Lab 3: Datasets

Note: Do not include any personal, identifying, or confidential information into the lab environment. Information entered may be visible to others.

Corrections, feedback, or other questions? Contact us at *AWS Training and Certification*.

Lab overview

You have just enhanced your first Amazon QuickSight dashboard and added interactivity. AnyCompany Software has many data sources, and you are investigating how to incorporate additional information from these sources into the sales dashboard.

In this lab, you pull in data from a new source and prepare the data for visualization.

OBJECTIVES

By the end of this lab, you should be able to do the following:

- Join two data sources together using Amazon QuickSight joins.
- Filter and transform the data during import.
- Create a field folder to consolidate related data points.
- Create a geospatial hierarchy.

DURATION

This lab requires approximately *30* minutes to complete.

ICON KEY

Various icons are used throughout this lab to call attention to different types of instructions and notes. The following list explains the purpose for each icon:

- **Expected output:** A sample output that you can use to verify the output of a command or edited file.
- **Note:** A hint, tip, or important guidance.
- **Learn more:** Where to find more information.
- **Consider:** A moment to pause to consider how you might apply a concept in your own environment or to initiate a conversation about the topic at hand.
- **Hint:** A hint to a question or challenge.
- **Answer:** An answer to a question or challenge.
- **Task complete:** A conclusion or summary point in the lab.

Start lab

1. To launch the lab, at the top of the page, choose Start lab.

Caution: You must wait for the provisioned AWS services to be ready before you can continue.

2. To open the lab, choose Open Console.

You are automatically signed in to the AWS Management Console in a new web browser tab.

WARNING: Do not change the Region unless instructed.

COMMON SIGN-IN ERRORS

Error: You must first sign out

If you see the message, **You must first log out before logging into a different AWS account:**

- Choose the **click here** link.
- Close your **Amazon Web Services Sign In** web browser tab and return to your initial lab page.
- Choose Open Console again.

Error: Choosing Start Lab has no effect

In some cases, certain pop-up or script blocker web browser extensions might prevent the **Start Lab** button from working as intended. If you experience an issue starting the lab:

- Add the lab domain name to your pop-up or script blocker's allow list or turn it off.
- Refresh the page and try again.

AWS SERVICES NOT USED IN THIS LAB

AWS service capabilities used in this lab are limited to what the lab requires. Expect errors when accessing other services or performing actions beyond those provided in this lab guide.

Task 1: Prepare a dataset for a join

Your team would like to add fiscal calendar information to the Amazon QuickSight dashboard. You choose to join a new fiscal calendar dataset to the current data.

In this task, you review datasets that your team added. You duplicate the data and change the data types before joining the two tables.

TASK 1.1: DUPLICATE A NEW DATASET

Review the datasets that were added through Amazon Simple Storage Service (Amazon S3). Then, retain an original copy of the imported sales data by duplicating your sales data.

3. At the top of the **AWS Management Console**, in the search bar, search for and choose

   QuickSight.

4. In the left navigation pane, choose **Datasets**.

Note: You should see at least two datasets. The *lab-3-quicksight-data-set* is your original dataset with the sales data. For date information, the sales dataset includes the *Order Date* and *Date Key* fields.

Note: The *lab-3-quicksight-fiscal-data-set* is a new dataset that your team added to QuickSight through Amazon S3. The fiscal dataset includes the *Date Key*, *Date*, *Day of Week*, *Fiscal Year*, and *Fiscal Quarter* fields. When joined with the sales dataset, you can organize orders by fiscal year and fiscal quarter. These fields help track seasonality.

5. From the **Datasets** list, choose **lab-3-quicksight-data-set**.
6. Choose the **EDIT DATASET** dropdown list, then choose **DUPLICATE DATASET**.
7. For the name of the duplicated dataset, enter

   lab-3-quicksight-data-set-duplicate.

8. Choose **DUPLICATE**.

Expected output: You now have a duplicate dataset. If you make changes to this duplicate, it will not impact the original data, giving you a way to retain a stable, original copy of the data while you work on joining a new data source.

TASK 1.2: CHANGE THE DATA TYPES IN THE NEW DATASET

Edit the data types for the *Country*, *City*, and *Region* fields to test out QuickSight's data type functionality.

9. Choose **EDIT DATASET**.
10. In the **Fields** pane, choose the ellipsis icon next to **Country**.

Consider: What is the *Country* data type currently set to?

11. Hover over **Change data type**, and then choose **String**.
12. Repeat the same steps to change the **City** field data type to **String**.
13. Repeat the same steps to change the **Region** field data type to **String**.
14. In the top menu bar, choose **SAVE & PUBLISH**.

- **Expected output:** Your geospatial data types are all changed to strings. Notice that there is no immediate change in the sample table. Changing the data type from geospatial to string has not changed the underlying data.

Note: Later in the lab, you'll adjust the *City* and *Country* fields back to geospatial fields and add them to a geospatial hierarchy. Leave the fields as string data types for now.

Task complete: You have completed **Task 1** by preparing your dataset for a join. You duplicated the original dataset and edited the data types.

Task 2: Join two tables together

Join together two data sources on the *Date Key* field.

In this task, you join together two tables using a primary key.

TASK 2.1: ADD A NEW DATA SOURCE

Add a new data source to the workspace.

15. Choose **Add data**.
16. From the dropdown list, choose **Data source**.
17. From the data source list, choose **lab-3-quicksight-fiscal-data-source**, and then choose **Select**.
18. In the **Table** section, choose **lab-3-quicksight-fiscal-data-source**, and then choose **Select**.

Expected output: Both tables appear in the data workspace.

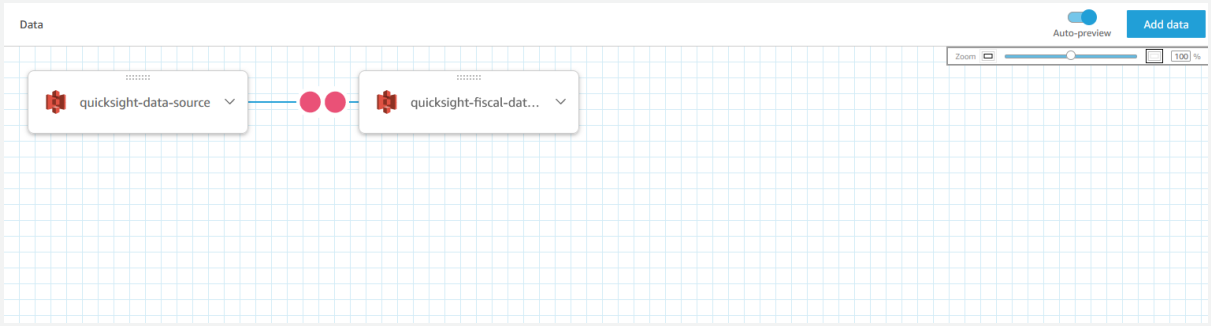

Image description: The previous image shows the two tables being joined in the workspace.

TASK 2.2: JOIN THE TABLES

Join the tables together using a left join.

19. Choose the two red dots between the data sources to configure the join.
20. Under **Join configuration**, in the **lab-3-quicksight-data-set** section, choose the **Date Key** field from the dropdown list to set the first primary key.
21. Under **Join configuration**, in the **lab-3-quicksight-fiscal-data-source** section, choose the **Date Key** field from the dropdown list to set the second primary key.

Note: For this join, you can leave the join type as *Left*.

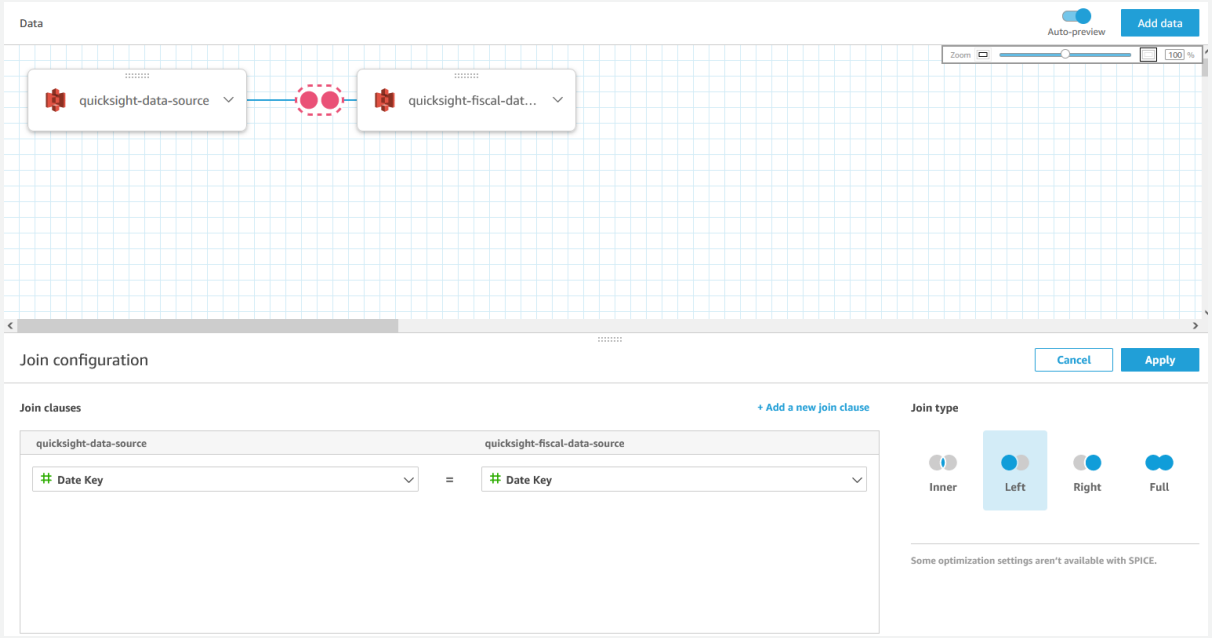

Image description: The previous image shows the join configuration menu.

Learn more: QuickSight provides many ways to join your data, and it will automatically recommend the optimum join type for your datasets. For more information about joining data in QuickSight, see Joining data in the *Amazon QuickSight User Guide*.

22. To finalize the join, choose **Apply**.
23. In the top menu bar, choose **SAVE & PUBLISH**.

Consider: What are some ways that you can use QuickSight joins to pull together tables from many sources that are ready for analysis? Would any other data sources augment your sales dashboard?

Task complete: You have completed **Task 2** by joining two tables together. You added another data source to the workspace and then joined the two tables using a left join.

Task 3: Filter and transform the ingested data

Your team wants customized sales calculations. The team would like data to be cleaned up before they add more visualizations to a new sheet on the dashboard. To meet this request, you apply preliminary filters to the dataset and create custom calculations as the data is ingested.

In this task, you create a filter and a calculated field, exclude fields from the dataset, and create a child dataset.

TASK 3.1: ADD A FILTER

To exclude any industry that is miscellaneous, you add a filter to exclude all of the rows that have an *Industry* value of *Misc*.

24. On the left side of the workspace, choose **Filters** to expand the **Filters** pane.
25. Choose **ADD FILTER**.
26. In the **Search value** box, enter

    ```
    Industry
    ```
    .
27. Choose **Industry**.
28. Choose **Industry** to expand it.
29. Choose **Include**, and choose **Exclude** instead.
30. In the **List** section, enter

    ```
    Misc
    ```
    .
31. To save the filter, choose **Apply**.

Expected output: The *Misc* value is now filtered out of the *Industry* field. All of the rows with an industry of *Misc* do not show up in the dashboard.

TASK 3.2: ADD A CALCULATED FIELD

Add a cost field to the dataset. You calculate cost by subtracting profit from sales.

32. Choose **Fields** to expand the **Fields** pane.
33. Choose the plus **+** icon and choose **Add calculated field**.
34. To change the field name, for **Add name**, enter

    ```
    Cost
    ```
    .
35. In the calculation workspace, enter

    ```
    {Sales} - {Profit}
    ```
    .
36. To create the **Cost** field, choose **Save**.

Expected output: After QuickSight validates the calculation syntax, the data workspace will reappear. Look through your *Dataset* at the bottom of the workspace. You can find your new *Cost* field at the end of the table.

TASK 3.3: EXCLUDE FIELDS FROM THE DATASET

You only need to retain the Order Date, Day of Week, Fiscal Year, and Fiscal Quarter in the dataset. You choose to exclude unnecessary fields from the dataset, retaining the relevant date fields for analysis while keeping the dataset clean.

37. In the **Fields** pane, in the **Select** section, choose **Date**, **Date Key**, and **Date Key[quicksight-fiscal-data-source]**.
38. Choose the ellipsis icon next to any of the fields you selected, and then choose **Exclude fields**.
39. In the top menu bar, choose **SAVE & PUBLISH**.

Expected output: The fields *Date*, *Date Key*, and *Date Key[quicksight-fiscal-data-set]* are now excluded from the dataset.

TASK 3.4: CREATE A CHILD DATASET

In QuickSight, you can create a child dataset from a parent dataset. In a child dataset, you can join and filter new data separate from changes to the parent dataset. You can adjust the data refresh schedule based on your team's needs.

40. To return to the main QuickSight menu, choose the **QuickSight** icon at the top of the page.
41. In the left navigation pane, choose **Datasets**.
42. From the **Datasets** list, choose **lab-3-quicksight-data-set-duplicate**.
43. Choose the **USE IN ANALYSIS** dropdown list, then choose **USE IN DATASET**.
44. In the top menu bar, for **Dataset Name**, enter

    lab-3-quicksight-data-set-child

    .
45. Choose **SAVE & PUBLISH**.

Expected output: You now have a child dataset. Changes to this dataset do not impact the parent. However, any changes made to the parent will update the child dataset.

Learn more: For more information about parent, child, and grandchild datasets, see Creating a dataset using an existing dataset in Amazon QuickSight in the *Amazon QuickSight User Guide*.

Task complete: You have completed **Task 3** by filtering and transforming the ingested data. You added a filter, created a calculated field, excluded fields from the dataset, and created a child dataset.

Task 4: Group fields with a field folder

To help with future dashboard development, you choose to group similar fields together. In QuickSight, you can organize fields in field folders.

In this task, you group similar fields in custom field folders.

TASK 4.1: ADD ALL OF THE DATE FIELDS TO A FIELD FOLDER

Add all of the fields that include date information to a field folder called *Date*.

46. In the **Fields** pane, choose **Order Date**, **Fiscal Year**, **Fiscal Quarter**, and **Day Of Week**.
47. Choose the ellipsis icon next to any of the fields you selected, and then choose **Add to folder**.
48. In the **Create a new folder** section, enter

    Date

    .
49. Choose **Apply**.

Expected output: All of the date fields are now grouped together in a *Date* field folder in the **Fields** pane.

Consider: Take a moment to consider what other fields you could group together. Would more field folders help organize fields for future analysis in the dashboard?

TASK 4.2: ADD ALL OF THE SALES METRICS TO A METRICS FOLDER

Add all of the fields that are sales metrics to a field folder called *Metrics* using the same process as the *Date* folder.

50. In the **Fields** pane, choose **Sales**, **Quantity**, **Discount**, **Profit** and **Cost**.
51. Choose the ellipsis icon next to any of the fields you selected, and then choose **Add to folder**.
52. In the **Create a new folder** section, enter

    Metrics

    .
53. Choose **Apply**.

Expected output: All of the metrics are now grouped together in a *Metrics* field folder in the **Fields** pane.

TASK 4.3: CREATE A GEOSPATIAL HIERARCHY

Change the *Country* and *City* data types back to geospatial values and create a geospatial hierarchy.

54. In the **Fields** pane, choose the ellipsis icon next to **City**.
55. Hover over **Change data type**, and then choose **City**.
56. Repeat the same steps to change the **Country** field data type to **Country**.

Image description: The previous image shows the Country and City fields as geospatial fields.

57. In the **Fields** section, next to **Country**, choose the ellipsis icon.
58. Choose **Add to a hierarchy**.
59. Choose **Create a new geospatial hierarchy** if it is not already selected.
60. Choose **Add**.
61. In the **Name your hierarchy** section, enter

    Country - City

    .
62. Choose **This hierarchy is for multiple countries** if it is not already selected.
63. From the dropdown list, choose **Country** if it is not already selected.
64. Choose **Create hierarchy**.
65. In the **Fields** pane, next to **City**, choose the ellipsis icon.
66. Choose **Add to a hierarchy**.
67. Choose **Add to existing geospatial hierarchy**.
68. Choose **Add**.
69. In the top menu bar, choose **SAVE & PUBLISH**.

Expected output: Your fields are now organized in field folders, and geospatial data is organized in a hierarchy. This organization helps you optimize the sales dashboard for future visualization creation.

Task complete: You have completed **Task 4** by grouping fields together. You added fields to two field folders and created a geospatial hierarchy.

CHALLENGE A: SWAP THE DATASET IN THE ANALYSIS

Your team loves the changes you made to the dataset. However, the dataset is not connected to the sales dashboard. Connect your sales dashboard to the **lab-3-quicksight-data-set-child** dataset you just configured so your team can use the new data.

Hint: You need to navigate to your **quicksight-lab-3** analysis, and then replace the dataset with your new **quicksight-data-set-child** dataset.

Hint: For help completing this challenge task, see the **Challeng A solution**.

Conclusion

Task complete: You have incorporated data from different sources into one dashboard. You have also cleaned up a dataset using the join, filter, and calculation options in QuickSight.

You have successfully done the following:

- Joined two data sources together using Amazon QuickSight joins
- Filtered and transformed the data during import
- Created a field folder to consolidate related data points
- Created a geospatial hierarchy

End lab

Follow these steps to close out the console and end your lab.

70. At the upper-right corner of the QuickSight console, choose the user icon and then choose **Sign out**.
71. On this screen, choose **End lab** and then confirm that you want to end your lab.

Appendix

CHALLENGE A SOLUTION

Answer: To complete this challenge, you change the underlying dataset in your **quicksight-lab-3** analysis to the child dataset.

72. To return to the main QuickSight menu, choose the **QuickSight** icon at the top of the page.
73. In the left navigation pane, choose **Analyses**.
74. Choose **quicksight-lab-3**.
75. In the top menu bar, choose **File** and then **Save as Analysis**.
76. For **Analysis name**, enter

    quicksight-lab-3-child

77. Choose **SAVE**.
78. Above the **Fields list** to the left of the dashboard, in the **Dataset** section, choose **Manage datasets**.
79. Choose the ellipsis icon for **lab-3-quicksight-data-set**.
80. Choose **Replace**.
81. From the list of datasets, choose **lab-3-quicksight-data-set-child**.

82. Choose **Select**.

A warning message should indicate a mismatched geographic role for the **Region** field. You can disregard the message because the mismatch will be fixed when you map the fields in the next step.

83. Choose **Update field mapping** to resolve the issue.
84. Choose **Confirm**.
85. To close the **Datasets in this analysis** pop-up window, choose **X**.

Expected output: Your dataset is now updated to include the fiscal year data and custom *Cost* calculation.

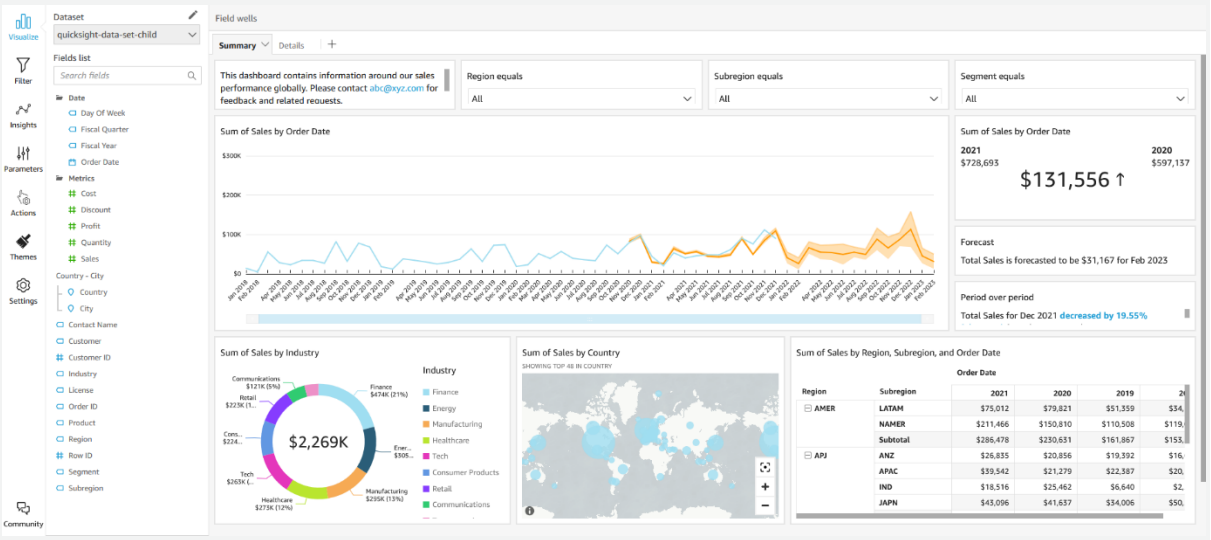

Image description: The previous image shows the dashboard.

To continue this lab, move on to the Conclusion.