

Module 2

Conventional Search VS Semantic Search :

Aspect	Conventional Search	Semantic Search
Search Method	Based on exact keyword matching.	Based on understanding the intent and context of the search.
Search Results	Results are based on the exact occurrence of the keyword in assets.	Results are based on the intent behind the search and the context of the terms.
Example	Keyword: "jsmith" → Results: All assets with "jsmith" (exact match).	Keyword: "jsmith" → Results: Assets with "jsmith" plus related content (e.g., emails, documents, etc.).
Result Scope	Limited to exact matches of keywords.	Broader, includes related content even if the exact keyword is not present.
Context	No understanding of the context or user's intent.	Takes into account the user's intent and the context of the search terms.

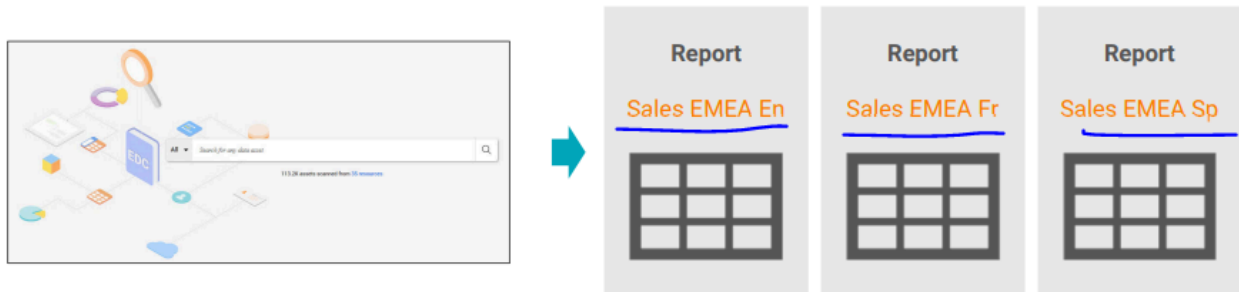
What is Semantic Search in EDC?

- EDC displays probable matches based on:
- Relationships: Primary Key – Foreign Key Relationships, and so on
- Data Domains: Columns inferred as part of the domain discovery
- Search Hints provide additional details on why it includes the data asset in the search results

Example :

Searching in EDC

- Consider an example where you want to search for sales report for the EMEA region
 - Keyword:** Sales report EMEA
 - Search Results:** Includes actual report resource types or objects with region attributes having value EMEA



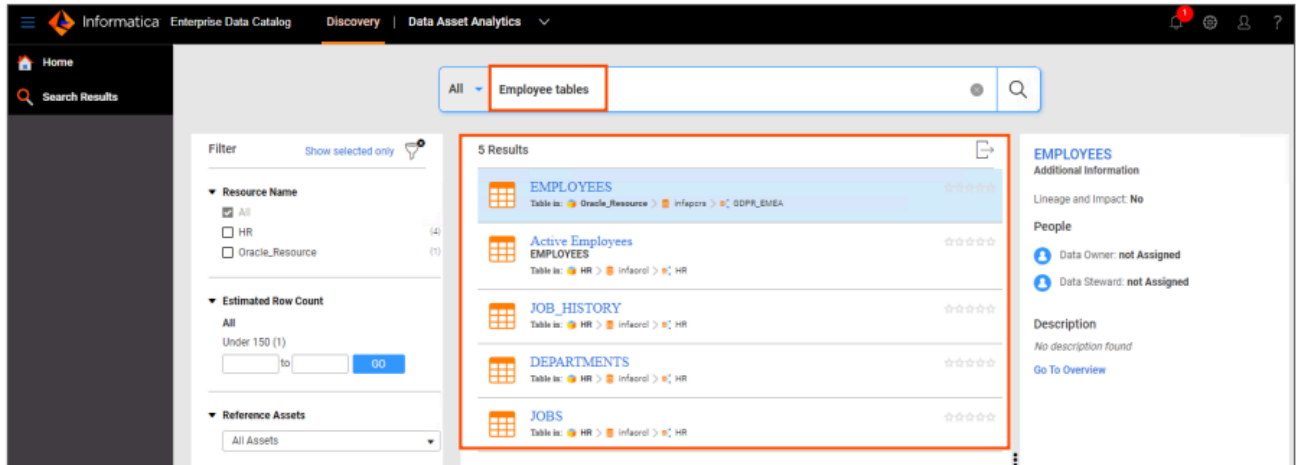
Example 1

Searching for Implied Assets

هنا عمل بحث علي كلمة delay جابلك كل ال Assets اللي تخص ال delay

Example 2

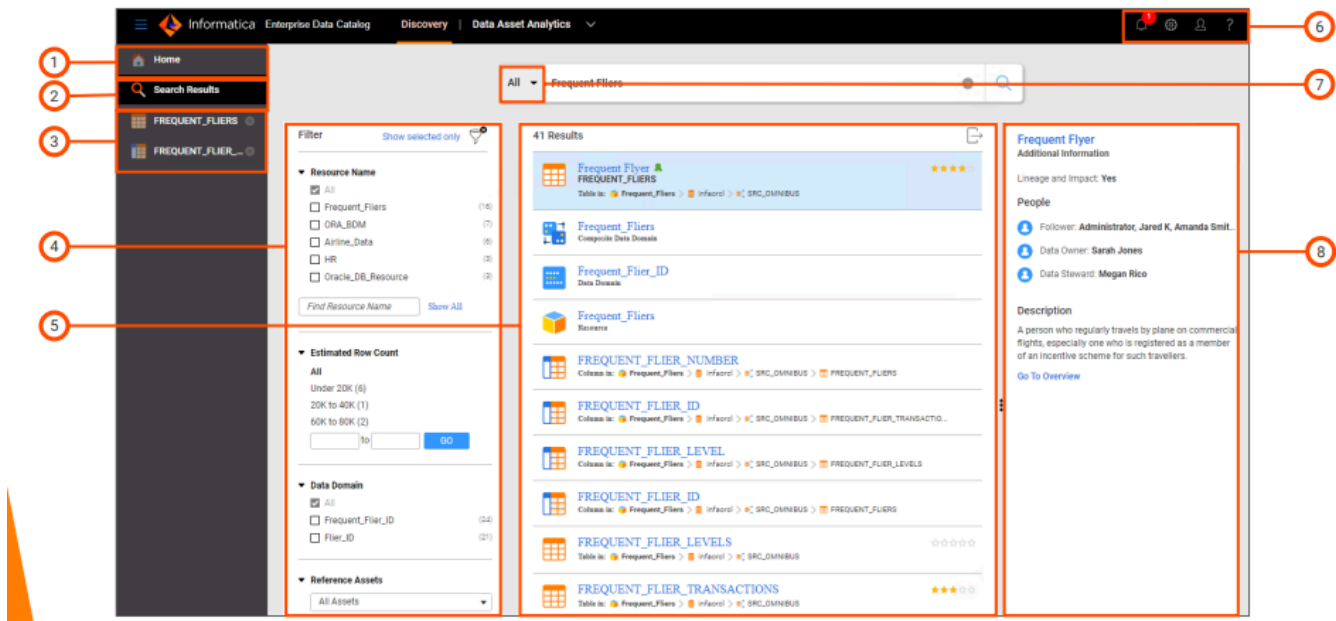
Searching for Known Assets



هنا عمل بحث علي Employee tables بحيث يدور علي الجداول بس مش كل ال Assets فا عرض كل الجداول اللي تخص البحث او ممكن بيها علاقه بمضمون الجدول

Explain User Interface for Search Results Page :

The Search Results Page



1. Go back to Home screen
2. Go to latest search results
3. List of opened assets. You can click on an asset in the list to go back to it.

4. Filter panel. Allows you to refine search results
 5. List of search result assets
 6. Main toolbar – this is available in all the pages
 7. Prefilters – you can use this to search for specific type of assets
 8. A quick look at the description and owners of the asset. You can also see if there is any lineage information for the asset.
-

Additional Search Options :

Available Search Options in EDC :

- Prefilters
 - Search by name
 - Wildcard search
 - Search suggestions
 - Keyword search
 - Double Quotes Search
 - Search Operators
 - Search Rank
 - Related Search
-

1-Prefilters :

- Predefined filters to be used before you search for assets

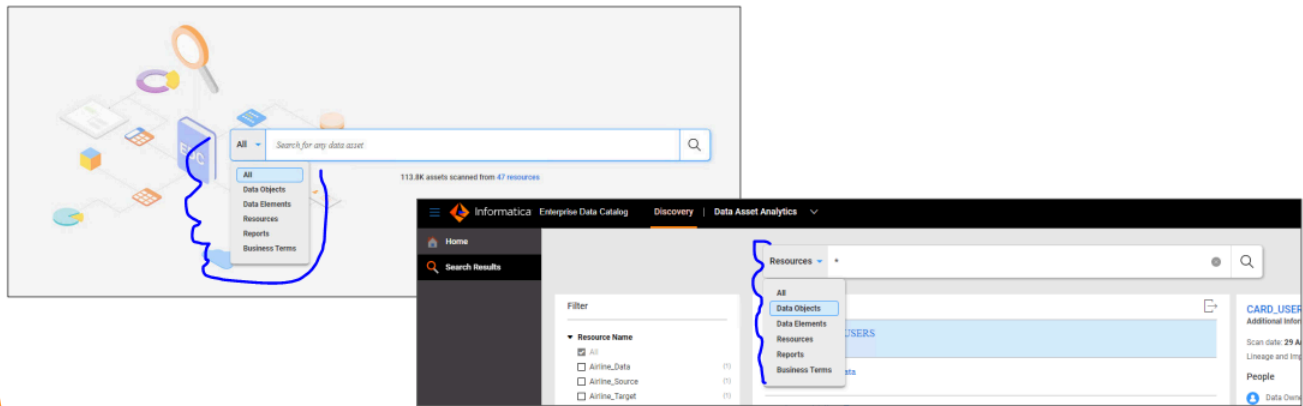
يعتبر فلتر تصفيه تحتد قبل البحث

- Used to get search results filtered according to a set criteria

تستخدم لتصفية البحث وفقا للفلتر اللي تم اختياره

- Available with the search bar in the Home page and the Search Results page

موجود في شريط البحث في الصفحة الرئيسية و صفحة عرض نتائج البحث



© Informatica. Proprietary and Confidential.



The default prefilters include **Resources** :

1-Data Objects

هو عبارة عن وحدة بيانات أكبر، ممكن تحتوي على مجموعة من Data Elements

مثال هيكل أو مجموعة من البيانات مرتبطة ببعضها (مثل جدول، مستند، ملف)

2-Data Elements : Cell & Single Information (تحمل قيمة واحدة او خلية و يمكن ان تكون نص او رقم)

3-Reports تقارير

4-Business Terms مصطلحات تجارية

Resources Displays all the available resources

Data Objects Displays the tabular and mapping assets in a resource

Data Elements Displays the assets within any data set

Business Terms Displays the business terms like Axon glossaries and Business glossaries

Reports Displays the reports based on resources

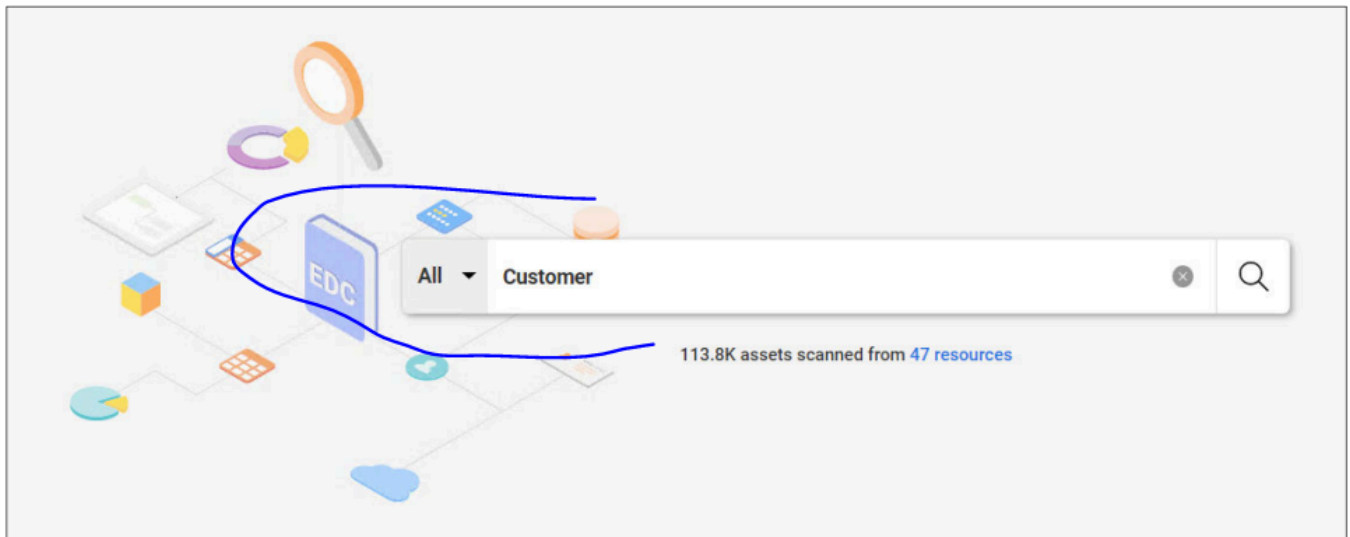
You can also create customized prefilters with the help of a combination of search filters

تقدر تنشئ فلتر مخصوص للبحث

2-Search by name :

Type the search keyword in the field to search for assets

- البحث من خلال الاسم Search by name
- البحث من خلال المرادفات للاسم Search by synonyms



This will display a list of results having a name (or other attributes) similar to the search keyword. You can also search using synonyms as the keyword. EDC will search the integrated Business Glossary for the synonym and return results.

3-Wildcard Search :

Use asterisk () character and question mark (?) character to perform wildcard search

- '' matches zero or more non-space characters

هيجيب البحث بظبط او الكلمة اللي تحتوي علي نفس الكلمة

EXAMPLE FOR *

Search keyword: sales_*

Filter

▼ Resource Name

☒ All (29)

☐ Employee_Resources (9)

☐ PCRS_SALES (9)

☐ BG_Resource (2)

☐ PCRS_SALES_GLOBAL (2)

☐ Sales_Data (2)

▼ Estimated Row Count

All

Under 15K (67)

15K to 30K (4)

30K to 45K (5)

44 Results

Global Customer Sales ★★★★★

SALES_GLOBAL

Table in: PCRS_SALES_GLOBAL > Infapcrs > SALES_GLOBAL

Sales Data Resource

Sales SALES_APJ ☆☆☆☆☆

Table in: PCRS_SALES > Infapcrs > SALES_APJ

Sales SALES_NA ☆☆☆☆☆

Table in: PCRS_SALES > Infapcrs > SALES_NA

SALES_API

- '?' Matches exactly one non-space character

مفيد لما تكون مش متأكد من **حرف واحد** فقط في الاسم او الاسم ممكن يكون لوحده او موجود في كلمتين الاول او الاخر

EXAMPLE FOR ?

Search keyword: sales_?

All sales_?
✕
🔍

Filter

▼ Resource Name

☒ All (75)

☐ Employee_Resources (75)

☐ BG_Resource (15)

☐ PCRS_SALES (10)

☐ PCRS_SALES_GLOBAL (5)

☐ Sales_Data (4)

[Show All](#)

▼ Estimated Row Count

All

Under 15K (1)

116 Results 🔗

Global Customer Sales ★★★★★

SALES_GLOBAL

Table in: PCRS_SALES_GLOBAL > infapers > SALES_GLOBAL

Sales

Term in: BG_Resource > Sales_Glossary

Sales

Term in: BG_Resource > Info_Demo_Glossary

Gross Sales

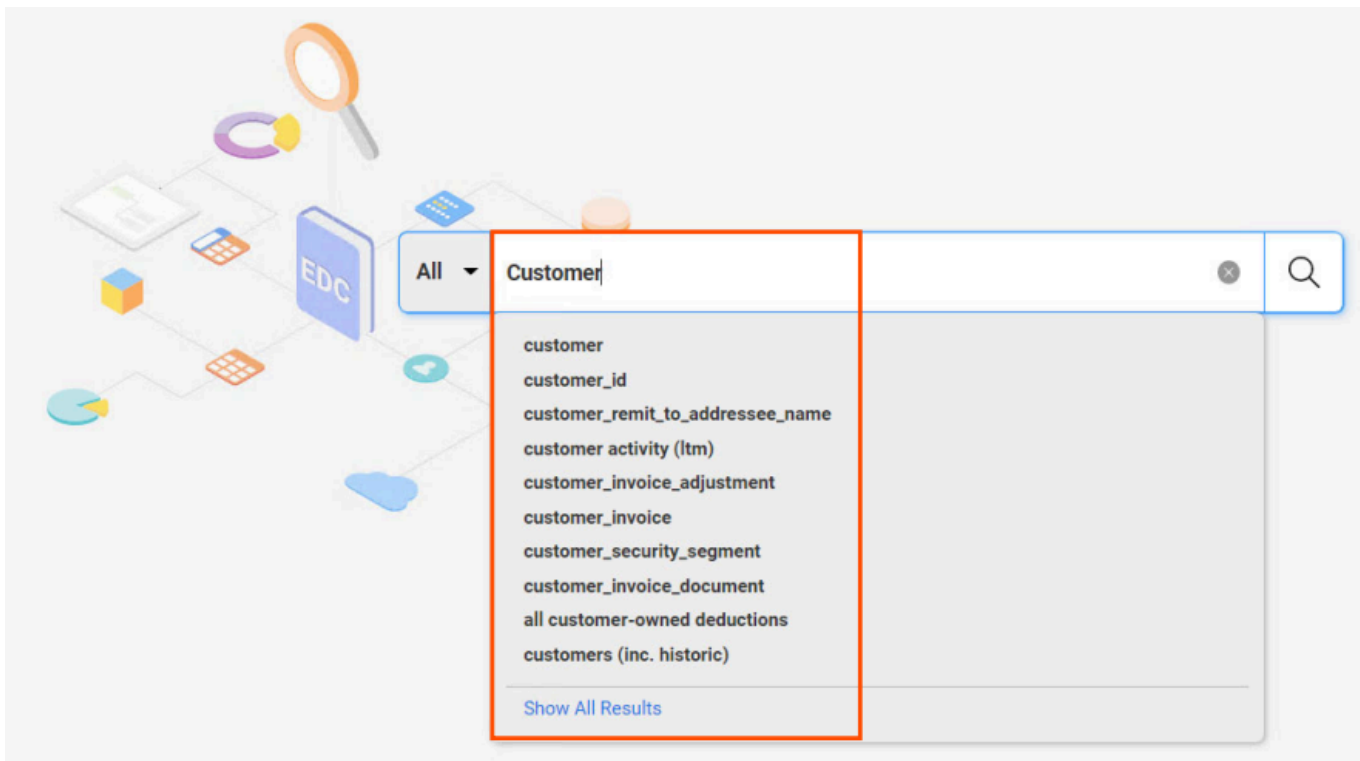
Term in: BG_Resource > Info_Demo_Glossary

Sales Amount

Term in: BG_Resource > Sales_Glossary

4-Search Suggestions :

When you type a search keyword, EDC provides suggestions matching the keyword



5-Keyword Search :

- Use keyword when you want to search for a specific type of asset
- Use keywords in regular sentences or phrases
- Example: “tables with sales”

The image shows a search interface with a search bar containing the text 'tables with sales'. The search results are displayed in a list format. The results are as follows:

Search Results
Global Customer Sales SALES_GLOBAL Table loc: PCRS_SALES_GLOBAL > Infapcrs > SALES_GLOBAL
Sales SALES_APJ Table loc: SALES_APJ
Sales SALES_NA Table loc: SALES_NA
Sales SALES_EMEA Table loc: SALES_EMEA

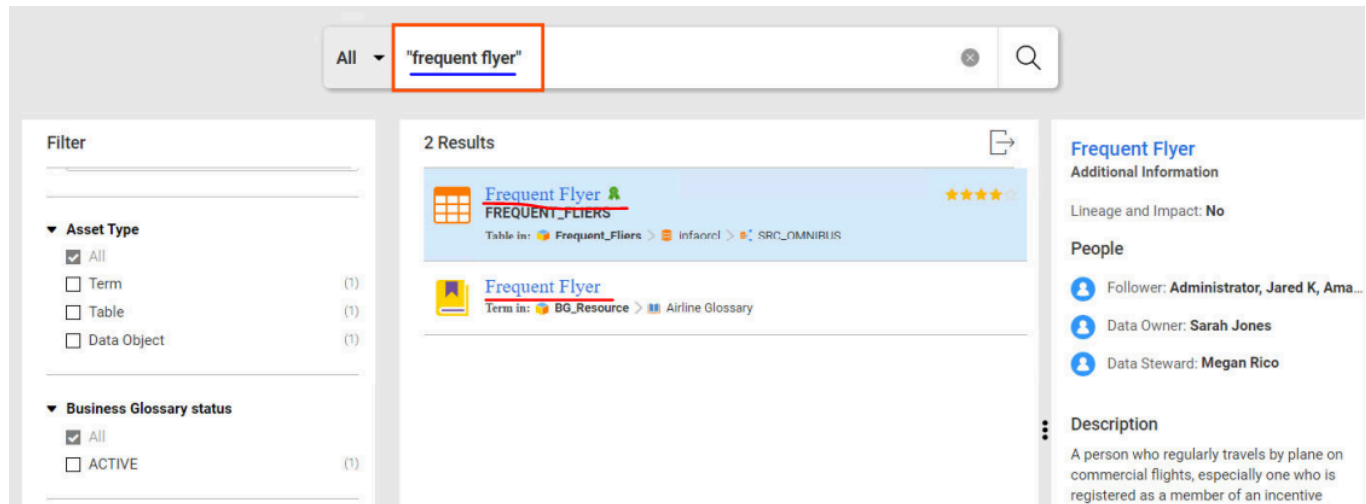
On the left side of the search results, there is a filter section. The filter section has two sections: 'Resource Name' and 'Estimated Row Count'. The 'Resource Name' section has a dropdown menu with 'All' selected. The 'Estimated Row Count' section has a dropdown menu with 'All' selected. The 'Asset Type' section has a dropdown menu with 'All' selected. The 'Asset Type' section also has a link that says 'Go To Overview'.

عرض كل الجداول التي تخص ال Sales

4-Double Quotes Search :

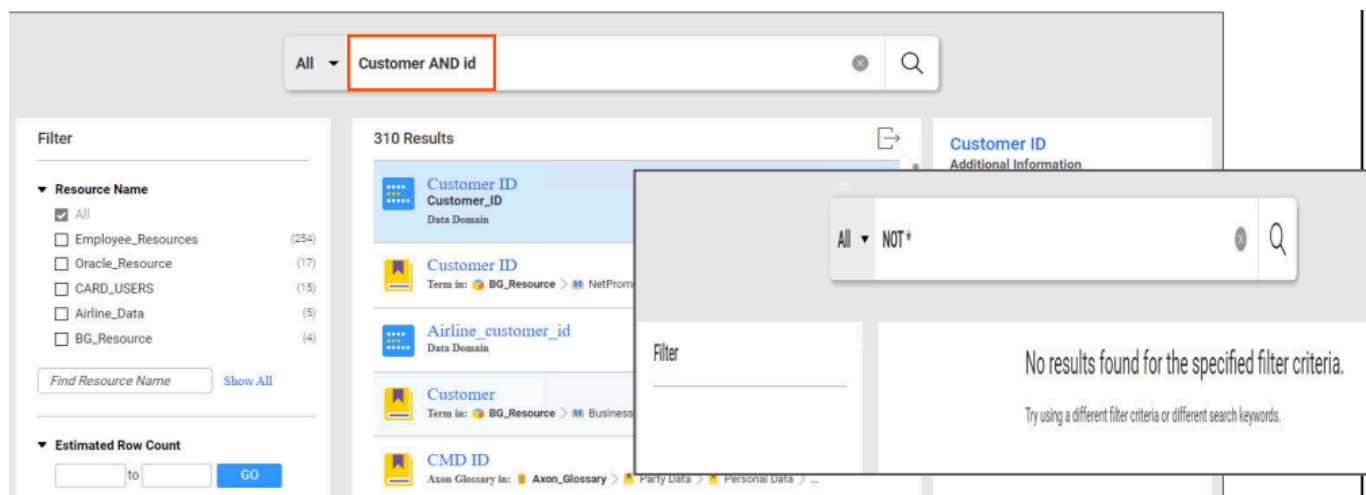
- With the help of double quotes (""), you can find the assets that exactly match the asset name within the double quotes
- This restricts the variations of the asset name in the catalog

لو استخدم "" هيعرض نتائج البحث نفس الكلمة اللي مكتوبة داخل ال Double Quotes



5-Search Operators :

- Use the search operators to make the search results accurate
- Examples: AND, OR, NOT, title, and description



نتائج تحتوي علي الكلمتين معا (أقل أولوية) AND

نتائج تحتوي علي كلمة من الاثنين (متوسط الاولوية) OR

هيستبعد اي كلمة (أعلي أولوية) NOT

6-Search Rank :

Search as per the precedence of an asset when compared to other assets that are a part of specific search results

Notice that a certified asset is listed first

ترتيب النتائج يكون كالتالي:

Certified Assets

- تظهر في أعلى النتائج لأنها موثوقة ومعتمدة من أصحاب البيانات.

أنواع الأصول التالية (بالترتيب):

Composite Data Domains

Data Domain Group

Data Domains

Axon Glossary و Business Glossary

Reports

Matching Title

الأصول التي تحتوي على الكلمة في العنوان تظهر قبل التي تحتويها في الوصف أو الخصائص.

User Interaction (تفاعل المستخدمين)

الأصول التي فيها تقييمات، مراجعات، تعليقات، أو خصائص مخصصة تظهر أعلى من غيرها.

Explore Search Results :

The screenshot displays a data catalog interface for 'Frequent Flyers'. On the left is a 'Filter' panel with sections for 'Resource Name' (All, BG_Resource, Frequent_Fliers), 'Estimated Row Count' (All, Under 1K (4)), 'Reference Assets' (All Assets), and 'Asset Type' (All, Term, Table). The central panel shows '2 Results' for the asset 'Frequent Flyer' (FREQUENT_FLIERS), including its lineage and a table icon. An orange callout points to an export icon in the top right of this panel. The right panel provides a 'Summary of the asset', including 'Additional Information' (Lineage and Impact: No), 'People' (Follower: Administrator, Jared K, Ama; Data Owner: Sarah Jones; Data Steward: Megan Rico), and a 'Description' (A person who regularly travels by plane on commercial flights...). An orange callout points to this summary panel. At the bottom, pagination shows '1 - 2 of 2' and a dropdown for '20' items per page.

- When you click on the asset name, it opens the asset details view
- When you click on the asset row, its summary is displayed in the right panel. By default, the first asset is selected and its information displayed in the panel.
- You can use the Filter panel to refine the search

1- Asset Summary :

The screenshot shows the 'Global Customer Sales' asset summary. Under 'Additional Information', it states 'Lineage and Impact: Yes'. The 'People' section lists three roles: 'Data Owner: John Smith', 'Data Steward: Sarah Jones', and 'Follower: Megan Rico'. The 'Description' section states 'FlyBird global customer sales with payment details.' At the bottom, there is a blue link that says 'Go To Overview'.

- Summary of the asset is available in the right panel.
- Information includes:
 - Whether the asset has lineage information
 - List of stakeholders for the asset: data owners, data steward, and followers
 - Description of the asset
 - Profiling data for column assets
- For more information on the asset, you can click the Go To Overview link below the description

2- Refine Results Using Search Filters :

The screenshot shows a 'Filter by' panel with several sections. Annotations point to specific features:

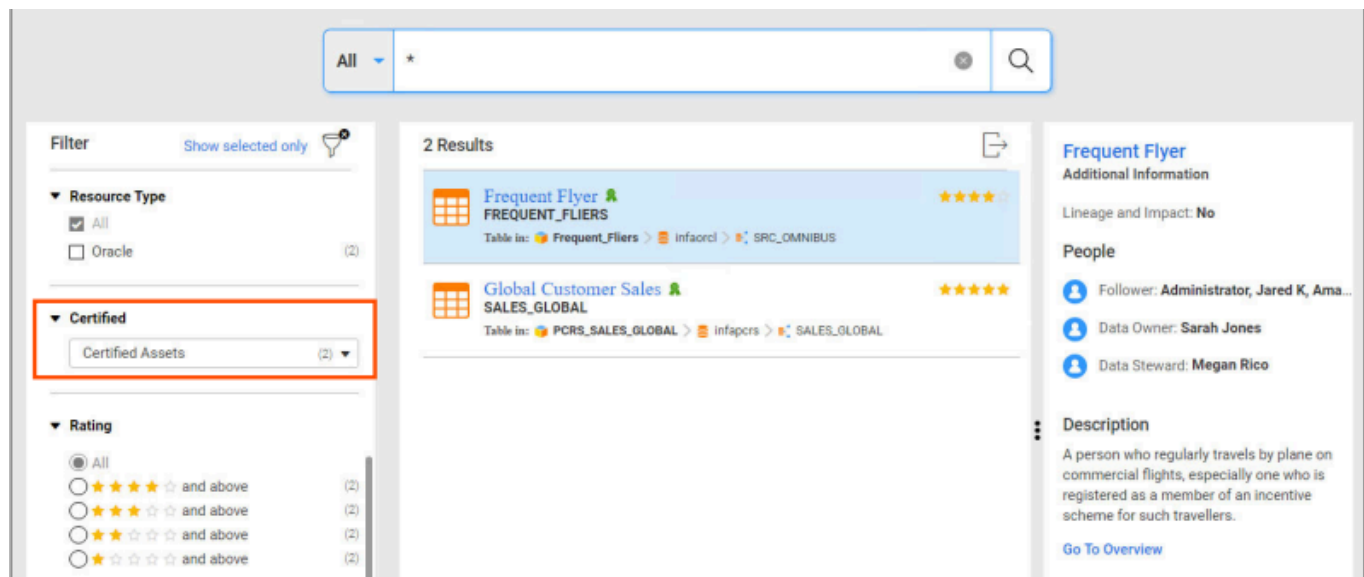
- Asset Type:** A list of asset types with their counts. An annotation points to the count '(59.39K)' for 'DataElement', stating 'Number of matching search results'. Another annotation points to the 'Add' button at the bottom of the list, stating 'Type in the asset type that is not available in the list'.
- Resource Type:** A list of resource types with their counts. An annotation points to the 'Show All' button at the bottom of the list, stating 'Click to view the complete list of resource types'.
- Estimated Row Count:** A section with radio buttons for different row count ranges. An annotation points to the input fields for the '600K to 800K (111)' range, stating 'Enter minimum and maximum range for size'.

The 'Filter by' panel includes the following sections and options:

- Asset Type:**
 - ☒ All (59.39K)
 - ☐ DataElement (54.54K)
 - ☐ Property (39.67K)
 - ☐ Data Object (28.72K)
 - ☐ Entity (12.61K)
 - ☐ City
 -
- Resource Type:**
 - ☒ All (94.00K)
 - ☐ workday (12.93K)
 - ☐ System Location (3507)
 - ☐ OneDrive (1463)
 - ☐ Oracle (571)
 - ☐ Business Glossary
 -
 -
- Last Updated:**
 -
- Estimated Row Count:**
 - ☐ N/A (4)
 - ☐ Under 200K (48)
 - ☐ 600K to 800K (111)
 - to
- Data Domain:**
 - ☒ All
 - ☐ Identification (58)

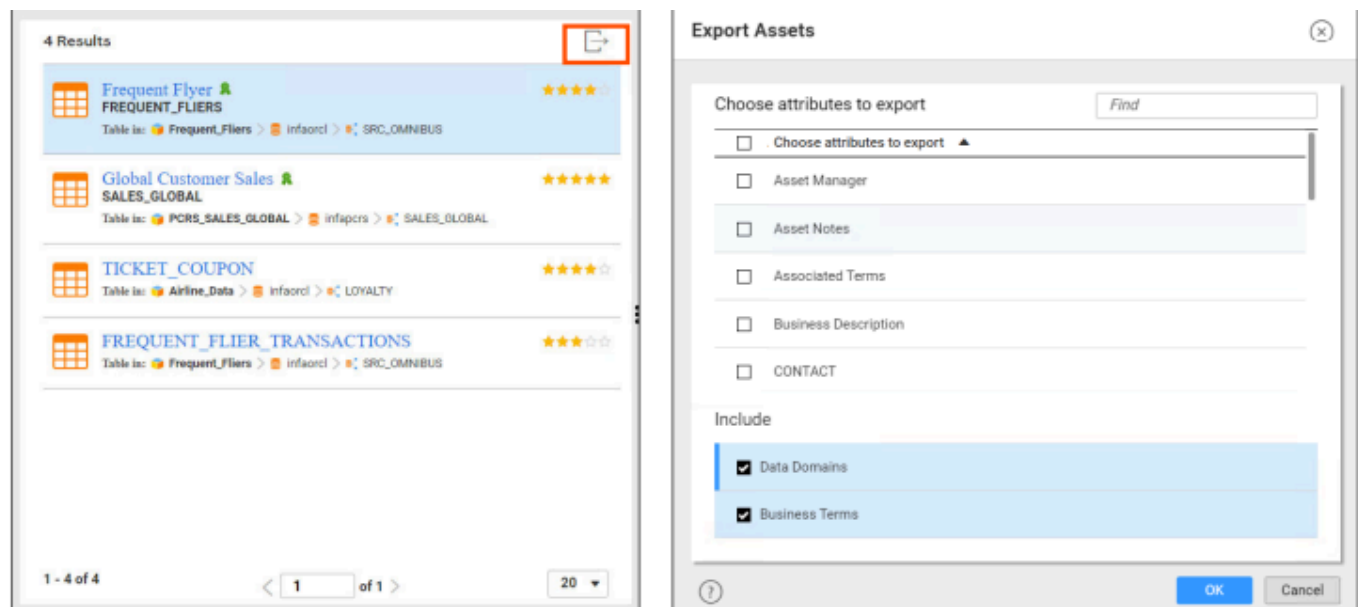
3- Search for Certified Assets :

- Search filters include a facet named Certified Assets
- Lists only certified assets related to the search string
- Toggle between All Assets and Certified Assets



4- Export Search Results :

- Export a list of objects from search results
- Import objects at a global level from multiple sources
- Monitor import and export jobs in a central UI



	A	B	C	D	E
1	id	core_name	core_classType	com.infa.ldm.axon.governedAxon	com.infa.ldm.bg.governedBgDescri
2	id	name	classType	Axon Glossary Description	Business Glossary Description
3	Frequent_Fliers://infaorcl/SRC_OMNIBUS/FRE	FREQUENT_FLIERS	Table		"A person who regularly travels by p
4	BG_Resource://BG/Term/7757459869942718	Frequent Flyer	Term		
5					

Types of Assets :

Axon Glossary Assets

Business Glossary Assets

Tabular Assets

Rule Assets

Scorecard Assets

Column and Field Assets

Data Domain Assets

Composite Data Domain Assets

Mapping Assets

Reference Assets

1- Axon Glossary Assets :

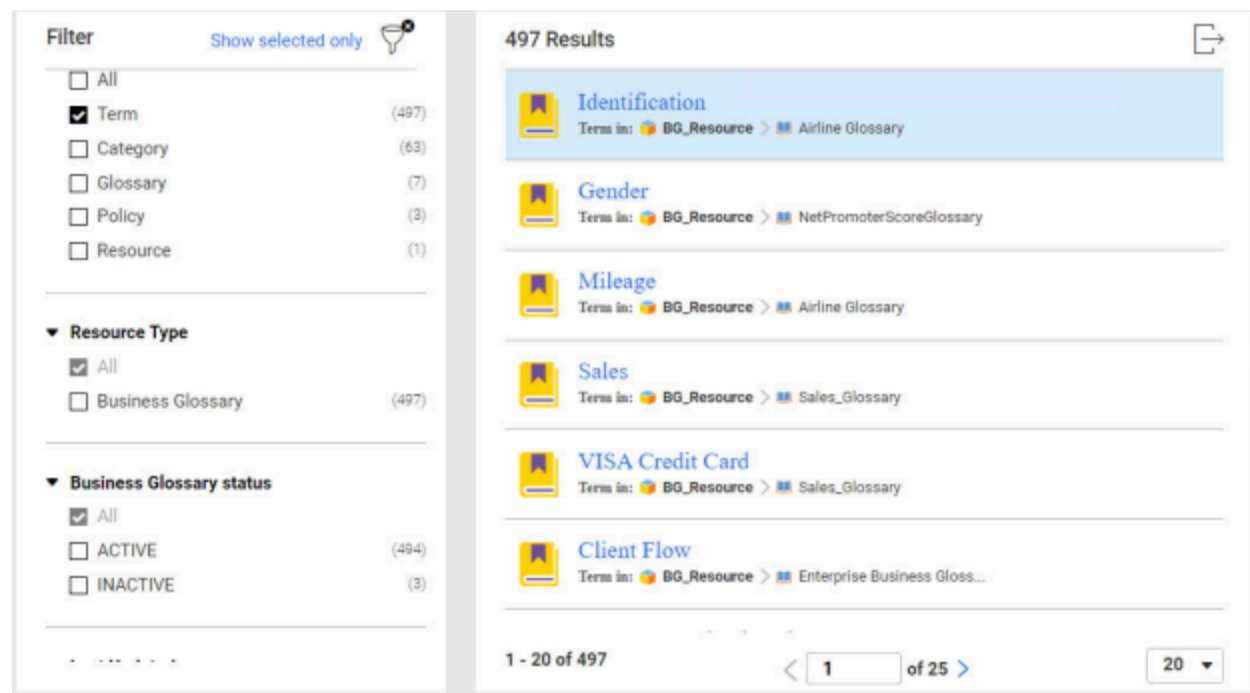
In an Axon Glossary asset, EDC displays the details, related to Axon Glossary assets, classified assets, and related technical assets

- You can assign multiple Axon glossaries as custom attribute values to an asset

The screenshot displays the Axon Glossary Assets interface. On the left, a 'Filter' sidebar contains several sections: 'Resource Name' with 'All' and 'Axon_Glossary' (270) options; 'Estimated Row Count' with input fields and a 'GO' button; 'Reference Assets' with a dropdown set to 'All Assets'; and 'Asset Type' with 'All', 'Axon Glossary' (270), and 'Resource' (1) options. The main area, titled '270 Results', lists assets with a blue header for 'Email' and subsequent entries for 'Insurer', 'Phone', 'Age', 'Country', and 'Transaction Data'. Each entry shows its path: 'Axon Glossary in: Axon_Glossary > Party Data > Party Data - Core Entitie...'. At the bottom, pagination shows '1 - 20 of 270' and a page number '1' of '14'.

2- Business Glossary Assets :

- The Business Glossary assets in the catalog come from the glossaries in the Analyst tool
- You can view the following Business Glossary assets in EDC:
 - Glossaries • Business Terms • Categories • Policies
- You cannot view the lineage and impact for Business Glossary assets



3-Tabular Assets :

You can view details about tabular assets, check the quality of data within it, rate and review them, and ask questions about them.

- Examples of tabular assets:
 - Business Intelligence dimension, logical, and presentation tables • CSV, JSON, and XML files
 - Hive tables • Relational tables, views, and synonyms • Salesforce objects •

You can check the quality of the column or field in a tabular asset if it has profile results.

Frequent Flyer
FREQUENT_FLIERS
★★★★☆ (5)
>> Frequent_Fliers > infoact > SRC_OMNIBUS

Following | Refresh | More | Close

Overview | Columns | Keys | Lineage and Impact | Relationships | Reviews | Questions

Description ⓘ
A person who regularly travels by plane on commercial flights, especially one who is registered as a member of an incentive scheme for such travellers.

Sample Columns [Show All](#)

Name	Business Title	Data Domains ⓘ
FREQUENT_FLIER_ID		Frequent_Flier_ID +1 more
FLIER_ID		Flier_ID
FREQUENT_FLIER_NUMBER		Frequent_Flier_ID(65%)
FREQUENT_FLIER_LEVEL_ID		Frequent_Flier_ID(80%) +1 more

People ⓘ

- Data Owner
Sarah Jones
- Data Steward
Megan Rico
- Followers
Administrator
Jared K
Amanda Smith
Stella Bradford
Richard Hopkins
James Butler
Sarah Jones

4- Rule Assets :

بيحدد شروط تُطَبَّق على بيانات الأعمدة (columns) عند تشغيل Column Profile

5- Scorecard Assets :

- A scorecard is the graphical representation of the valid values for a column or an output of a rule in profile results تعرض نسبة القيم الصحيحة للعمود او المخرجات
- You can configure data quality resource in the Catalog Administrator to view a Scorecard asset in EDC

Add Business Title
Bank_Details_Quality
>> Bank_Branch > Customer > Customer_Data

Follow | Refresh | More | Close

Overview | Where Used | Metric Groups | Lineage and Impact | Relationships | Reviews | Questions

1 Metric Group

99.71%

Default

People ⓘ

- Data Owner
not assigned
- Data Steward
not assigned
- System Attributes**

The **Where Used** tab displays the list of tables and path to the tables linked to the scorecard.

You can check the quality of the data in the data source through the **Metric Group** tab

Add Business Title
Bank_Details_Quality
Bank_Branch > Customer > Customer_Data

Overview **Where Used** Metric Groups Lineage and Impact Relationships Reviews Questions

Name
1 SC_ACCOUNTS

Add Business Title
Bank_Details_Quality
Bank_Branch > Customer > Customer_Data

Overview Where Used **Metric Groups** Lineage and Impact Relationships Reviews Questions

Name	Quality Score	Metrics	Metric Score
Default	99.71%	5	
		ACCT_TYPE	100%
		ACCT_NR	99%
		ACTIVE	100%
		Active_Status	0%
		ACCT_OPENED_BRANCH	100%

6- Column and Field Assets :

يمكنك عرض تفاصيل أكثر عن العمود و الخلية نفسها

Add Business Title
ACCT_OPENED_BRANCH
ORA_BDM > Infoord > ORA_BDM > SC_ACCOUNTS

Overview **Value Frequency** Lineage and Impact Relationships

Description
No description found. Click the icon to add a description for this asset.

Value Frequency
Total Rows: 100
Min: 1002
Max: 9049
Null | Distinct | Non Distinct

Value	Frequency	Percentage
1298	18	18.00

People
Data Owner: not assigned
Data Steward: not assigned

Data Domains
Frequent_Flier...
Flier_ID(69%)

System Attributes

Pattern
9(4) 100 rows 100.00%

7- Data Domain Assets :

- In EDC, you can view details about Data Domains and accept or reject assets for a data domain

من خلال الداتا دومين اقدر اوافق او ارفض الAssets ده عن طريق الشرح ومعرفة كل عمود يرمز لاية و كل خلية بتتكم عن ايه

- Data Domains are inferred for Column and Field assets based on the profiles that users run in Catalog Administrator
- When you curate a data domain, you make the data domain more accurately reflect the type of data that belongs to the data domain

Add Business Title
Airline_Ticket

Overview

Lineage and Impact

Relationships

Description

A document or electronic record, issued by an airline or a travel agency, that confirms that an individual is entitled to a seat on a flight on an aircraft.

Accepted (1)

find

Name	Business Term	Dataset	Assigned By	Inference Type
<input type="checkbox"/> TICKET_NUM		TICKET	Administrator	

People

Data Owner

not assigned

Data Steward

not assigned

System Attributes

8-Composite Data Domain Assets:

مجال البيانات المركب هو مجموعة من مجالات البيانات أو مجالات البيانات المركبة الأخرى

Add Business Title
Customer

Overview

Lineage and Impact

Relationships

Description

Composite data domain to find the customer details.

Found in

Name	Resource Type	Business Term
<input type="checkbox"/> PCRS_SALES_GLOBAL	Oracle	

People

Data Owner

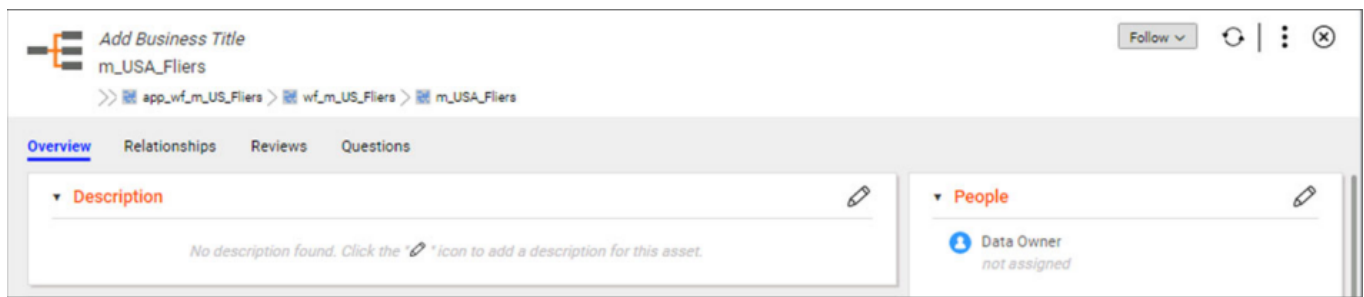
not assigned

Data Steward

not assigned

System Attributes

9-Mapping Assets :



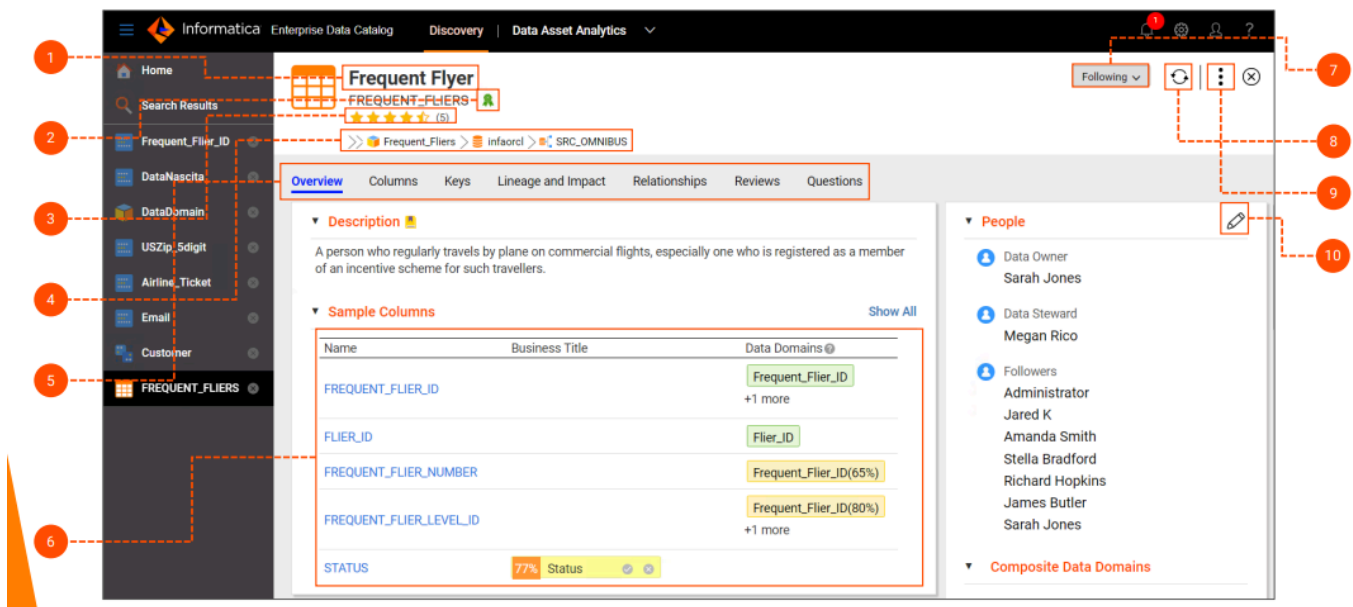
10-Reference Assets

- EDC displays reference resources and reference assets after configuration
- You can configure a resource to get references from other data sources in Catalog Administrator

View Assets :

The Overview Page

The Overview Page

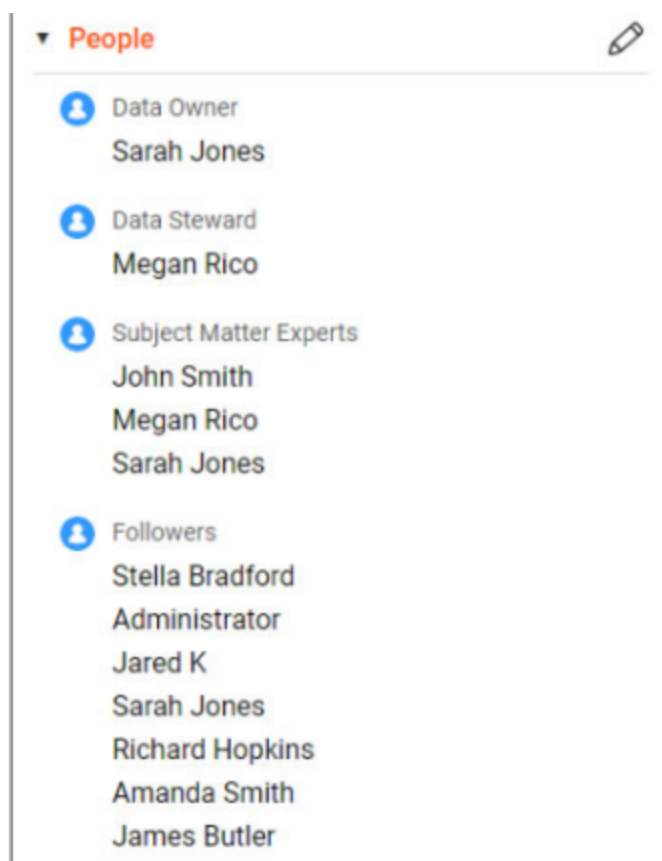


- 1 – Business title associated to the asset
- 2 – Icon indicating asset certification
- 3 – Average rating of the asset
- 4 – Asset breadcrumb
- 5 – Tabs for viewing various pages of the asset

- 6 – Sample columns contained in the table asset
 - 7 – Follow/unfollow the asset
 - 8 – Refresh the asset
 - 9 – Menu options to export asset or certify/decertify asset
 - 10 – Editable properties
-

The People Section :

- Data Owner – User who owns the asset
- Data Steward – User who performs data steward tasks
- Subject Matter Experts – Users who are knowledgeable in the asset contents
- Followers – List of users who follow the asset



The Custom Attributes Section :

- Displays any additional information on the asset

▼ Business Terms

Associated Terms

Frequent Flyer

Travel Type

▼ Custom Attributes

CONTACT

Address

Department

NA

Priority

1

Quality Rating

5

REGION

USA

▼ Asset Notes

This asset is tagged for updates on Lineage and Impact.

▼ Asset Notes

This asset is tagged for updates on Lineage and Impact.

The System Attributes Section :

- Un-editable information extracted at resource level
- Provides information such as:
- Resource name
- Data type
- Last modified time
- Location
- Output format

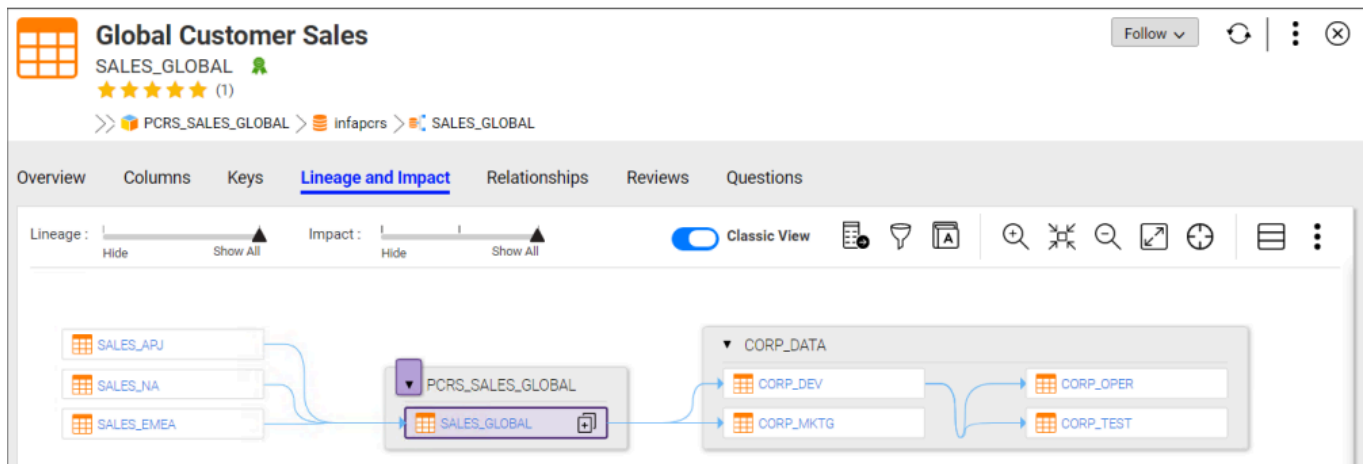
▼ System Attributes	
Input Format	org.apache.hadoop.mapred.TextInputFormat
Last Modified	Feb 26, 2019, 1:22 PM
Location	hdfs://EICBETA38001.informatica.com:8020/data/nps
Output Format	org.apache.hadoop.hive ql.io.HiveIgnoreKeyTextOutputFormat
Resource Name	LakeProduction
Resource Type	Hive
Rows covered by profiling	129880 sample rows.
SerDe Library	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
Asset Type	Table

Types of Assets :

- EDC has many types of assets – as many as 75 asset types
- Common assets:
- Resource
- Database
- Schema
- Table
- Column
- Field
- Report
- BusinessObjects
- Terms

75 Asset Type		find
Name	Count	
<input type="checkbox"/> BusinessObject	13064	
<input type="checkbox"/> Report	9365	
<input type="checkbox"/> Term	6763	
<input type="checkbox"/> Policy	6408	
<input type="checkbox"/> Table	4430	
<input type="checkbox"/> Operation	3480	
<input type="checkbox"/> DataSource	2822	
<input type="checkbox"/> View Column	2365	
<input type="checkbox"/> Show only assigned		

View Data Lineage :



The data flow for an asset has two components:

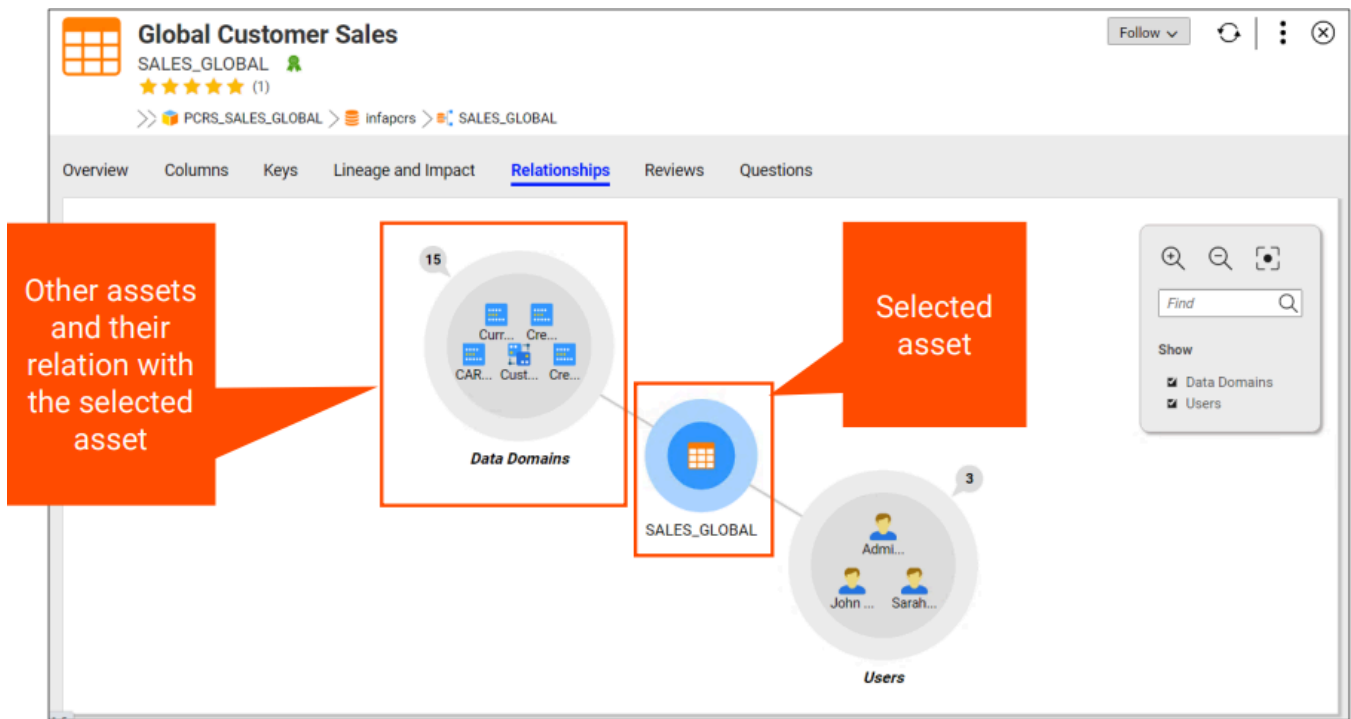
- Lineage – describes the flow of data from the origins to an asset

تقول الداتا جايه منين ورايحه فين

- Impact – describes the flow of data from an asset to its destination

بنقول لو حصل حاجه في الداتا ممكن تأثر علي ايه

View Relationships :



Module Summary :

This module showed you how to:

- Navigate through the user interface of Enterprise Data Catalog (EDC)
- Explore the advantages of semantic search
- List additional search features
- View search results
- Explore the Assets view