

Mar 06, 2020

# Intelligent Data Catalog for IICS, EDP and Axon

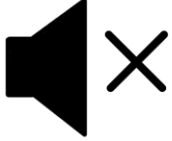
Srinivasa Gopal

*Principal Technologist – Customer Success*



Informatica™

# 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 to view on our **INFASupport YouTube channel** and **Success Portal**. The link will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

# Success Portal

<https://success.informatica.com>

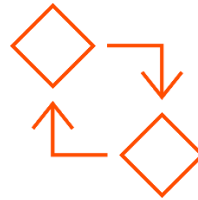
Learn. Adopt. Succeed.



Bootstrap product  
trial experience



Enriched Onboarding  
experience



**FREE** Product  
Learning Paths  
and weekly Expert  
sessions



Informatica  
Concierge with  
Chatbot integrations



Tailored training and  
content  
recommendations

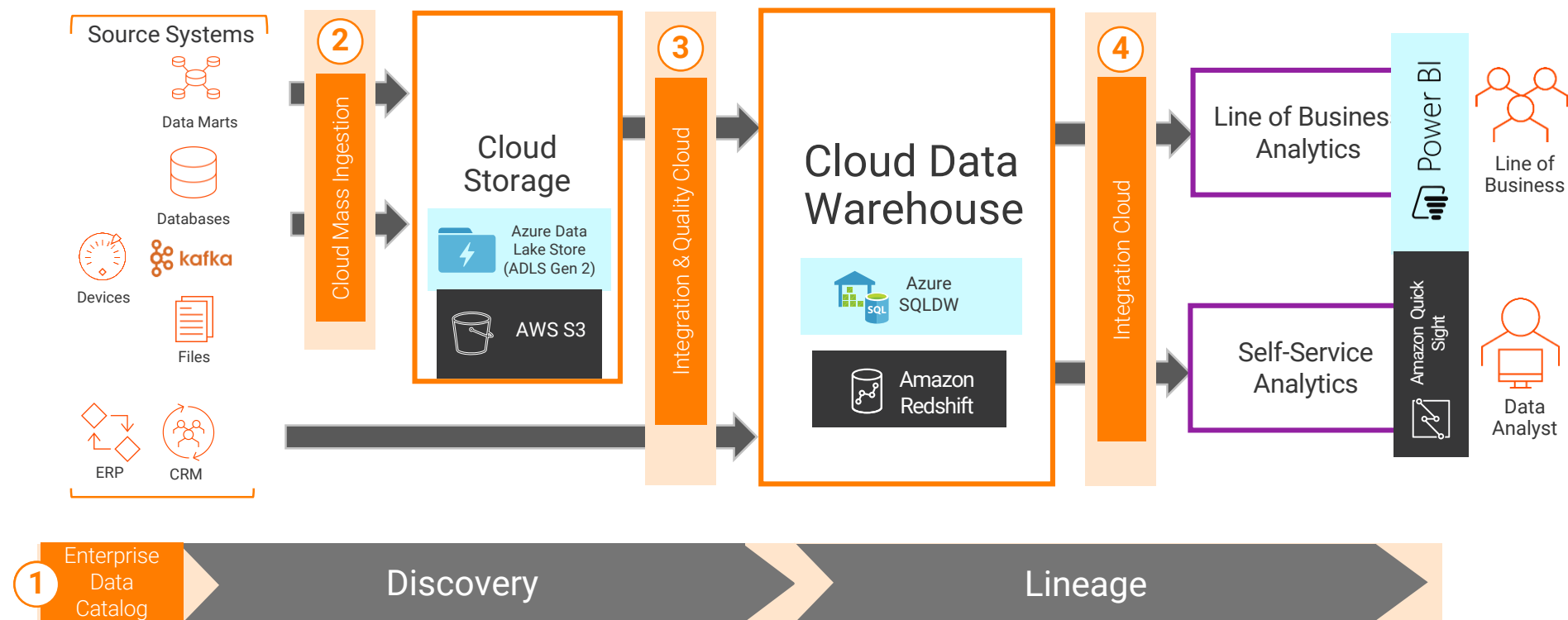
# Agenda

- Cloud Data Integration UseCases
- Enterprise Data Preparation Integration UseCases
- Data Governance Integration UseCases
- Demo
- Q&A

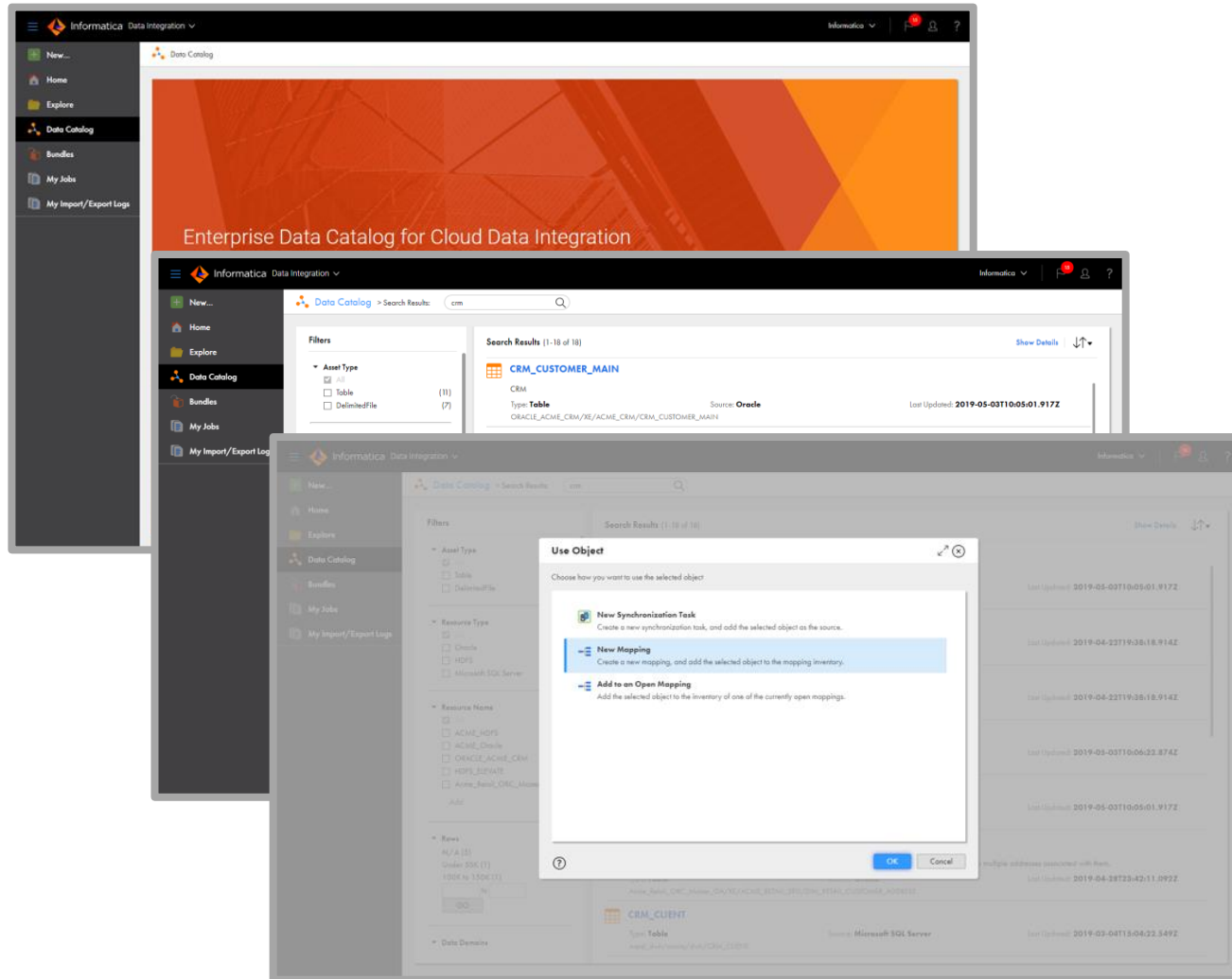


# Data Catalog for Cloud Journey

# Cloud Data Warehouse



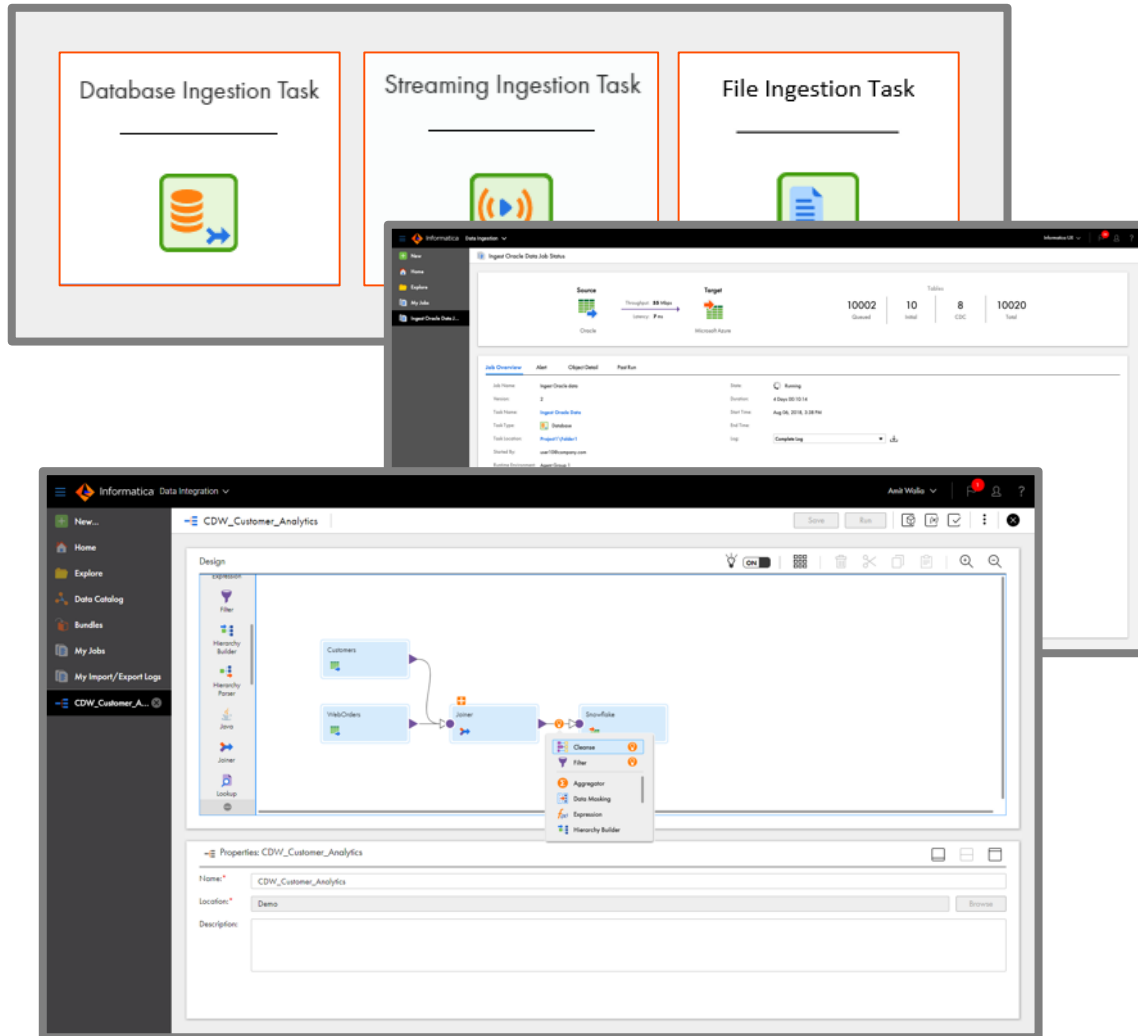
# 1. Find data to ingest



Enterprise Data Catalog – industry's #1 data catalog to find all your enterprise data, detailed data lineage, certified data, social collaboration

Automate data provisioning from the data catalog

## 2. Ingest data to landing zone



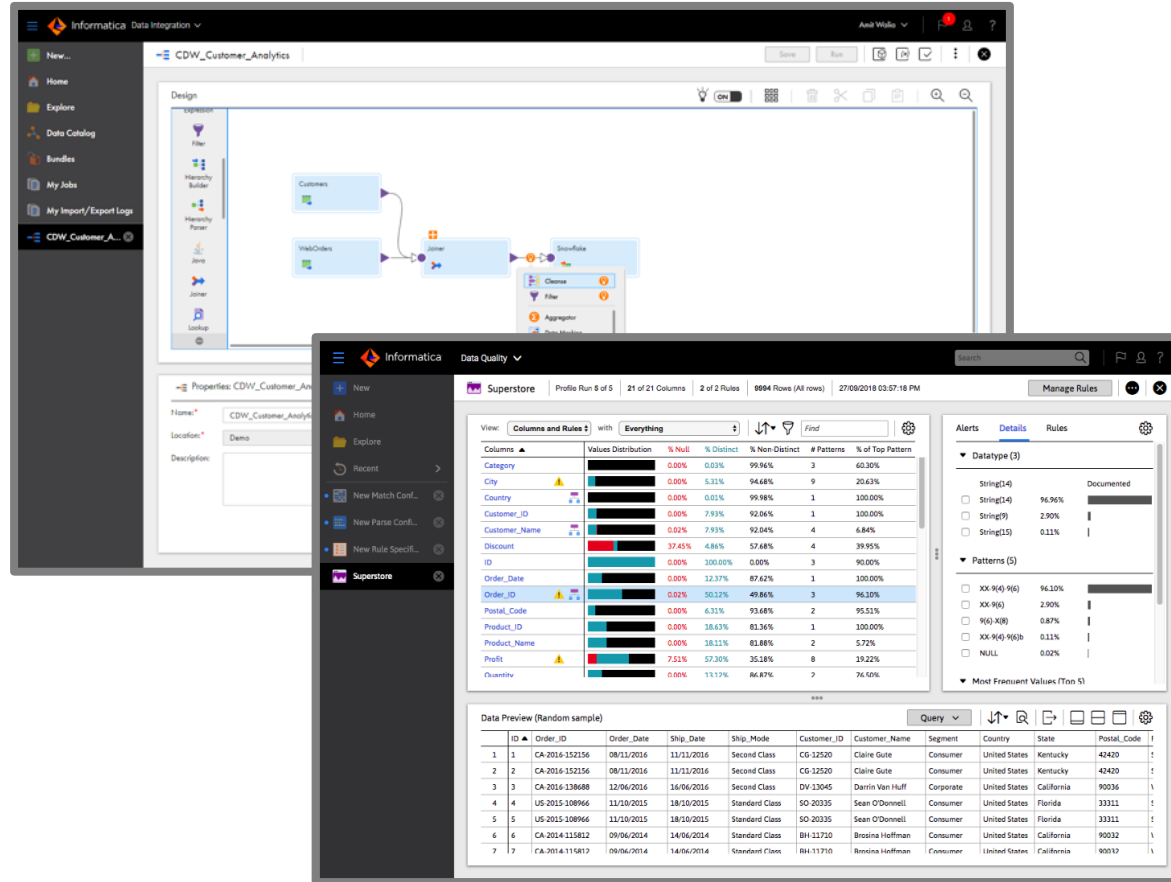
Cloud Mass Ingestion – one tool to quickly do all types of ingestion at scale

Cloud Data Integration – connect to all types of data, easy to use ETL, integrated to the catalog for a shopping cart experience

- Files
- Databases
- CDC / Schema drift
- Real-time streaming
- API



### 3. Integrate and cleanse data into cloud data warehouse

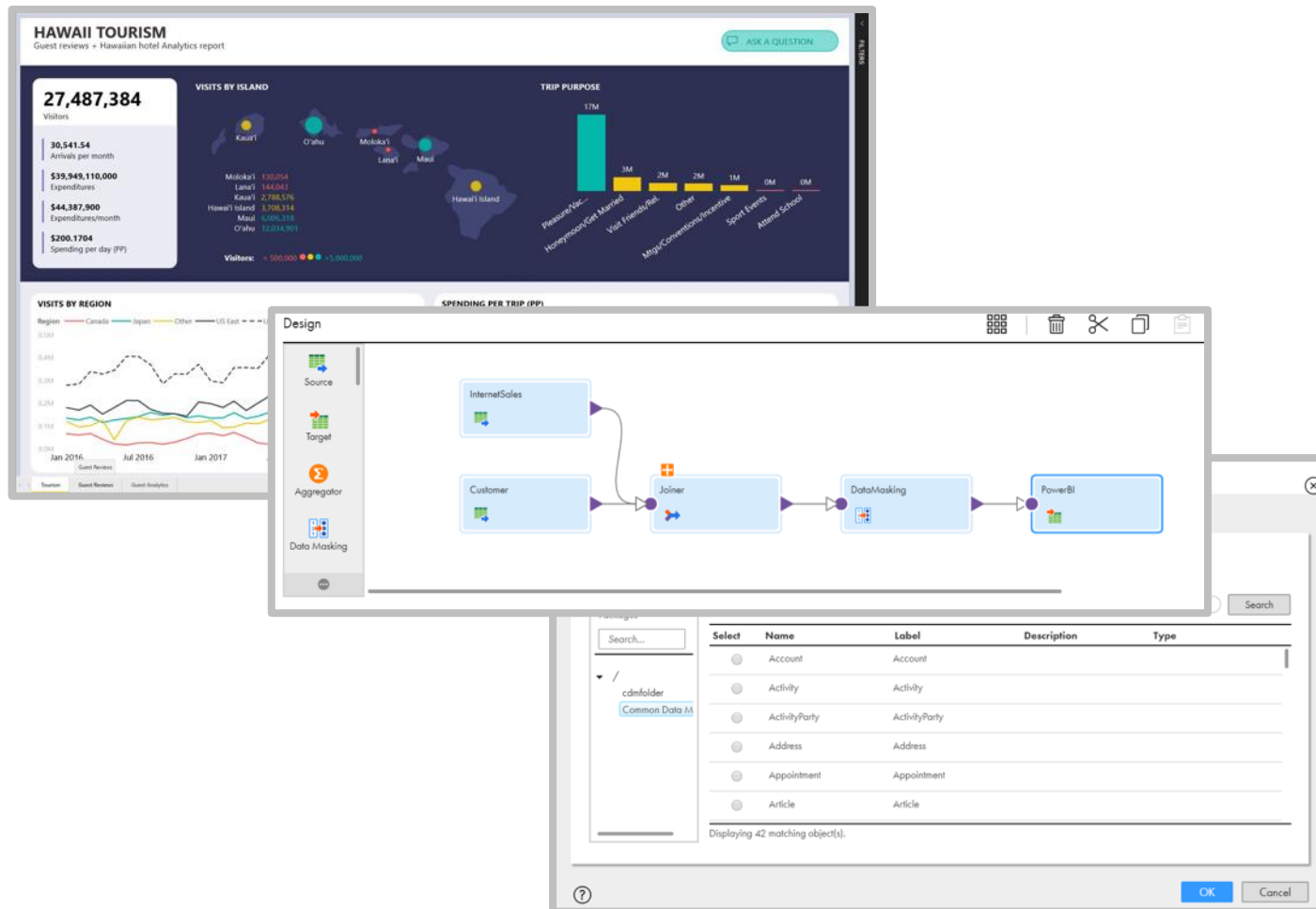


Cloud Data Integration (CDI) – simply and easy-to-use ETL for any source, pushdown optimization to leverage target engine

Cloud Data Integration Elastic (CDIe) – high performance cloud-native large volume processing

Cloud Data Quality – the industry's only cloud-native data profiling and cleansing tool

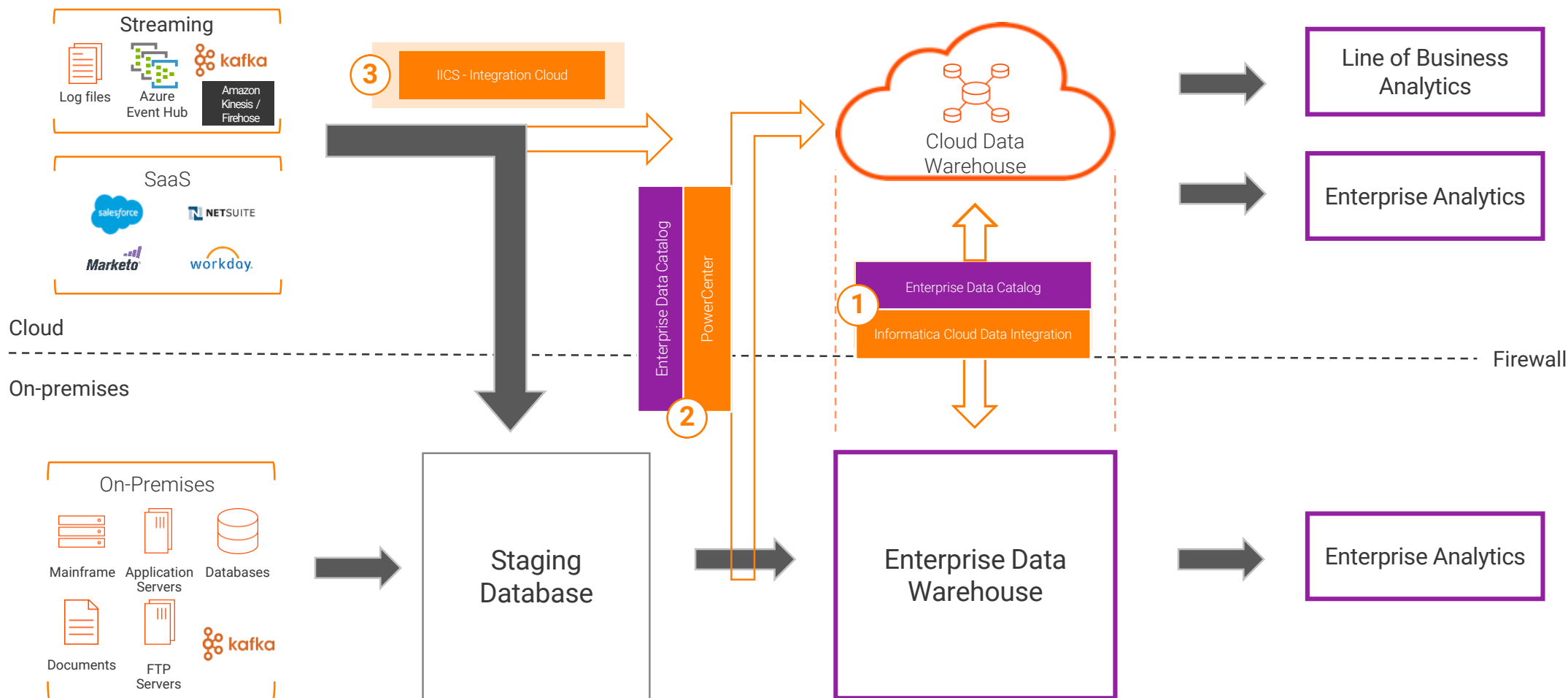
# 4. Self Service BI with embedded Cloud DI and Cataloging



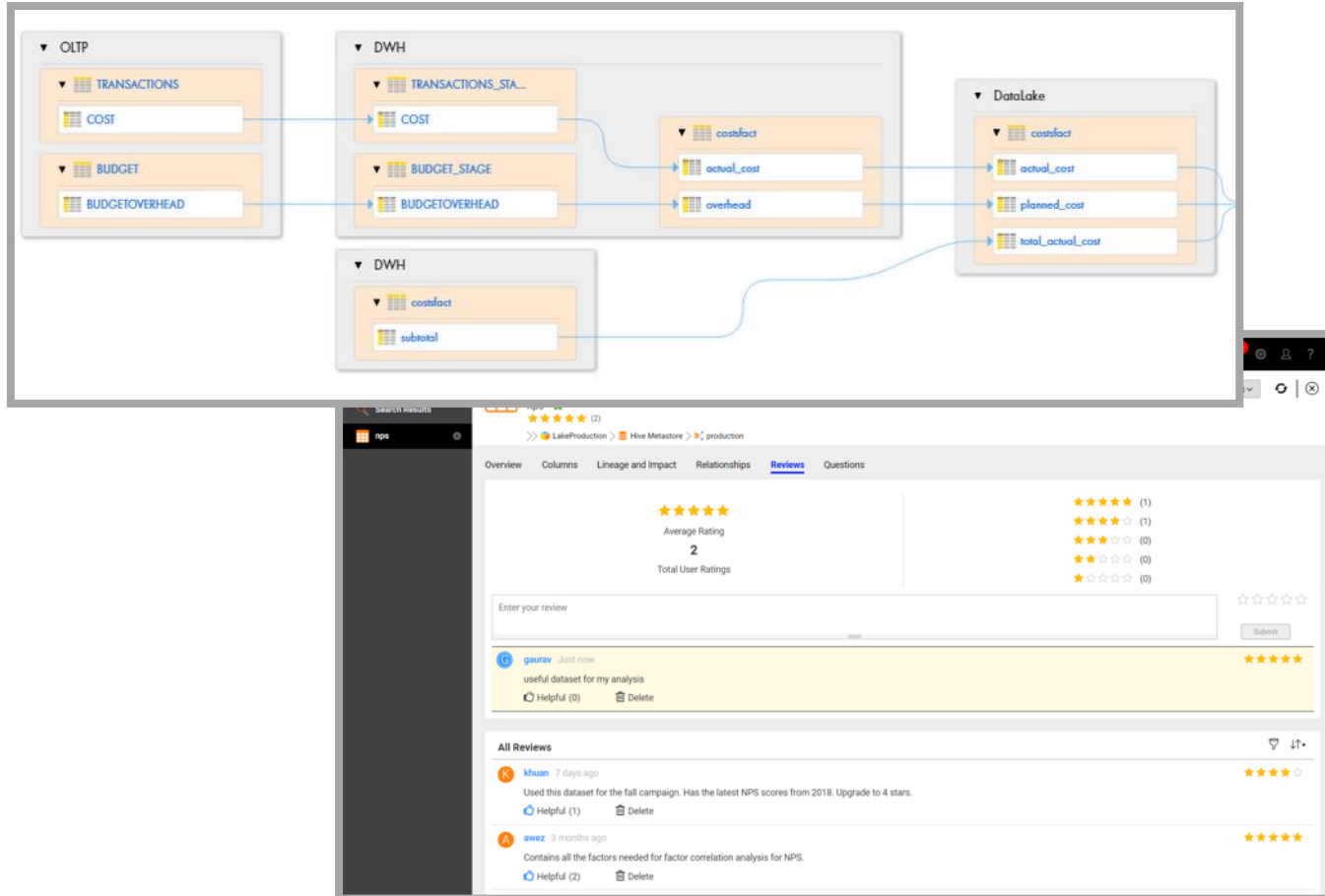
➔ Embedded BI tool capabilities (i.e. Tableau, PowerBI):

- Cloud Data Integration – automated data provisioning
- Enterprise Data Catalog – catalog at your fingertips and detailed lineage to show where data came from

# Cloud Data Warehouse Migration



# 1. Discover and prioritize data that needs to migrate

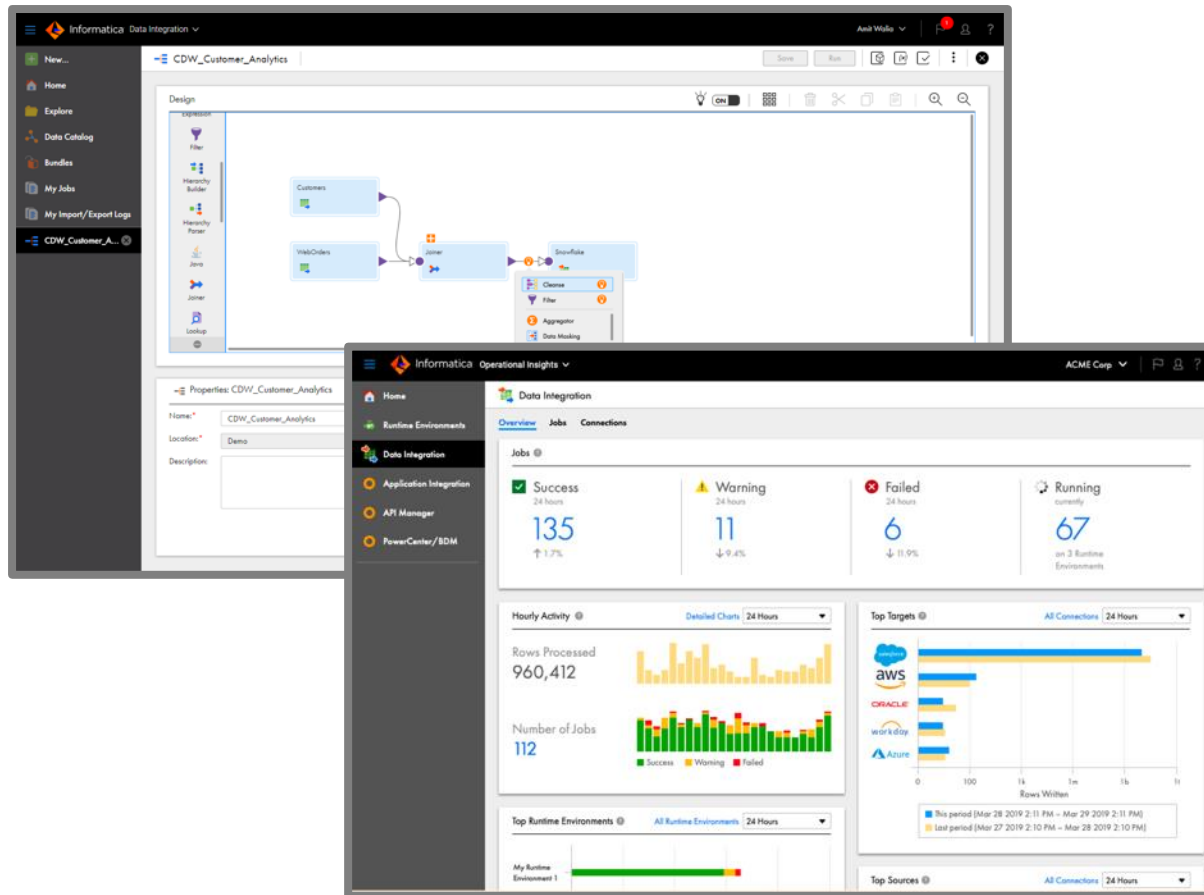


Enterprise Data Catalog – industry's #1 data catalog that provides detailed lineage to determine what data and data pipelines needs to be migrated

Automate data provisioning from the catalog across the source system landscape

View relevance and quality of data before migrating

## 2. Migrate data from on-prem DW to Cloud DW



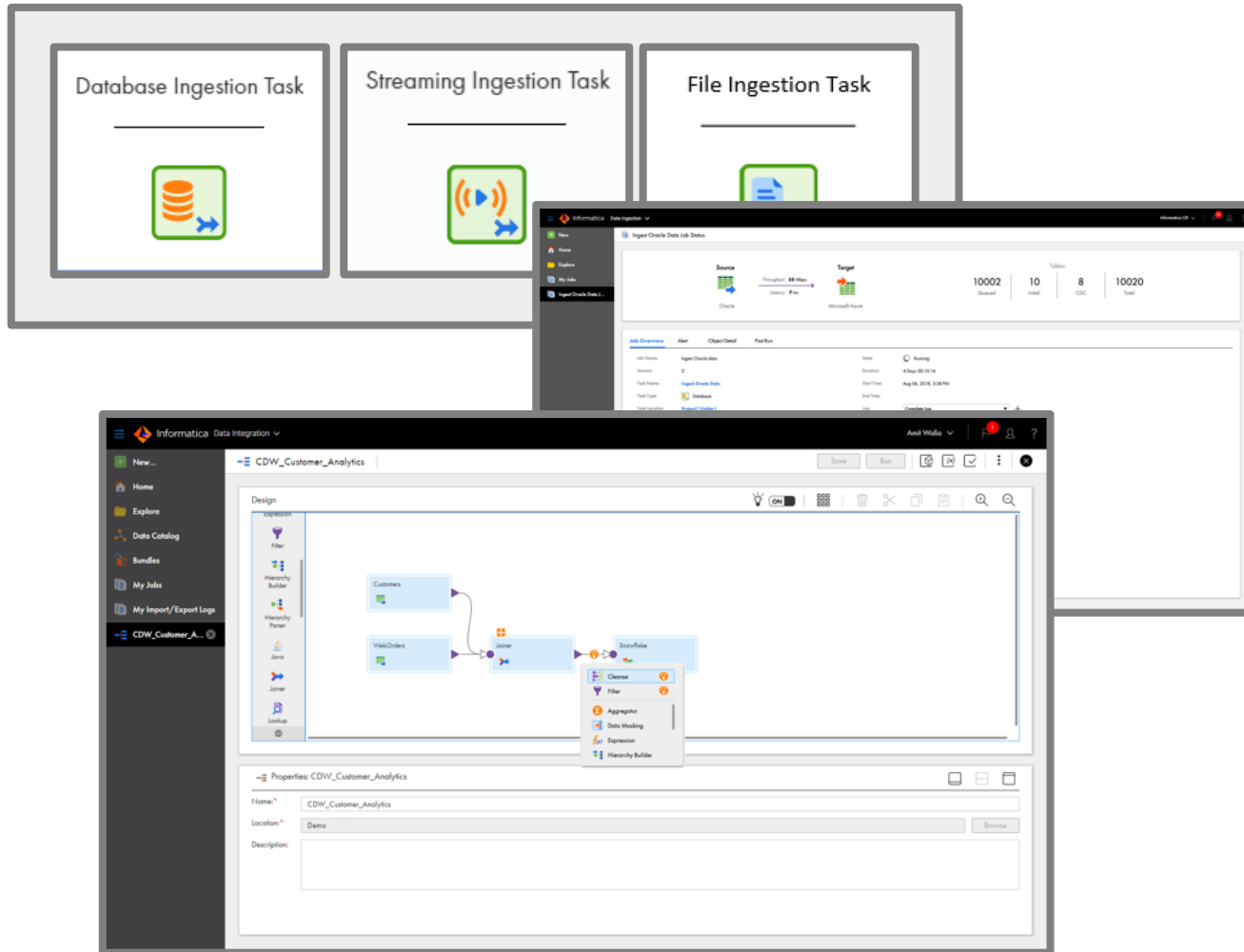
Cloud Data Integration – connect to all types of data, easy to use ETL, integrated to the catalog for a shopping cart experience

### Key connectors

- AWS Redshift
- Azure SQL DW
- Snowflake
- Google Big Query

### Operational Insights

### 3. Ingest data from new sources

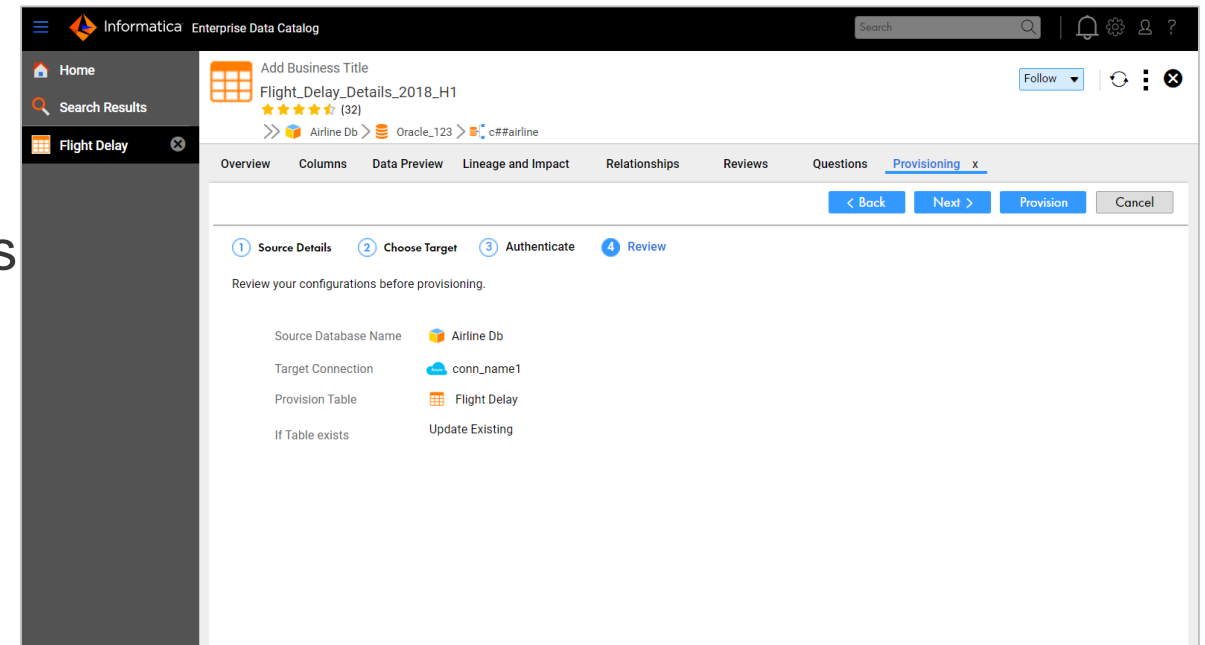
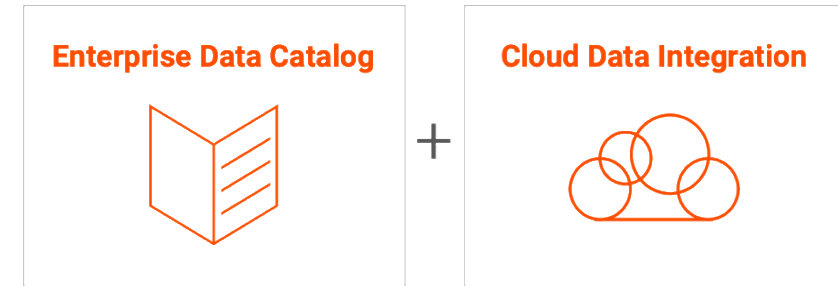


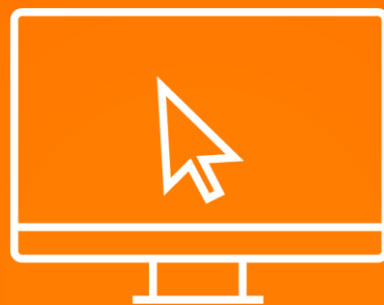
Cloud Mass Ingestion – one tool to quickly do all types of ingestion at scale

Cloud Data Integration – connect to all types of data, easy to use ETL, integrated to the catalog for a shopping cart experience

# Self-Service Data Provisioning

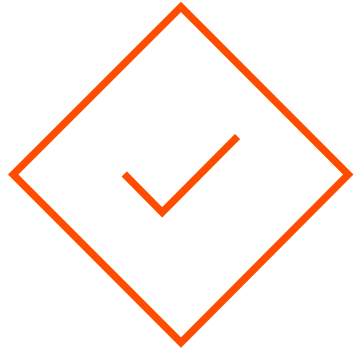
- Empower analysts to access the data they need
- Simple, click-through data provisioning to easily deliver your data to desired target
- Broad choice of sources & targets, including AWS Redshift, Azure SQL DW, Google BigQuery, Snowflake and BI tools like Tableau





Demo





# Data Catalog for Data Preparation

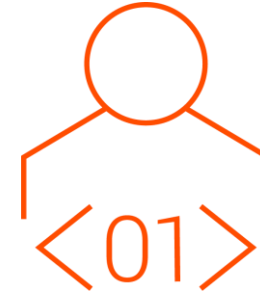
# Typical Data Preparation Challenges



## **Business/Data Analysts**

---

- Difficulty finding trusted data
- Limited access to the data
- Frustrated by slow response from IT
- Constrained by disparate tools, manual steps
- No way to collaborate, share, and update curated datasets, reuse knowledge

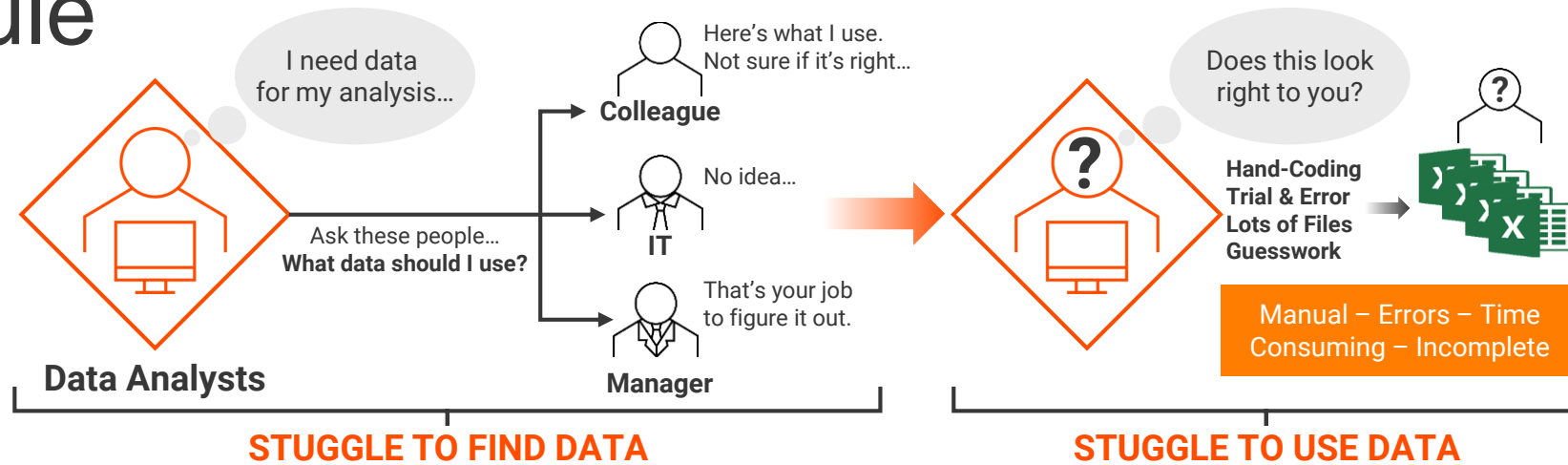


## **IT/Data Engineers**

---

- Can't cope with growing demand for data from the business
- No visibility into what the business is doing with the data
- Struggling to deliver value to the business
- Losing the ability to govern and manage data as an asset

# Enterprise Data Preparation (EDP) Flips the 80/20 Rule



80% of time spent on data prep

20% on analytics

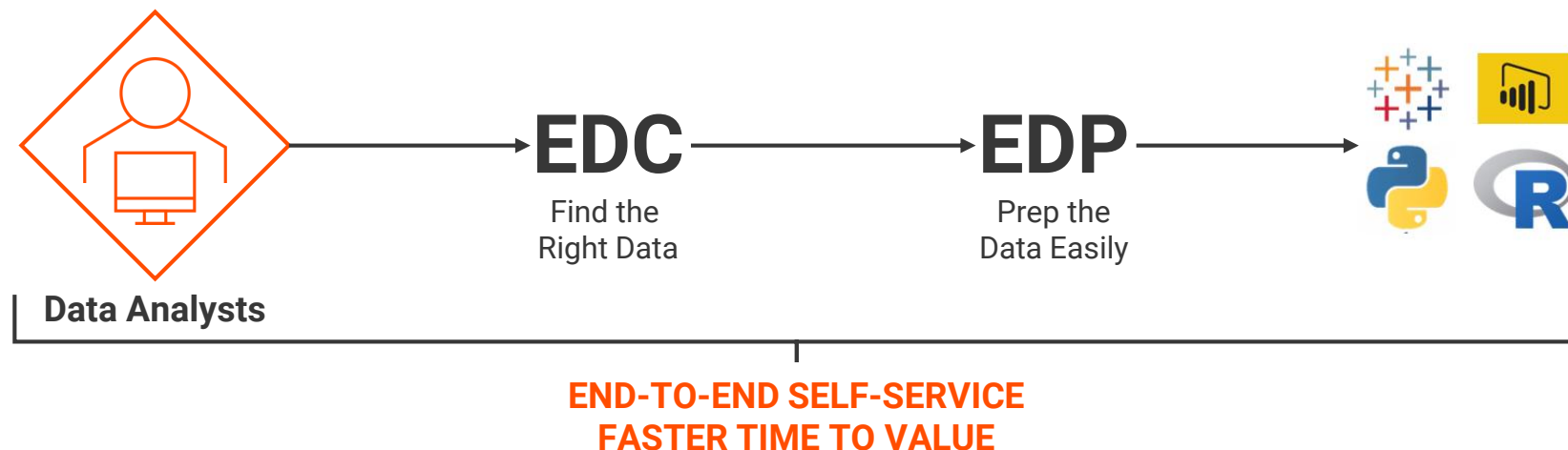
Months / Weeks

VS

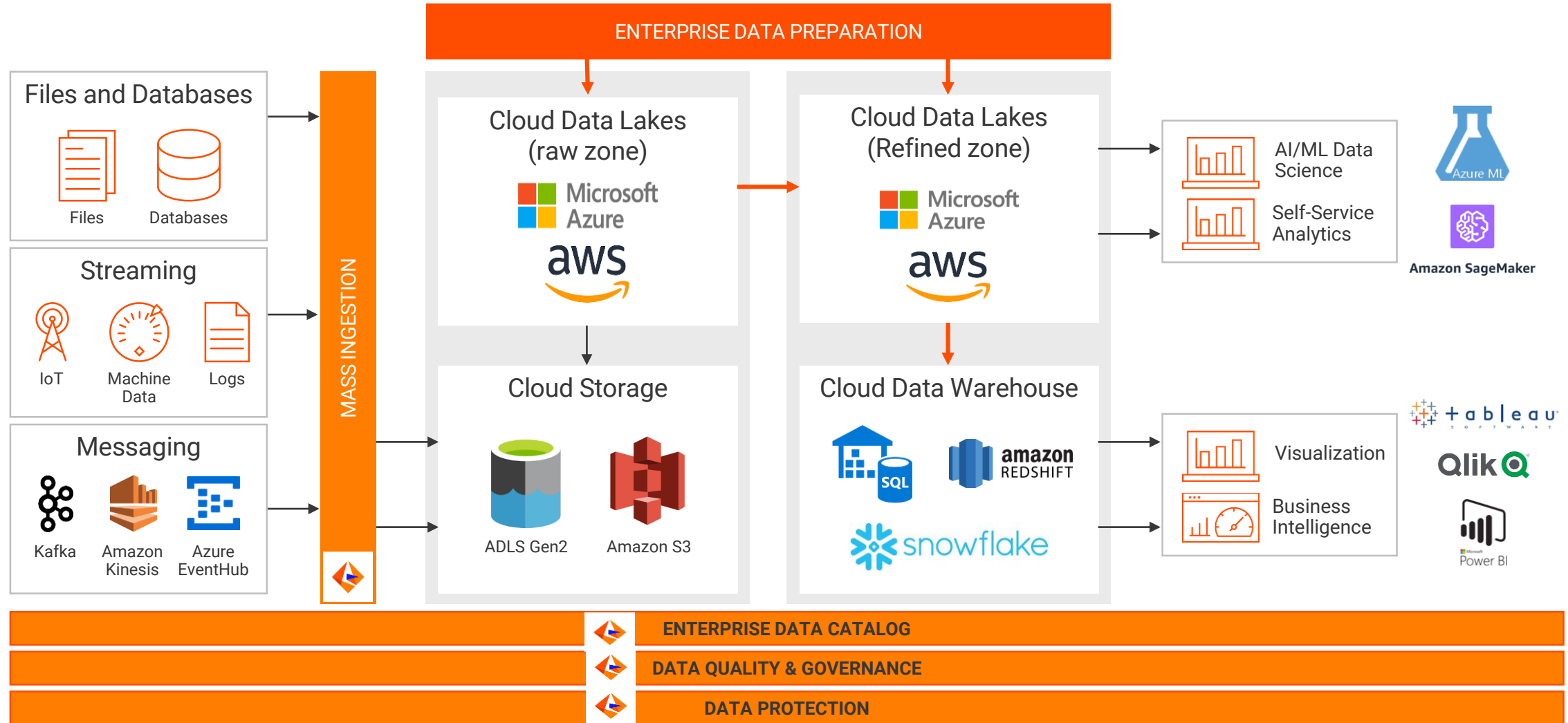
Hours / Minutes

20% of time spent on data prep

80% on analytics



# Enterprise Data Preparation Reference Architecture



# Enterprise Data Preparation Primary Use cases



## **Data Preparation for Self-service Analytics on Data Lakes**

---

Find, explore and prepare data quickly in the data lake to support ad-hoc analytics followed by operationalization in a governed environment



## **Data Preparation for Advanced Analytics (AI/ML Projects)**

---

Find, explore and prepare data that quickly so that it is ready for Data Science projects followed by operationalization of Data Science process

# Enterprise Data Preparation Empowers DataOps Teams

Democratize data pipelines for analytics and AI/ML workloads at scale

## Data Engineer

How can I quickly prepare data for advanced analytics and operationalize data pipelines for reliable data delivery?



## Data Analyst

How can I quickly and easily find, explore and prepare data for my analysis and create a repeatable process without depending on IT?



## Data Preparation Process Agile, Iterative, Collaborative

## Data Scientist

How can I extract completely new patterns and models that detect or predict behavior?



## Data Steward

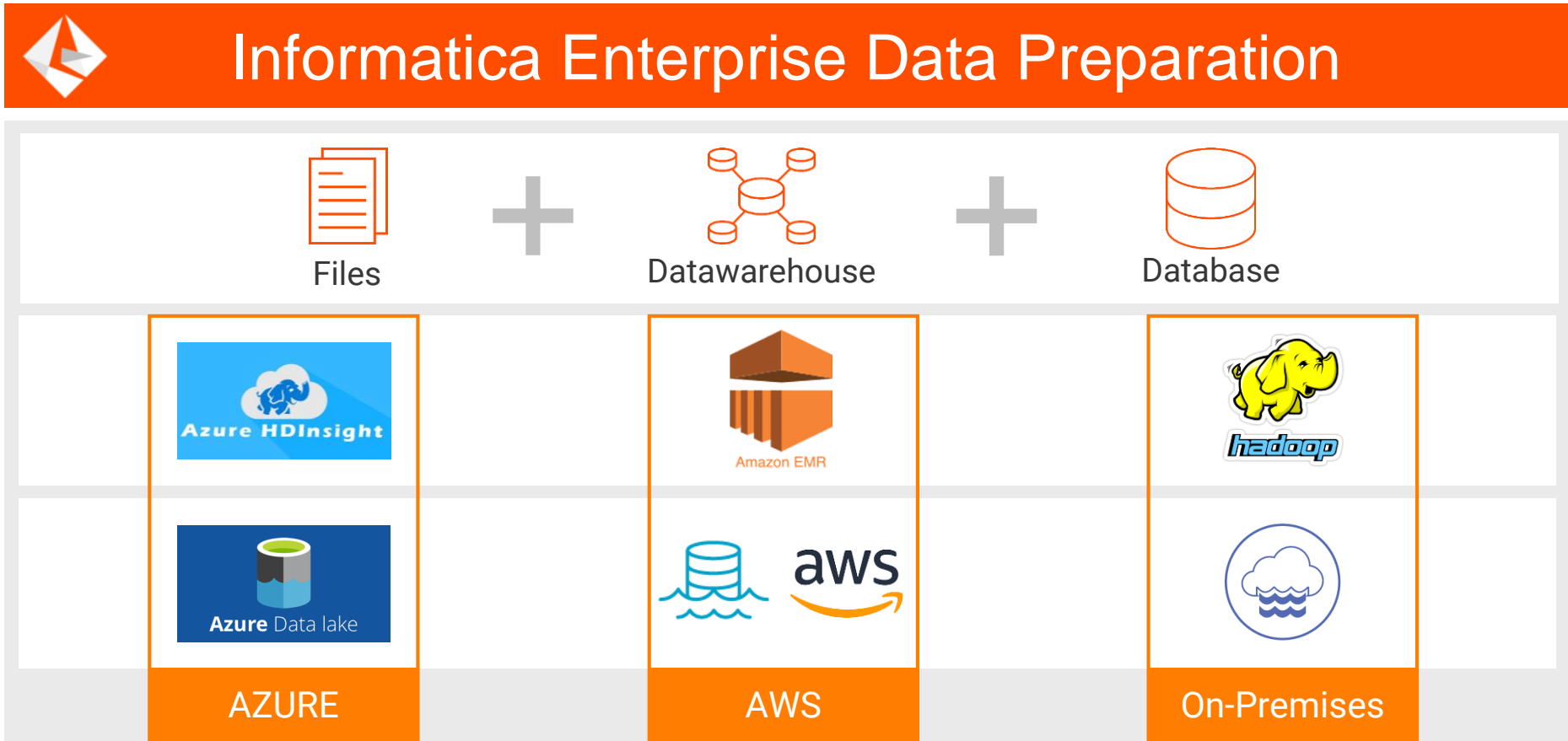
How can I enable self-service and ensure right data is available to right people in trusted and secure manner?

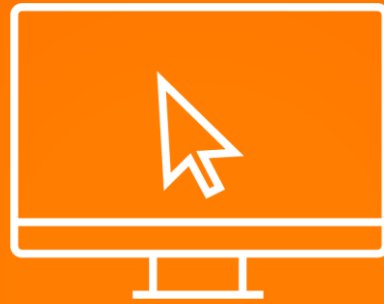


# Data Preparation on Modern Data Lakes with EDP 10.4

Enable citizen integrators to be self-served on ADLS/S3 data lakes

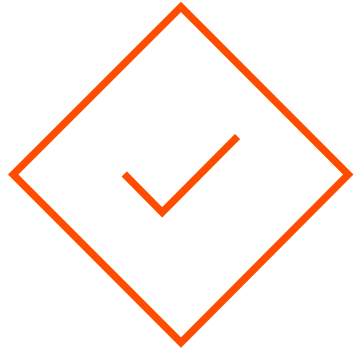
Self-Serve Data Prep





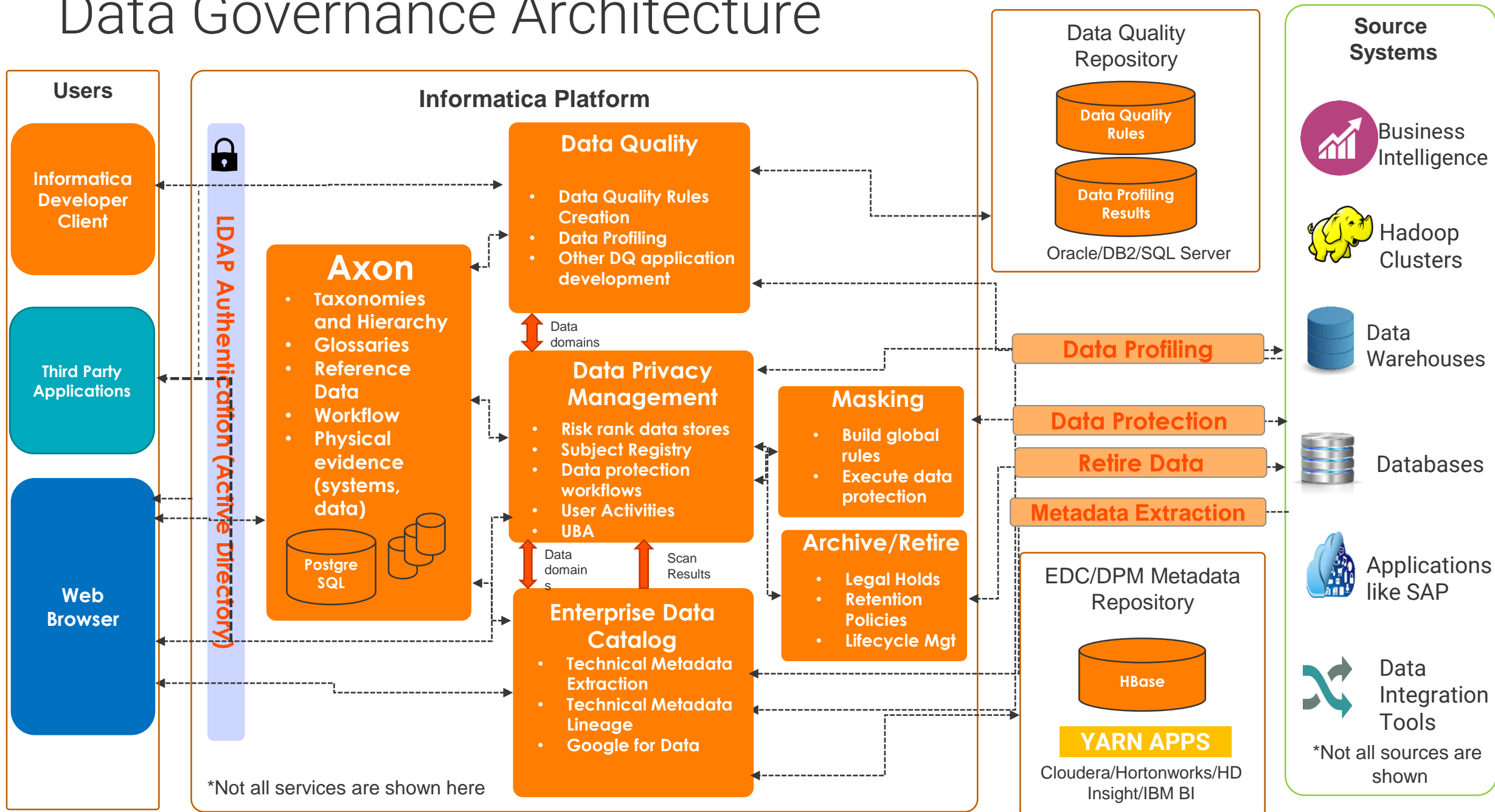
Demo





# Data Catalog for Data Governance Journey

# Data Governance Architecture

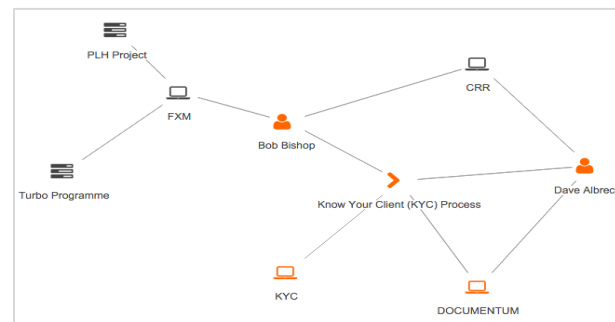


# Solution Approach for Actionable Understanding

Glossary

Data Sets 13 of 17	System 10 of 11	Glossary 63 of 78	Data Quality 25 of 29
Process 5 of 12	Capability 0 of 50	Business 0 of 19	Project 86 of 98
Product 6 of 9	Policy 8 of 50	Change 4 of 4	Regulation 0 of 105
Role 501 of 501	People 8 of 8	Committee 0 of 19	Jurisdiction 4 of 4
Client 3 of 3	Org Unit 5 of 9	Legal Entity 0 of 5	Regulator 7 of 7

Relate



Control



Collaborate

# Automate Data Discovery, Cataloging, and Context

Leverage ML and AI to find critical data across structured and unstructured sources

Onboard discovered data automatically with oversight and control

Automatically tag data with business context to help users assess relevance

The screenshot displays the Informatica Data Catalog interface, organized into three main sections:

- Top Section: Data Discovery**
  - Filters: Location, Customer, Person.
  - Buttons: Show All.
  - Table with columns: Name, Business Title, Data Domains, Null | Distinct | Non-Distinct %, Source Data Type, Inferred Data Types.
  - Rows: cust\_email (Email), cust\_firstname (FirstName), cust\_gender (Gender), cust\_id (CMD-ID).
- Middle Section: DATA ATTRIBUTE INSTANCES**
  - Review Status: Accepted, CLAIRE Discovered.
  - Table with columns: Ref, Name, Definition, Review Status, Field Name, Confidence Score (%), Data Set.
  - Rows: Attribute Name 01 to 33, showing various attributes and their review status.
- Bottom Section: Customer (customer\_master\_CS)**
  - Overview tab selected.
  - Description: A party that receives or consumes products (goods or services) and has the ability to choose between different products and suppliers.
  - Sample Columns: cust\_firstname (FirstName), cust\_middlename (Middle Name), cameo\_category\_desc (CAMEO Category Descrip...).
  - People: Data Owner (Gaurav Pathak), Data Steward (Khan Tan), Experts (Abhiram Gujjuwar).

# AXON - EDC - Business & Technical Lineage

Informatica  
Axon

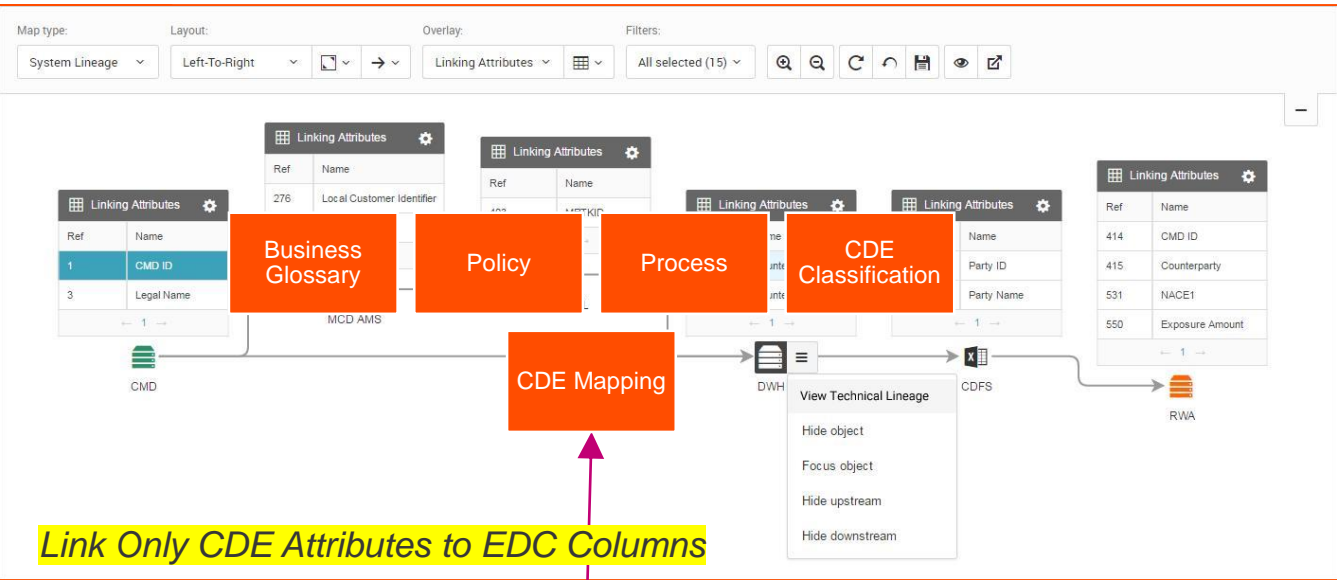
Conceptual



Business  
SME

Maps

Document/Enforce



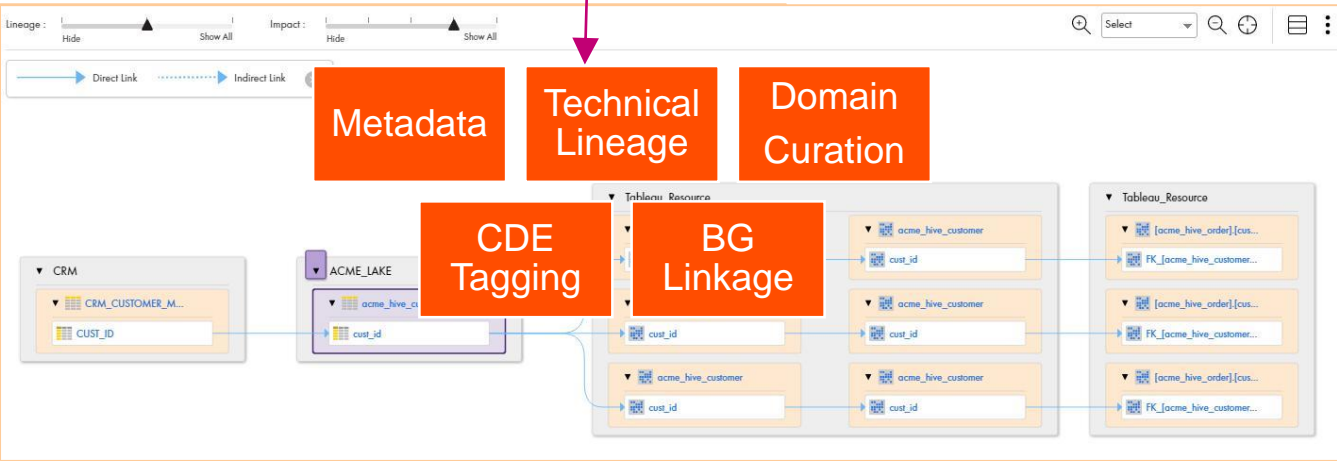
Enterprise  
Data  
Catalog

Physical



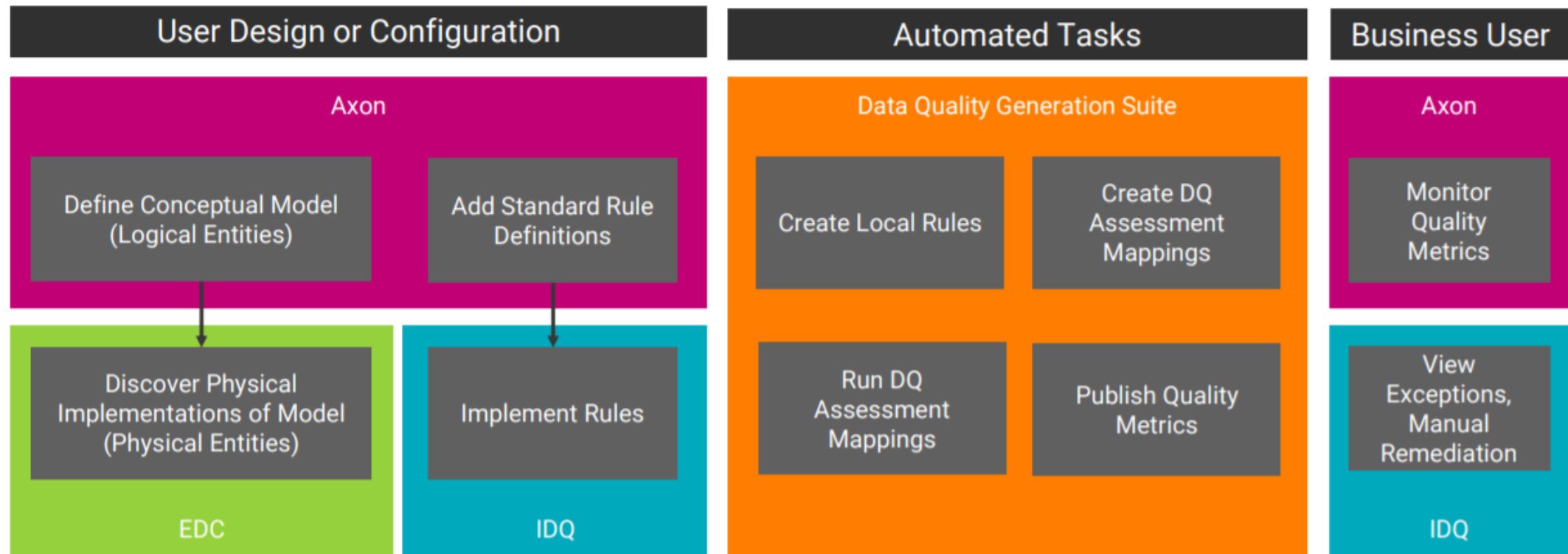
Enterprise  
Architect

Lineage



Validate/Recommend

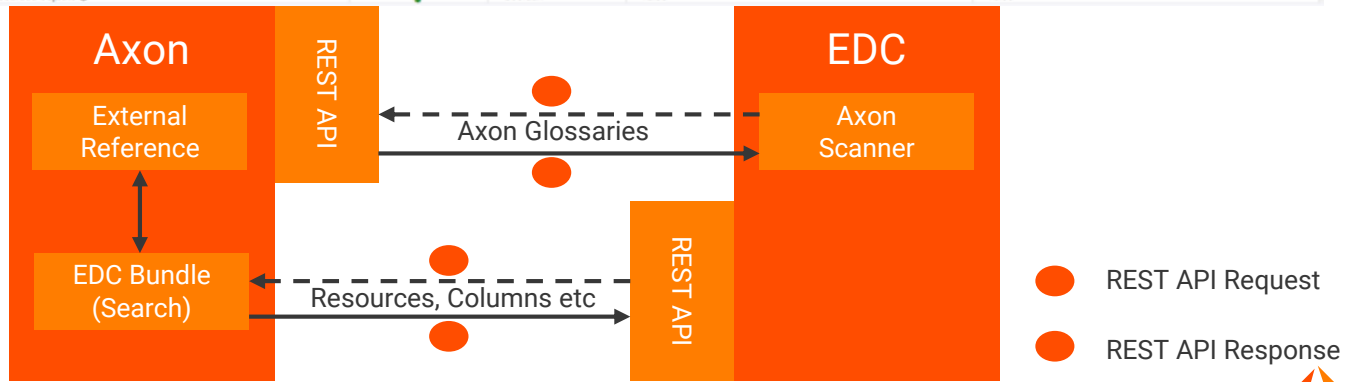
# Automating Data Quality for Data Governance

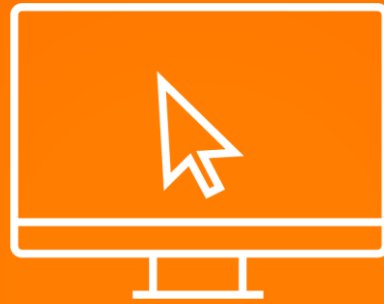


# Axon integration with EDC

## Configuration

Display Name	Customizable	Value Type	Default Value	Value
Enterprise Data Catalog Server Host	✓	string		
Enterprise Data Catalog Server Port	✓	string		
Enterprise Data Catalog Server Login User Name	✓	string		Administrator
Enterprise Data Catalog Server Login Password	✓	string	*****	*****
Enterprise Data Catalog Server Login Namespace	✓	string		Native
Enterprise Data Catalog Resource Name	✓	string		axonresource
Enable Automatic Onboarding	✓	boolean	false	true
Confidence Score Threshold	✓	string	95	95
Accept Onboarded Objects	✓	boolean	false	false





Demo



Questions?





Thank You!

---

# References

- IICS – EDC Integration Properties : <https://docs.informatica.com/integration-cloud/cloud-platform/current-version/administrator/organizations/organization-properties/enterprise-data-catalog-integration-properties.html>
- EDC Managing Cloud Org : <https://docs.informatica.com/data-catalog/enterprise-data-catalog/10-4-0/catalog-administrator-guide/managing-cloud-organization.html>
- EDC Data Preview & Provisioning : <https://docs.informatica.com/data-catalog/enterprise-data-catalog/10-4-0/enterprise-data-catalog-user-guide/data-preview-and-provisioning.html>
- Using Axon with Enterprise Data Catalog  
<https://kb.informatica.com/h2l/HowTo%20Library/1/1150-UsingAxonDataGovernancewithEnterpriseDataCatalog-H2L.pdf>
- Axon Checklist to Automate Data Quality Rules  
<https://docs.informatica.com/data-quality-and-governance/axon-data-governance/h2l/checklist-to-automate-data-quality-rules-in-axon-data-governance/abstract.html>
- 7 Best Practices to Drive Data Catalog Adoption  
[https://www.informatica.com/lp/7-best-practices-to-drive-data-catalog-adoption\\_3712.html#fbid=PYYnGF22B9T](https://www.informatica.com/lp/7-best-practices-to-drive-data-catalog-adoption_3712.html#fbid=PYYnGF22B9T)
- Managing the Data Lake using EDP : <https://docs.informatica.com/data-catalog/enterprise-data-preparation/10-4-0/enterprise-data-preparation-administrator-guide/managing-the-data-lake/data-lake-management-overview.html>
- New Features in EDP v10.4 : <https://docs.informatica.com/data-catalog/enterprise-data-preparation/10-4-0/new-features-guide/new-features--10-4-0-/enterprise-data-preparation.html>