Split

- Overview
- Step Configuration
- Complete Example

Overview

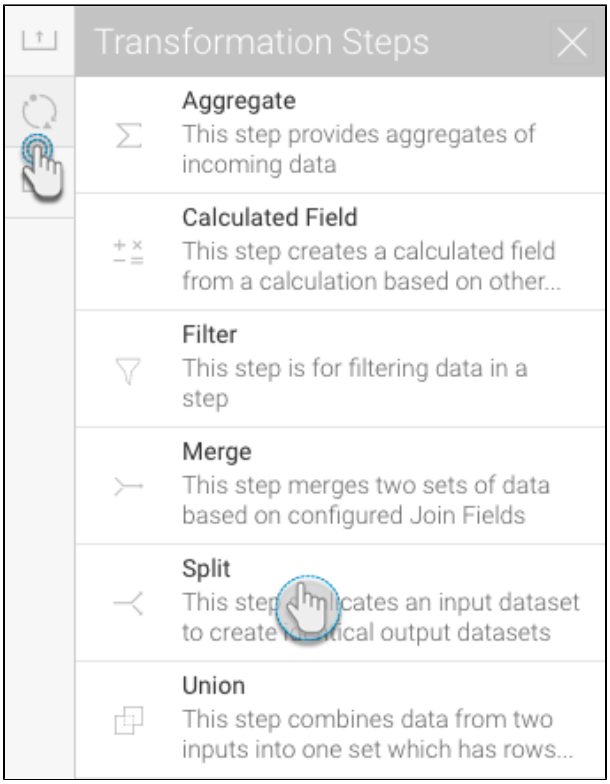
This type of transformation duplicates an input dataset to create identical output datasets. This is done by using the Split step in the Data Transformation module. This step does not need to be configured, and can be used multiple times to split a data set.

Note: This is a built-in step, and therefore will be available in the Transformations List by default.

Step Configuration

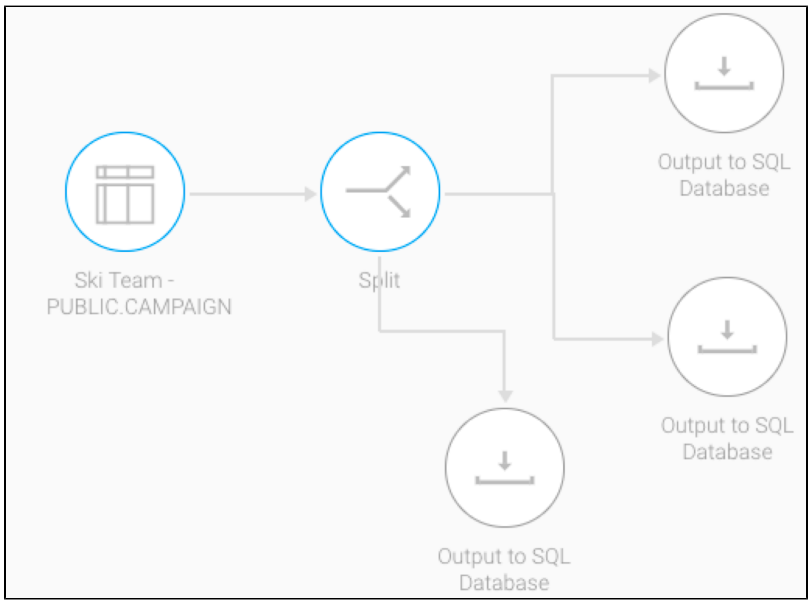
Follow the instructions below to configure a split step:

1. Expand the Transformation Steps button on the left side of the Transformation Flow builder, to view a list of transformation steps.
2. Drag the Split step from the list of transformation steps.



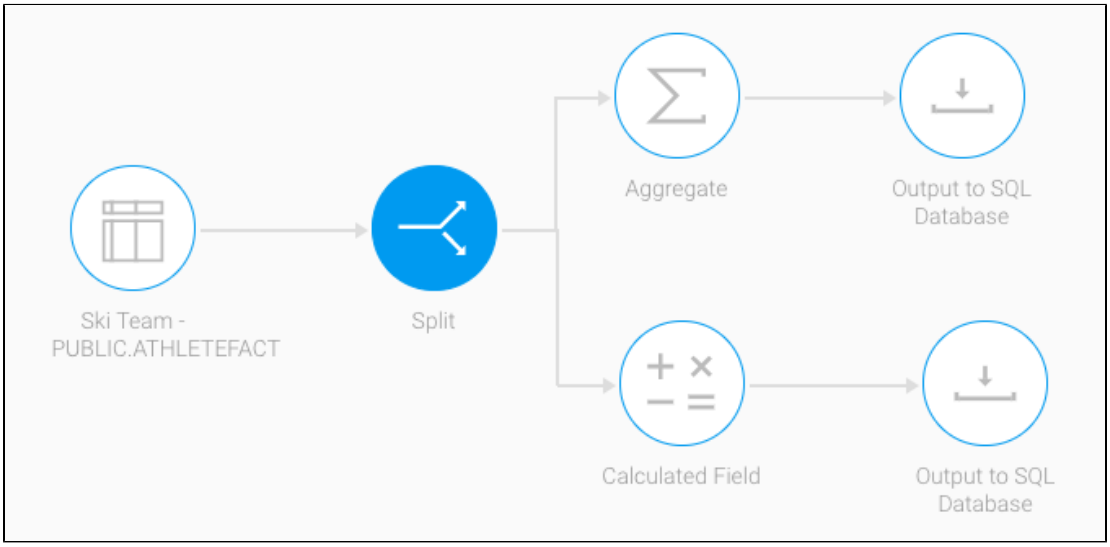
3. Connect this step to the previous step in the flow.
4. Add more steps to the flow, and one by one connect them to the split step.

5. For example, you can split data from a single data source and store it in multiple databases.



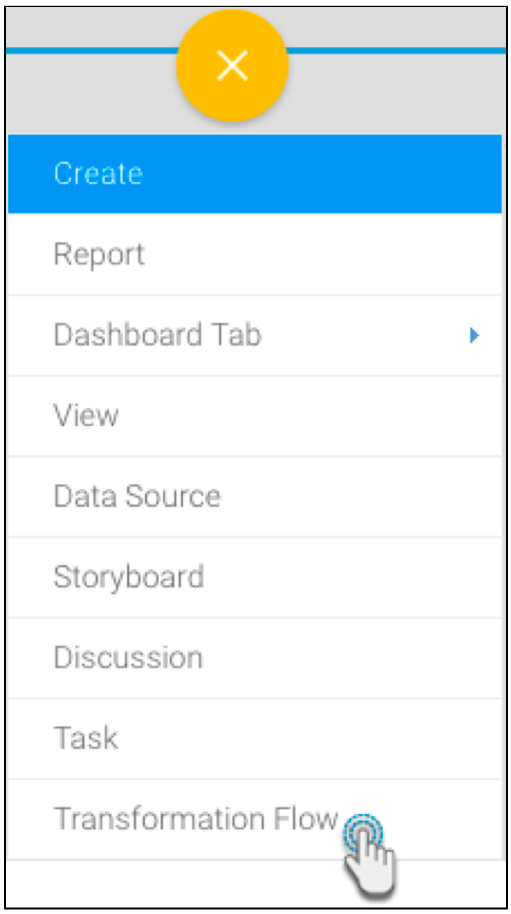
Complete Example

In this example, we will cover how to create a simple transformation flow that involves a split step. Our flow will involve extracting data from a data source, splitting it into two parts, with each part getting a different type of transformation applied to it. Each of the results are then saved into separate outputs. You could always include additional steps in your own transformation flows.



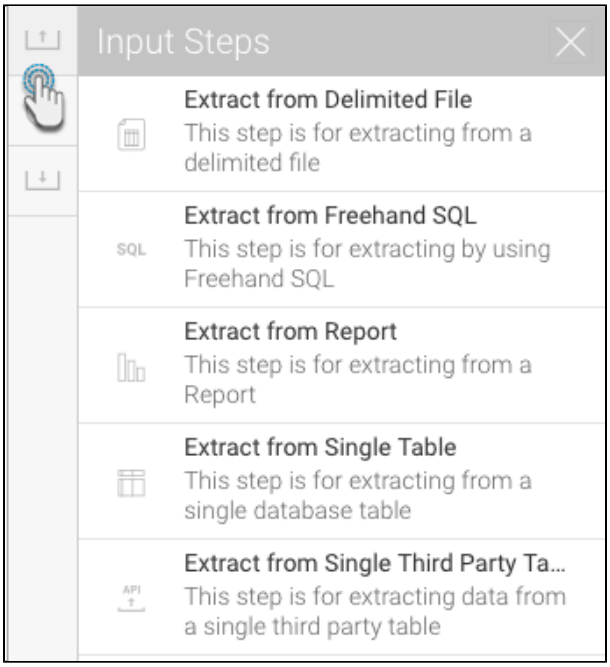
1. Click on the Create button in the top-right corner.
2. Then select Transformation Flow.

⚠ If you do not see this option, you may not have security access to transformation flows. Learn how to get access here.

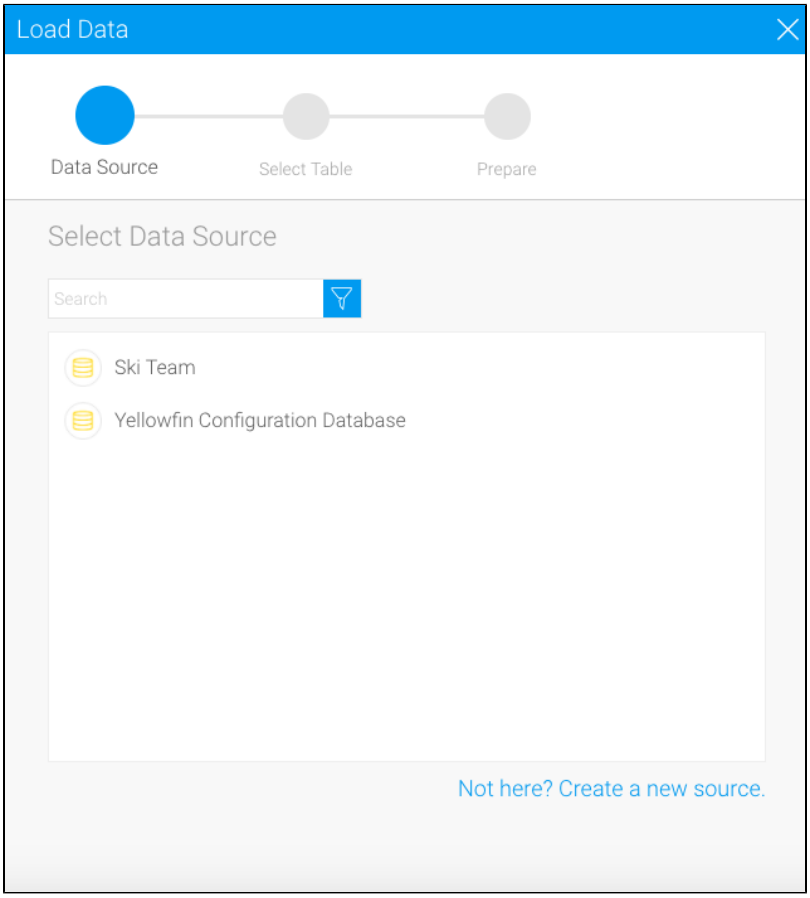


3. You will be taken to the transformation flow builder.

4. Hover your cursor over the input steps button on the left side. A panel with a list of all data extraction steps will appear.

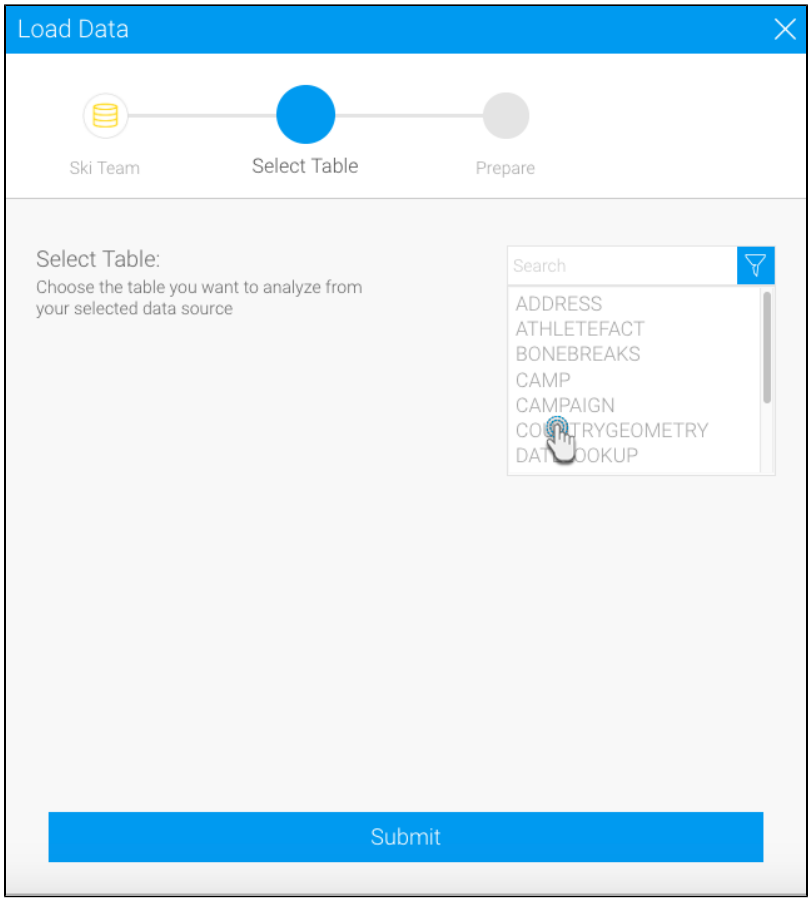


5. Drag one of these steps onto the canvas. For this procedure, we will use the single table step as an example. (Click here to learn about all the different input steps.)
6. On doing so, a popup will appear to load data from a data source.



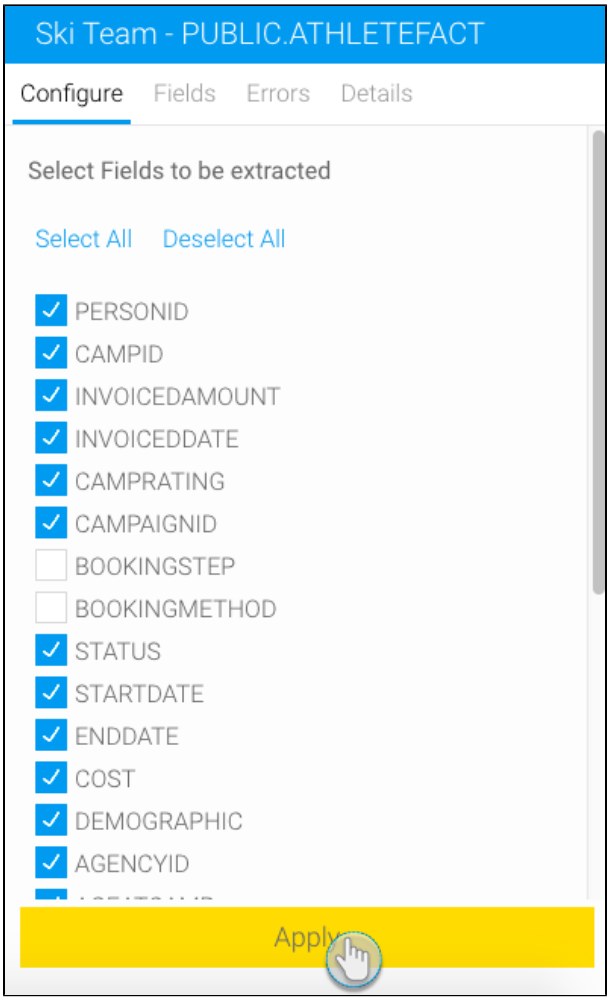
7. Click on the data source that you require.

8. Then choose the database table, and click on Submit.



9. The selected table's fields will appear in the transformation flow panel to be configured.

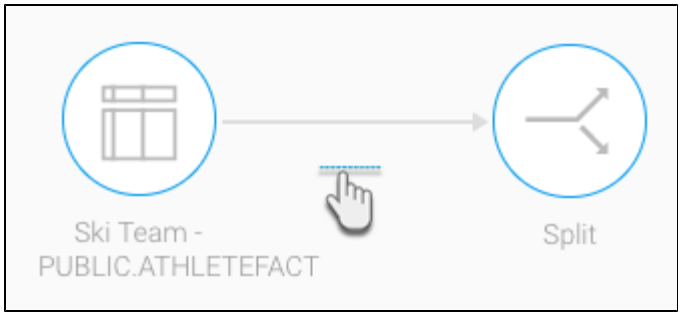
10. Select only the fields that you want data to be extracted from.



11. You can make further changes to the step, such as renaming it, adding a description, etc.
12. Once you're done with the step configuration, click on the Apply button.
13. On doing so, the data preview panel will display the data extracted from the configured database table.

	CAMPAIGN... NUMERIC	CAMPAIGN... TEXT	MEDIANA... TEXT	TARGETDE... TEXT
1	51797	Ski Racing Maga...	WebsiteButton	Relaxation
2	51798	SnowWatch Online	DirectMail	Family
3	51799	Ski Racing Maga...	MagazineAd	Luxury
4	51800	Nastar Discussio...	WebBoardPost	Family

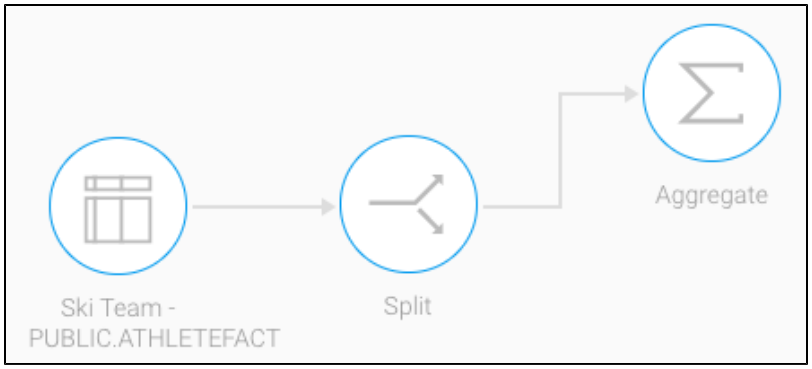
14. Once you are ready to split your data, extend the transformation steps panel, and drag in the Split step.
15. Then create a connection between the database input step (or the previous step) and this split step.



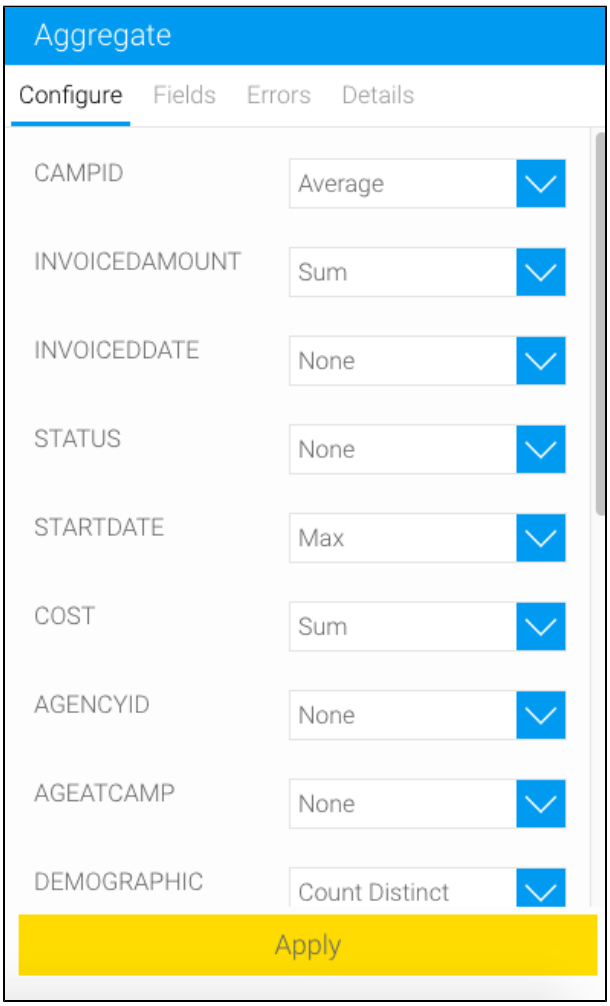
16. There is no need to configure the split step, however you can still choose which fields will be carried onto the next steps through the Field tab.



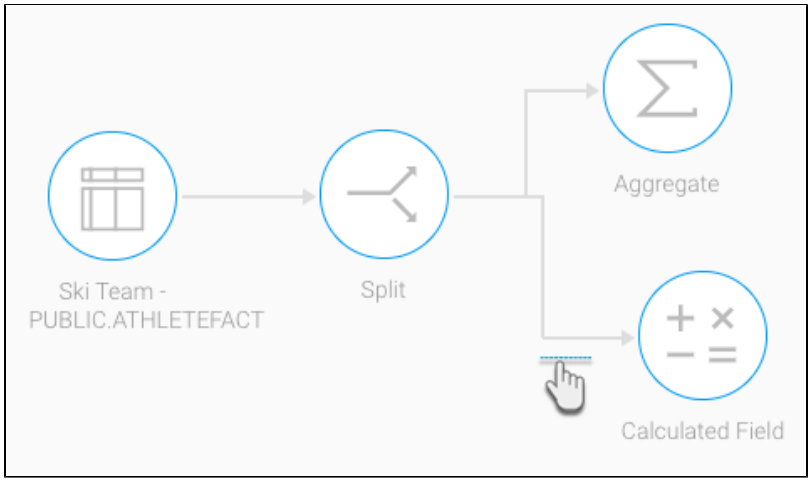
17. Now you can add multiple steps simultaneously and perform different transformations using the same dataset.
18. For this example, we will aggregate this dataset and separately create a custom calculated field on its copy.
19. Drag in the Aggregate step from the transformation steps panel, and connect the split step to it.



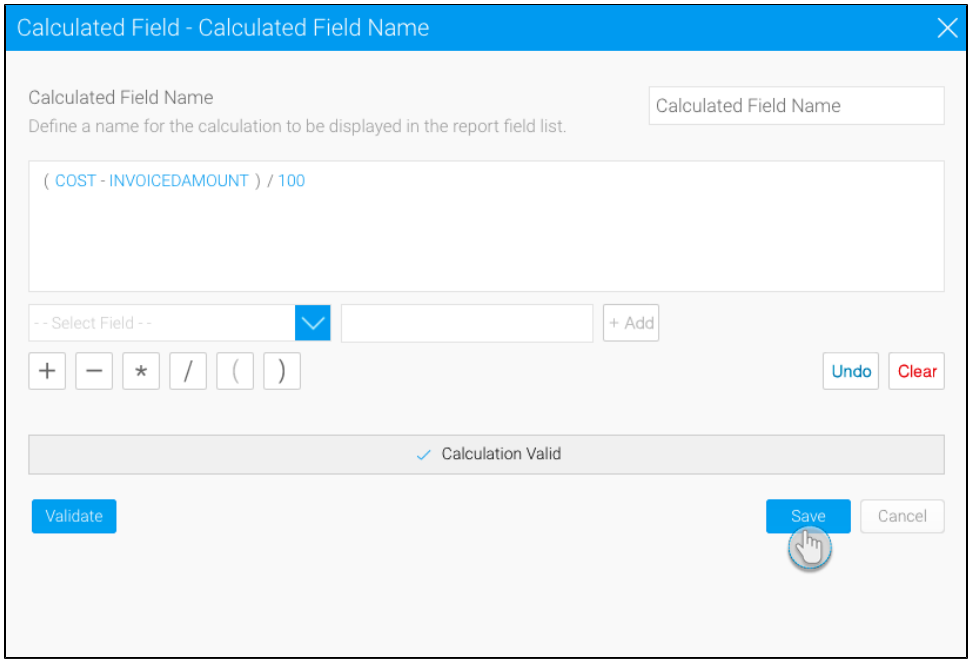
20. Click on the aggregate step, and in the transformation flow panel, select the aggregations to be applied on each of the fields. Then click Apply.



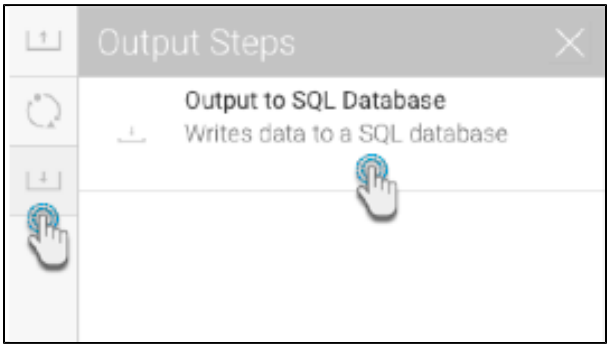
21. The result of this step will appear in the data preview panel.
22. Next, include the calculated field transformation step to the flow, and connect the split step to it.



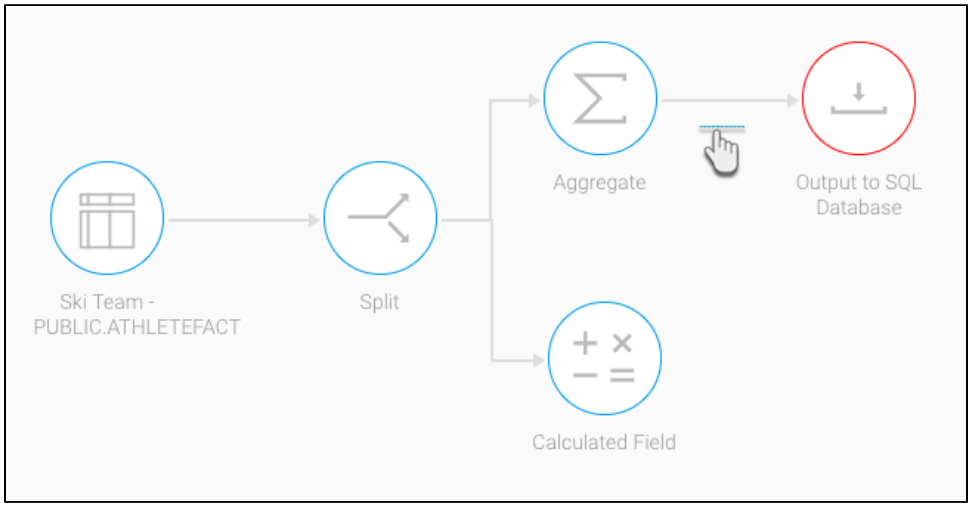
23. Click on this newly added step and configure it through the transformation flow panel. First click Add Item, and then in the popup window, create a custom calculated field that you want your data to generate.
24. You can confirm the validation of your formula by using the Validate button.
25. Then click on the Save button.



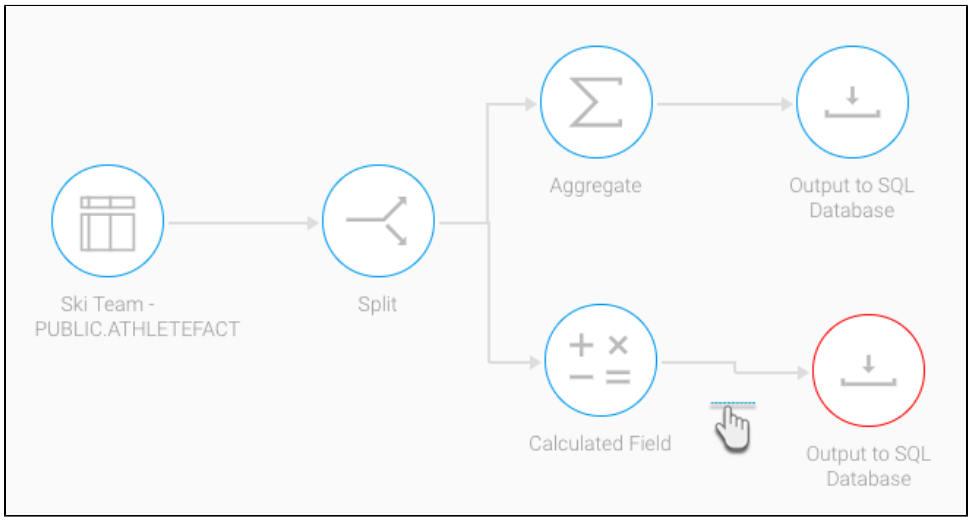
26. The result of this step will be generated in the data preview panel.
27. Now we will save the data from each of these steps into separate output steps.
28. Extend the output steps panel by hovering on its icon, and drag the SQL database output step onto the canvas.



29. Connect the aggregate step to this output step by creating a connection.
    Note: By default, the output step will be highlighted as red to signify that it contains errors. This is because it has not been configured yet.



30. And then configure the output step through the panel on the right-side. Click here to learn more about configuring this step.
31. Similarly, add another SQL database output step, and create a connection with the calculated field step.



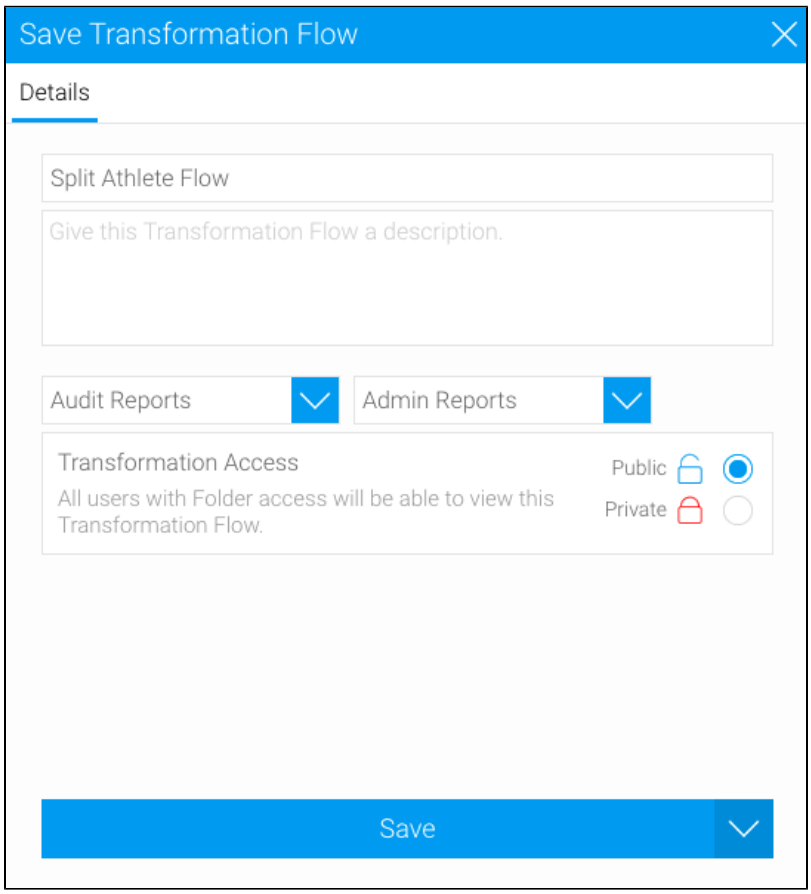
32. Configure this output step as well.
33. You can now execute the draft flow by clicking on the run button in the top header menu. (This does a quick execution of the data rows in the data preview panel.)



34. Or save the flow for a full execution. To do that, click on the Publish button.



35. Then provide details in the popup that appears, such as providing a proper name, and selecting rights to access the flow.



36. Finally, click on the Save button.