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

Enterprise Data Catalog

Knowledge Graph + AI/ML

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

- Analytics
- Data Governance
- Master Data Management
- Cloud Modernization
- Metadata Intelligence
- Data Integration
- Data Quality

- Discovery
- Profiling
- Lineage
- Impact Analysis
- Semantic Search
- Domain Discovery
- Similarity Clustering
- Business Term Association
- Relationships
- Business Context
- Glossary Integration
- Custom Annotations
- Reviews/Ratings
- Questions/Answers
- Data Certifications
- Change Notifications
The Catalog of Catalogs: One Vendor, One Solution

Open APIs, Full Access

Enterprise Data Catalog + Compact Solutions

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
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

Broadest and Most Complete Metadata Connectivity
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

---

*Feature in Tech Preview*
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
DAA: Privileges

Three new privileges are added into existing CS privileges

<table>
<thead>
<tr>
<th>Permissions: v</th>
<th>DAA Privileges: &gt;</th>
<th>Dashboard Management: Visualize</th>
<th>Report Management: View and Download Raw Data</th>
<th>Admin Management : Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Dashboards</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Download Dashboard Chart Data</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Create Reports &amp; Download Report Data from Monitor tab</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Edit Asset Value Calculator</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>DAA application configuration</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

*Note:*

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

Informatica
DAA Architecture

Enterprise Data Catalog
- Scanner/Ingestion
- Catalog UI
- Auto Assignment / Enrichments

REST API
- Event Capture Service

Event Staging Tables
- Raw Event Tables

Data Asset Analytics
- DAA UI
- REST API (Internal)
- Charts/Reports/Monitoring/Download

CRUD
- Event Tables, MetadataTables, Views

Data Access Object Layer
- Event Processing Engine
- Reporting Service
- Chart Service
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
Key Upgrades in 10.4.1 (New)

- Asset Type Filter
- Filter by time periods instead of individual scans
- All changes shown chronologically by default
- Search for a specific asset by name
- Add Asset Type column
- Export Change Summary to CSV file
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 Library**
  - **Read**
    - View resources in library
  - **Scan**
    - Cancel, Rollback and Purge of resources
  - **Write**
    - Edit resource configuration
    - Delete Resource

- **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

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

**Resource Library**
- **Read**
  - Users can view list of connections extracted by resource
- **Write**
  - Users can generate Missing link report

**Connection Assignment**
- **Read**
  - Users can view list of connections extracted by resource
- **Write**
  - Users can assign/unassign connections

**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

© Informatica. Proprietary and Confidential.
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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Asset Name</th>
<th>Matched Term</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle Finance</td>
<td>FIRST DATE ACTIVATED</td>
<td>DATE FIRST ACTIVATED</td>
<td>100%</td>
<td>Out-of-order Term matched</td>
</tr>
<tr>
<td>Oracle Finance</td>
<td>LAST REJECTED DATE</td>
<td>DATE LAST REJECTED</td>
<td>100%</td>
<td>Out-of-order Term matched</td>
</tr>
</tbody>
</table>
## Auto Business term association – example (Tf-Idf)

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

Sample_Custom_DataDomain

Data Domain

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

General

Name*: Sample_Custom_DataDomain

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: ^([a-zA-Z0-9\._]+)@[a-zA-Z0-9\._]+ X

Column Name Rule: \"mail\" X

Trump Rules:  
Both metadata and data conditions match

Data Conformance:

Minimum conformance:

Auto accept if more than:
Bulk curation of custom attributes

Select Asset Types
- Select Asset Types
- Column
- Database
- Foreign Key Constraint
- Primary Key Constraint
- Resource
- Schema
- Table
- UDP
- UDPs
- Unique Constraint
- View
- View Column

Choose attributes to export
- Business Description
- Data Owner
- Data Steward
- Display Name
- URL

Include
- Data Domains
- Business Terms
### Bulk curation of custom attributes – CSV Example

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>name</td>
<td>classType</td>
<td>bulkimport-DataDomainsAccepted</td>
<td>bulkimport-DataDomainsInferred</td>
<td>bulkimport-DataDomainsRejected</td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>PROVISION3</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>PROVISION2</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>PROVISION1</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>CONTACT_NO</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>EMAIL</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>PROVISION1</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>PROVISION3</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>PROVISION2</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>FIRST_NAME</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>STRING_APP</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>AGE</td>
<td>Column</td>
<td></td>
<td>/AlphaNumeric_SpecialCharacter,Age/</td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>BLOOD_GROUP</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>AMBULANCE_STATUS</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>DOC_NAME</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>MED_NAME</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>EMP_ID</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>ORACLE_ACCESSCONTROL</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>DERIVER_STATUS</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>ID</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>CATEGORY</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>ORACLE_ACCESSCONTROL1</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>CITY</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>FATHERS_NAME</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>COLUMN1</td>
<td>Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle_Test://SCALA11GR2/HO</td>
<td>EMP_TYPE</td>
<td>Column</td>
<td>AlphaNumeric_SpecialCharacter</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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)
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

© Informatica. Proprietary and Confidential.
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

<table>
<thead>
<tr>
<th>Resource Name</th>
<th>Connection</th>
<th>Resource Type</th>
<th>Schema/End-point</th>
<th>Resource Name</th>
<th>Resource Type</th>
<th>Assignment Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>WorldImporterDB</td>
<td>External sys</td>
<td>Microsoft SQL Server</td>
<td>WideWorldImportersDW</td>
<td>sql_WideWorldImportersDW</td>
<td>Microsoft SQL Server</td>
<td>In Progress</td>
</tr>
</tbody>
</table>
Cloud storage supportability enhancements
Partitioned and similar schema file detection

- For Parquet file formats
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
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!