Exercise 1. Creating a Project and a Parallel job

Estimated time

01:00

What this exercise is about

This exercise covers the whole process of creating project and creating, compiling, running, and monitoring a DataStage parallel job.

What you should be able to do

At the end of this exercise, you should be able to:

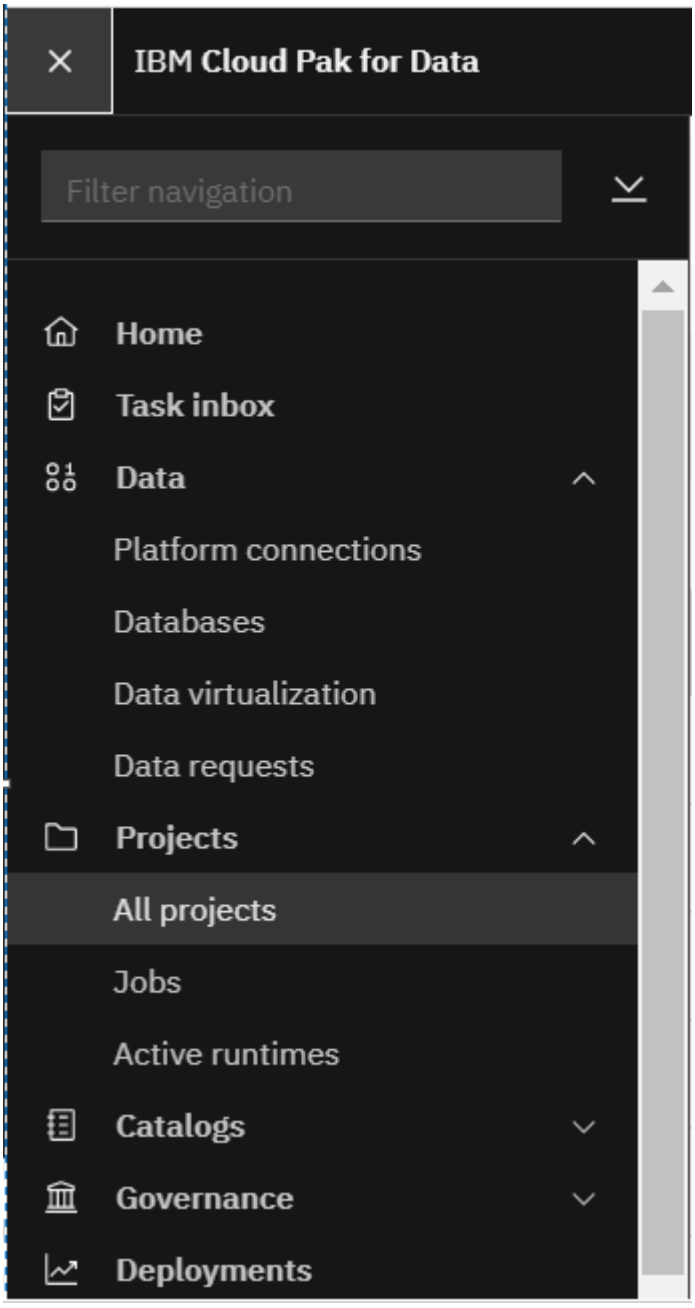
- Create a Transform Project
- Design a simple DataStage parallel job
- Compile a job
- Run a job
- View messages written to the job log

Introduction

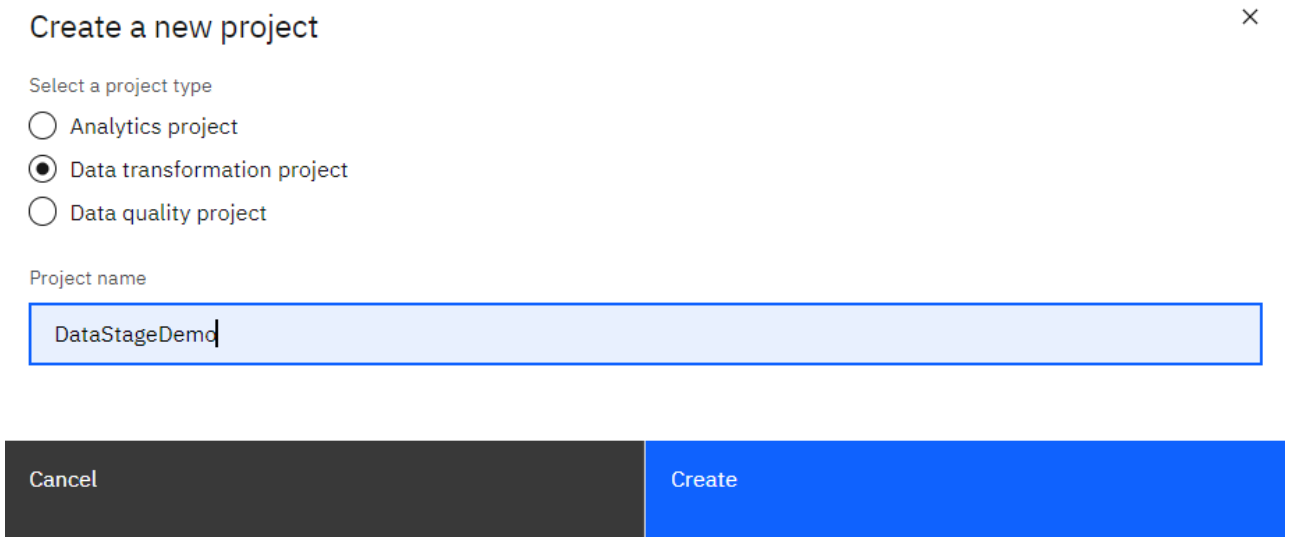
Building a DataStage parallel job, however complex, involves the same basic workflow. This exercise introduces you to that workflow. Later exercises will introduce additional functionality into the workflow.

Task: Create Project

1. Expand **Projects** and click on **All Projects**



2. Click on **New Project** button on top-right of the screen and below window will appear. Select **Data transformation project** radio button and give **DataStageDemo<YourFisrName><InitialOfLastName>** as the name for the project. So if your name is **John Smith**, then the project name will be **DataStageDemoJohnS.** Then click on **Create**.

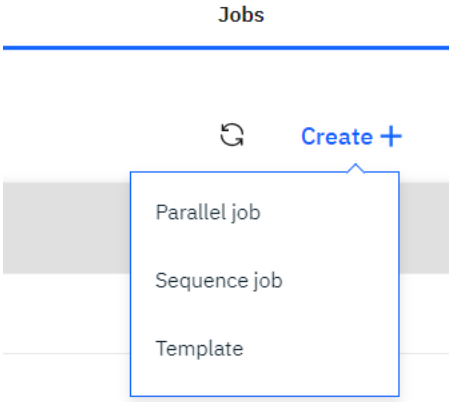


3. Wait for some time and then refresh the page and you shall see the project is created.

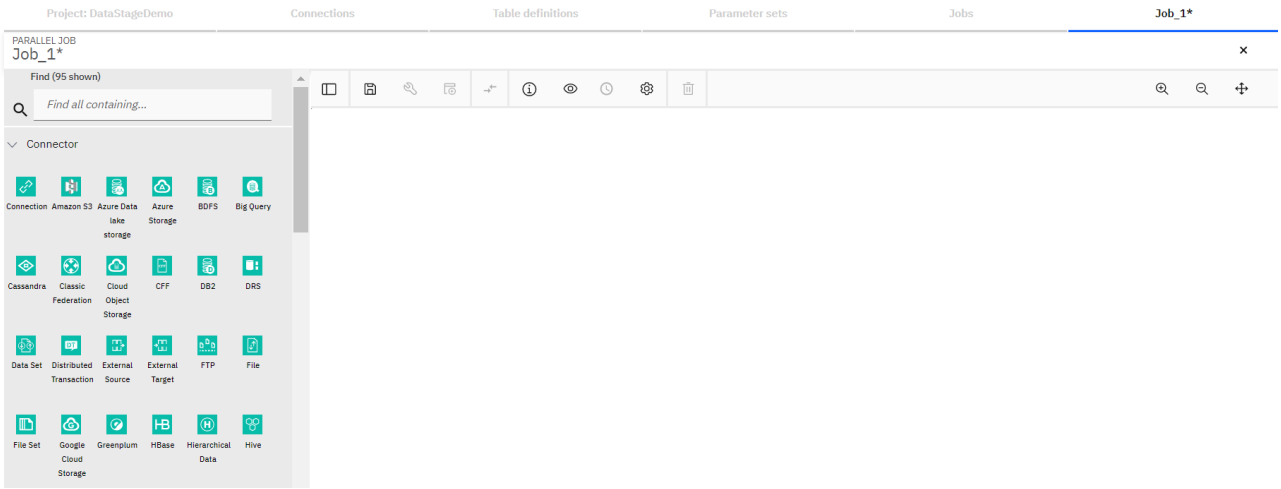
Task: Create a parallel job

1. Click on your project created in the previous task to open the project.
2. Click on **Jobs** tab after the project is opened.

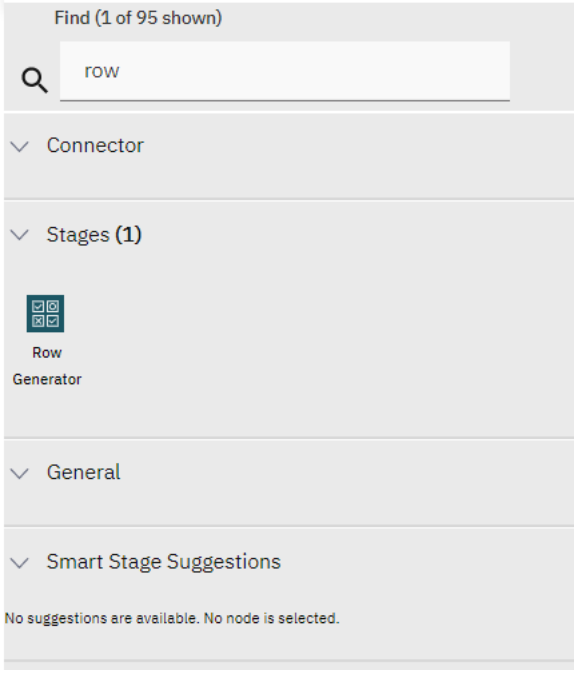
3. While in Jobs tab click on **Create +** button. A drop will appear, now click on **Parallel job**.



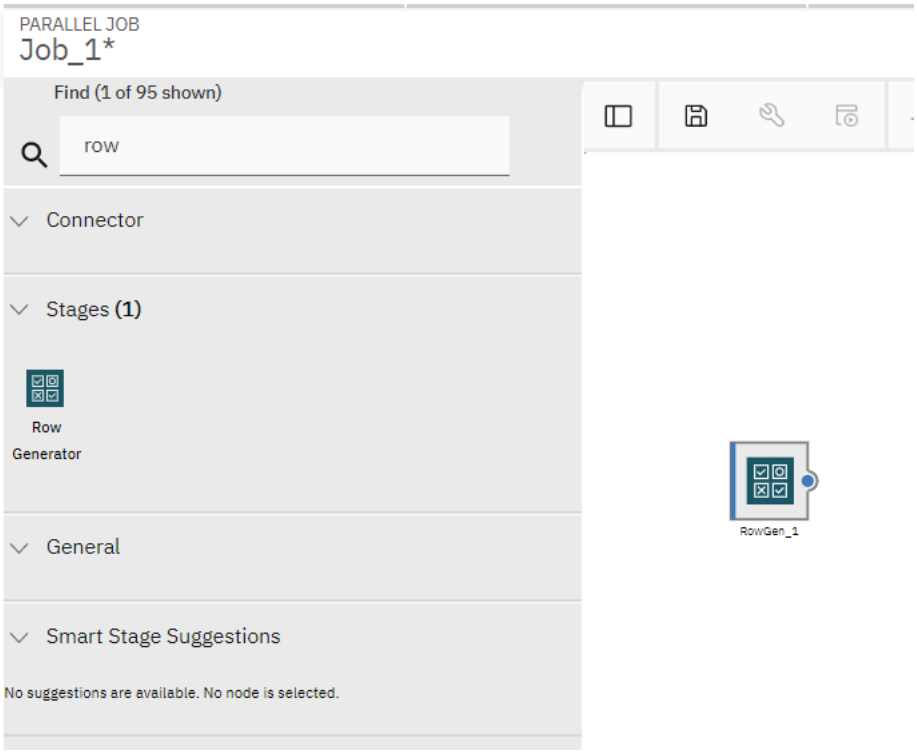
4. Now you can see the below screen



5. Now type **row** in the **Find** search box and you can see **Row Generator** stage appears in the palette



6. Drag the **Row Generator** stage onto the middle blank area called canvas

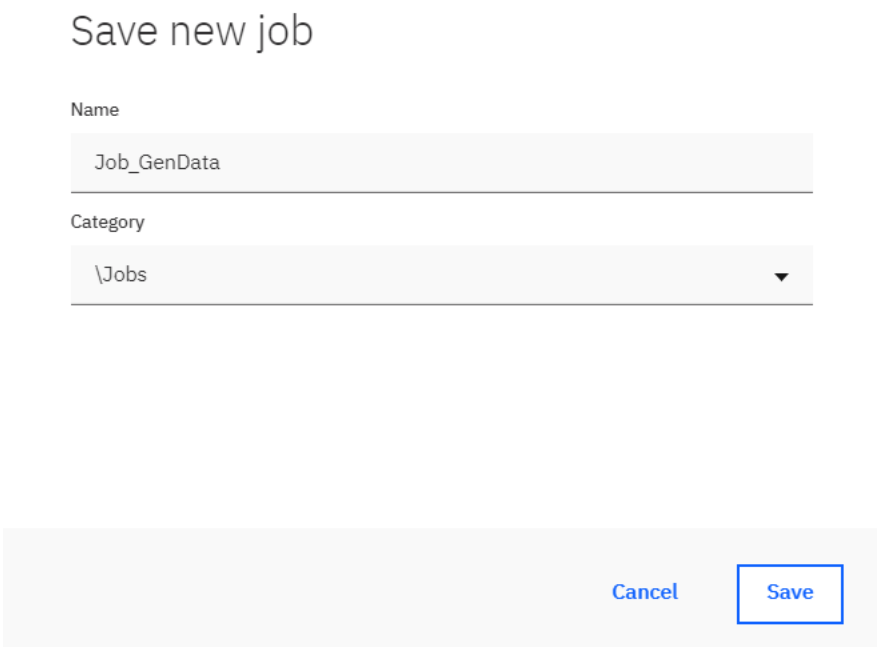


7. You might see a Copy stage to the right of Row Generator stage in a dotted outline connected with a dotted line. The built-in machine learning feature suggests the next stage based on the usage pattern of the stage we place on the canvas. It can be turned off by unchecking Smart stage suggestions option from the Settings.

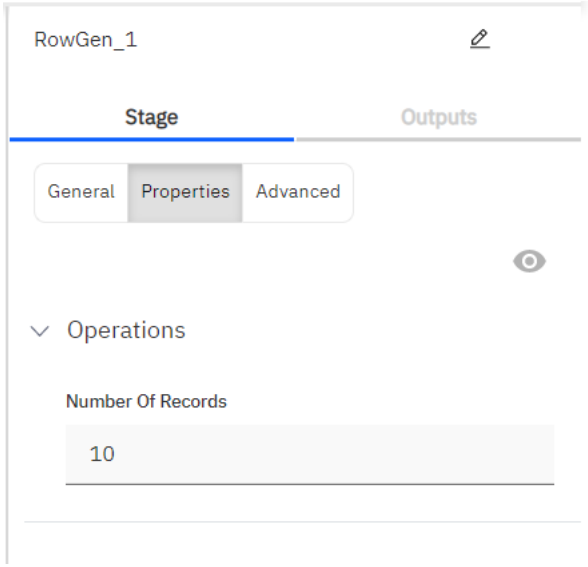
8. Now type **peek** in the **Find** search box and Peek stage will appear in the Palette.

9. Drag the **Peek** stage onto the canvas and place it to the right of **Row Generator** stage. Now also you might see few stages as part of Smart Suggestion. If you click on anywhere on the canvas it'll go away

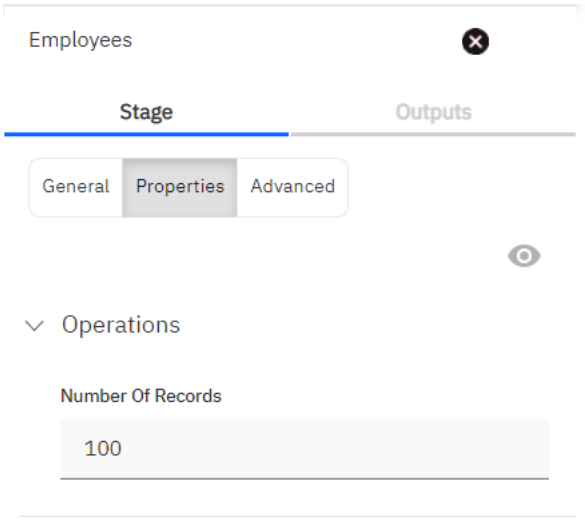
10. If both the stages are not joined by a link, then click on small blue dot that appears to the right of the Row Generator stage and then click on the blue dot that appears on the left of Peek stage. Now you shall have the two stages linked.
11. Now save the job by clicking on the **save** icon. Save new job window will appear. Give job name as **Job_GenData** and let the Category be **\Jobs** and click on the **Save** button.



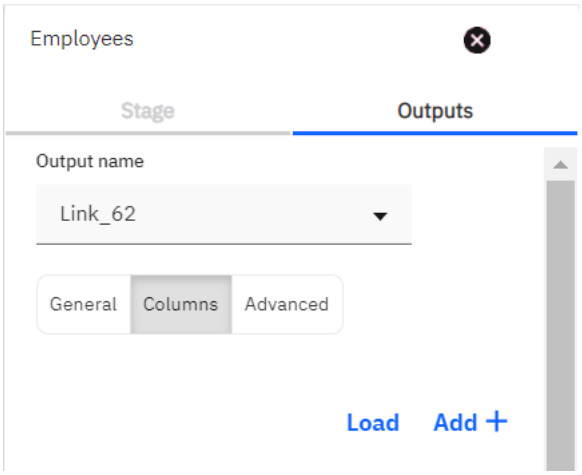
12. Now Double click on Row generator stage and one window will appear on the right side of the screen. Here you can specify different properties related to the stage.



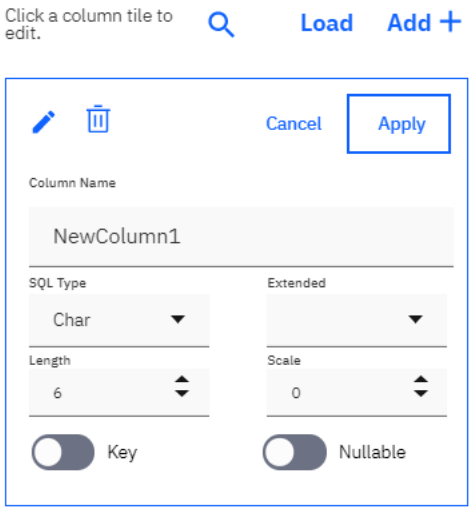
13. Change the name of the **Row Generator** stage from **RowGen_1** to **Employees.** You can rename it at the top of the **Row Generator** stage window.
14. In the **Properties** section set **Number Of Records** to **100** from **10**.



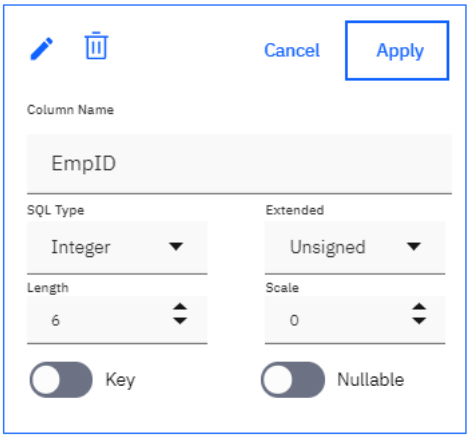
15. Now go to the **Outputs** tab and click on **Add+** link to add columns to the stage.



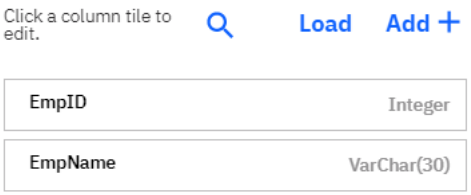
16. Click on the NewColumn1 section and you'll see option to define the column details.



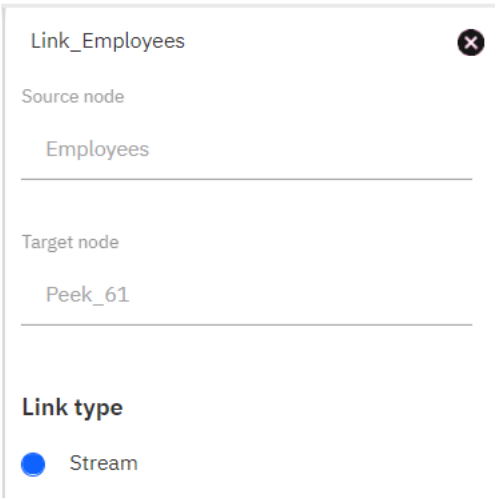
17. Change the **Column Name** to **EmpID** and **SQL Type** to **Integer**. Click on **Apply**



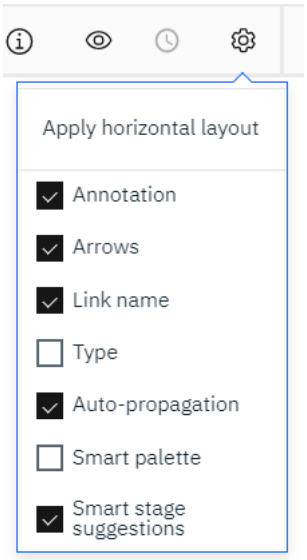
18. Click on Add+ link and add another column as described in the previous steps. Name the column as **EmpName**, set **SQL Type** to **Varchar** and **Length** to **30**. Click on **Apply**.



19. Now click on **OK** button to save the stage changes and come back to the canvas.
20. Now double click on the link and change its name to **Link_Employees** and click on **OK**.



21. This step is optional. If you need the name of the link to appear on the links on the canvas then go to **Settings** and check on **Link name** checkbox.

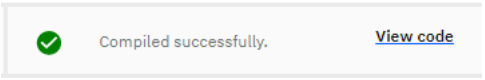


22. Now double click on the Peek stage and change its name to **Peek_Employees** and click on **OK**.
23. Click on the save icon to save the job

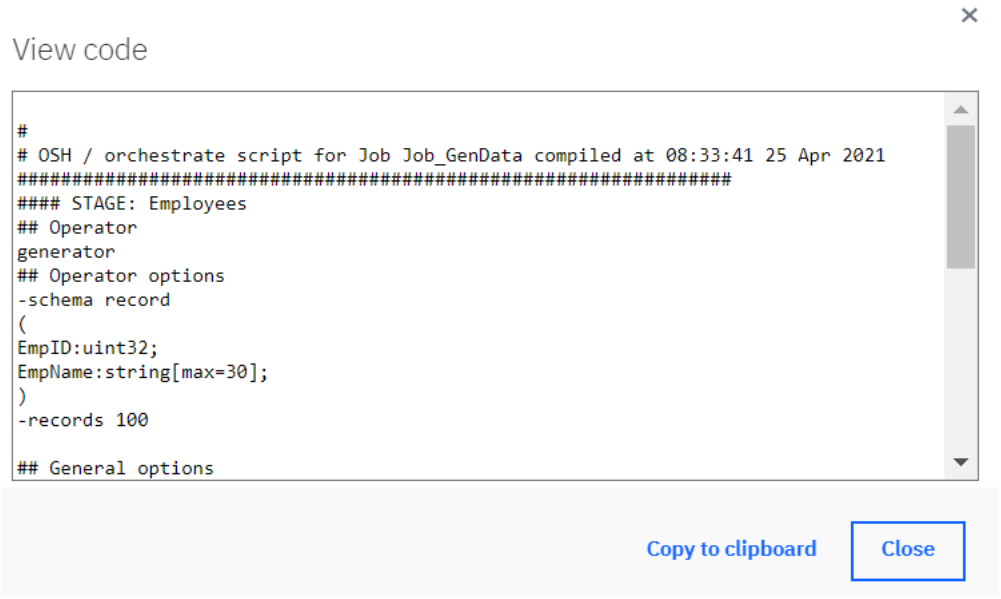
Task: Compile, run, and monitor the job

1. Click on the **Compile** icon to compile your job. If your job compiles with errors, fix the errors before continuing.

2. If the job compiles successfully, you'll see the message **Compiled Successfully** and **View Code** link will appear.

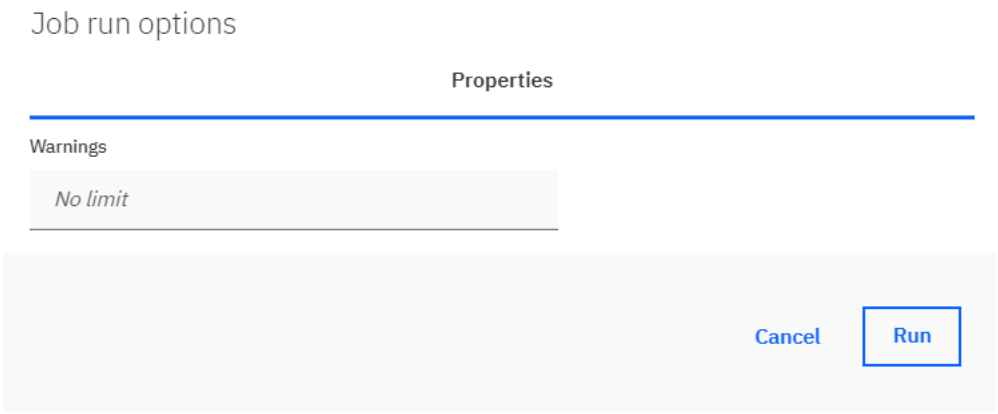


3. This step is optional. Click on the **View code** link to see the osh code generated because of compilation.

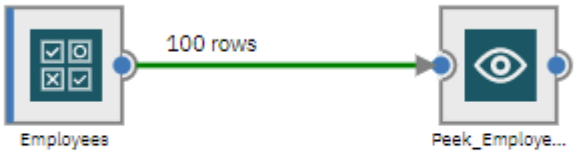


4. Click the **Run** icon to run the job

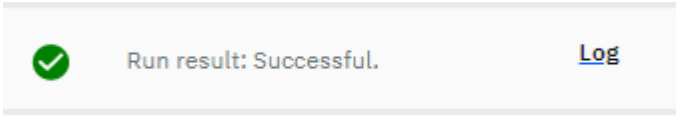
5. Once you click on the **Run** icon, **Job run options** window will appear. Click on **Run** button.



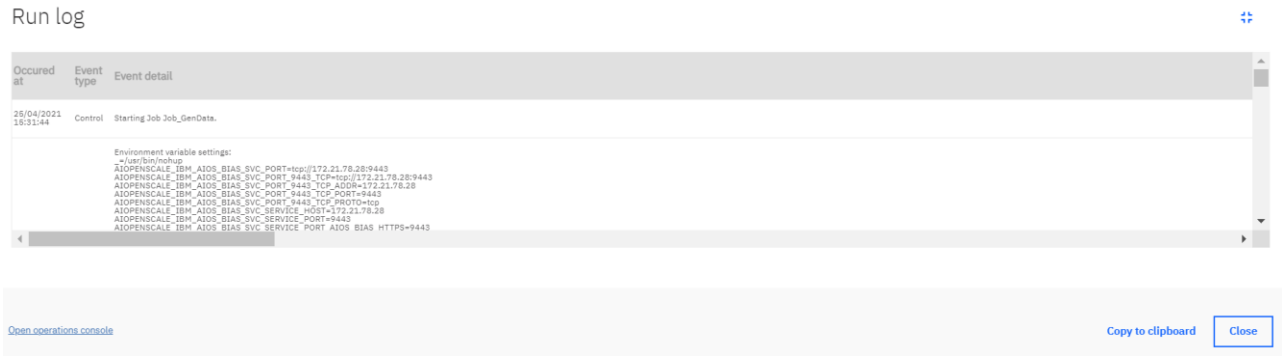
6. When the job is running the link color will turn blue, once the job run completes successfully it'll turn green. If the job run fails, then the link will turn red. Also, the number of records transferred through the link will appear on the link. In our demo, it should show 100 rows.



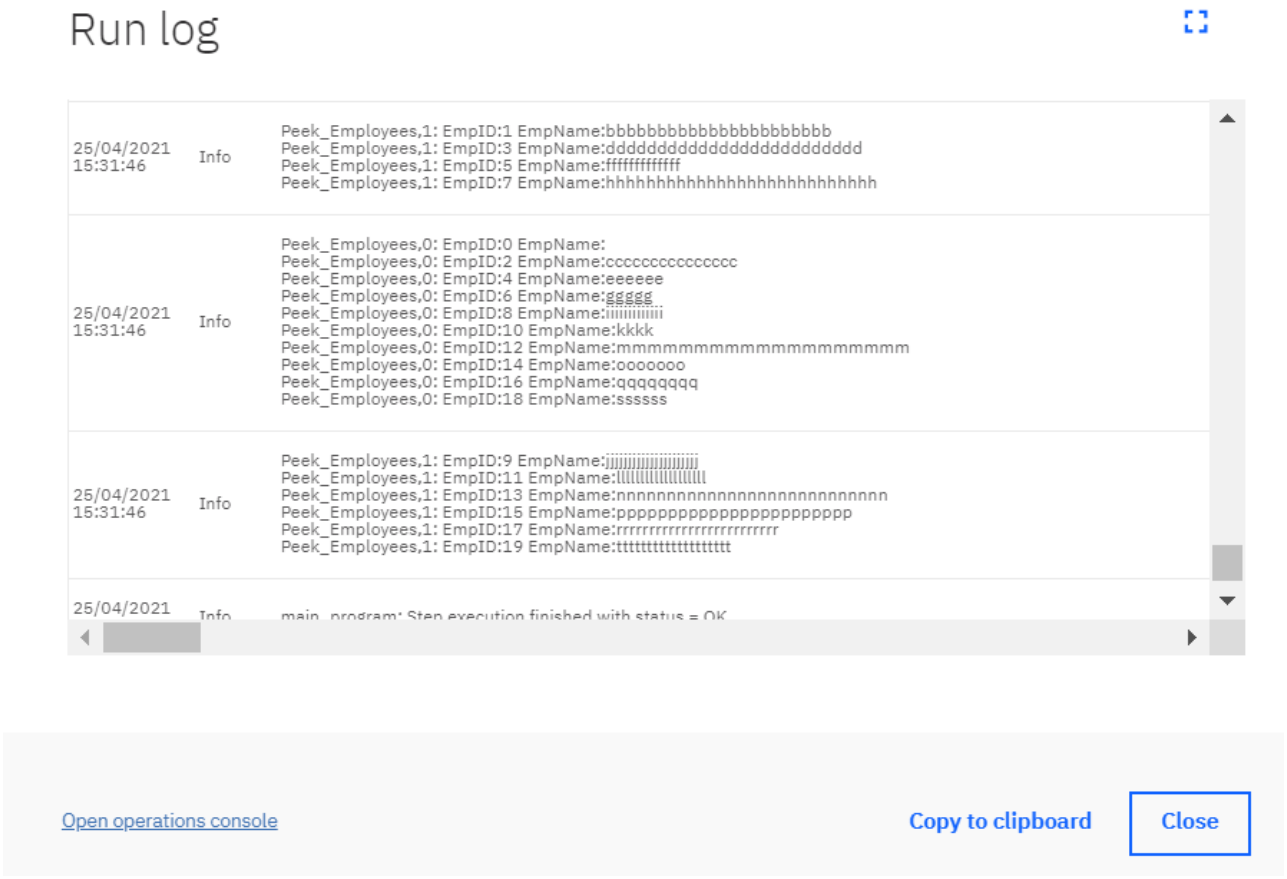
7. After the job is run, you can see whether the job ran successfully or not along with the **Log** link.



8. Click on the **Log** link to view the logs in Data Flow Designer.



9. Scroll down to see the Peek stage output in the log.



10. Right click and open **Open operations console** link in a new tab to open it. It has many options to control and monitor jobs.