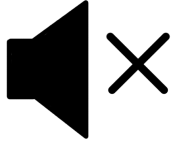


Nov 23, 2021

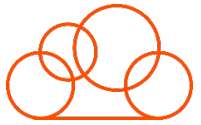
Search Intelligence: Rank Assets based on Usage Popularity

Srinivasa Gopal, Principal Customer Success Technologist

Housekeeping Tips



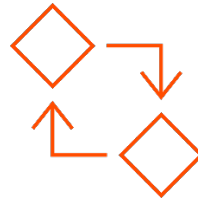
- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available on our **INFASupport YouTube channel** and **Success Portal** - where you can download the **slide deck** for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.



Bootstrap trial and
POC Customers



Enriched Customer
Onboarding
experience



Product Learning
Paths and Weekly
Expert Sessions

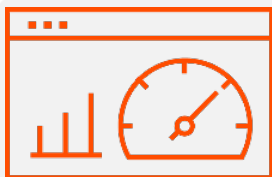


Informatica
Concierge



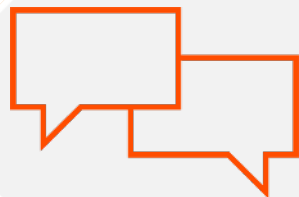
Tailored training and
content
recommendations

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

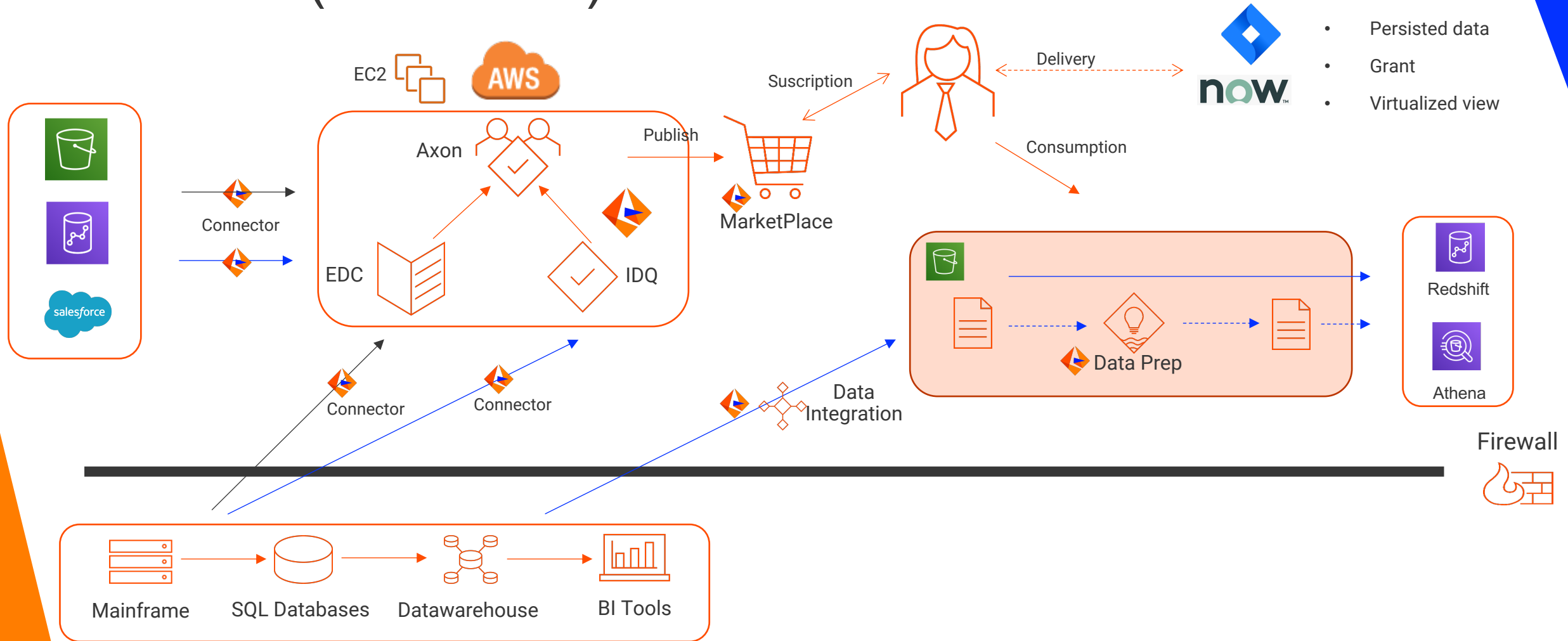
The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

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

- UseCase Overview
- Gathering the Usage Statistics
- Loading content into EDC
- Application Configuration
- Demo
- Q&A

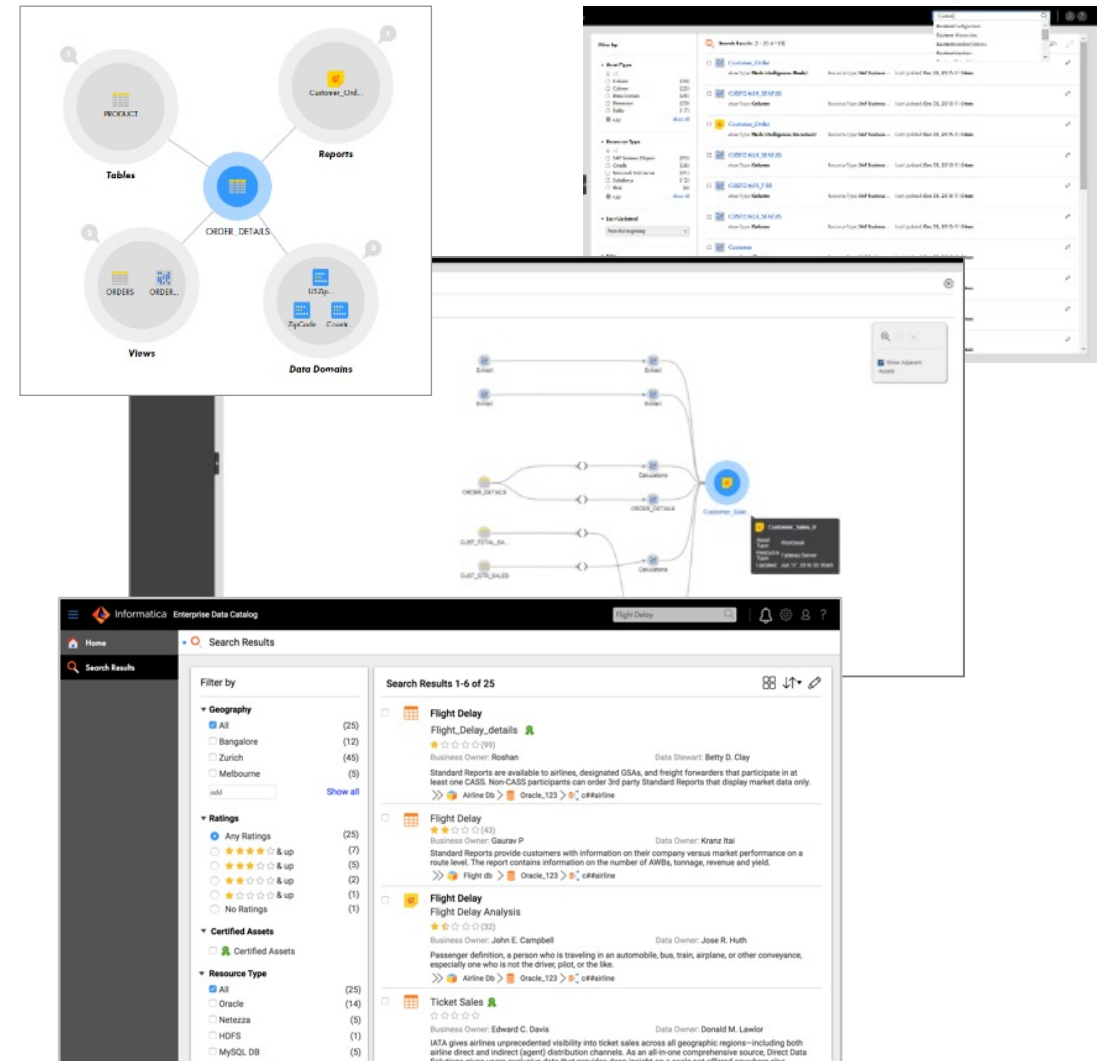
Data Flow (with AWS)



- Persisted data
- Grant
- Virtualized view

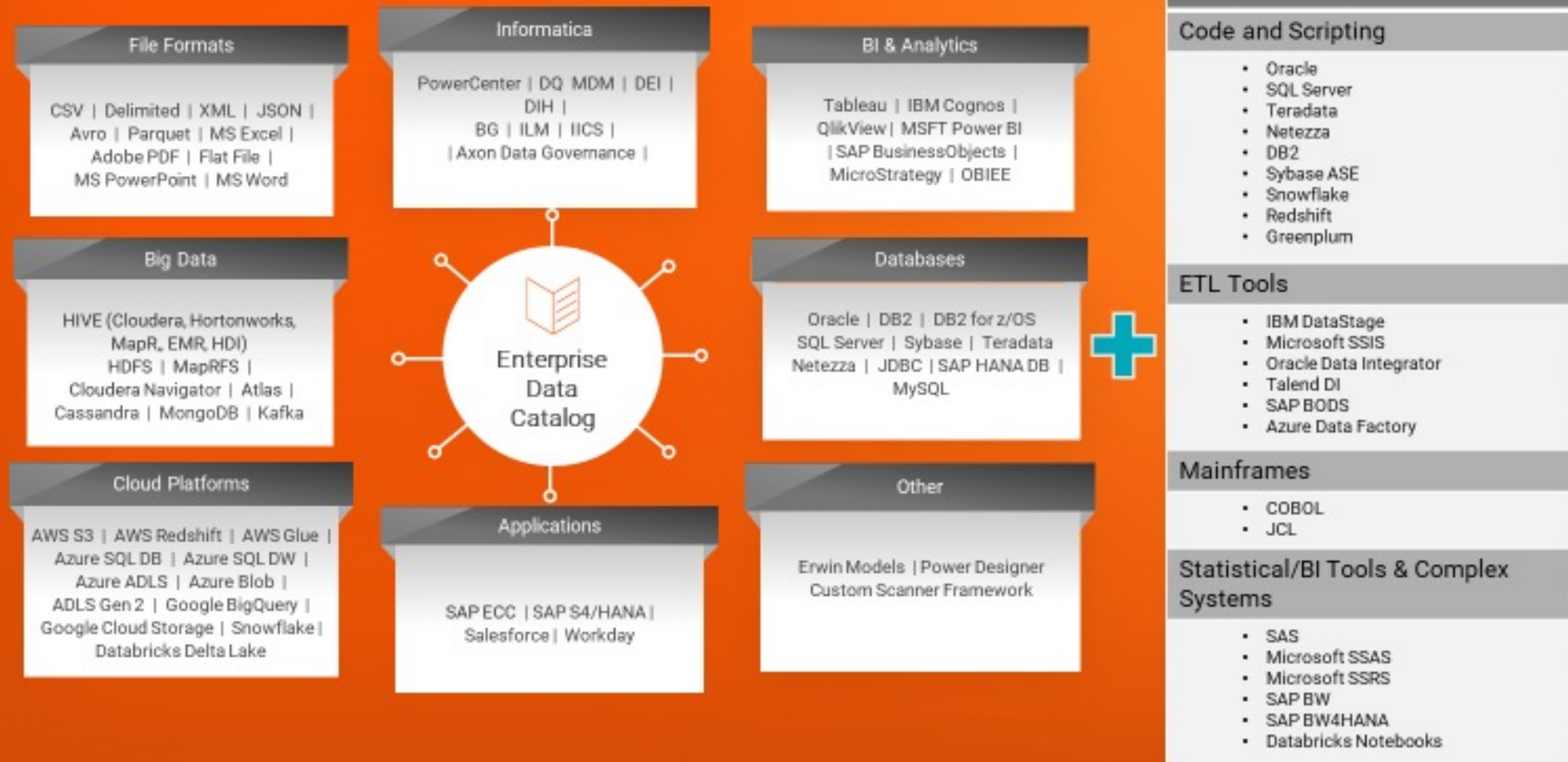
Enterprise Data Catalog

- **Search** and discovery
- **End-to-end** data lineage & impact analysis
- **AI-powered** automatic discovery, classification and business context
- **Open framework** for custom extensions and integrations
- **Collaboration & social curation** to tap into shared data knowledge



Broadest and Most Complete Metadata Connectivity

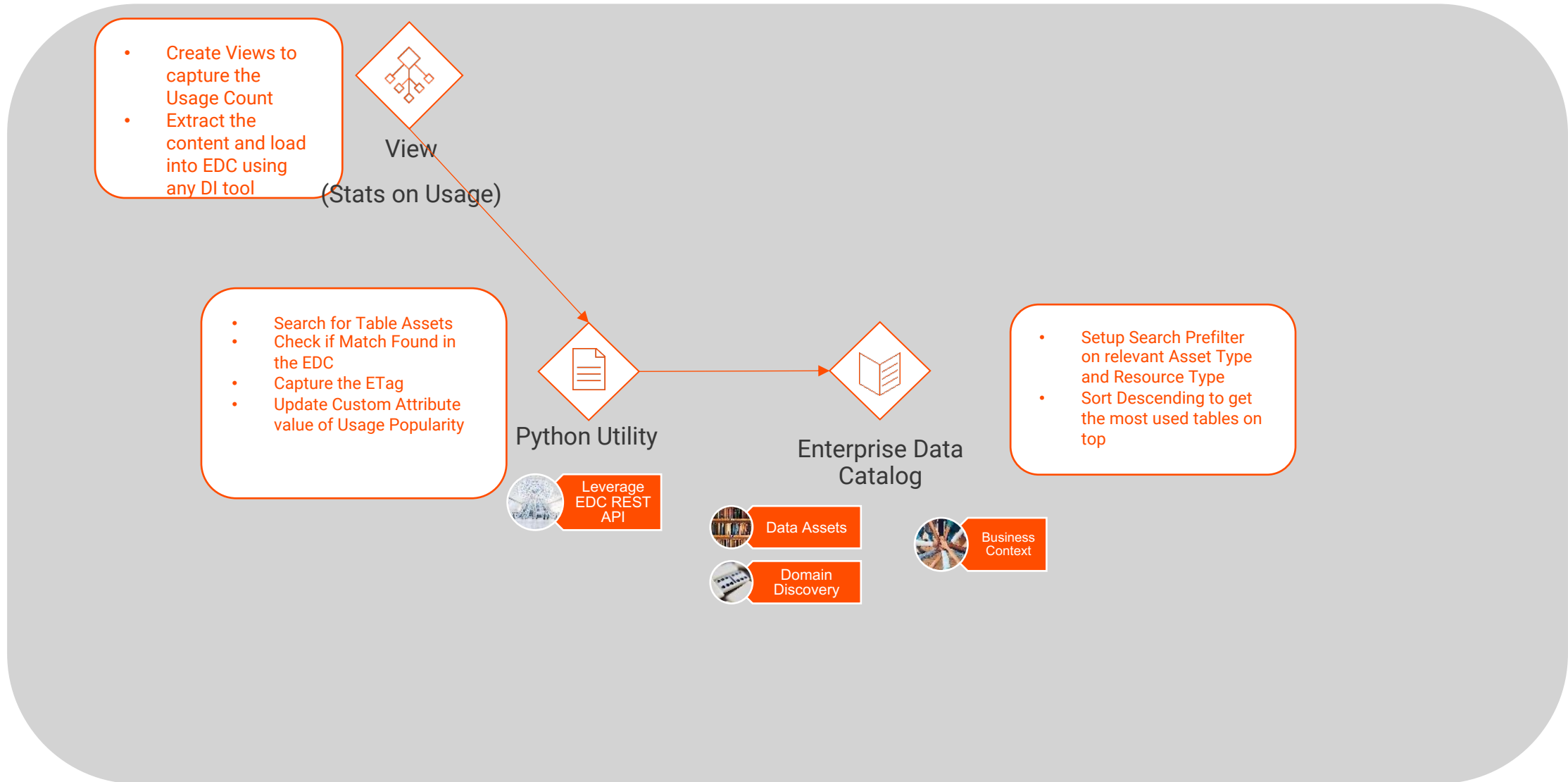
EDC 10.5.1



Ranking Data Assets

- How to Rank Assets based on Usage Stats?
- What are the top most / least used Assets in the actual Source?
- As the Usage varies overtime, how to keep the ranking up-to-date?
- How to automate the process, to avoid manual overhead?

Load Usage Stats into EDC



Create Custom Attribute

Informatica | New

Overview Library Resource Monitoring **Attributes** X

Add, edit or delete custom attributes. You can also edit the search configuration settings for the System Attributes.

Custom Attributes System Attributes

Custom Attributes

Name

App_link

Application_Link

Asset Links

Business Description

Business Unit

Data Center

Data Owner

Data Steward

Department

deployment environment

Display Name

EDC_RO_Classifications

Location

Region

Related Business Terms

resourceOnly

Sensitive Assets

Subject Matter Experts

Tag

URL

Usage Popularity

Usage Popularity

Page 1 of 1

Usage Popularity

General Properties

Name: Usage Popularity

Description:

Data Type: Integer

Allow Multiple Value Selection:

Search Configuration

Search Rank: High

Allow Filtering:

Class Assignment

Assign To: All Classes Specific Classes

Name	Originates From
<input checked="" type="checkbox"/> Table	com.infa.idm.relational
<input type="checkbox"/> Alias Table	com.infa.idm.obiee
<input type="checkbox"/> Alias Table	com.infa.idm.businessobject
<input type="checkbox"/> Analyze Tables Component	com.idm.custom.matillion
<input type="checkbox"/> BO Table	com.infa.idm.mdm
<input type="checkbox"/> BO Table Column	com.infa.idm.mdm

Show object types in hierarchy

Resource Assignment

Assign To: All Resources Specific Resources None

Name	Type
<input type="checkbox"/> CDW_Redshift	Amazon Redshift
<input type="checkbox"/> aws_glue_GlueScanner_REDSHIFT_infaeu...	Amazon Redshift
<input checked="" type="checkbox"/> cdg_redshift_retail_dwh	Amazon Redshift
<input type="checkbox"/> redshift_infaetest	Amazon Redshift
<input type="checkbox"/> redshift_styagi	Amazon Redshift
<input type="checkbox"/> redshift_styagi1	Amazon Redshift

Enable Sort/Filter – Custom Attribute

The screenshot displays the 'Application Configuration' page in Informatica Enterprise Data Catalog. The 'Custom Attributes' tab is active, showing a list of attributes. The 'Usage Popularity' attribute is highlighted with a red box. In the 'Settings' panel on the right, the 'Allow Sorting' and 'Allow Filtering' checkboxes are checked and highlighted with a red box. The 'Allow Searching' checkbox is also checked, and the 'Search Rank' is set to 'High'. The 'Allow Editing' checkbox is unchecked, and the 'Display in Object Overview' checkbox is checked.

Informatica Enterprise Data Catalog Discovery | Data Asset Analytics

Application Configuration

Custom Attributes System Attributes Asset Types Filter Settings Search Prefilter Settings

Select the custom attributes that you want to display as search filters and specify the required properties.

Attributes Add

Name ▾

- ▶ App_link
- ▶ Application_Link
- ▶ Data Center
- ▶ Data Centers
- ▶ EDC_RO_Classifications
- ▶ Party Information
- ▶ Region
- ▶ Related Business Terms
- ▶ Sensitive Assets
- ▶ Tag
- ▶ URL
- ▶ Usage Popularity

Settings

- Allow Searching
- Search Rank: High ▾
- Allow Sorting
- Allow Filtering
- Allow Editing
- Display in Object Overview

Application Configuration – Create Prefilter

The screenshot shows the 'Application Configuration' page for 'Enterprise Data Warehouse'. The page is divided into two main sections: '1 Create Prefilter' and '2 Prefilter Permissions'. The 'Create Prefilter' section is active and contains four configuration panels:

- Resource Name:** A list of resources with checkboxes. 'All' is selected. Other resources include SAP_S4HANA_S4H_EMEA_FULL_METADA... (6.86M), Cognos_OA (280.10K), SAP_S4HANA_S4H_EMEA (141.40K), SAP_S4HANA_EH7_TMP (139.41K), and teradata_express (83.79K). There is an 'Add' input field and a 'Show All' link.
- Data Set Used By:** A list of users with checkboxes. 'All' is selected. Other users include Administrator (10), Fahd Boulbellout (4), Loosli Vincent (2), Siddall, Anthony (2), and Bezzari Anass (1).
- Data Set Modified By:** A list of users with checkboxes. 'All' is selected. Other users include Administrator (21), Siddall, Anthony (8), Fahd Boulbellout (3), Bezzari Anass (2), and Loosli Vincent (2). There is an 'Add' input field and a 'Show All' link.
- Sensitive Assets:** A list of asset types with checkboxes. 'All' is selected. Other types include true (124) and false (117).

Navigation buttons include '< Back', 'Next >', 'Save', and 'Cancel'.

Application Configuration – Create Prefilter

Informatica Enterprise Data Catalog | Discovery | Data Asset Analytics

Application Configuration

Custom Attributes | System Attributes | Asset Types | Filter Settings | Search Prefilter Settings

1 Create Prefilter | 2 Prefilter Permissions

Certified
All Assets

Rating
 All (43)
 ★★★★★ and above (45)
 ★★★★☆ and above (47)
 ★★★☆☆ and above (47)

Asset Type
 All (6.73M)
 DataElement (3.04M)
 Reference Data Element (1.92M)
 Function Module Table Field (1.66M)
 Table Field (1.35M)
 Table (164.83K)
Add Show All

Resource Type
 All (7.23M)
 SAP S/4HANA (280.10K)
 Cognos (84.44K)
 Teradata (61.67K)
 SAPBW (44.30K)
 Amazon Redshift (2430)
Add Show All

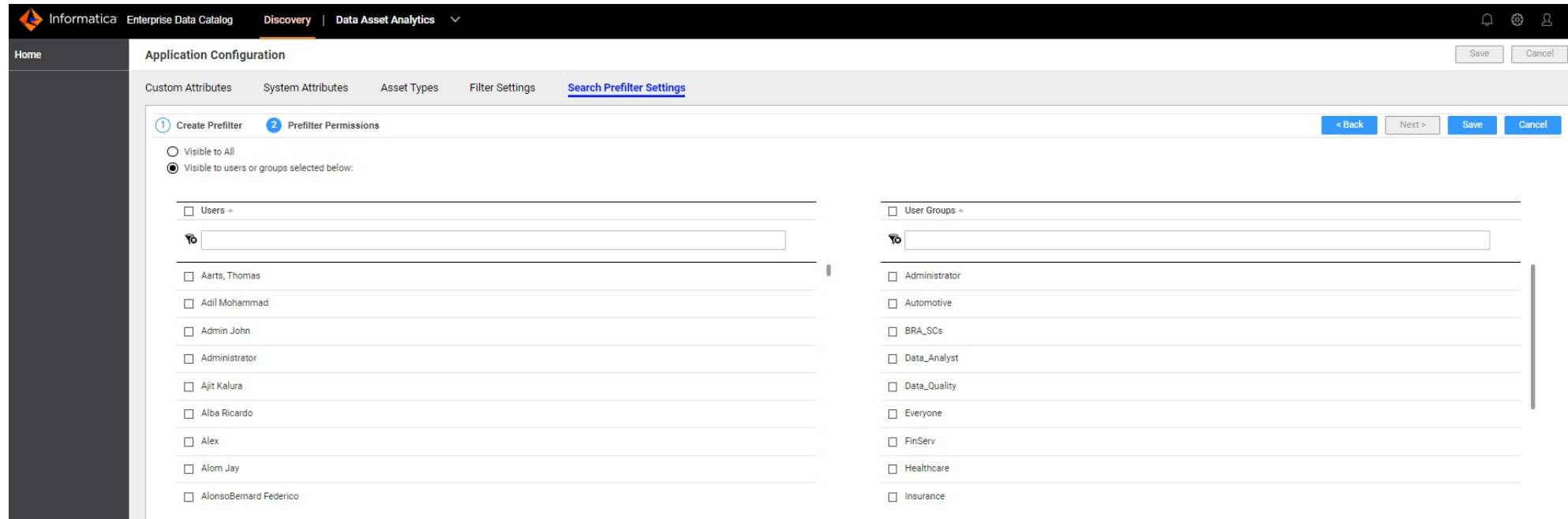
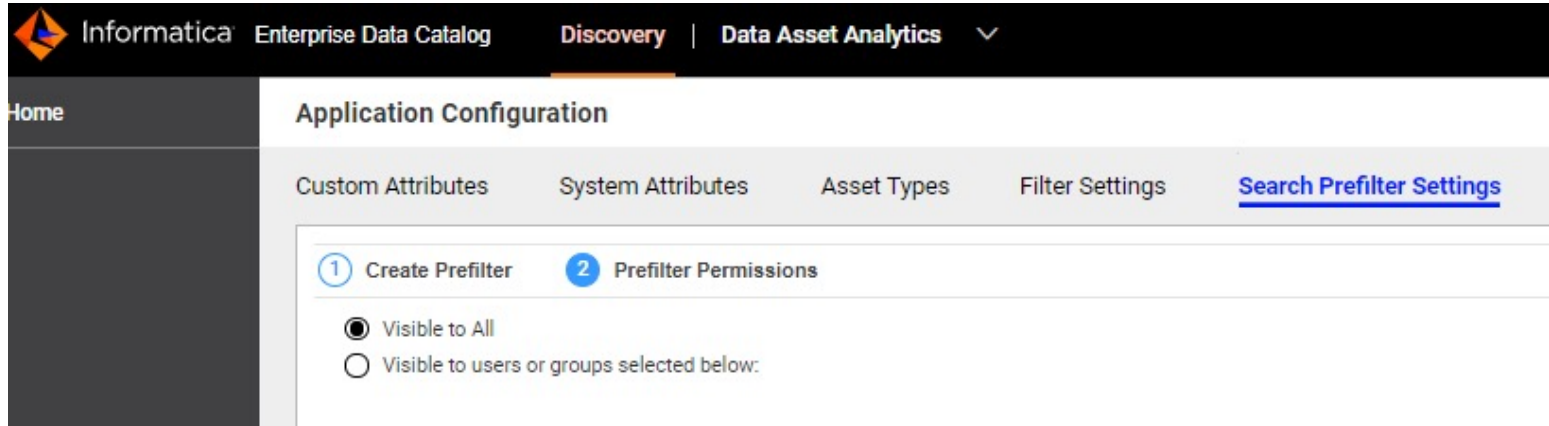
Last Updated
From the beginning

Usage Popularity
Under 150 (4)
 to GO

Location
 All (5214)
 Data Lakehouse (3226)
 Data Lake

Data Domain
 All (2307)
 Address (838)
 Country (810)
 City (751)
 LastName (718)
 Date_MM_DD_YYYY
Add Show All

Application Configuration – Create Prefilter



Redshift – Example

- Create View to capture the Usage Count for each SELECT Query on a particular Table, num_qs provides the number of times DISTINCT Queries are made

```
CREATE SCHEMA admin;
CREATE OR REPLACE VIEW admin.v_get_tbl_scan_frequency
AS
SELECT
    database,
    schema AS schemaname,
    table_id,
    "table" AS tablename,
    size,
    sortkey1,
    NVL(s.num_qs,0) num_qs
FROM svv_table_info t
LEFT JOIN (SELECT
    tbl, perm_table_name,
    COUNT(DISTINCT query) num_qs
FROM
    stl_scan s
WHERE
    s.userid > 1
    AND s.perm_table_name NOT IN ('Internal Worktable','S3')
GROUP BY
    tbl, perm_table_name) s ON s.tbl = t.table_id
AND t."schema" NOT IN ('pg_internal')
ORDER BY 7 desc;
```

EDC Swagger – REST API

swagger Select a spec default

Catalog Administrator REST Documentation. Version 10.5.1

[Base URL: `infa1051-axon721-demo-cat.infacloud.eu:9085/access`]
<http://infa1051-axon721-demo-cat.infacloud.eu:9085/access/v2/api-docs>

This document contains the details of all the APIs that are exposed to user. Please find additional documentation [here](#).

Copyright © 1993-2021 Informatica LLC. All rights reserved
[Informatica LLC](#).

catalog-api-controller the catalog API >

Catalog

- GET** `/2/catalog/configuration/app_events` Gets the application configuration. 🔒
- PUT** `/2/catalog/configuration/app_events` Updates the application configuration 🔒
- GET** `/2/catalog/configuration/email_events` Gets the email configuration. 🔒
- PUT** `/2/catalog/configuration/email_events` Updates the email configuration 🔒
- GET** `/2/catalog/data/app_events` Events for the application 🔒
- GET** `/2/catalog/data/app_events_class_types` Class types of Events for the application 🔒
- GET** `/2/catalog/data/app_events/count` Event count for application 🔒
- GET** `/2/catalog/data/objects` Objects 🔒
- PUT** `/2/catalog/data/objects` Update Objects 🔒

[http\(s\)://host:port/access/swagger-ui.html#/](http(s)://host:port/access/swagger-ui.html#/)

Python Utility – Design / Pseudocode

- Input – Redshift_Popularity_Report.csv → Extracted from AWS Usage Stats View
 - Database,Schema,Table,Usage Popularity
 - test,cloudetldemo,product,69 → Sample Row Format, product table as part of cloudetldemo Schema, test Database has been Queried 69 times
- processCSVFile (csvFile)
 - Process the CSV file, for each row call the findAndUpdateCatalog function to pass the Database, Schema, Table and Usage Popularity Values
- findAndUpdateCatalog (databaseName, schemaName, datasetName, attrDict)
 - Execute the search on /access/2/catalog/data/objects and extract the result in JSON
 - Get the eTag using the edcUtils get the Facts
 - Prepare the Parameters with the new URL and Query string along with Custom Attribute ID to be updated
 - Execute the request put, to update the Usage Popularity Custom Attribute
- listCustomAttributes ()
 - Get all the Attribute defined using /access/2/catalog/models/attributes
 - Filter down Custom Attributes with Package prefix: com.infa.appmodels.Idm.“
 - Create and Add : Attribute ID and Name into Array
- replaceExistingAttrValue (theItem, theAttribute, theValue)
 - facts are represented as a list of dictionaries
 - each fact will have: attributeld and value for the given attributeld (in 'theAttribute') we want to replace the value with theValue

Setup API User Access

Edit Roles and Privileges

admin
Select a domain or service to edit the assigned roles and privileges.

Roles **Privileges**

- Scheduler Service
- DIS - Data Integration Service
- CS - Catalog Service
 - API Privileges
 - REST API Privilege
 - REST API User Privilege
 - Catalog Privileges
 - Catalog Management
 - Catalog View
 - Catalog Edit
 - Application Configuration**
 - Domain Creation
 - Domain Curation
 - Resource Management
 - Admin - View Resource
 - Admin - Edit Profiling
 - Admin - Edit Resource
 - Domain Management
 - Admin - View Domain and Domaingroup
 - Admin - Edit Domain and Domaingroup
 - Data Privileges
 - View Data
 - View Sensitive Data
 - Admin - Create Attribute
 - Admin - Monitoring

Download Here



[GitHub - UpdateUsagePopularity](#)

Note: The package will be published in INFA Marketplace soon

Demo



Thank You!

References

AWS System Tables

- **SVV_TABLE_INFO** ([Click here](#)) which shows the summary information for tables in your database. And does not behave as STL, since it contains snapshots of the current system data.
- **STL_SCAN** ([Click Here](#)) which analyzes the table scan steps for queries. And being a STL log views, it may only retain approximately two to five days of log history

EDC REST Resources

- EDC GitHub REST API Utilities: [Click here](#)
- EDC REST API Getting Started Guide: [Click here](#)
- EDC REST API Reference Guide: [Click here](#)