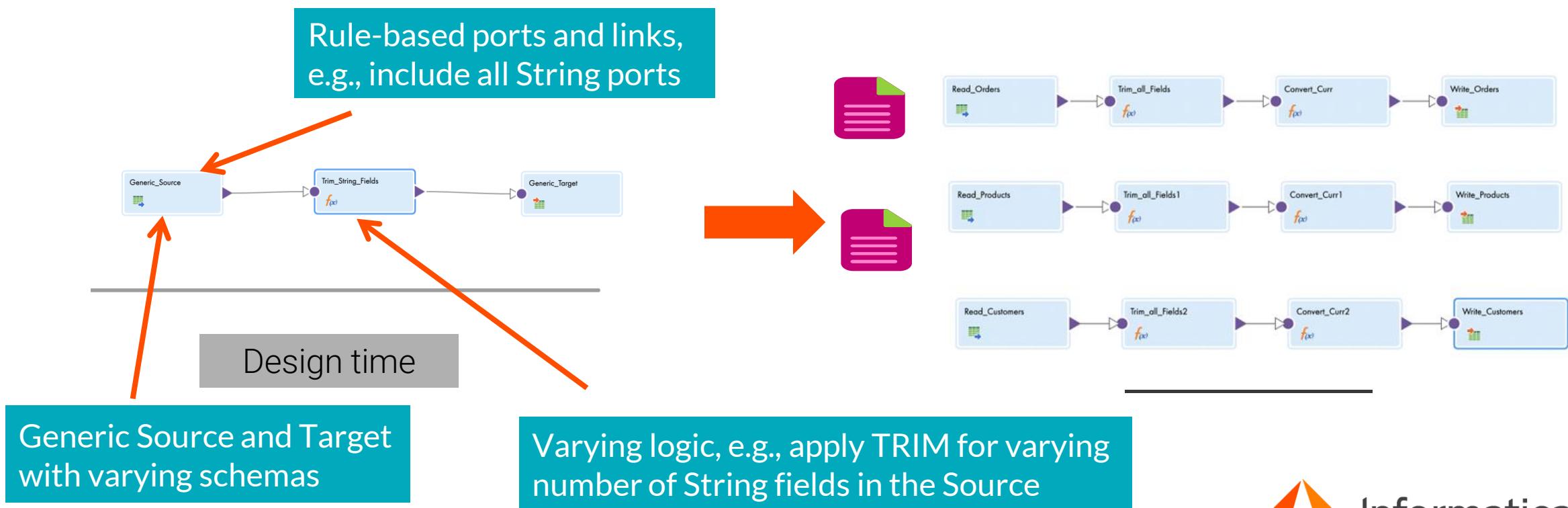
Multiple mappings/workflows are created, tested, maintained for each source

Dynamic Mapping – Goals

- **Support Any Data Integration Pattern**
 - Give customers the ability to develop a highly parameterized mapping
- **Schema Drift**
 - Use one mapping to support multiple file formats
 - Discover the schema at run-time
- **Simplify Maintenance**
 - Turn hundreds of mappings into 1
 - Support table changes without changing the related mappings

Efficiency & Flexibility with Dynamic Mapping

- **Data Integration:** Build a template once – automate mapping execution for 1000's of sources with different schemas automatically
- **Mapping *self-adjusts dynamically*** to external schema changes and column characteristics



Dynamic Mapping – Features

- Major Enhancements
 - Parameterization
 - Dynamic Schema
 - Dynamic Target Creation
 - Concurrent Mapping Execution
 - Dynamic Expressions
- Other Enhancements
 - Dynamic Ports
 - Dynamic Transformations
 - Run-time Linking

DEMO

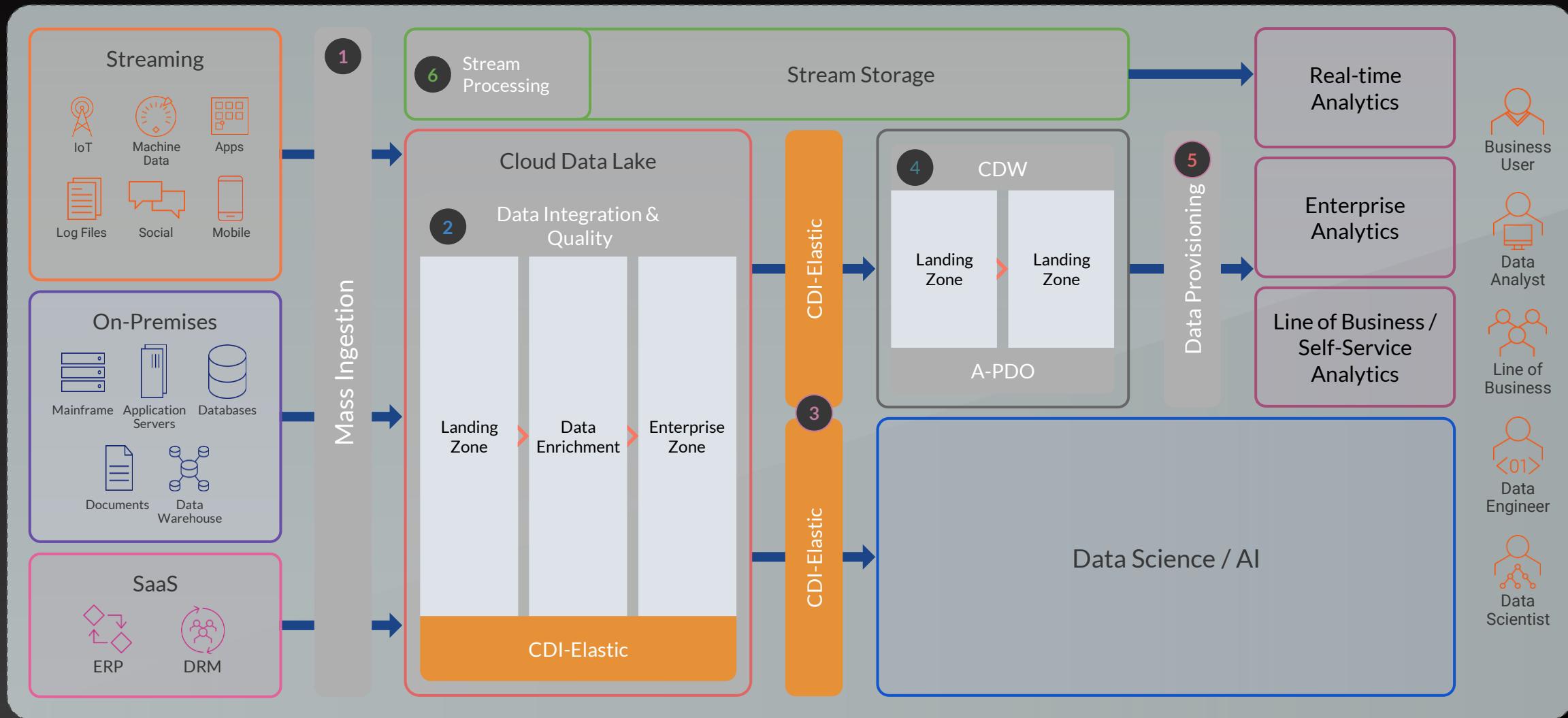
Mappings





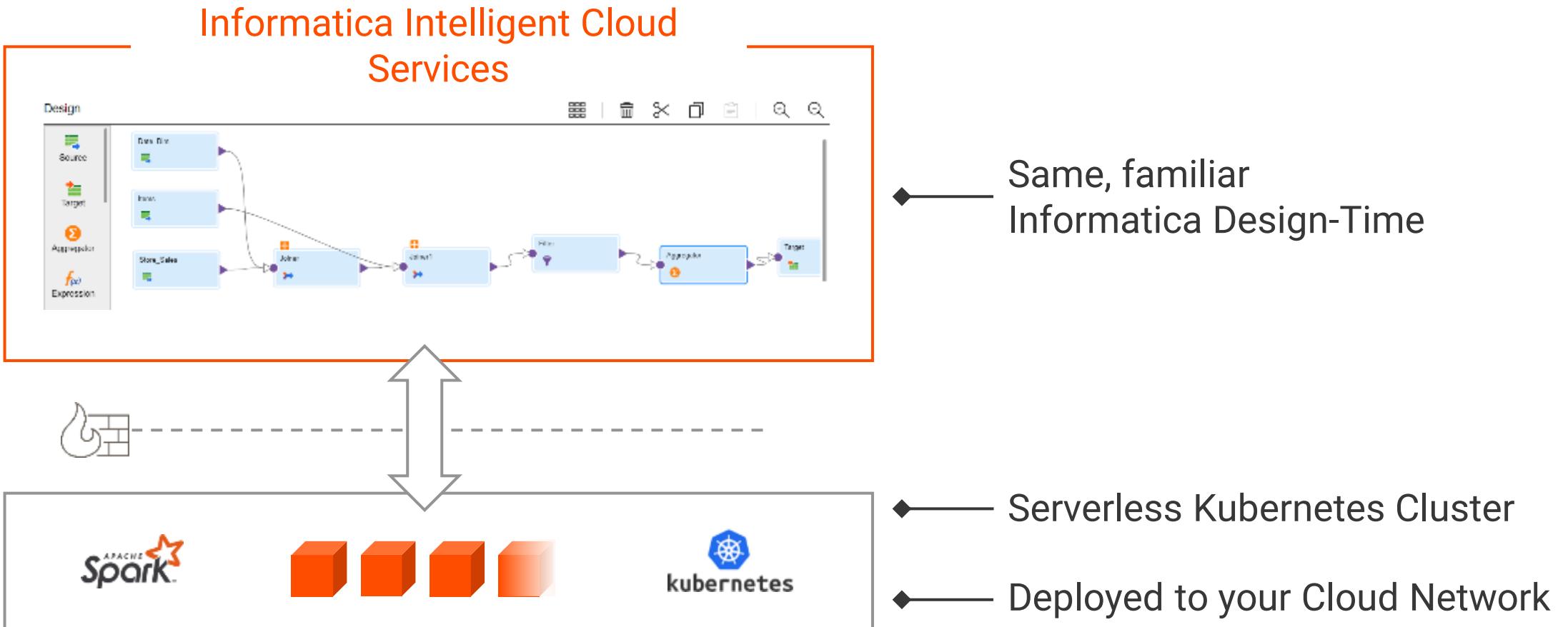
Cloud Data Integration Elastic

Informatica Data Warehouse and DataLake Architecture

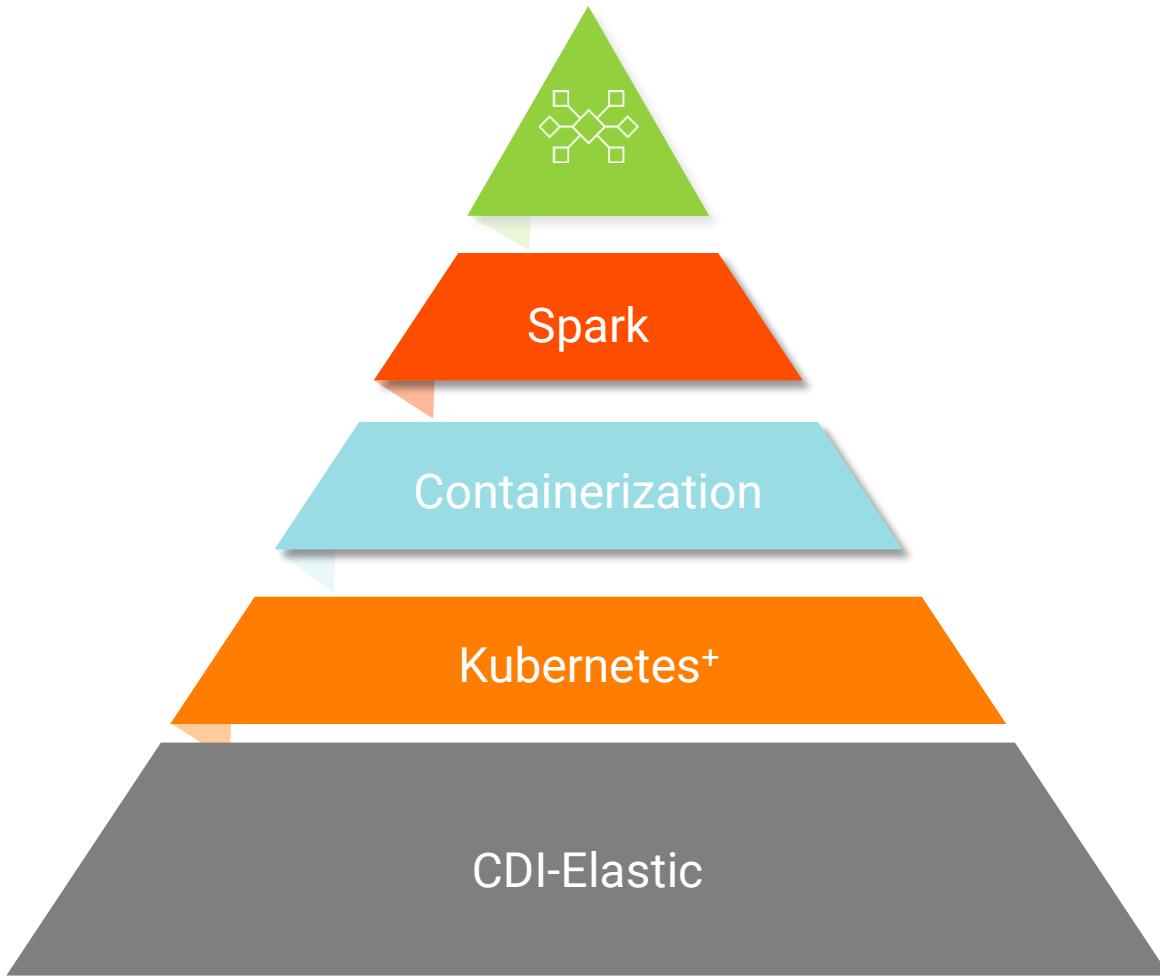


CDI-Elastic

Enabling Kubernetes for auto-scaling and provisioning

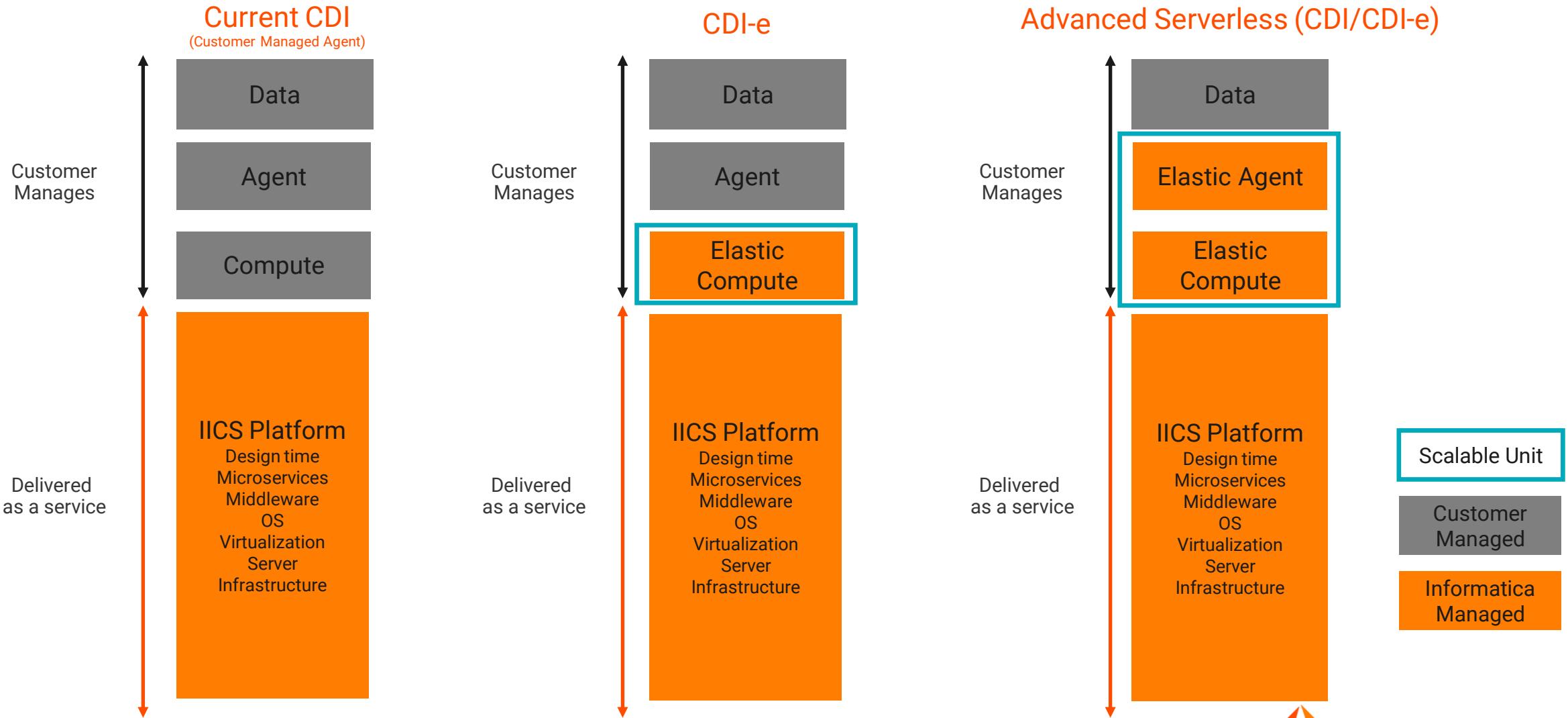


Architecture



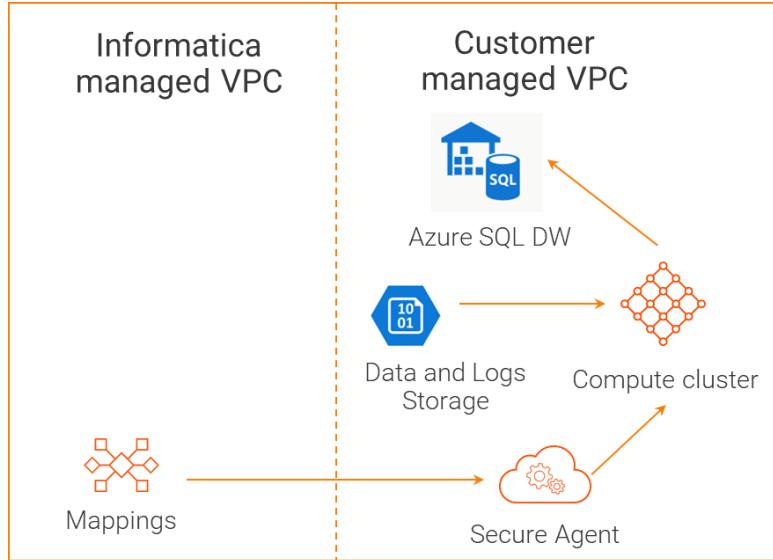
- IICS-based Spark serverless solution
 - Cutting edge technology
 - Open source and best-in-class
 - Built on the cloud, built for the future
 - Lower the overall TCO for customers with Claire-based auto scaling and provisioning
 - Informatica will manage the compute cluster
 - Vendor neutral architecture
 - Ready for multi-cloud from the get-go
- + – Kubernetes based orchestration for Serverless

Standard vs. Advanced Serverless

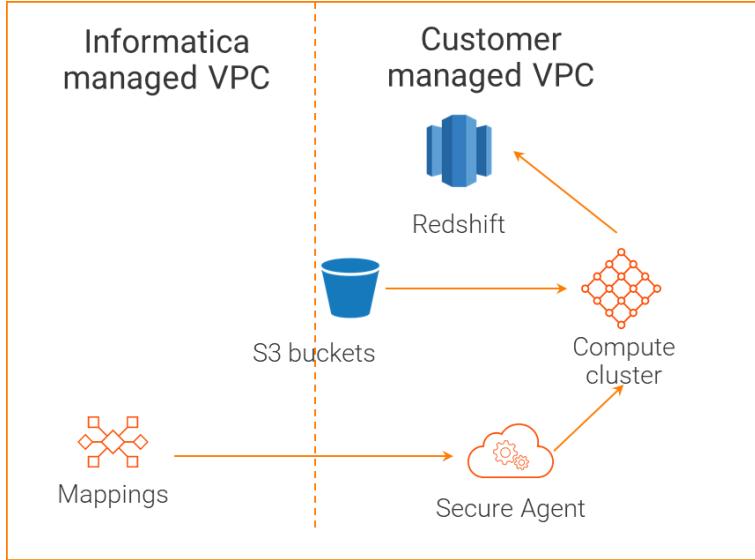


CDI-Elastic Deployment Options

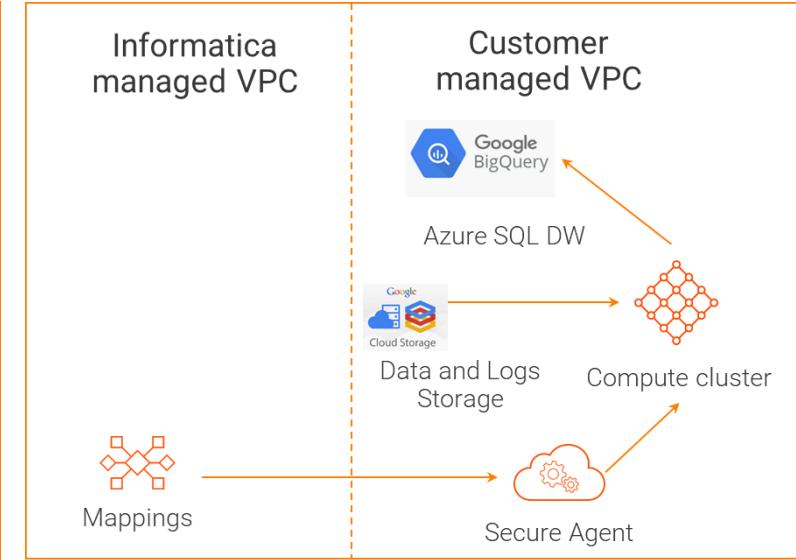
Azure



AWS

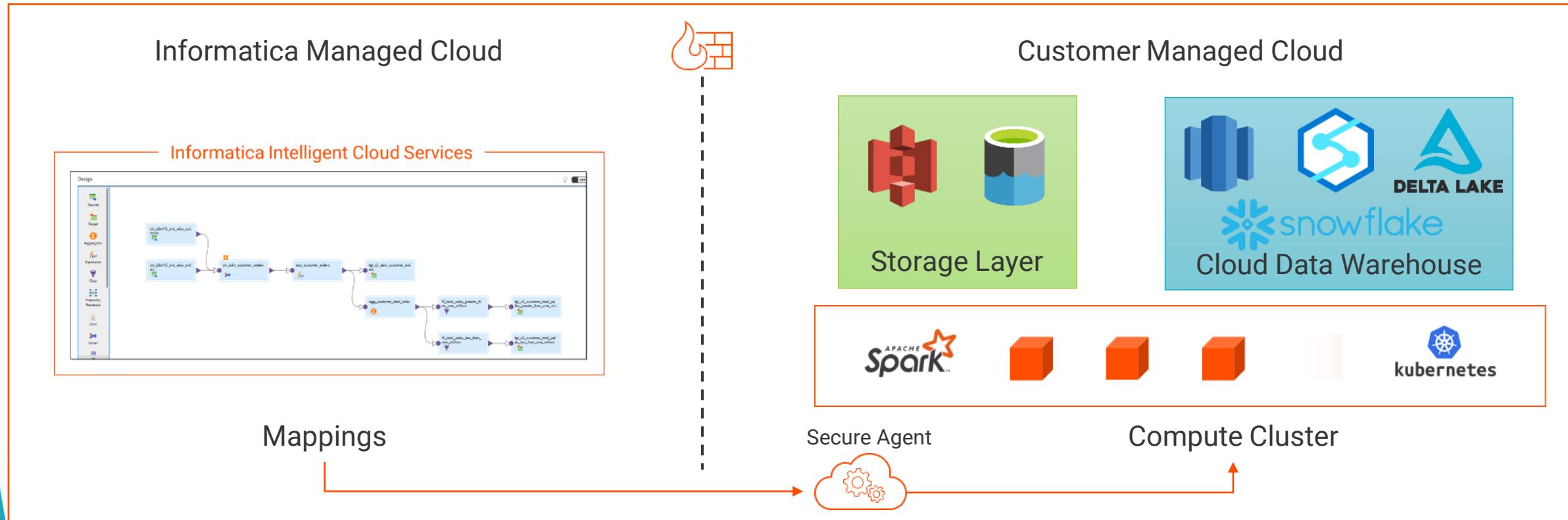


GCP

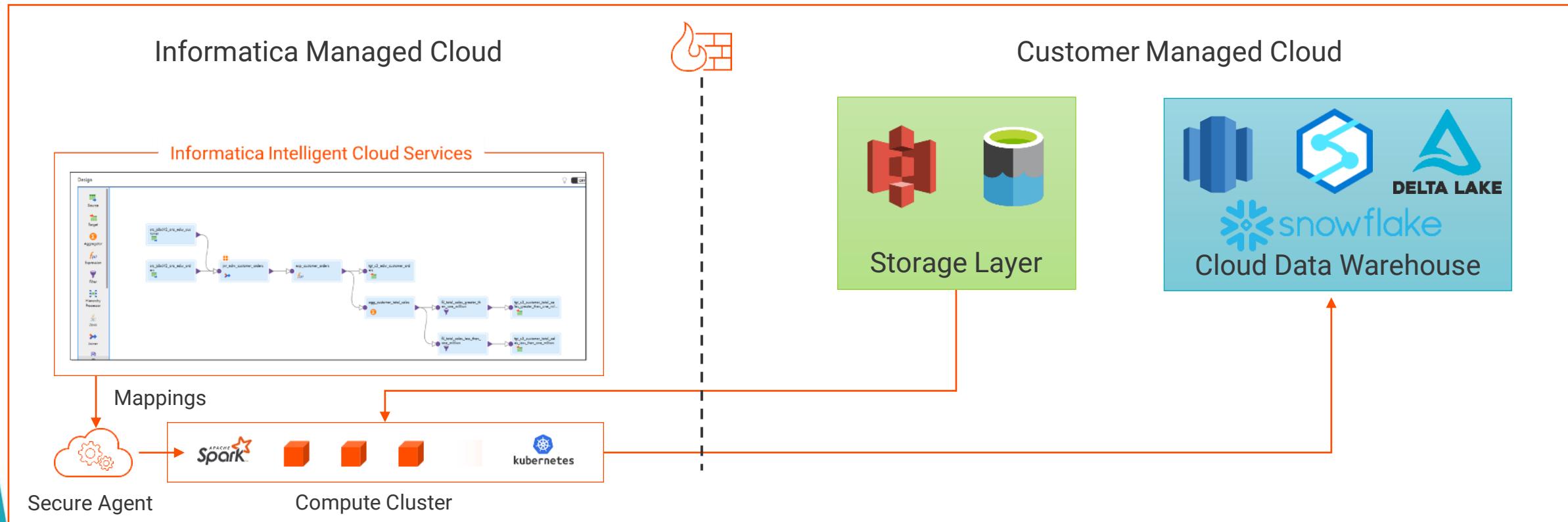


- Compute cluster is launched by Secure Agent in the customer network
- Customer has complete control on network peering, assigning roles and privileges

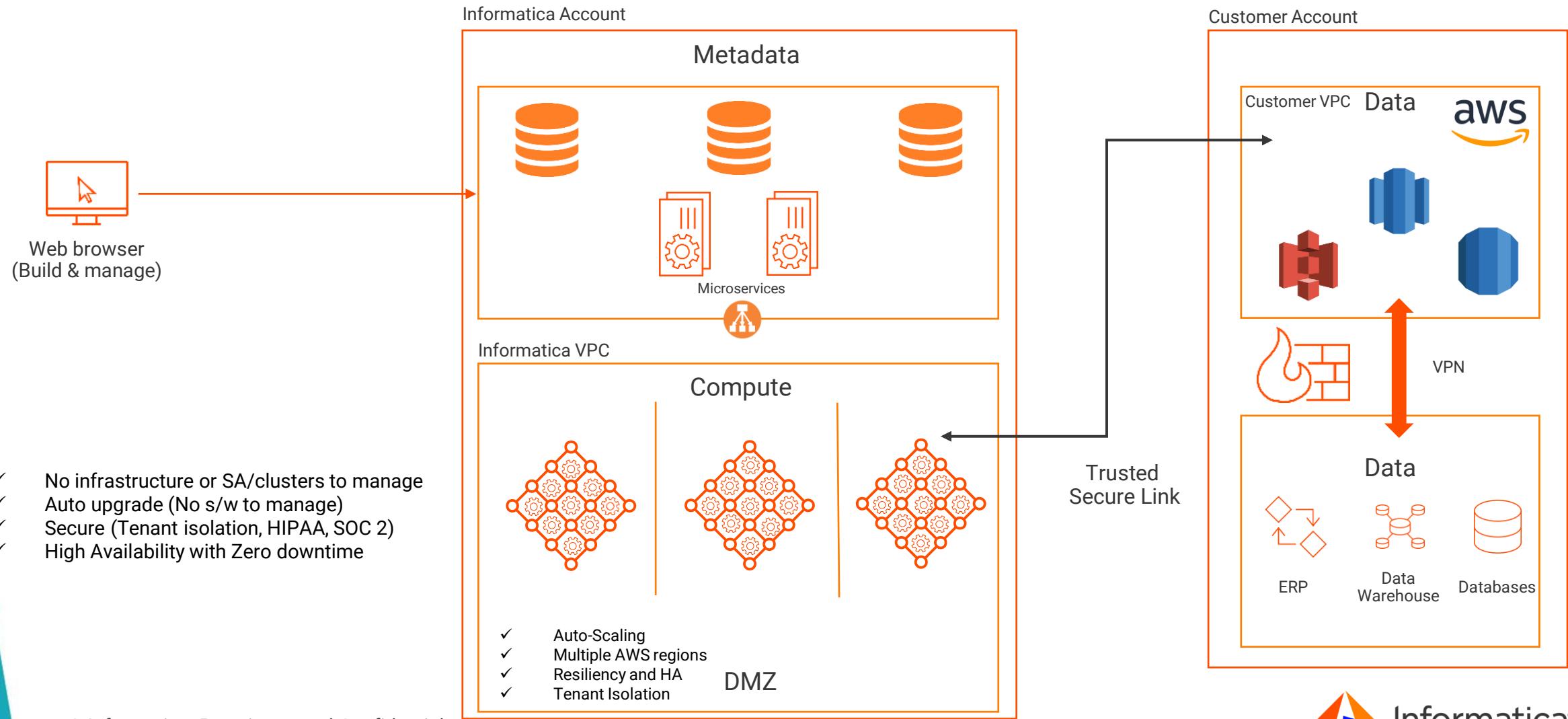
CDI-E Standard Serverless Architecture



CDI-E Advanced Serverless Architecture



CDI-E Advanced Serverless Architecture: Details



CDI-e: Automated Performance Tuning

Powered by CLAIRE

The screenshot illustrates the Informatica CLAIRE automated performance tuning process. It consists of two main windows:

- Configure Tuning Window:** This window is titled "Configure Tuning" and contains instructions for tuning a mapping task. It includes a text area explaining the tuning process, a configuration input for "Number of Mapping Task Runs" set to 25, and a "Tune" button. Below the input field, there is a diagram showing three data flows: "src_jdbcV2_ora_edw_customer" to "fl_customers_by_market_segment", and "fl_customers_by_market_segment" to "tgt_s3_edw_customer".
- Tuning Result Window:** This window is titled "Tuning Result" and displays the results of the tuning process. It shows a summary of performance improvements and a detailed table of tuning recommendations. The table lists four properties and their recommended values:

| Property Name | Property Value | Tuning Recommendation |
|---------------------------|----------------|-----------------------|
| spark.executor.memory | 2G | ? |
| spark.rdd.compress | false | ? |
| spark.shuffle.file.buffer | 29k | ? |
| spark.executor.cores | 4 | ? |

CDI-e: Why tune?



Manual work

30% of your Engineers time



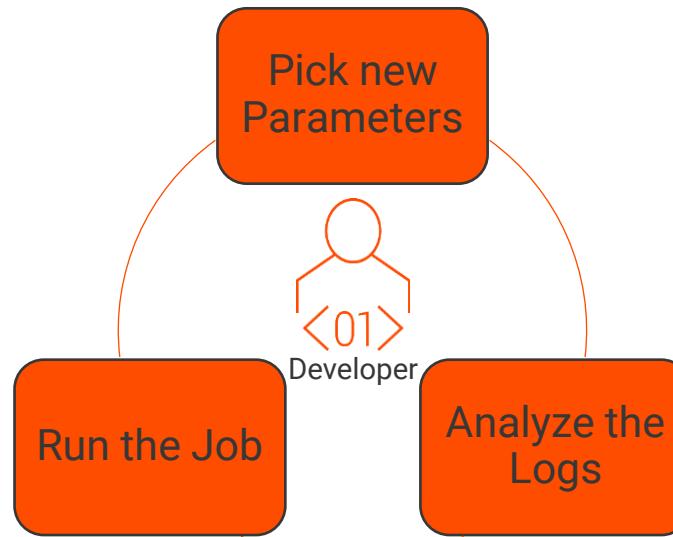
Frequent Outages

Pager ringing at 3 AM



Slow and expensive

Missing SLA's every week.



CDI-e: What is tuned?

Optimal cluster parameters

- Size
- Instance Type
- # of processors
- # of memory
- Disks
- ...

Optimal Spark Configuration

- Parallelism
- Shuffle
- Storage
- JVM Tuning
- Feature Flags
- ...

CDI-e: Auto Scaling

- Auto scaling to meet your SLA at least possible costs
- Dynamically respond to changes to environment and workloads to meet the data volume requirements and compute requirements
 - Algorithm to scaling up/down effectively
 - Auto adjust based on concurrency
 - Horizontal and vertical scaling
- Increase/decrease parallelism by arriving at the optimum number of nodes and spark executors based on the job demands

CDI-e: Incremental File Load

- Challenge:
 - I want to load data (different flat files) into Cloud Storage
 - Files which already have been processed should be ignored.
 - I cannot just delete them, since they are used by other processes as well.



- Solution:
 - CDI-Elastic can track data that has been processed during a previous run of an MCT by persisting the state information of the job run.
 - Incremental File Load is a feature of CDI-Elastic which will maintain the state information and prevent reprocessing of old data.
 - Time travel will help to go back in time and re-process files

DEMO

CDI-e

