

There are three hive actions in **movie_ratings_wf.xml** workflow:

movie_ratings table creation under movie_lens_data database.

Truncate existing data.

Populate new data.

Validating Oozie workflow:

Steps to validate Oozie workflow **movie_ratings_wf.xml**:

Download Oozie workflow `hdfs://EETeamJ1/user/oozie/oozie-jobs/job-1/movie_ratings_wf.xml` from hdfs to machine 172.26.60.16 at any location.

Delete the existing **movie_ratings** table if it is present and test whether table is created by the oozie workflow and try to insert data into the table and see workflow is truncating it or not.

Execute below command to validate.

```
oozie validate <LocalFileLocation>/movie_ratings_wf.xml??
```

Workflow execution and Job status Commands:

Oozie workflow execution command:

```
oozie job --oozie http://impetus-i0161.impetus.co.in:11000/oozie -Doozie.wf.application.path=hdfs://EETeamJ1/user/oozie/oozie-jobs/job-1/movie_ratings_wf.xml -config/home/IMPETUS/omprakash.soni/ActualWF/MovieRating-WF/job.properties -run
```

This will return a job Id in format job: <Job ID>

2. oozie job status check command.

```
oozie job --oozie http://impetus-i0161.impetus.co.in:11000/oozie -info<JobId>
```

Output Verification:

After successfully execution of Oozie workflow a new table **movie_ratings** created.

Verify the data of new table from original tables. Count and data should be match.

Related queries:

List all the movies and the number of ratings from **movie_ratings** table:

```
select count (*) from mov_rating_count;
select * from mov_rating_count;
```

List all the movies and the number of ratings from original tables:

```
SELECT movies.movie_id, movies.title, count(ratings.movie_id) as count FROM
movies INNER JOIN ratings ON (movies.movie_id = ratings.movie_id) group by
movies.movie_id ,title;
```

oozie workflow validation.

oozie validate <WorkFlow Path>

Hive table drop command.

drop table <table name>;

Getting hive table info.

describe <table name>;

Inserting data into hive table.

insert into table <table name> values ('<col1_value>','<col2_value>'..., '<coln_value>');