

August 11, 2020

# What's New in Enterprise Data Catalog 10.4.1

Keshav Ramarao

# 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.

# Feature Rich Success Portal

---



Bootstrap trial and  
POC Customers



Enriched Customer  
Onboarding  
experience



Product Learning  
Paths and Weekly  
Expert Sessions



Informatica  
Concierge with  
Chatbot integrations



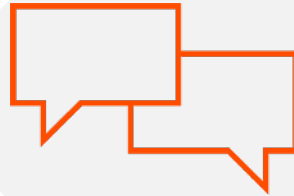
Tailored training and  
content  
recommendations

# More Information



## Success Portal

<https://success.informatica.com>



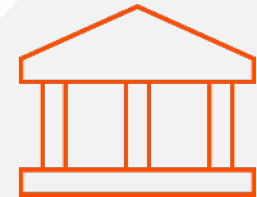
## Communities & Support

<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

- New Features in 10.4.1
- Demo
- Q&A



*Enabling the next frontier of meta-data connectivity*

# Informatica + Compact Solutions

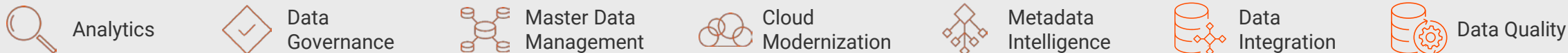
## Informatica Acquires Compact Solutions to Extend Industry Leading Enterprise Data Catalog

*Industry's first and only catalog of catalogs across all enterprise data with the broadest metadata connectivity*

- **Extends the industry's most comprehensive and detailed data lineage** (i.e. no “black boxes”) to understand the provenance of all enterprise data and impact of changes as companies digitally transform and modernize.
- **Expands the industry's broadest metadata connectivity** to catalog all types of data in support of analytics, data governance and privacy, customer experience, and data warehouse modernization initiatives.
- **Establishes the industry's only single vendor metadata management solution** to simplify procurement, deployment, maintenance and support.



# The Catalog of Catalogs



Open APIs, Full Access

- Discovery
- Profiling
- Lineage
- Impact Analysis

- Semantic Search
- Domain Discovery
- Similarity Clustering
- Business Term Association



## Enterprise Data Catalog

- Relationships
- Business Context
- Glossary Integration
- Custom Annotations
- Reviews/Ratings
- Questions/Answers
- Data Certifications
- Change Notifications

Knowledge Graph + AI/ML

**CLAIRE™**

Breadth of Active Metadata



On-prem  
Databases



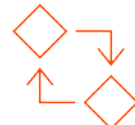
Data  
Warehouses



Data Lakes



File  
Systems



Code and  
Scripting



Statistical & BI  
Tools



Analytics  
Apps



On-prem/  
SaaS Apps



ETL



Mainframes



Platform as a  
Service

# The Catalog of Catalogs: One Vendor, One Solution



Analytics



Data Governance



Master Data Management



Cloud Modernization



Metadata Intelligence



Data Integration



Data Quality

Open APIs, Full Access

- Discovery
- Profiling
- **Lineage**
- **Impact Analysis**
- Semantic Search
- Domain Discovery
- Similarity Clustering
- Business Term Association



Enterprise Data Catalog  
+ Compact Solutions

- Relationships
- Business Context
- Glossary Integration
- Custom Annotations
- Reviews/Ratings
- Questions/Answers
- Data Certifications
- Change Notifications

Knowledge Graph + AI/ML

**CLAIRE™**

Breadth of Active Metadata



On-prem Databases



Data Warehouses



Data Lakes



File Systems



Code and Scripting



Statistical & BI Tools



Analytics Apps



On-prem/SaaS Apps



ETL

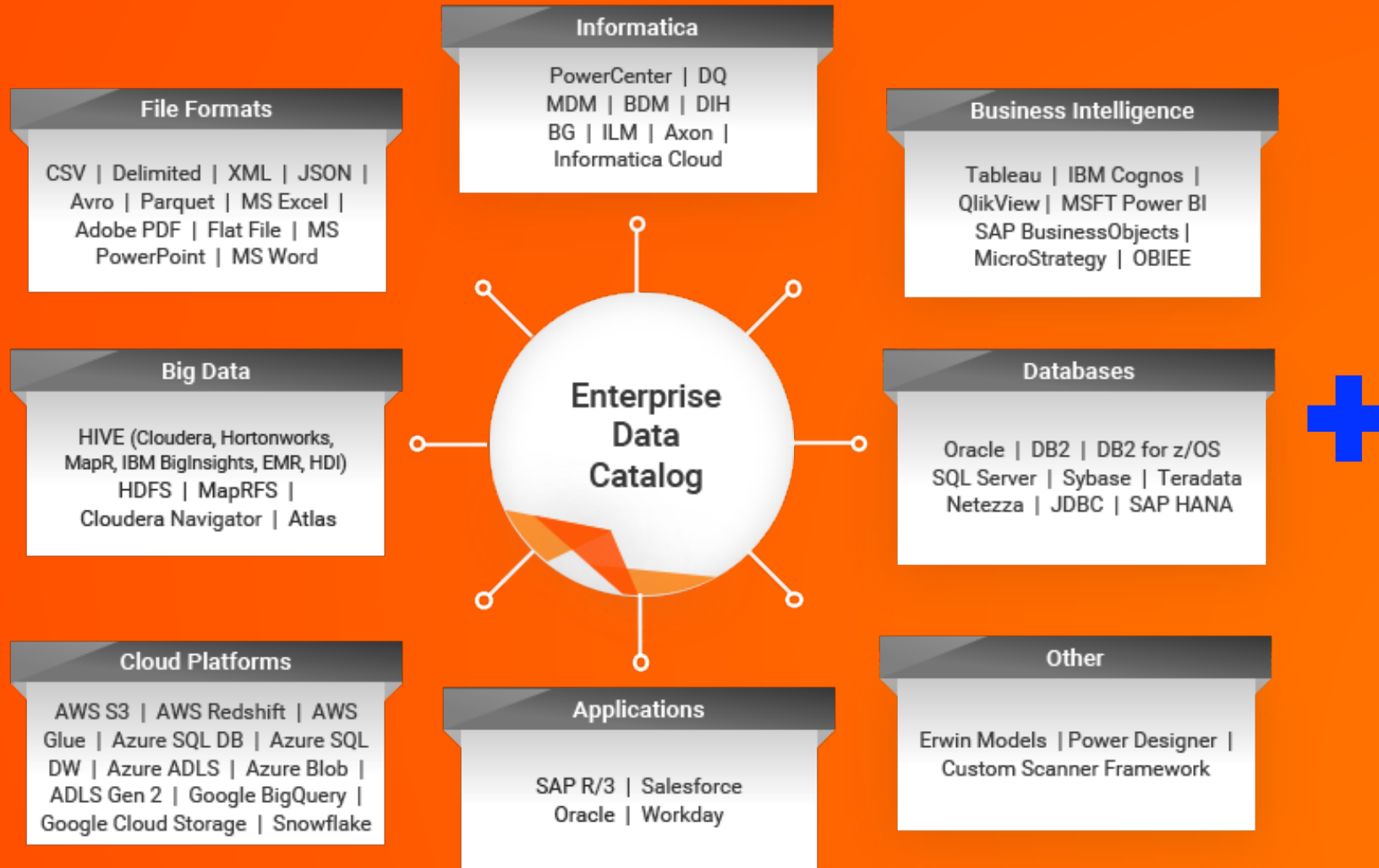


Mainframes



Platform as a Service

# Broadest and Most Complete Metadata Connectivity



## EDC Advanced Scanners

### Code and Scripting

- Oracle
- SQL Server
- Teradata
- Netezza
- DB2
- Sybase

### ETL Tools

- IBM Datastage
- Microsoft SSIS

### Mainframes

- COBOL
- JCL

### Statistical and BI Tools

- SAS
- Microsoft SSAS
- Microsoft SSRS

# EDC Advanced Scanners General Availability

*Available  
NOW!  
in EDC 10.4.1*



# Informatica EDC 10.4.1 (Summary)



## User Experience

- Data Asset Analytics
- Walkthrough – On-line Contextual Tutorial
- Field sort by position in source system
- Execution contextual lineage support



## CLAIRE

- Domain Discovery – Rule Enhancements
- Business Term Association enhancements



## Deploy at Enterprise Scale

- Resource Configuration Permissions
- Hot Backup\*
- Rest API based log aggregation\*



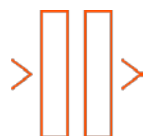
## Scanners/Connectivity

- SAP BW & SAP BW/4 HANA (GA)
- SAP ECC & S/4 HANA\*
- SSIS detailed lineage
- DataStage (GA)
- ETLs: Field level lineage for Flat files
- IICS: File lineage on Cloud storage
- Partitioned File Support
- Parquet Profiling on S3 and ADLS
- New MDM Scanner
- MicroStrategy enhancements
- ADLS Gen 2 profiling
- Cassandra Profiling
- Reference object for Hive and MS SQL Server

# Data Asset Analytics

Industry's first data asset analytics solution

**Optimize data asset value by using analytics on data asset inventory, enrichment, usage, and user collaboration in your Enterprise Data Catalog**



Measure



Analyze



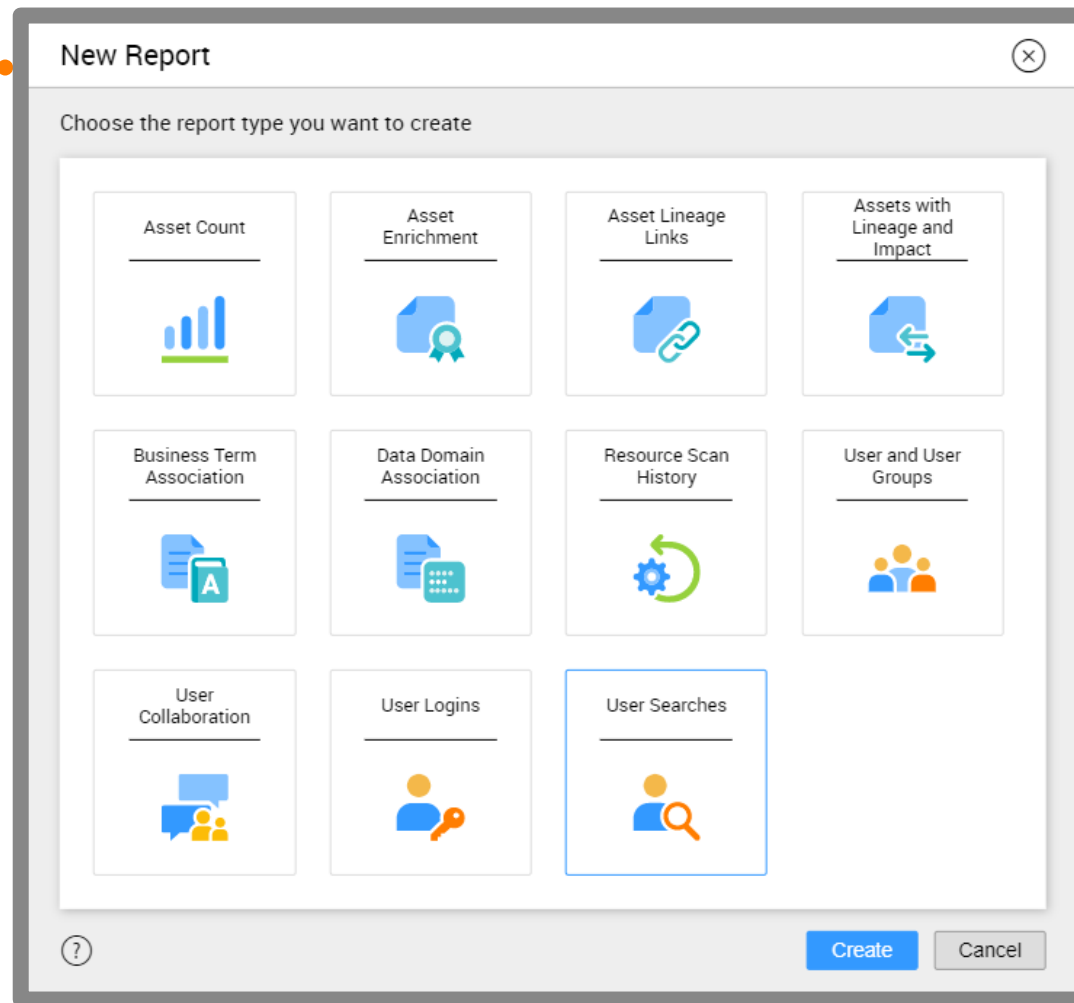
Optimize

# Data Asset Analytics (DAA)

## Highlights

### Comprehensive Report Types

- Predefined event (audit) history datasets
- Preview and download event history data
- Access event datasets directly from BI Tools



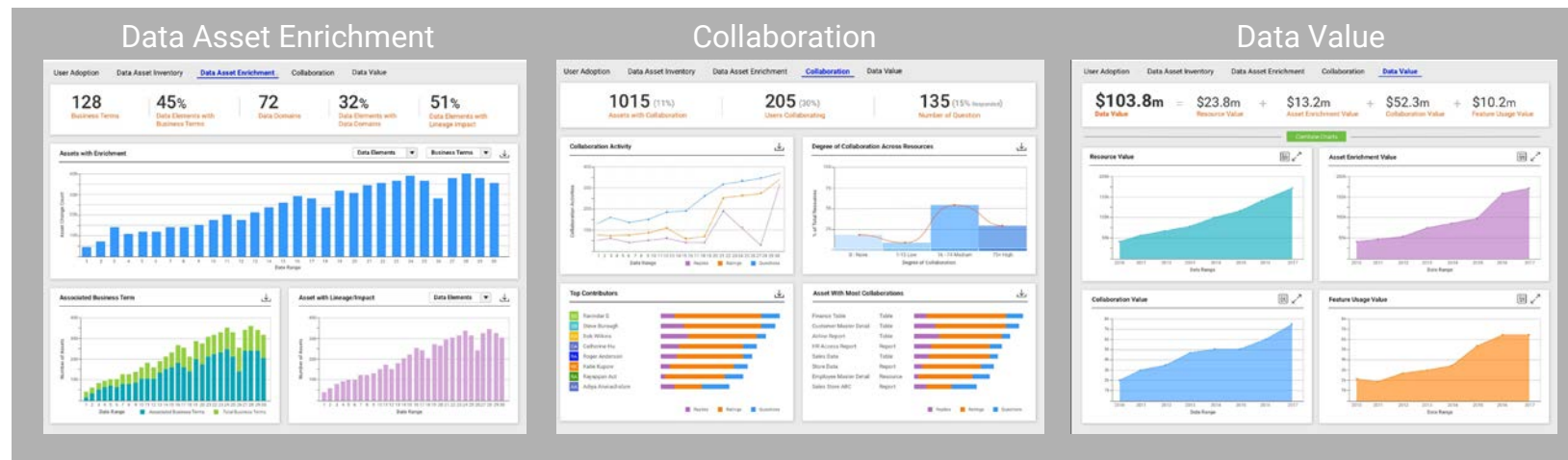
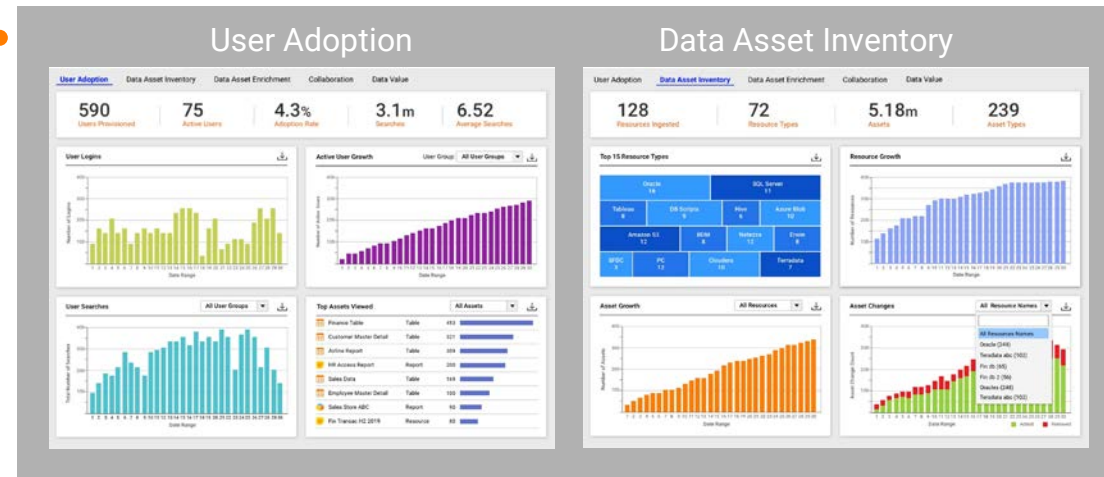


# Data Asset Analytics (DAA)

## Highlights

### Out-of-Box Data Asset Analytics Dashboards

- Instant visibility to data asset usage, inventory, enrichment, collaboration, and data value
- Near real-time key metrics and trends
- Quick analysis with built-in filters
- Storytelling for data assets





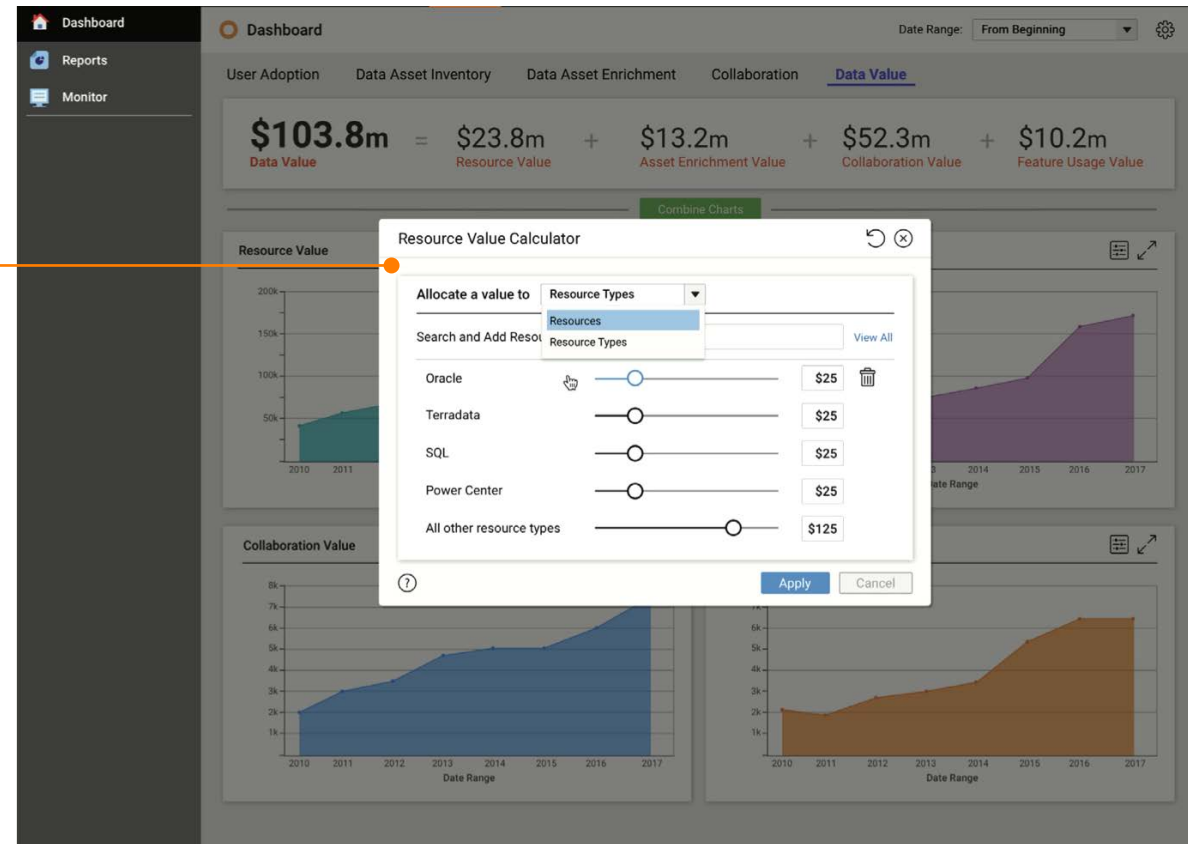
# Data Asset Analytics (DAA)

## Highlights

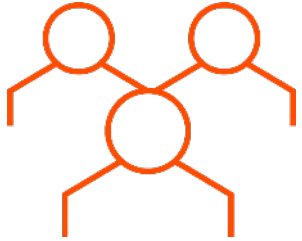
### Automated Data Value Calculator

- Allows an enterprise to estimate the data value
- Takes into account four key metrics:
  - Resource value
  - Asset enrichment value
  - Collaboration value
  - Feature usage value\*

Note: \* 10.4.1.1



# Key Benefits



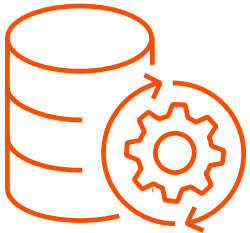
## Gain Insight into Data Usage and User Adoption

- Insight into what data asset is in demand by which group of user
- Visibility into who are the top contributors, influencers and subject matter experts
- Spot trends and uncover usage patterns to ensure relevant data assets are available



## Identify and Socialize Best Practices to Empower Data-Driven Decision Making

- Identify data usage best practices, and socialize them across functions
- Diagnose and remedial instances of low data usage and adoption
- Catalyst to empower and expand data-driven decision making



## Understand and Maximize Data Asset Value

- Understand the value of any data asset within the catalog
- Focus organizations on their most valued assets
- Drive greater data value with data knowledge, collaboration, and usage

# Data Asset Analytics: Opening Door to New Possibilities

Visibility to data asset usage, inventory, enrichment, collaboration and data value



Track Key Metrics



Monitor Activities



Undercover Patterns



Spot Trends



Improve Data Value



Capture Mis-use



Answers to Questions

- Who is using, what asset, what feature, and when?
- What is the catalog adoption rate?
- What are the top searches, and average searches per user?
- What is the most access catalog module or assets?
- Who is doing DG and curating asset?
- What is % of objects with lineage, owner, description, bus. term?
- Who are the top contributors and SME inferred from collaboration activities?
- Which user groups, datasets need more attention and why?
- Who modified and scanned sensitive data sources?
- How to grow data asset value?



# DAA: License, Installation, Configuration

## License

- DAA is included as of part EDC

## Installation

- DAA is part of Catalog Service (CS) configuration in the INFA Admin Console

## DAA Configuration

- For new EDC 10.4.1 install, by default DAA is enabled
- For upgrade from EDC 10.4.0 to 10.4.1, by default DAA is disabled
- DAA configuration can be modified post EDC 10.4.1 installation or upgrade

**New Catalog Service - Step 5 of 5**

Fields marked with an asterisk (\*) are required.

**Specify the properties for this new CatalogService**

**Data Asset Analytics Configuration Properties**

☒ Enable Data Asset Analytics

Database Type to Select Database \*

Username \*

Password \*

Database Connection String \*

Secure JDBC Parameters

# DAA: Privileges

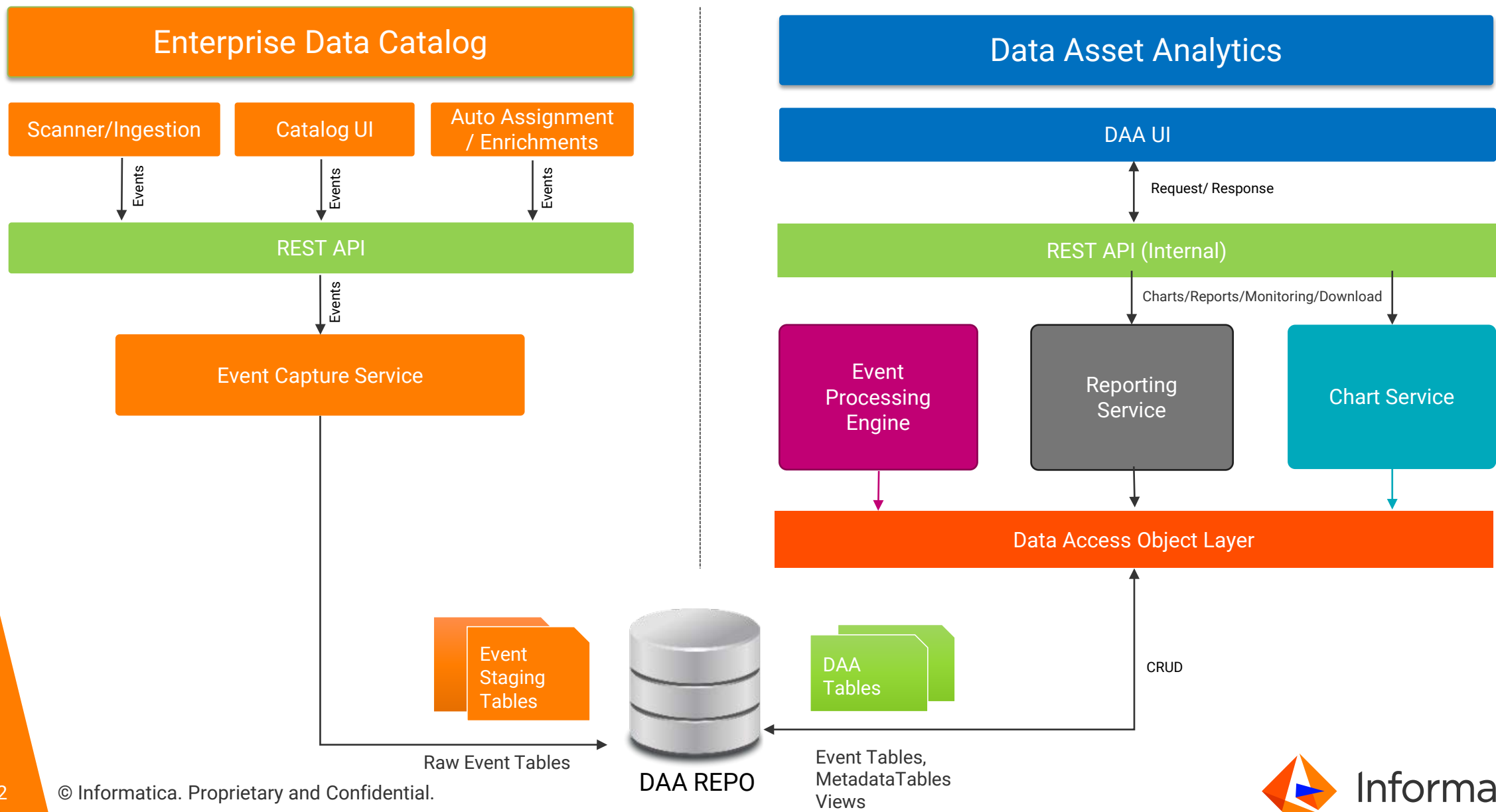
Three new privileges are added into existing CS privileges

DAA Privileges: >  Permissions: v	Dashboard Management: Visualize	Report Management: View and Download Raw Data	Admin Management : Configuration
View Dashboards	Y	Y	Y
Download Dashboard Chart Data	N	Y	Y
Create Reports & Download Report Data from Monitor tab	N	Y	Y
Edit Asset Value Calculator	N	N	Y
DAA application configuration	N	N	Y

**Note:**

By default, Native "Administrator" User or user having system defined "Administrator" group or role will have access to everything in Analytics Tab

# DAA Architecture



# Recommendations

- It is highly recommended to enable DAA always
- Separate & Dedicated Schema is recommended in Oracle and PostgreSQL. Similarly, separate and dedicated Database for MS SQL Server
- As a best practice, better to create dedicated Table Space/File Group for DAA schema/Database
- Minimum Disk space requirement on the Database Server:
  - For 10 Million Assets: 15-20 GB
  - For 20 Million Assets: 30-40 GB
  - For 30 Million Assets: 40-50 GB
  - Similarly, Database hosting server should have minimum of 24GB RAM and 16 CPUs/Cores, if server is hosting only the Database
- Run the Table Statistics on periodic basis for optimum performance



# Demonstration

Data Asset Analytics

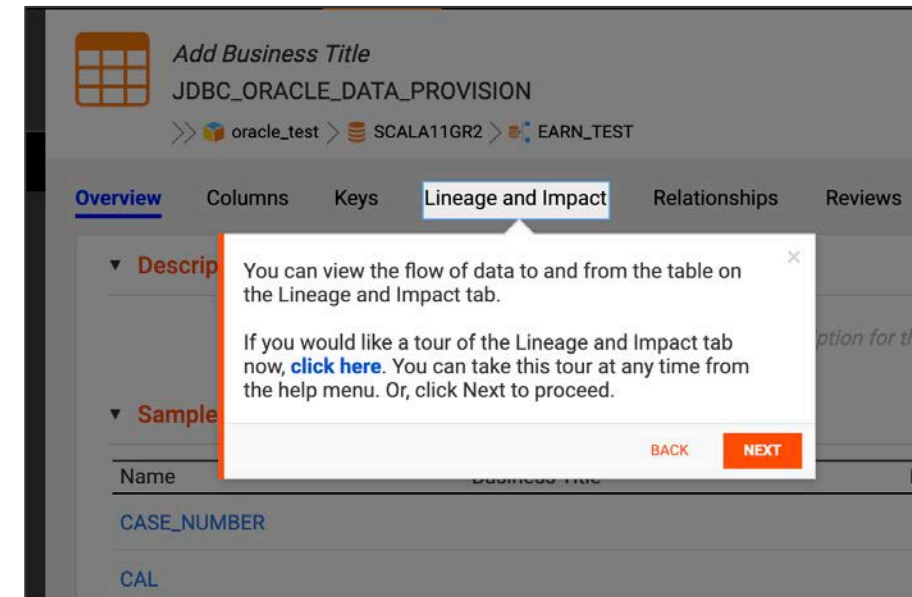
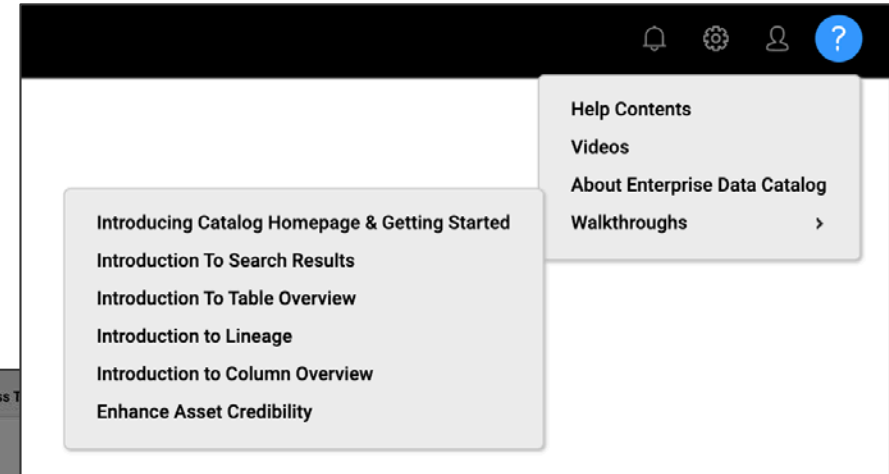
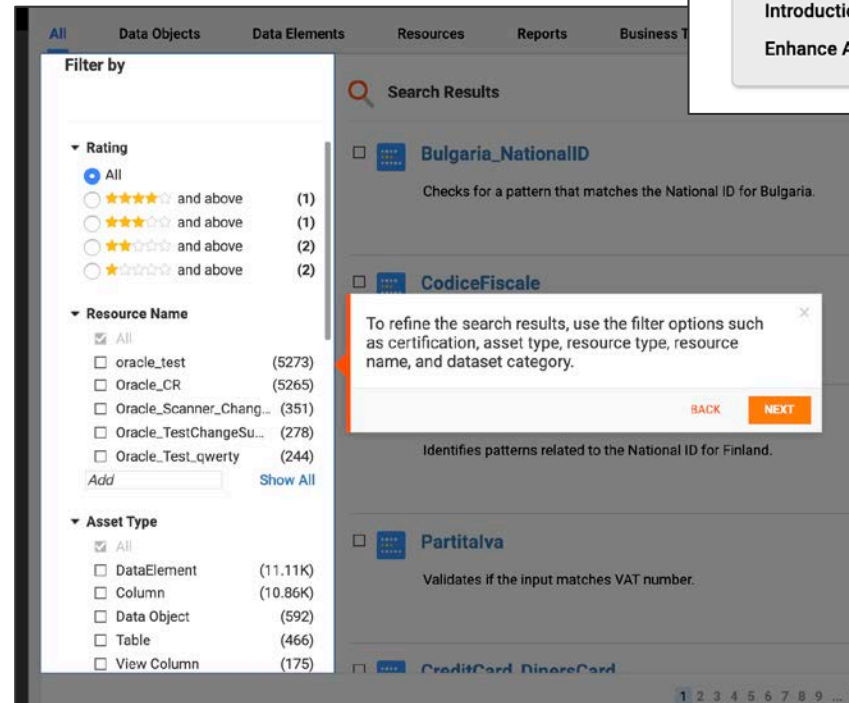


# Reference to Useful DAA Resources

- All documentation (User Guide, Admin Guide, Sizing Guidelines, etc.):  
<https://docs.informatica.com/data-catalog/enterprise-data-catalog/10-4-1.html>

# WalkThroughs

- On-demand, Context-based, in app product walkthroughs
- New “Walkthroughs” option added to help menu
- Teach users about:
  1. Homepage
  2. Search Results
  3. Tables & Columns
  4. Lineage
  5. Asset Enrichment



# Change Summary Enhancements

- Key Enhancements:

1. All changes displayed chronically by default
2. Filter by asset name, asset type, and scan time (now allows for time periods)
3. Enhanced sorting capabilities
4. User can now export change summary into Excel

The screenshot displays the Informatica Change Summary interface. On the left, a 'Filter by' panel includes sections for 'Asset Type' (with checkboxes for All, Column, Database, External Database, External Schema, and Foreign Key Constraint), 'Scan Time' (with a 'Custom' dropdown and date pickers for 'From' and 'To'), and 'Changes Followed' (with a 'Source Changes' dropdown). An 'Apply' button is at the bottom of the filters. A calendar for May 2020 is overlaid on the right side of the filter panel, with the 19th highlighted. The main area shows a table of changes with columns: Asset Name, Scan Time, Asset Type, Source Changes, and Path. The first row shows an asset named ALPHA123, scanned on May 19, 2020, at 12:42 AM, identified as a Table. The 'Source Changes' column contains a detailed log of system actions, including profile additions, checksums, and row counts. The 'Path' column shows the location of the change summary file. At the bottom, there is a pagination bar indicating '1 - 20 of 222' items, a page selector for '1 of 12', and an 'Assets per page' dropdown set to '20'.

Asset Name	Scan Time	Asset Type	Source Changes	Path
ALPHA123	May 19, 2020, 12:42 AM	Table	Asset Modified System added the Last Profile Time For File: 2020-05-19T07:41:48 to asset ALPHA123. System added the Column Checksum: 6dc701b0ed1ce3225127fe1a... to asset ALPHA123. System added the Estimated Row Count: 5 to asset ALPHA123. System added the Profiled Row Count: 5 to asset ALPHA123. System added the Rows covered by profiling: 100.00% (5	Oracle_Scanner_ChangeSummary/SCALA11G/EARN_TEST1

# Key Upgrades in 10.4.1 (New)

All changes shown chronologically by default

Search for a specific asset by name

Add Asset Type column

Export Change Summary to CSV file

Asset Type Filter

Filter by time periods instead of individual scans

**Oracle\_ChangeSummaryArpit**

Overview Lineage and Impact Relationships **Change Summary**

Last Refresh: at 12:44 PM

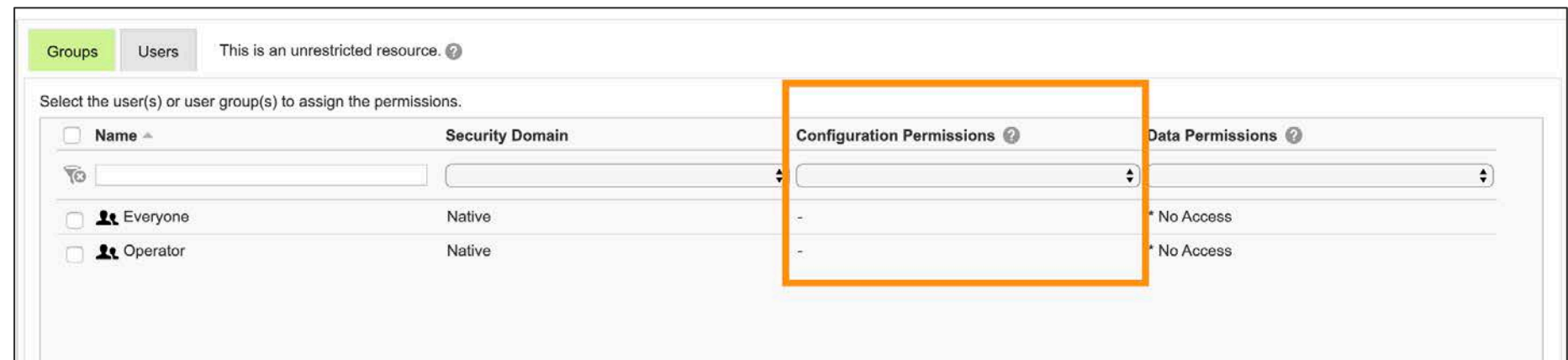
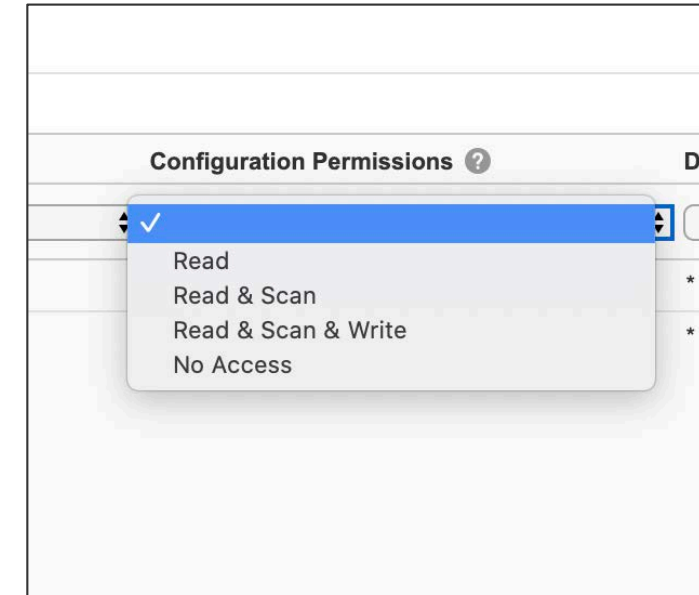
Search by Asset Name

Asset Name	Scan Time	Asset Type	Source Changes	Path
SCALA11GR2_[783]M...	May 27, 2020, 1:39 AM	External Schema	Asset Modified System updated value of name from <b>External Schema_0_63</b> to <b>External Schema_0_65</b> in asset SCALA11GR2_[783]MYLINK.INFOR... <a href="#">+ 1 More</a>	Oracle_ChangeSummaryArpit/SCALA11GR2/MYLINK.INFORMATICA.COM
ALPHA123_PK	May 27, 2020, 1:39 AM	Primary Key Constraint	Asset Added <a href="#">+ 2 More</a>	Oracle_ChangeSummaryArpit/SCALA11GR2/ARPIT/ALPHA123
ALPHA123_UN	May 27, 2020, 1:39 AM	Unique Constraint	Asset Added <a href="#">+ 2 More</a>	Oracle_ChangeSummaryArpit/SCALA11GR2/ARPIT/ALPHA123
COL1	May 27, 2020, 1:39 AM	Column	Asset Modified System updated value of Primary Key Column from <b>false</b> to <b>true</b> in asset COL1. <a href="#">+ 2 More</a>	Oracle_ChangeSummaryArpit/SCALA11GR2/ARPIT/ALPHA123
TEST140	May 27, 2020, 1:23 AM	View	Asset Modified	Oracle_ChangeSummaryArpit/SCALA11GR2/ARPIT/TEST140

Assets per page: 20

# Resource Configuration Permissions

- Edit permissions on Resources can now be set at the user level
- New “Configuration Permissions” column in Security library
- Creator of resource will get all permissions by default



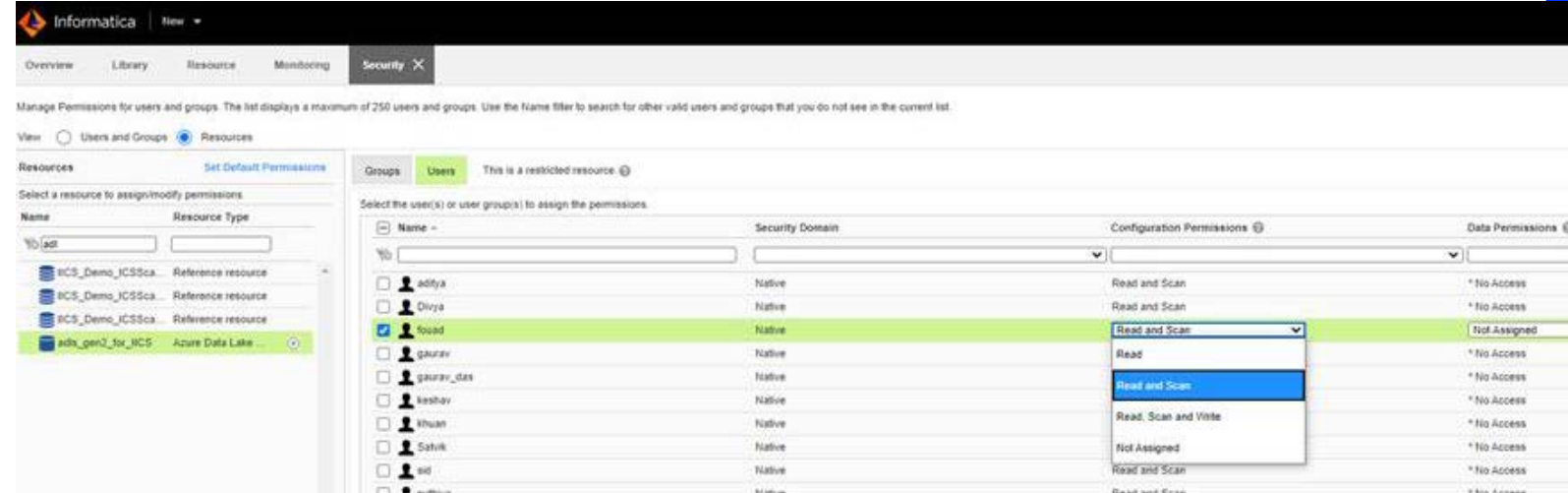
# Resource Configuration Permissions

## Resource Configuration Permissions

- Read
- Write
- Scan

## Default Permissions

- By default creator of the resource will get read permissions on the catalog objects as well as write permissions on resource config



01

### Resource Library

- Read
  - View resources in library
- Scan
  - Cancel, Rollback and Purge of resources
- Write
  - Edit resource configuration
  - Delete Resource

02

### Security Library

- Read
  - Users with read permissions can view who all have permissions on this resource
- Write
  - Users require this permission to assign permissions to other users/groups

03

### Connection Assignment

- Read
  - Users can view list of connections extracted by resource
  - Users can generate Missing link report
- Write
  - Users can assign/unassign connections

04

### Monitoring

- Read
  - Only users with this permission can view the scans done for this resource and statistics of this run
- Login page
  - This page shows overall statistics

# Discovery enhancements in 10.4.1

# Auto Business term association – Limitations Now

Finds match only when words appear in the same sequence in an Asset and the corresponding Business Term.

Example: `ISSUE_COUNTRY_CODE` does not match with `Country Code Issue`

All words in a business term are given the same weight.

Missing words – other than stop words – are not tolerated when matching with assets.

For example, in the following association, the word “Code” in the business term doesn’t align with any token in the asset name:

`CAREER_LEVEL_SUBGROUP` does not match with `Career Level Subgroup Code`

Table Name is not considered when finding Business terms



# Auto Business term association - enhancements

The out of order matches can now be captured by Bag of Words matching (word-based search).

New TF-IDF based scoring scheme can assign weights to words based on their importance in that business term.

This facilitates ignoring of trivial words while performing term-association


Taking table name into account for column type assets, while performing term-association.

Dataset	Asset Name	Matched Term	Score	Description
Oracle Finance	FIRST DATE ACTIVATED	DATE FIRST ACTIVATED	100%	Out-of-order Term matched
Oracle Finance	LAST REJECTED DATE	DATE LAST REJECTED	100%	Out-of-order Term matched

# Auto Business term association – example (Tf-Idf)

Dataset	Asset Name	Matched Term	Score	Description
Michigan Dataset	ACAD CAR FIRST TERM	ACADEMIC CAREER FIRST TERM CODE	84.8%	“CODE” is ignored because it has a low Tf-Idf Score
Franklin	CHEMICAL SAFETY QUARTILE	CHEMICAL SAFETY SCORE QUARTILE	83.3%	“SCORE” is ignored because it has a low Tf-Idf Score.
Anthem Columns	PRMRY_PHYSN_RQRD_OPTNL_CD	PRIMARY CARE PHYSICIAN REQUIRED OPTIONAL CODE	80.9%	“CARE” is ignored because it has a low Tf-Idf Score
Anthem Columns	PAYER_VAL_AMT	PAYER VALUE CODE AMOUNT	87.1%	“CODE” is ignored because it has a low Tf-Idf Score
Anthem Columns	HSA_TRNSCTN_DESC	HEALTH SAVINGS ACCOUNT TRANSACTION CODE DESCRIPTION	87.9%	“CODE” is ignored because it has a low Tf-Idf Score
Anthem Columns	MAPD_PDP_CD	MEDICARE ADVANTAGE PRESCRIPTION DRUG PRESCRIPTION DRUG PLAN INDICATOR CODE	86.2%	“INDICATOR” is ignored because it has a low Tf-Idf Score
Anthem Columns	SRC_CLM_NASCO_PAR_CD	SOURCE CLAIM NATIONAL ACCOUNTS SERVICE COMPANY NASCO PAR CODE	80.7%	Business term has NASCO as well as full form of NASCO. “NASCO” in Business Term does not align with any asset letter.
Michigan ERPM	LAST_SCR_WITHDRAW_TOT	LAST SCHEDULE CONTINUING REVIEW WITHDRAWAL TOTAL COUNT	80.5%	“COUNT” is ignored because it has a low Tf-Idf Score
Michigan ERPM	SPONSOR_OWN_EQP	SPONSOR OWNED EQUIPMENT INDICATOR	83.9%	“INDICATOR” is ignored because it has a low Tf-Idf Score

# Domain Discovery - Override rules

 **Sample\_Custom\_DataDomain**

**Data Domain**

Specify the name, description, rules, and Data domains group for the Data domain.

**General**

**Name\*:**

**Description:**

**Specify rules for this data domain:** ☐ No ☒ Yes

You must specify either the Data Rule or the Column Name Rule created using the Informatica Analyst Tool or the Informatica Developer Tool.

**Data Rule:**

**Column Name Rule:**

**Trump Rules:**

Both metadata and data conditions match ▼

In case of a conflict, data overrides metadata condition

In case of a conflict, metadata overrides data condition

Both metadata and data conditions match

**Data Conformance**

**Minimum conformance:**

**Auto accept if more than**

# Bulk curation of custom attributes

Export Custom Attributes(1/2) ⓧ

Select Asset Types find

☒ Select Asset Types ▲

- ☒ Column
- ☐ Database
- ☐ Foreign Key Constraint
- ☐ Primary Key Constraint
- ☐ Resource
- ☐ Schema
- ☐ Table
- ☐ UDP
- ☐ UDPs
- ☐ Unique Constraint
- ☐ View
- ☐ View Column

? Next > Cancel

Export Custom Attributes(2/2) ⓧ

Choose attributes to export find

☐ Choose attributes to export ▲

- ☐ Business Description
- ☐ Data Owner
- ☐ Data Steward
- ☐ Display Name
- ☐ URL

Include

- ☒ Data Domains
- ☐ Business Terms

? < Back OK Cancel

# Bulk curation of custom attributes – CSV Example

A	B	C	D	E	F
id	core.name	core.classType	bulkimport-DataDomainsAccepted	bulkimport-DataDomainsInferred	bulkimport-DataDomainsRejected
id	name	classType	DataDomains Accepted	DataDomains Inferred(Read Only)	DataDomains Rejected
Oracle_Test://SCALA11GR2/HO	PROVISION3	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	PROVISION2	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	PROVISION1	Column			
Oracle_Test://SCALA11GR2/HO	CONTACT_NO	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	EMAIL	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	PROVISION1	Column			
Oracle_Test://SCALA11GR2/HO	PROVISION3	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	PROVISION2	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	FIRST_NAME	Column			
Oracle_Test://SCALA11GR2/HO	STRING_APP	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	AGE	Column			"/AlphaNumeric_SpecialCharacter,/Age"
Oracle_Test://SCALA11GR2/HO	BLOOD_GROUP	Column			
Oracle_Test://SCALA11GR2/HO	AMBULANCE_STATUS	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	DOC_NAME	Column			
Oracle_Test://SCALA11GR2/HO	MED_NAME	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	EMP_ID	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	ORACLE_ACCESSCONTROLVS	Column			
Oracle_Test://SCALA11GR2/HO	DERIVER_STATUS	Column	/AlphaNumeric_SpecialCharacter		
Oracle_Test://SCALA11GR2/HO	ID	Column			
Oracle_Test://SCALA11GR2/HO	CATEGORY	Column			
Oracle_Test://SCALA11GR2/HO	ORACLE_ACCESSCONTROLVI	Column			
Oracle_Test://SCALA11GR2/HO	CITY	Column			
Oracle_Test://SCALA11GR2/HO	FATHERS_NAME	Column			
Oracle_Test://SCALA11GR2/HO	COLUMN1	Column			
Oracle_Test://SCALA11GR2/HO	EMP_TYPE	Column	/AlphaNumeric_SpecialCharacter		



# Deployment at enterprise scale

# Hot Backup

- Take a backup without shutting down the catalog service
- CLI based (infacmd)
- Rest API based

GET	/2/catalog/jobs/backup	Backup Jobs
POST	/2/catalog/jobs/backup	Submit job to get catalog backup.
GET	/2/catalog/jobs/backup/{id}	Backup Job Configuration
PUT	/2/catalog/jobs/backup/{id}	Change backup job config
DELETE	/2/catalog/jobs/backup/{id}	Delete job to collect catalog backup
GET	/2/catalog/jobs/backup/{id}/result	Download backup zip file

# Rest API based log aggregation

- Most of the users do not have access to servers.
- Typically takes longer to get the right set of logs from customers. Helps with supportability
- Rest APIs (or UI <http://invrh74alpha05.informatica.com:9085/access/ui/index>)

GET	/2/catalog/jobs/serviceLogs	Job detail of service log jobs
POST	/2/catalog/jobs/serviceLogs	Submit a job to get service logs
GET	/2/catalog/jobs/serviceLogs/{id}	Service Logs Job Configuration
PUT	/2/catalog/jobs/serviceLogs/{id}	Service logs job config
DELETE	/2/catalog/jobs/serviceLogs/{id}	Delete service logs job
GET	/2/catalog/jobs/serviceLogs/{id}/result	Download service logs
GET	/2/catalog/jobs/resourceScanLogs	Resource scan log jobs
POST	/2/catalog/jobs/resourceScanLogs	Submits a job to get Resource scan logs.
GET	/2/catalog/jobs/resourceScanLogs/{id}	Resource scan log configuration
PUT	/2/catalog/jobs/resourceScanLogs/{id}	Change resource scan logs job config
DELETE	/2/catalog/jobs/resourceScanLogs/{id}	Delete resource scan logs jobs

Monitoring Status .

### Jobs Monitoring

Job Type	Status	Created on	Start Time	End Time	Result	Log URL
resourceScanLogs	Running	1589914589682	1589914589709	0	<a href="#">Download</a>	<a href="#">Log Location</a>

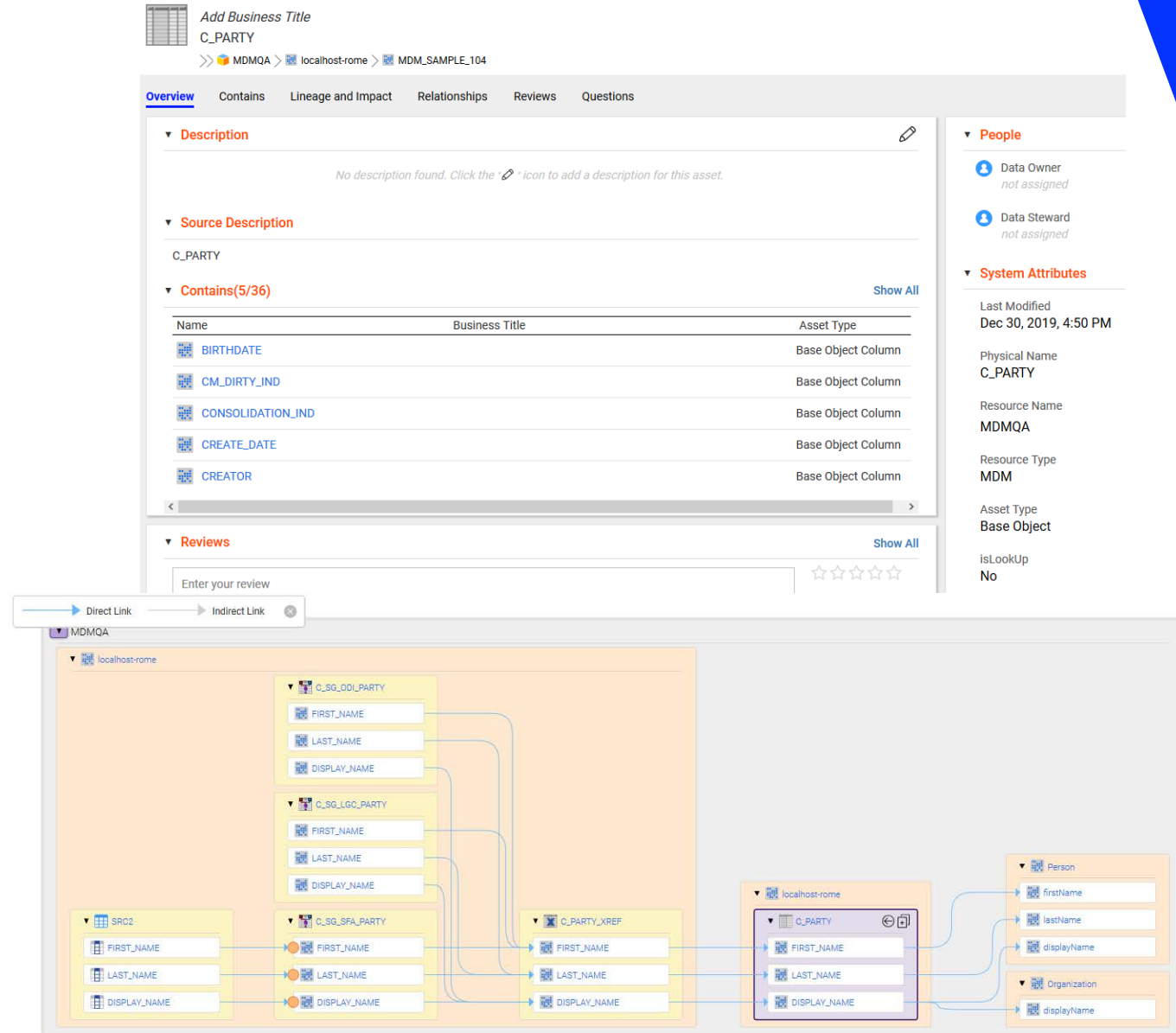
« Prev 1 Next »



New scanners in 10.4.1

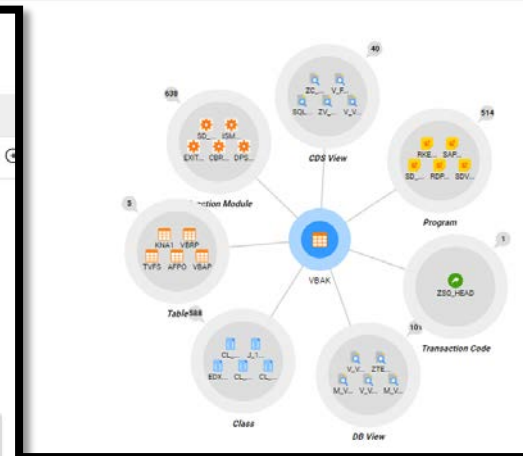
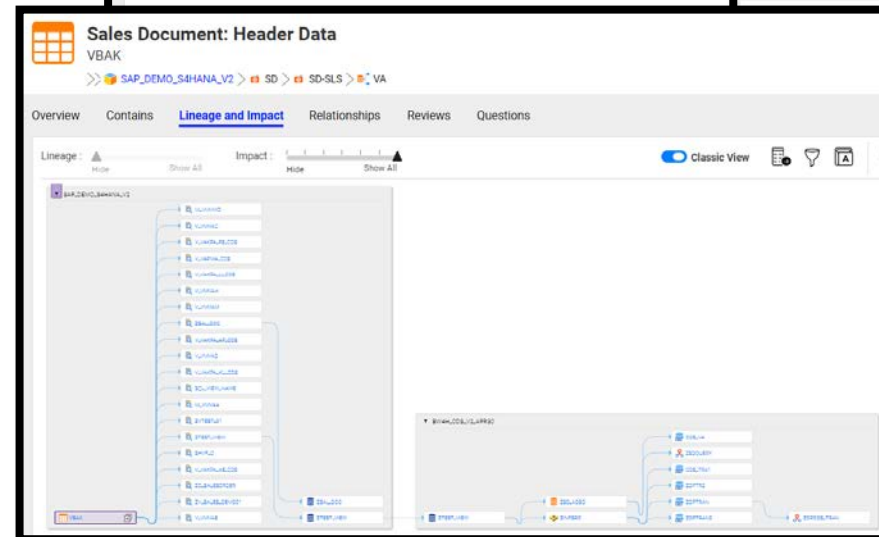
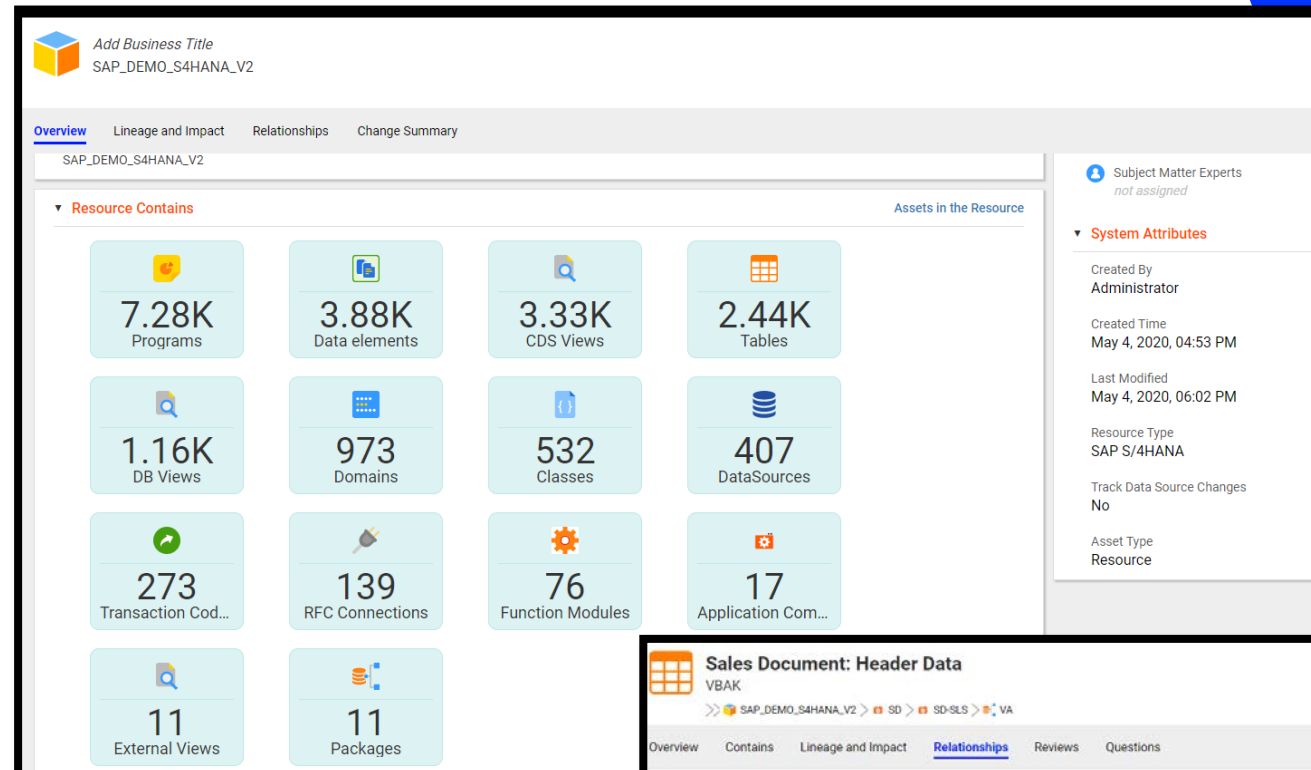
# New MDM Scanner

- Based on API (work both for on-prem and hosted MDM)
- Extract Base objects, Landing and Staging tables including field level information and attributes
  - Data Type
  - Nullable, Trust, system column indicator
- Internal lineage from landing tables all the way to business entities
- Detailed lineage from platform applications



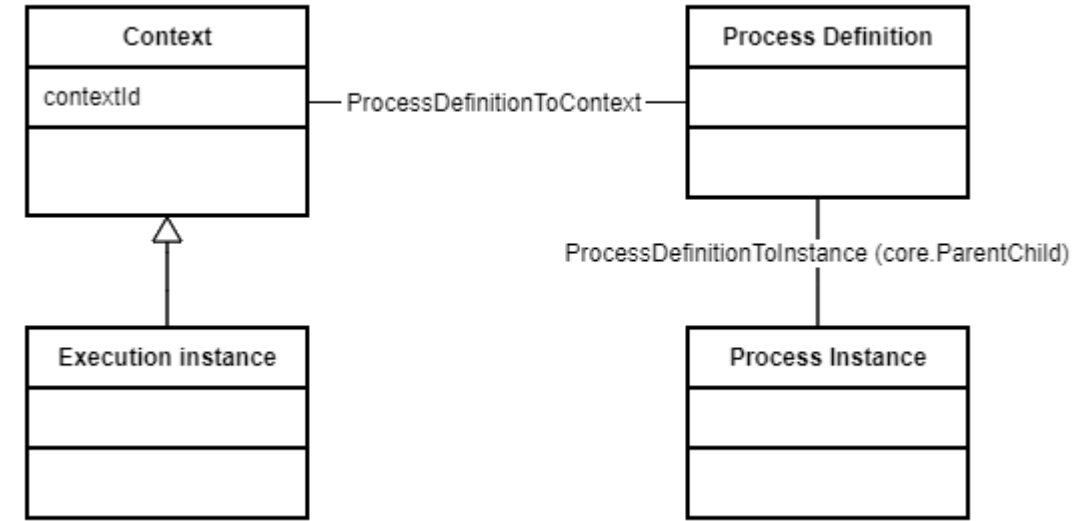
# SAP ECC and S/4 Hana (Tech Preview)

- Extract SAP objects with Complete business descriptions
- View to table lineage
- Relationship with transaction code, programs, Function Modules
- Direct lineage with SAP BW
- Support for Load Balancer connection



# Contextual lineage support for custom scanner

- Allow representation of lineage for the same object with multiple execution instance
- New **Process flow package** to derive custom model from
  - Abstract classes **Process Definition** and **Process Instance** to define custom classes
  - Process Definition and Process Instance derived classes will have a special handling in the UI
  - Lineage link are attached to the process instance
  - **Context** class to allow naming and description of execution instance



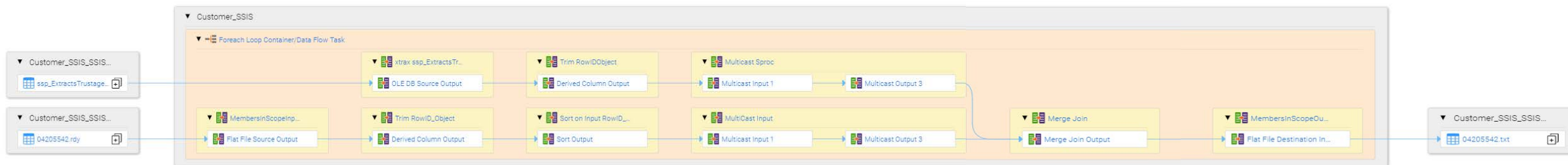
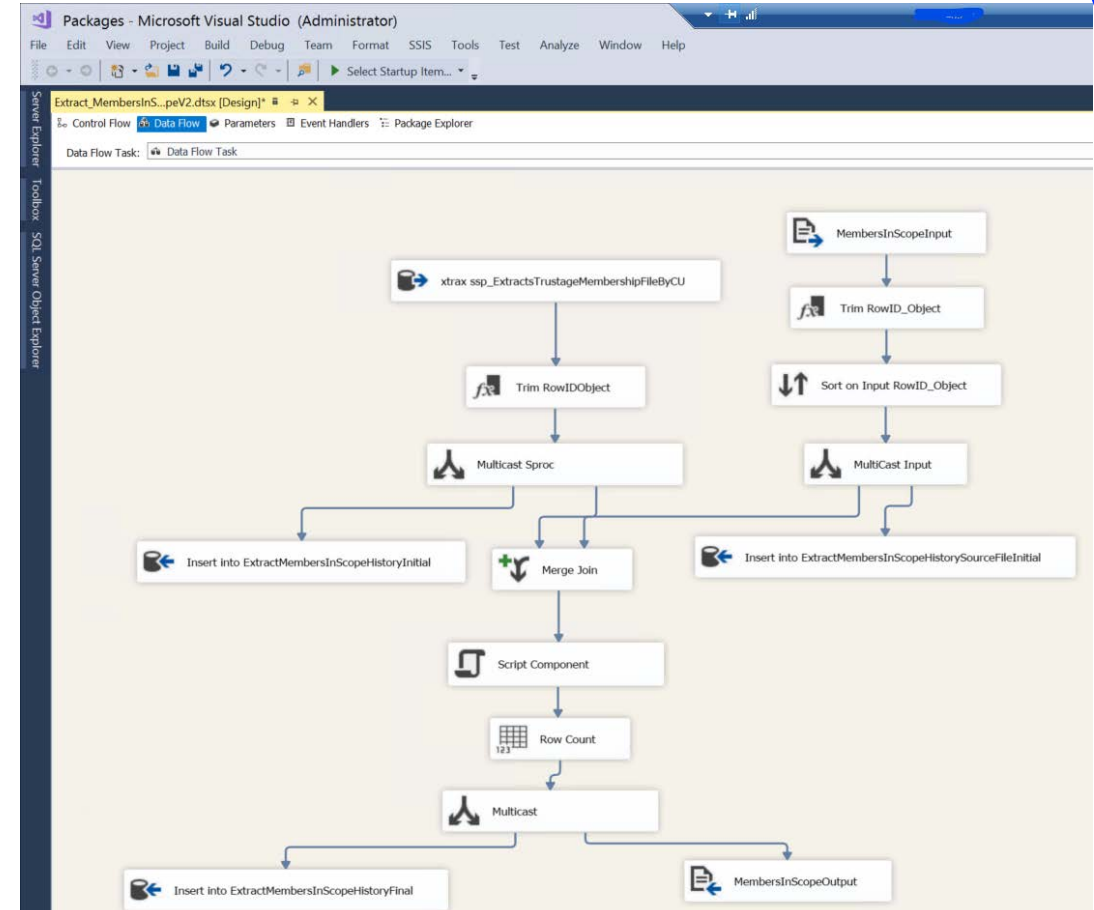
The screenshot shows the Informatica UI for a Stored Procedure Definition. The breadcrumb path is my\_SP\_1 > Database0 > Schema0. The 'Overview' tab is selected. The 'Context' section shows a table with two rows of lineage information.

Name	Source(s)	Source Type	Target(s)	Target Type
Context0	SRM1_D0_I0_TABLE2 SRM1_D0_I0_TABLE1 SRM1_D0_I0_TABLE0	Table Table Table	TGTM1_D0_I0_TABLE0	Table
Context1	SRC100_D0_I0_TABLE27	Table	TGTM1_D0_I0_TABLE0	Table

# Scanner enhancements in 10.4.1

# SSIS Scanner enhancements

- Support for transformation logic (detailed lineage)
- Support for metadata extraction from SSIS DB
- Field level lineage for flat files
- Control lineage links extraction




# MicroStrategy Scanner enhancements



- Retain enrichment across loads
- Incremental load
- Lineage simplification
  - Page level containers removed (page header, page footer, page body...)
  - Grouping within reports removed (added as attribute)
- Lineage available at report level
- Xdocs partitioning to avoid large memory requirements

# Reference object for Hive and MS SQL Server

- Before: External database object were created when not part of the object scanned
  - Table part of another database queried in a view (or stored proc)
  - Table part of another schema not part of the scan queried in a view (or stored proc)
- After: External database content will be created as reference objects
  - Allow stitching DB resource to DB resource
  - New connections available in connection assignment page

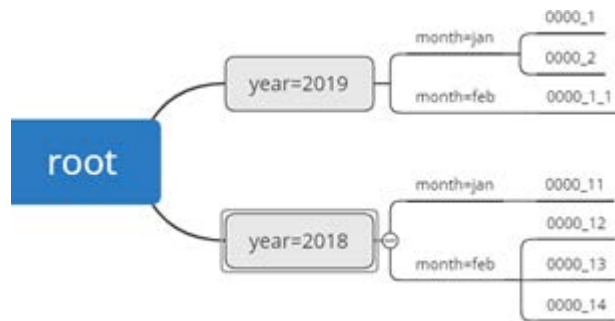
<input type="checkbox"/> Resource Name	Connection	Resource Type	→ Schema/End-point	Resource Name	Resource Type	Assignment Type
 <input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
▼ <input type="checkbox"/> WorldImporterDB						
<input type="checkbox"/>	External.sys	Microsoft SQL Server	→ WideWorldImportersDW	sql_WideWorldImportersDW	Microsoft SQL Server	In Progress



# Cloud storage supportability enhancements

# Partitioned and similar schema file detection

- For Parquet file formats



## ▼ Source Metadata

☐ Enable Source Metadata

## Additional properties

File Types\*:

☐ All ☒ Select

Specific File Types\*:

parquet text

Enter File Delimiter:

Other File Types:

☐

Treat Files Without  
Extension As\*:

Parquet ▼

Enable Partitioned File  
Discovery:

☒

First Level Directory:

All

Include Subdirectory:

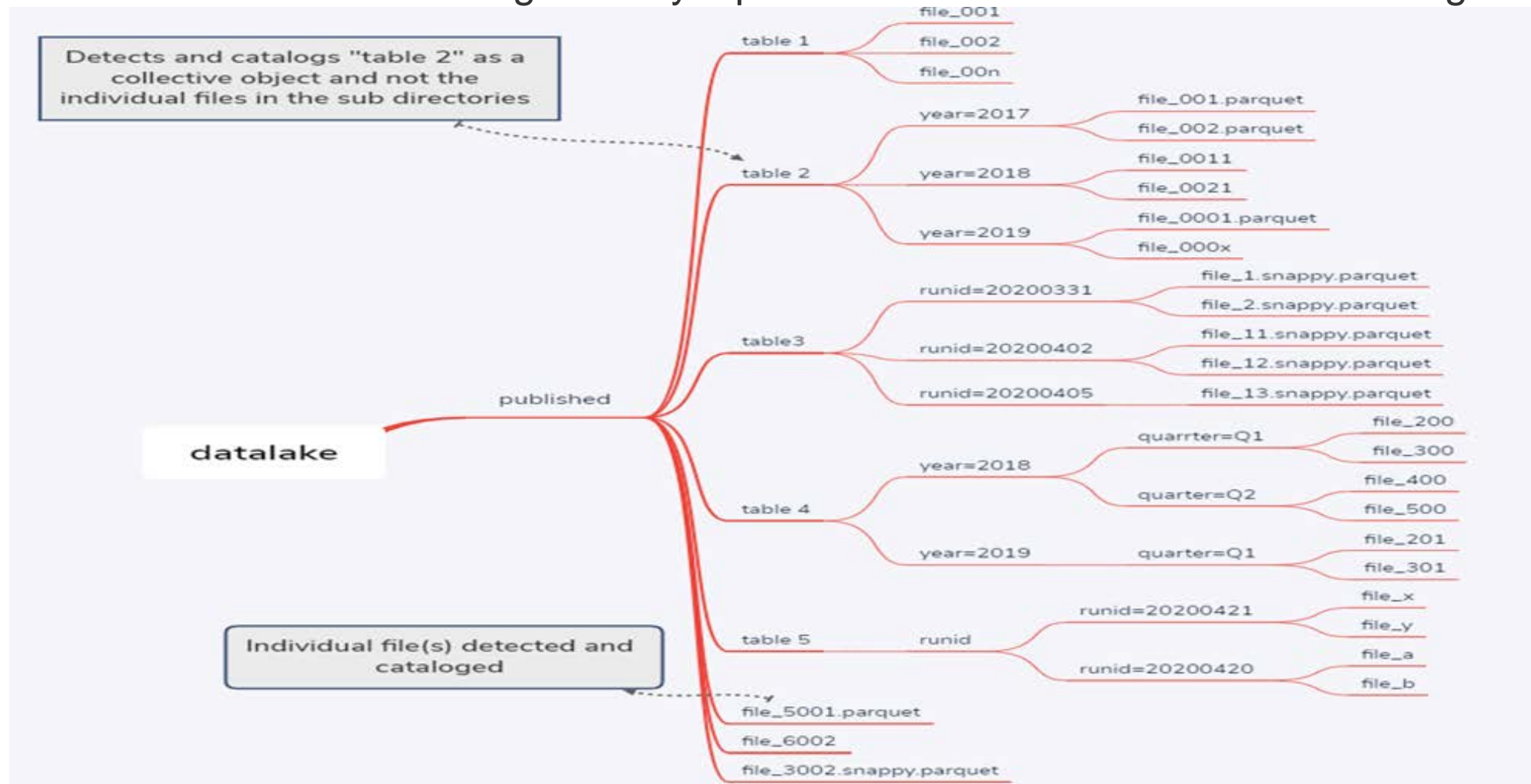
☒

Case Sensitive:

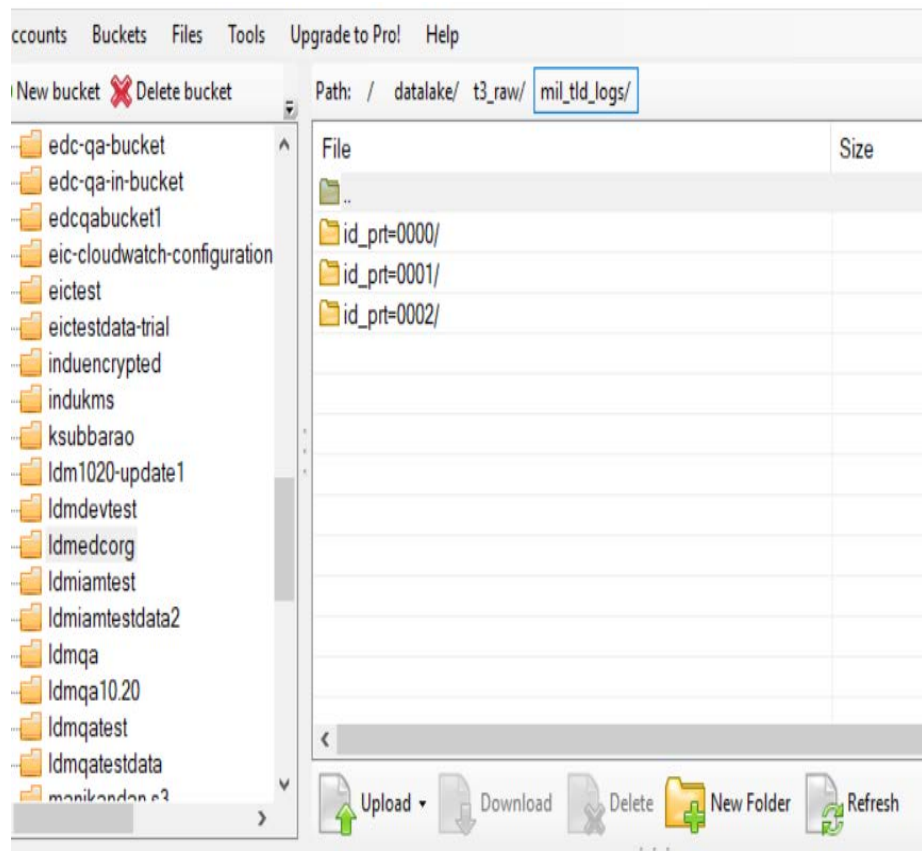
☒


# Parquet file profiling on non-hdfs file systems

- EDC can detect and catalog hive-style partitioned files on AWS s3 and ADLS gen 2











# What gets catalogued ? – partitioned files



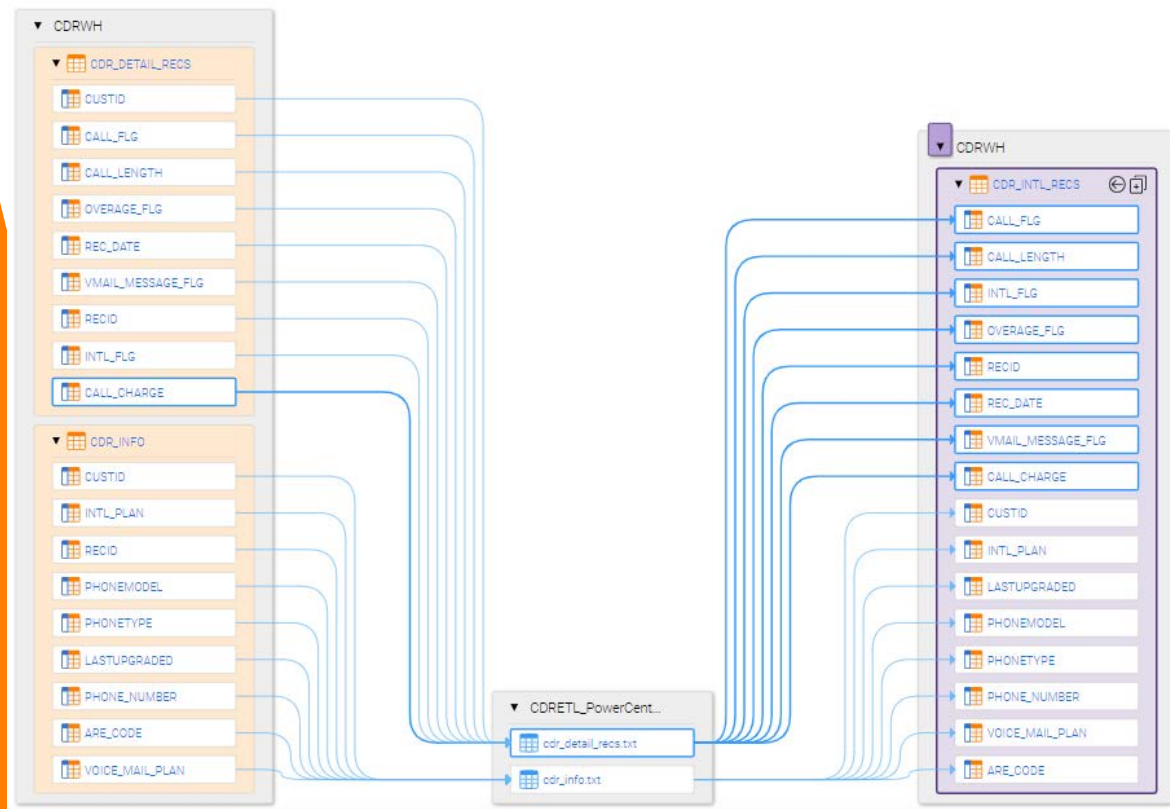
 Add Business Title  
mil\_tld\_logs  
>> s3\_partition\_detection > amazons3 > datalake > t3\_raw

Overview **Fields** Lineage and Impact Relationships Reviews Questions

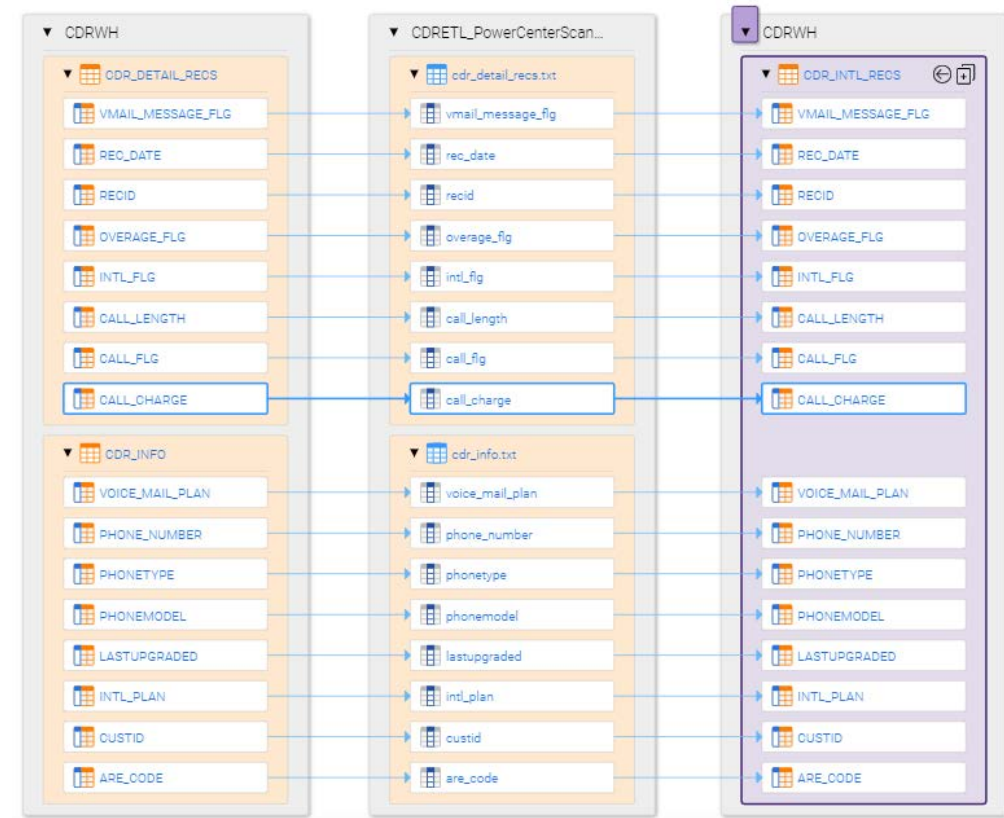
Name	Business Title	Data Domains ?	Null   Distinct   Non-Distinct %	Source Data Type ? Inferred Data Types
 comment				string
 creation_date				int96
 creator_user_id				string
 criticality				string
 diff				string
 end_dt				int96
 expected				string
 id_prt				int32

# Support for field level lineage for flat file

Before 10.4.1

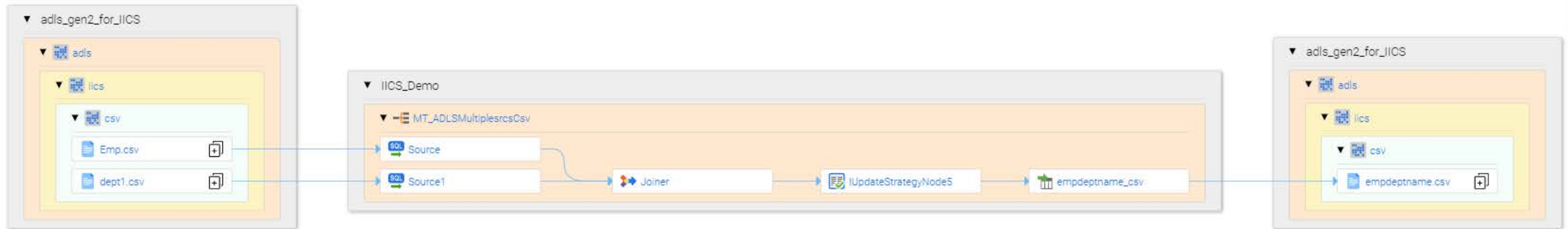


With 10.4.1



# IICS: File lineage on Cloud storage

- Support for file lineage for cloud storage in IICS scanner





# Questions?





Thank You!

---