Oracle® Data Foundation Cloud Service

Data Platform

Release 24C
G13191-01
August 2024

Oracle Data Foundation Cloud Service Data Platform, Release 24C

G13191-01

Primary Authors: (primary author), (primary author)

Contributing Authors: (contributing author), (contributing author)

Contributors: (contributor), (contributor)

Contents

4 Data Pipeline

Index

1
Data Catalog Overview

Data Catalog provides the Data Model for the Financial Services Industry. The key capabilities of the Data Catalog are as follows:

Catalog Browser: Data Catalog consists of a Viewing Framework. The Data Catalog Browser API-based Interface allows you to view the Data Catalog Components. The Catalog Browser enables the Users to go through the Data Catalog Contents and view the Business Terms by Domain, Subject Areas, and Entity. The Catalog Browser also shows the Properties of Business Terms, Contextual Definition in a User-friendly language, List of Vaues, Data Sourcing Components, and Data Quality Rules.

In-built Data Modeling Capabilities: Data Catalog contains all the required information to establish services underlying the Data Model, which is Entities, Attributes, and the relationship between Entities and Attributes. Similar to any Entity Relationship Modeling Tool used for this purpose, the Data Catalog can build, manage and hold Data Model for the deployment purpose.

Data Quality Checks Rule: Data Catalog Contents include Data Quality Checks Rule so the system incoming data can be verified and validated.

Catalog Extension: Data Catalog supports a Framework to extend the Data Catalog called as the Data Catalog Extension or catalog Extension. The Catalog Extension allows user to extend the Seeded Catalog Contents to support a new or client-specific business use cases. You can add new Business Terms or customize the existing definitions when the Business Term is enforced by the external entities.

Comprehensive Coverage: Data Catalog provides a collection of comprehensive Business Terms across the Business Lines and Use Cases.

Data Movement: Data Catalog provides the mechanism of ‘Stage to Standardize to Process’ to move the Data along to the Result for analytical consumption. The Catalog Services are accessed through API calls used by the Data Services Module to move data.

Data Catalog comprises of elements called Business Terms supporting the business needs of the banking and financial services industry across finance, risk, and regulatory compliance functions. Entities and Subject Areas simplify your understanding of the business relevance of an Element and the associated Data Definition grain.

Data Platform Deployment

In the OFS Cloud Services, the Service-specific Functions (For example, Accounting Integration, Balance Computation, Reconciliation, and so on) and their relevant Business Terms are deployed.

The Deployment Process includes the Data Model upload, metadata deployments and creation of synonyms, and SLA (Sub-ledger Application) related ADI refresh tasks. The Data Platform is the Data Model, which is grouped into Domain and Subject Areas. All the necessary Entities are mapped to the Subject Areas for the selected Domain.

Data Catalog Key Capabilities

Data Catalog provides the Data Model for the Financial Services Industry. The key capabilities of the Data Catalog are as follows:

- Catalog Browser: Data Catalog consists of a Viewing Framework. The Data Catalog Browser API-based Interface allows you to view the Data Catalog Components. The Catalog Browser enables the Users to go through the Data Catalog Contents and view the Business Terms by Domain, Subject Areas, and Entity. The Catalog Browser also shows the Properties of Business Terms, Contextual Definition in a User-friendly language, List of Vaues, Data Sourcing Components, and Data Quality Rules.
- In-built Data Modeling Capabilities: Data Catalog contains all the required information to establish services underlying the Data Model, which is Entities, Attributes, and the relationship between Entities and Attributes. Similar to any Entity Relationship Modeling Tool used for this purpose, the Data Catalog can build, manage and hold Data Model for the deployment purpose.
- Data Quality Checks Rule: Data Catalog Contents include Data Quality Checks Rule so the system incoming data can be verified and validated.
- Catalog Extension: Data Catalog supports a Framework to extend the Data Catalog called as the Data Catalog Extension or catalog Extension. The Catalog Extension allows user to extend the Seeded Catalog Contents to support a new or client-specific business use cases. You can add new Business Terms or customize the existing definitions when the Business Term is enforced by the external entities.
- Comprehensive Coverage: Data Catalog provides a collection of comprehensive Business Terms across the Business Lines and Use Cases.
- Data Movement: Data Catalog provides the mechanism of ‘Stage to Standardize to Process’ to move the Data along to the Result for analytical consumption. The Catalog Services are accessed through API calls used by the Data Services Module to move data.

Data Catalog Architecture

Data Catalog is at the center of sources and uses of data. Along with providing comprehensive understanding of data through the available metadata about sources and uses, it is necessary to be able to enrich or modify as required. In this direcrion OFS Data Catalog is designed to support integration with Data Sources and downstream uses, which means predefined for known uses and extendable for the unknown. The Data Catalog by Subject Area can be further segregated as Sources, Results (calculated), and Master Data that are sourced and seeded, which helps to plan the sourcing and supplying of data based on the Enterprise needs. Considering the need to serve a conglomerate of Users, the Data Catalog is embedded with a sound Governance Framework assuring the security and privacy.

Data Mapping Illustrations

The Data Foundation Cloud Service provides various features for best-class user experience. One of the key activities in Data Foundation on-prem product was the data mapping between core-banking and data foundation. This was one of the foundation activities to implement OFSAA applications. As data foundation had a rich set of data model, the data mapping exercise was very complex and expert professional services were mandatory to complete this activity. In Data Foundation Cloud Service, there is a feature Data Mapping Illustrations getting introduced to give a real time experience on this complex activity. The Data Mapping

Illustrations will enable the users to experience the entire data mapping experience through various real life banking use cases.

The Data Mapping Illustrations cover various real-life events happening in banking and how each event creates data. This feature helps users to reduce the time and effort required to map core-banking systems to data warehousing systems.

Field	Description
Domain	Banking
Subject Area	Product Processors
Party Type	Corporate, Individual, Retail
Product Type	Bills Contracts, Borrowing Contracts, Credit Cards, Current or Saving Accounts, Forward Contracts, Future contracts, Investment Contracts, Leases Contracts, Letter of Credit, Loan Contracts

Business Terms

A Business Term is a Functional Keyword that represents a unique functional aspect of the Financial Artefacts (Entities and Attributes). There is a business meaning to each Financial Artefact, and therefore, a business name is created in the form of a Business Term. Then the Business Term is used to find and fit into a purpose of different contexts. Therefore, Business Terms are expressions of the Participants/Actors, their Activities, and requirements conveyed in common business.

The approach to arrive at the Logical View of the Physical Model has inverted in the Data Catalog. Contrary to the previous practices, first the Functional Keyword or Artefacts are identified, collated, and then expressed in common business language, and then the corresponding Physical Model is developed.

Multiple Business Terms are a part of each Subject Area. A Business Term is a singular term irrespective of multiple places it appears in and the Business Term will have a recognizable naming pattern.

Metadata is attached to the Business Terms.

The Content Structure allows you to define relationship between related Business Terms for better discovery and usage in the aspect of sourcing.

Properties of Buisiness Terms:

- References
- Related Terms
- Contextual Definition
- Data Quality
- List of Values
- Use Cases
- Comments

Business Term Glossary

For each Business Term, there is a corresponding Glossary. The Glossary explains the meaning and purpose of each Business Term, which is the context. The Glossary consists of the information such as description, usage, sample values, and usage examples of the Business Terms.

Business Term List of Values

Data Platform defines and seeds the List of Values for all the Business Terms that are Flags. The List of Values help you with data sourcing and creating Data Quality Checks.

Related Business Terms

Platform has a provision to store the metadata to represent the relationship between two Business Terms along with the type of relationship.

Entities

Data Platform has a comprehensive list of entities to support a very wide range of analytical business use cases across Banking and Insurance domains. These entities are mapped to the subject area and the domain which helps in deploying the model, either completely or selectively, based on the services the user has subscribed.

Entity Type

Entities in the model are grouped into the following types based on data layers they are tied to:

- Download or Staging
- Result
- Dimension
- Preparation
- Integration

Preparation and Integration Entities or layers are internal to a product. A Preparation layer contains entities used for processing while an Integration layer contains placeholders for other engines or applications to publish their processed output.

Stage Entity

Stage Entities are a part of the Unified Data-Sourcing layer for data for a wide variety of analytical needs. The Staging layer faces the operational (OLTP) and Front Office Systems of a bank. It consists of entities to capture key business and operational data from these systems, which is then processed by one or more Analytical Engines.

Since these entities are designed to facilitate loading efficiency, they are denormalized. Typically, this data is extracted from source systems and placed in the object store and then loaded to Staging Entities using data services.

Result Entity

It is a Unified Reporting or a Consumption layer. Analytical results can be simple to complex, contain quantitative and qualitative measures of a bank's Risk, Compliance, Customer and

Financial Performance. The Reporting layer is a Dimensional Data Model spanning these key analytical functions. It forms the foundation of OFSAA Business Intelligence Applications but can clearly be used as the result data store for any equivalent engines and processes provided by other vendors, or custom built solutions. By providing a Single Data Repository for reporting needs, this layer provides a foundation for departmental as well as Cross-Departmental and Cross-Functional Reporting. Additionally, the schema of the reporting area is specifically built for Financial Services Analytics. As an illustration, the reporting area has a 'Run Dimension' that is shared across all BI/reporting use cases. Similarly, the aggregation of shared measures/ reporting into a common set of output structures (Account Summary) allows for Cross-Functional Reporting, which is increasingly the norm in Financial Institutions.

It is also the layer where all the input data sourced by the user through Staging Entities are persisted and available for reporting and consumption.

Along with the definition or structure of these entities, the catalog also provides mapping metadata to populate these entities from sourcing to PMF Data Pipeline.

Currently, this layer has entities to persist Journal, General Ledger Account, Transactions and Contract Data to cater to the needs of the DFCS solution.

Note:

For any existing extractions defined on Journal, Journal Header, and General Ledger Account Data, re-select the grain, entity, and attributes and save. This will enable extraction from the Persistence layer.

Note:

This is applicable only for the Banking domain. The PMF pipeline Account Load Run Map Population execution should have completed before executing the below step. This step is required for Common Account Summary and Transaction Summary data population.

Based on the domain (Life Insurance, Annuity, Retirement, General insurance (Property and Casualty), Health insurance, and Reinsurance issued) selected in the Add Domain page under Administration, domain specific Result Area Entities are populated.

The user must manually execute a PMF Pipeline Result Area Entities Data Population to populate data in the Result Area tables.

Note:

This is specific to Banking domain. The following Target tables Accounting Entries, Accounting Entries Header, Currency Exchange Rates, General Ledger Data, Common Account Summary and Transaction Summary are used for data population in the result area.

The user must manually execute a PMF Pipeline PPN to Result Area Entities Population to retrieve the data in the Result Area tables. The Decomposed Journals to Accounting Entries data population is performed through this pipeline.

The user must manually execute a PMF Pipeline Data Service Result Area Data Population to retrieve the data in the Result Area tables. The persistence of day and period-to-date movement information is performed through this pipeline.

Properties

Entity consists of a set of Attributes. Each Attribute is mapped to a unique Business Term. Each Attribute derives its characteristics such as Data Type, Constraints, and so on from the Business Definitions.

Granularity

Data Catalog defines granularity of each of the Entity it holds.

V_GRAIN_ID	V_GRAIN_NAME
DATE	Date
ACCT	Customer Account
ACCT_TXN	Customer Account Transactions
ACCT_ENT	Accounting Entries
GL_DATA	General Ledger Data
EXCHANGE_RATE	Exchange Rates
PARTY_CONSENT	Party Consent

For example, as the Loan Contracts Entity contains information at the Account Level, the grains of Loan Entity is defined as ACCT.

Personal Identifiable Information (PII)

Use the Personal Indetifiabnle Information tab to Manage PII data lifecycle.

On the **Attributes** Tab, you can search and see the list of Attributes of the Entity. Select the required Attribute to see its Details and List of Values. The Details List displays the Data Type and Description of the selected Attribute. In the List of Values Section, you can sort the Value and Description information by ascending or descending order. The Attribute, which is a PII (Personally Identifiable Information), is marked with a Red symbol.

2 Data Integration

Data Integration enables data exchange between OFSAA and external systems through a logical abstraction of the OFSAA Data Foundation (Financial Services Data Foundation and Insurance Data Foundation) exposed as Application Data Interfaces (ADI). External Data Sources (EDS) and External Data Descriptors (EDD) are defined via the DIH user interface, using which you can map EDDs to ADIs to form Connectors.

External Data Descriptor

External Data Descriptor External Data Descriptors (EDD) are definitions of specific data content from External Data Sources (EDS). Each EDS may have a number of EDDs defined against it. EDD definitions can be used for ingesting data into DFCS Data Foundation or extracting data from it. DFCS also supports the usage of control information while ingesting data into the Data Foundation. Such control information may also be defined as EDDs.

1. From the **Subledger Applications** page, select the required SLA.
2. Click **Data Ingestion** to define External Data Descriptors, Connectors, and view data. You can manage data ingestion by configuring sources, mapping to catalog, and required transformation.
3. Click **External Data Descriptors**. The Summary page is displayed.
4. Use the search functionality if you are looking for an existing EDD.
5. Click **Add** to configure additional EDDs. Provide a name and description, and select the Datastore. The Data, Transformation, and Control tabs appear based on your selection. For details, see section Defining an External Data Descriptor.
6. Click **Delete** to delete an existing EDD. For details, see section Deleting an External Data Descriptor.

Ingest Connector

To create a Connector for ingesting data into DFCS, follow these steps:

1. On the **Data Ingestion** page, click **Connector**.
   The connectors which are defined are listed here.
2. Click **Add**.
   The **New Connectors Definition** window is displayed.
3. To define a connector, you must have a source with EDD and a target, which is ADI.
4. Click **Source** to select the required EDD from the list defined in the default EDS of the ERP File Extracts.
5. Select the default EDS-ERP File Extracts File Type EDD and drag it to the canvas.
6. Click **Target**.
7. Click **Search** to search for a particular ADI.
8. Select the required ADI. Drag it to the canvas and then link the input and output nodes.

9. Click the input white circle.
The anchor symbol is displayed. Drag and drop the line to link it to the required component.
10. At any given time, you can right-click the node to either delink or remove links/outline or delete a node.
11. To edit or view the properties, on the **Connector** window, click **>**
12. In **Connector Details**, enter the name and description for the connector.

> **Note:**
>
> The metadata key will be available in the **Connector Properties** tab.
>
> In **View** or **Edit** mode, you can copy the Metadata Key of the Connector using the **Copy** icon.

13. In **Pre Load Options,** select the truncate option to be defined in the target.
To remove data from the table as per the truncate option specified, select **Truncate**.

    Select **No**, if you do not wish to truncate the table before loading.

    The service will display an option to perform Partial Truncate. This option is not enabled as of 22C and should not be chosen.

    > **Note:**
    >
    > For multi-target loads, the truncate type must be the same for all targets. However, truncate expression may vary.

    - Select **Full Truncate** to fully truncate. Here no expression is required. If you want to truncate the entire table, the Full Truncate option must be selected.
    - Select **Selected Rows** to truncate on the selected expressions. If you remove specific rows, the selected rows option must be selected. Specify the filter condition for the rows to be deleted. Specific rows are removed from the table before load
        – Click **Edit** to filter the selected rows.
        – Select the required entity and click **Validate**. This validates the expression.
        – Click **Ok** once the expressions are selected.
14. In **Properties**, enter the value in the **Target Rejection Threshold** field. You can define the value in number or percentage of error records.

Data Pipelines

Process Orchestration is a design and execution service that enables process pipeline developers to implement pipelines modeled by business analysts. Process pipeline developers use this framework to orchestrate Run Pipelines within DFCS, and also to design the artifacts that are used in the pipelines.

The Process Modeller and the Process Monitor are two key parts of Process Orchestration. The Process Modeller is used to model pipelines. It aids in representing the various artifacts required for modeling and provides implementation details of the DFCS process artifacts. The Process Monitor is used to monitor instantiated pipelines of DFCS.

DFCS supports orchestration of the **Run Pipeline**, which is any orchestrated pipeline consisting of DFCS tasks and service calls that run within the DFCS context.

Process Modeller

The Process Modeller page displays existing pipelines along with their details such as: process ID, process name, process description, version, instance, application, and last modified date, and last modified by.

On this page, you have options to:

1. Click + to create a new pipeline.
2. Click the process name link to launch and edit the pipeline.
3. Launch the process in a new window.
4. Delete a pipeline.
5. Click the **Menu** button and do the following for each pipeline:
   - View a process flow
   - Copy a process flow
   - Monitor the pipeline in the **Process Flow Monito**r window
   - Execute a pipeline
   - Apply a filter condition to a Run pipeline

Use the **Search** grid to search for a pipeline by providing a keyword containing the process ID, process name, or description. Click the Reset icon to reset the search fields.

You can narrow down your search results by additionally selecting the Pipeline Filter options to filter pipelines based on pipeline type. Example: To view only Run Pipelines, click inside the Pipeline Filter list, and select Run Pipeline. Remove the other pipeline types if already selected and click Search.

You can also sort the pipelines on this page based on the Process ID, Process Name, or Application. Click the Sort By drop-down and select the sort criterion.

Click to go to the **Process Monitor** window.

Data Extraction

You can access data extraction definitions by accessing the DFCS User Interface. On the Home page, from the Left Hand Side menu, click Data Extraction. This displays basic details of the Extract definitions. DFCS supports user-configured extraction of data from Staging and Results entities via Connectors.

> **Note:**
>
> If the Extraction Entity contains the As Of Date column, the data will be filtered based on the PMF execution date. When no date is selected during PMF Execution, the data will be filtered based on the current date.

To extract all the active nodes, the default filter applied for the dimension extract is "Execution Date between Record Start Date and Record End Date" instead of "As of date".

As the PMF execution date will be picked up as the default date for the extraction, you are not expected to provide a filter on As of Date during connector creation.

Adding a Data Extraction

1. From the DFCS Home page, select Data Extraction.
   The Summary page is displayed with a list of data extraction definitions.
2. Click **Add**.
   The **Specify Details** window appears.
3. Enter the **Name** for the Data Extraction.
   For Example: Extract CASA.
4. Enter the **Description** of the Data Extraction.
   This field is optional.
5. Select the **Grain** from the drop-down list.
   For Example: Customer Account.
6. Select the **Entity** from the drop-down list.
   For Example: Casa

   Enter the required filter expression for the Data Extraction. In addition, you can also add Parameters.

   This field is optional.

   For Example: [Casa].[Account or Contract Number] is not null.

   This filters out rows where the account or contract number is not null.

   > **Note:**
   >
   > For filtering data on Dimension Codes, you can specify member values for dimension surrogate key attributes only for the result table entities.

7. Click **Validate** to verify the correctness of the SQL Expression and click **OK**.
   The filter expression is displayed in the Expression text box.
8. Navigate to the **Select Attributes** page.
   The **Attributes list** displays the list of attributes that can be associated with the Data Extraction. The attribute list shows the combined attributes for all the tables selected.
9. Select the attributes and click **OK**.
   Select or unselect the attributes by clicking the respective icons. You can also drag and drop to rearrange the attributes in the Selected Attributes section.

   > **Note:**
   >
   > You can select a maximum of 998 attributes.

> **Note:**
>
> You can click Edit icon and rename the header attribute name, if required.

> **Note:**
>
> Post upgrade, attributes of existing result entity extract connectors must be de-selected and selected again to auto rename attributes and validation on duplicate columns in selected attributes.

For new connectors, attributes will be automatically renamed when selected. Auto-renaming of attributes to its corresponding dimension code attributes in the user interface is done only for Result table surrogate keys.

10. Navigate to the **Describe File** page.
11. In Specify Data file names, enter the file name. This is a mandatory field.
    For example: test_file_data.txt

> **Note:**
>
> The file extension must be either .txt or .csv.

12. Select Column delimiter from the drop down list.
    For Example: Comma
13. Specify the Text qualifier.
    For Example: Double quotes can be used, prefixed, and suffixed with text.
14. Select the file format.
    For Example: Fixed Length or Delimited.
15. Select the Record delimiter.
    For Example: Unix
16. Toggle the Header button if you want a header to be displayed in the extracted file.
    This field is optional.

> **Note:**
>
> The extracted file will show Result, Dimension, Integration, and Staging tables.
> Currently, Processing layer tables are not supported.
> In case two physical entities contain the same logical name, the order of preference for the entity will be in the following order:
>
> a. Result
> b. Dimension
> c. Integration
> d. Staging

17. Toggle the **Generate a control file** button if you want to generate a control file. The control file provides verification/validation information for use with extracted data files.
    This field is optional.

    In the Specify control file name field, enter the file name. This is a mandatory field. The file extension must be either .txt or .csv.

    The parameters you entered for Specify Data file names are copied to the Control file properties. You can modify based on your requiements. By default, the name is added as < Specify Data file names >_control.csv/.txt . For more details on fields such as Select file format, Select Column delimiter, Select Record delimiter, and Specify Text qualifier, see the above step.

    To add the controls, click **Add** button.

    Add the control name in the text box.

    Select the aggregation type from the dropdown. The supported options are Count, Avg, Sum, Min, and Max.

    Select the attribute from the drop-down. The Attribute field is disabled when the aggregation type is selected as Count. The numerical attributes supports the aggregation type Sum, Average, Min, and Max. The Date attribute supports the aggregation type Min and Max only.

    You can add multiple controls to the file by clicking on Add button.

    The Control file is available in the Archived Extract file.

    Note:The following are the behavior of the Control file in this release:

    There is no validations performed while generating a control file.

    Header will not be generated in case of control file during data extraction and you will not have an option to add the header.
18. Click **Save**.
    The Data Extraction Definition is created.

    Note:The maximum extracted archive file size must be less than or equal to 7.5 GB.
19. After the Data Extraction is saved, the PMF process needs to be configured and executed. For more details, see Executing Connector using Process Modelling Framework.
    The Data Extraction file is created.
20. Download the file from the File Operations screen. For more details, see File Operations.

File Operations

The file operation process enables you to add files that must be processed to the object store. You can upload any CSV, XLSX, and Text file that you wish to be stored in the object store, which can be processed later.

To upload a file, complete the following steps:

1. On the DFCS Home page, click the **My Profile** icon and select **Administration**.
2. Click **FileOperations**.
3. Click **UploadFile**.
   The **Generate Pre-Authenticated URL** window is displayed.
4. Enter the **FileName**.
5. Select the file format from the drop-downlist.

The supported file types are: CSV, XLSX, Text, and Zip.

6. Enter the file size (in bytes) as seen on the disk. To add more files, click **Add** and repeat the steps.
7. Click **Generate**.
This generates a PAR (pre-authenticated request) URL which is valid for 24 hours. This PAR URL is used to upload the file into the object store.
8. An entry for the PAR URL you just created appears on the File Operations page.
The search feature provides you multiple options to search for a file based on name, creation date, and so on. You can also sort the list by name and creation date in ascending or descending order.

Data Controls

Data Quality consists of a scalable, rule-based engine that uses a single-pass integration process to standardize, match, and duplicate information across global data. This framework within the Infrastructure system facilitates you to define rules and execute them to query, validate, and correct the transformed data existing in an Information Domain.

3

Data Quality Check

Data Platform Contents include Data Quality Check Rules. These Rules are defined at the Business Term and Entity Level, and seeded as a part of the Data Platform Content.

Types of Data Quality Checks

The following are the types of Data Quality Checks and their definitions:

Data Quality Check	Definition
Blank Value Check	Identifies if the base column is empty considering the blank space.
Column Reference/Specific Value Check	Compares the base column data with another column of the base table or compare with any attribute of compatible data type from a referenced dimension of a base entity.
Data Length Check	Checks for the length of the base column data by using a minimum and maximum value, and identifies if it falls outside the specified range.
Duplicate Check	Is used when a combination of the column is unique and identifies all duplicate data of a base table in terms of the columns selected for the duplicate check.
List of Value Check	It can be used to verify values where a dimension/ master table is not present. This check identifies if the base column data does not match with a value or specified code in a list of values.
NULL Value Check	Identifies if NULL is specified in the base column.
Referential Integrity Check	Identifies all the base column data that has not been referenced by the selected column of the referenced table. Here, the user specifies the reference table and columns.
Range Check	Identifies if the base column data falls outside a specified range of a Minimum and Maximum value. Value Needs to be between 0 and 100.

Data Quality Check	Definition
Uniqueness Check for Numeric Identifiers in Dimension	Check to identify duplicates in Numeric Identifier Attribute for a Dimension Entity. Check to identify changes in Numeric Identifier Attribute for a Dimension Entity for the same Business Key member. **Note:** Threshold option is currently not supported for custom check.
Special Character Check	Identify business term contains only the allowed set of special characters. Currently, DFCS has preconfigured rules for the following Business Terms: Legal Entity Code Legal Entity Description Legal Entity Name Data Source Code Data Source Description

Data Quality Group

Data Quality Groups facilitate the logical grouping of the defined DQ definitions and to schedule their execution.

Create a Group

The following capabilities are supported in Data Quality Group:

- Ability to create Custom DQ groups based either on an entity or on an existing group.
- Ability to map the following three types of rules to the group:
  - Seeded DQ Rules
  - Published Custom DQ Rules
  - Unpublished Custom DQ Rules of the same action

> **Note:**
>
> A custom DQ group can be created only from Rules of same Entity. If you want to change the entity and rule selected, close the Create Group screen and open it again.
> No DQ rule is deleted or created while editing a Custom Group. Only the mapping is updated.
>
> - Ability to delete Custom DQ groups in Draft or Returned status.
> - Ability to edit group name, description, and DQ rules mapped to the Custom DQ group in Draft or Returned status.
>
> To create a Data Quality Group, complete the following steps:

1. From the **Inbox** page, click the action for which you want to create a DQ Group. Now click Control Extensions from the LHS menu.

   > **Note:**
   >
   > You can create a Data Quality Group only from the action which is in New or Returned status.

2. Click **Create Group**.
   The **Create Group** window is displayed.
3. Enter the Name of the group. The name can consist of alphanumeric characters and an underscore (_).
4. Enter a Description or related information about the group.

   > **Note:**
   >
   > The description field only supports the following special characters: comma (,) , full-stop/dot (.) and underscore (_).

5. Select an **Entity**.
6. However, if you want to create a new group from an existing group, enable the Create Data Quality Group from an existing DQ group option and select the required group from the Copied from **Data Quality Group** drop-down list.
7. Click **+** to link the DQ rules for the selected entity. Select the required rules and click Link.
8. Click **Save**.
   The group code is automatically populated, and the group is displayed in the **Control Extensions** page.

User can subscribe to add-on services available in DFCS. Each customer gets a prescribed set of credits along with DFCS subscription. These credits can be used to subscribe add-on services. Add-on services can be a Data application, Recipe, External Dataset or Analytical Model.

4
Data Pipeline

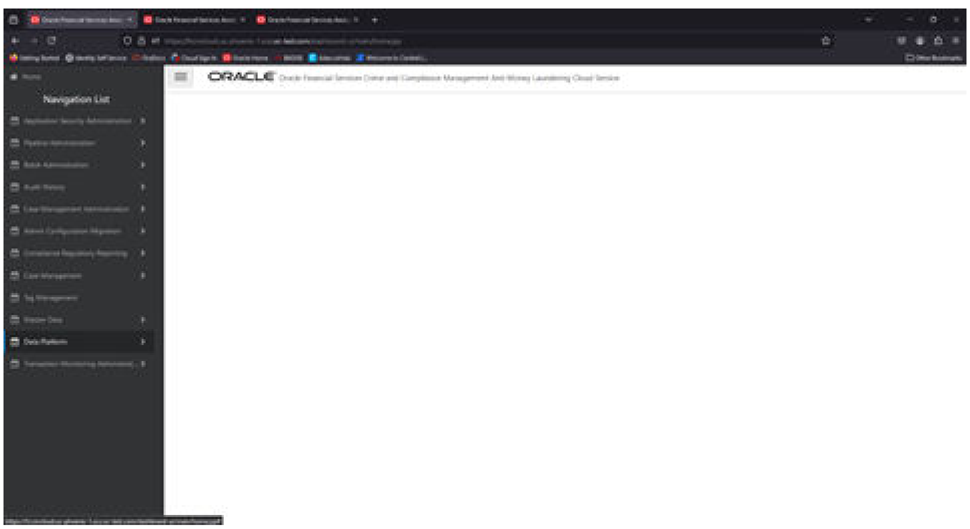
PMF Execution

When there are multiple pipelines running in parallel, monitoring the processes to find out if they have executed successfully or otherwise, information about the tasks within pipelines that have completed, failed, canceled, and in-progress yet becomes important.

The Process Modeling Framework (PMF) displays granular level execution details of all PMF processes belonging to Loader components and Data Quality executions within an environment (tenancy). Only authorized users have access to the PMF Dashboard. For each PMF process, these authorized users can drill down and view the related activities and further into the events recorded for each activity along with information such as the timestamp, event messages, and event error codes.

To add Process Modeller to Data Pipeline, follow these steps:

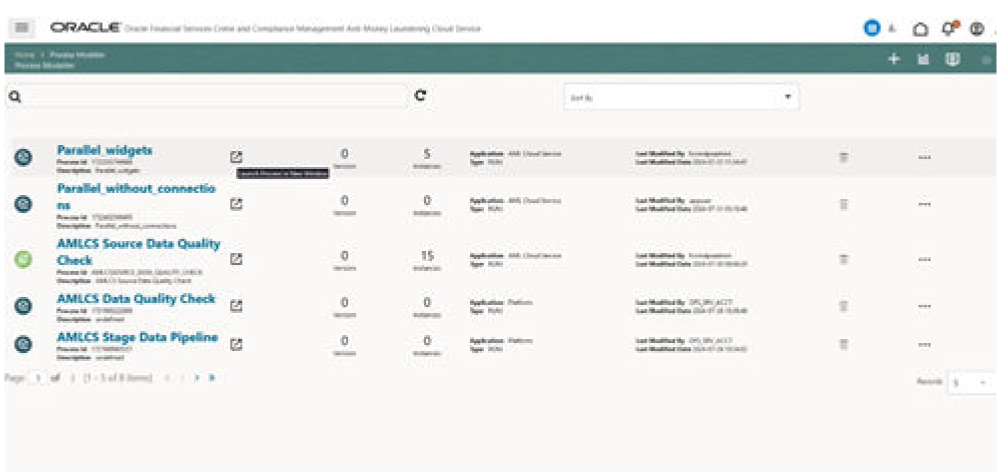
1. Navigate to Data Platform feature:



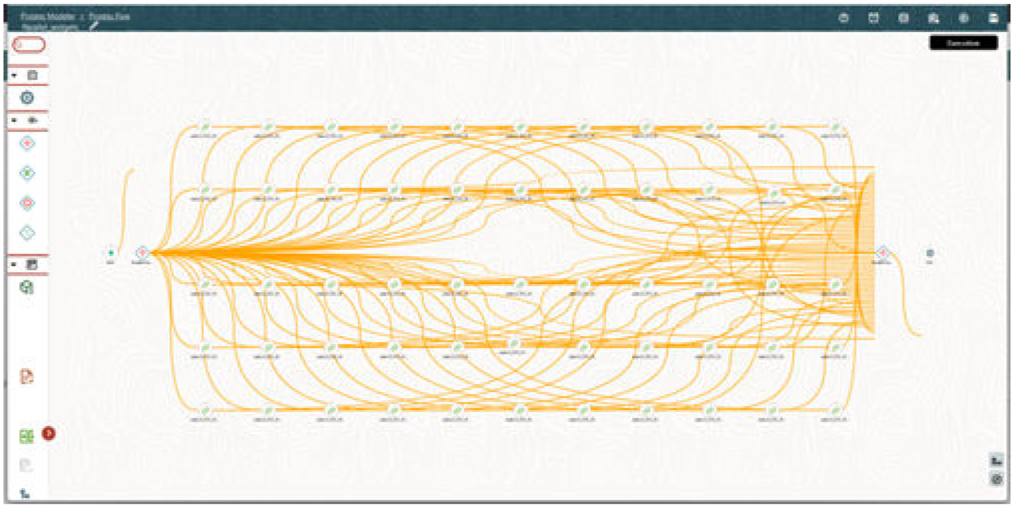
2. Select Data pipeline

3. Click on **Process Modeller**:

4. Click to open the pipeline in new window:



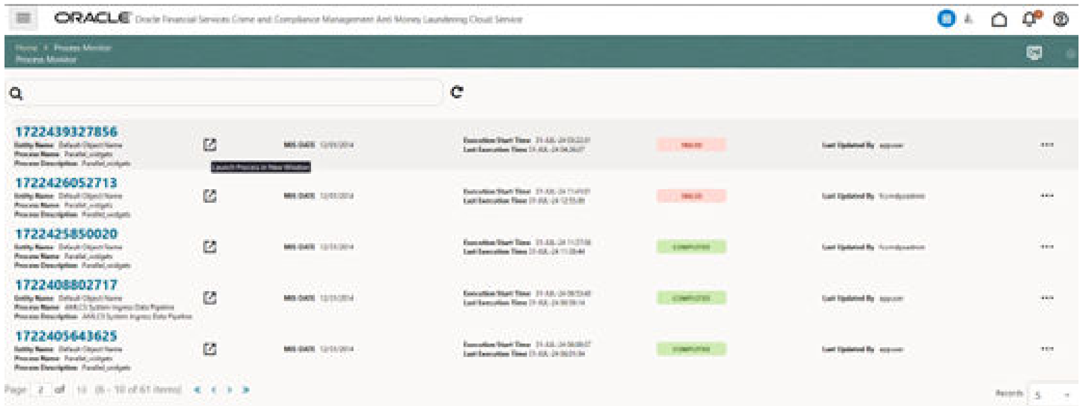
5. Click on the **Execute** button on top:



Check Execution Logs

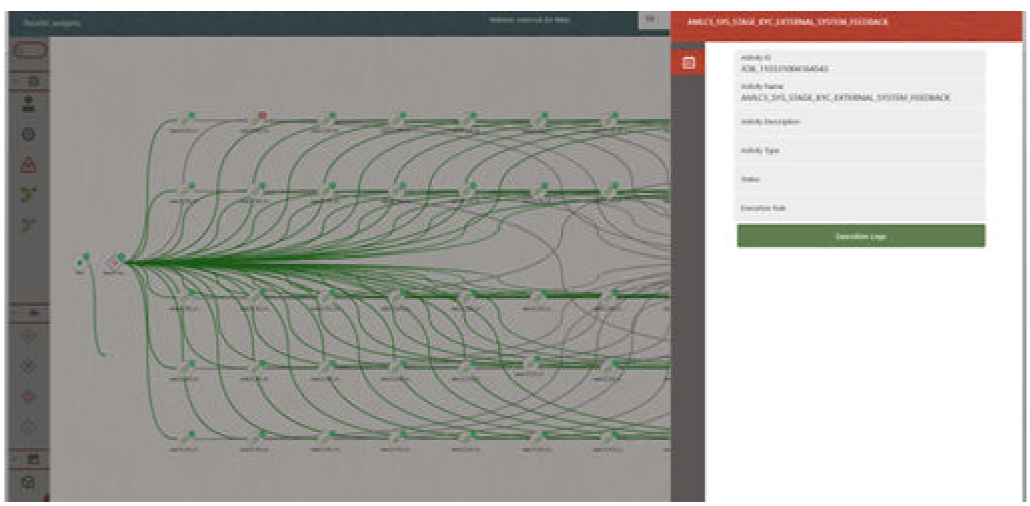
1. After PMF Execution, from the main window, select **Process Monitor**:

2. Click **Launch Process in New Window** icon open in a new tab:

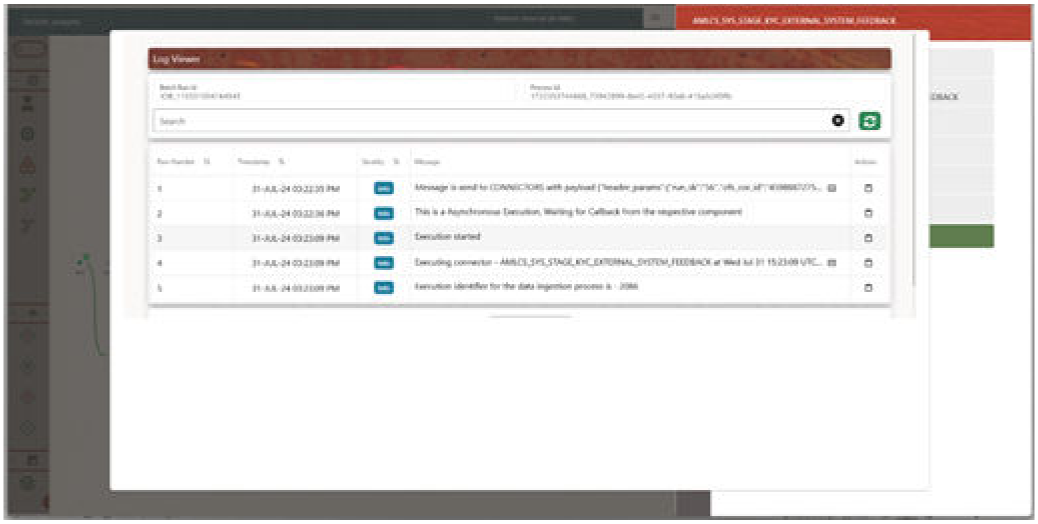


3. Double click on the failed widget to view logs:

4. Click **Execution Logs**



5. The **Log Viewer** window shows the log messages.


Log Viewer
Search
Execution started

Glossary

Index