

Deleting Rows from a Table

This article summarizes options for deleting row-level data from an Treasure Data table. The current best practice is to use the Presto DELETE command.

The older Partial Delete method is documented for customers who are still using it but it is not recommended.

Limitations

The Presto DELETE statement has several known limitations. Review [Presto DELETE limitations](#) for more information.

Continue to these tasks.

- [Prerequisites](#)
- [Options for Row Delete](#)
 - [Using Presto DELETE Statements](#)
 - [Partial Delete of Data in a Time Range](#)

Prerequisites

- Basic knowledge of Treasure Data, including the TD [Toolbelt](#)
- Understanding of Presto or Hive

Options for Row Delete

As you use the procedures in this article, refer to the following articles:

- [Presto DELETE Statement Syntax](#).
- [Partial Delete](#): Efficient for small tables because it reloads the entire table
- [Partial Delete based on TIME Column](#): Efficient for huge tables. You take a superset of rows, which includes rows to be deleted and filtered based on the `time` column.

Using Presto DELETE Statements

Presto Delete enables you to issues DELETE statement queries against any table in Treasure Data.

Use the following syntax to isolate the row you want to delete and delete it:

```
DELETE FROM table_name [ WHERE condition ]
```

For example:

```
DELETE FROM emp WHERE name = 'beekeeper1'
```

Partial Delete of Data in a Time Range

TD previously offered and still supports a technique known as "partial delete" that deleted all rows in a specific time range by dropping the partitions covering that time range. This technique is no longer recommended because it is more complex and more likely to lead to user error. Customers should use Presto DELETE instead.

One way to use this feature to delete a subset of rows was to:

1. Extract rows in the desired time range into a temporary table using:

```
CREATE TABLE saved_rows_tbl AS SELECT * FROM _source_table_ WHERE _condition_ = 0 AND td_time_range(. . .)
```

2. Delete the whole time range using the partial delete command in TD Toolbelt:

```
td table:partial_delete example_db source_table --from _start_time_ --to _end_time_
```

You can also [delete imported data within a specific table](#).

3. Re-insert the saved rows into the table

```
INSERT INTO source_table SELECT * FROM saved_rows_table
```

4. Delete the saved_rows_table once you have confirmed that the result of the INSERT INTO has restored the desired rows.