

---

# 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

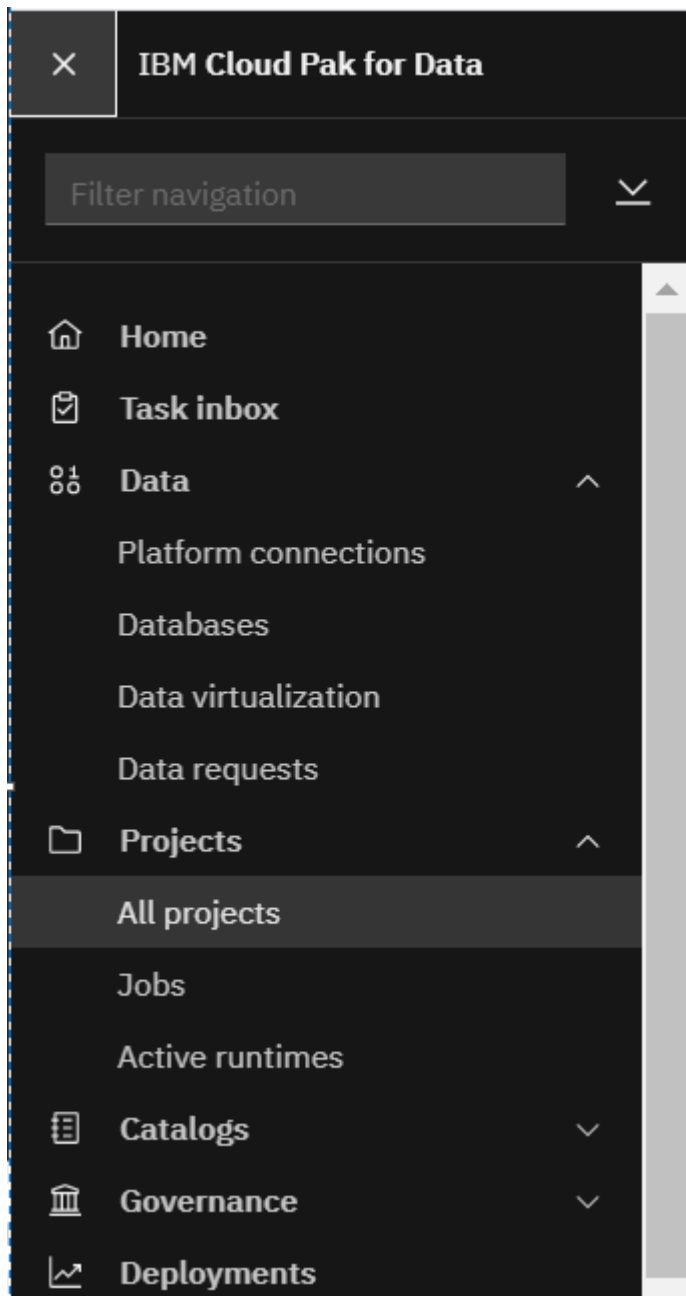
## Introduction

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**.

---

## Create a new project ×

Select a project type

- ☐ Analytics project
- ☒ Data transformation project
- ☐ Data quality project

Project name

DataStageDemd

Cancel

Create

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

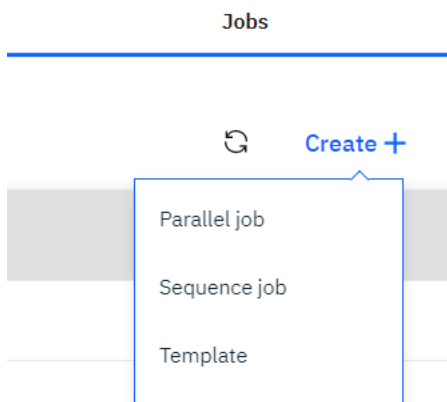
---

## Task: Create a parallel job

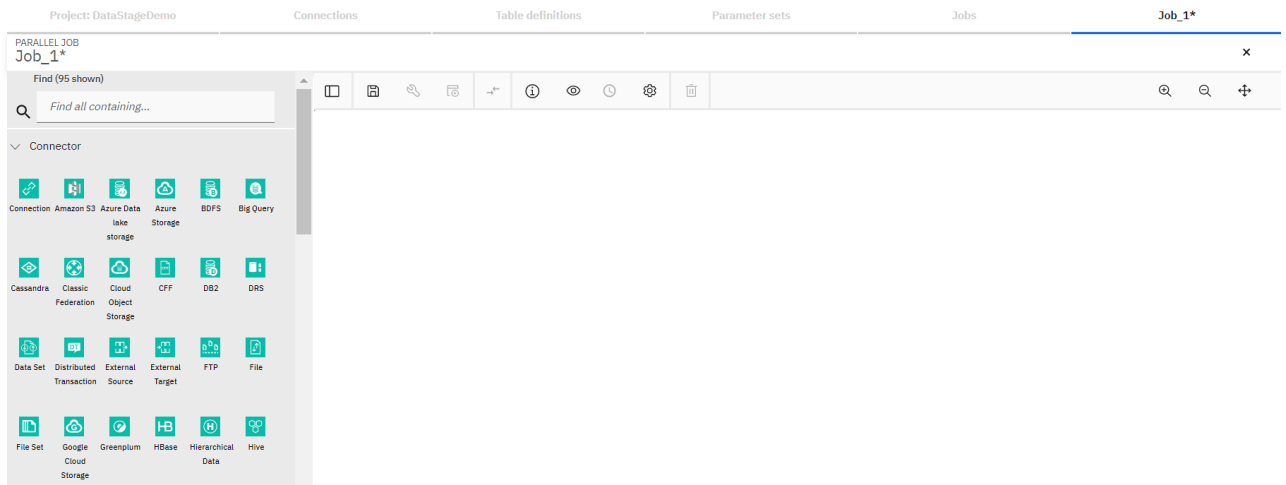
1. Click on your project created in the previous task to open the project.
2. Click on **Jobs** tab after the project is opened.



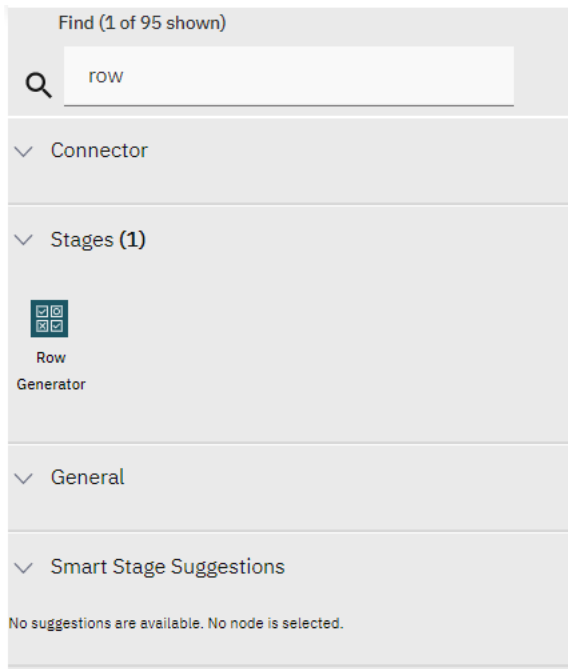
3. While in Jobs tab click on **Create +** button. A drop will appear, now click on **Parallel job**.



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



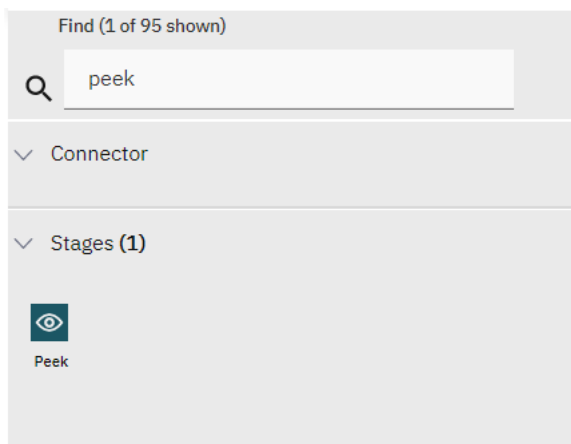
6. Drag the **Row Generator** stage onto the middle blank area called canvas



- 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.



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



- 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



- 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.
- 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.

---

## Save new job

Name

Job\_GenData

Category

\Jobs

Cancel

Save

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.

The screenshot shows the configuration window for the 'RowGen\_1' stage. At the top, the name 'RowGen\_1' is displayed with an edit icon. Below this, there are two tabs: 'Stage' and 'Outputs'. The 'Stage' tab is active, and within it, there are three sub-tabs: 'General', 'Properties', and 'Advanced'. The 'Properties' sub-tab is selected. Under the 'Properties' sub-tab, there is a section titled 'Operations' with a dropdown arrow. Below 'Operations', there is a label 'Number Of Records' followed by a text input field containing the value '10'. There is also a small eye icon to the right of the 'Operations' section.

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**.

Employees ✕

**Stage** Outputs

General **Properties** Advanced

Operations

Number Of Records

100

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

Employees ✕

Stage **Outputs**

Output name

Link\_62 ▼

General **Columns** Advanced

Load Add +

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

Click a column tile to edit. 🔍 Load Add +

✎ 🗑 Cancel **Apply**

Column Name

NewColumn1

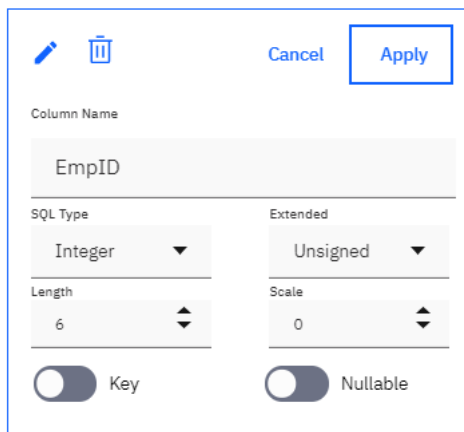
SQL Type Char ▼ Extended ▼

Length 6 ▼ Scale 0 ▼

☐ Key ☐ Nullable



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



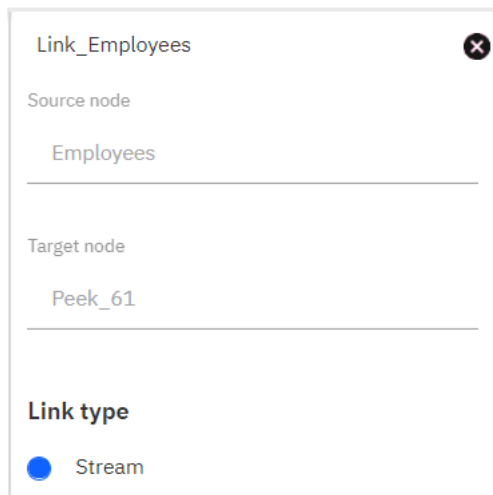
18. Click on Add+ link and add another column as described in the previous steps. Name the column as **EmpName**, set **SQL Type** to **Varchar** and **Length** to **30**. Click on **Apply**.

Click a column tile to edit. 🔍 Load Add +

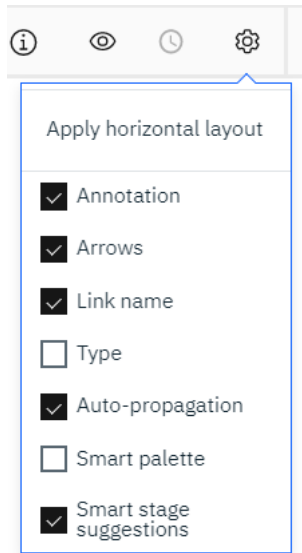
EmpID	Integer
EmpName	VarChar(30)

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

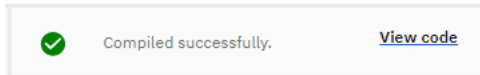
---

## Task: Compile, run, and monitor the job

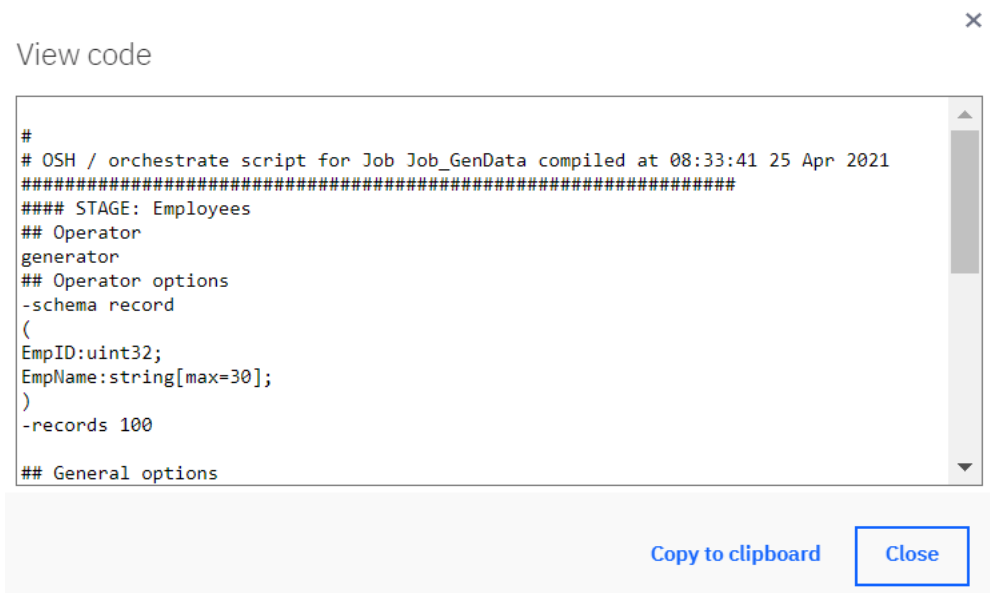
1. Click on the **Compile** icon to compile your job. If your job compiles with errors, fix the errors before continuing.



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.

## Job run options

### Properties

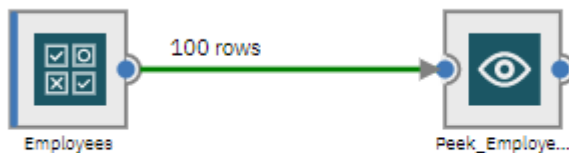
#### Warnings

No limit

Cancel

Run

- 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.



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



Run result: Successful.

[Log](#)

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

#### Run log

Occurred at	Event type	Event detail
25/04/2021 18:31:44	Control	Starting Job Job_GenData.  Environment variable settings: -user/bin/nohup AIOPEMSCALE IBM_AIOS_BIAS_SVC_PORT=https://172.21.78.28:9443 AIOPEMSCALE IBM_AIOS_BIAS_SVC_PORT_9443_TCP=https://172.21.78.28:9443 AIOPEMSCALE IBM_AIOS_BIAS_SVC_PORT_9443_TCP_ADDR=172.21.78.28 AIOPEMSCALE IBM_AIOS_BIAS_SVC_PORT_9443_TCP_PORT=9443 AIOPEMSCALE IBM_AIOS_BIAS_SVC_PORT_9443_TCP_PROTO=tcp AIOPEMSCALE IBM_AIOS_BIAS_SVC_SERVICE_HOST=172.21.78.28 AIOPEMSCALE IBM_AIOS_BIAS_SVC_SERVICE_PORT=9443 AIOPEMSCALE IBM_AIOS_BIAS_SVC_SERVICE_PORT_AIOS_BIAS_HTTPS=9443

[Open operations console](#)

[Copy to clipboard](#)

[Close](#)

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

## Run log



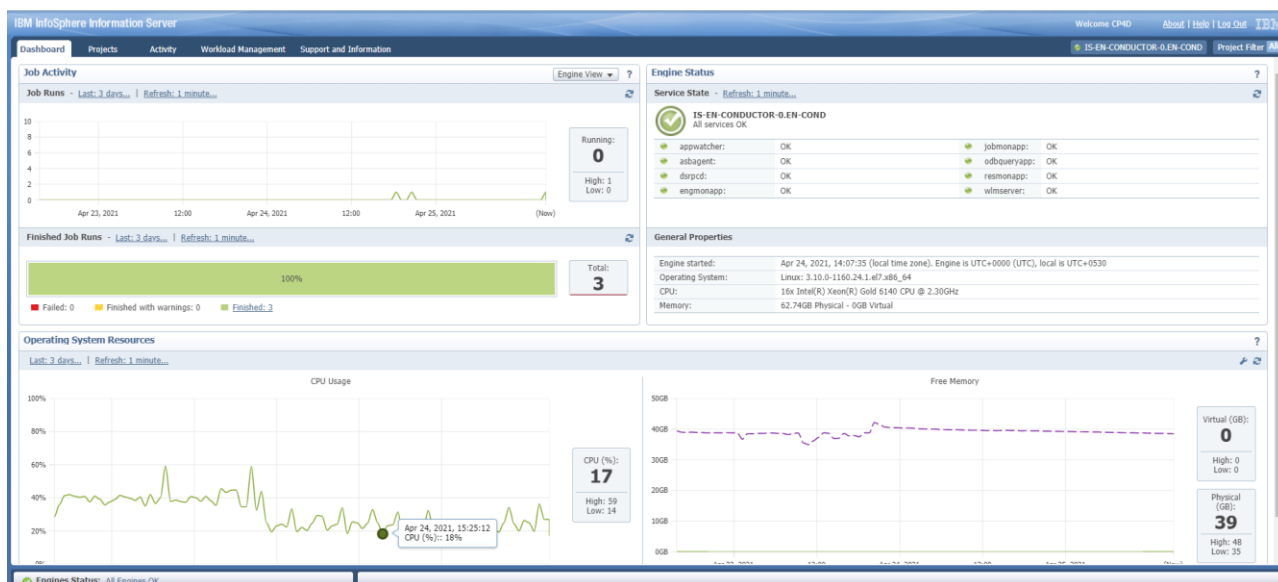
25/04/2021 15:31:46	Info	Peek_Employees,1: EmpID:1 EmpName:bbbbbbbbbbbbbbbbbbbb Peek_Employees,1: EmpID:3 EmpName:dddddddddddddddddddd Peek_Employees,1: EmpID:5 EmpName:ffffffffffff Peek_Employees,1: EmpID:7 EmpName:hhhhhhhhhhhhhhhhhhhhhhhhhhhh
25/04/2021 15:31:46	Info	Peek_Employees,0: EmpID:0 EmpName: Peek_Employees,0: EmpID:2 EmpName:cccccccccccccc Peek_Employees,0: EmpID:4 EmpName:eeeeee Peek_Employees,0: EmpID:6 EmpName:ggggg Peek_Employees,0: EmpID:8 EmpName:iiiiiii Peek_Employees,0: EmpID:10 EmpName:kkkk Peek_Employees,0: EmpID:12 EmpName:mmmmmmmmmmmmmmmmmmmm Peek_Employees,0: EmpID:14 EmpName:ooooooo Peek_Employees,0: EmpID:16 EmpName:qqqqqqqq Peek_Employees,0: EmpID:18 EmpName:ssssss
25/04/2021 15:31:46	Info	Peek_Employees,1: EmpID:9 EmpName:jjjjjjjjjjjjjjjjjj Peek_Employees,1: EmpID:11 EmpName:llllllllllllllllll Peek_Employees,1: EmpID:13 EmpName:nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn Peek_Employees,1: EmpID:15 EmpName:pppppppppppppppppppppppppppp Peek_Employees,1: EmpID:17 EmpName:rrrrrrrrrrrrrrrrrrrrrrrrrrrrrr Peek_Employees,1: EmpID:19 EmpName:tttttttttttttttttt
25/04/2021	Info	main_program: Step execution finished with status = OK

[Open operations console](#)

[Copy to clipboard](#)

[Close](#)

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.



The screenshot displays the IBM InfoSphere Information Server interface. The top navigation bar includes 'Dashboard', 'Projects', 'Activity', 'Workload Management', and 'Support and Information'. The 'Projects' tab is active, showing a tree view on the left with 'Job\_GenData' selected. The main panel shows the 'Details' for 'Job\_GenData'.

**Properties:**

- Name: Job\_GenData
- Short description:
- Created: CSD Apr 25, 2021, 12:44:50
- Last modified: CSD Apr 25, 2021, 13:28:16
- Last compiled: Apr 25, 2021, 14:03:42
- Job type: Parallel
- Multi-instance: No
- Information Services enabled: No
- Used by job(s): 0
- Depends on job(s): 0

**Latest Activity:**

- Last started: @adm Apr 25, 2021, 15:31:44
- Invocation ID:
- Status: Finished, Apr 25, 2021, 15:31:46
- Elapsed: 00:00:02
- CPU time: 00:00:00
- Rows in: 100 (50 per second)
- Rows out: 20 (10 per second)
- Messages: 15
- Queue: MediumPriorityJobs

**Job Runs:**

1 - 1 Page 1 of 1 Items Per Page: 25

Started	Finished	Invocation ID	Queue	Status	Elapsed	Rows In / Second	Rows Out / Second
Apr 25, 2021, 15:31:44	Apr 25, 2021, 15:31:46		MediumPriorityJobs	Finished	00:00:02	50	10

Engines Status: All Engines OK