

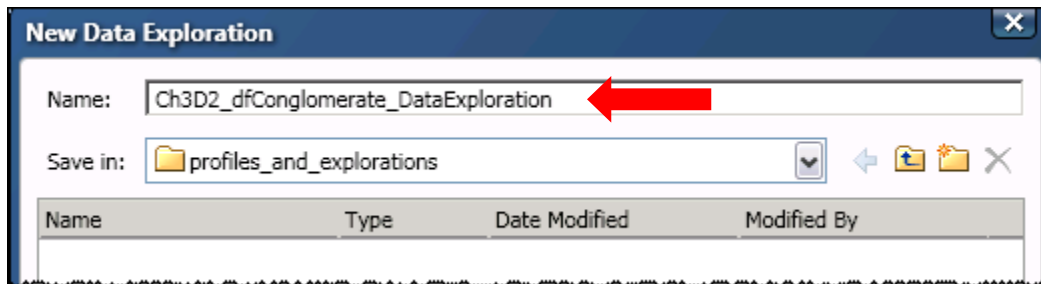
Demonstration:
Creating and Working with a
Data Exploration



Creating and Working with a Data Exploration

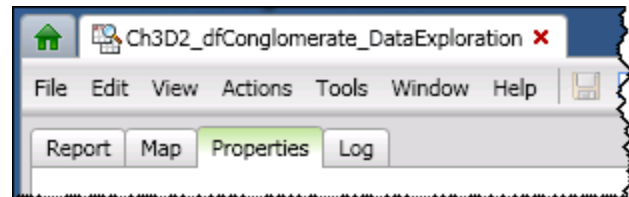
This demonstration illustrates the steps that are necessary to create a data exploration and explore the results.

1. If necessary, access Data Management Studio.
 - a. Select **Start** ⇒ **All Programs** ⇒ **DataFlux** ⇒ **Data Management Studio 2.7**.
 - b. Click **Cancel** to close the Log On window.
2. Verify that the **Home** tab is selected.
3. Click the **Folders** riser bar.
4. If necessary, expand the **Basics Demos** repository.
5. Right-click the **profiles_and_explorations** folder and select **New** ⇒ **Data Exploration**.
 - a. Enter **Ch3D2_dfConglomerate_DataExploration** in the **Name** field.



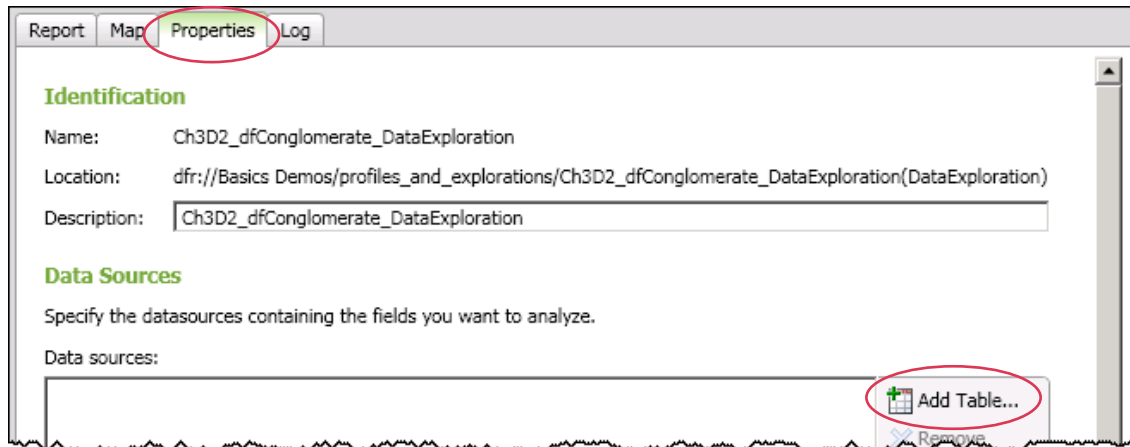
- b. Click **OK**.

The new data exploration appears on a primary tab.



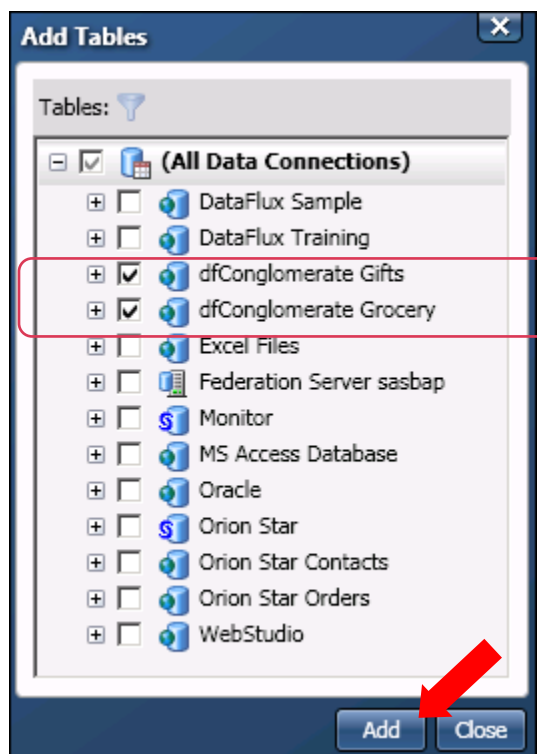
6. Verify that the **Properties** tab is selected.

7. Define the data to be explored.
 - a. In the Data Sources area, click **Add Table**.

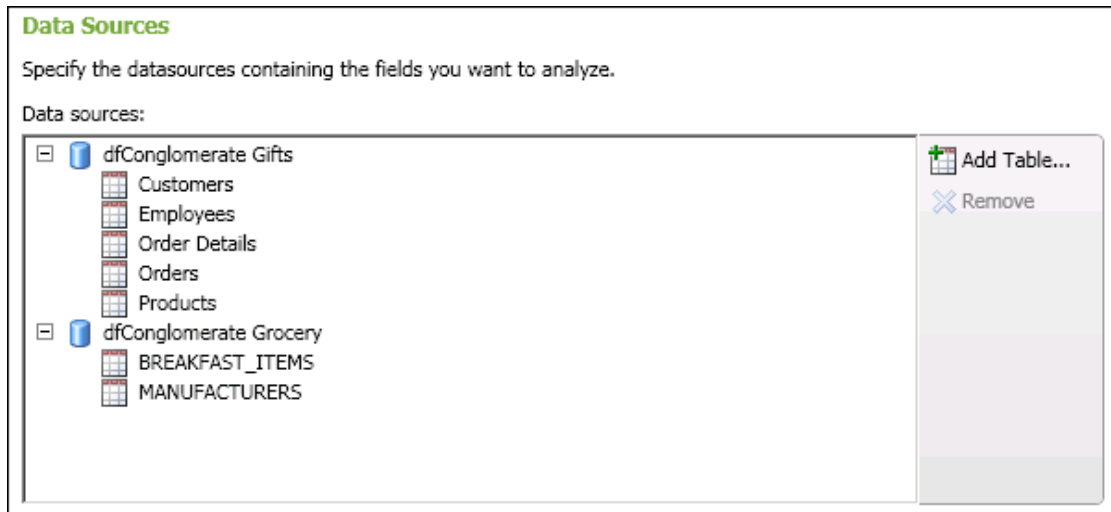


The Add Tables window appears.

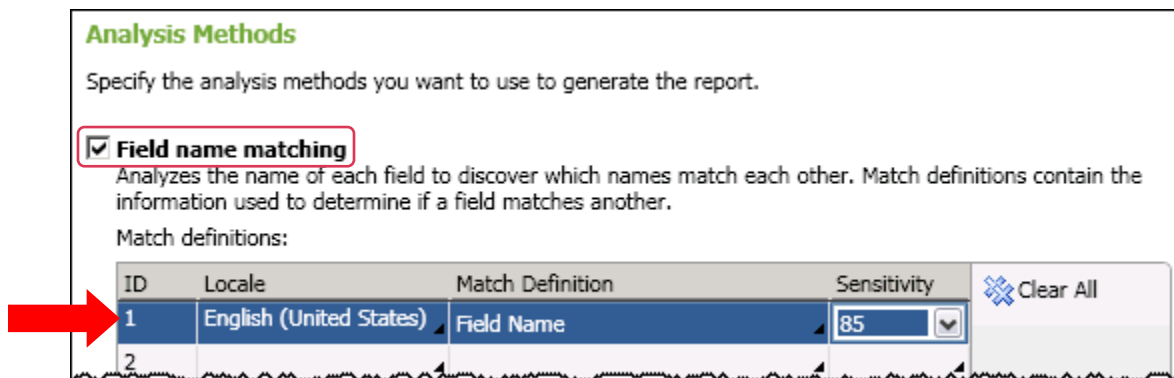
- 1) Expand **(All Data Connections)**.
- 2) Click the **dfConglomerate Gifts** check box.
- 3) Click the **dfConglomerate Grocery** check box.



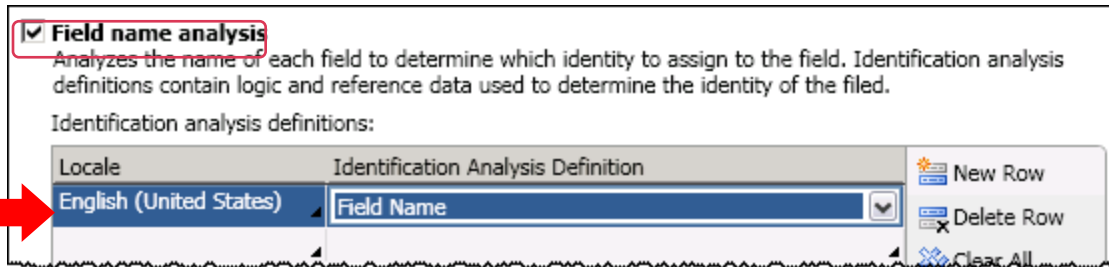
- 4) Click **Add**.
 - 5) Click **Close** to close the Add Tables window.
- b. If necessary, expand each of the data sources.
 - c. Verify that seven tables appear in the Data Sources area.



8. Specify the settings for field name matching analysis.
 - a. In the Analysis Methods area, click **Field name matching**.
 - b. Click once in the **Locale** field. This reveals a selection tool.
 - c. Click in the **Locale** field and select **English (United States)**.
 - d. Click once in the **Match Definition** field. This reveals a selection tool.
 - e. Click in the **Match Definition** field and select **Field Name**.
 - f. Click once in the **Sensitivity** field. This reveals a selection tool.
 - g. Click in the **Sensitivity** field and select **85**.

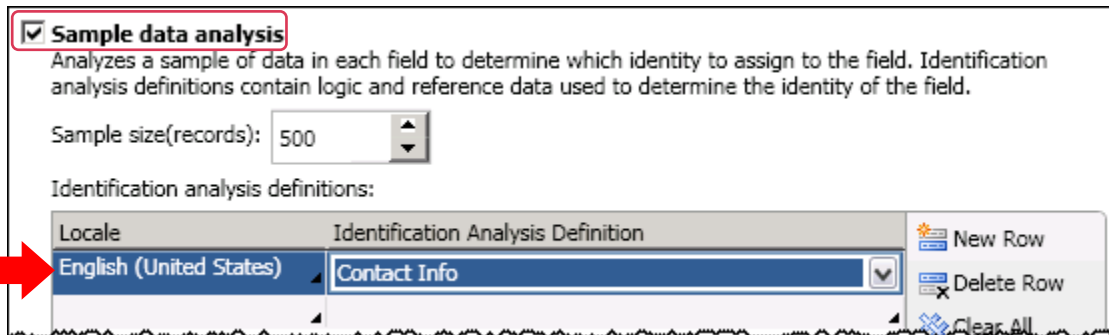


9. Specify the settings for field name analysis.
 - a. In the Analysis Methods area, click **Field name analysis**.
 - b. Click once in the **Locale** field. This reveals a selection tool.
 - c. Click in the **Locale** field and select **English (United States)**.
 - d. Click once in the **Identification Analysis Definition** field. This reveals a selection tool.
 - e. Click in the **Identification Analysis Definition** field and select **Field Name**.



10. Specify the settings for sample data analysis.

- a. In the Analysis Methods area, click **Sample data analysis**.
- b. Enter **500** for **Sample size(records)**.
- c. Click once in the **Locale** field. This reveals a selection tool.
- d. Click in the **Locale** field and select **English (United States)**.
- e. Click once in the **Identification Analysis Definition** field. This reveals a selection tool.
- f. Click in the **Identification Analysis Definition** field and select **Contact Info**.

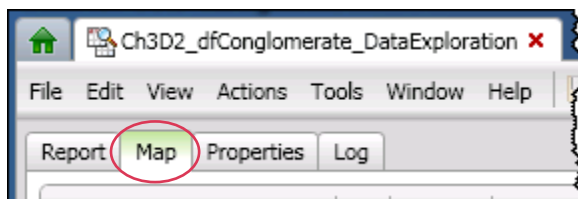


11. Select **Actions** ⇒ **Run** to execute the data exploration.

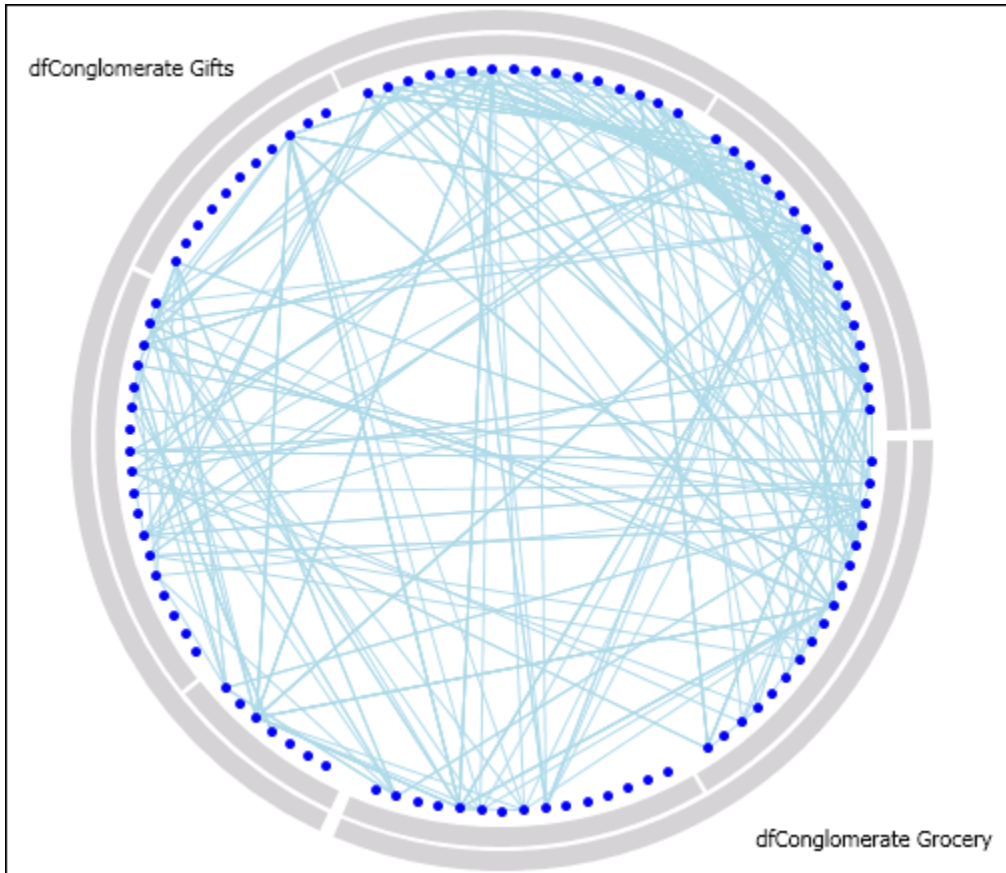
Note: You can also click (**Run Exploration**) on the toolbar to execute the data exploration.

12. Explore the field relationship map results.

- a. Click the **Map** tab.



The field relationship map appears:

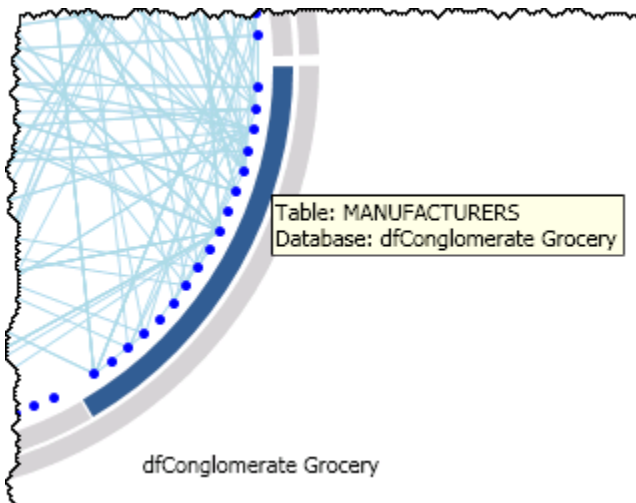


The outer ring segments represent the selected data connections: dfConglomerate Gifts and dfConglomerate Grocery.

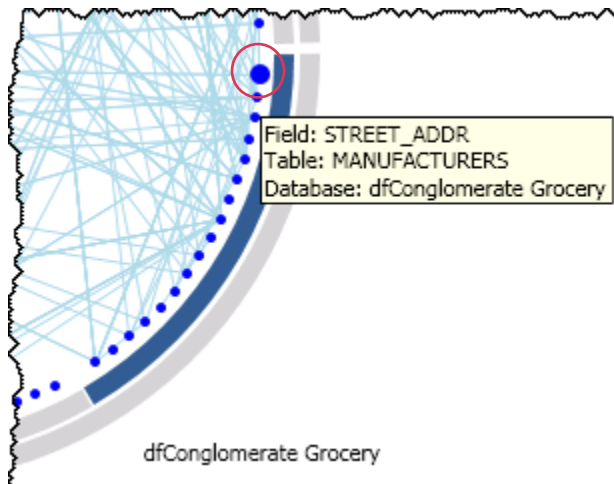
The inner ring segments represent the selected tables from each data connection. Moving the cursor over a segment of the inner ring displays the table name.

The dots represent the fields in each table.

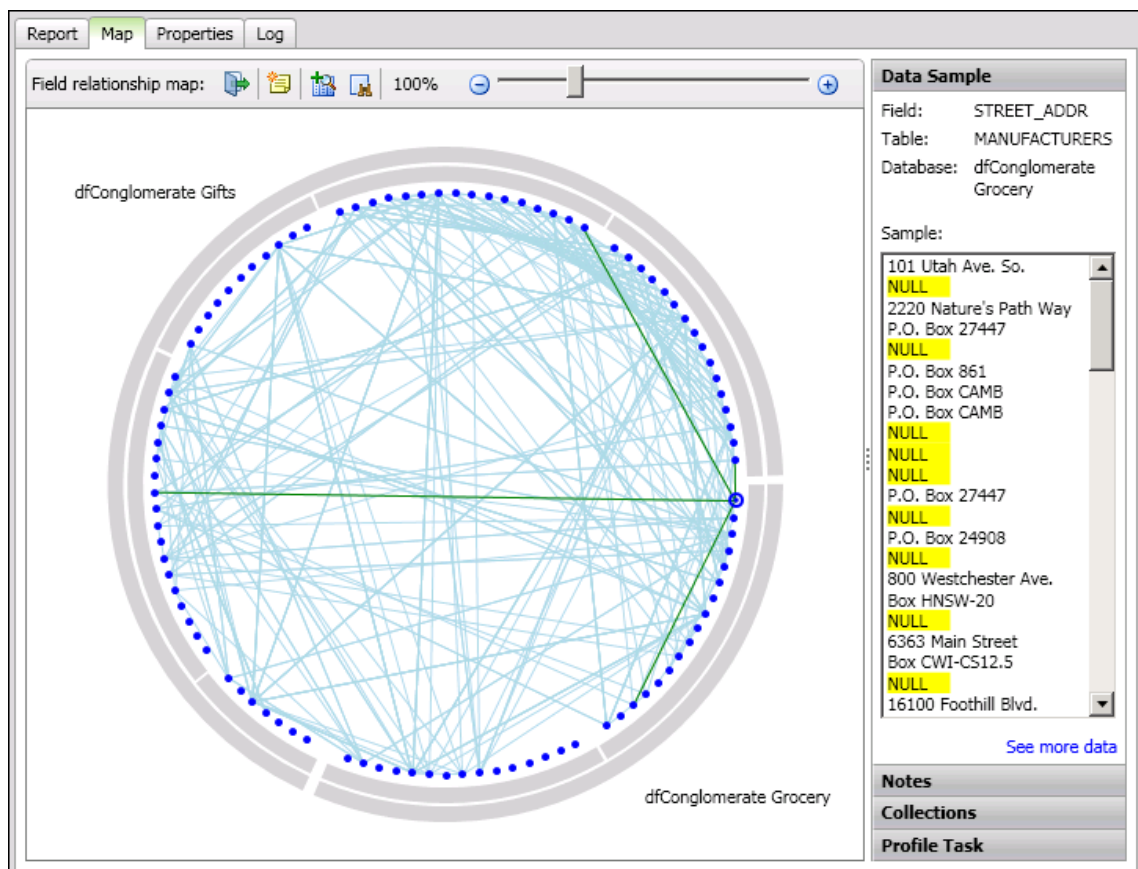
- b. Locate the **MANUFACTURERS** table from the dfConglomerate Grocery data connection.



c. Locate the dot that represents the **STREET_ADDR** field.



d. Click the dot that represents the **STREET_ADDR** field.

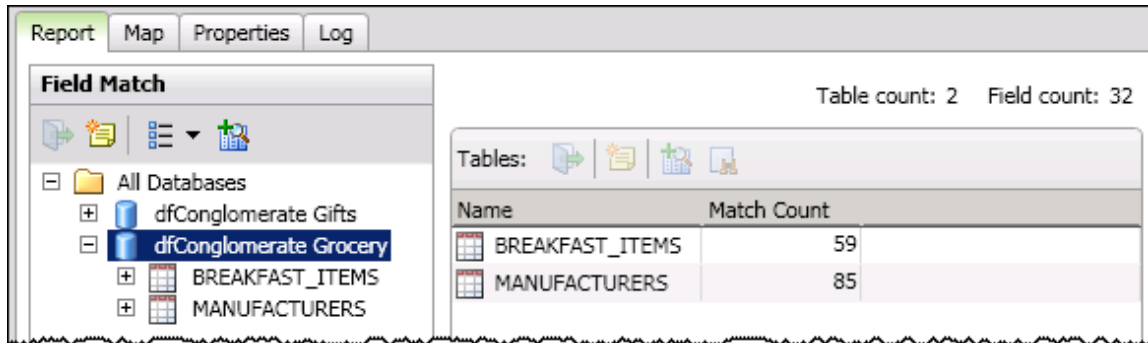


A sample of data values appears in the pane on the right.

The green lines indicate that four other fields are related to the **STREET_ADDR** field.

- CONTACT_ADDRESS in the **Manufacturers** table
- SHIP ADDRESS in the **Orders** table
- ADDRESS in the **Employees** table
- ADDRESS in the **Customers** table

13. Explore the data exploration Field Match report results.
 - a. Click the **Report** tab.
 - b. Verify that the **Field Match** riser bar is selected.
 - c. Expand **All Databases**.
 - d. Expand **dfConglomerate Grocery** to display a list of tables for this data connection.
 - e. Click **dfConglomerate Grocery**.

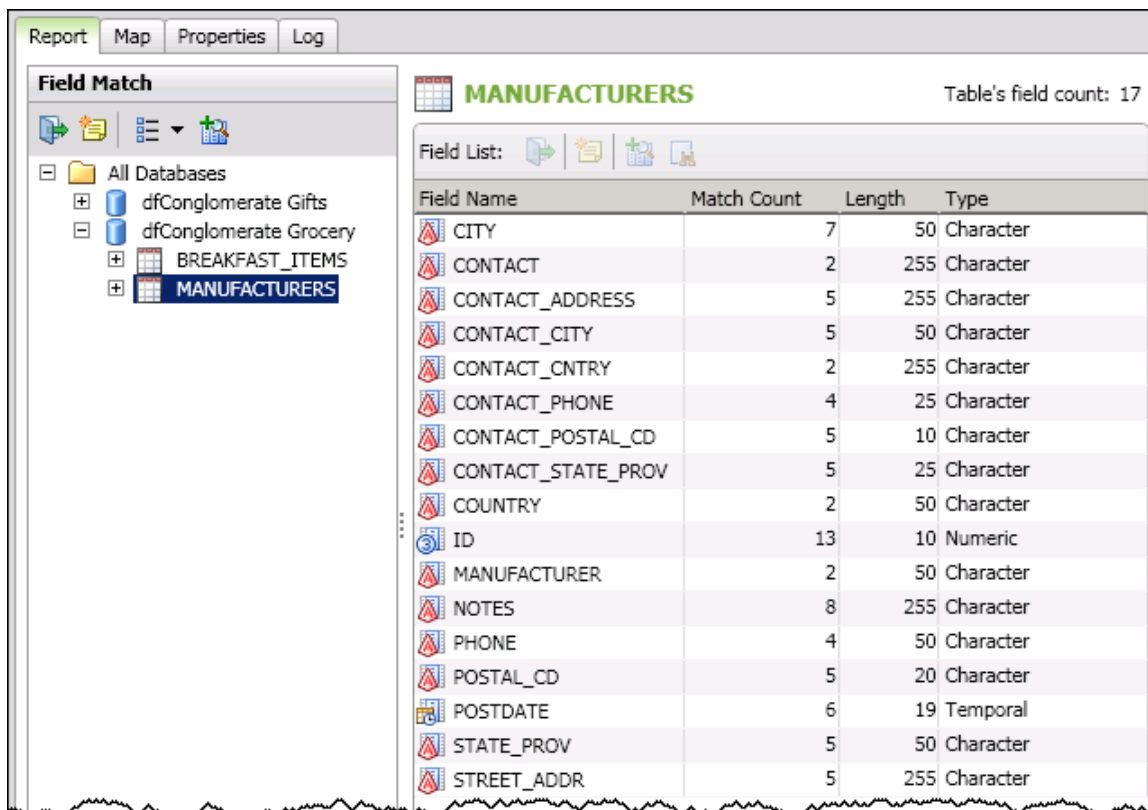


For the two tables in the dfConglomerate Grocery connection, there are 32 fields.

There are 59 fields that match the fields in the BREAKFAST_ITEMS table.

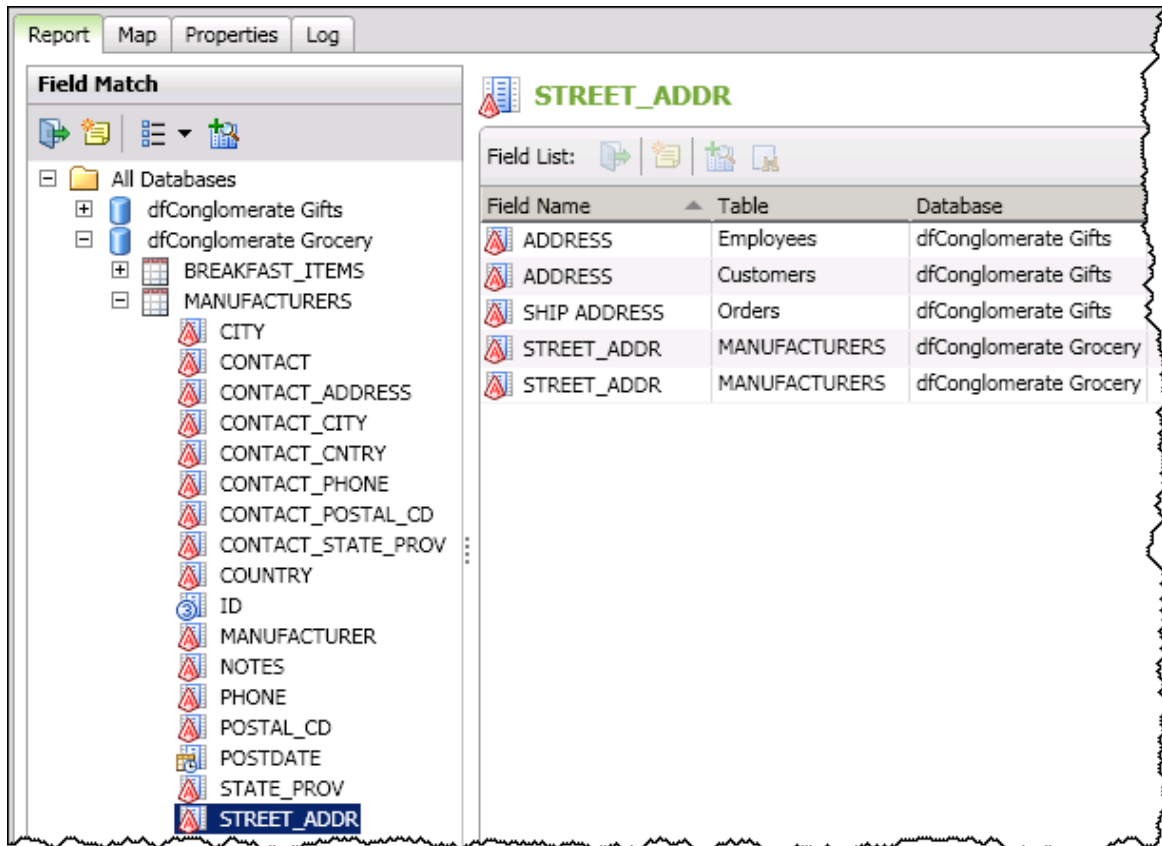
There are 85 fields that match the fields in the MANUFACTURERS table.

- f. Click the **MANUFACTURERS** table.



The 17 fields from the MANUFACTURERS table are displayed with the number of matching fields found for each.

- g. Expand the **MANUFACTURERS** table to display a list of fields for this table.
- h. Click the **STREET_ADDR** field.



Information about the fields related to the **STREET_ADDR** field appears.

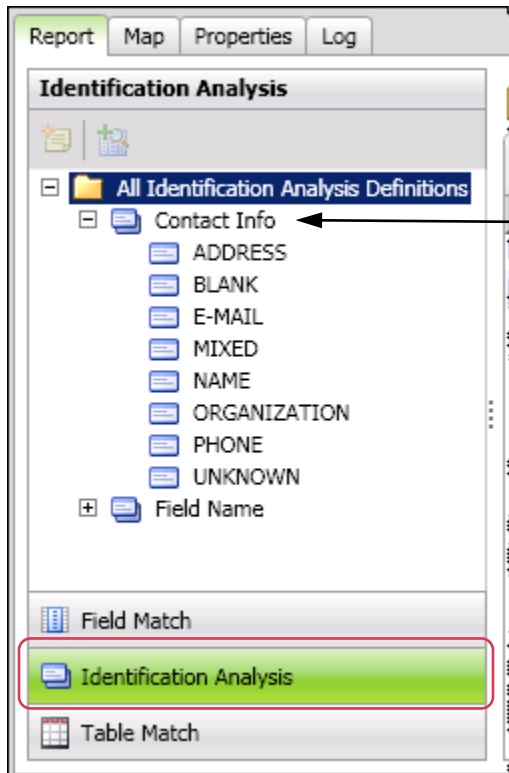
For each of the fields related to STREET_ADDR, you can discover if the relationship is listed as an EXACT MATCH (the field names are spelled exactly the same) or as FIELD NAME (the field names matched according to the rules in the Field Name match definition selected for field name matching analysis).

The below view is an arranged view to see the field names and the method of matching:

Field Name	Method
ADDRESS	Field Name
ADDRESS	Field Name
SHIP ADDRESS	Field Name
STREET_ADDR	Field Name
STREET_ADDR	EXACT MATCH

14. Explore the data exploration Identification Analysis report results.

- a. If necessary, click the **Report** tab.
- b. Click the **Identification Analysis** riser bar.
- c. Expand **All Identification Analysis Definitions**.
- d. Expand **Contact Info**.



The Contact Info identification analysis definition was chosen on the Properties tab for **Sample data analysis**.

The Contact Info identification analysis definition has eight (8) defined categorizations for data (ADDRESS, BLANK, E-MAIL, MIXED, NAME, ORGANIZATION, PHONE, UNKNOWN). This definition inspects the data values (we chose a sample size of 500) to see if the data seem to be representative of ADDRESS data, or BLANK data or E-MAIL data or ...

- e. Click the **ADDRESS** category.

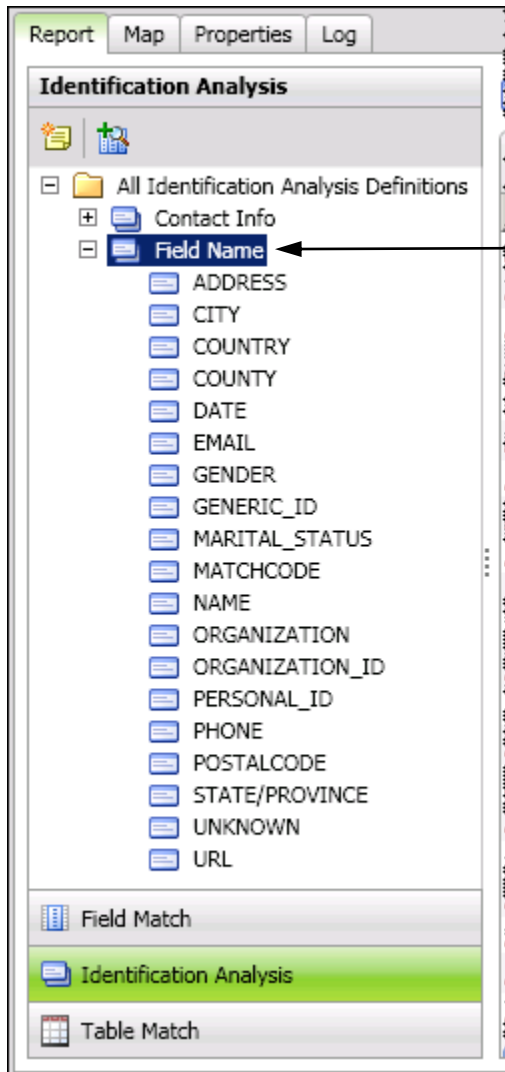
Each of the listed fields (26) is identified or categorized as **ADDRESS**, one of the categories from the Contact Info Identification Analysis definition, based on a sample of data values.

The screenshot shows the 'Identification Analysis' tool interface. The left-hand navigation pane is expanded to show the 'Contact Info' category, with 'ADDRESS' selected. The main window displays a table titled 'Identity matches:' with the following data:

Field Name	Table	Database						
ADDRESS	Customers	dfConglomerate Gifts	107	Ch	100	Sai		
ADDRESS	Employees	dfConglomerate Gifts	107	Ch	100	Sai		
BRAND	BREAKFAST_ITEM	dfConglomerate Grocei	16	Ch	0%	Sai		
CITY	Customers	dfConglomerate Gifts	50	Ch	5%	Sai		
CITY	MANUFACTURERS	dfConglomerate Grocei	50	Ch	7%	Sai		
CONTACT_ADDRESS	MANUFACTURERS	dfConglomerate Grocei	255	Ch	99%	Sai		
CONTACT_CITY	MANUFACTURERS	dfConglomerate Grocei	50	Ch	8%	Sai		
CONTACT_CNTRY	MANUFACTURERS	dfConglomerate Grocei	255	Ch	79%	Sai		
CONTACT_POSTAL_CD	MANUFACTURERS	dfConglomerate Grocei	10	Ch	92%	Sai		
CONTACT_STATE_PRO	MANUFACTURERS	dfConglomerate Grocei	25	Ch	2%	Sai		
COUNTRY	MANUFACTURERS	dfConglomerate Grocei	50	Ch	81%	Sai		
COUNTRY/REGION	Customers	dfConglomerate Gifts	50	Ch	76%	Sai		
COUNTRY/REGION	Employees	dfConglomerate Gifts	50	Ch	100	Sai		
MANUFACTURER	MANUFACTURERS	dfConglomerate Grocei	50	Ch	1%	Sai		
NAME	BREAKFAST_ITEM	dfConglomerate Grocei	32	Ch	1%	Sai		
POSTAL_CD	MANUFACTURERS	dfConglomerate Grocei	20	Ch	87%	Sai		
QUANTITY PER UNIT	Products	dfConglomerate Gifts	50	Ch	14%	Sai		
SHIP ADDRESS	Orders	dfConglomerate Gifts	107	Ch	100	Sai		

Additional categorizations can be examined similarly.

- f. Collapse **Contact Info**.
- g. Expand **Field Name**.



The Field Name identification analysis definition was chosen on the Properties tab for Field name analysis.

The Field Name identification analysis definition has nineteen (19) defined categorizations for data (ADDRESS, CITY, COUNTRY, COUNTY, DATE, EMAIL, GENDER, GENERIC_ID, MARITAL_STATUS, MATCHCODE, NAME, ORGANIZATION, ORGANIZATION_ID, PERSONAL_ID, PHONE, POSTALCODE, STATE/PROVINCE, UNKNOWN, URL). This definition inspects the field names to see if a field name seems to be an ADDRESS field name, or a CITY field name, or a COUNTRY field name, or ...

h. Click the **ADDRESS** category.

Each of the listed fields is identified as **ADDRESS**, one of the categories from the Field Name identification analysis definition.

The screenshot shows the 'Identification Analysis' tool interface. On the left, a tree view under 'All Identification Analysis Definitions' has 'Field Name' expanded, with 'ADDRESS' selected. On the right, the 'ADDRESS' category is active, showing an 'Identity match count: 5'. Below this, a table lists the following identity matches:

Field Name	Table	Database				
ADDRESS	Customers	dfConglomerate Gifts	107	Ch	Fie	
ADDRESS	Employees	dfConglomerate Gifts	107	Ch	Fie	
CONTACT_ADDRESS	MANUFACTURERS	dfConglomerate Grocery	255	Ch	Fie	
SHIP ADDRESS	Orders	dfConglomerate Gifts	107	Ch	Fie	
STREET_ADDR	MANUFACTURERS	dfConglomerate Grocery	255	Ch	Fie	

Additional categorizations of fields can be examined similarly.

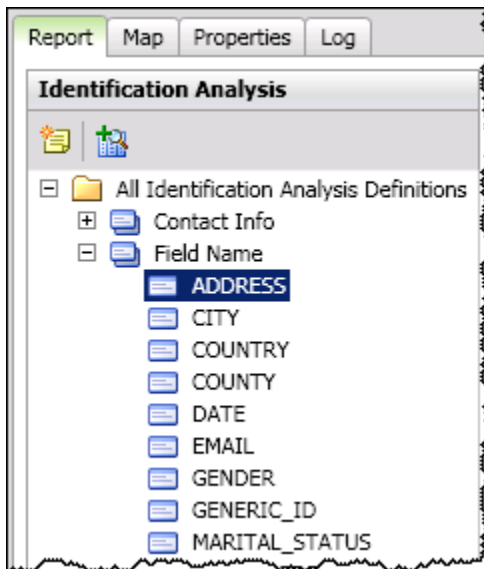
End of Demonstration



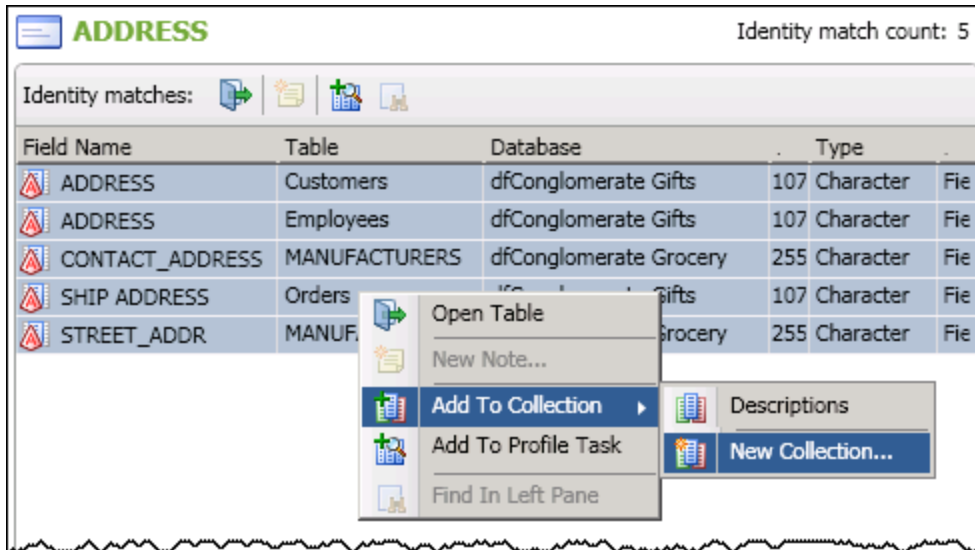
Creating a Collection from a Data Exploration

This demonstration illustrates the steps that are necessary to create a data collection when reviewing the results of a data exploration.

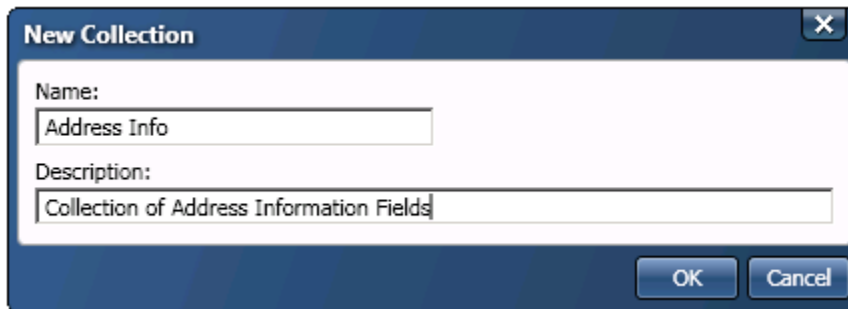
1. If necessary, access Data Management Studio.
 - a. Select **Start** ⇒ **All Programs** ⇒ **DataFlux** ⇒ **Data Management Studio 2.7**.
 - b. Click **Cancel** to close the Log On window.
2. If necessary, access **Ch3D2_dfConglomerate_DataExploration**.
 - a. Verify that the **Home** tab is selected.
 - b. Click the **Folders** riser bar.
 - c. Expand **Basics Demos**.
 - d. Click **profiles_and_explorations**.
 - e. In main information area, double-click **Ch3D2_dfConglomerate_Data_Exploration**.
The data exploration opens on a primary tab.
6. If necessary, click the **Report** tab.
7. Click the **Identification Analysis** riser bar.
8. Expand the **All Identification Analysis Definitions** folder.
9. Expand **Field Name**.
10. Click **ADDRESS**.



11. In the main area, select all the fields that are identified as **ADDRESS** fields.
 - a. Click the first field.
 - b. Hold down the Shift key.
 - c. Click the last field.
12. Right-click one of the selected fields and select **Add To Collection** ⇒ **New Collection**.

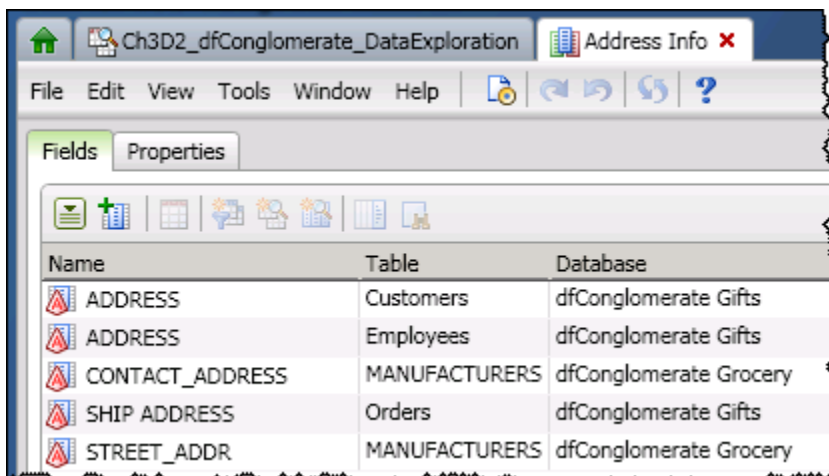


13. Enter information for the new collection.
 - a. Enter **Address Info** in the **Name** field.
 - b. Enter **Collection of Address Information Fields** in the **Description** field.



- c. Click **OK**.

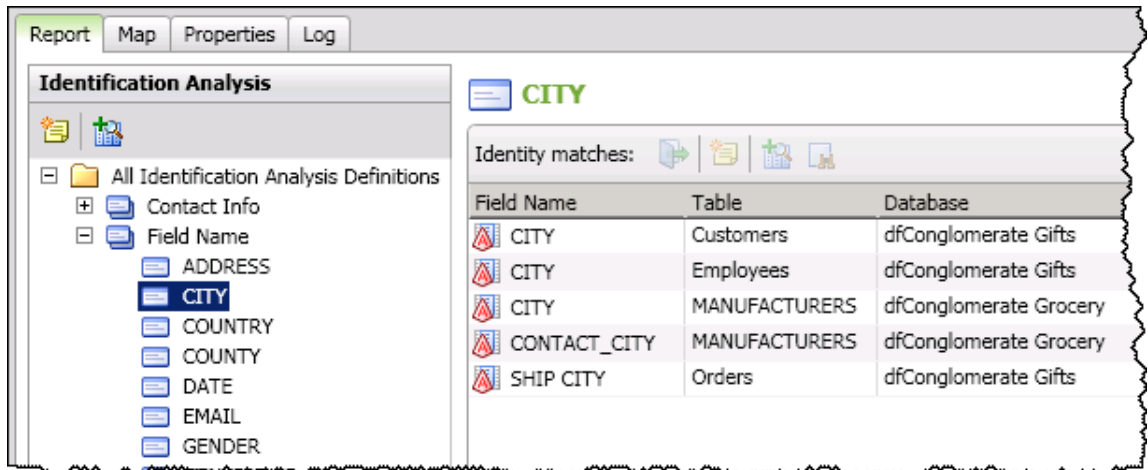
The new collection appears on a separate tab.



14. Click the data exploration tab (**Ch3D2_dfConglomerate_Data_Exploration**).
15. Add city fields to the collection.

- a. Click **CITY** (under Field Name identification definition).

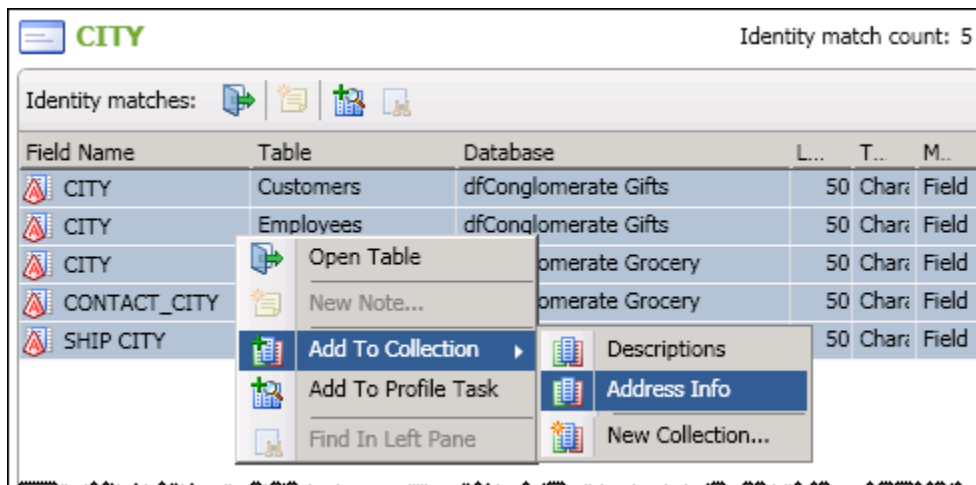
The fields identified as **CITY** fields appear in the main area.



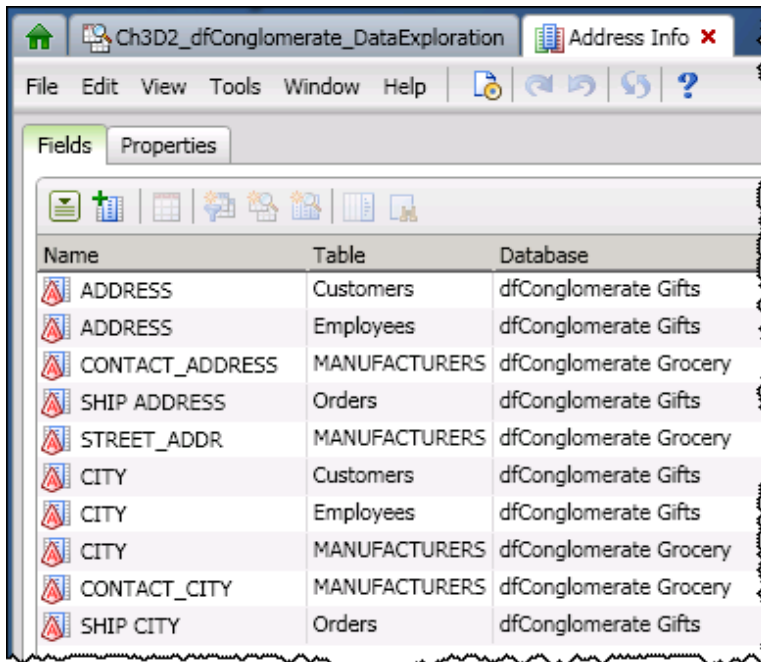
- b. Select all the fields that are identified as **CITY** fields.

- 1) Click the first field.
- 2) Hold down the Shift key.
- 3) Click the last field.

- c. Right-click one of the selected fields and select **Add To Collection** ⇒ **Address Info**.



16. Click the collection tab (**Address Info**).
17. Verify that the city fields were added to the collection.



18. Click the data exploration tab (**Ch3D2_dfConglomerate_Data_Exploration**).
19. Add country fields to the collection.
 - a. Click **COUNTRY** (under Field Name identification definition).
The fields identified as **COUNTRY** fields appear in the main area.
 - b. Select all the fields that are identified as **COUNTRY** fields.
 - 1) Click the first field.
 - 2) Hold down the Shift key.
 - 3) Click the last field.
 - c. Right-click one of the selected fields and select **Add To Collection** ⇨ **Address Info**.
20. Click the collection tab (**Address Info**).
21. Verify that the fields that are identified as **COUNTRY** are now part of the Address Info collection.
22. Click the data exploration tab (**Ch3D2_dfConglomerate_Data_Exploration**).
23. Add postal code fields to the collection.
 - a. Click **POSTALCODE** (under Field Name identification definition).
The fields identified as **POSTALCODE** fields appear in the main area.
 - b. Select all the fields that are identified as **POSTALCODE** fields.
 - 1) Click the first field.
 - 2) Hold down the Shift key.
 - 3) Click the last field.
 - c. Right-click one of the selected fields and select **Add To Collection** ⇨ **Address Info**.

24. Click the collection tab (**Address Info**).
25. Verify that the fields that are identified as **POSTALCODE** are now part of the Address Info collection.
26. Click the data exploration tab (**Ch3D2_dfConglomerate_Data_Exploration**).
27. Add State/Province fields to the collection.
 - a. Click **STATE/PROVINCE** (under Field Name identification definition).
The fields identified as **STATE/PROVINCE** fields appear in the main area.
 - b. Select all the fields that are identified as **STATE/PROVINCE** fields.
 - 1) Click the first field.
 - 2) Hold down the Shift key.
 - 3) Click the last field.
 - c. Right-click one of the selected fields and select **Add To Collection** ⇒ **Address Info**.

28. Click the collection tab (**Address Info**).
29. Verify that the fields that are identified as **ADDRESS, CITY, COUNTRY, POSTALCODE, and STATE/PROVINCE** are now part of the Address Info collection.

Name	Table	Database
ADDRESS	Customers	dfConglomerate Gifts
ADDRESS	Employees	dfConglomerate Gifts
CONTACT_ADDRESS	MANUFACTURERS	dfConglomerate Grocery
SHIP ADDRESS	Orders	dfConglomerate Gifts
STREET_ADDR	MANUFACTURERS	dfConglomerate Grocery
CITY	Customers	dfConglomerate Gifts
CITY	Employees	dfConglomerate Gifts
CITY	MANUFACTURERS	dfConglomerate Grocery
CONTACT_CITY	MANUFACTURERS	dfConglomerate Grocery
SHIP CITY	Orders	dfConglomerate Gifts
CONTACT_CNTRY	MANUFACTURERS	dfConglomerate Grocery
COUNTRY	MANUFACTURERS	dfConglomerate Grocery
CONTACT_POSTAL_CD	MANUFACTURERS	dfConglomerate Grocery
POSTAL_CD	MANUFACTURERS	dfConglomerate Grocery
SHIP ZIP/POSTAL CODE	Orders	dfConglomerate Gifts
ZIP/POSTAL CODE	Customers	dfConglomerate Gifts
ZIP/POSTAL CODE	Employees	dfConglomerate Gifts
CONTACT_STATE_PROV	MANUFACTURERS	dfConglomerate Grocery
COUNTRY/REGION	Customers	dfConglomerate Gifts
COUNTRY/REGION	Employees	dfConglomerate Gifts
SHIP COUNTRY/REGION	Orders	dfConglomerate Gifts
SHIP STATE/PROVINCE	Orders	dfConglomerate Gifts
STATE/PROVINCE	Customers	dfConglomerate Gifts
STATE/PROVINCE	Employees	dfConglomerate Gifts
STATE_PROV	MANUFACTURERS	dfConglomerate Grocery

30. Select **File** ⇒ **Close Collection**.
The data exploration tab should be active.
31. Select **File** ⇒ **Close Exploration**.

End of Demonstration