



Free Training Transcript: Cross-database Joins

Welcome to this video on cross-database joins. You can download the datasets to follow along in your own copy of Tableau.

Data from different sources

Many organizations have different data in different systems. Maybe there's financial data in a SQL Server database and product data in Amazon Redshift. The data may live in different environments, but we want to do analysis on everything together.

For this video, we have two different file-based data sources – one Excel, one CSV. We're only using flat files for ease of you following along.

In this instance, our data is structured such that our sales data has a product ID but no other information about the product sold. Our product data in the CSV file is a simple table of product ID, category, sub-category, and the product name.

This is a great setup for a cross-database join. If we perform a left join Product data to the Sales data on Product ID, we'll add that product information to each relevant sales transaction. For more information on joins, check out the video on join types.

Cross-database joins

To create this integrated data source with a cross-database join, we'll start with our Sales – click Excel, and open the Sales 2016 file. To keep track of what's what, let's double click and name this table “Sales”. Next, we'll add a data connection to this source. Click the Add button by the Connections area, NOT the new data source button in the toolbar: We'll click Add, and Text File, and open Products 2016. The sheet has already been added to the canvas. Let's rename it just “Products”. And we'll name the data source more descriptively Sales and Products 2016.

Just like any join from within the same data source, we can edit the join – click on the icon, and we'll select a left join. We can see that our data has been joined at the row level on Product ID. In the grid, we can see the product information (in yellow) has been added to each row from the Sales data (in blue).

Using Integrated Data Sources

Now that we have a new, single data source, let's test it out. And we see we have a single data source in the data pane, but our fields are broken out by table—just like with a join from within the same data source. If desired, we can turn that off, by right clicking and say Group by Folder instead of Group by Data Source Table. And make a hierarchy for Sub-Category and Category, add in Product ID and Product Name, and we can hide the Product ID from the CSV table as it's just there for the join clause.

Let's bring Category to rows, and expand it out. Bring Sales to Columns. And break it out by order date. An integrated data source functions like any other single-source data source—it can even be published and saved, unlike blends.

Conclusion

Thank you for watching this video on cross-database joins. We invite you to continue with the Free Training videos to learn more about using Tableau.