IBM **Guardium**
Discover and Classify

SUPERVISED AI USER GUIDE

IBM GUARDIUM DISCOVER AND CLASSIFY

VERSION 4.1.1

TABLE OF CONTENTS

SUPERVISED AI OVERVIEW

ABOUT SUPERVISED AI

Supervised AI is a wizard-based machine learning tool enabling SQL database and data lake trainers to build, modify, and train AI models to identify personal and sensitive information based on automatically discovered results.

The Supervised AI module visualizes the possible sources of personal data (candidate virtual views) and assists in training the system to accept or reject candidates as trusted data. You can also create root data assets and apply system-detected virtual views as root data assets, as well as review and export the candidates discovered by IGDC in various data sources.

Data training using the Supervised AI tool improves the quality of personal information identification during IGDC automatic discovery and supports customization of the system according to network and business specifics.

Supervised AI is based on the master catalog of sensitive personal data and is available for the cataloged data elements when global data cataloging is enabled in the master catalog engine settings (Settings > Data recognition > Data element configuration).

For details, see the Analytic Engine and Console Management Administrator Guide.

DATA TO BE TRAINED

The source of information of the Supervised AI is an automatically discovered group of candidates featuring an identical set of data elements and identified in the same virtual view and SQL database/data lake. A *candidate* is a set of personal data retrieved by the SQL database/data lake analyzer, and not confirmed against any RDA.

If no RDAs are configured, all personal data is considered candidates.

DATA TRAINING RESULTS

Data training improves personal data identification accuracy and allows setting a standard (RDA) to verify discovered personal information instances. The Supervised AI module allows achieving one of the four results as described in the table below.

Table 1: Data Training Results

#	RESULT	DESCRIPTION
1	**Valid virtual view.**	User accepts the candidate virtual view as verified trusted data. The system will map the associated personal information instances to the existing RDA records. If a personal information instance is mapped to the existing RDA, it will be a verified data standard for newly discovered data subjects. If a personal information instance is mapped to the default data asset, it will be available in the data asset manager and personal information search.
2	**Ignored virtual view.**	User identifies the candidate virtual view as a false-positive discovery and rejects its usage by IGDC. This action prevents retrieval of the same false positives, as the IGDC DB analyzer will ignore the rejected virtual view in subsequent analysis cycles.
3	**Virtual view marked as RDA.**	User approves the candidate virtual view as a root data asset. The data subjects will be used as a standard of verified data for newly discovered candidates. It will also be available in the Data Asset Manager for custom data asset configuration.
4	**RDA/virtual view**	User creates a virtual view entry from scratch and accepts/rejects it as valid

#	RESULT	DESCRIPTION
	created from scratch.	information or approves it as a root data asset.

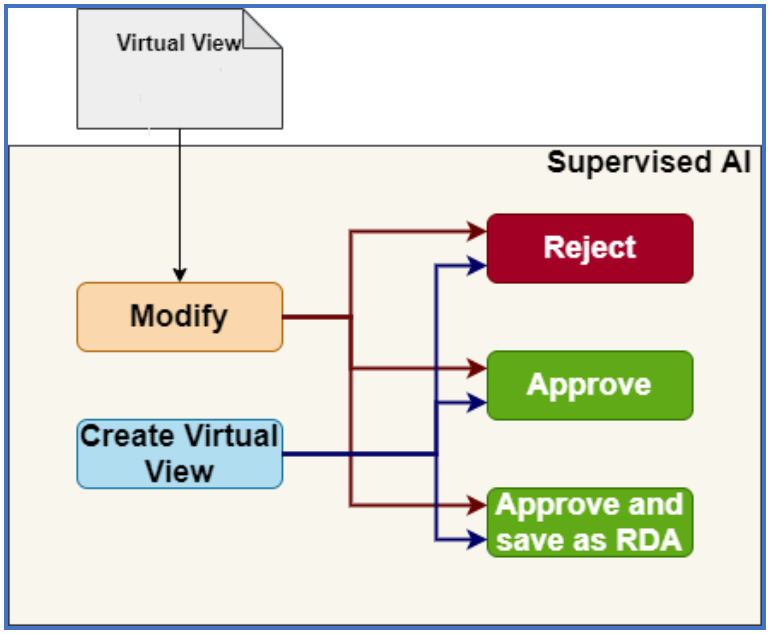


Figure 2: Data Training Results

CHAPTER 1: VISUALIZATION OF GROUPS OF CANDIDATES

The Supervised AI landing page shows the results of automatic SQL database/data lake analysis within the organizational network and assists in selecting data to be trained for improved subsequent discovery activities.

IGDC is designed for operation with the latest version of Google Chrome. Using other browsers is not recommended and may affect performance and functionality

The page consists of two parts: the main window with groups of candidate virtual views (bubbles) and the sidebar with filters for sorting the presented data. The **Create RDA/Virtual View** button in the upper right corner redirects you to the page for creating virtual views and mappings from scratch.

Figure 1: Supervised AI landing page

MAIN WINDOW

The main window provides a visual display of candidate virtual views for data training in two formats: "bubble" and list.

A virtual view is a set of tables or schemas used as a source for a specific group of candidates. You can see the total number of groups of candidates available for training at the top of the page. This number changes in accordance with the selected filters or search results.

Table 2: Supervised AI landing page formats

FORMAT	DESCRIPTION
Bubble format	Each bubble represents a candidate virtual view from a data source (SQL database or data lake), you can view a number of bubbles with the same data source but different virtual views. The bubble visualizes metadata in normal and hover views. The main window shows a maximum of 27 bubbles of three sizes - large, medium, and small (nine of each size) with brief information on the group of candidates in each bubble. The candidate group size is relative and depends on other filters: e.g. a specific group may be considered small relative to all groups, but the same group will be considered "large" if you select a filter such as MySQL. To see additional information, hover over the desired bubble. To begin training the desired group of candidates, hover over the desired group of candidates (bubble) and click the **Train** button.
List format	The list format shows all available candidate virtual views.

FORMAT	DESCRIPTION
	Each row represents a candidate virtual view from a data source (SQL database or data lake), you can view a number of bubbles with the same data source but different virtual views. The list provides metadata visualized in collapsed and expanded views. To begin training the desired group of candidates, click the desired data source name.

The table below describes the bubble information provided in normal and hover views. When you hover over the bubble (hover view), the bubble shows more parameters. When you do not hover over the bubble (normal view), the bubble shows less parameters.

Table 3: Information in bubbles - normal and hover view

PARAMETER	DESCRIPTION	NORMAL VIEW	HOVER VIEW
Bubble name	Name of the subject data source.	✓	✓
Schema name	Name of the schema related to the candidate virtual view. You can click the (**Copy**) icon to copy the schema name to the buffer.	✓	✓
Hostname	Hostname of the candidate virtual view. You can click the (**Copy**) icon to copy the hostname to the buffer	✓	✓
Estimated candidates	Estimated number of personal information instances retrieved from the data source but not confirmed against RDA. **Note**: This value is propagated only for the Apache Hive product. For other products, the field shows "**-**".	✓	
Estimated unique candidates	Estimated number of unique personal information instances retrieved from the data source but not confirmed against RDA. **Note**: This value is propagated only for the Apache Hive product. For other products, the field shows "**-**".	✓	
Actual Candidates	Total actual number of personal information instances retrieved from the data source but not confirmed against RDA.	✓	✓
Actual Data Subjects	Total number of personal information instances retrieved from this specific data source and validated against at least one RDA. Such instances are considered trusted and are not included in data training. This number is provided to give additional visibility to the data inside this data source		✓
Product	Name of the data source product.		✓
Type	Source data source type (SQL database or data lake).		✓
Processing	Type of data processing in the data source. **Data at rest**: Data subjects that were identified in the data source stored content.		✓
Analytic Engine	Name of the IGDC analytic appliance that connected to the data source and discovered the candidate virtual view.		✓
Last Analysis	Date and time when the data source was last analyzed. Format: Month dd, yyyy hh:mm AM/PM.		✓
Train button	Button redirecting to the data training page of the desired virtual view.		✓

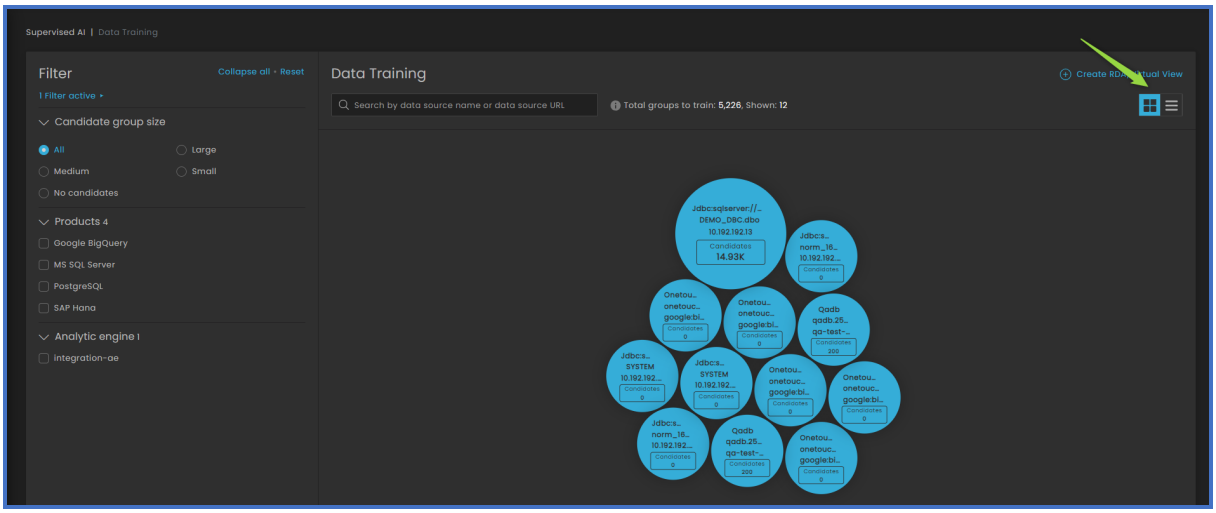


Figure 4: Landing page: Bubble format

Table 5: Information in list format

PARAMETER	DESCRIPTION
Collapsed View	
Data source name	Name of the subject data source. Click the data source name to go to the data training page.
Estimated candidates	Estimated number of personal information instances retrieved from the data source but not confirmed against RDA. **Note**: This value is propagated only for the Apache Hive product. For other products, the field shows "**N/A**".
Estimated unique candidates	Estimated number of unique personal information instances retrieved from the data source but not confirmed against RDA. **Note**: This value is propagated only for the Apache Hive product. For other products, the field shows "**N/A**"
Actual Data Subjects	Total actual number of personal information instances retrieved from this specific data source and validated against at least one RDA. Such instances are considered trusted and are not included in data training.
Actual candidates	Total number of personal information instances retrieved from the data source but not confirmed against RDA.
Schema name	Name of the schema related to the candidate virtual view.
Data source URL	URL of the data source related to the candidate virtual view.
Expanded View	
Data Subjects	Total actual number of personal information instances retrieved from this specific data source and validated against at least one RDA. Such instances are considered trusted and are not included in data training.
Data source type	Data source type (SQL database or data lake).
Processing type	Type of data processing in the data source
Last analysis time	Date and time when the data source was last analyzed. Format: Month dd, yyyy hh:mm AM/PM.
Analytic engine	Name of the IGDC analytic appliance that connected to the data source and discovered the candidate virtual view.
Data source ID	Unique identifier of the data source in IGDC.

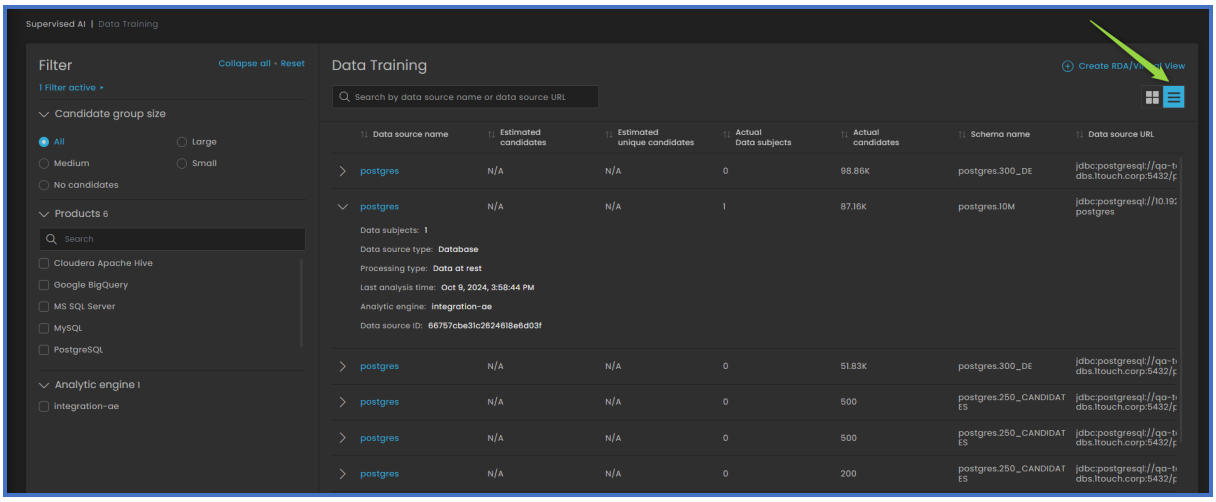


Figure 6: Landing page: List format

SORTING VISUALIZED GROUPS OF CANDIDATES (BUBBLES)

You can search for a group of candidates by data source name or data source URL via the search box at the top of the main window - or sort the groups of candidates (bubbles) using the filters on the sidebar.

CANDIDATE GROUP SEARCH

To find a specific group of candidates to train, use the **Search** box. Begin typing the data source name or the data source URL and select the desired option from the dropdown list. You can enter multiple values for each search option. The main window will show the associated groups of candidates (bubbles) or the **No results** screen if no group matches the selected parameter.

Table 7: Search parameters

PARAMETER	DESCRIPTION
Data source name	Data source name shown at the top of the bubble.
Data source URL	URL for connection to the data source. The bubble shows it under the data source name.

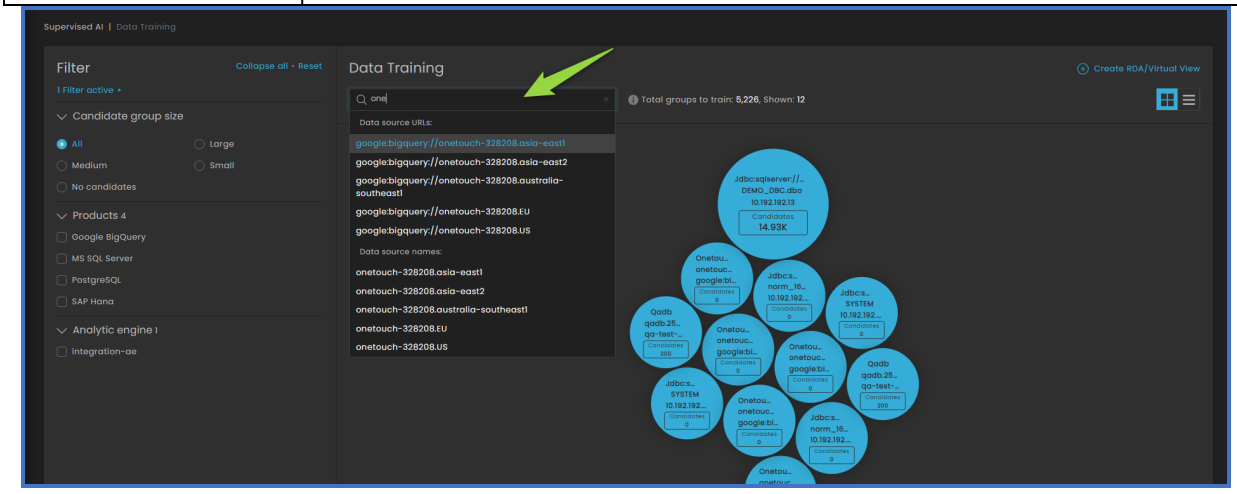


Figure 8: Search for system-detected candidate virtual views

SIDEBAR FILTERS

The **Sidebar** shows the options for filtering the groups of candidates (bubble and list view) by size, product, and analytic engine. The options shown in the filters (except for the candidate group size) depend on the specific properties of the discovered candidates and may change along the data training process as new groups of candidates are discovered.

To disable all selected filters, click the **Reset** button (**4**).

Table 9: Sidebar filters

NAME	DESCRIPTION
Candidate group size	Relative size of the group of candidates to train. The system divides all groups of candidates (bubbles or selected groups) into three groups by the number of identified candidates, visualized by the bubble's size. You can choose to start data training with larger or smaller groups. Options: All, Large, Medium, Small.
Product	Name of the data source product. The filter shows the products of the candidate virtual views discovered by IGDC in your network.
Analytic engine	Name of the IGDC analytic engine (appliance) connected to the data source. The filter shows the analytic engines that discovered the candidate virtual views in your network.

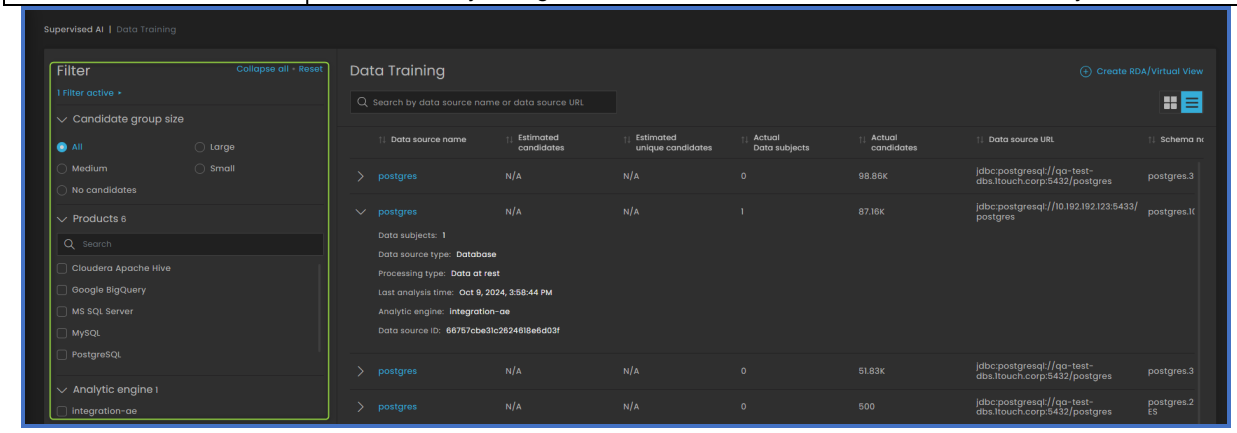


Figure 10: Filters on sidebar

CHAPTER 2: SUPERVISED AI DATA TRAINING

The Supervised AI module allows you to train data as follows:

- Modify the automatically identified data lineage and map the discovered data elements to IGDC-supported entities within a group of candidates.
- Create virtual views and mappings from scratch.

CANDIDATE VIRTUAL VIEW TRAINING

Training of a system-detected candidate virtual view is implemented in three steps:

1. Review & verify: Review candidate virtual view structure and field-to-data element mappings.

2. Modify: Edit field-to-data element mappings, add or delete fields from the candidate virtual view, add table relationships.

3. Train: Set the virtual view as valid, ignored or save it as an RDA.

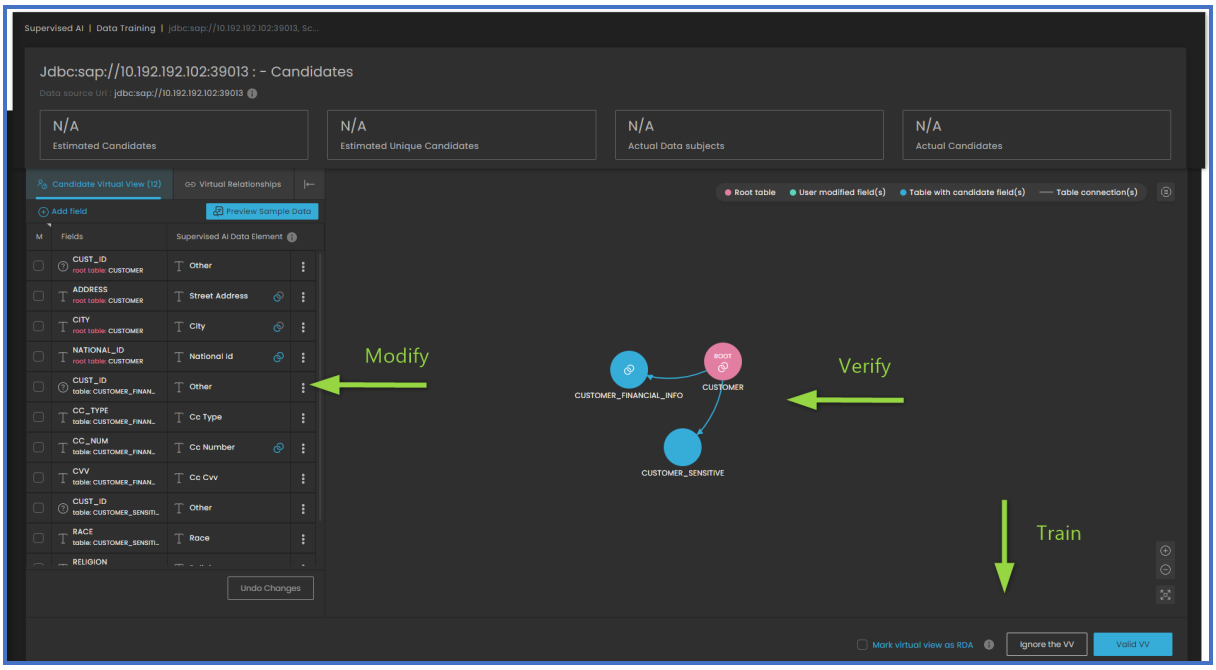


Figure 1: Data training steps

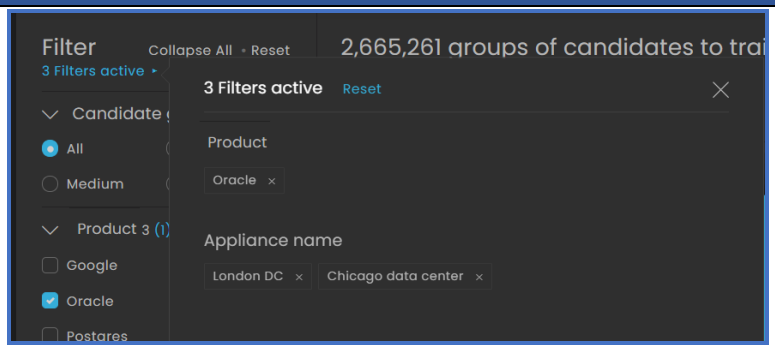
VERIFY

To begin training a group of candidates, hover over the group (bubble) on the Supervised AI landing page and click the **Train** button.

You will be redirected to the **Data Training** page.

Table 2: Data Training page elements

ELEMENT	DESCRIPTION
Candidate group info	Name of the subject data source and the number of candidates in the group to be trained. Hover over the (Info) icon to see the auxiliary information, including the data source vendor and type, appliance name, last analysis time, and schema name. This information is also available in the hover view of the candidate group (bubble).

ELEMENT	DESCRIPTION
	 *Figure 3: Figure 7: Auxiliary information on the candidate virtual view in the Data Training page*
Visualization	Graph (map) of the table links in the virtual view.
Mapping	Table of the virtual view columns mapping to the IGDC data elements.
Actions	Buttons for rejecting the candidate virtual view or approving it as verified personal data and a root data asset.

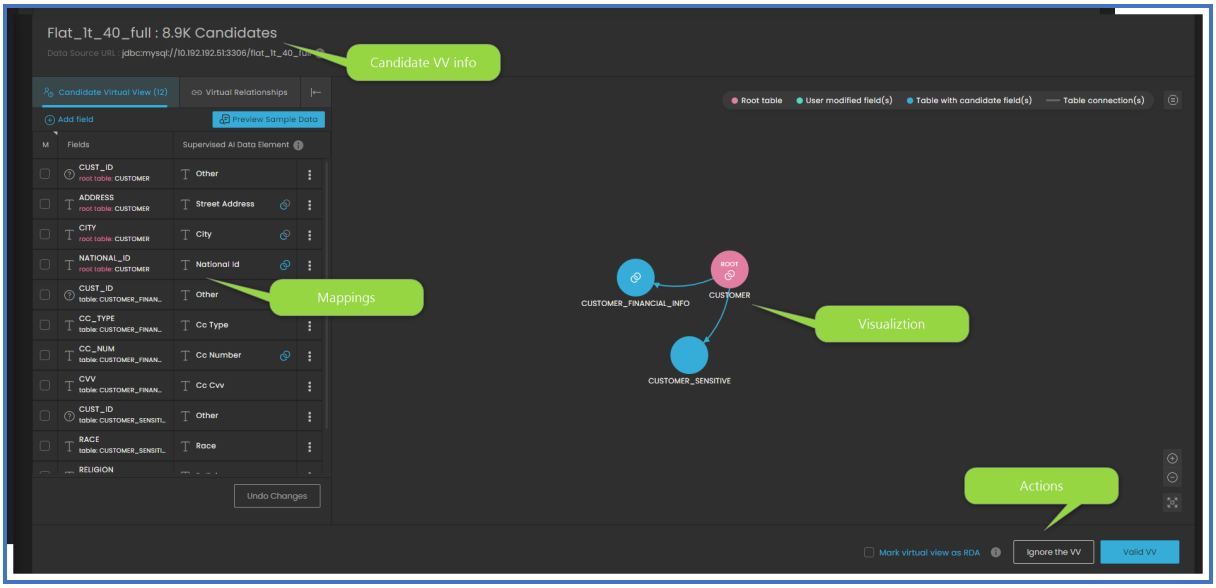


Figure 4: Data Training page elements

Explore The Virtual View Structure

To facilitate virtual view reviews, the right part of the **Data Training** page visualizes the candidate virtual view as a map of tables and links between them.

Each circle (**3**) on the map represents a table in the virtual view with the table name below the circle. The circle color indicates whether the associated table has been modified by a user in the **Candidate Virtual View** (**1**) or **Virtual Relationships** tabs (**2**). The circle color indicates its status.

Table 5: Circle colors in the virtual view map

CIRCLE COLOR	DESCRIPTION
Pink	Root table of the candidate virtual view. Each virtual view must contain at least one root table.
Blue	No user modifications. All circles are blue when you open the **Data Training** page.
Green	Table fields have been modified by users in the **Candidate Virtual View** tab (**1**). You can highlight the edited rows by clicking the map's associated circle (**3**).

CIRCLE COLOR	DESCRIPTION
Grey	Table that is not included in the virtual view. It can be manually added by a user in the **Virtual Relationships** tab (**2**). Grey circles are only shown with the **Virtual Relationships** tab selected and are excluded from the **Candidate Virtual View** tab.

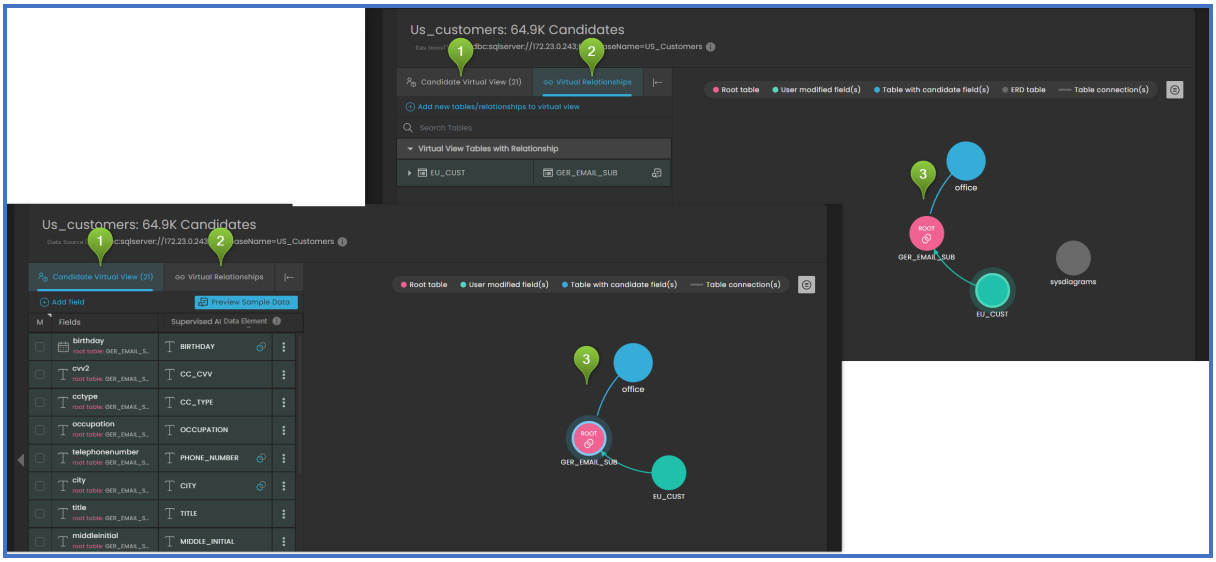


Figure 6: Virtual view map with the selected Candidate Virtual View tab (back) or Virtual Relationships tab (front)

If the virtual view is large, you can hide the mapping tables by clicking the (**Hide**) icon and explore the map full-screen. To recover the mapping tab, click the (**Expand**) icon.

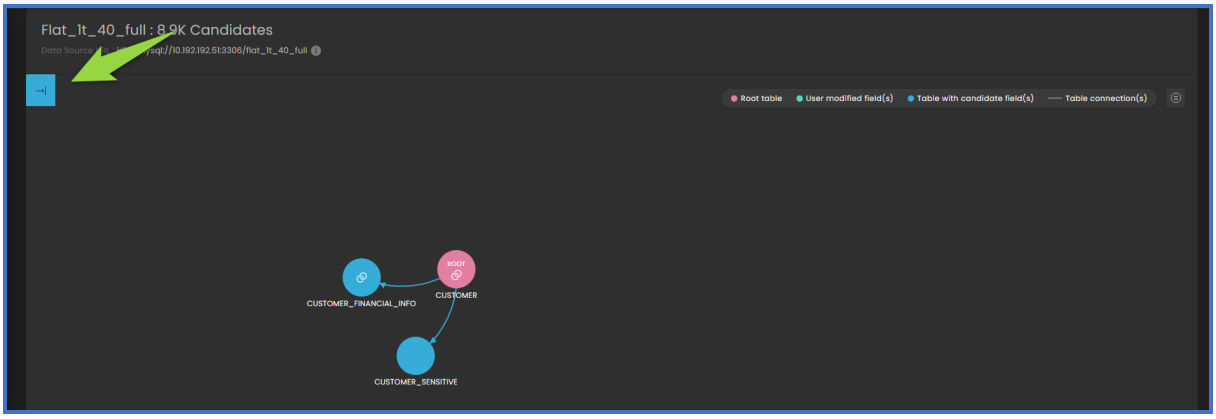


Figure 7: Virtual view map. Expanded view

Verify The Virtual View

To improve the quality of data identified by a candidate virtual view, you can review and modify the virtual view structure in the left part of the **Data Training** page.

Verify Mappings

When retrieving data from a specific field (=column) in the table, the system assigns a specific type (=data element) such as given name, tax ID, gender, etc.

The **Candidate Virtual View** tab displays mapping of table fields (=column headers) to data elements supported by IGDC.

Table 8: Candidate Virtual View tab elements

ELEMENT	DESCRIPTION
Field column (**1**)	Column header and table name of the field in the candidate virtual view. The fields from the root table are additionally highlighted: root table:.The icon defines the value type. : Integer data (numbers). : String data (text). : Timestamps (date and time).
Supervised AI Data Element column (**2**)	Name used by Supervised AI and IGDC for specific personal data types. For example, CC_NUMBER for credit card numbers. To see the supported data elements hover over the (**Info**) icon (**5**).
M columns (**3**)	Mandatory checkboxes tag the desired field as mandatory (not empty) for data retrieval. By default, all checkboxes are disabled. **Enabled**: The system will verify that the field is not empty before retrieving the data subject's info. If the mandatory field is empty, the system will skip the whole row. **Disabled**: If the field is not marked as mandatory and is empty, the system will retrieve other values from the row.
Constraint icons (**4**)	Icons indicating that the data element is a single-data element constraint or as a part of a combined constraint. A constraint specifies a unique person in IBM Guardium inventory of personal data. To create a high-quality inventory of unique persons, the virtual view must contain at least one constraint. To see the IGDC constraints hover over the (**Info**) icon (**5**). You can configure the constraints in the Data element configurator. : Icon that tags the data element as a constraint. : Icon that tags the data element as a part of a combined constraint. Hover over the icon to see the other data elements in this constraint.

Table 9: Candidate Virtual View tab actions

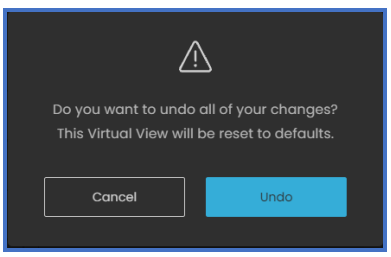
ACTION	DESCRIPTION
Add field (**6**)	Opens a form for adding a new mapping to the candidate virtual view.
Preview Sample Data (**7**)	Shows 30 rows with sample data instead of the virtual view map.
Edit (**8**)	Open a form for modifying the selected mapping (> **Edit**).
Delete (**9**)	To delete the selected mapping from the candidate virtual view (> Delete).
Undo Changes(**10**)	Resets all user mapping modifications and returns to the default virtual view detected by IGDC. The system will ask to confirm the action. Once you click **Yes**, the circles on the virtual view map will turn blue, and the column names in the sample data table will turn grey.  *Figure 10: Confirmation popup to undo changes in the Update Virtual View tab*

Figure 11: Candidate Virtual View tab elements and actions

Verify Relationships

The group of candidates is based on a virtual view. The **Virtual Relationships** tab shows the following:

- List of established relationships (**1**) between the virtual view tables. Click the  (**Expand**) icon to see the field-to-field matches within the table relationship. The field type is visualized by the icon ( stands for integer data (numbers),  stands for string data (text),  stands for timestamps (date and time). The  (Link) icon color indicates if the field-to-field match is included in the virtual view (blue - included, grey - not included). If a match has been modified, it is highlighted by a vertical green line - 

.
- List of potential relationships (**2**) matching the virtual view tables to other schema tables. The left column shows the tables from the virtual view, and the right part - schema tables not included in the view.

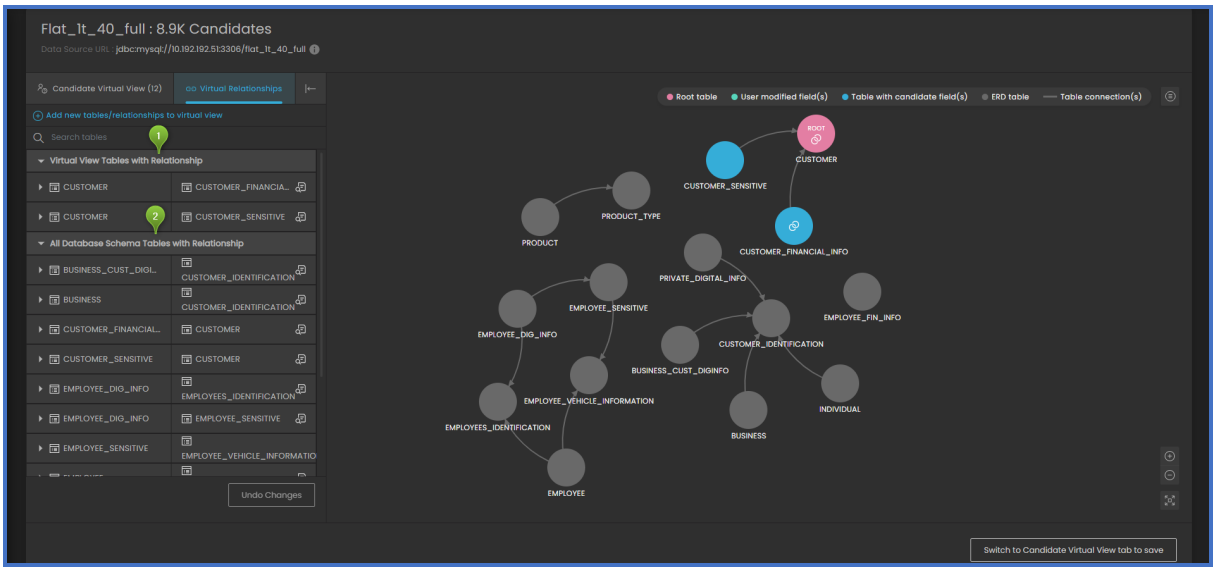


Figure 12: Virtual Relationships tab elements

The **Undo Changes** button below the tab cancels all modifications made in the tab and returns table relationships to the default state. The system will request confirmation for this action.

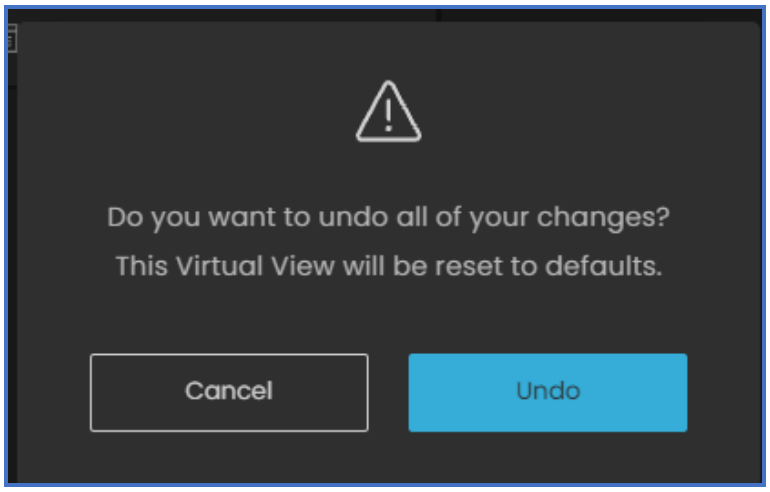


Figure 13: Confirmation popup to undo changes in the Virtual Relationships tab

Once you click **Yes**, the circles on the virtual view map will turn blue and grey; column names in the sample data table will turn grey.

Check Real Sample Data

To facilitate verification of the virtual view table relationships and field-to-data element mappings, IGDC offers data samples in the **Candidate Virtual View** and **Virtual Relationships** tabs.

To extract high-quality and useful sample data, the system skips the rows with many empty fields.

In the **Candidate Virtual View** tab, the sample data enables you to confirm that the field-to-data element mappings are correct. Click **Preview Sample Data** (**1**) to open 30 rows of sample data for 30 unique persons instead of the virtual view map in the right part of the screen. The sample data is grouped by table name (**2**) and column name (**3**) within the table and is useful when modifying the mappings.

All data samples in the row relate to the same individual only if the candidate virtual view has not been modified by the user.

If you modify a mapping in the **Candidate Virtual View** tab, the system will not re-match the values in the associated sample data column, meaning that the fields in the column will not be related to unique persons in the rows but will actually represent the sample of data from the data source.

The header of the modified field is highlighted green (**4**).

Note: If the data element is encrypted, the value will be masked (******) or hashed, with the ability to unlock and view the value by clicking the (**Locked**) icon. This user action will be logged.

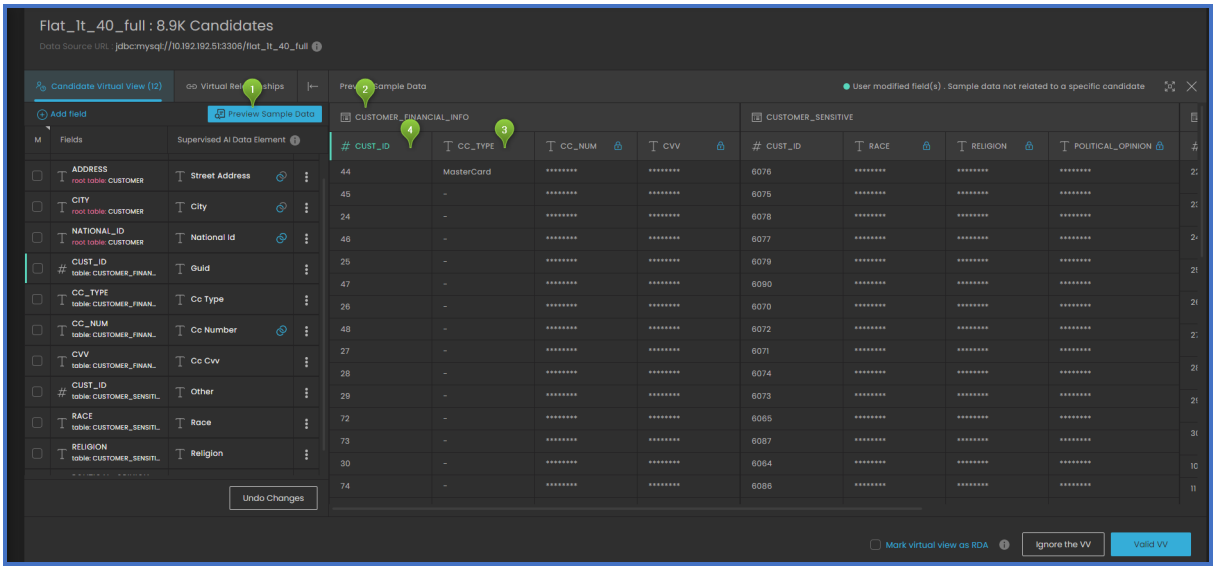


Figure 14: Sample data for the Candidate Virtual View tab

The sample data in the **Virtual Relationships** tab enables you to verify that field-to-field matches are correct for a specific table relationship.

Click the icon (**1**) for the desired table relationship, or click the (**Options**) icon (**2**) and select **Preview data** (**3**). The right part of the screen will open 30 rows of sample data instead of the virtual view map.

The sample data is grouped in pairs by related tables (**4**) and matching fields (**5**) from the table. If a mapping has been modified, the field's header is highlighted **green**.

In addition to the matching fields from the virtual view, the sample data provides combinations of other fields, simplifying the search for other field-to-field matches to add them to the virtual view.

The (**Link**) icon color indicates if the field-to-field match is included in the virtual view (**blue** - included, **grey** - not included). The sample data begins with field matches followed by the potential combinations.

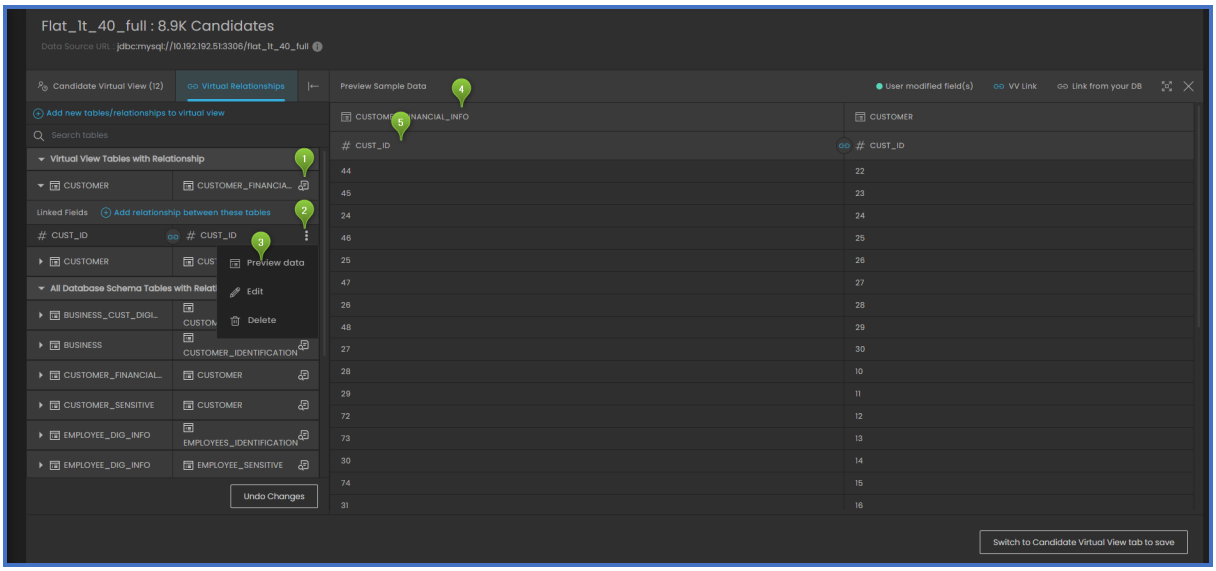


Figure 15: Sample data overview

To hide the sample data, click the **X** (**Close**) icon. If the virtual view is large, expand the table by clicking the  (**Expand**) icon and explore the sample data full-screen.

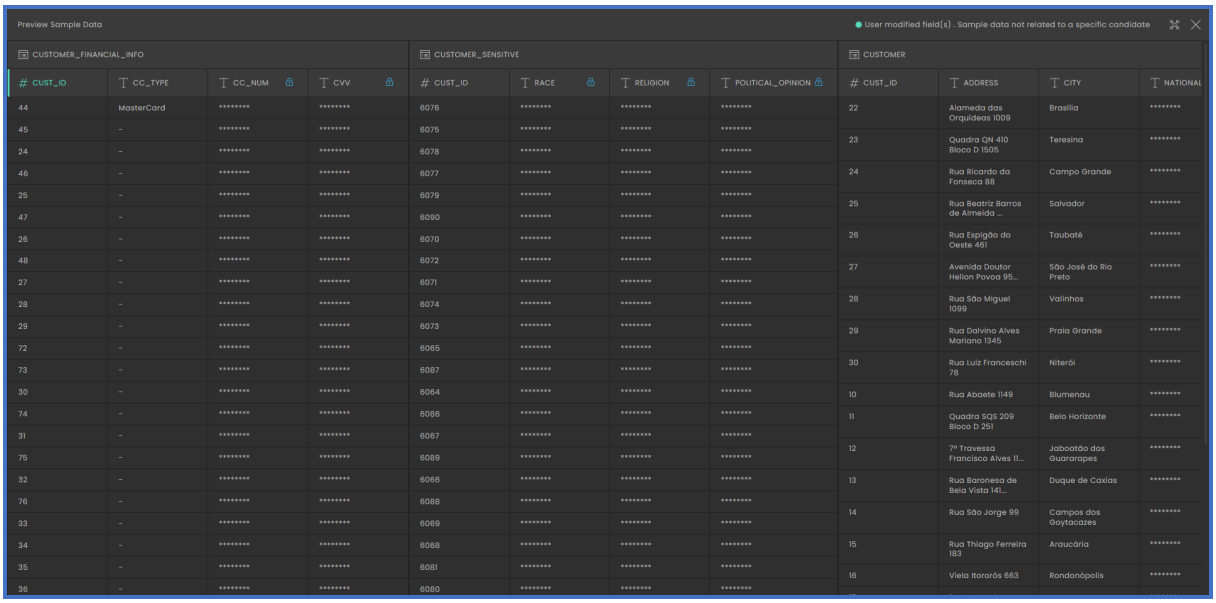


Figure 16: Sample Data of the Candidate Virtual View. Expanded view

MODIFY

Users can modify the virtual view structure on a mapping and structural level.

Mapping Modification

The Candidate Virtual View tab enables you to modify candidate quality by editing the field-to-data element mapping.

You can always return to the initial state of the virtual view mappings by clicking **Undo Changes**. The system will cancel all the user modifications - added, edited, or deleted mappings.

Edit Mapping

The **Candidate Virtual View** tab allows you to edit mappings if the virtual view field does not match the Supervised AI data element. For example, the system mis-identified a passport number as a user id.

To edit a mapping, click the (**Options**) icon and select **Edit**. In the popup window, change the source table, field, or the matching Supervise AI data element. Click **Save** to apply the new mapping or **Cancel** to exit without saving changes.

Table 17: Editable mapping parameters

PARAMETER	DESCRIPTION
Table dropdown list (**1**)	Virtual view table with the desired field. The dropdown list contains only tables from the detected virtual view (not all tables). To configure this table as a root table, check the **Set as a root table** checkbox under the field.
Field dropdown list (**2**)	Virtual view field. The dropdown list contains only fields from the table selected above.
Supervised AI Data Element dropdown list (**3**)	data element supported by the Supervised AI, which matches the virtual view field. The data element can be labeled by a constraint icon. : Data element is a constraint. : Data element is a part of a combined constraint.

Figure 18: Editing the field-to-data element mapping

Add Mapping

To add a mapping to the virtual view structure, click **Add field** in the **Candidate Virtual View** tab. In the popup, select the virtual view table & field and the matching Supervised AI data element. The content of the dropdown lists is identical to the mapping editing popup.

Once you save a new mapping, the affected table's circle on the virtual view map will turn green, and the relative column will be added to the sample data table.

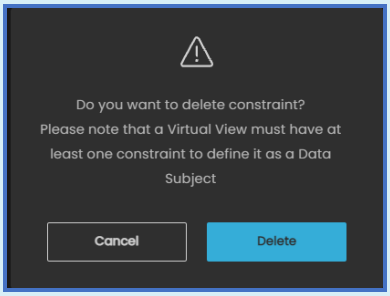
Figure 19: Adding a field-to-data element mapping

Delete Mapping

To delete a mapping from the virtual view structure, click the (**Options**) icon and select **Delete**. The relative column will be deleted from the sample data table. If other mappings exist in the affected table, the related circle on the virtual view map will turn green. If the deleted mapping was the last one in the table, the related circle will disappear from the map.

The candidate virtual view must contain at least one constraint.

If you delete a single-data element constraint ( ) or a part of a combined constraint ( ), configure a new constraint before approving or rejecting the candidate virtual view. Before deleting a constraint, the system requests confirmation. In the popup, click **Delete** to remove the mapping or **Cancel** to exit without deletion.



Confirmation popup for constraint deletion

Figure 20: Field-to-data element mapping deletion

Table Relationships Modification

The **Relationships** tab presents relationships between candidate virtual view tables and adds relationships to the virtual view.

Relationships are based on common fields in two tables.

Add Table Relationship

If the system has not automatically linked one or more tables to the candidate virtual view, you may need to add a table relationship. In a new relationship, one table must belong to the virtual view, and the second can be any table (grey circle in the map).

1. Go to the **Virtual Relationships** tab and add a table relationship by clicking **Add new tables/relationships to Virtual view** (**1**).

2. In the **Add new relationship** popup window, fill all fields. If necessary, begin typing the table or field name in the **Search** box to find the desired item quickly.

Table 21: Fields in the Add new relationship popup window

FIELD	DESCRIPTION
Select Table from dropdown list (**2**)	List of tables belonging to the virtual view (blue and green circles in the map), excluding the tables not linked to the virtual view (grey circles).
Select Field from dropdown list (**3**)	List of all fields in the table selected above (**2**).
Select Table to dropdown list (**4**)	List of all tables from the schema (blue, green, and grey circles in the map).
Select Field to dropdown list (**5**)	List of all fields in the table selected above (**4**).

3. Click the **Save** button (**6**) to add the relationship to the list or the **Cancel** button to return to relationship management without saving the changes.

Once you click **Save**:

- The table relationship will be added to the list in the **Virtual Relationships** tab.
- The table will be added to the candidate virtual view, and the related circle will turn green in the virtual view map, and it will be shown in the map when the **Candidate Virtual View** tab is selected (grey circles are excluded in such cases).
- The table added to the virtual view will be available in the **Select Table from** dropdown list (**2**) when adding subsequent relationships. You can use any data elements from these tables for data training. However, the linking fields are not automatically included in the Candidate Virtual View, as such fields are usually not mapped to the IGDC data elements.

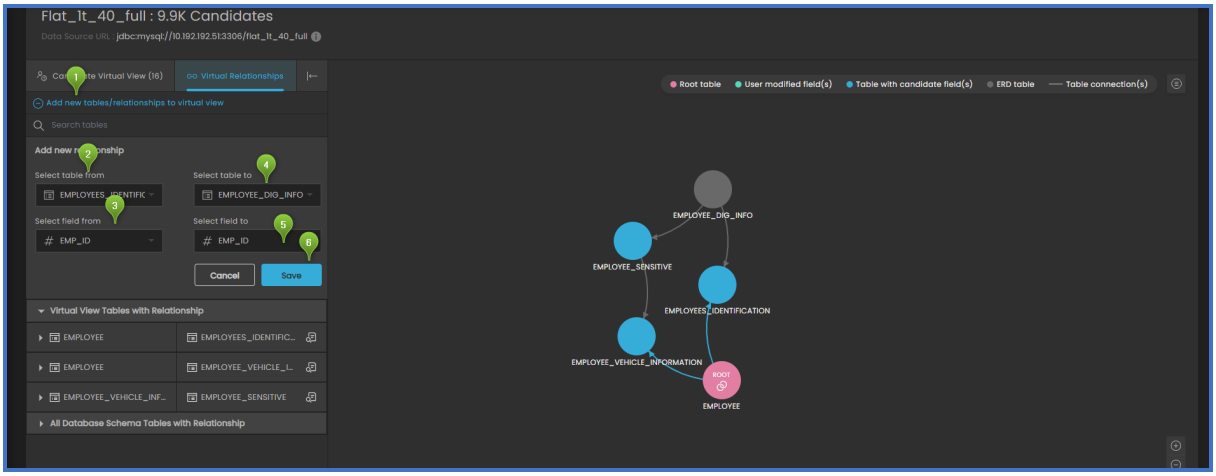


Figure 22: Adding a table relationship in the Virtual Relationships tab

Add Field Relationship

You can add a field-to-field match to the table relationship of the candidate virtual view.

Click **Add relationships between these tables** (**1**) in the **Virtual Relationships** tab. The tables in the **Table from** and **Table to** fields will be preselected and non-editable.

Select the matching fields from the **Select Field from** (**2**) and **Select Field to** (**3**) dropdown lists. Click **Save** to add the match or **Cancel** to close the popup without saving changes.

Once you click **Save**:

- The field-to-field match will be added to the table relationship in the **Virtual Relationships** tab.
- Table-related circles will turn green in the virtual view map.
- The table fields will be available for mappings in the Candidate Virtual View tab at the top of the list for mapping to the Supervised AI data element.

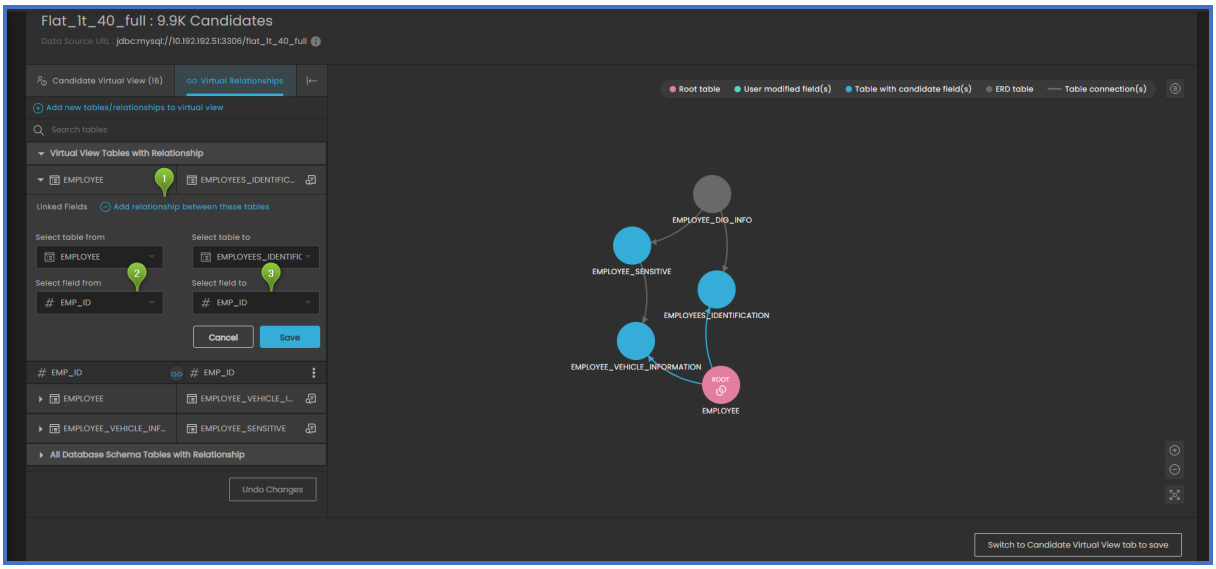


Figure 23: Adding a field-to-field match to the existing table relationship

Edit Relationship

To modify a field-to-field match, expand the desired table relationship, click the  (**Options**) icon to edit the match, and select **Edit**. In the popup window, change the **Select Field from** and/or **Select Field to**. Click **Save** to modify the match or **Cancel** to exit without saving changes. Once you click **Save**, the related circles in the virtual will turn green.

When adding a new relationship, one table must already exist in the virtual view.

Create relationships between tables excluded from the virtual view by linking excluded tables to those included in the virtual view you are working with.

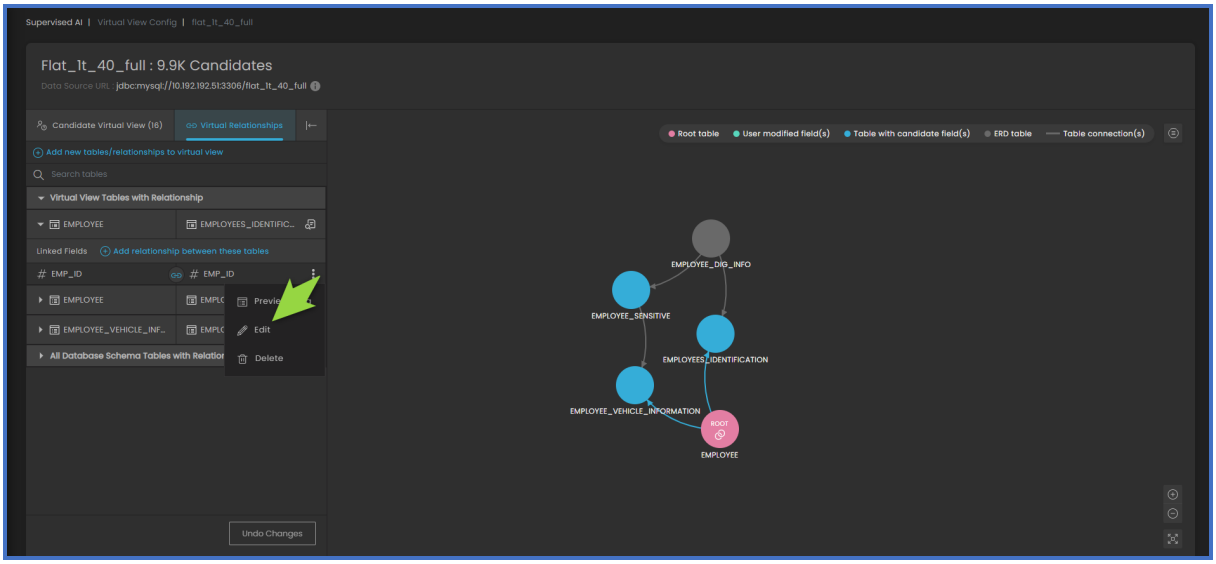


Figure 24: Editing field relationships

Delete Relationship

To delete a field-to-field match from the table relationship, expand the desired table relationship, click the **Options** icon for the match to be deleted, and select **Delete**.

If other field-to-field matches exist in the affected tables, the related circles on the virtual view map will turn green. If the deleted match was the last one in the table relationship, the related circles will turn grey in the virtual view map.

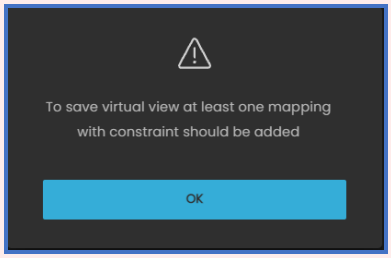
Deletion of relationships does not require confirmation. Once you click **Delete**, the relationship will disappear from the subject virtual view.

Figure 25: Deleting field relationships

TRAIN

After verifying and modifying the candidate virtual view, you can approve or reject the associated group of candidates. Once you approve/reject the virtual view, the SAI will train the system, and the corresponding group of candidates (bubble) will disappear from the main window.

At least one mapping must be configured as a constraint in the **Candidate Virtual View** tab to accept/reject the virtual view. If the tab lacks a constraint, you will see the corresponding notification:



Approve The Virtual View

To accept the candidate virtual view as verified trusted data, click the **Valid VV** button. All mappings must be configured. If there are any empty fields in the **Candidate Virtual View** tab's mappings, they will be

highlighted with a red border and moved to the top of the list.

Once you click **Valid VV**, you will be redirected to the training page, and the virtual view will appear on the **RDAs & Virtual Views** page as valid.

After the subsequent data source analysis, the system will map the associated personal information instances to existing RDA records and the master catalog of valid personal data.

If a personal information instance is mapped to an existing RDA, it will be a verified data standard for newly discovered personal information instances.

If a personal information instance is mapped to the master catalog, it will be available as data subjects in the appropriate IGDC modules like Data Asset Manager and Personal Information Search.

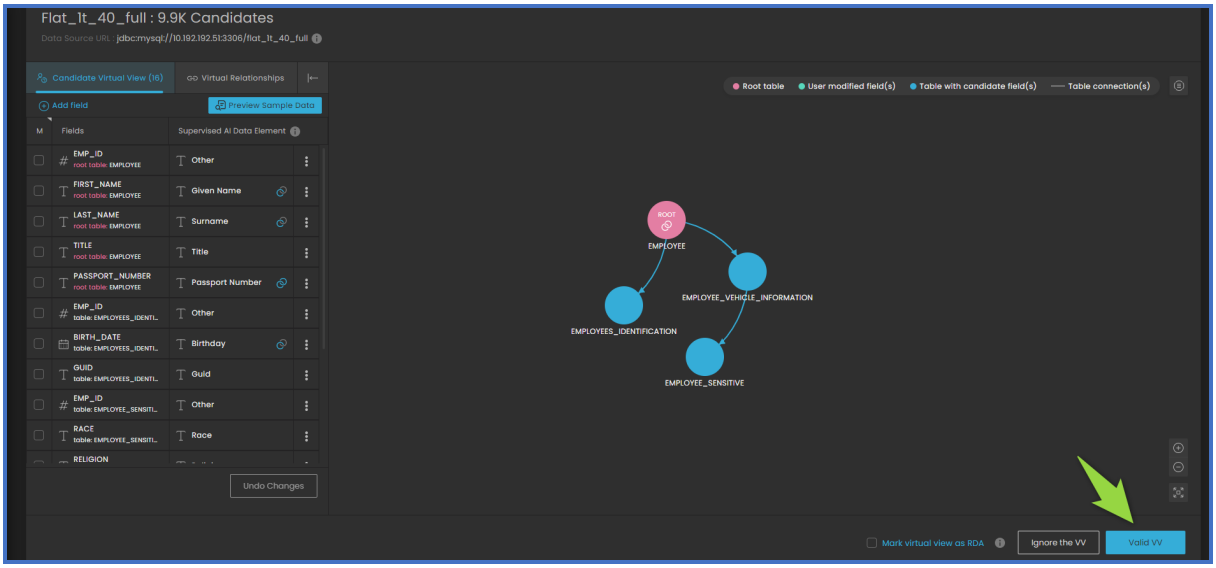


Figure 26: Accepting a candidate virtual view

Create An RDA

To approve the candidate virtual view as a root data asset, enable the **Mark Virtual View as RDA** checkbox (**1**) and enter the RDA name (**2**) and description (**3**) in the popup window that opens.

The name and description will be shown when configuring data assets and on the **RDAs & Virtual Views** page. Click **Save** (**4**) to enable the checkbox or **Cancel** to exit, leaving the checkbox disabled.

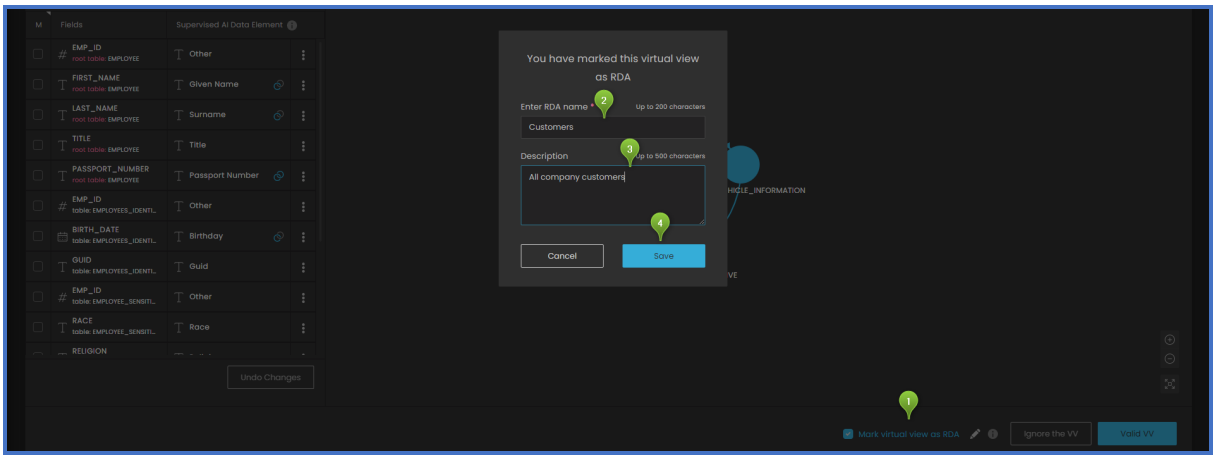


Figure 27: Marking a candidate virtual view as RDA

Once you click **Save**, the **Mark Virtual View as RDA** checkbox (**1**) becomes enabled. Click the **Valid VV** button (**3**) to accept the virtual view as RDA.

> You can return to editing the RDA name and description by clicking the (**Edit**) icon (**2**) next to the **Mark Virtual View as RDA** checkbox.

Once you click **Valid VV**, you will be redirected to the training page, and the virtual view will appear on the **RDAs & Virtual Views** page as an RDA. After the subsequent data source analysis, the system will map the associated personal information instances to the master catalog.

The system will verify the discovered candidates against the new RDA instances.

The created RDA will also be available in the **Data Asset Manager** for custom data asset configuration (**CM UI > Inventory > Data Asset Management > Create/Edit Data Asset**).

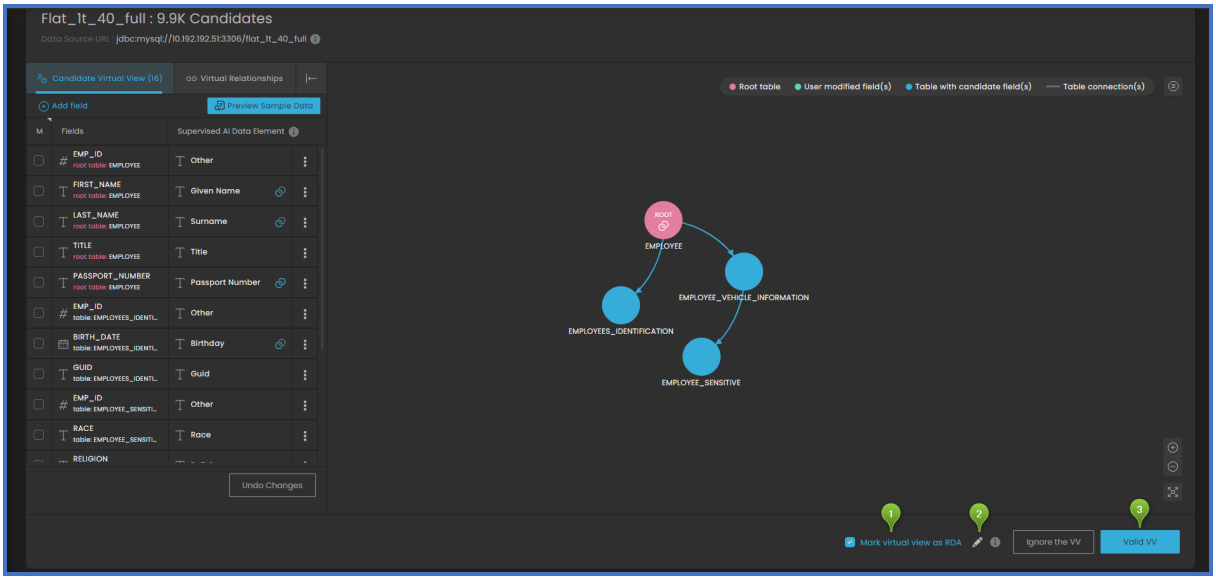


Figure 28: Accepting a candidate virtual view as RDA

Reject The Virtual View

If you identified the candidate virtual view as a false-positive discovery, reject its further usage by IGDC.

This action prevents retrieval of the same false positives, as the IGDC DB analyzer will ignore the rejected virtual view in subsequent analysis cycles.

However, the same personal information instances can be discovered in another data source or if a new field is added to the schema forming a new virtual view.

To reject the candidate virtual view, click the **Ignore the VV** button (**1**). In the confirmation popup, select the reason for ignoring the VV from the dropdown (**2**) or enter your reason (**3**). Then click **Ignore the VV** (**4**) to complete rejection of the VV or **Cancel** to exit, leaving the checkbox disabled.

Once you click **Ignore the VV**, you will be redirected to the training page, and the virtual view will appear on the **RDAs & Virtual Views** page as ignored.

After the subsequent data source analysis, IGDC will skip candidates from this virtual view.

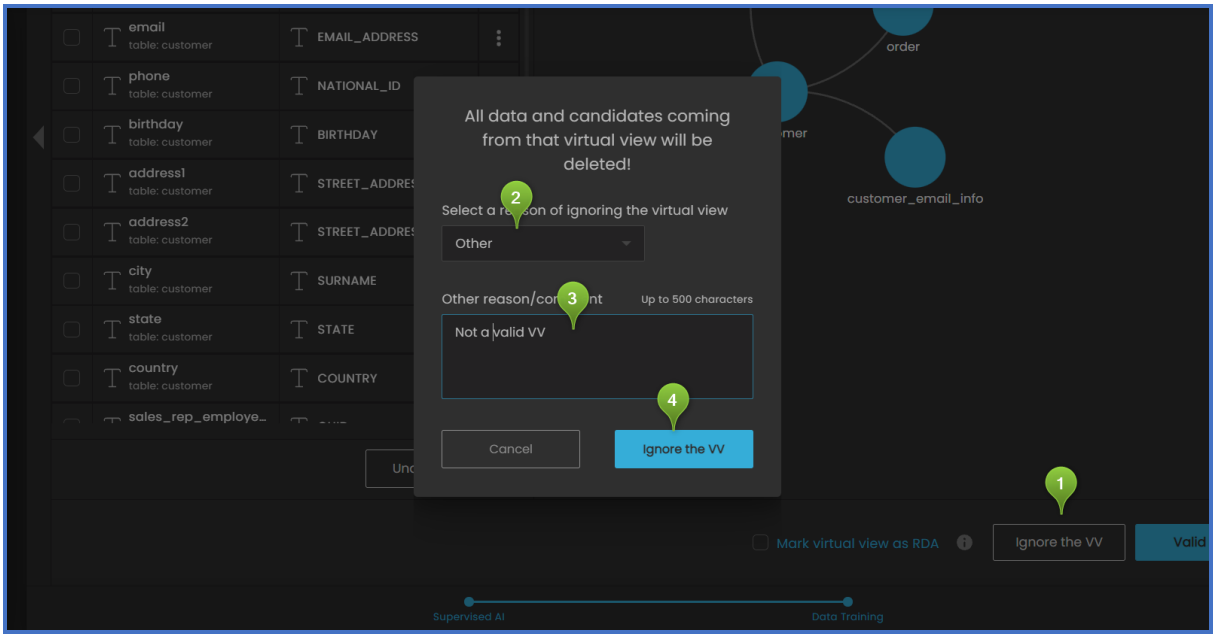


Figure 29: Rejecting a candidate virtual view

Training Page

Once you approve or reject the virtual view, the system switches to training mode. The system will train IGDC analyzers to identify information according to the new structure.

The training process is sequential, so the system begins processing a changed virtual view only after completing training of the previous group of candidates (bubble). However, you can return to the landing page and train other candidate virtual views.

After accepting or rejecting the virtual view, changes will be applied after the next scheduled analysis of the relevant data source.

The training page provides information on the virtual view data source, which can be used to find this data source in the Data Source Catalog, and monitor or initiate its analysis (**CM UI > Inventory > Data Source Catalog**).

Table 30: Fields in the Add new relationship popup window

PARAMETER	DESCRIPTION
Data source name	Name of the data source with the subject virtual view
Data source URL	URL for connection to the data source. The URL format depends on the data source type

PARAMETER	DESCRIPTION
	and vendor. You can copy the URL to use it as a search parameter in the Data Source Catalog.
Vendor	Company that offers the data source solution.
Data source type	Type of the Virtual View/RDA - SQL database or data lake.
Appliance name	Name of the IGDC analytic engine that discovered the virtual view being trained.
Schema name	Name of the schema associated with the virtual view being trained.

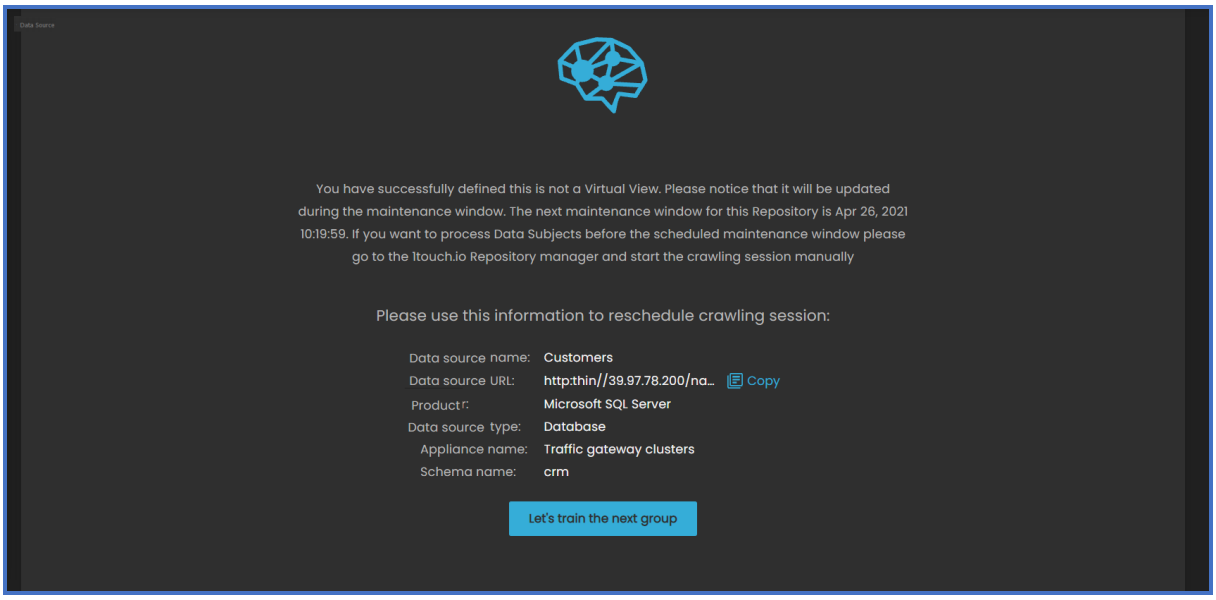


Figure 31: Training mode

Upon completion of data training:

- Groups of candidates (bubbles) are available for training on the landing page.
- Group of candidates (bubble) related to the trained group of data subjects disappears from the landing page.
- Virtual View/RDA entry appears in the RDA/Virtual View list page (**Supervised AI > RDA & Virtual Views**).

Training results are available after a full analysis of the related virtual view. The system analyzes data sources according to the schedule configured in the Data Source Catalog.

To speed up the training process, you can copy the data source URL on the training page, find it in the Data Source Catalog (**CM UI > Inventory > Data Source Catalog**) and manually initiate the data source analysis - or shift the analysis window.

CHAPTER 3: ROOT DATA ASSET/VIRTUAL VIEW CREATION

The Supervised AI enables users to create root data assets or virtual views from scratch without any automatically discovered candidate groups of personal information.

You can create a VV/RDA in three steps:

1. Select the data source associated with the virtual view.

2. Select the schema associated with the virtual view.

3. Modify and save the virtual view/RDA.

To start creating a VV/RDA, go to **Supervised AI > RDAs & Virtual Views.** On the **RDA/Virtual View list** page, click the **Create RDA/Virtual View** button.

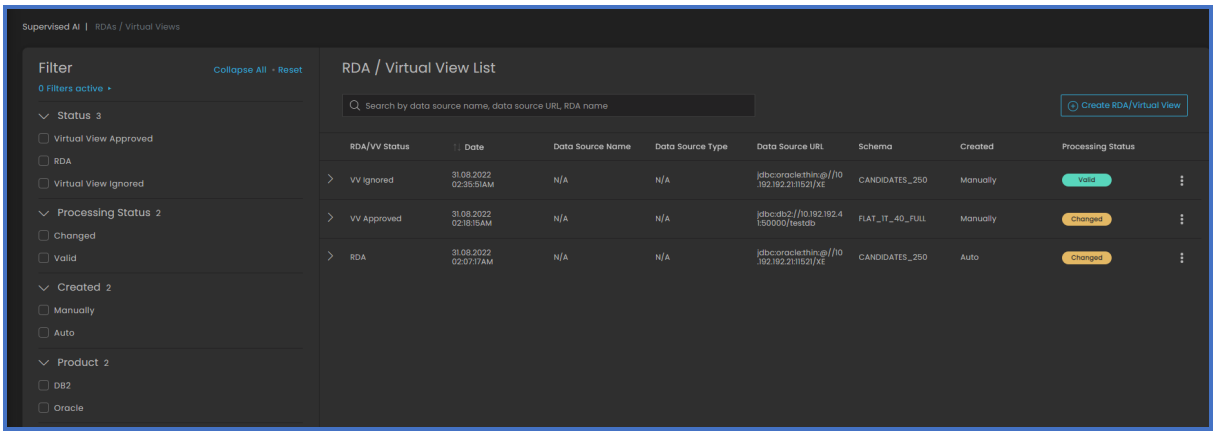


Figure 1: Initiating creation of a root data asset or virtual view from scratch

STEP 1: SELECT DATA SOURCE

The **Select Data Source** page shows a list of data sources analyzed by IGDC and supported by the Supervised AI platform. You can search for data sources by data source name or URL. Click **Select** for the desired data source to continue the virtual view/RDA creation.

The counters in the Supervised AI may be different from the counters in Data Source Catalog due to data processing lags. The time needed to load depends on the amount of data.

Table 2: Data source parameters on the Select Data Source page

PARAMETER	DESCRIPTION
Collapsed view	
Data Source Name	Name of the data source (data storage).
Data Source URL	URL for connection to the data source. The URL format depends on the data source type and product.
Data Subjects	Total number of personal information instances retrieved from this specific data source and validated against at least one RDA. Such personal information instances are considered trusted and are not included in data training. This number is provided to give additional visibility to the data inside this data source
Candidates	Total number of instances of personally identifiable information retrieved from the data source but not confirmed against the root data asset (RDA).
Product	Company that offers the data source solution.
Expanded view	

PARAMETER	DESCRIPTION
Data Source Type	Type of the data source - SQL database or data lake.
Last Analysis Time	Date and time of the last data source analysis by IGDC.

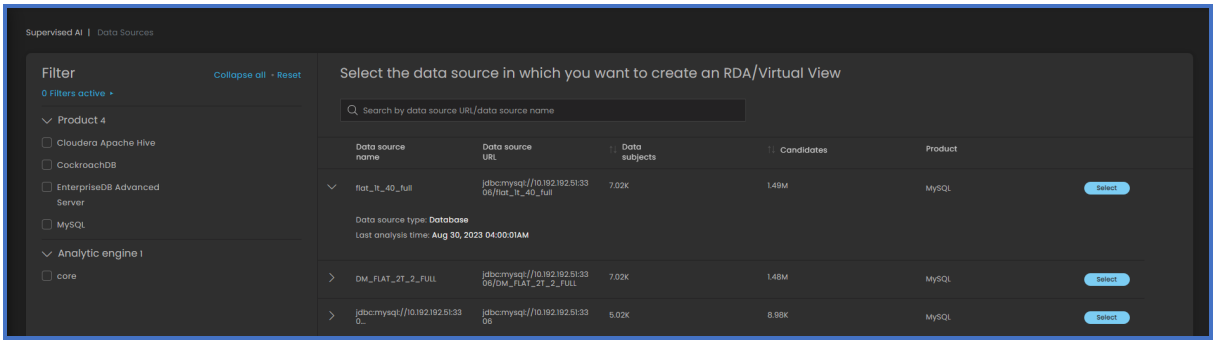


Figure 3: Select Data Source page

To find the desired data source, begin typing the data source name or URL in the **Search** box and select the desired item from the dropdown list. The main window will show the associated entry or the **No results** screen if none of the entries matches the selected parameter.

You can also sort the displayed data sources using the sidebar filters: product and appliance name. The filter options change dynamically relative to the data sources available in the main window.

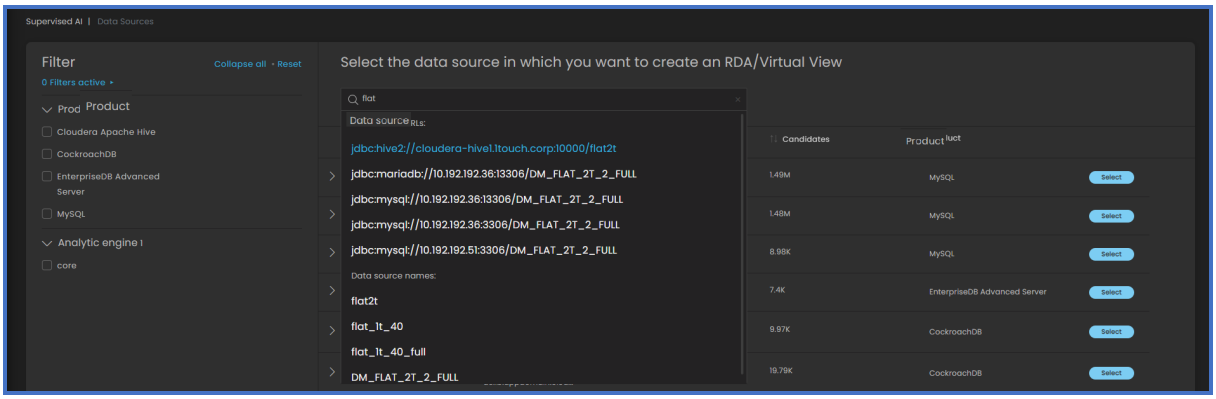


Figure 4: Search in the Select Data Source page

STEP 2: SELECT SCHEMA

The **Select Schema** page shows a list of schemas (**1**) available in the selected data source (SQL database or data lake). Hover over the (**Info**) icon (**2**) to see the data source details like product, type, appliance name, date, and time of the last data source analysis.

Click the **Select** button (**3**) for the schema with the desired virtual view and then **Next** (**4**) to continue the virtual view/RDA creation.

The Select button is disabled if no schema is selected.

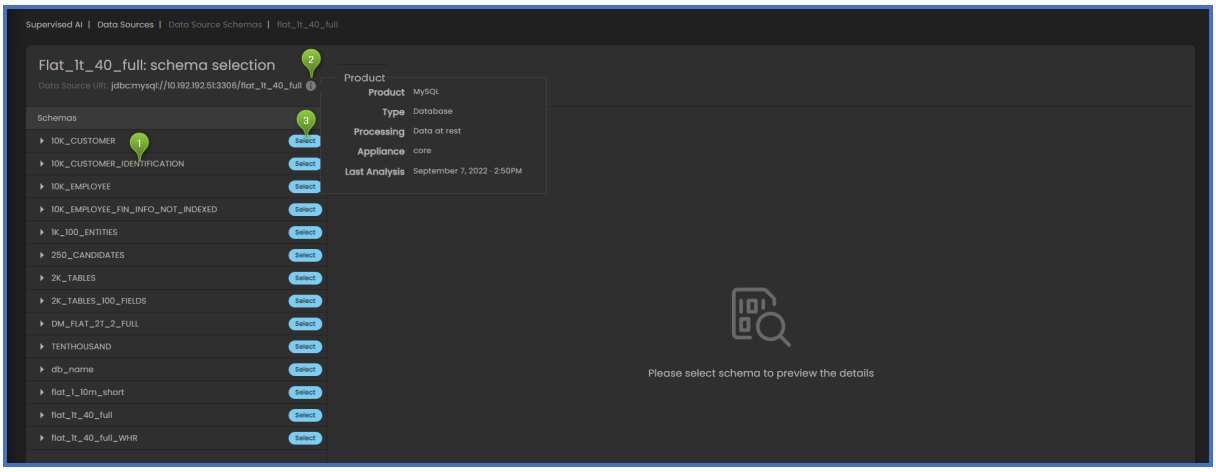


Figure 5: Select Schema page

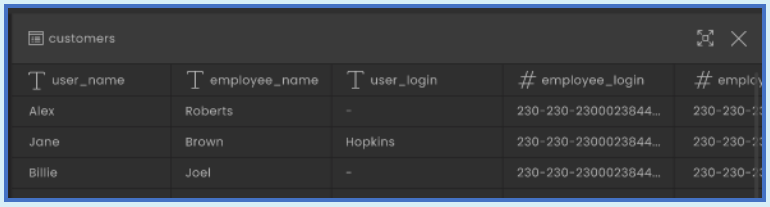
To review the schema details, click the **>** (**Expand**) icon or the **Select** button for the desired entry.

Figure 6: Table 14: Schema details in the Select Schema page

Table 7: Schema details in the Select Schema page

PARAMETER	DESCRIPTION
Visualization (**1**)	Map of the table links in the schema in the right part of the screen.
Tables (**2**)	List of all schema tables.
Columns (**3**)	List of all table columns available by clicking the **>** (**Expand**) icon for the desired table. The field icon represents the data type.  : Integer data (numbers).  : String data (text).  : Timestamps (date and time).

You can review sample data extracted from the table by clicking the  icon (**4**). To extract high-quality and useful sample data, the system skips the rows with many empty fields.



Sample data of a schema

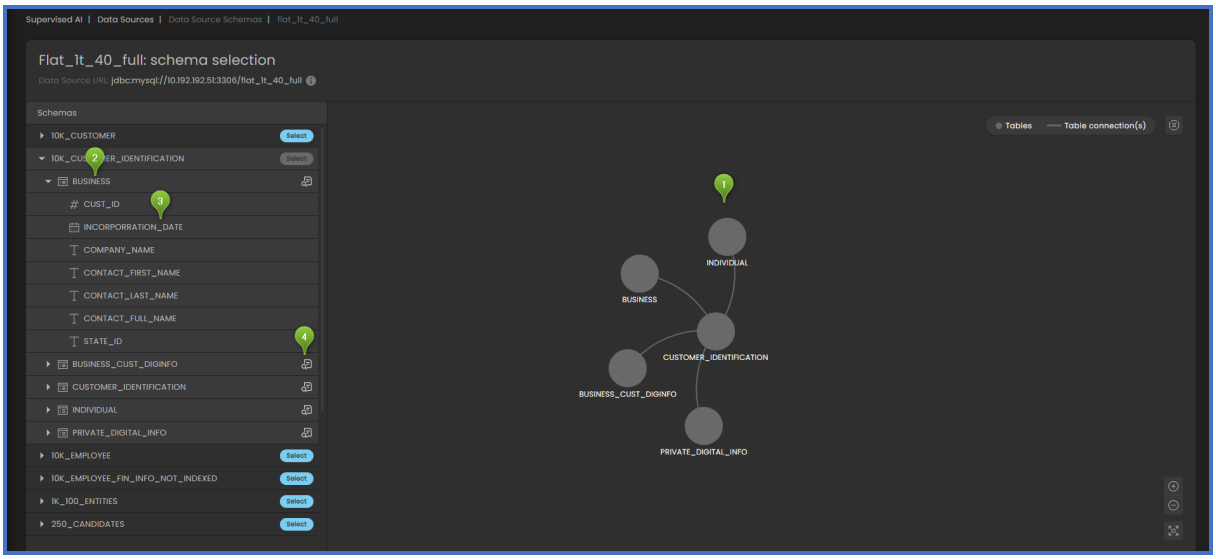


Figure 8: Schema details on the Select Schema page

STEP 3: MODIFY AND SAVE

After selecting the data source and schema, click **Next** to go to the **Edit/Create Virtual View/RDA** page, similar to the Data Training page. The virtual view configurations are made in the **Virtual Relationships** and **Candidate Virtual View** tabs and visualized in the virtual view map.

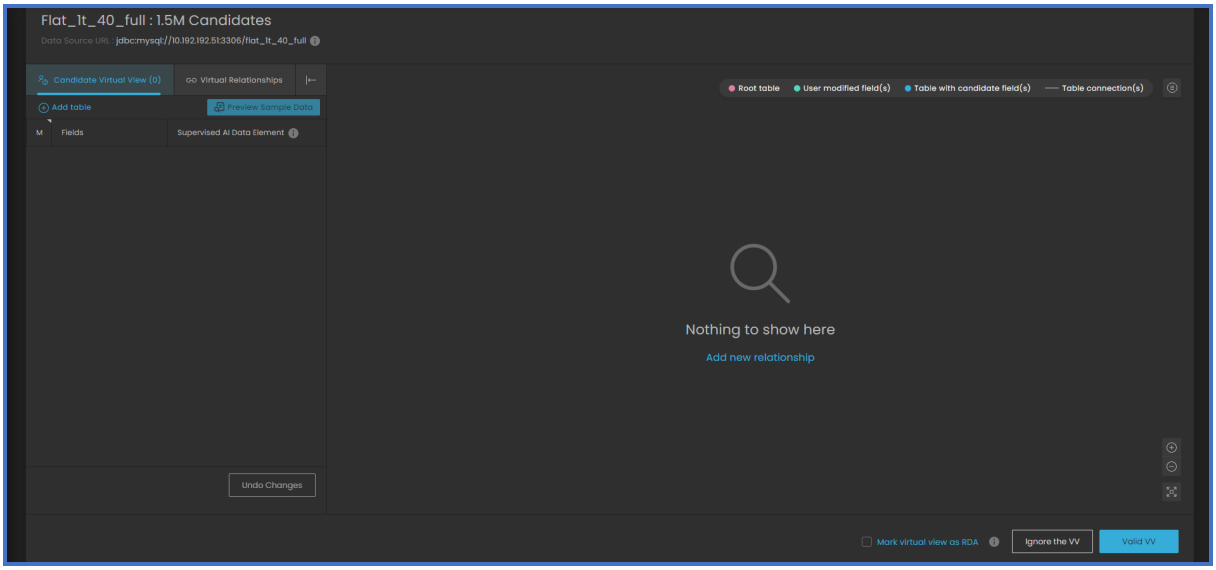


Figure 9: Edit/Create Virtual View/RDA page

The **Edit/Create Virtual View/RDA** page enables users to create a virtual view from the selected schema and save it as follows:

1. Configure the virtual view.
2. Configure field-to-data element mappings.
3. Train IGDC against the configured virtual view.

Configure The Virtual View

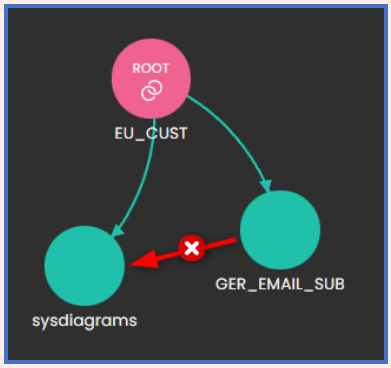
Configuring the virtual relationships is not required for RDA/VV based on a single table: setting a root table in the **Candidate Virtual View** tab is sufficient.

In the **Virtual Relationships** tab, configure the desired table relationships and field relationships (the procedure is identical to the Data Training page).

The grey circles related to the affected tables will turn green in the visualization map, and the direction of the relationship will be visualized by arrows (left field -> right field).

When adding the first table relationship, the system will offer all available tables in the schema in both fields. When adding the next relationships, the left field will show only tables from the virtual view being created.

The system does not accept "looped" virtual views. For example, in the virtual view below, a user cannot create a relationship between the sysdiagrams and GER_EMAIL_SUB tables.



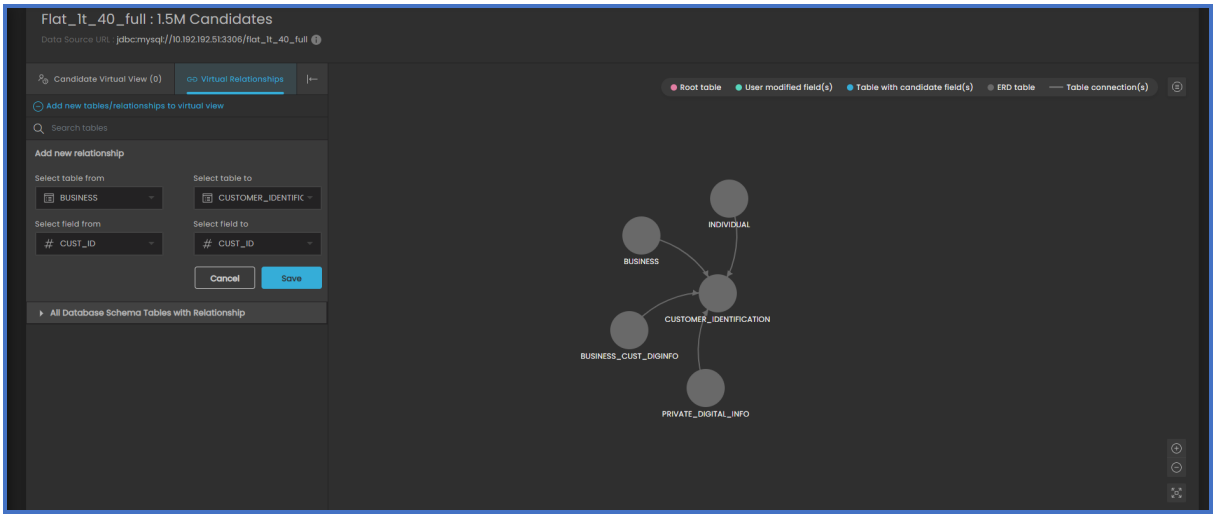


Figure 10: VV/RDA creation from scratch in Virtual Relationships tab

2. In the **Candidate Virtual View** tab, find the root table and enable the **Root** checkbox. The related circle on the virtual view map will turn pink.

Configure The Mappings

1. In the **Candidate Virtual View** tab, configure the field-to-data element mappings - a relation between a specific field (column) in the table and a specific type (data element) like given name, tax ID, gender, and others supported by the Supervised AI application.

Ensure that the root table is set and at least one constraint is configured.

2. When mapping fields from the root table, enable the **Root** checkbox. The related circle on the virtual view map will turn pink.

3. Ensure at least one constraint is configured.

Figure 11: Mapping the new VV/RDA data elements

Train

Save the created entry as a virtual view/RDA or reject the virtual view. The system will redirect you to the training page.

CHAPTER 4: VIRTUAL VIEW & RDA MANAGEMENT HISTORY

The **RDA/Virtual View list** page shows all the virtual views and root data assets created, modified, or rejected in the Supervised AI application and their metadata (**Supervised AI > RDAs & Virtual Views**) and enables users to create root data assets from scratch.

REVIEWING DATA TRAINING HISTORY

The **RDA/Virtual View list** page shows a list of the following Supervised AI entries:

- Valid groups of candidates.
- Approved groups of candidates saved as RDA.
- Rejected groups of candidates.
- Root data assets created from scratch in the Supervised AI application.

The candidate virtual views (groups of candidates) are out of scope and are only available on the landing page as "bubbles".

Table 1: Virtual View/RDA information in a collapsed state

ELEMENT	DEFINITION
Status	Result of the last virtual view training. **VV Valid**: Candidate Virtual View has been approved as a trusted group of personal information instances. **RDA**: Virtual View has been approved and saved as a root data asset, or a root data asset has been created from scratch. **VV Ignored**: Candidate Virtual View has been rejected for usage in the system.
Date	Date and time of the Virtual View/RDA creation.
Data Source Name	Name of the data source with the subject virtual view
Data Source Type	Type of the Virtual View/RDA - SQL database or data lake.
Data Source URL	URL for connection to the data source. The URL format depends on the data source type and vendor.
Schema	The schema associated with the Virtual View/RDA.
Created	Type of user involvement in the Virtual View/RDA creation and modification before approval/rejection. **Automatic**: Candidate Virtual View or RDA has been automatically identified by the system, and a user has not modified any mappings or table relationships. **Manual**: User has modified the Virtual View mappings or table relationships or has created the RDA from scratch.
Processing status	Virtual view status in the Supervised AI platform. **Valid**: Virtual view is considered valid for processing according to the Data element configuration. **Invalid**: Virtual view is considered invalid for processing due to changed Data element configuration. For example, the passport number data element has been deleted from the list of applied data elements. If a passport number is the only constraint in the virtual view, the system will stop its processing. However, if the virtual view includes at least one other constraint, the system will continue its processing. **Changed**: Data element configuration has been changed, and the changes have affected the virtual view. For example, the full name data element has been deleted from the list of applied data elements. If the given name is used in a virtual view, its status will become Changed.

ELEMENT	DEFINITION
Options	Click the ⋮ (**Options**) icon to see the available actions. **Edit** button: Redirects to the virtual view editing page with a possibility to modify mappings and table relationships and then approve/reject the virtual view. **Ignore** button: Configures an RDA as ignored, meaning that the system will stop using it as a set of trusted data for comparison with newly retrieved data.

By default, all entries in the list are collapsed. Click the **>** (**Expand**) icon to see details for the desired entries. The auxiliary information varies depending on the entry type - virtual view or RDA.

Table 2: Virtual View/RDA information in an expanded state

ELEMENT	DESCRIPTION	VV	RDA
Virtual View ID	Unique identifier of the virtual view in IGDC platform.	✓	✗
Ignored reason	Reason for rejecting the virtual view	✓	✗
RDA Name	Root data asset name entered when marking the virtual view as RDA.	✗	✓
RDA Description	Additional information entered when marking the virtual view as RDA.	✗	✓
Data elements	List of IGDC data elements associated with the virtual view/RDA.	✓	✓
Product	Company that offers the data source solution.	✓	✓
Processing	Processed data - data at rest.	✓	✓
Appliance	Name of the appliance, which discovered the virtual view/RDA data source.	✓	✓
Last analysis	Date and time of last data source analysis.	✓	✓

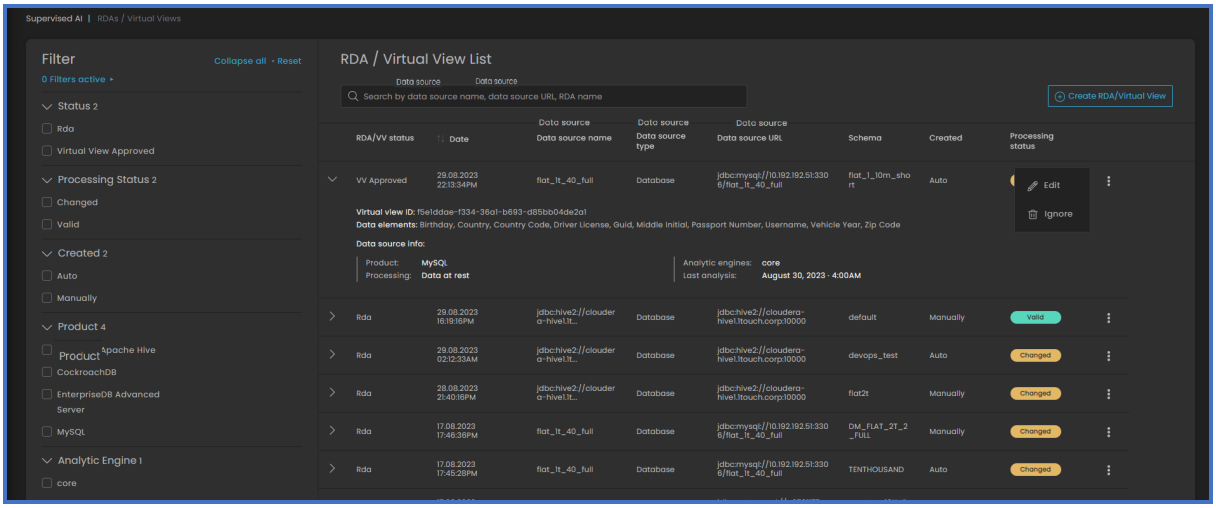


Figure 3: Virtual View/RDA information in a collapsed state in the RDA/Virtual View List page

SORTING VIRTUAL VIEWS & RDAS

You can select the virtual view & RDA entries to be shown on the page via the Search box or by using the filters on the sidebar.

SEARCH

To find a specific virtual view or RDA, use the **Search** box. Begin typing the data source name, data source URL, or RDA name, and select the desired option from the dropdown list.

The main window will show the associated entries or the **No results** screen if none of the entries matches the selected parameter.

Table 4: Search parameters in the RDA/Virtual View list page

PARAMETER	DESCRIPTION
Data source name	Name of the data source.
Hostname	URL for connection to the data source.
RDA name	Root data asset name entered when marking the virtual view as RDA.

SIDEBAR FILTERS

The **Sidebar** shows the options for filtering the virtual view & RDA entries by status, creation type, vendor, and appliance name.

The options shown in the filters depend on the virtual view/RDA properties in the page and may change along with the data training process.

To disable all selected filters, click the **Reset** button. The sidebar filters' display modes are identical to the filters on the landing page.

Table 5: Sidebar filters in the RDA/Virtual View list page

NAME	DESCRIPTION
Status	Result of the last virtual view training. **Virtual View Approved**: Candidate Virtual View has been approved as a trusted group of personal information instances. **RDA**: Virtual View has been approved and saved as a root data asset, or a root data asset has been created from scratch. **Virtual View Ignored**: Candidate Virtual View has been rejected for usage in the system.
Created	Type of user involvement in the Virtual View/RDA creation and modification before approval/rejection. **Automatic**: Candidate Virtual View or RDA has been automatically identified by the system, a user has not modified mappings or table relationships. **Manual**: User has modified the Virtual View mappings or table relationships or has created the RDA from scratch.
Product	Name of the data source product.
Analytic engine	Name of the IGDC analytic engine (appliance) connected to the data source.

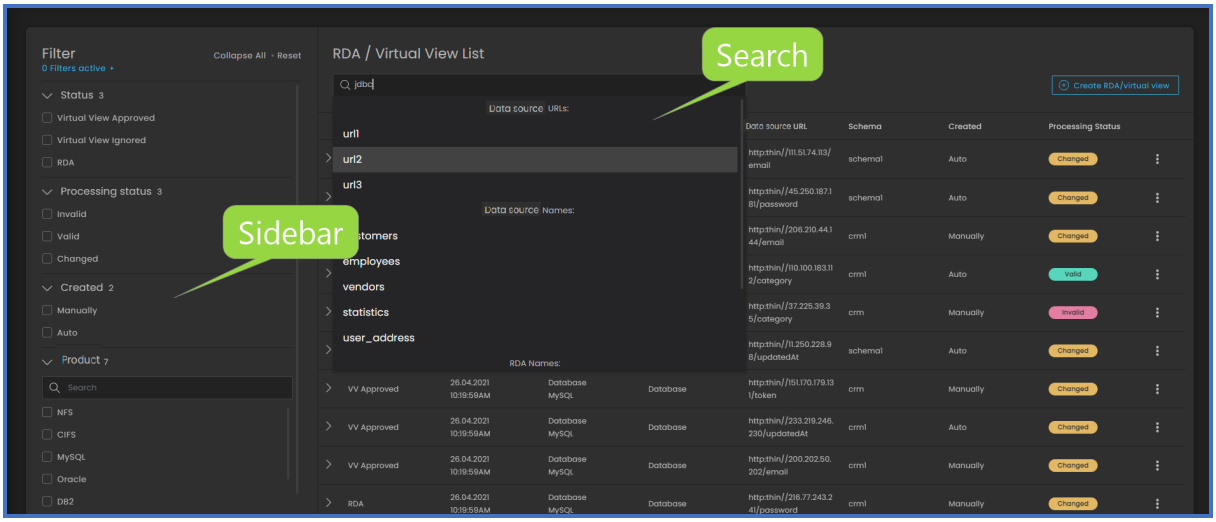


Figure 6: Sorting of virtual views & RDAs

MANAGING VIRTUAL VIEWS & RDAS

The **RDA/Virtual View list** page enables you to create an RDA/virtual view from scratch, and edit or ignore the "trained" RDAs and virtual views.

1. To create an RDA or virtual view from scratch, click the **Create RDA/Virtual View** button (**1**). Then follow the procedure described in Chapter 3.

2. To **edit an RDA or virtual view**, click the ⋮ (**Options**) icon (**2**) for the desired entry and select **Edit** (**3**). You will be redirected to the **Data Training** page to make the desired modifications and select how to train the modified virtual view. After editing, the entry will move to the top of the list as the latest event of the data training history.

3. To **stop applying the virtual view or RDA** by IGDC, click the ⋮ (**Options**) icon (**2**) for the desired entry and select **Ignore** (**4**).

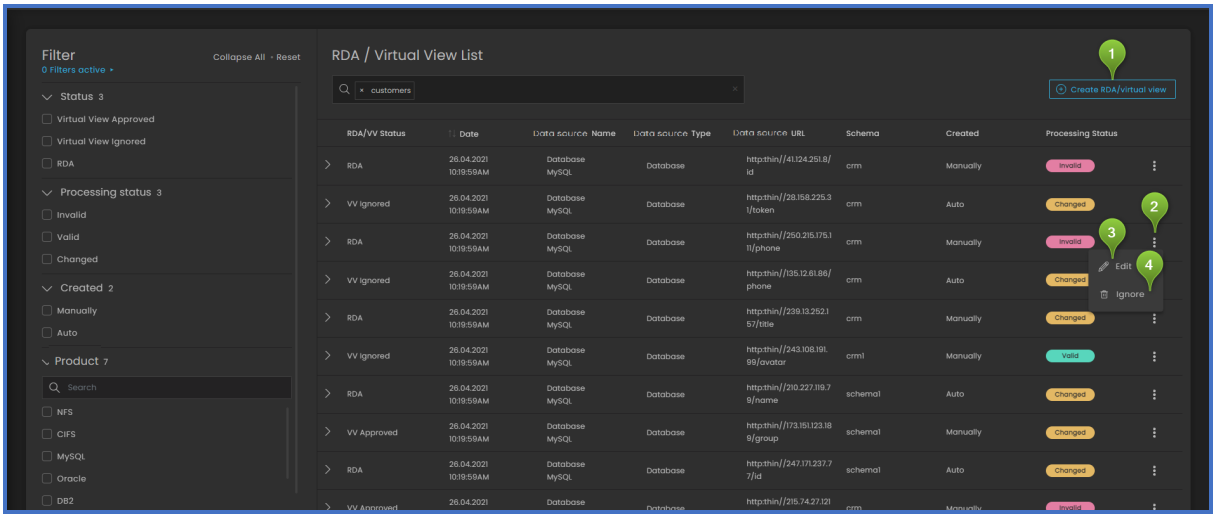


Figure 7: Managing virtual views & RDAs

CHAPTER 5: REVIEW CANDIDATES

The **Candidates** page allows you to view, filter, and export candidates data - data pertaining to instances of personal data retrieved from the data source but not confirmed against the root data asset (RDA). The page shows candidates discovered in all data sources supported by IGDC, including file shares, databases, cloud storages, etc.

VIEWING CANDIDATE DATA

To view all Candidates' Identified by the IGDC Discovery module, go to **Supervised AI > Candidates** to access the **Candidates** page:

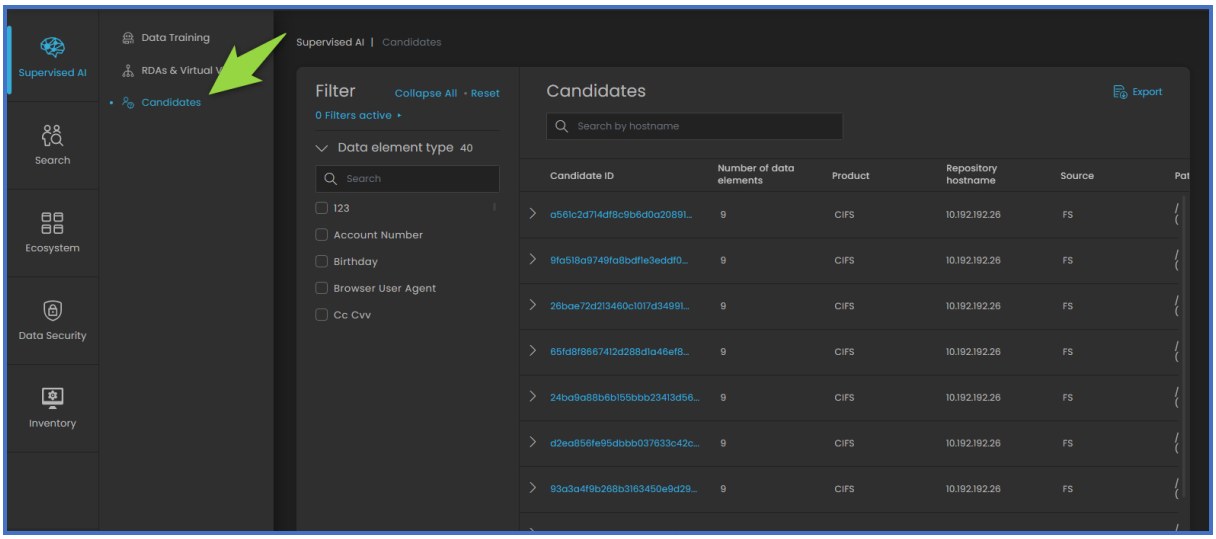


Figure 1: Selecting Candidates page

The information displayed in the **Candidates** page includes the following:

Table 2: Information displayed in the Candidates page

PARAMETER	DESCRIPTION
Candidate ID	Candidate unique identifier generated by the system.
Number of data elements	Number of data elements associated with the candidate (e.g. If a candidate information includes only Name and Address - they will have 2 data elements).
Product	Product name (e.g. CIFS, Oracle, Google Drive).
Data source hostname	URL of data source in which Candidate was discovered.
Source	Data source file/record type.
Path	Path to file.
Data source type	The type of data source in which the candidate was discovered (SQL database, data lake, central storage, cloud storage, etc.). Shown only in the candidate expanded view. Click  to expand the candidate record.
Data elements	List of data elements associated with the candidate. Shown only in the candidate expanded view. Click  to expand the candidate record.

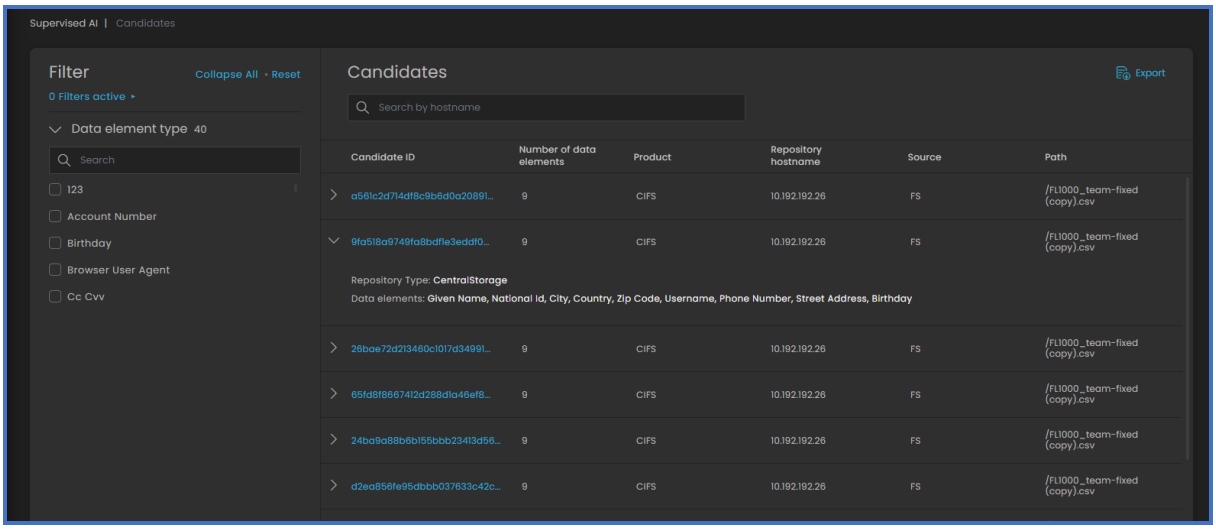


Figure 3: Candidates page

To view all the data discovered for a specific Candidate, click on the **Candidate ID** link. An auxiliary popup window will open with the candidate profile, showing all the discovered data elements, their values and confidence level (discovered).

Note: If the data element is encrypted, the value will be masked (******) or hashed, with the ability to unlock and view the value by clicking the (**Locked**) icon. This user action will be logged.

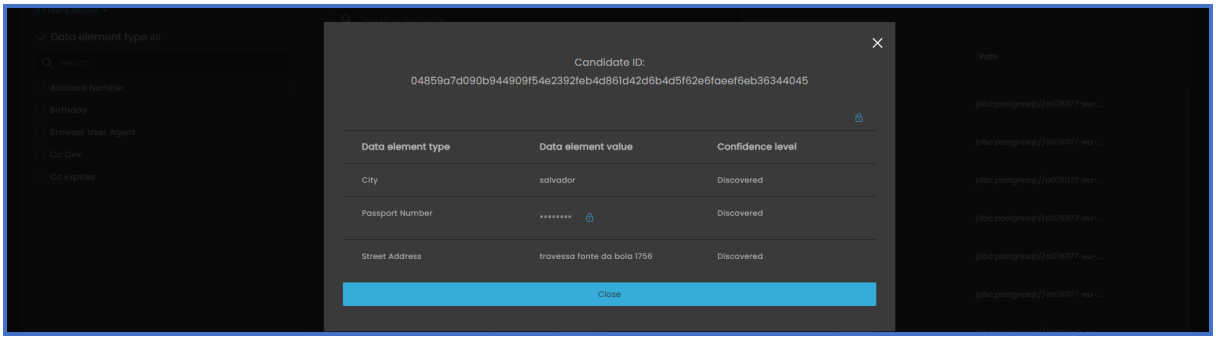


Figure 4: Candidate profile

FILTERING CANDIDATES

The default view of the **Candidates** page contains all the candidates discovered by IGDC. To view the specific candidate/s, you may use the page data filters. After selecting the relevant filter data In the filter fields, the page will automatically update to reflect the filtered data.

Table 5: Candidates page filters

FILTER	DESCRIPTION
Filter by **hostname** (**1**)	View only candidates located in data sources associated with the entered hostname.
Filter by data element(s) (**2**)	Check one or more data elements from the filter pane, and view only the candidates with data that includes those data elements.

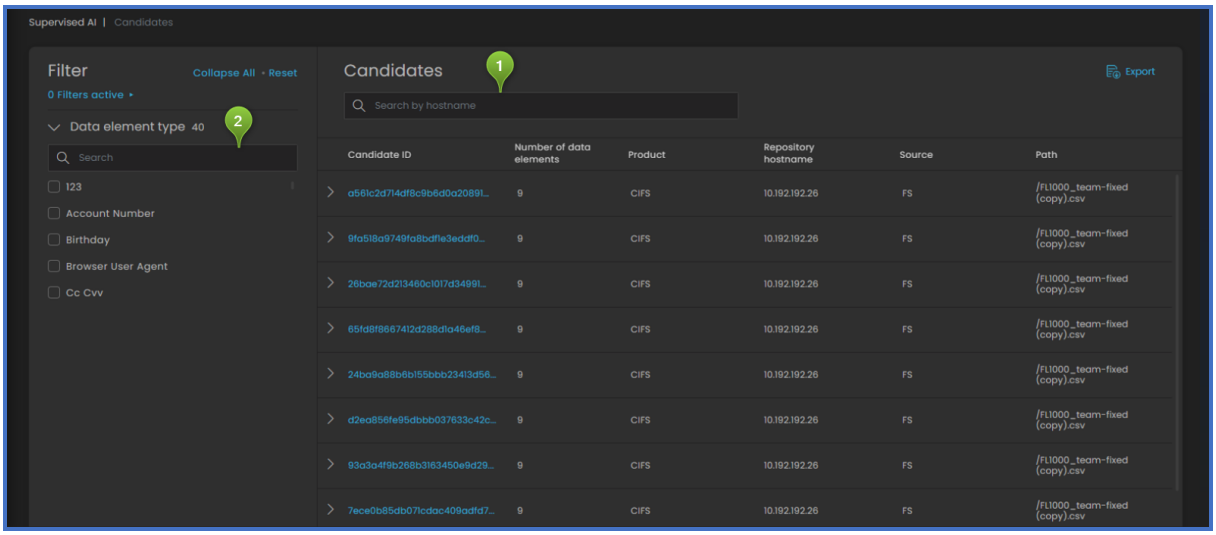


Figure 6: Candidates page filters

CANDIDATES REPORT

To generate a candidates report, access the **Candidates** page as detailed above, and then either click the **Export** button to export all candidate data - or filter the data as detailed above and then click **Export** to generate a report pertaining to the selected candidates.

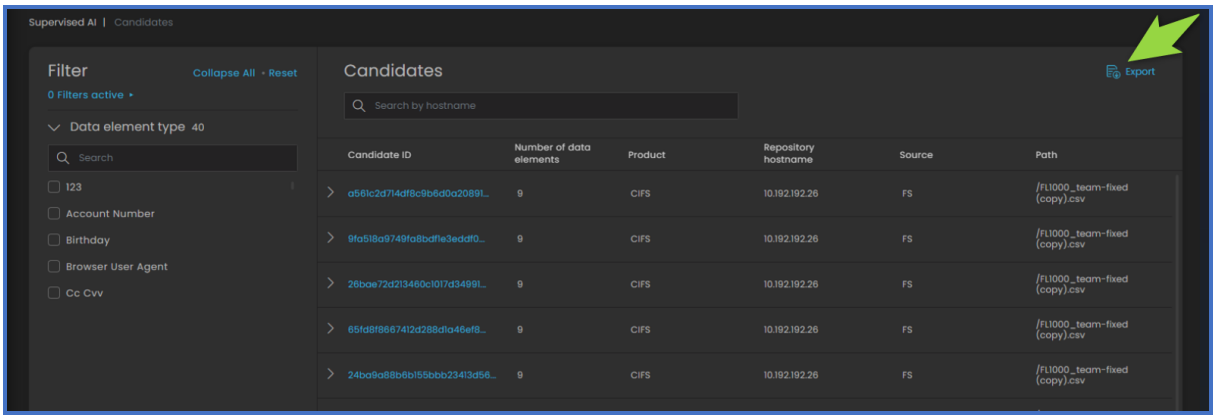


Figure 7: Initiating the candidates report

In the **Export** as popup window that will open, select the exporting parameters.

Table 8: Export popup window elements

PARAMETER	DESCRIPTION
Format options	Format of the exported file. Supported formats: XLSX; CSV.
File name field	Name of the exported file. For example, *IGDC candidates report*.
Limit field	The maximum number of copies to be included in the report. Default limit: 10,000 copies. Maximum limit: 100,000 copies.

If the discovered candidates exceed the value in the **Limit** field, the system will randomly select the copies to be included in the report.

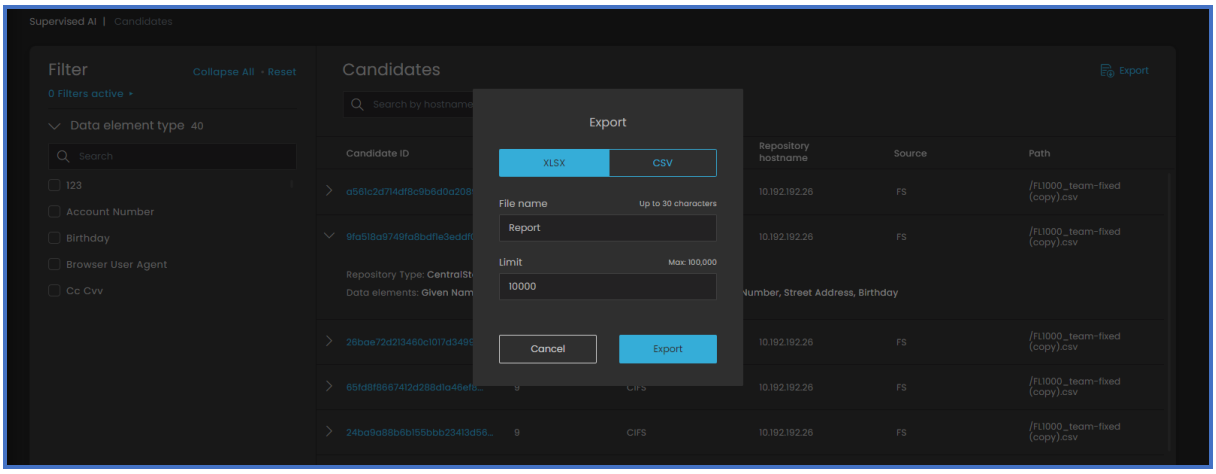
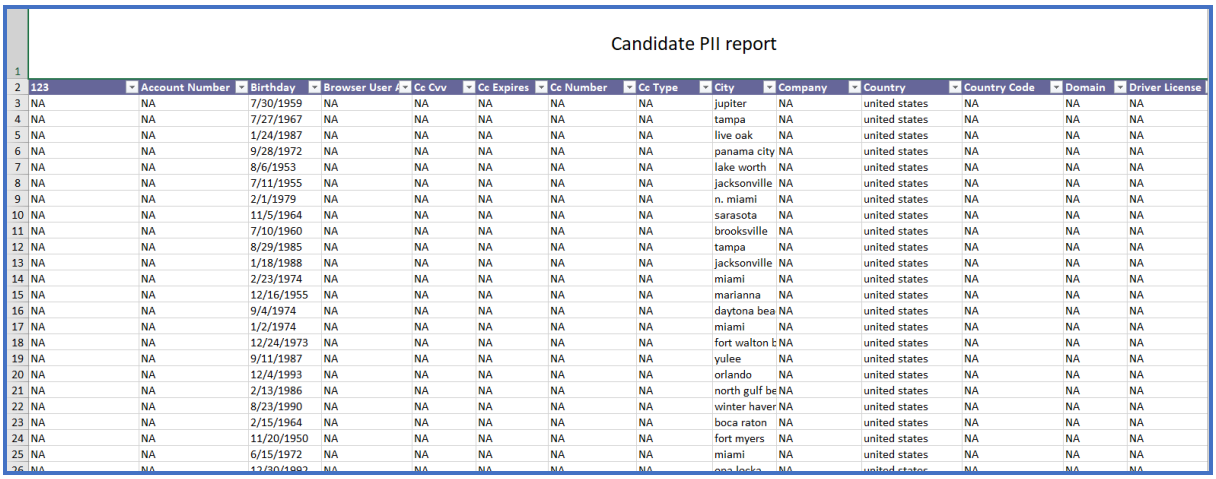


Figure 9: Export popup for candidates report

Click the **Export** button to download the file to your computer. The report contains a list or data element values in columns per unique candidate in rows. If the data element is encrypted, the value will be masked (******) or hashed.



Candidate PII report

123	Account Number	Birthday	Browser User A	Cc Cvv	Cc Expires	Cc Number	Cc Type	City	Company	Country	Country Code	Domain	Driver License
NA	NA	7/30/1959	NA	NA	NA	NA	NA	jupiter	NA	united states	NA	NA	NA
NA	NA	7/27/1967	NA	NA	NA	NA	NA	tampa	NA	united states	NA	NA	NA
NA	NA	1/24/1987	NA	NA	NA	NA	NA	live oak	NA	united states	NA	NA	NA
NA	NA	9/28/1972	NA	NA	NA	NA	NA	panama city	NA	united states	NA	NA	NA
NA	NA	8/6/1953	NA	NA	NA	NA	NA	lake worth	NA	united states	NA	NA	NA
NA	NA	7/11/1955	NA	NA	NA	NA	NA	jacksonville	NA	united states	NA	NA	NA
NA	NA	2/1/1979	NA	NA	NA	NA	NA	n. miami	NA	united states	NA	NA	NA
NA	NA	11/5/1964	NA	NA	NA	NA	NA	sarasota	NA	united states	NA	NA	NA
NA	NA	7/10/1960	NA	NA	NA	NA	NA	brooksville	NA	united states	NA	NA	NA
NA	NA	8/29/1985	NA	NA	NA	NA	NA	tampa	NA	united states	NA	NA	NA
NA	NA	1/18/1988	NA	NA	NA	NA	NA	jacksonville	NA	united states	NA	NA	NA
NA	NA	2/23/1974	NA	NA	NA	NA	NA	miami	NA	united states	NA	NA	NA
NA	NA	12/16/1955	NA	NA	NA	NA	NA	marianna	NA	united states	NA	NA	NA
NA	NA	9/4/1974	NA	NA	NA	NA	NA	daytona bea	NA	united states	NA	NA	NA
NA	NA	1/2/1974	NA	NA	NA	NA	NA	miami	NA	united states	NA	NA	NA
NA	NA	12/24/1973	NA	NA	NA	NA	NA	fort walton b	NA	united states	NA	NA	NA
NA	NA	9/11/1987	NA	NA	NA	NA	NA	yulee	NA	united states	NA	NA	NA
NA	NA	12/4/1993	NA	NA	NA	NA	NA	orlando	NA	united states	NA	NA	NA
NA	NA	2/13/1986	NA	NA	NA	NA	NA	north gulf be	NA	united states	NA	NA	NA
NA	NA	8/23/1990	NA	NA	NA	NA	NA	winter haver	NA	united states	NA	NA	NA
NA	NA	2/15/1964	NA	NA	NA	NA	NA	boca raton	NA	united states	NA	NA	NA
NA	NA	11/20/1950	NA	NA	NA	NA	NA	fort myers	NA	united states	NA	NA	NA
NA	NA	6/15/1972	NA	NA	NA	NA	NA	miami	NA	united states	NA	NA	NA

Figure 10: Example of XLS candidates report

APPENDIX A GLOSSARY OF TERMS AND ACRONYMS OF THE SUPERVISED AI MODULE

Table 1: Glossary of Terms

TERM	DEFINITION
Appliance name	Name of the IGDC analytic appliance or CM appliance.
Bubble	Visualization of a group of candidates combined by common properties - data source and virtual view.
Candidate	Personal data instances that were not confirmed against any of the RDAs and therefore cannot be considered "trusted" (verified).
Candidate Virtual View Tab	Tab for the identified virtual view structure modification and adjustment of the column mapping to the Supervised AI data elements (Supervised AI > Data Training).
Constraint	Single data element or a combination of data elements that specify a unique person in the IBM Guardium inventory of personal data.
Data source	Network element that stores data in a structured or unstructured format. The IGDC uses them as a data source to create an inventory of the personal information.
Data subject	Instance of retrieved personal data confirmed as a unique data subject due to match with RDA.
Database	Structured data storage serving as a source of personal data.
Database name	Database vendor name. For example, Oracle, MySQL, MariaDB.
Group (complex)constraint	Combination of data elements that specify a unique person in IBM Guardium inventory of personal data.
Hostname	Data source URL, which format depends on the data source type and vendor.
Last time analysis	Date and time of the last data sourceanalysis.
Mandatory field	Attribute of a virtual view column specifying it as mandatory for personal information instance retrieval. The rows with a mandatory blank field (potential data subject record) will be ignored during virtual view analysis.
RDA/VV list	List of root data assets and approved virtual views created or modified in the Data Training page.
Relationship Tab	Tab for virtual view creation from scratch, including table links and column-data element mappings (Supervised AI > Data Training).
Root Data Asset (RDA)	Set of structured data used by IBM Guardium as a source for comparison with detected personal information.
Sample Data	Table with data elements of 30 data subjects retrieved by the modified virtual that facilitate verification of the column-data element mappings.
Schema	Logical collection of objects (tables, views, indexes) in a database.
Supervised AI	Wizard-based machine learning tool enabling a "non-data scientist" to build, modify, and train AI models on how to identify personal and sensitive information based on discovered results. It assists users in the analysis of the identified candidates based on database virtual views and training the system to accept or reject the candidates as trusted data, creation of RDAs, and applying the system-detected virtual views as RDAs.
Supervised AI data element	Name used by IGDC software for specific personal data types. For example, CC_ NUMBER for credit card numbers. The Supervised AI data elements are identical to the data elements across the IGDC modules like Personal Information Search, Advanced Search and others. You can change the list of data elements and modify data element recognition in the Data Element Configurator (CM UI > Settings > Data Recognition > Data Element Configuration).
Training mode	Mode triggered by changes in a virtual view when the system trains IGDC analyzers to identify Information according to the new structure. The training process is sequential, meaning that the system begins processing a changed virtual view only after finishing training the previous group of candidates (bubble).
Virtual View (VV)	Set of database tables or schema used as a source of a specific group of candidates

TERM	DEFINITION
	forming a collection of identical data elements designated to the subject candidate group.

APPENDIX B: SUPERVISED AI SUPPORTED PRODUCTS

Table 1: Supported products - Supervised AI

PRODUCT	DATA SOURCE TYPE
Amazon Athena	Database
Amazon Aurora MySQL	Database
Amazon Aurora PostgreSQL	Database
Amazon RDS MySQL	Database
Amazon RDS MariaDB	Database
Amazon RDS PostgreSQL	Database
Amazon RDS Oracle	Database
Amazon RDS SQL Server	Database
Amazon Redshift	Database
Azure MySQL	Database
Azure MariaDB	Database
Azure PostgreSQL	Database
Apache Hive	Database
CockroachDB	Database
EnterpriseDB	Database
Google Cloud MySQL	Database
Google Cloud PostgreSQL	Database
Google Cloud SQL Server	Database
Greenplum	Database
IBM DB2 (type4)	Database
IBM DB2 for z/OS	Database
IBM Informix	Database
MariaDB	Database
MS SQL Server	Database
MySQL	Database
Oracle	Database
Oracle Exadata	Database
PostgreSQL	Database
SAP HANA	Database
Databricks	Data lake
Snowflake	Data lake