

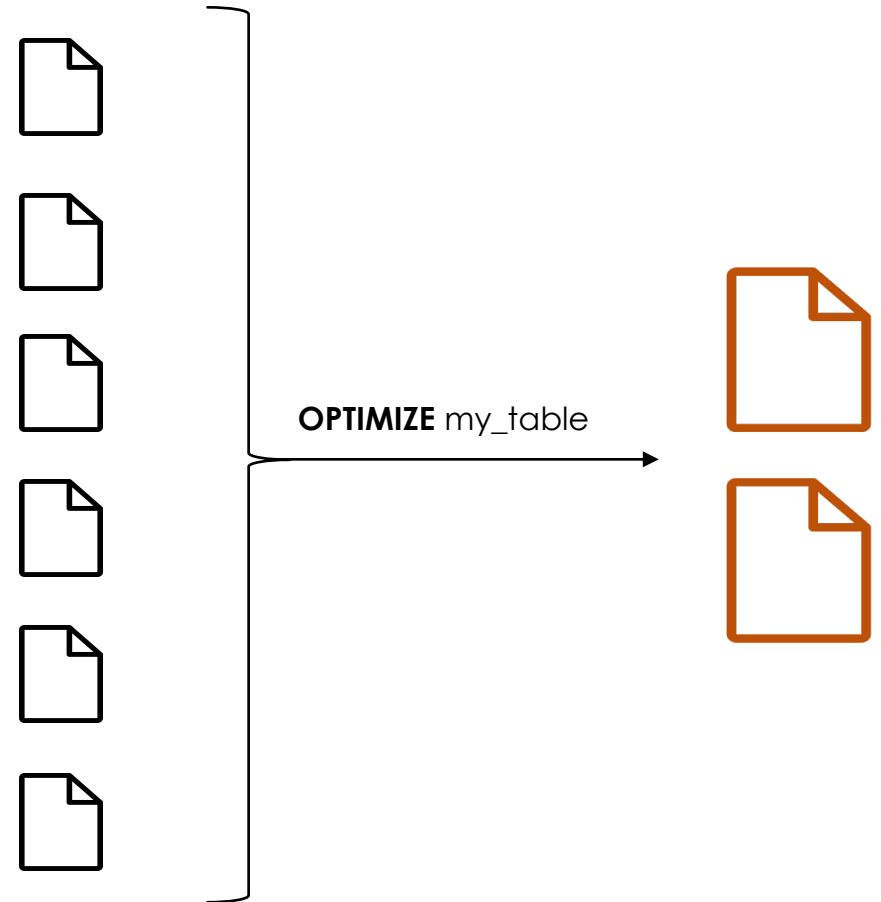
Auto Optimize

Learning Objectives

- ▶ Auto Optimize
- ▶ Enabling Auto Optimize
- ▶ Autotune file size

Optimize command

- ▶ Compacting Small Files
- ▶ The target file size is 1 GB



Auto Optimize

- ▶ Automatically compacts small files during individual writes to a table
- ▶ 2 complementary features
 1. Optimized writes: attempts to write 128 MB files
 2. Auto compaction:
 - ▶ After the write completes, it checks if files can further be compacted
 - ▶ If yes, it runs an **OPTIMIZE** job toward a file size of **128 MB** (*instead of 1 GB*)
- ▶ No Z-Ordering

Enabling Auto Optimize

- ▶ New tables

- ▶ **CREATE TABLE** myTable (id INT, name STRING)
TBLPROPERTIES (delta.autoOptimize.optimizeWrite = **true**,
delta.autoOptimize.autoCompact = **true**)

- ▶ Existing table

- ▶ **ALTER TABLE** myTable
SET TBLPROPERTIES (delta.autoOptimize.optimizeWrite = **true**,
delta.autoOptimize.autoCompact = **true**)

- ▶ Spark sessions

- ▶ spark.databricks.delta.optimizeWrite.enabled
 - ▶ spark.databricks.delta.autoCompact.enabled

Autotune file size based on workload

- ▶ Having many small files is not always a problem!
 - ▶ it can lead to better data skipping, and help minimize rewrites during merges and deletes.
- ▶ Databricks can automatically tune the file size of Delta tables, based on workloads operating on the table.
 - ▶ For Frequent *MERGE* operations, *Optimized Writes* and *Auto Compaction* will generate data files **smaller than 128 MB**. This helps in reducing the duration of future *MERGE* operations.