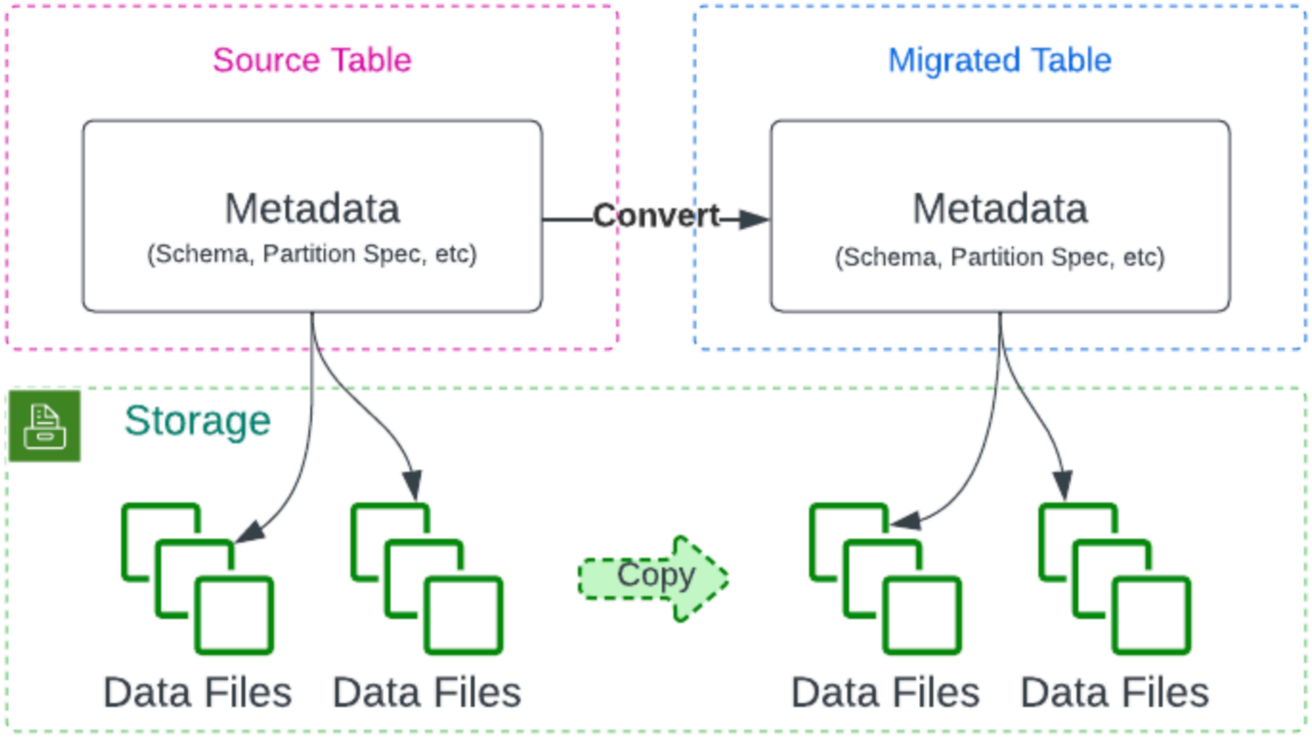
Migration Approaches

There are two main approaches to perform table migration: CTAS (Create Table As Select) and in-place migration.

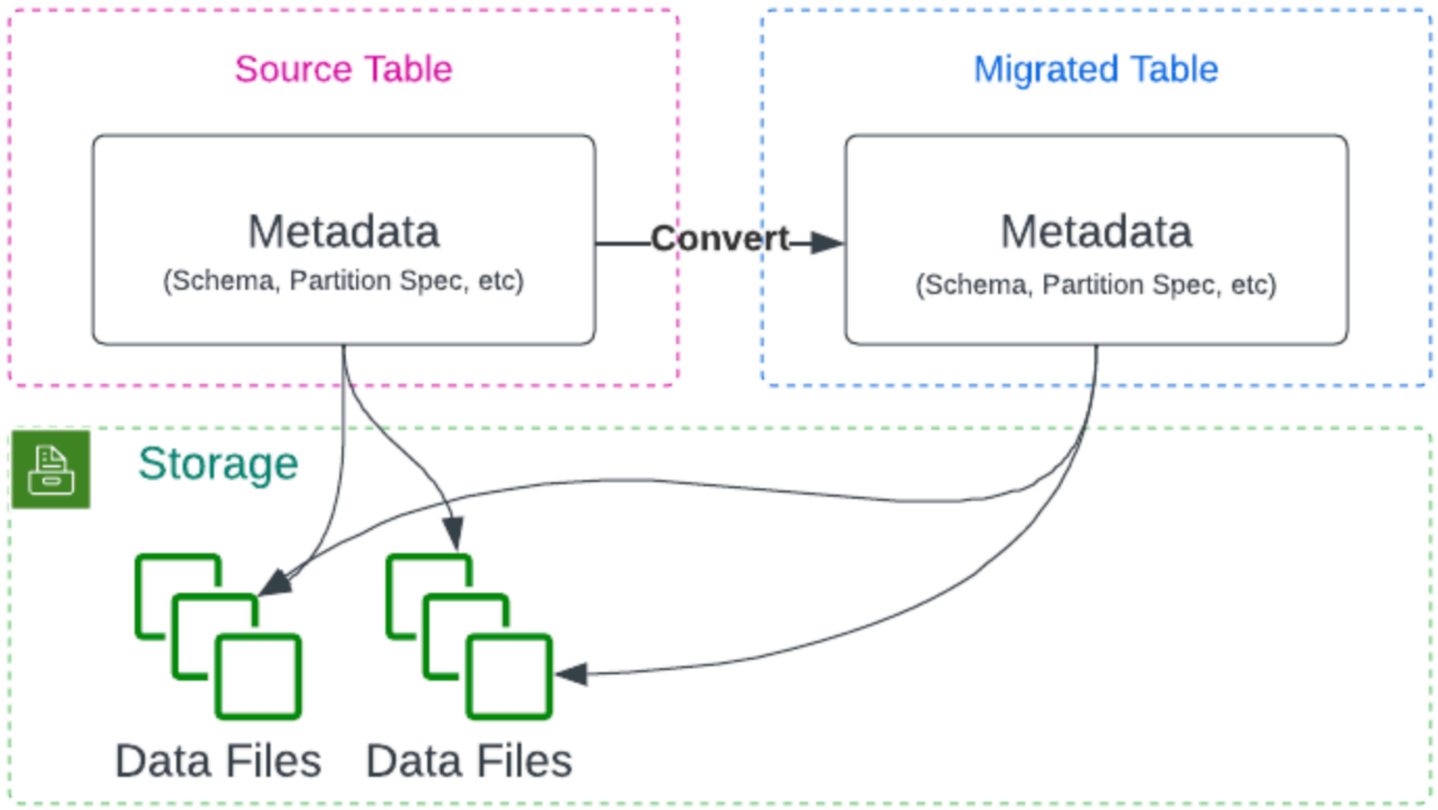
Create-Table-As-Select Migration

CTAS migration involves creating a new Iceberg table and copying data from the existing table to the new one. This method is preferred when you want to completely cut ties with your old table, ensuring the new table is independent and fully managed by Iceberg. However, CTAS migration may require more time to complete and might not be suitable for production use cases where downtime is not acceptable.



In-Place Migration

In-place migration retains the existing data files but adds Iceberg metadata on top of them. This approach is faster and does not require copying data, making it more suitable for production use cases.



In-Place Migration Actions

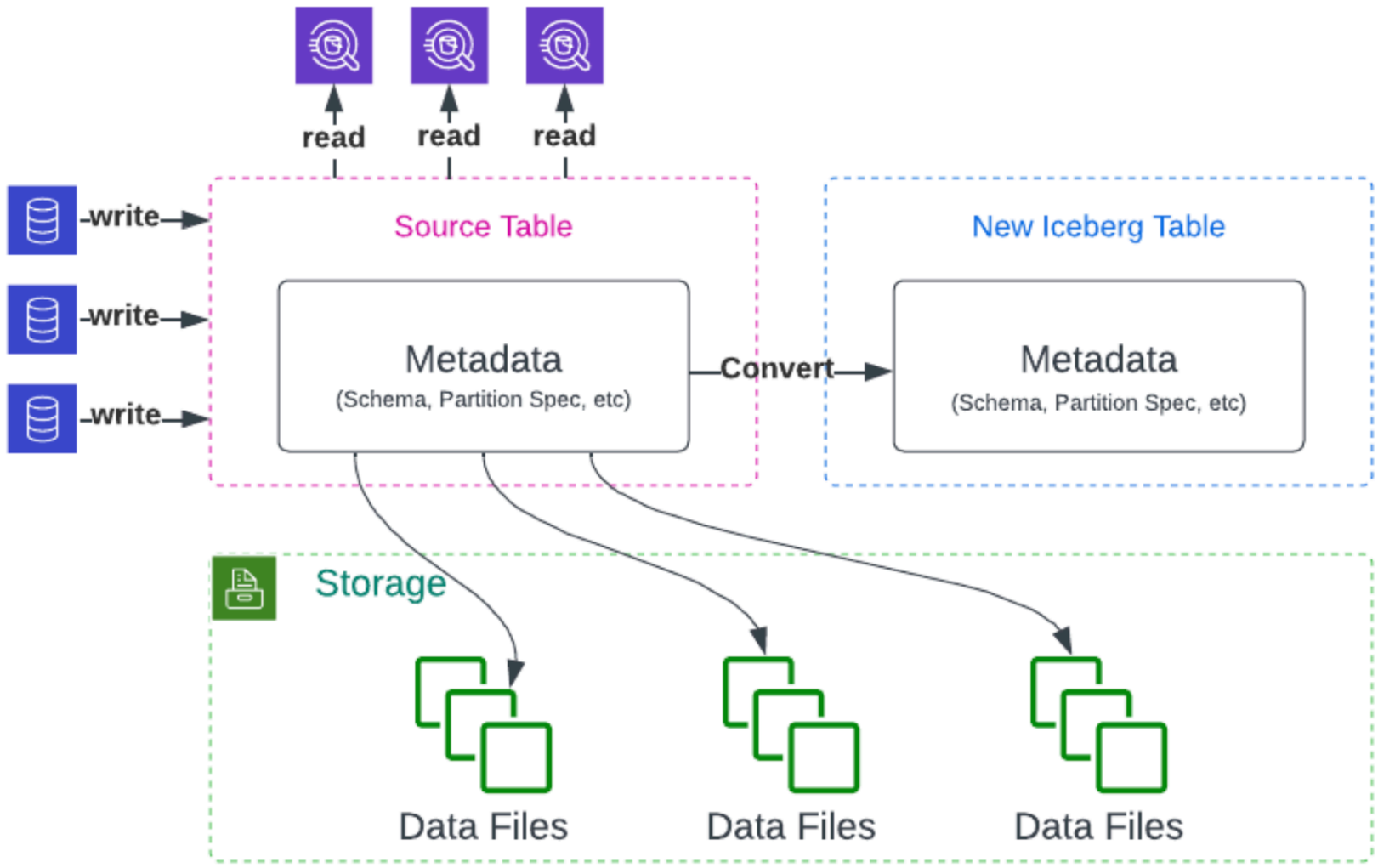
Apache Iceberg primarily supports the in-place migration approach, which includes three important actions:

1. Snapshot Table
2. Migrate Table
3. Add Files

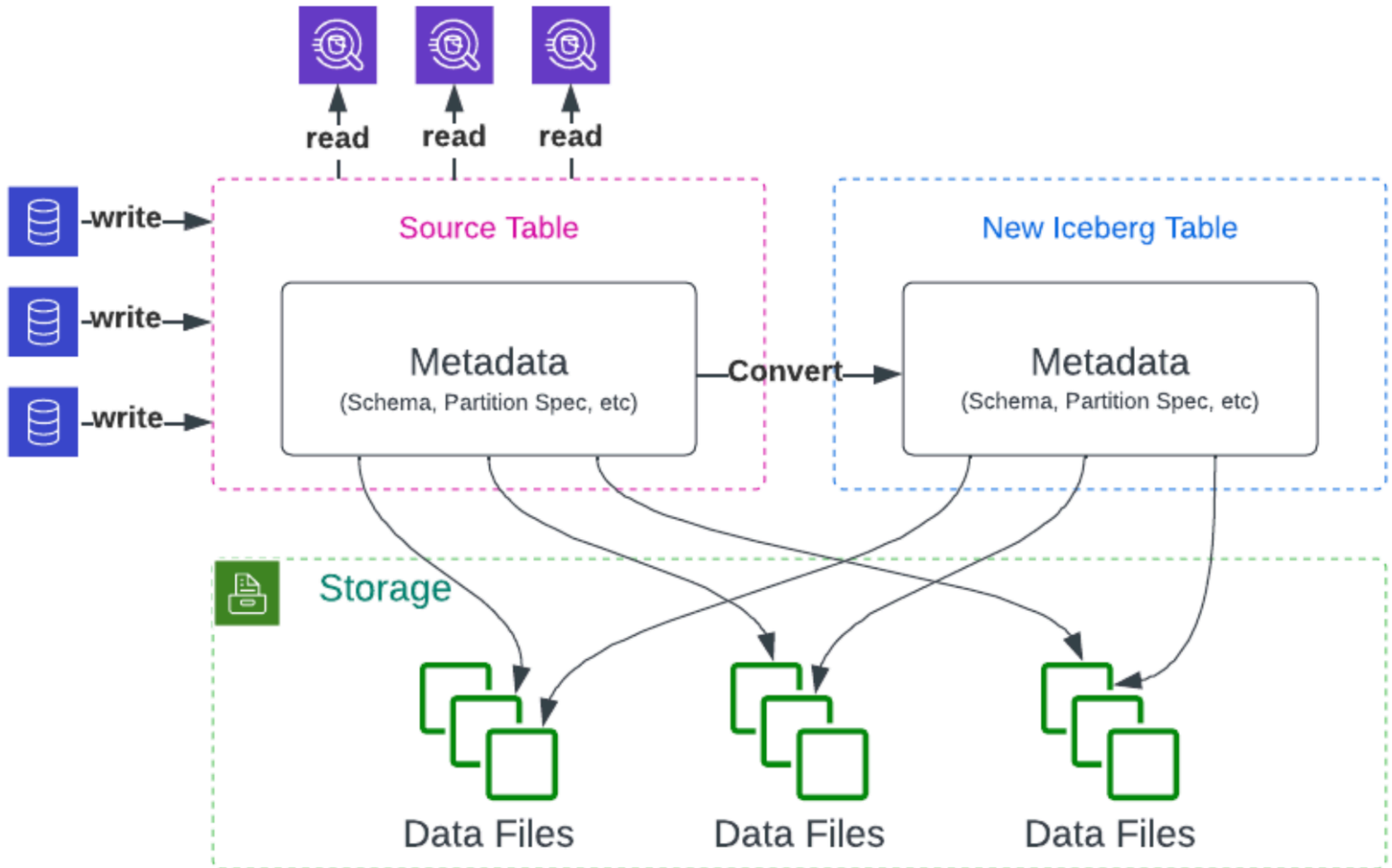
Snapshot Table

The Snapshot Table action creates a new iceberg table with the same schema and partitioning as the source table, leaving the source table unchanged during and after the action.

Step 1: Create a new Iceberg table with the same metadata (schema, partition spec, etc.) as the source table



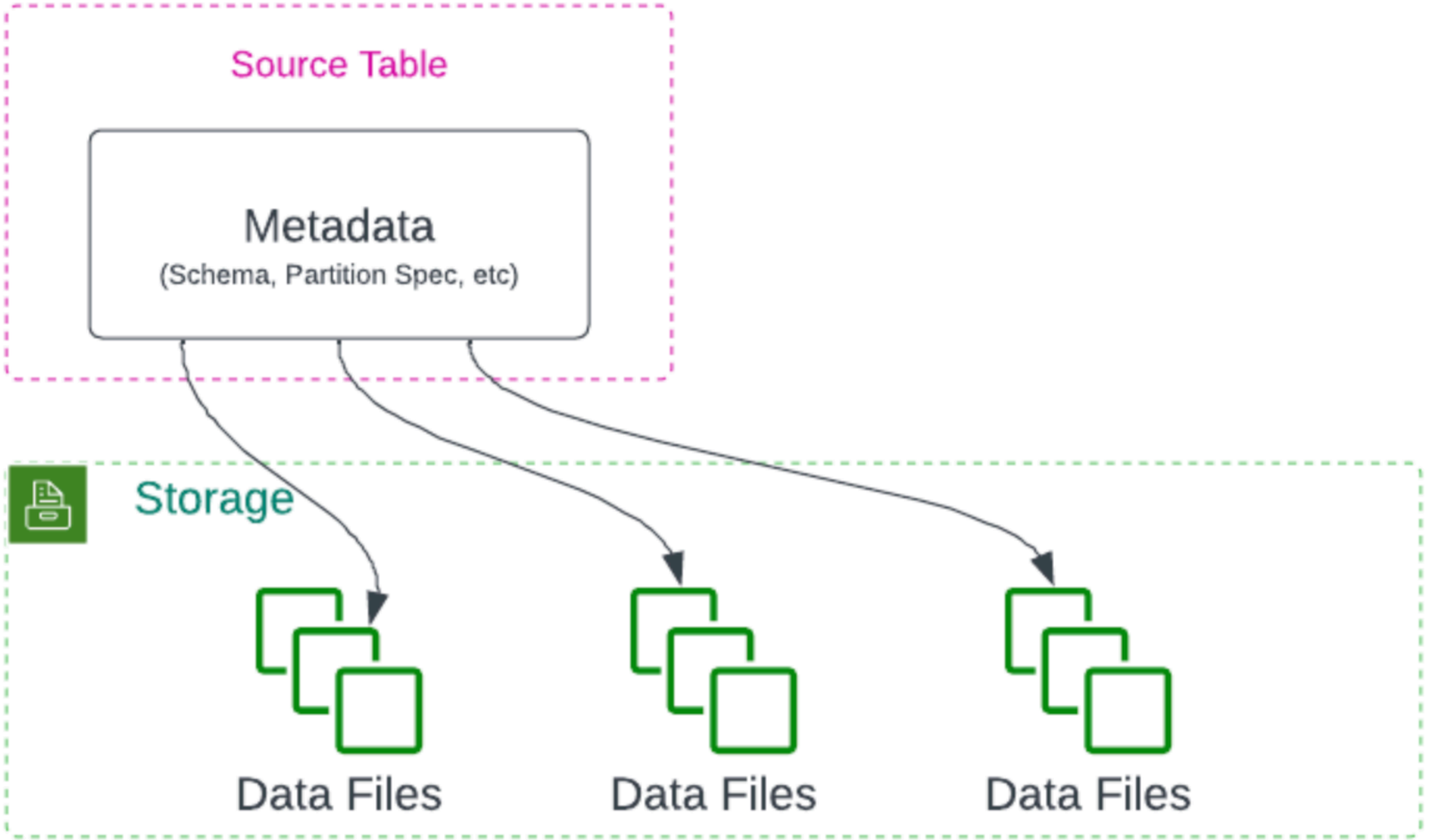
Step 2: Commit all data files across all partitions to the new Iceberg table. The source table remains unchanged.



Migrate Table

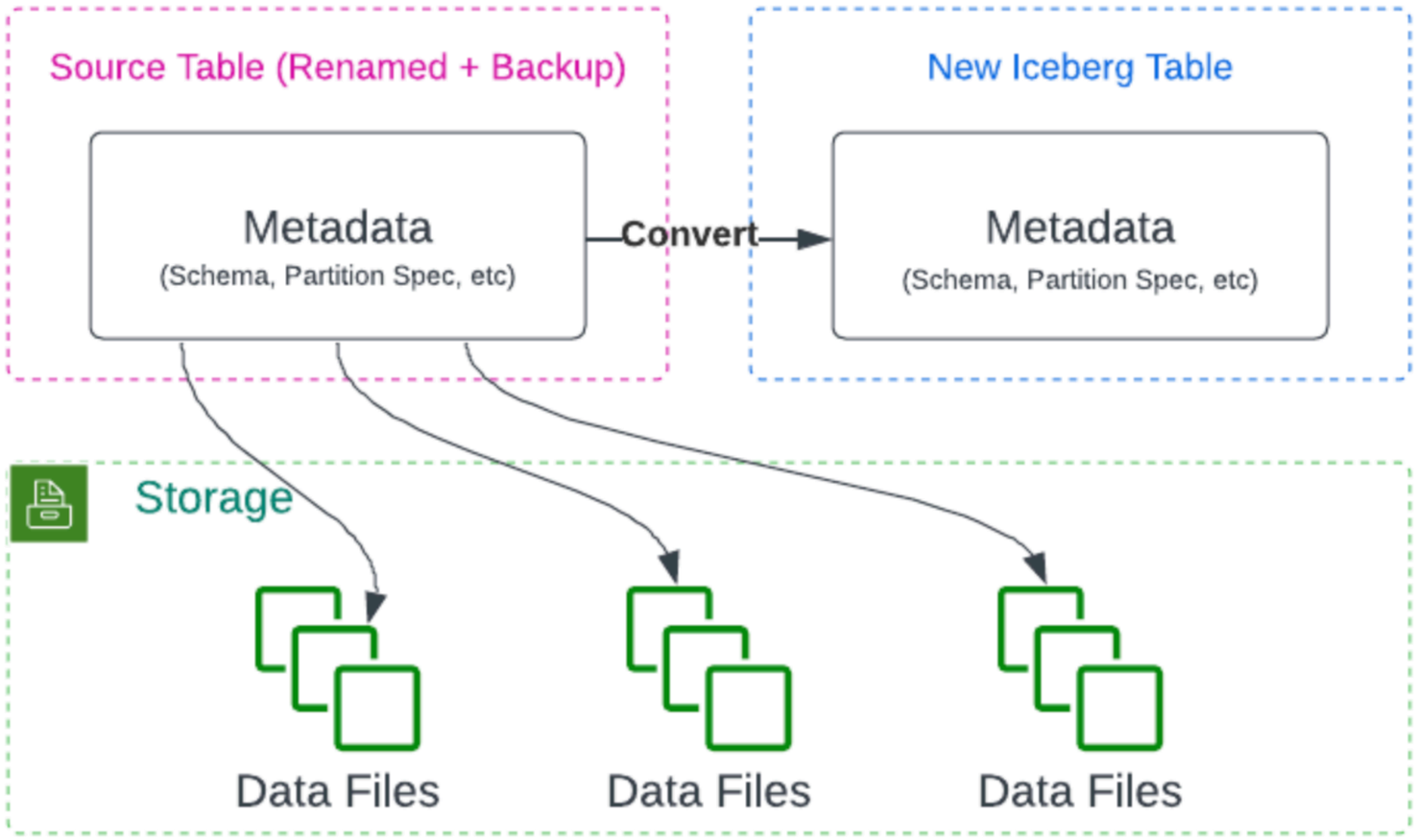
The Migrate Table action also creates a new Iceberg table with the same schema and partitioning as the source table. However, during the action execution, it locks and drops the source table from the catalog. Consequently, Migrate Table requires all readers and writers working on the source table to be stopped before the action is performed.

Step 1: Stop all readers and writers interacting with the source table

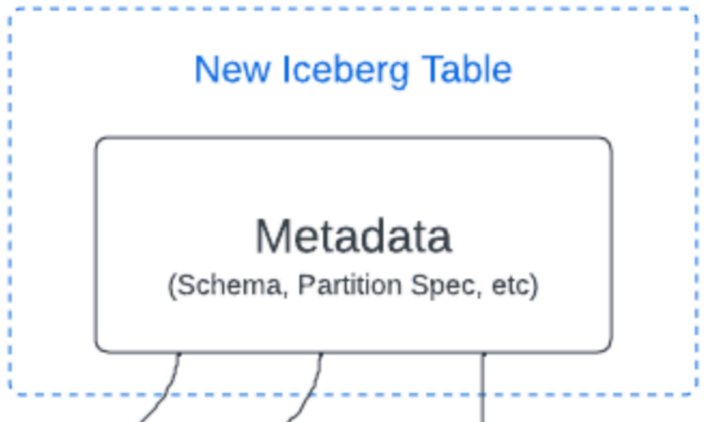


Step 2: Create a new Iceberg table with the same metadata (schema, partition spec, etc.) as the source table. Rename the source table for a backup in case of failure and rollback.

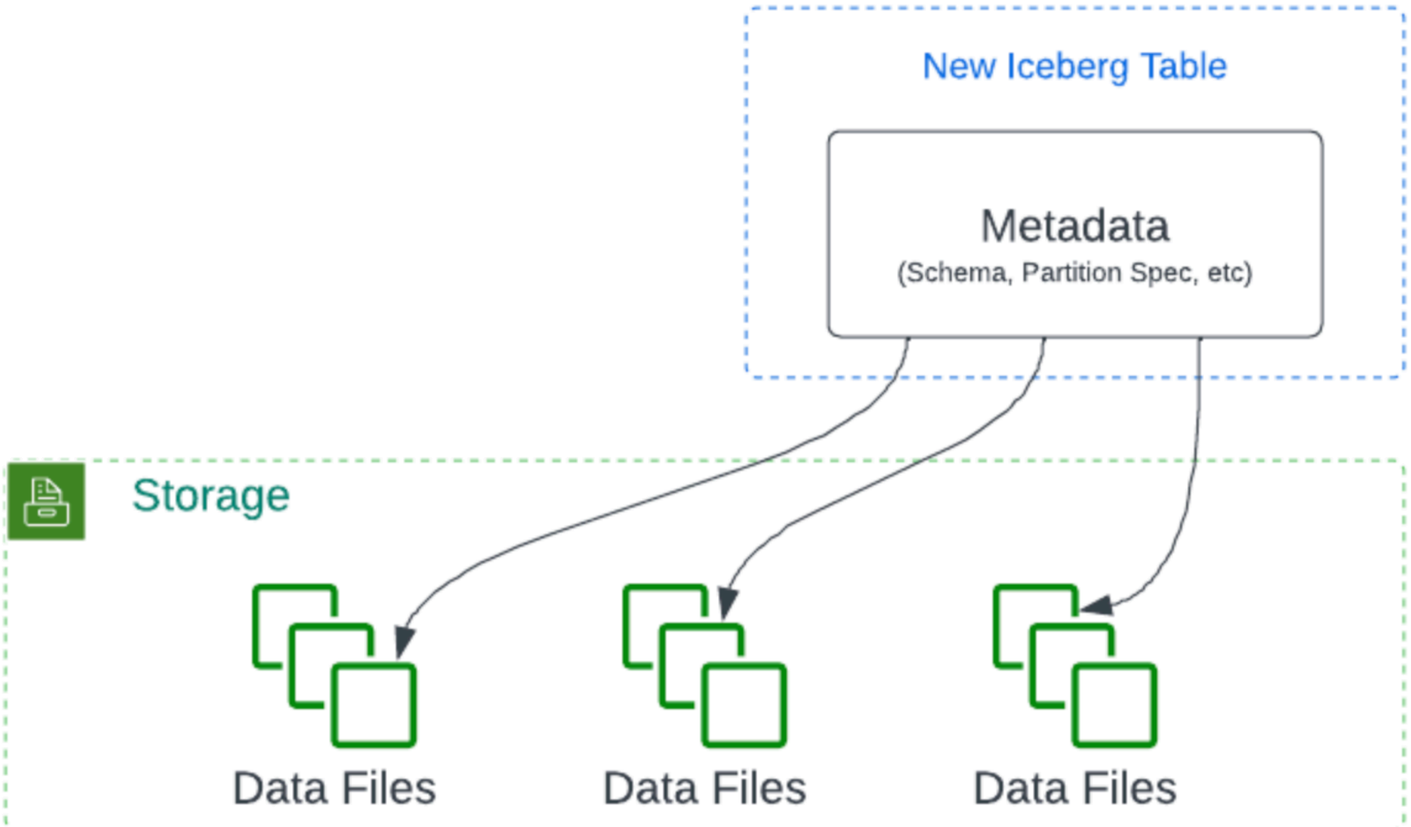
Step 2: Create a new Iceberg table with the same metadata (schema, partition spec, etc.) as the source table. Rename the source table for a backup in case of failure and rollback.



Step 3: Commit all data files across all partitions to the new Iceberg table. Drop the source table.



Step 3: Commit all data files across all partitions to the new Iceberg table. Drop the source table.



Add Files

After the initial step (either Snapshot Table or Migrate Table), it is common to find some data files that have not been migrated. These files often originate from concurrent writers who continue writing to the source table during or after the migration process. In practice, these files can be new data files in Hive tables or new snapshots (versions) of Delta Lake tables. The Add Files action is essential for incorporating these files into the Iceberg table.