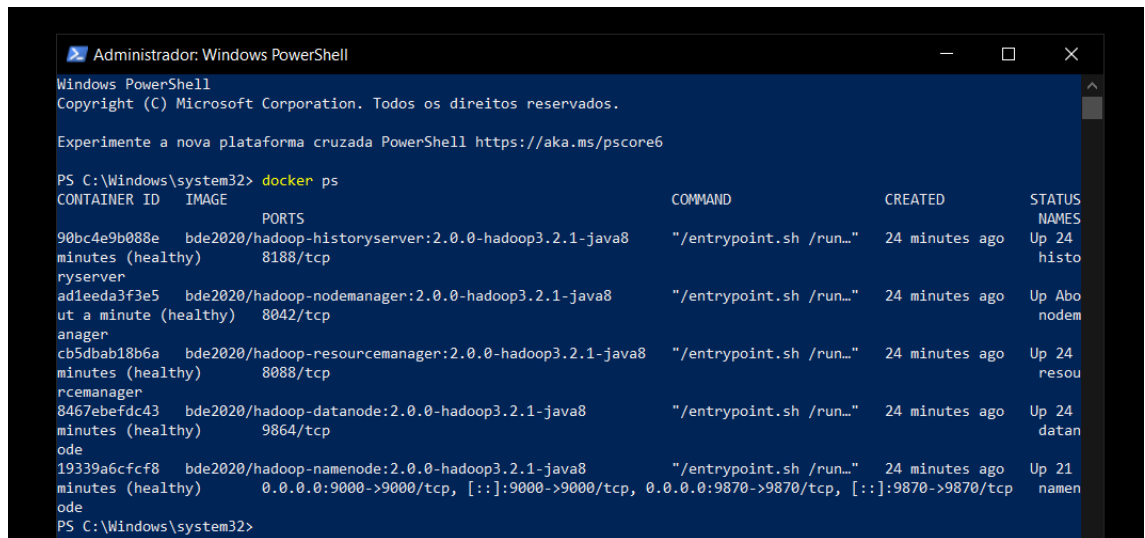


Aluna: Samila Garrido

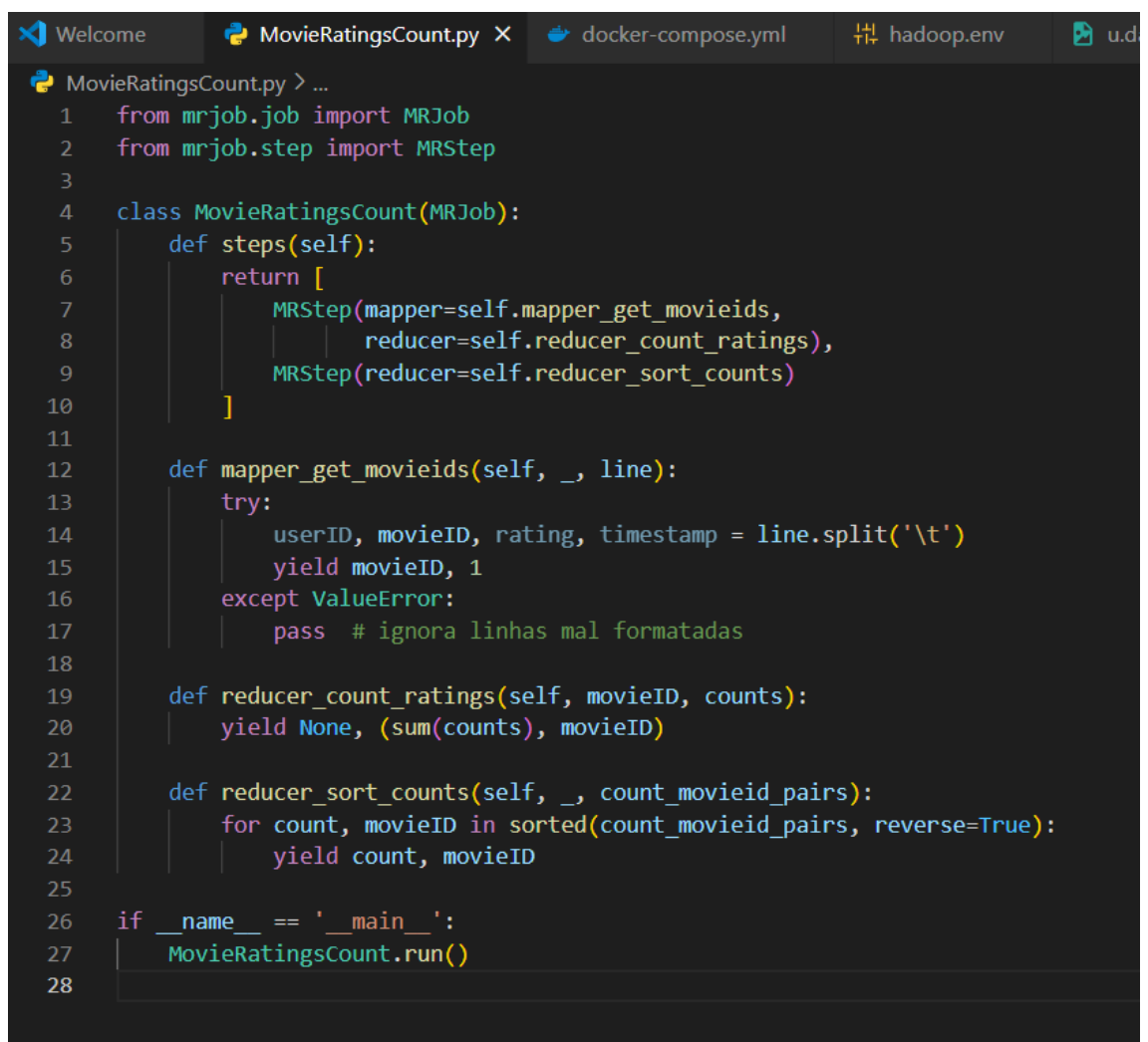
- hadoop funcionando com Docker:



A screenshot of a Windows PowerShell window titled "Administrador: Windows PowerShell". The window shows the output of the command `docker ps`. The output is a table with columns: CONTAINER ID, IMAGE, PORTS, COMMAND, CREATED, and STATUS. It lists five Hadoop containers: historyserver, nodemanager, resourcemanager, datanode, and namenode, all running the `bde2020/hadoop-*` images. The status for all containers is "Up 24 minutes ago".

CONTAINER ID	IMAGE	PORTS	COMMAND	CREATED	STATUS
90bc4e9b088e	bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8	8188/tcp	"/entrypoint.sh /run..."	24 minutes ago	Up 24 minutes (healthy)
ad1eeda3f3e5	bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8	8042/tcp	"/entrypoint.sh /run..."	24 minutes ago	Up 24 minutes (healthy)
cb5dbab18b6a	bde2020/hadoop-resourcemanager:2.0.0-hadoop3.2.1-java8	8088/tcp	"/entrypoint.sh /run..."	24 minutes ago	Up 24 minutes (healthy)
8467ebefdc43	bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8	9864/tcp	"/entrypoint.sh /run..."	24 minutes ago	Up 24 minutes (healthy)
19339a6cfcf8	bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8	0.0.0.0:9000->9000/tcp, [::]:9000->9000/tcp, 0.0.0.0:9870->9870/tcp, [::]:9870->9870/tcp	"/entrypoint.sh /run..."	24 minutes ago	Up 21 minutes (healthy)

- Código Python



A screenshot of a code editor showing a Python script named `MovieRatingsCount.py`. The script uses the `mrjob` library to implement a MapReduce job for counting movie ratings. It defines a `MovieRatingsCount` class that inherits from `MRJob`. The class has three main methods: `steps` (returns a list of `MRStep` objects), `mapper_get_movieids` (maps movie IDs to 1), and `reducer_count_ratings` (reduces the counts for each movie ID). The script also includes a `reducer_sort_counts` method and a main block to run the job.

```
1 from mrjob.job import MRJob
2 from mrjob.step import MRStep
3
4 class MovieRatingsCount(MRJob):
5     def steps(self):
6         return [
7             MRStep(mapper=self.mapper_get_movieids,
8                   reducer=self.reducer_count_ratings),
9             MRStep(reducer=self.reducer_sort_counts)
10        ]
11
12    def mapper_get_movieids(self, _, line):
13        try:
14            userID, movieID, rating, timestamp = line.split('\t')
15            yield movieID, 1
16        except ValueError:
17            pass # ignora linhas mal formatadas
18
19    def reducer_count_ratings(self, movieID, counts):
20        yield None, (sum(counts), movieID)
21
22    def reducer_sort_counts(self, _, count_movieid_pairs):
23        for count, movieID in sorted(count_movieid_pairs, reverse=True):
24            yield count, movieID
25
26 if __name__ == '__main__':
27     MovieRatingsCount.run()
28
```

- Saída

```
Administrador: Windows PowerShell
Running step 2 of 2...
job output is in /tmp/MovieRatingsCount.root.20250913.013341.423591/output
Streaming final output from /tmp/MovieRatingsCount.root.20250913.013341.423591/output...
584 "50"
509 "258"
508 "100"
507 "181"
485 "294"
481 "286"
478 "288"
452 "1"
431 "300"
429 "121"
420 "174"
413 "127"
394 "56"
392 "7"
390 "98"
384 "237"
378 "117"
368 "172"
365 "222"
350 "313"
350 "204"
344 "405"
336 "79"
331 "210"
326 "151"
324 "173"
321 "69"
316 "748"
316 "168"
315 "269"
303 "257"
301 "195"
300 "423"
299 "9"
298 "318"
298 "276"
297 "302"
297 "22"
295 "96"
295 "328"
293 "25"
293 "15"
293 "118"
291 "183"
290 "216"
284 "176"
283 "64"
```

```
Administrador: Windows PowerShell
1 "1546"
1 "1543"
1 "1536"
1 "1533"
1 "1526"
1 "1525"
1 "1520"
1 "1515"
1 "1510"
1 "1507"
1 "1505"
1 "1498"
1 "1494"
1 "1493"
1 "1492"
1 "1486"
1 "1482"
1 "1476"
1 "1461"
1 "1460"
1 "1458"
1 "1457"
1 "1453"
1 "1452"
1 "1447"
1 "1414"
1 "1373"
1 "1366"
1 "1364"
1 "1363"
1 "1352"
1 "1349"
1 "1348"
1 "1343"
1 "1341"
1 "1340"
1 "1339"
1 "1329"
1 "1325"
1 "1320"
1 "1310"
1 "1309"
1 "1236"
1 "1235"
1 "1201"
1 "1156"
1 "1130"
1 "1122"
Removing temp directory /tmp/MovieRatingsCount.root.20250913.013341.423591...
root@19339a6cfcf8:~#
```