This page describes how to import Cloud Spanner databases into Cloud Spanner using the Google Cloud Console. To import Avro files from another source, see Importing data from non-Cloud Spanner databases (/spanner/docs/import-non-spanner).

The process uses Dataflow (/dataflow/); it imports data from a Cloud Storage (/storage/) bucket folder that contains a set of Avro files (https://en.wikipedia.org/wiki/Apache_Avro) and JSON manifest files. The import process supports only Avro files exported from Cloud Spanner.

To import a Cloud Spanner database using the REST API or the `gcloud` command-line tool, complete the steps in the Before you begin (#before-you-begin) section on this page, then see the detailed instructions in Cloud Storage Avro to Cloud Spanner (/dataflow/docs/guides/templates/provided-batch#gcsavrotocloudspanner).

To import a Cloud Spanner database, first you need to enable the Cloud Spanner, Cloud Storage, Compute Engine, and Dataflow APIs:

Enable the APIs (https://console.cloud.google.com/flows/enableapi?apiid=spanner.googleapis.com,storage_component,compute,dataflow)

You also need enough quota and the required Cloud IAM permissions.

The quota requirements for import jobs, by Google Cloud service, are as follows:

- **Cloud Spanner**: You must have enough nodes to support the amount of data that you are importing. No additional nodes are required to import a database, though you might need to add more nodes so that your job finishes in a reasonable amount of time. See Optimizing jobs (#optimize-slow) for more details.

- **Cloud Storage**: To import, you must have a bucket containing your previously exported files. You do not need to set a size for your bucket.

- **Dataflow**: Import jobs are subject to the same CPU, disk usage, and IP address Compute Engine quotas (/dataflow/quotas#compute-engine-quotas) as other Dataflow jobs.

- **Compute Engine**: Before running your import job, you must set up initial quotas (https://support.google.com/cloud/answer/6075746) for Compute Engine, which Dataflow uses. These quotas represent the *maximum* number of resources that you allow Dataflow to use for your job. Recommended starting values are:

  - **CPUs**: 200

  - **In-use IP addresses**: 200

  - **Standard persistent disk**: 50 TB

  Generally, you do not have to make any other adjustments. Dataflow provides autoscaling so that you only pay for the actual resources used during the import. If your job can make use of more resources, the Dataflow UI displays a warning icon. The job should finish even if there is a warning icon.

To import a database, you also need to have Cloud IAM roles with sufficient permissions to use all of the services involved in an import job. For information on granting roles and permissions, see Applying IAM roles (/spanner/docs/grant-permissions).

To import a database, you need the following roles:

- At the Google Cloud project level:

  - Cloud Spanner Viewer

  - Dataflow Admin

  - Storage Admin

- At the Cloud Spanner database or instance level, or at the Google Cloud project level:

  - Cloud Spanner Reader

  - Cloud Spanner Database Admin (required only for import jobs)

To find the folder that contains your exported database in the Cloud Console, navigate to the Cloud Storage browser and click on the bucket that contains the exported folder.

Go to the Cloud Storage browser (https://console.cloud.google.com/storage/browser)

The name of the folder that contains your exported data begins with your instance's ID, database name, and the timestamp of your export job. The folder contains:

- A `spanner-export.json` file.

- A `TableName-manifest.json` file for each table in the database you exported.

- One or more `TableName.avro-#####-of-#####` files. The first number in the extension `.avro-#####-of-#####` represents the index of the Avro file, starting at zero, and the second represents the number of Avro files generated for each table.

  For example, `Songs.avro-00001-of-00002` is the second of two files that contain the data for the `Songs` table.

To import your Cloud Spanner database from Cloud Storage to your instance, follow these steps.

1. Go to the Cloud Spanner **Instances** page.

   Go to the instances page (https://console.cloud.google.com/spanner/instances)

2. Click the name of the instance that will contain the imported database.

3. Click the **Import/Export** tab and then click the **Import** button.

4. Under **Choose a source folder**, click **Browse**.

5. Find the bucket that contains your export in the initial list, or click **Search**

   🔍

   to filter the list and find the bucket. Double-click the bucket to see the folders it contains.

6. Find the folder with your exported files and click to select it.

★   **Note:** Be sure to select the folder created by the export job and not a higher-level folder that contains the exported folder.

7. Click **Select**.

8. Enter a name for the new database, which Cloud Spanner creates during the import process. The database name cannot already exist in your instance.

9. Select a region in the **Choose a region for the import job** drop-down menu.

★   **Note:** To avoid network egress charges (/storage/pricing#network-pricing), choose a region that overlaps with your Cloud Storage bucket's location. See Choosing a region (#choose-region) below for more information.

10. Select the checkbox under **Confirm charges** to acknowledge that there are charges in addition to those incurred by your existing Cloud Spanner nodes.

11. Click **Import**.

    The Cloud Console displays the **Database details** page, which now shows a box describing your import job, including the job's elapsed time:

    ⚠ Importing data from Google Cloud Storage into example-target-db.   Stop import
       Elapsed time: 41 sec. View job details in Cloud Dataflow

When the job finishes or terminates, the Cloud Console displays a message on the **Database details** page. If the job succeeds, a success message appears:

✓ Import from Google Cloud Storage into example-target-db complete.

After the Dataflow import job successfully finishes, Cloud Spanner creates indexes for your imported database. When index creation is still in ...ss, the Cloud Console shows an in-progress icon to indicate that a long-running operation is occurring. The icon is next to the index's name, in ...ase hierarchy to the left of the **Database details** page. When the in-progress icon changes to the index icon

...ion of that index is complete.

If the job does not succeed, a failure message appears:

🚫 Import failed. For more information check the Cloud Dataflow job logs. To retry your import, delete   **Delete database**
    this database and start over with a new import job. Learn more

If your job fails, check the job's Dataflow logs (#dataflow-job-logs) for error details.

You might want to choose a different region based on whether your Cloud Storage bucket uses a regional or multi-regional configuration. To avoid network egress charges (/storage/pricing#network-pricing), choose a region that overlaps with your Cloud Storage bucket's location.

If your Cloud Storage bucket location is regional (/storage/docs/bucket-locations#location-r), choose the same region for your import job if that region is available to take advantage of free network usage (/storage/pricing#network-buckets).

If the same region is not available, egress charges will apply. Refer to the Cloud Storage network egress (/storage/pricing#network-pricing) pricing to choose a region that will incur the lowest network egress charges.

If your Cloud Storage bucket location is multi-regional (/storage/docs/bucket-locations#location-mr), choose one of the regions that make up the multi-regional location to take advantage of free network usage (/storage/pricing#network-buckets).

If an overlapping region is not available, egress charges will apply. Refer to the Cloud Storage network egress (/storage/pricing#network-pricing) pricing to choose a region that will incur the lowest network egress charges.

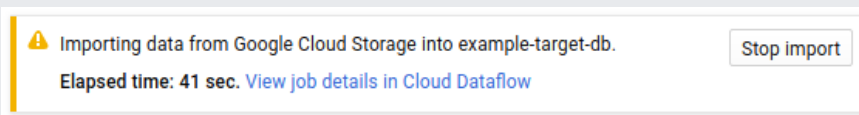After you start an import job, you can view details of the job, including logs, in the Dataflow section of the Cloud Console.

To see details for a currently running job:

1. Navigate to the **Database details** page for the database.

2. Click **View job details in Dataflow** in the job status message, which looks similar to the following:

The Cloud Console displays details of the Dataflow job.

To view a job that you ran recently:

1. Navigate to the **Database details** page for the database.

2. Click the **Import/Export** tab.

3. Click your job's name in the list.

   The Cloud Console displays details of the Dataflow job.

To view a job that you ran more than one week ago:

1. Go to the Dataflow jobs page in the Cloud Console.

Go to the jobs page (https://console.cloud.google.com/dataflow)

2. Find your job in the list, then click its name.

   The Cloud Console displays details of the Dataflow job.

Jobs of the same type for the same database have the same name. You can tell jobs apart by the values in their **Start time** or **End time** column

To view a Dataflow job's logs, navigate to the job's details page as described above, then click **Logs** to the right of the job's name.

If a job fails, look for errors in the logs. If there are errors, the error count displays next to **Logs**:

❗ 5   ☰ **LOGS**
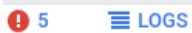
To view job errors:

1. Click on the error count next to **Logs**.

   The Cloud Console displays the job's logs. You may need to scroll to see the errors.

2. Locate entries with the error icon
   ‼️
   .

3. Click on an individual log entry to expand its contents.

For more information about troubleshooting Dataflow jobs, see Troubleshooting your pipeline
(/dataflow/pipelines/troubleshooting-your-pipeline#basic-troubleshooting-workflow).

If you have followed the suggestions in initial settings (#quota), you should generally not have to make any other adjustments. If your job is running slowly, there are a few other optimizations you can try:

- **Optimize the job and data location**: Run your Dataflow job in the same region (#choose-region) where your Cloud Spanner instance and Cloud Storage bucket are located.

- **Ensure sufficient Dataflow resources**: If the relevant Compute Engine quotas (/dataflow/quotas#compute-engine-quotas) limit your Dataflow job's resources, the job's Dataflow page (#dataflow-job-details) in the Google Cloud Console displays a warning icon
  ⚠️ 4
  and log messages:

  > ⚠️ 2018-06-28 (17:39:14) Autoscaling: Unable to reach resize target in zone us-central1-f. QUOTA_EXCEEDED: Quota 'IN_USE_ADDR...

  In this situation, increasing the quotas (https://support.google.com/cloud/answer/6075746) for CPUs, in-use IP addresses, and standard persistent disk might shorten the run time of the job, but you might incur more Compute Engine charges.

- **Check the Cloud Spanner CPU utilization**: If you see that the CPU utilization for the instance is over 65%, you can increase the number of nodes in that instance. The extra nodes add more Cloud Spanner resources and the job should

speed up, but you incur more Cloud Spanner charges.

Several factors influence the time it takes to complete an import job.

- **Cloud Spanner database size**: Processing more data takes more time and resources.

- **Cloud Spanner database schema (including indexes)**: The number of tables, the size of the rows, and the number of secondary indexes influence the time it takes to run an import job. Note that index creation continues after the Dataflow import job completes.

- **Data location**: Data is transferred between Cloud Spanner and Cloud Storage using Dataflow. Ideally all three components are located in the same region. If the components are not in the same region, moving the data across regions slows the job down.

- **Number of Dataflow workers**: By using autoscaling, Dataflow chooses the number of workers for the job depending on the amount of work that needs to be done. The number of workers will, however, be capped by the quotas for CPUs, in-use IP addresses, and standard persistent disk. The Dataflow UI displays a warning icon if it encounters quota caps. In this situation, progress is slower, but the job should still complete.

- **Existing load on Cloud Spanner**: An import job adds significant CPU load on a Cloud Spanner instance. If the instance already has a substantial existing load, then the job runs more slowly.

- **Number of Cloud Spanner nodes**: If the CPU utilization for the instance is over 65%, then the job runs more slowly.