Advanced Delta Lake Features

Learning Objectives

▶ Time travel

▶ Compacting Small Files and Indexing

▶ Vacuum

Time travel

▶ Audit data changes

▶ **DESCRIBE HISTORY** command

Time travel

▶ Query older versions of the data

▶ Using a timestamp

    ▶ SELECT * FROM my_table **TIMESTAMP AS OF** "2019-01-01"

▶ Using a version number

    ▶ SELECT * FROM my_table **VERSION AS OF** 36

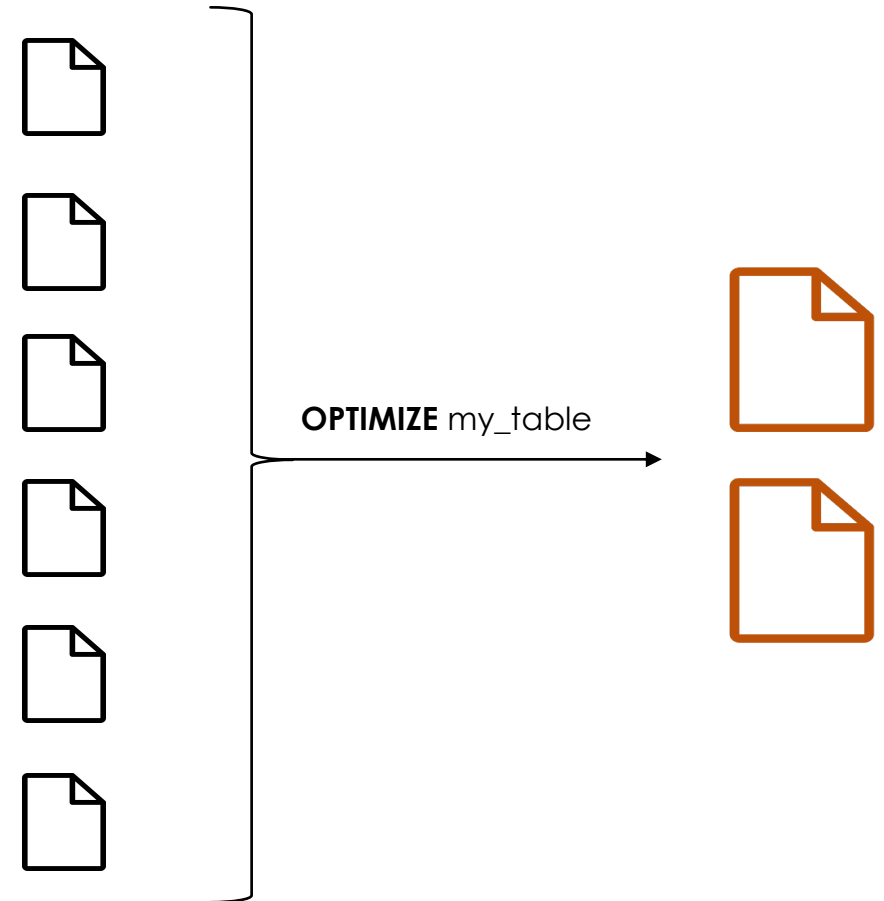   ▶ SELECT * FROM my_table**@v**36

Time travel

▶ Rollback Versions

▶ **RESTORE TABLE** command:

    ▶ **RESTORE TABLE** my_table **TO TIMESTAMP AS OF** "2019-01-01"

    ▶ **RESTORE TABLE** my_table **TO VERSION AS OF** 36
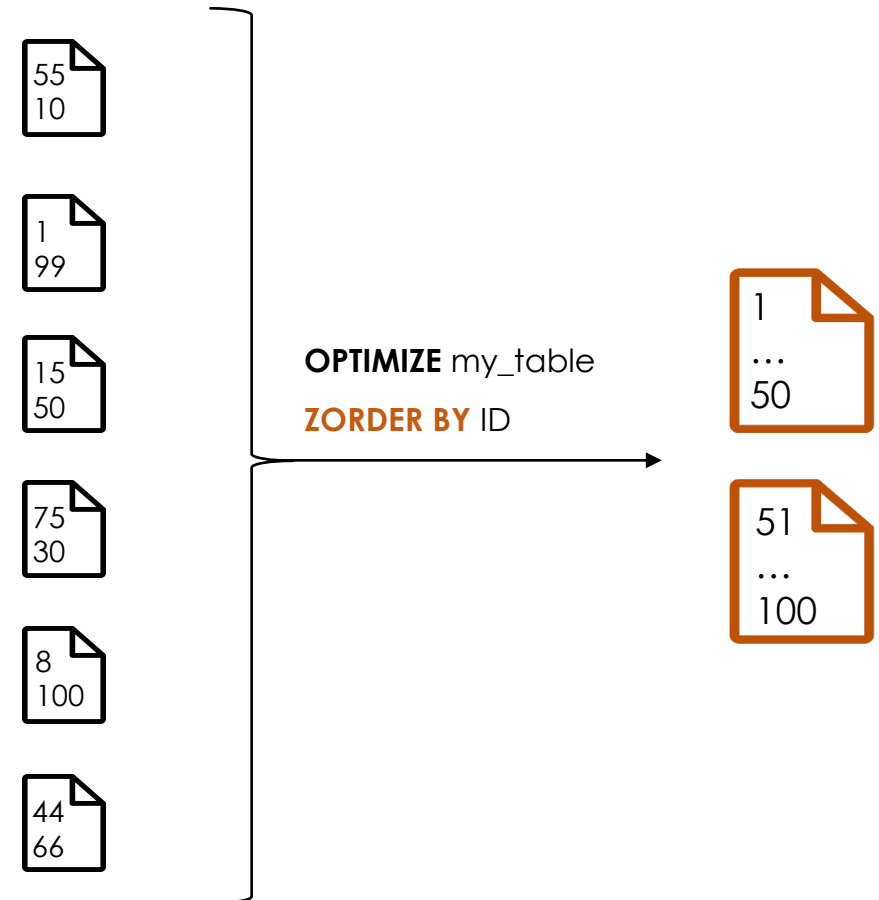
Compaction

▶ Compacting Small Files

▶ **OPTIMIZE** command:

    ▶ **OPTIMIZE** my_table

OPTIMIZE my_table

Indexing

▶ Co-locate column information

▶ **OPTIMIZE** command:
  ▶ OPTIMIZE my_table
    ZORDER BY column_name

55
10

1
99

15
50

75
30

8
100

44
66

OPTIMIZE my_table

ZORDER BY ID

1
...
50

51
...
100

Vacuum a Delta table

▶ Cleaning up unused data files
  ▶ uncommitted files
  ▶ files that are no longer in in latest table state

▶ **VACUUM** command
  ▶ VACUUM table_name [retention period]
  ▶ Default retention period: 7 days

▶ **Note: Vacuum = no time travel**