Replication package for paper "Data migration for column family database evolution"

This document contains the instructions to replicate the results that are described and discussed in the paper "Data migration for column family database evolution" as well as the instructions of how to change the initial data and connection parameters. The project code and all files and folders referenced in this document are hosted in the following repository: https://github.com/giis-uniovi/modevo

Java 17 was used to obtain the results described in this paper.

The experimentation is performed as follows:

1. **Creation of data migration models (Transform module)**. These models are obtained by running the test cases through the following command:

```
mvn test -Dtest='*' -pl modevo-transform -am -Dmaven.test.failure.ignore=true -U --no-transfer-progress
```

   Input: Inputs models are located in: "modevo-transform/dat/inp". Output: output models are saved in: "modevo-transform/dat/out"

2. **Creation of data migration scripts (Script module)**.
   2.1. **Initialize Cassandra database**. The following command can be used to start it in a Docker container and bind port 9042 to the host so MoDEvo can access it:

```
docker run --name cassandra -d -p 9042:9042 cassandra:4.1.4
```

   2.2. **Create Cassandra schema** by executing the script contained in "modevo-script/dat/inp/creationSchema.cql".
   2.3. **Run test cases**. The following can be used for this:

```
mvn test -Dtest='!TestTransform*' -pl modevo-script -am -Dmaven.test.failure.ignore=true -U --no-transfer-progress
```

   Input: Inputs models are located in: "modevo-transform/dat/inp".
   Output: output scripts are stored in: "modevo-script/dat/out"

3. **Verification of data integrity maintenance (Consistency module)** .
   3.1. **Initialize MySQL database**. The following command can be used to start it in a Docker container:

```
docker run --name test-mysql -e MYSQL_ROOT_PASSWORD=rootpassword -d -p 3306:3306 mysql:9.0
```

   3.2. **Initialize the schema and data** of SQL database by executing SQL files stored in setup folder at the root of the project. The following script be executed from the root of the project:

```
docker exec -i test-mysql sh -c "exec mysql -uroot -prootpassword" < ./setup/custom.sql
docker exec -i test-mysql sh -c "exec mysql -uroot -prootpassword" < ./setup/minds.sql
docker exec -i test-mysql sh -c "exec mysql -uroot -prootpassword" < ./setup/wire.sql
docker exec -i test-mysql sh -c "exec mysql -uroot -prootpassword" < ./setup/thingsboard.sql
```

3.3. **Initialize Cassandra database**. As described in 2.1 and 2.2.

3.4. **Run test cases**. If they are successful, data integrity is verified. The following command can be used to execute them:

```
mvn test -Dtest='!TestTransform,!TestExecutionScript' -pl modevo-consistency -am -Dmaven.test.failure.ignore=true -Dsurefire.failIfNoSpecifiedTests=false -U --no-transfer-progress
```

3.5. **Initial data generation**: Initial data was created using Benerator[i]. The scripts used to generate these data are stored in "setup/benerator/". These scripts can be modified to change the number of rows generated to each table. If this is done, the SQL scripts stored in the setup folder must be replaced with scripts that contain these new data.

4. Connections with databases.

4.1. **Cassandra**: The connection parameters can be modified in the properties file located in "modevo-script/src/test/resources/dbconnection.properties"

4.2. **MySQL**: The connection parameters can be modified in the properties file located in "modevo-consistency/src/test/resources/sqlconnection.properties"

[i] https://github.com/rapiddweller/rapiddweller-benerator-ce