

Loading CSV data from Cloud Storage

Loading CSV files from Cloud Storage

When you load CSV data from Cloud Storage, you can load the data into a new table or partition, or you can append to or overwrite an existing table or partition. When your data is loaded into BigQuery, it is converted into columnar format for [Capacitor](#)

(<https://cloud.google.com/blog/big-data/2016/04/inside-capacitor-bigquerys-next-generation-columnar-storage-format>)

(BigQuery's storage format).

When you load data from Cloud Storage into a BigQuery table, the dataset that contains the table must be in the same regional or multi-regional location as the Cloud Storage bucket.

For information about loading CSV data from a local file, see [Loading data into BigQuery from a local data source](#) (/bigquery/docs/batch-loading-data#loading_data_from_local_files).

Try it for yourself

If you're new to Google Cloud, create an account to evaluate how BigQuery performs in real-world scenarios. New customers also get \$300 in free credits to run, test, and deploy workloads.

[Try BigQuery free \(https://console.cloud.google.com/freetrial\)](https://console.cloud.google.com/freetrial)

Limitations

When you load CSV files into BigQuery, the following

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

issues.

ding

Query is slower than loading uncompressed data. See

Loading compressed and uncompressed data

(/bigquery/docs/batch-loading-data#loading_compressed_and_uncompressed_data).

- You cannot include both compressed and uncompressed files in the same load job.
- The maximum size for a gzip file is 4 GB.
- When you load CSV or JSON data, values in DATE columns must use the dash (-) separator and the date must be in the following format: YYYY-MM-DD (year-month-day).
- When you load JSON or CSV data, values in TIMESTAMP columns must use a dash (-) separator for the date portion of the timestamp, and the date must be in the following format: YYYY-MM-DD (year-month-day). The hh:mm:ss (hour-minute-second) portion of the timestamp must use a colon (:) separator.

Required permissions

When you load data into BigQuery, you need permissions to run a load job and permissions that let you load data into new or existing BigQuery tables and partitions. If you are loading data from Cloud Storage, you also need permissions to access to the bucket that contains your data.

BigQuery permissions

At a minimum, the following permissions are required to load data into BigQuery. These permissions are required if you are loading data into a new table or partition, or if you are appending or overwriting a table or partition.

- `bigquery.tables.create`
- `bigquery.tables.updateData`
- `bigquery.jobs.create`

The following predefined IAM roles include both `bigquery.tables.create` and `big`

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

The following predefined IAM roles include `bigquery.jobs.create` permissions:

- `bigquery.user`
- `bigquery.jobUser`
- `bigquery.admin`

In addition, if a user has `bigquery.datasets.create` permissions, when that user creates a dataset, they are granted `bigquery.dataOwner` access to it. `bigquery.dataOwner` access lets the user create and update tables in the dataset by using a load job.

For more information on IAM roles and permissions in BigQuery, see [Access control](#) (`/bigquery/access-control`).

Cloud Storage permissions

To load data from a Cloud Storage bucket, you must be granted `storage.objects.get` permissions. If you are using a URI [wildcard](#) (`/bigquery/docs/batch-loading-data#load-wildcards`), you must also have `storage.objects.list` permissions.

The predefined IAM role [storage.objectViewer](#) (`/storage/docs/access-control/iam`) can be granted to provide both `storage.objects.get` and `storage.objects.list` permissions.

Loading CSV data into a table

You can load CSV data from Cloud Storage into a new BigQuery table by:

- Using the Cloud Console
- Using the bq command-line tool's `bq load` command
- Calling the `jobs.insert` API method and configuring a load job

To

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

Con

[More details](#) [OK](#)

[Python](#) (#python

For step-by-step guidance on this task directly in Cloud Shell Editor, click **Guide me**:

[Guide me](https://console.cloud.google.com/?tutorial=bigquery_import_data_from_cloud_storage) (https://console.cloud.google.com/?tutorial=bigquery_import_data_from_cloud_storage)


The following sections take you through the same steps as clicking **Guide me**.

1. In the Cloud Console, open the BigQuery page.

[Go to BigQuery](https://console.cloud.google.com/bigquery) (https://console.cloud.google.com/bigquery)

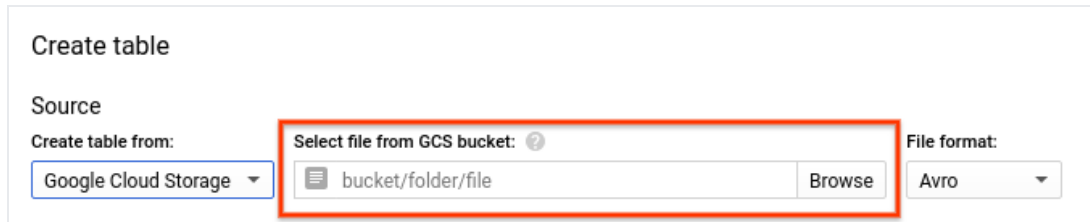
2. In the **Explorer** panel, expand your project and select a dataset.

3. Expand the **Actions** option and click **Open**.

4. In the details panel, click **Create table** .

5. On the **Create table** page, in the **Source** section:

- For **Create table from**, select Cloud Storage.
- In the source field, browse to or enter the [Cloud Storage URI](#) (/bigquery/docs/batch-loading-data#gcs-uri). Note that you cannot include multiple URIs in the Cloud Console, but [wildcards](#) (/bigquery/docs/batch-loading-data#load-wildcards) are supported. The Cloud Storage bucket must be in the same location as the dataset that contains the table you're creating.



Create table

Source

Create table from: Google Cloud Storage

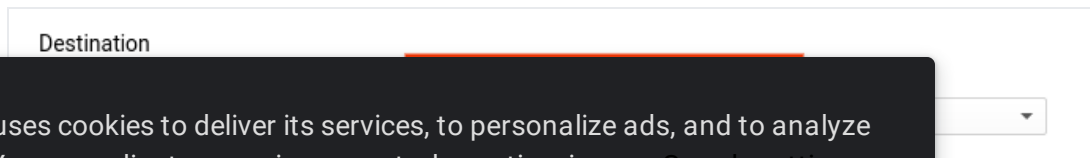
Select file from GCS bucket: Browse

File format: Avro

- For **File format**, select **CSV**.

6. On the **Create table** page, in the **Destination** section:

- For **Dataset name**, choose the appropriate dataset.



Destination

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) **OK**

ery.

7. In the **Schema** section, for **Auto detect**, check **Schema and input parameters** to enable schema [auto detection](/bigquery/docs/schema-detect) (/bigquery/docs/schema-detect). Alternatively, you can manually enter the [schema](/bigquery/docs/schemas) (/bigquery/docs/schemas) definition by:

- Enabling **Edit as text** and entering the table schema as a JSON array.

```
[
  {
    "description": "quarter",
    "mode": "REQUIRED",
    "name": "qtr",
    "type": "STRING"
  },
  {
    "description": "sales representative",
    "mode": "NULLABLE",
    "name": "rep",
    "type": "STRING"
  },
  {
    "description": "total sales",
    "mode": "NULLABLE",
    "name": "sales",
    "type": "INTEGER"
  }
]
```

★ **Note:** You can view the schema of an existing table in JSON format by entering the following command in the **bq** command-line tool: `bq show --format=prettyjson dataset.table`.

- Using **Add field** to manually input the schema.

Name	Type	Mode	
qtr	STRING	REQUIRED	X
+ Add field			

8. (Optional) To partition the table, choose your options in the **Partition and cluster settings**. For more information, see [Creating partitioned tables](/bigquery/docs/creating-partitioned-tables) (/bigquery/docs/creating-partitioned-tables)

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) **OK**

users to
filter may
on tables
is option is

10. (Optional) To cluster (/bigquery/docs/creating-clustered-tables) the table, in the **Clustering order** box, enter between one and four field names.

11. (Optional) Click **Advanced options**.

- For **Write preference**, leave **Write if empty** selected. This option creates a new table and loads your data into it.
- For **Number of errors allowed**, accept the default value of 0 or enter the maximum number of rows containing errors that can be ignored. If the number of rows with errors exceeds this value, the job will result in an **invalid** message and fail.
- For **Unknown values**, check **Ignore unknown values** to ignore any values in a row that are not present in the table's schema.
- For **Field delimiter**, choose the character that separates the cells in your CSV file: **Comma**, **Tab**, **Pipe**, or **Custom**. If you choose **Custom**, enter the delimiter in the **Custom field delimiter** box. The default value is **Comma**.
- For **Header rows to skip**, enter the number of header rows to skip at the top of the CSV file. The default value is 0.
- For **Quoted newlines**, check **Allow quoted newlines** to allow quoted data sections that contain newline characters in a CSV file. The default value is **false**.
- For **Jagged rows**, check **Allow jagged rows** to accept rows in CSV files that are missing trailing optional columns. The missing values are treated as nulls. If unchecked, records with missing trailing columns are treated as bad records, and if there are too many bad records, an invalid error is returned in the job result. The default value is **false**.
- For **Encryption**, click **Customer-managed key** to use a Cloud Key Management Service key (/bigquery/docs/customer-managed-encryption). If you leave the **Google-managed key** setting, BigQuery encrypts the data at rest (/security/encryption/default-encryption).

12. Click **Create table**.

Note: When you load data into an empty table by using the Cloud Console, you cannot add a label, description, table expiration, or partition expiration.

After the table is created, you can update the table's expiration, description, and labels, but you cannot add a partition expiration after a table is created using the Cloud Console. For more information, see Managing tables.

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your Google settings.

[More details](#) **OK**

Loading CSV data into a table that uses column-based time partitioning

To load CSV data from Cloud Storage into a BigQuery table that uses column-based time partitioning:

[Go](#) [Java](#) (#java) [Node.js](#) (#node.js) [Python](#) (#python) [Go](#) (#go)

Before trying this sample, follow the Go setup instructions in the [BigQuery quickstart using client libraries](#) (/bigquery/docs/quickstarts/quickstart-client-libraries). For more information, see the [BigQuery Go API reference documentation](#) (<https://godoc.org/cloud.google.com/go/bigquery>).

4a0939603bb9f0a74eb2b815aa6577/bigquery/snippets/loadingdata/bigquery_load_table_partitioned.go)

```
import (  
    "context"  
    "fmt"  
    "time"  
  
    "cloud.google.com/go/bigquery"  
)  
  
// importPartitionedTable demonstrates specifying time partitioning for a BigQu  
// CSV data from Cloud Storage.  
func importPartitionedTable(projectID, destDatasetID, destTableID string) erro  
    // projectID := "my-project-id"  
    // datasetID := "mydataset"  
    // tableID := "mytable"  
    ctx := context.Background()  
    client, err := bigquery.NewClient(ctx, projectID)  
    if err != nil {  
        return fmt.Errorf("bigquery.NewClient: %v", err)  
    }  
}
```

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

obr", Type: bigquery.StringFieldType},

```

        {Name: "date", Type: bigquery.DateFieldType},
    }
    loader := client.Dataset(destDatasetID).Table(destTableID).LoaderFrom(
    loader.TimePartitioning = &bigquery.TimePartitioning{
        Field:      "date",
        Expiration: 90 * 24 * time.Hour,
    }
    loader.WriteDisposition = bigquery.WriteEmpty

    job, err := loader.Run(ctx)
    if err != nil {
        return err
    }
    status, err := job.Wait(ctx)
    if err != nil {
        return err
    }

    if status.Err() != nil {
        return fmt.Errorf("job completed with error: %v", status.Err())
    }
    return nil
}

```

Appending to or overwriting a table with CSV data

You can load additional data into a table either from source files or by appending query results.

In the Cloud Console, use the **Write preference** option to specify what action to take when you load data from a source file or from a query result.

You have the following options when you load additional data into a table:

Console

BigQuery API

opti

Writ

emp

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

(Default

to table -- (/bigquery/docs/reference/rest/v2/Job#JobConfigurationLoa
replace=false;) Appends the data to the end of the table.
if --
[no] replace is
unspecified, the
default is append

Overwrite--replace or --WRITE_TRUNCATEErases all existing data in a table before writing the new data. Th
table replace=true table schema and removes any Cloud KMS key.

If you load data into an existing table, the load job can append the data or overwrite the table.

You can append or overwrite a table by:

- Using the Cloud Console
- Using the bq command-line tool's bq load command
- Calling the jobs.insert API method and configuring a load job
- Using the client libraries

This page does not cover appending or overwriting partitioned tables. For information on appending and
riting partitioned tables, see: [Appending to and overwriting partitioned table data](#)
query/docs/managing-partitioned-table-data#append-overwrite).

[Console](#) ([#console](#)) [bq](#) ([#bq](#)) [API](#) ([#api](#)) [Go](#) ([#go](#)) [Java](#) ([#java](#)) [Node.js](#) ([#node.js](#)) [Python](#) ([#python](#))

1. In the Cloud Console, open the BigQuery page.

[Go to BigQuery](https://console.cloud.google.com/bigquery) (https://console.cloud.google.com/bigquery)

2. In the **Explorer** panel, expand your project and select a dataset.

3. Expand the **Actions** option and click **Open**

Google uses cookies to deliver its services, to personalize ads, and to analyze
traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

- In the source field, browse to or enter the [Cloud Storage URI](#) (#gcs-uri). Note that you cannot include multiple URIs in the Cloud Console, but [wildcards](#) (/bigquery/docs/batch-loading-data#load-wildcards) are supported. The Cloud Storage bucket must be in the same location as the dataset that contains the table you're appending or overwriting.

Create table

Source

Create table from: Google Cloud Storage

Select file from GCS bucket: ? Browse

File format: Avro

- For **File format**, select **CSV**.

6. On the **Create table** page, in the **Destination** section:

- For **Dataset name**, choose the appropriate dataset.

Destination

Project name example-project

Dataset name example_dataset

Table type ? Native table

- In the **Table name** field, enter the name of the table you're appending or overwriting in BigQuery.
- Verify that **Table type** is set to **Native table**.

7. In the **Schema** section, for **Auto detect**, check **Schema and input parameters** to enable schema [auto detection](#) (/bigquery/docs/schema-detect). Alternatively, you can manually enter the [schema](#) (/bigquery/docs/schemas) definition by:

- Enabling **Edit as text** and entering the table schema as a JSON array.

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) OK

```
[
  {
    "description": "quarter",
    "mode": "REQUIRED",
    "name": "qtr",
    "type": "STRING"
  },
  {
    "description": "sales representative",
    "mode": "NULLABLE",
    "name": "rep",
    "type": "STRING"
  },
  {
    "description": "total sales",
    "mode": "NULLABLE",
    "name": "sales",
    "type": "INTEGER"
  }
]
```

★ **Note:** You can view the schema of an existing table in JSON format by entering the following command: `bq show --format=prettyjson dataset.table`.

- Using **Add field** to manually input the schema.

Name	Type	Mode	
qtr	STRING	REQUIRED	X
+ Add field			

★ **Note:** It is possible to modify the table's schema when you append or overwrite it. For more information on supported schema changes during a load operation, see [Modifying table schemas](#) (/bigquery/docs/managing-table-schemas).

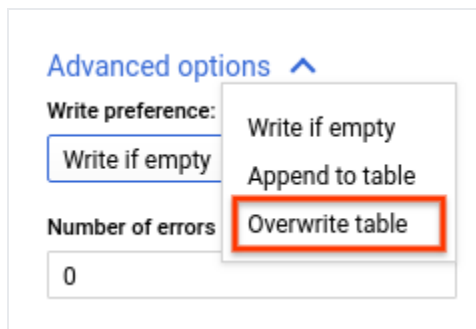
8. For **Partition and cluster settings**, leave the default values. You cannot convert a table to a partitioned or clustered table by appending or overwriting it, and the Cloud Console does not support appending to or overwriting partitioned or clustered tables in a load job.

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) **OK**

Maximum
s with errors

- For **Unknown values**, check **Ignore unknown values** to ignore any values in a row that are not present in the table's schema.
- For **Field delimiter**, choose the character that separates the cells in your CSV file: **Comma**, **Tab**, **Pipe**, or **Custom**. If you choose **Custom**, enter the delimiter in the **Custom field delimiter** box. The default value is **Comma**.
- For **Header rows to skip**, enter the number of header rows to skip at the top of the CSV file. The default value is **0**.
- For **Quoted newlines**, check **Allow quoted newlines** to allow quoted data sections that contain newline characters in a CSV file. The default value is **false**.
- For **Jagged rows**, check **Allow jagged rows** to accept rows in CSV files that are missing trailing optional columns. The missing values are treated as nulls. If unchecked, records with missing trailing columns are treated as bad records, and if there are too many bad records, an invalid error is returned in the job result. The default value is **false**.
- For **Encryption**, click **Customer-managed key** to use a [Cloud Key Management Service key](#) (/bigquery/docs/customer-managed-encryption). If you leave the **Google-managed key** setting, BigQuery [encrypts the data at rest](#) (/security/encryption/default-encryption).



10. Click **Create table**.

Loading hive-partitioned CSV data

BigQuery supports loading hive-partitioned CSV data stored on Cloud Storage and will populate the table with the data. For more information, see [Loading hive-partitioned CSV data](#). For more information, see [Loading CSV data](#).

Google uses cookies to deliver its services, to personalize ads, and to analyze usage. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) **OK**

This section describes how BigQuery handles various CSV formatting options.

Encoding

BigQuery expects CSV data to be UTF-8 encoded. If you have CSV files with data encoded in ISO-8859-1 (also known as Latin-1) format, you should explicitly specify the encoding so that BigQuery can properly convert the data to UTF-8.

If you don't specify an encoding, or if you specify UTF-8 encoding when the CSV file is not UTF-8 encoded, BigQuery attempts to convert the data to UTF-8. Generally, your data will be loaded successfully, but it may not match byte-for-byte what you expect. To avoid this, specify the correct encoding by using the `--encoding flag` (`#csv-options`).

By default, if the CSV file contains the ASCII `0` (NULL) character, you can't load the data into BigQuery. If you want to load CSV files containing ASCII `0` and/or other ASCII Control characters, please contact BigQuery customer support.

If BigQuery can't convert a character other than the ASCII `0` character, BigQuery converts the character to the standard Unicode replacement character: `U+FFFD`.

Field delimiters

Delimiters in CSV files can be any single-byte character. If the source file uses ISO-8859-1 encoding, any character can be a delimiter. If the source file uses UTF-8 encoding, any character in the decimal range 1-127 (U+0001-U+007F) can be used without modification. You can insert an ISO-8859-1 character outside of this range as a delimiter, and BigQuery will interpret it correctly. However, if you use a multibyte character as a delimiter, some of the bytes will be interpreted incorrectly as part of the field value.

Generally, it's a best practice to use a standard delimiter, such as a tab, pipe, or comma. The default is a comma.

Da

Boo

or f

(/b

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

true or false, t

[More details](#) [OK](#)

1.

Bytes. Columns with BYTES types must be encoded as Base64.

Date. Columns with DATE types must be in the form YYYY-MM-DD.

Datetime. Columns with DATETIME types must be in the form YYYY-MM-DD HH:MM:SS[.SSSSSS].

Time. Columns with TIME types must be in the form HH:MM:SS[.SSSSSS].

Timestamp. BigQuery accepts a variety of timestamp formats. The timestamp must include a date portion and a time portion.

- The date portion can be formatted as YYYY-MM-DD or YYYY/MM/DD.
- The timestamp portion must be formatted as HH:MM[:SS[.SSSSSS]] (seconds and fractions of seconds are optional).
- The date and time must be separated by a space or 'T'.
- Optionally, the date and time can be followed by a UTC offset or the UTC zone designator (Z). For more information, see [Time zones](#) (/bigquery/docs/reference/standard-sql/data-types#time_zones).

For example, any of the following are valid timestamp values:

- 2018-08-19 12:11
- 2018-08-19 12:11:35
- 2018-08-19 12:11:35.22
- 2018/08/19 12:11
- 2018-07-05 12:54:00 UTC
- 2018-08-19 07:11:35.220 -05:00
- 2018-08-19T12:11:35.220Z

If you provide a schema, BigQuery also accepts Unix Epoch time for timestamp values.

How... as a numeric or
stri...

Exa...

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

- 1.534680695e11

CSV options

To change how BigQuery parses CSV data, specify additional options in the Cloud Console, the `bq` command-line tool, or the API.

For more information on the CSV format, see [RFC 4180](https://tools.ietf.org/html/rfc4180) (<https://tools.ietf.org/html/rfc4180>).

CSV option	Console option	bq tool flag	BigQuery API property
Field delimiter	Field delimiter: Comma, Tab, Pipe, Custom	<code>-F</code> or <code>--field_delimiter</code>	<code>fieldDelimiter</code> (Java) (Python)

Header rows	Header rows to skip	<code>--skip_leading_rows</code>	<code>skipLeadingRows</code> (Java) (Python)
-------------	---------------------	----------------------------------	---

Num
bad
recd
allo

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

Newline Allow -- **allowQuotedNewlines** ([Java](#)
charactersquoted **allow_quoted_newlines** (<https://googleapis.dev/java/google-cloud-bigquery/latest/c>
newlines (<https://googleapis.dev/python/bigquery/latest/generated/c>

Custom None --**null_marker** **nullMarker** ([Java](#)
null values (<https://googleapis.dev/java/google-cloud-bigquery/latest/c>
(<https://googleapis.dev/python/bigquery/latest/generated/c>

Trail
opti
colu

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

Unknown Ignore -- [ignoreUnknownValues \(Java](#)
values [unknownignore_unknown_values \(https://googleapis.dev/java/google-cloud-bigquery/latest/c](#)
values [Python](#)
([https://googleapis.dev/python/bigquery/latest/generated/c](#)
)

Quote None --quote [quote \(Java \(https://googleapis.dev/java/google-cloud-bigq](#)
([https://googleapis.dev/python/bigquery/latest/generated/c](#)

Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your [Google settings](#).

[More details](#) [OK](#)

Encoding None -E or --encoding

encoding ([Java](https://googleapis.dev/java/google-cloud-b) (<https://googleapis.dev/java/google-cloud-b>)
[Python](https://googleapis.dev/python/bigquery/latest/gene) (<https://googleapis.dev/python/bigquery/latest/gene>

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the [Apache 2.0 License](#) ([https://www.apache.org/licenses/LICENSE-2.0](#)). [Privacy Policies](#)

([https://www.google.com/policies/](#)) and our [affiliates](#).

Last updated: [Google uses cookies to deliver its services, to personalize ads, and to analyze traffic. You can adjust your privacy controls anytime in your \[Google settings\]\(#\).](#)

Last

[More details](#) [OK](#)