

ID: 3367315

TESTS

The following are the tests that I ran on the code to ensure its correctness.

1. Input Parsing

I tested whether the mapper code could read and parse input files correctly by printing the lines of code through `.readLines()`.

2. Split the genres

Using the smaller file as input, I tested whether the lines were correctly based on ``` separator.

3. Verify filters

I ran the code to verify that it only included titles in the years and genres specified in the input file.

4. Empty filters

I ran the code to verify that it correctly handles empty filters and returns all titles when no filters are specified.

5. Similar Input and Output

I ensured both the combiner and reducer code could handle similar input formats. If Hadoop skips the combiner, the reducer should still work correctly.

6. Error Handling

I tested whether the code could handle errors gracefully, such as missing input files or invalid input formats.

7. Scalability

I tested whether the code could handle large input files efficiently by running it on a cluster with multiple nodes. I edited the `runhadoop.sh` script to test with 1 or more nodes.

8. Testing Spark

I tested the spark code on the university cluster and it ran well.

9. Comparing Hadoop, Spark and Bash

I compared the results of the Hadoop, Spark and Bash implementations to ensure they produced the same output. They did.