

Setting up Cloud Dataflow in Eclipse

This page describes how to create a Dataflow project and run an example pipeline from within Eclipse.

The Dataflow Eclipse plugin works only with the Dataflow SDK distribution versions 2.0.0 to 2.5.0. The Dataflow Eclipse plugin does not work with the Apache Beam SDK distribution.

Before you begin

1. [Sign in](https://accounts.google.com/Login) (https://accounts.google.com/Login) to your Google Account.

If you don't already have one, [sign up for a new account](https://accounts.google.com/SignUp) (https://accounts.google.com/SignUp).

2. In the Cloud Console, on the project selector page, select or create a Cloud project.

★ **Note:** If you don't plan to keep the resources that you create in this procedure, create a project instead of selecting an existing project. After you finish these steps, you can delete the project, removing all resources associated with the project.

[Go to the project selector page](https://console.cloud.google.com/projectselector2/home/dashboard) (https://console.cloud.google.com/projectselector2/home/dashboard)

3. Make sure that billing is enabled for your Google Cloud project. [Learn how to confirm billing is enabled for your project](/billing/docs/how-to/modify-project) (/billing/docs/how-to/modify-project).

4. Enable the Cloud Dataflow, Compute Engine, Stackdriver Logging, Cloud Storage, Cloud Storage JSON, BigQuery, Cloud Pub/Sub, Cloud Datastore, and Cloud Resource Manager APIs.

[Enable the APIs](https://console.cloud.google.com/flows/enableapi?apiid=dataflow,compute_compo) (https://console.cloud.google.com/flows/enableapi?apiid=dataflow,compute_compo)

5. [Install and initialize the Cloud SDK](/sdk/docs) (/sdk/docs).

6. Ensure you have installed [Eclipse](https://eclipse.org/) (https://eclipse.org/) IDE version 4.7 or later.

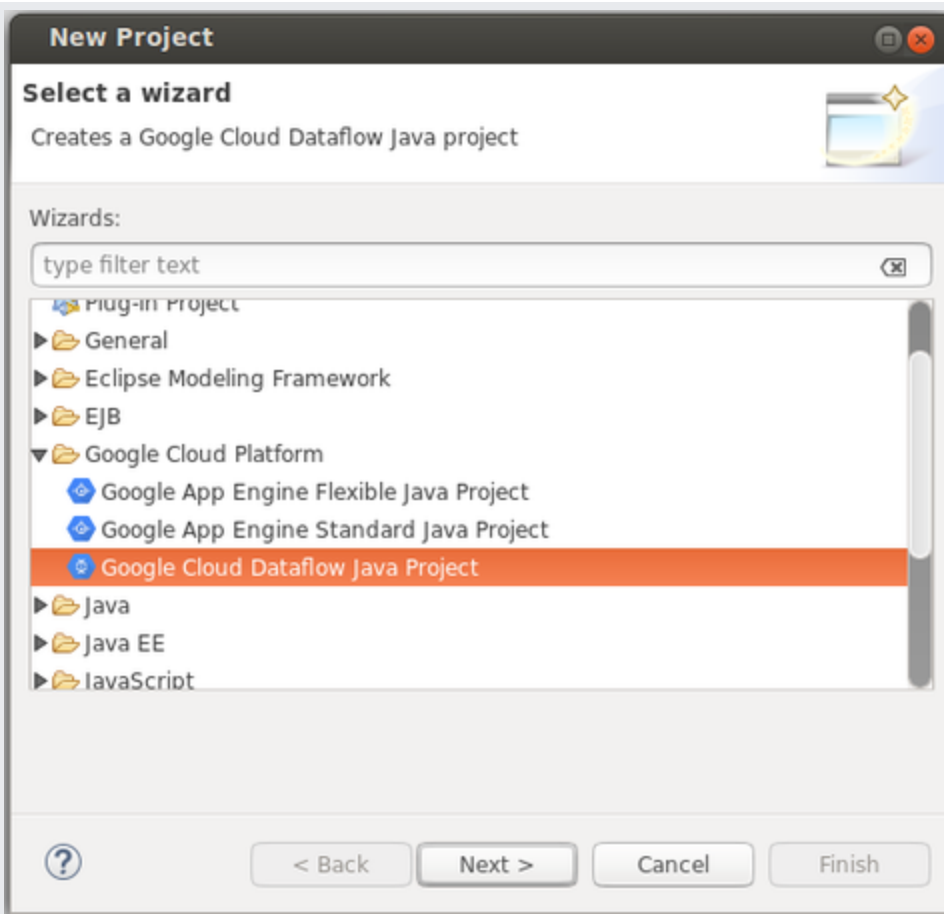
7. Ensure you have installed the [Java Development Kit \(JDK\)](http://www.oracle.com/technetwork/java/javase/downloads/index.html) (<http://www.oracle.com/technetwork/java/javase/downloads/index.html>) version 1.8 or later.
8. Ensure you have installed the latest version of the Cloud Tools for Eclipse plugin.
 - a. If you have not done so already, follow the [Cloud Tools for Eclipse Quickstart](/eclipse/docs/quickstart) (</eclipse/docs/quickstart>) to install the plugin.
 - b. Or, select **Help > Check for Updates** to update your plugin to the latest version.

Create a Dataflow project in Eclipse

To create a project, use the **New Project** wizard to generate a template application that you can use as the start for your own application.

If you don't have an application, you can run the **WordCount** sample app to complete the rest of these procedures.

1. Select **File -> New -> Project**.
2. In the **Google Cloud Platform** directory, select **Cloud Dataflow Java Project**.



3. Enter the **Group ID**.
4. Enter the **Artifact ID**.
5. Select the **Project Template**. For the **WordCount** sample, select **Example pipelines**.
6. Select the **Project Dataflow Version**. For the **WordCount** sample, select **2.5.0**.
7. Enter the **Package** name. For the **WordCount** sample, enter **com.google.cloud.dataflow.examples**.

Create a Cloud Dataflow Project
This wizard creates a new Google Cloud Dataflow project.

Group ID:

Artifact ID:

Project Template:

Project Dataflow Version:

Package:

Use default Workspace location

Location:

▶ Advanced

8. Click **Next**.

Configure execution options

You should now see the **Set Default Cloud Tools for Eclipse Run Options** dialog.

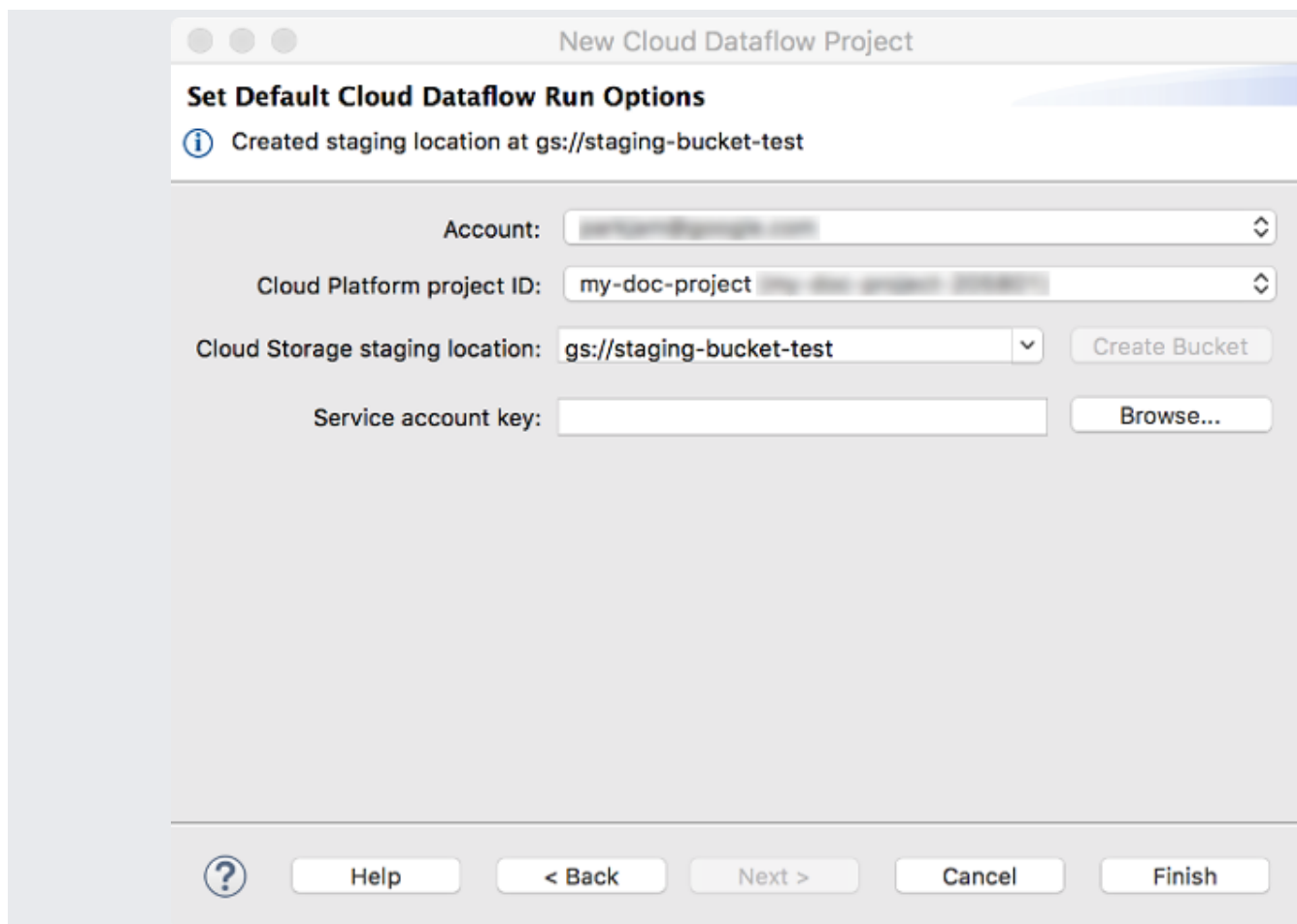
1. Select the account associated with your Google Cloud project or add a new account. To add a new account:
 - a. Select **Add a new account...** in the **Account** drop-down menu.
 - b. A new browser window opens to complete the sign-in process.
2. Enter your **Google Cloud Platform project ID**.
3. Select a **Cloud Storage staging location** or create a staging location. To create a staging location:
 - a. Enter a **unique name** for **Cloud Storage staging location**. Location name must include the bucket name and a folder. Objects are created in your Cloud Storage

bucket inside the specified folder. Do not include sensitive information in the bucket name because the bucket namespace is global and publicly visible.

⊕ See bucket naming requirements.

- Bucket names must contain only lowercase letters, numbers, dashes (-), underscores (_), and dots (.). Spaces are not allowed. Names containing dots require verification (/storage/docs/domain-name-verification).
- Bucket names must start and end with a number or letter.
- Bucket names must contain 3-63 characters. Names containing dots can contain up to 222 characters, but each dot-separated component can be no longer than 63 characters.
- Bucket names cannot be represented as an IP address in dotted-decimal notation (for example, 192.168.5.4).
- Bucket names cannot begin with the "goog" prefix.
- Bucket names cannot contain "google" or close misspellings, such as "g00gle".

b. Click **Create Bucket**.

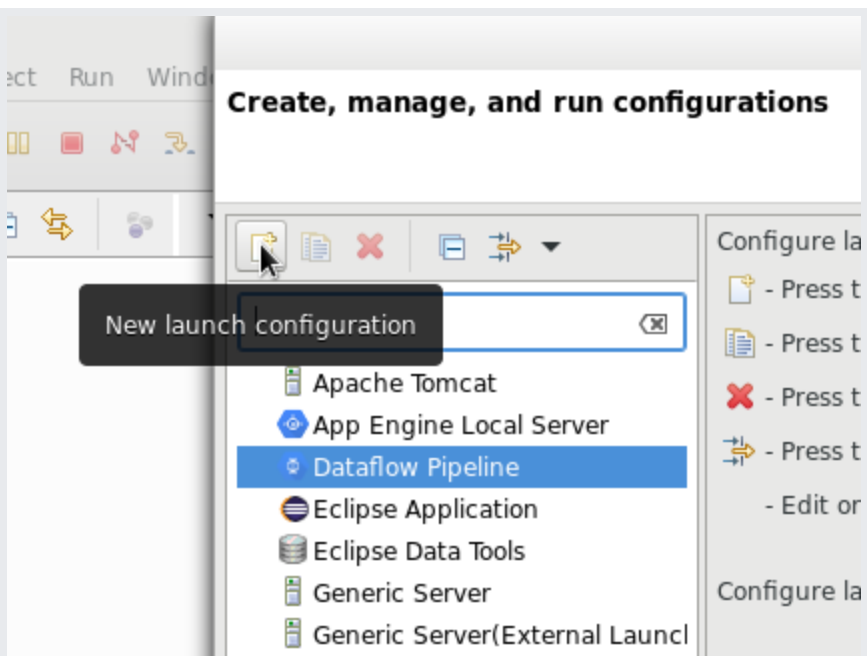


4. Click **Browse** to navigate to your service account key.
5. Click **Finish**.

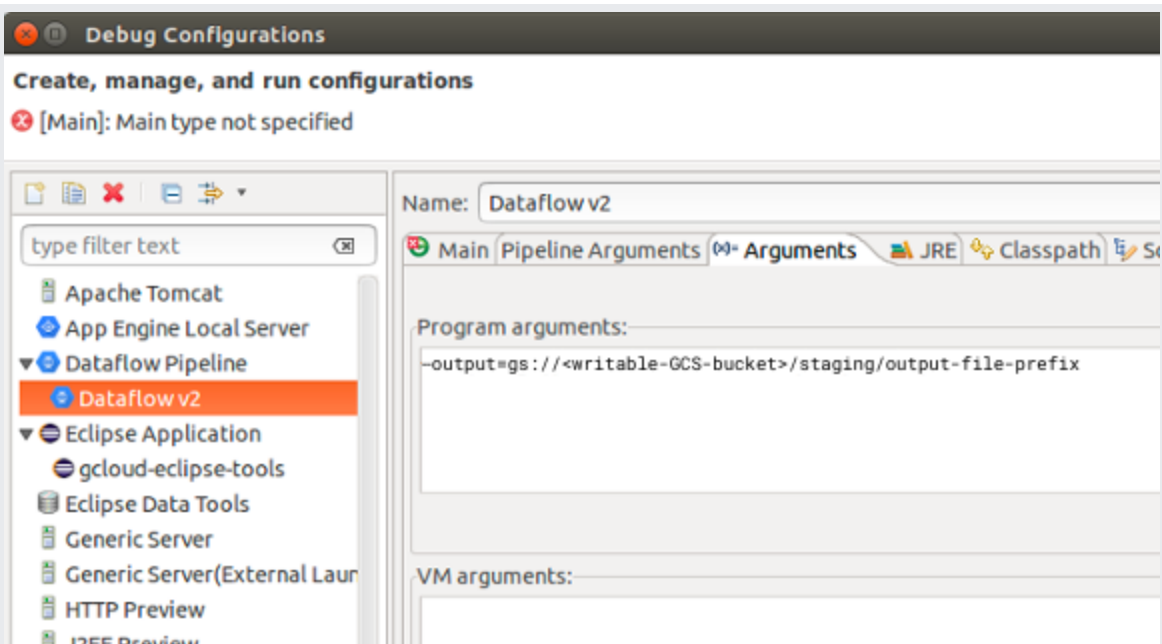
Run the WordCount example pipeline on the Dataflow service

After creating your Cloud Tools for Eclipse project, you can create pipelines that run on the Dataflow service. As an example, you can run the **WordCount** sample pipeline.

1. Select **Run -> Run Configurations**.
2. In the left menu, select **Dataflow Pipeline**.
3. Click **New Launch Configuration**.



4. Click the **Main** tab.
5. Click **Browse** to select your Dataflow project.
6. Click **Search...** and select the **WordCount Main Type**.
7. Click the **Pipeline Arguments** tab.
8. Select the **DataflowRunner** runner.
9. Click the **Arguments** tab.
10. In the **Program arguments** field, set the **output** to your **Cloud Storage Staging Location**.
The staging location must be a folder; you can't stage pipeline jobs from a bucket's root directory.



11. Click **Run**.

12. When the job finishes, among other output, you should see the following line in the Eclipse console:

```
Submitted job: <job_id>
```

Clean up

To avoid incurring charges to your Google Cloud account for the resources used in this quickstart, follow these steps.

1. Open the [Cloud Storage browser](https://console.cloud.google.com/storage/browser) (<https://console.cloud.google.com/storage/browser>) in the Google Cloud Console.
2. Select the checkbox next to the bucket that you created.
3. Click **DELETE**.
4. Click **Delete** to confirm that you want to permanently delete the bucket and its contents.

What's next

- [Read about the Dataflow Programming Model \(/dataflow/model/programming-model-beam\)](/dataflow/model/programming-model-beam).
- [Learn how to design and create your own pipeline \(/dataflow/pipelines/creating-a-pipeline-beam\)](/dataflow/pipelines/creating-a-pipeline-beam).
- [Work through the WordCount and Mobile Gaming examples \(/dataflow/examples/examples-beam\)](/dataflow/examples/examples-beam).

All rights reserved. Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2020-07-10 UTC.