

Lookup Wizard

Contents:

- *Lookup Wizard - Step 1*
 - *Lookup Wizard - Step 2*
 - *Column Cleanup*
 - *Auto-updating Lookups*
-

You can perform lookups from one set of values in your dataset into another set of values in another dataset. A **lookup** compares each value in the selected column against the values in a selected column of the target dataset. Where a match is found, the values in other columns of the target dataset are inserted as new columns in the dataset from which the lookup was executed.

For example, your enterprise is changing the names of all of your products. Instead of performing a complex set of replace transforms, you can perform a lookup from your `productName` column into a two-column dataset, which contains the original name and the new name in separate columns. When the new name is inserted into your source dataset via lookup, you can delete the source column and continue transforming your data with the new names.

- You cannot perform lookups on columns of Object or Array data type.
- A lookup essentially performs a left join between the first dataset and the second one. However, lookups are less flexible in terms of defining and editing them.

NOTE: If column values are non-unique, the resulting dataset can be significantly larger than the original dataset.

This workflow is best demonstrated by example. In this case, your raw sales data records product information in internal numeric identifiers. For analysis, you may want to integrate data from your products master data based on the internal identifier, so that you have a product description and other useful information as part of your dataset.

To perform a lookup, select the caret next to a column title, and then select **Lookup....**

Lookup Wizard - Step 1

In the first step, you select the dataset against which you would like to perform your lookup for matching data for the `Item_Nbr` column. In this example, the products dataset is selected, since it contains the list of recognized products:

Tip: You can search your available flows and datasets. When you search for flows, all datasets in the flow are matched.

Step 1 of 2 Select data to Lookup ?



Search...

1 2

All (29) Imported (27) Reference (2) Recipe (0)

NAME	SOURCE	LAST UPDATED
POS-r02.txt	HDFS	Today at 10:17 AM
POS-r03.txt	HDFS	Today at 10:17 AM
<input checked="" type="checkbox"/> REF_PROD.txt	HDFS	Today at 10:17 AM
REF_CAL.txt	HDFS	Today at 10:17 AM
USDA Farmers Market 2014	HDFS	Today at 10:15 AM
BOH August.csv	HDFS	Today at 9:49 AM

Cancel Select

Figure: Lookup Wizard - Step 1

Lookup Wizard - Step 2

After you select the dataset against which to perform the lookup, you select the field in the target dataset to use as the lookup key. The **lookup key** provides the set of identifiers for which you are trying to find a match for each value in the source column. In this case, the lookup key column has the same name as the source column: ITEM_NBR.

Step 2 of 2 Select Lookup Key

ITEM_NBR

< Back Cancel Execute Lookup

Figure: Lookup Wizard - Step 2

Column Cleanup

When the lookup is executed, for each value in the source `item_nbr` column that can be found in the target dataset's `ITEM_NBR` column, all of the other columns in the corresponding row of the second dataset are inserted as separate columns in the first dataset. These columns are inserted to the immediate right of the column that was used for the lookup:

#	Store_Nbr	#	Item_Nbr	RBC	PRODUCT DESC	#	CAT_CD	#	GRP_CD	#	CLASS_CD	#	FL_CLASS_CD
1	1	381000	ACME LAWN GARDEN BAG CLEAR	07	75	05							
2	2	325000	ACME COOKIES CHOC CHIP	81	10	25							
2	2	325000	ACME COOKIES CHOC CHIP	81	10	25							
2	2	403000	ACME SANDWICH BAG	07	70	05							
2	2	449000	ACME SODAS SALTED	81	30	15							
2	2	490000	ACME SCENTED OIL REFILL-CTRY SUN	07	65	20							
2	2	560000	ACME LARGE FUDGE GRAHAMS COOKIES	81	10	25							
2	2	573000	ACME SUGAR ICE WAFERS VANILLA	81	10	25							
3	3	486000	ACME ZOO ANIMAL FRUIT SNACKS 6'S	81	70	30							
3	3	488000	ACME WAFERS SUGER ICE	81	10	25							
3	3	490000	ACME SCENTED OIL REFILL-CTRY SUN	07	65	20							
3	3	490000	ACME RICE CRACKERS ONION	81	20	30							
3	3	503000	ACME GARBAGE BAG BLACK	07	75	15							
3	3	530000	ACME FUDGE DIP CHOC CHIP COOKIE	81	10	25							
3	3	560000	ACME LARGE FUDGE GRAHAMS COOKIES	81	10	25							
3	3	573000	ACME SUGAR ICE WAFERS VANILLA	81	10	25							
4	4	325000	ACME COOKIES CHOC CHIP	81	10	25							
4	4	325000	ACME COOKIES CHOC CHIP	81	10	25							
4	4	326000	ACME DIGESTIVE RICH TEA BISCUITS	81	10	25							
4	4	327000	ACME ASSORTED COOKIES DRP	81	10	25							
4	4	328000	ACME KITCHEN BAG	07	75	10							

Figure: Lookup Wizard - Results

NOTE: If the second dataset contains multiple matching entries for individual lookup key values from the first dataset, rows from the first dataset are duplicated in the results.

NOTE: You may need to delete some of the columns that have been imported into your dataset.

Auto-updating Lookups

After you have added a lookup to your recipe, subsequent changes to that reference data are automatically reflected in the dataset.

Tip: If you must freeze the data in the dataset that you are using for a lookup, you should create a copy of the dataset as a snapshot. See *Dataset Details Page*.

To use the copy, delete the lookup and rebuild it using the copied version. See *Fix Dependency Issues*.