

# Working with Data Enrichment in a Data Job

<b>Section 1: Introduction</b> .....	<b>3</b>
<b>Section 2: Data Enrichment Jobs</b> .....	<b>7</b>
Demonstration: Working with Address Verification and Geocoding Nodes .....	10
Exercise .....	21
<b>Section 3: Solutions</b> .....	<b>23</b>
Solutions to Exercises .....	23
Solutions to Activities and Questions .....	31




## Section 1: Introduction

### Address Verification

*Address verification* identifies, corrects, and enhances address information.

Original Address	
940 Cary Parkway	
27513	



Verified Address Information	
Street Address	940 NW CARY PKWY
City	CARY
State	NC
ZIP	27513
ZIP+4	2792
County Name	WAKE
Congressional District	2

3




The address verification lookup process requires only a valid street address and postal code, or a valid street address with the corresponding city and state values. If these values match an address in the lookup database, then the data can be enriched with a multitude of additional data fields. In the example shown, the address 940 Cary Parkway with a zip code of 27513 is passed into the address verification process. Because this is a valid address, additional information can be added to the data row (for example, City, State, ZIP+4, County Name, and Congressional District).

## Geocoding

*Geocoding* enhances address information with latitude and longitude values.

Original Address	
940 NW CARY PKWY	
CARY , NC 27513-2792	



Verified Geocode Information	
Latitude	35.818339
Longitude	-78.797333

4

Copyright © SAS Institute Inc. All rights reserved.



Geocoding latitude and longitude information can be used to map locations and plan efficient delivery routes. Geocoding can be licensed to return this information for the centroid of the postal code or at the roof-top level.

## Address Verification Reference Data Sources

Reference data sources to be used in address verification processes include the following sources:

- US Address Verification (USPS – supports CASS certification)
- Canada Address Verification (Canada Post – supports SERP certification)
- North America Postal Level Geocode (includes PhonePlus)
- US Street Level Geocode \*\*
- Loqate Data (worldwide address verification and geocoding)

\*\* requires North American Postal Level Geocode data pack

5

Copyright © SAS Institute Inc. All rights reserved.



The address verification nodes in Data Management Studio require that one (or more) third-party reference data sources be licensed. These include the following data sources:

US Address Verification	United States Postal Service database, which supports CASS (Coding Accuracy Support System) certification
Canada Address Verification	Canada Post database, which supports SERP (Software Evaluation and Recognition Program) certification
North America Postal Level Geocode (includes PhonePlus)	US and Canada database that includes Geocoding and PhonePlus data
US Street Level Geocode	Additional level of detail to the North American Postal Level Geocode database added
Loqate	Worldwide databases, which also support worldwide geocoding

**Note:** The US Street Level Geocode data requires that the North American Postal Level Geocoding data pack be installed first.

**Note:** More information about the reference data sources can be found by navigating to <https://support.sas.com/downloads/index.htm> and selecting **DataFlux Data Updates** on the Product Category tab.

## x.01 Poll

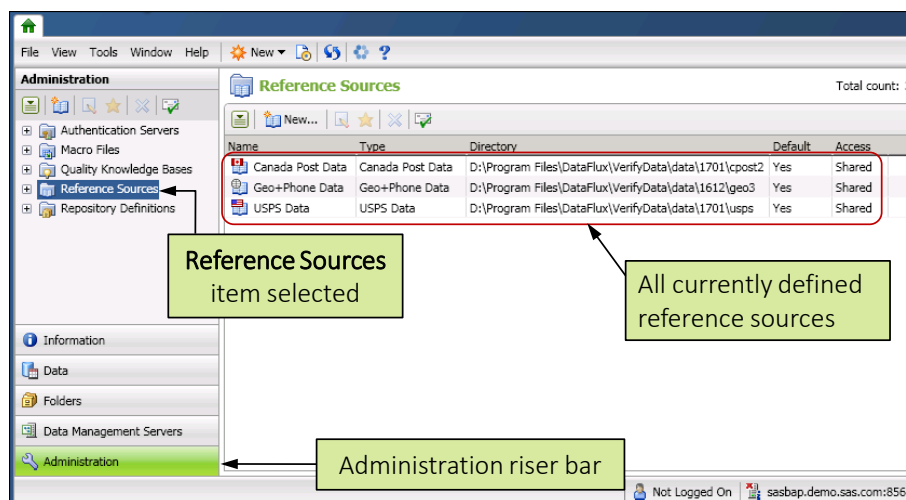
Address Verification and Enrichment reference sources exist only for the United States and Canada.

- True
- False

## Section 2: Data Enrichment Jobs

### Registering Reference Data Sources

Reference data sources are registered on the Administration riser bar.

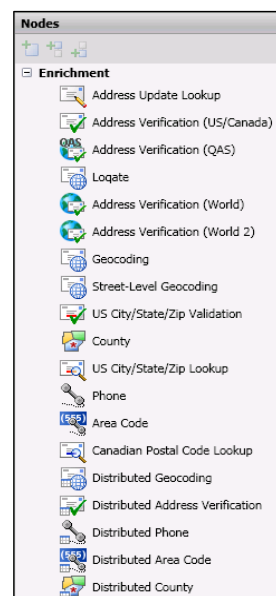


Reference data sources are registered in Data Management Studio metadata from the Administration riser bar. In the example shown, you can see that three reference data sources are registered for use.

### Using Address Verification Nodes in a Data Job

The nodes for performing address verification are found in the Enrichment grouping of nodes.

These nodes require the third-party reference data packs.



## Using Address Verification Nodes in a Data Job

The diagram illustrates the process of using address verification nodes in a data job. It shows a job flow starting with a 'Customers Table' data source, followed by an 'Address Verification' node, and then a 'Geocoding using Verifie...' node. A red arrow points from the 'Address Verification' node to a screenshot of its properties dialog.

The 'Address Verification (US/Canada) Properties' dialog shows the following input fields:

Field Name	Field Type
ID	
COMPANY	
LAST NAME	
FIRST NAME	
ADDRESS	Address Line 1
CITY	City
STATE/PROVINCE	State
ZIP/POSTAL CODE	Zip
COUNTRY/REGION	

The dialog also shows a list of available output fields and a table of selected output fields:

Available:	Selected:
Addresssee	
Firm	
Address Line 1	Address Line 1
Address Line 2	
City	City_V
State	State_V
ZIP/Postal Code	ZIP/Postal_Code_V
Country	
US ZIP	US_ZIP
US ZIP4	US_ZIP4
US Street Number	US_Street_Number
US Street Pre-direction	
US Street Name	US_Street_Name
US Street Suffix	
US Street Post-direction	
US Street Secondary Abbrev	

The SAS logo is visible in the bottom right corner of the dialog screenshot.

The address verification process is performed by one of many available nodes in the Enrichment group of nodes. After a node is added to the data job flow, you can connect the node's input fields to the reference data source to validate the address. If the address is found in the reference data source, then there are several available fields that can be used to enrich the existing record.

For example, the job flow diagram shown contains an Address Verification node. When examining the properties of this node, we can see the list of fields being passed in to this node, or the input fields. The input fields can be matched to fields from the reference source using the Field Type area. Depending on the specified reference source, there will be a number of output fields that can be generated from the reference source.

It is important to note that these output fields are fields added to the result set \*in addition to\* the original fields that may be passing through this node. (Recall the Additional Outputs button can control which of the input fields will pass out of this node.) In the example shown, the original data has a CITY field (as shown in the Input listing). One of the available output fields is also by default named City. One must change this output field name to some name that work conflict with another field that is passing through the node. In the example the added output field City has been renamed to City\_V to avoid the potential replication of field names.



## x.02 Poll

The address verification and enrichment features for a data job require the use of third-party reference data.

- True
- False

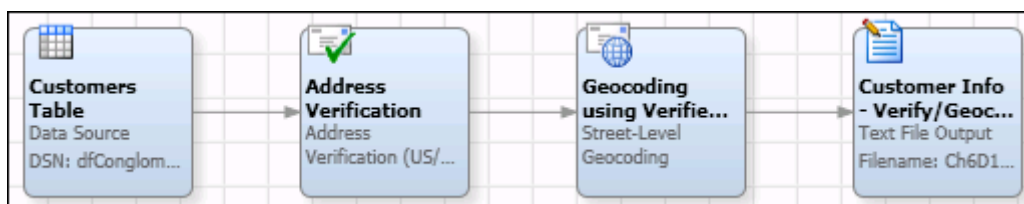


## Working with Address Verification and Geocoding Nodes

This demonstration uses an existing data job in the Basics Solutions repository. It illustrates the use of an Address Verification node to verify addresses and a Geocoding node to enhance addresses with latitude and longitude information.

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. Open a data job from the Basics Solutions repository.
  - a. Click the **Folders** riser bar.
  - b. Expand **Basics Solutions**.
  - c. Click **batch\_jobs**.
  - d. Double-click **Ch6D1\_AddressVerification**.

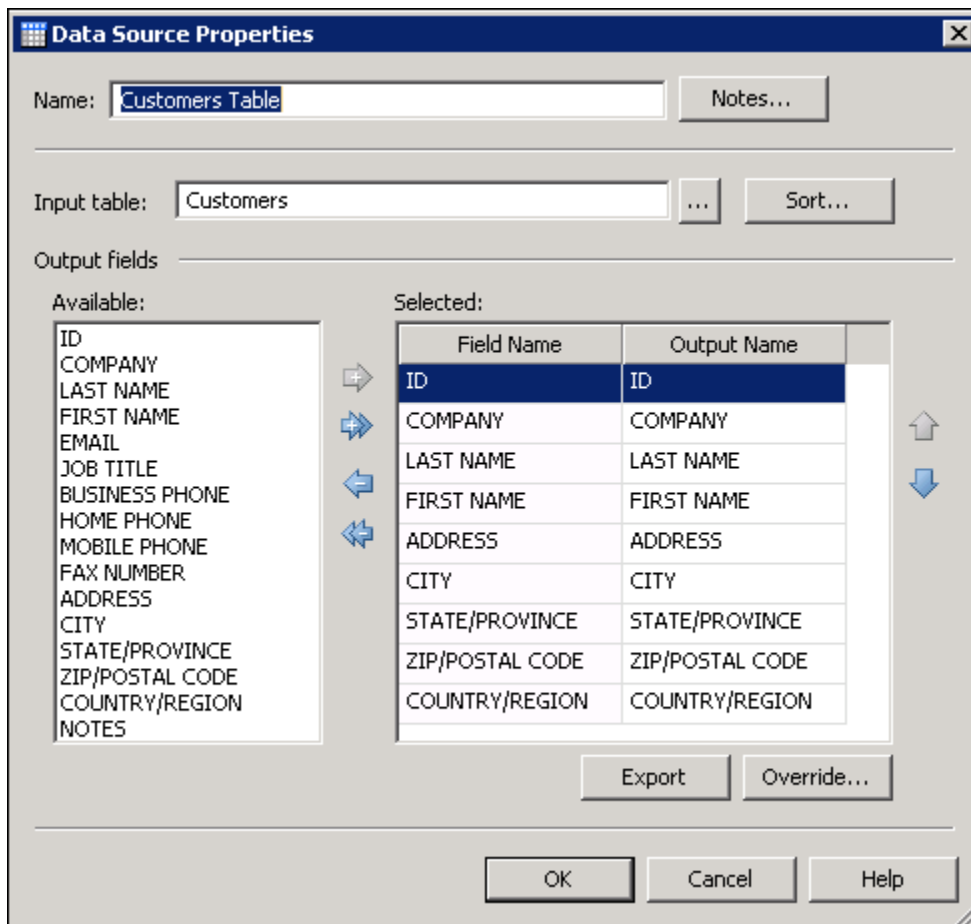
The data job appears on a new primary tab. The job flow should resemble the following:



3. Examine the properties of the Data Source node in the job flow.
  - a. Double-click the **Customers Table** Data Source node.
  - b. Verify that one the following fields were added to the Selected list:

**ID**  
**COMPANY**  
**LAST NAME**  
**FIRST NAME**  
**ADDRESS**  
**CITY**  
**STATE/PROVINCE**  
**ZIP/POSTAL CODE**  
**COUNTRY/REGION**

The Data Source Properties window should resemble the following:



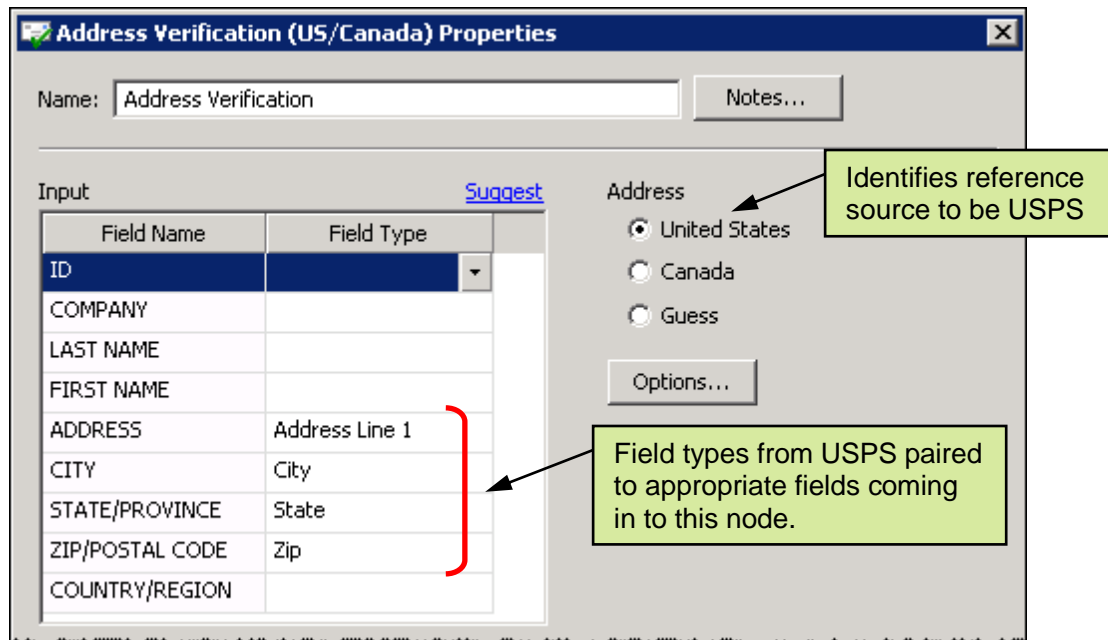
- c. Click **Cancel** to close the Data Source Properties window.
4. Right-click the **Customers Table** Data Source node and select **Preview**.

A sample of records appears on the Preview tab of the Details pane.

ID	COMPANY	LAST NAME	FIRST NAME	ADDRESS	CITY	STATE/PROVINCE
1	1 Transamerica Financial Group	Bedecs	Anna	3720 Wimbledon Lane	Birmingham	AL
2	2 DataFlux	Ramos	Antonio	4013 Winston Way	Birmingham	AL
3	3 Transamerica Occidental	Axen	Thomas	2437 Mountain Vista	Birmingham	AL
4	4 Transamerica Financial Svcs	Lee	Christina	6629 East Meadowlark Ln	Paradise Valley	AZ
5	5 Transamerica Occidental	O'Donnell	Martin	7921 E Parkview Lane	Scottsdale	AZ
6	6 (null)	Pérez-Olaeta	Francisco	15609 E Melrose St.	Gilbert	AZ
7	7 Transamerica Financial Svc	Xie	Ming-Yang	157 N. Sierra Vista Drive	Tuscon	Az.
8	8 Applied Computer Research	Andersen	Elizabeth	10116 East Winter Sun Drive	Scottsdale	AZ
9	9 Applied Data Svcs	Mortensen	Sven	1815 Loma Roja Dr	Santa Ana	CA
10	10 Applied Data Svcs	Wacker	Roland	16600 Calle Haleigh	Pacific Palisades	Ca.
11	11 Aratex Svcs Inc	Smith	Peter	969 Mc Kenzie Ave.	Watsonville	CA

5. Examine the properties for the Address Verification (US/Canada) node.
  - a. Double-click the **Address Verification** node to view the Properties window.
  - b. Verify that **United States** is selected in the **Address** area.
  - c. Verify the following input specifications:
    - 1) The **Address Line 1** field type is assigned to **ADDRESS**.
    - 2) The **City** field type is assigned to **CITY**.
    - 3) The **State** field type is assigned to **STATE/PROVINCE**.
    - 4) The **Zip** field type is assigned to **ZIP/POSTAL CODE**.

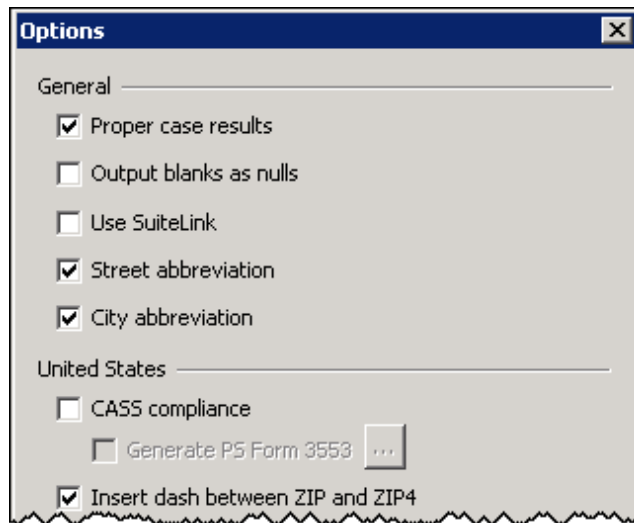
The Input information should resemble the following:



**Note:** You can click **Suggest** to attempt to match the fields from the data source with the fields in the reference source.

- d. Investigate options for address verification.
  - 1) Click **Options**.
  - 2) Verify that **Proper case results** is selected.
  - 3) Verify that **Street abbreviation** is selected.
  - 4) Verify that **City abbreviation** is selected.
  - 5) Verify that **CASS compliance** is *not* selected.
  - 6) Verify that **Insert dash between ZIP and ZIP4** is selected.

The Options window should resemble the following:

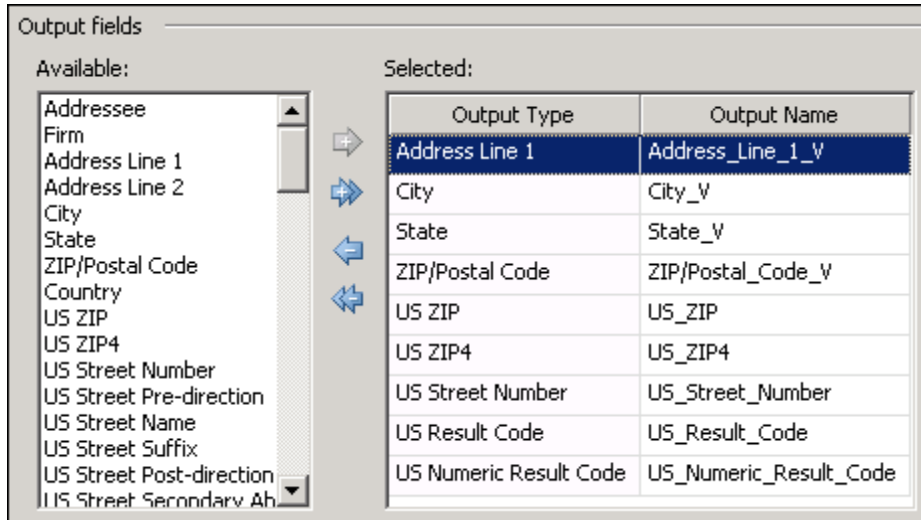


- 7) Click **Cancel** to close the Options window.
- e. Verify the output information.
- 1) In the Output fields area, verify that the following fields are in the Selected list:
 

<b>Address Line 1</b>	<b>US ZIP</b>
<b>City</b>	<b>US ZIP4</b>
<b>State</b>	<b>US Street Number</b>
<b>ZIP/Postal Code</b>	<b>US Result Code</b>
	<b>US Numeric Result Code</b>
  - 2) In the **Output Name** field, verify that **\_V** is after the following fields:
 

<b>Address_Line_1</b>
<b>City</b>
<b>State</b>
<b>ZIP/Postal_Code</b>

The Output field area should resemble the following:



- f. Click **Cancel** to close the Address Verification (US/Canada) Properties window.
6. Preview the data.
    - a. Right-click the **Address Verification** node and select **Preview**.
    - b. Scroll to the right to view verified address information.

	City_V	State_V	ZIP/Postal_Code_V	US_ZIP	US_ZIP4	US_Street_Number	US_Result_Code	US_Numeric_Result_Code
1	Mountain Brk	AL	35223-2732	35223	2732	3720	OK	0
2	Birmingham	AL	35213-3941	35213	3941	4013	OK	0
3	Vestavia	AL	35243-2856	35243	2856	2437	OK	0
4	Paradise Vly	AZ	85253-3620	85253	3620	6629	OK	0
5	Scottsdale	AZ	85255-2708	85255	2708	7921	OK	0
6	Gilbert	AZ	85297-1121	85297	1121	15609	OK	0
7	Tucson	AZ	85719-3840	85719	3840	157	OK	0
8	Scottsdale	AZ	85262-3105	85262	3105	10116	OK	0
9	Santa Ana	CA	92705-3332	92705	3332	1815	OK	0
10	Pacific Plsds	CA	90272-1969	90272	1969	16600	OK	0
11	Watsonville	CA	95076-3527	95076	3527	969	OK	0

Notice the following:

- The street types from the original **Address** field are abbreviated in the verified **Address** field.
- Some verified **City** values are updated to match to the ZIP code information.
- Some cities from the original **City** field are abbreviated in the verified **City** field.
- Some values from the **State** field are standardized to the pattern AA in the verified field.
- Most records in the preview have a US Result Code = **OK** and US Numeric Result Code = **0**.

**Note:** The **US\_Result\_Code** field indicates whether the address was successfully verified. If the address was not successfully verified, the code indicates the cause of failure.

**Note:** The **US\_Numeric\_Result\_Code** field provides a numeric value for the result. Possible values for both fields are as follows:

Text Result Code	Numeric Result Code	Description
OK	0	Address was verified successfully.
PARSE	11	Error parsing address. Components of the address might be missing.
CITY	12	Could not locate city, state, or ZIP code in the USPS database. At least, city and state or ZIP code must be present in the input.
MULTI	13	Ambiguous address. There are two or more possible matches for this address with different data.
NOMATCH	14	No matching address is found in the USPS data.
OVER	15	One or more input strings is too long (maximum 100 characters).

7. Examine the properties for the Street-Level Geocoding node in the job flow.
  - a. Double-click the **Street-Level Geocoding** node (named **Geocoding using Verified Zip and Street Number**) to view the Properties window.
  - b. Verify that **United States** is selected as the address type.
  - c. In the Input fields area, verify that the **ZIP/Postal\_Code\_V** field type is set to **Postal/ZIP Code**.
  - d. Verify that the **Street number** field type is set to **US\_Street\_Number**.
  - e. In the Output fields area, verify that the following fields are in Selected list:
    - Geocode Result Code**
    - Geocode Latitude**
    - Geocode Longitude**
    - Geocode FIPS**

The Street-Level Geocoding Properties window should resemble the following:

**Street-Level Geocoding Properties**

Name:

---

Input fields:

Field Name	Field Type
ZIP/Postal_Code_V	Postal/ZIP Code
US_ZIP	
US_ZIP4	
US_Street_Number	Street number
US_Result_Code	

Address type:

United States  
 Canada  
 Guess

---

Output fields:

Available:

- Geocode Result Code
- Geocode Latitude
- Geocode Longitude
- Geocode Country Code
- Geocode FIPS
- Geocode Census Tract
- Geocode Census Block
- Geocode FIPS MCD Cod
- Geocode MSA or CMSA
- Geocode PMSA (for CM...

Selected:

Output Type	Output Name
Geocode Result Code	Geocode_Result_Code
Geocode Latitude	Geocode_Latitude
Geocode Longitude	Geocode_Longitude
Geocode FIPS	Geocode_FIPS

- f. Click **Cancel** to close the Street-Level Geocoding Properties window.



8. Preview the data.
  - a. Right-click the **Street-Level Geocoding** node and select **Preview**.  
A sample of records appears on the Preview tab of the Details pane.
  - b. Scroll to the right to view the geocode information.

Result_Code	Geocode_Result_Code	Geocode_Latitude	Geocode_Longitude	Geocode_FIPS
0	DP	33.468273	-86.736395	01073
0	DP	33.506826	-86.737779	01073
0	DP	33.417105	-86.767985	01073
0	DP	33.540252	-111.938414	04013
0	DP	33.706838	-111.909543	04013
0	DP	33.281572	-111.702949	04013
0	DP	32.246592	-110.940993	04019
0	DP	33.824599	-111.862584	04013
0	DP	33.750377	-117.792707	06059
0	DP	34.072775	-118.547787	06037
0	DP	36.924139	-121.741623	06087
0	DP	37.801000	-122.421100	06075
0	DP	33.816317	-84.385207	13121
0	DP	33.957391	-84.353285	13121
0	DP	32.415009	-81.726688	13031
0	DP	39.572430	-104.981794	08005
0	DP	39.555432	-104.952603	08035
14	ZIP	39.642179	-105.004305	08005
0	DP	39.776228	-105.048593	08031

Notice the following:

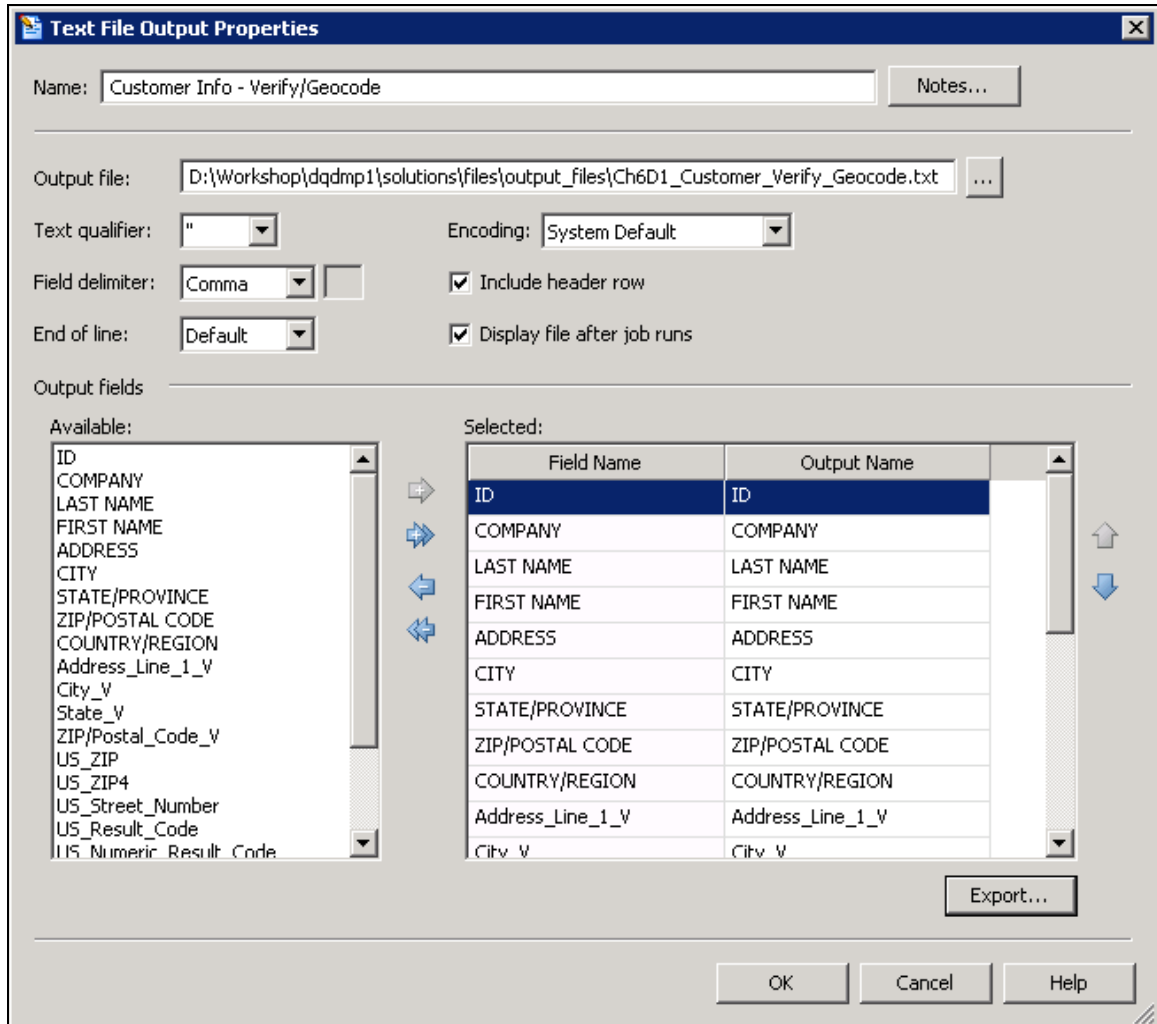
- Most records in the preview have a Geocode\_Result\_Code of **DP (Delivery Point)**.
- Records where the verified ZIP code has only five digits have a Geocode\_Result\_Code of **ZIP**. These records have less precise values for the Geocode\_Latitude and Geocode\_Longitude columns.

**Note:** When you use the Street-Level Geocoding node, a license check is performed to confirm the license type. If you have a ZIP+4 database, you see only results reflecting that license type. If you have a database that includes street-level data, you see street-level results.

The field definitions are shown below:

Available Field	Description
Geocode Result Code	<p>The result code indicates whether the record was successfully geocoded. Other possible codes are as follows:</p> <ul style="list-style-type: none"> <li>• <b>DP</b> – The match is based on the delivery point.</li> <li>• <b>PLUS4</b> – The match failed on the delivery point, so the match is based on ZIP+4.</li> <li>• <b>ZIP</b> – The ZIP+4 match failed, so the match is based on the ZIP code.</li> <li>• <b>NOMATCH</b> – The first three checks failed. There is no match in the geocoding database.</li> </ul>
Geocode Latitude	The numerical horizontal map reference for address data.
Geocode Longitude	The numerical vertical map reference for address data.
Geocode Country Code	The country code for the address.
Geocode FIPS	The U.S. Federal Information Processing Standards (FIPS) number used by the U.S. Census Bureau to refer to geographical areas.
Geocode Census Tract	A US Census Bureau reference number is assigned using the centroid latitude and longitude. This number contains references to the State and County codes.
Geocode Census Block	The last four digits of the State/County/Tract/Block value.
Geocode FIPS MCD Code	The FIPS Minor Civil Division (MCD) code refers to a subsection of the US Census county subdivision statistics. This number includes census data for county divisions, census subareas, minor civil divisions, unorganized territories, and incorporated areas.
Geocode MSA or CMSA	A US Census Bureau code referencing Metropolitan Statistical Areas (MSA) or County Metropolitan Statistical Areas (CMSA).
Geocode PMSA (for CMSA)	A US Census Bureau code referencing Principal Metropolitan Statistical Areas (PMSA) or Consolidated Metropolitan Statistical Areas (CMSA).
Geocode CBSA	The code for a Core-Based Statistical Area.

9. Investigate the properties for the Text File Output node.
  - a. Double-click the **Customer Info - Verify/Geocode** Text File Output node.
  - b. Verify the Text File Output Properties window resembles the following:



- c. Click **Cancel** to close the Text File Output Properties window.

10. Run the job.
  - a. Verify that the text file appears.

```

Ch6D1_Customer_Verify_Geocode.txt - Notepad
File Edit Format View Help
"ID","COMPANY","LAST NAME","FIRST NAME","ADDRESS","CITY","STATE/PRO
1,"Transamerica Financial Group","Bedecs","Anna","3720 Wimbledon Ln","Birmingham","AL
2,"DataFlux","Ramos","Antonio","4013 Winston Way","Birmingham","AL
3,"Transamerica Occidental","Axen","Thomas","2437 Mountain Vista","Birmingham","AL
4,"Transamerica Financial Svcs","Lee","Christina","6629 East Meadow","Birmingham","AL

```

- b. Select **File** ⇒ **Exit** to close the Notepad window.

11. View the detailed log.

- a. Click the **Log** tab.
  - b. Review the information for each of the nodes.

Row	Node Name	Node ID	Node Type	Status	Run Time
0	Ch6D1_AddressVerification		Data Job	Completed successfully	00:00.678
1	Customers Table	1	Data Source	DSN: DSN=dfConglomerate Gifts;DF SQL: SELECT "ID","COMPANY","LAST 63 rows read	00:00.678
2	Address Verification	2	Address Verification (US/Canada)	DPV disabled RDI disabled eLOT disabled LACS disabled CASS compliant mode disabled Proper casing results enabled SuiteLink disabled Street address abbreviation enabled City name abbreviation enabled Skip_null_csz_line disabled	00:00.678
3	Geocoding using Verified ;	5	Street-Level Geocoding	63 row(s) read 63 row(s) geocoded	00:00.678
4	Customer Info - Verify/Ge	6	Text File Output	Wrote 63 rows to text file D:\Worksh	00:00.678

12. Close the data job.
  - a. Click the **Data Flow** tab.
  - b. Select **File** ⇒ **Close**.

**End of Demonstration**



## Exercise

### 1. Performing Address Verification on the MANUFACTURERS Table

Create a data job that uses the Address Verification node. The final job flow should resemble the following display:



- Create a new data job named **Ch6E1\_Manufacturers\_Verify** in the **batch\_jobs** folder of the **Basics Exercises** repository.
- Add the **MANUFACTURERS** table from the dfConglomerate Grocery data connection as the data source.
- Verify the address fields. Use the following specifications:
  - Map the following input fields to the specified field type:

Field Name	Field Type
<b>MANUFACTURER</b>	Firm
<b>STREET_ADDR</b>	Address Line 1
<b>CITY</b>	City
<b>STATE_PROV</b>	State
<b>POSTAL_CD</b>	Zip

- Specify the following options:
  - ✓ **Proper case results**
  - ✓ **Output blanks as nulls**
  - ✓ **Street abbreviation**
  - ✓ **City abbreviation**
  - ✓ **Insert dash between ZIP when ZIP4**

**Note:** CASS compliance should *not* be selected.

- Specify the following output fields:

Output Type	Output Name	Output Type	Output Name
Firm	Firm_V	ZIP/Postal Code	ZIP_V
Address Line 1	Address_V	US County Name	US_County_Name
City	City_V	US Result Code	US_Result_Code
State	State_V		

- Retain only the following list of original fields as additional outputs:

**ID**                                **STATE/PROV**  
**MANUFACTURER**                **POSTAL\_CD**  
**STREET\_ADDR**                    **COUNTRY**  
**CITY**                                **PHONE**

- Add a Text File Output node to the job flow. Use the following specifications:

**Output file:**            **Ch6E1\_Manufacturers\_Verify.txt** in the directory  
**D:\Workshop\ldqtmp1\Exercises\files\output\_files**

**Text qualifier:**    " (double quotation mark)

**Field delimiter:**   **Comma**

✓ **Include header row**

✓ **Display file after job runs**

Specify the following fields as output for the text file with the specified output names:

Field Name	Output Name
<b>ID</b>	ID
<b>Firm_V</b>	Manufacturer
<b>Address_V</b>	Address
<b>City_V</b>	City
<b>State_V</b>	State
<b>ZIP_V</b>	ZIP
<b>US_County_Name</b>	US_County_Name
<b>US_Result_Code</b>	US_Result_Code

- Save and run the job.
- Verify that the text file contains the verified address information.

## Section 3: Solutions

---

### Solutions to Exercises

#### 1. Performing Address Verification on the MANUFACTURERS Table

- a. If necessary, invoke Data Management Studio.
  - 1) Select **Start** ⇒ **All Programs** ⇒ **DataFlux** ⇒ **Data Management Studio 2.7**.
  - 2) Click **Cancel** in the Log On window.
- b. Create a new data job.
  - 1) Verify that the **Home** tab is selected.
  - 2) Click the **Folders** riser bar.
  - 3) Expand the **Basics Exercises** repository.
  - 4) Right-click the **batch\_jobs** folder and select **New** ⇒ **Data Job**.
    - a) Enter **Ch6E1\_Manufacturers\_Verify** in the **Name** field.
    - b) Click **OK**. The new data job appears on a tab.
- c. Add a Data Source node to the job flow.
  - 1) Verify that the **Nodes** riser bar is selected in the Resource pane.
  - 2) Expand the **Data Inputs** grouping of nodes.
  - 3) Double-click the **Data Source** node. The node is added to the job flow, and the properties window for the node appears.
- d. Specify properties for the Data Source node.
  - 1) Enter **Manufacturers Table** in the **Name** field.
  - 2) Click  next to the **Input table** field.
    - a) Expand the **dfConglomerate Grocery** data connection.
    - b) Click the **MANUFACTURERS** table.
    - c) Click **OK** to close the Select Table window.
  - 3) Click **OK** to close the Data Source Properties window.
- e. Edit the basic settings for the Data Source node and preview the data.
  - 1) Verify that the Details pane is displayed.
  - 2) If necessary, select the **Data Source** node in the job diagram.
  - 3) If necessary, click the **Basic Settings** tab.

- 4) Enter **Data Source** in the **Description** field.
- 5) Right-click the **Data Source** node and select **Preview**.

A sample of the records appears on the Preview tab of the Details pane.

ID	MANUFACTURER	STREET_ADDR	CITY	STATE_PROV	POSTAL_CD	COUNTRY
1	Dr. McDougall's Right Foods	101 Utah Ave. So.	San Francisco	CA	94080	USA
2	Certified Grocers of California, LTD.	(null)	Los Angeles	CA	(null)	(null)
3	Lifestream Natural Foods	2220 Nature's Path Way	Blaine	WA	98230	USA
4	American Procurement & Logistics Company	P.O. Box 27447	Salt Lake City	UT	84127-0447	USA
5	Kraft Foods North America Inc.	(null)	East Hanover	NJ	07936	US
6	Sunfresh Inc.	P.O. Box 861	West Seneca	New York	14224	USA
7	Kellogg USA Inc.	P.O. Box CAMB	Battle Creek	MI	49016-1986	USA
8	Kellogg's	P.O. Box CAMB	Battle Creek	Mich	49016-1986	USA
9	Albertson's Inc.	(null)	Boise	ID	83726	US

- f. Add an Address Verification (US/Canada) node to the job flow.
  - 1) Verify that the **Nodes** riser bar is selected in the Resource pane.
  - 2) Collapse the **Data Inputs** grouping of nodes.
  - 3) Expand the **Enrichment** grouping of nodes.
  - 4) Double-click the **Address Verification (US/Canada)** node.

The node is added to the job flow, and the properties window for the node appears.

- g. Specify properties for the Address Verification (US/Canada) node.
  - 1) Enter **Verify US Addresses** in the **Name** field.
  - 2) Verify that **United States** is selected in the Address area.
  - 3) Specify input information.
    - a) Click  under **Field Type** for the **MANUFACTURER** field and select **Firm**.
    - b) Click  under **Field Type** for the **STREET\_ADDR** field and select **Address Line 1**.
    - c) Click  under **Field Type** for the **CITY** field and select **City**.
    - d) Click  under **Field Type** for the **STATE\_PROV** field and select **State**.
    - e) Click  under **Field Type** for the **POSTAL\_CD** field and select **Zip**.



The Input area should resemble the following:

Field Name	Field Type
ID	
MANUFACTURER	Firm
STREET_ADDR	Address Line 1
CITY	City
STATE_PROV	State
POSTAL_CD	Zip
COUNTRY	

- 4) Specify options for address verification.
  - a) Click **Options**.
  - b) Click **Proper case results**.
  - c) Click **Output blanks as nulls**.
  - d) Click **Street abbreviation**.
  - e) Click **City abbreviation**.
  - f) Clear **CASS compliance**.
  - g) Verify that **Insert dash between ZIP and ZIP4** is selected.

- h) Click **OK** to close the Options window.


- 5) Specify the output information.
- In the Output fields area, double-click the following fields to move them to the Selected list:
 

<b>Firm</b>	<b>ZIP/Postal Code</b>
<b>Address Line 1</b>	<b>US County Name</b>
<b>City</b>	<b>US Result Code</b>
<b>State</b>	
  - Rename the fields below. Click the **Output Name** cell, enter the new name, and press Enter.

Output Type	Output Name
Firm	Firm_V
Address Line 1	Address_V
City	City_V
State	State_V
ZIP/Postal Code	ZIP_V

The Selected list should resemble the following:

Selected:	
Output Type	Output Name
Firm	Firm_V
Address Line 1	Address_V
City	City_V
State	State_V
ZIP/Postal Code	ZIP_V
US County Name	US_County_Name
US Result Code	US_Result_Code

- Retain only some of the original fields as additional outputs.
  - Click **Additional Outputs**.
  - Click  to remove all the fields from the Output fields list.
  - Double-click the following fields to move them from the Available list to the Output list:
 

<b>ID</b>	<b>STATE/PROV</b>
<b>MANUFACTURER</b>	<b>POSTAL_CD</b>
<b>STREET_ADDR</b>	<b>COUNTRY</b>
<b>CITY</b>	<b>PHONE</b>
  - Click **OK** to close the Additional Outputs window.

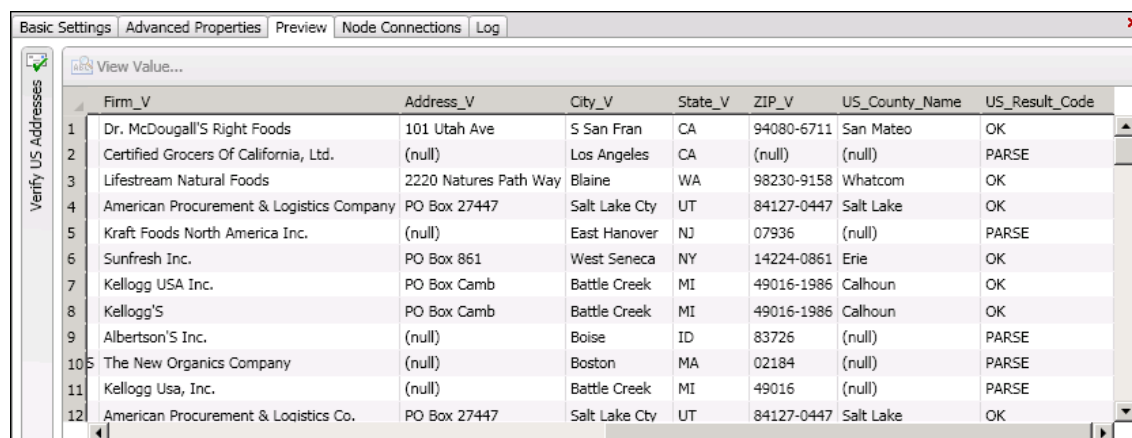
7) Click **OK** to close the Address Verification (US/Canada) Properties window.

h. Preview the data.

1) Right-click the **Address Verification (US/Canada)** node and select **Preview**.

A sample of the records appears on the Preview tab of the Details pane.

2) Scroll to the right to view the verified address information.



	Firm_V	Address_V	City_V	State_V	ZIP_V	US_County_Name	US_Result_Code
1	Dr. McDougall'S Right Foods	101 Utah Ave	S San Fran	CA	94080-6711	San Mateo	OK
2	Certified Grocers Of California, Ltd.	(null)	Los Angeles	CA	(null)	(null)	PARSE
3	Lifestream Natural Foods	2220 Natures Path Way	Blaine	WA	98230-9158	Whatcom	OK
4	American Procurement & Logistics Company	PO Box 27447	Salt Lake Cty	UT	84127-0447	Salt Lake	OK
5	Kraft Foods North America Inc.	(null)	East Hanover	NJ	07936	(null)	PARSE
6	Sunfresh Inc.	PO Box 861	West Seneca	NY	14224-0861	Erie	OK
7	Kellogg USA Inc.	PO Box Camb	Battle Creek	MI	49016-1986	Calhoun	OK
8	Kellogg'S	PO Box Camb	Battle Creek	MI	49016-1986	Calhoun	OK
9	Albertson'S Inc.	(null)	Boise	ID	83726	(null)	PARSE
10	The New Organics Company	(null)	Boston	MA	02184	(null)	PARSE
11	Kellogg Usa, Inc.	(null)	Battle Creek	MI	49016	(null)	PARSE
12	American Procurement & Logistics Co.	PO Box 27447	Salt Lake Cty	UT	84127-0447	Salt Lake	OK

i. Add a Text File Output node to the job flow.

1) Verify that the **Nodes** riser bar is selected in the Resource pane.

2) Collapse the **Enrichment** grouping of nodes.

3) Expand the **Data Outputs** grouping of nodes.

4) Double-click the **Text File Output** node.

The node is added to the job flow, and the properties window for the node appears.

j. Specify properties for the Text File Output node.

1) Enter **Manufacturers Info - Verify** in the **Name** field.

2) Specify the output file information.

a) Click  next to the **Output file** field.

b) Navigate to **D:\Workshop\ldqmp1\Exercises\files\output\_files**.

c) Enter **Ch6E1\_Manufacturers\_Verify.txt** in the **File name** field.

d) Click **Save**.

3) Specify attributes for the file.


a) Specify " (double quotation mark) as the text qualifier.

b) Verify that the **Field delimiter** field is set to **Comma**.

c) Click **Include header row**.

d) Click **Display file after job runs**.

4) Specify desired fields as output for the text file.

a) In the Output fields area, click  to remove all fields from the Selected list.

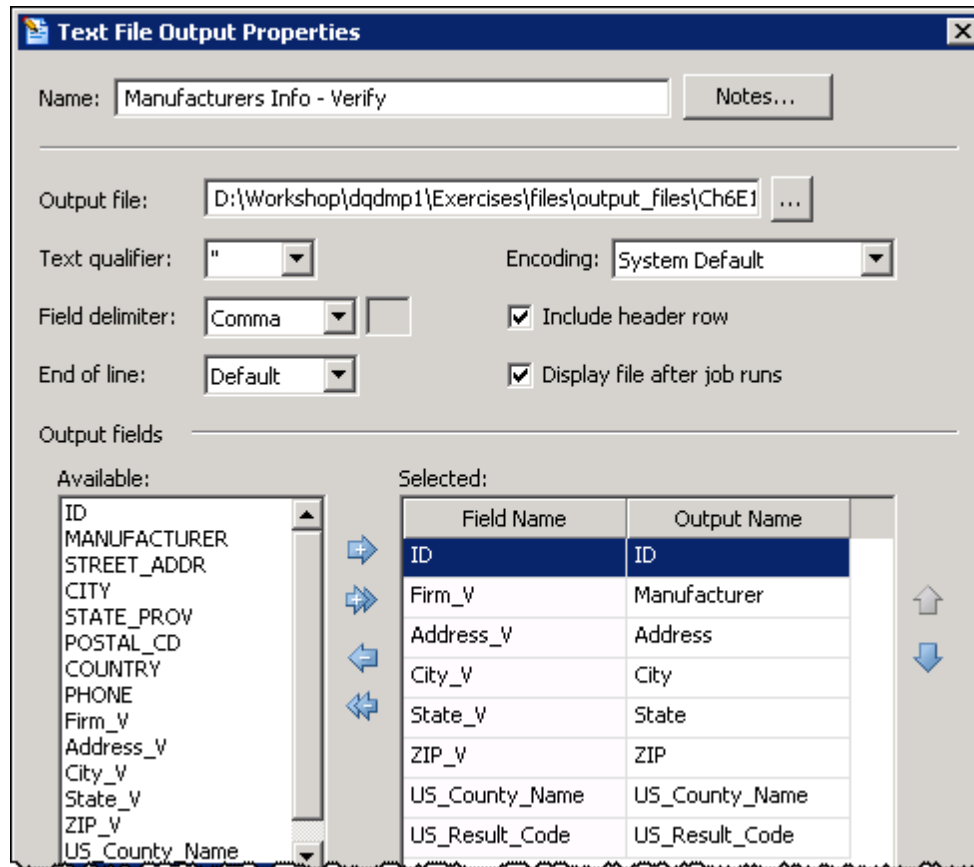
b) Double-click the following fields to move them from the Available list to the Selected list:

<b>ID</b>	<b>State_V</b>
<b>Firm_V</b>	<b>ZIP_V</b>
<b>Address_V</b>	<b>US_County_Name</b>
<b>City_V</b>	<b>US_Result_Code</b>

c) Rename the following fields.

Field Name	Output Name
<b>Firm_V</b>	Manufacturer
<b>Address_V</b>	Address
<b>City_V</b>	City
<b>State_V</b>	State
<b>ZIP_V</b>	ZIP

The Text File Output Properties should resemble the following display:



5) Click **OK** to close the Text File Output Properties window.

The final job flow should resemble the following:



- k. Save and run the data job.
  - 1) Select **File** ⇒ **Save** to save the data job.
  - 2) Verify that the **Data Flow** tab is selected.
  - 3) Select **Actions** ⇒ **Run Data Job**.
- l. Verify that the text file contains the verified address information.

```

Ch6E1_Manufacturers_Verify.txt - Notepad
File Edit Format View Help
|ID", "Manufacturer", "Address", "City", "State", "ZIP", "US_County_Name", "US_Result_Code"
1, "Dr. McDougall's Right Foods", "101 Utah Ave", "S San Fran", "CA", "94080-6711", "San M
2, "Certified Grocers Of California, Ltd.", "Los Angeles", "CA", "", "PARSE"
3, "Lifestream Natural Foods", "2220 Natures Path way", "Blaine", "WA", "98230-9158", "whs
4, "American Procurement & Logistics Company", "PO Box 27447", "Salt Lake Cty", "UT", "84
5, "Kraft Foods North America Inc.", "East Hanover", "NJ", "07936", "PARSE"
6, "Sunfresh Inc.", "PO Box 861", "West Seneca", "NY", "14224-0861", "Erie", "OK"
7, "Kellogg USA Inc.", "PO Box Camb", "Battle Creek", "MI", "49016-1986", "Calhoun", "OK"
8, "Kellogg's", "PO Box Camb", "Battle Creek", "MI", "49016-1986", "Calhoun", "OK"
9, "Albertson's Inc.", "Boise", "ID", "83726", "PARSE"
10, "The New Organics Company", "Boston", "MA", "02184", "PARSE"
11, "Kellogg USA, Inc.", "Battle Creek", "MI", "49016", "PARSE"
12, "American Procurement & Logistics Co.", "PO Box 27447", "Salt Lake Cty", "UT", "8412
13, "Hodgson Mill, Inc., Cbm", "Gainesville", "MO", "65655", "PARSE"
14, "Lowe's Foods, Inc.", "PO Box 24908", "Winston Salem", "NC", "27114-4908", "Forsyth", "C
15, "Meijer, Inc.", "Grand Rapids", "MI", "49544", "PARSE"
16, "Kraft Foods, Inc.", "800 Westchester Ave", "Rye Brook", "NY", "10573-1354", "Westches
17, "Kraft Foods, Inc.", "PO Box HNSW-20", "Tarrytown", "NY", "10591", "NOMATCH"
18, "Kellogg Inc.", "Battle Creek", "MI", "49016", "PARSE"
19, "Tops Markets, Inc.", "6363 Main St", "Williamsville", "NY", "14221-5855", "Erie", "OK"

```

- m. Select **File** ⇒ **Exit** to close the Notepad window.
- n. Select **File** ⇒ **Close** to close the data job.

**End of Solutions**

## Solutions to Activities and Questions

### x.01 Poll – Correct Answer

Address Verification and Enrichment reference sources exist only for the United States and Canada.

- True
- False

The Loqate reference data source contains worldwide address verification and enrichment data.

7

Copyright © SAS Institute Inc. All rights reserved.



### x.02 Poll – Correct Answer

The address verification and enrichment features for a data job require the use of third-party reference data.

- True
- False

The address verification and enrichment nodes in a data job do require a third-party reference data source to look up valid address information.

18

Copyright © SAS Institute Inc. All rights reserved.

