**1. Trung bình giá các món ăn trong quán**

Mapper\_AvgPrice.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  for line in sys.stdin:  line = line.strip()  fields = line.split('\t')  if len(fields) != 4:  continue  try:  idquan = fields[0]  price = float(fields[3])  print(f"{idquan}\t{price}\t1")  except ValueError:  continue |

Reducer\_AvgPrice.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  current\_id = None  total\_price = 0  count = 0  for line in sys.stdin:  line = line.strip()  parts = line.split('\t')  if len(parts) != 3:  continue  idquan, price\_str, count\_str = parts  try:  price = float(price\_str)  cnt = int(count\_str)  except ValueError:  continue  if current\_id == idquan:  total\_price += price  count += cnt  else:  if current\_id is not None and count > 0:  avg\_price = total\_price / count  print(f"{current\_id}\t{avg\_price:.2f}")  current\_id = idquan  total\_price = price  count = cnt  if current\_id is not None and count > 0:  avg\_price = total\_price / count  print(f"{current\_id}\t{avg\_price:.2f}") |

Chạy chương trình

$ hadoop jar /home/hadoopnhom3/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar \

-file /home/hadoopnhom3/Mapreduce/Mapper\_AvgPrice.py -mapper /home/hadoopnhom3/Mapreduce/Mapper\_AvgPrice.py \

-file /home/hadoopnhom3/Mapreduce/Reducer\_AvgPrice.py -reducer /home/hadoopnhom3/Mapreduce/Reducer\_AvgPrice.py \

-input /user/hadoopnhom3/BigData/DOAN \

-output /user/hadoopnhom3/BigData/output/avg\_price\_per\_quan

Hiển thị kết quả

$ hdfs dfs -cat /user/hadoopnhom3/BigData/output/avg\_price\_per\_quan/part-00000

A computer screen with a purple background

AI-generated content may be incorrect.

**2. Số lượng món ăn trong mỗi quán**

Mapper\_count.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  for line in sys.stdin:  line = line.strip()  fields = line.split('\t')  if len(fields) != 4:  continue  idquan = fields[0]  print(f"{idquan}\t1") |

Reducer\_count.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  current\_id = None  count = 0  for line in sys.stdin:  line = line.strip()  idquan, value = line.split('\t')  try:  value = int(value)  except ValueError:  continue  if current\_id == idquan:  count += value  else:  if current\_id is not None:  print(f"{current\_id}\t{count}")  current\_id = idquan  count = value  # output cuối cùng  if current\_id is not None:  print(f"{current\_id}\t{count}") |

$ hadoop jar /home/hadoopnhom3/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar \

-file /home/hadoopnhom3/Mapreduce/Mapper\_count.py -mapper /home/hadoopnhom3/Mapreduce/Mapper\_count.py \

-file /home/hadoopnhom3/Mapreduce/Reducer\_count.py -reducer /home/hadoopnhom3/Mapreduce/Reducer\_count.py \

-input /user/hadoopnhom3/BigData/DOAN \

-output /user/hadoopnhom3/BigData/output/count\_items\_per\_quan

Hiển thị kết quả

$ hdfs dfs -cat /user/hadoopnhom3/BigData/output/count\_items\_per\_quan/part-00000

A computer screen with white text

AI-generated content may be incorrect.

**3. Món ăn giá cao nhất ở mỗi quán**

Mapper\_maxPrice.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  for line in sys.stdin:  line = line.strip()  parts = line.split('\t')  if len(parts) != 4:  continue  idquan, \_, item, price = parts  try:  price = float(price)  print(f"{idquan}\t{item}\t{price}")  except ValueError:  continue |

Reducer\_maxPrice.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  current\_id = None  max\_price = -1  max\_item = ""  for line in sys.stdin:  line = line.strip()  parts = line.split('\t')  if len(parts) != 3:  continue  idquan, item, price\_str = parts  try:  price = float(price\_str)  except ValueError:  continue  if current\_id == idquan:  if price > max\_price:  max\_price = price  max\_item = item  else:  if current\_id is not None:  print(f"{current\_id}\t{max\_item}\t{max\_price}")  current\_id = idquan  max\_price = price  max\_item = item  # Output cuối cùng  if current\_id is not None:  print(f"{current\_id}\t{max\_item}\t{max\_price}") |

$ hadoop jar /home/hadoopnhom3/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar \

-file /home/hadoopnhom3/Mapreduce/Mapper\_maxPrice.py -mapper /home/hadoopnhom3/Mapreduce/Mapper\_maxPrice.py \

-file /home/hadoopnhom3/Mapreduce/Reducer\_maxPrice.py -reducer /home/hadoopnhom3/Mapreduce/Reducer\_maxPrice.py \

-input /user/hadoopnhom3/BigData/DOAN \

-output /user/hadoopnhom3/BigData/output/max\_price\_per\_quan

Hiển thị kết quả

$ hdfs dfs -cat /user/hadoopnhom3/BigData/output/max\_price\_per\_quan/part-00000

A computer screen shot of a computer

AI-generated content may be incorrect.

**4. Quán ăn có giá trung bình các món cao nhất với giá bao nhiêu**

Mapper\_max\_avg\_price.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  for line in sys.stdin:  line = line.strip()  parts = line.split('\t')  if len(parts) != 4:  continue  idquan, \_, \_, price = parts  try:  price = float(price)  print(f"{idquan}\t{price}\t1")  except ValueError:  continue |

Reducer\_max\_avg\_price.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  current\_id = None  total\_price = 0.0  count = 0  for line in sys.stdin:  line = line.strip()  parts = line.split('\t')  if len(parts) != 3:  continue  idquan, price\_str, cnt\_str = parts  try:  price = float(price\_str)  cnt = int(cnt\_str)  except ValueError:  continue  if current\_id == idquan:  total\_price += price  count += cnt  else:  if current\_id is not None and count > 0:  avg\_price = total\_price / count  print(f"{current\_id}\t{avg\_price}")  current\_id = idquan  total\_price = price  count = cnt  if current\_id is not None and count > 0:  avg\_price = total\_price / count  print(f"{current\_id}\t{avg\_price}") |

Reducer\_find\_max\_avg.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  max\_id = None  max\_avg = -1  for line in sys.stdin:  line = line.strip()  idquan, avg\_str = line.split('\t')  try:  avg = float(avg\_str)  except ValueError:  continue  if avg > max\_avg:  max\_avg = avg  max\_id = idquan  if max\_id is not None:  print(f"{max\_id}\t{max\_avg}") |

Tính giá trung bình

$ hadoop jar /home/hadoopnhom3/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar \

-file /home/hadoopnhom3/Mapreduce/Mapper\_max\_avg\_price.py -mapper /home/hadoopnhom3/Mapreduce/Mapper\_max\_avg\_price.py \

-file /home/hadoopnhom3/Mapreduce/Reducer\_max\_avg\_price.py -reducer /home/hadoopnhom3/Mapreduce/Reducer\_max\_avg\_price.py \

-input /user/hadoopnhom3/BigData/DOAN \

-output /user/hadoopnhom3/BigData/output/avg\_price\_per\_quan

Tìm quán ăn trung bình cao nhất

$ hadoop jar /home/hadoopnhom3/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar \

-input /user/hadoopnhom3/BigData/output/avg\_price\_per\_quan1 \

-output /user/hadoopnhom3/BigData/output/max\_avg\_price\_quan \

-mapper cat \

-reducer /home/hadoopnhom3/Mapreduce/Reducer\_find\_max\_avg.py \

-file /home/hadoopnhom3/Mapreduce/Reducer\_find\_max\_avg.py

Hiển thị kết quả

$ hdfs dfs -cat /user/hadoopnhom3/BigData/output/max\_avg\_price\_quan/part-00000

A computer screen shot of a computer screen

AI-generated content may be incorrect.

**5. Quán ăn có giá trung bình các món thấp nhất với giá bao nhiêu**

Reducer\_find\_min\_avg.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  min\_avg = float('inf')  min\_quan = ""  for line in sys.stdin:  line = line.strip()  if not line:  continue  quan, avg\_str = line.split()  try:  avg = float(avg\_str)  if avg < min\_avg:  min\_avg = avg  min\_quan = quan  except ValueError:  continue  print(f"{min\_quan}\t{min\_avg}") |

Chạy chương trình

$ hadoop jar /home/hadoopnhom3/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar \

-input /user/hadoopnhom3/BigData/output/avg\_price\_per\_quan1 \

-output /user/hadoopnhom3/BigData/output/min\_avg\_price\_quan \

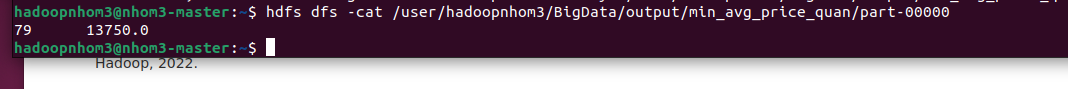
-mapper cat \

-reducer /home/hadoopnhom3/Mapreduce/Reducer\_find\_min\_avg.py \

-file /home/hadoopnhom3/Mapreduce/Reducer\_find\_min\_avg.py

Hiển thị kết quả

$ hdfs dfs -cat /user/hadoopnhom3/BigData/output/min\_avg\_price\_quan/part-00000



**6. Quán ăn có số lượng món ăn nhiều nhất**

Reducer\_find\_max\_items.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  max\_count = -1  max\_quans = []  for line in sys.stdin:  line = line.strip()  if not line:  continue  quan, count\_str = line.split()  try:  count = int(count\_str)  if count > max\_count:  max\_count = count  max\_quans = [quan]  elif count == max\_count:  max\_quans.append(quan)  except ValueError:  continue  for quan in max\_quans:  print(f"{quan}\t{max\_count}") |

Chạy chương trình

$ hadoop jar /home/hadoopnhom3/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar \

-input /user/hadoopnhom3/BigData/output/count\_items\_per\_quan \

-output /user/hadoopnhom3/BigData/output/max\_items\_quan \

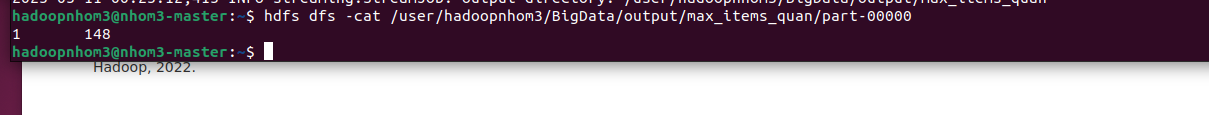
-mapper cat \

-reducer /home/hadoopnhom3/Mapreduce/Reducer\_find\_max\_items.py \

-file /home/hadoopnhom3/Mapreduce/Reducer\_find\_max\_items.py

Hiển thị kết quả

$ hdfs dfs -cat /user/hadoopnhom3/BigData/output/max\_items\_quan/part-00000



**7. Quán ăn có số lượng món ăn ít nhất**

Reducer\_find\_min\_items.py

|  |
| --- |
| #!/usr/bin/env python3  import sys  min\_count = None  min\_quans = []  for line in sys.stdin:  line = line.strip()  if not line:  continue  quan, count\_str = line.split()  try:  count = int(count\_str)  if min\_count is None or count < min\_count:  min\_count = count  min\_quans = [quan]  elif count == min\_count:  min\_quans.append(quan)  except ValueError:  continue  for quan in min\_quans:  print(f"{quan}\t{min\_count}") |

Chạy chương trình

$ hadoop jar /home/hadoopnhom3/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar \

-input /user/hadoopnhom3/BigData/output/count\_items\_per\_quan \

-output /user/hadoopnhom3/BigData/output/min\_items\_quan1 \

-mapper cat \

-reducer /home/hadoopnhom3/Mapreduce/Reducer\_find\_min\_items.py \

-file /home/hadoopnhom3/Mapreduce/Reducer\_find\_min\_items.py

Hiển thị kết quả

$ hdfs dfs -cat /user/hadoopnhom3/BigData/output/min\_items\_quan1/part-00000

A screenshot of a computer

AI-generated content may be incorrect.