

211275026-陈畅-作业5

[git仓库链接](#)

[设计思路](#)

[程序运行结果](#)

[进一步分析](#)

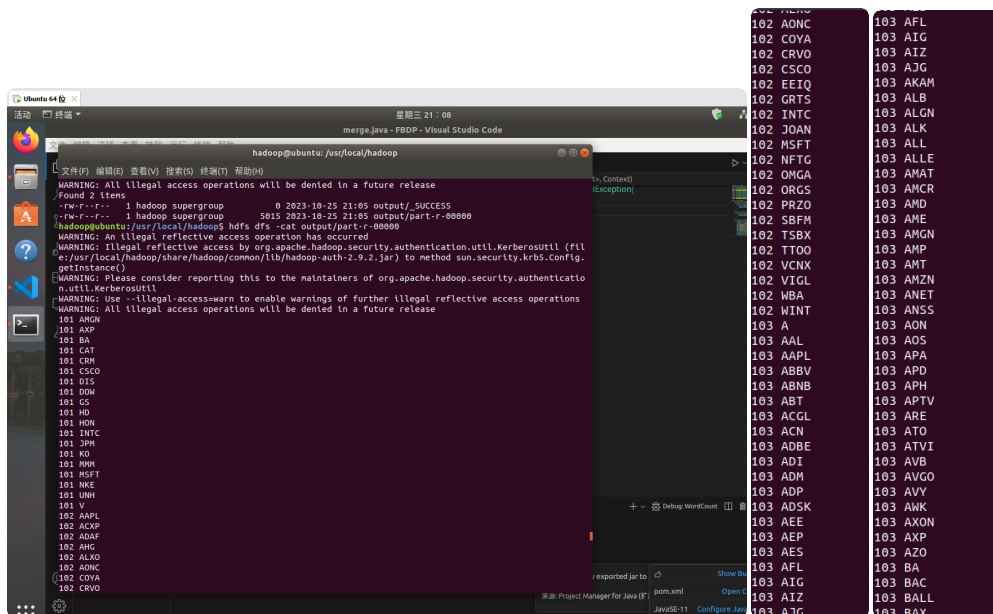
[git仓库链接](#)

<https://github.com/Cc17952/FBDP/tree/main>

设计思路

- map阶段：读取两个文件中的内容，每行内容为key，value取空值，保存为<key,value>的形式。
- reduce阶段：对所有的<key,value>进行reduce，由key的唯一性自然保证了去重并且合并，仍然保存为<key,value>的形式。

程序运行结果



```
hadoop@ubuntu: /usr/local/hadoop
WARNING: All illegal access operations will be denied in a future release
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2023-10-25 21:05 output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 5015 2023-10-25 21:05 output/part-r-00000
hadoop@ubuntu: /usr/local/hadoop$ hdfs dfs -cat output/part-r-00000
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file
es:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.9.2.jar) to method sun.security.krb5.Config.
getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication
n.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
101 AMGN
101 AXP
101 BA
101 CAT
101 CRM
101 CSO
101 DIS
101 DOW
101 GS
101 HJ
101 HON
101 INTC
101 JPM
101 KO
101 KRO
101 MSFT
101 NKE
101 UNH
101 V
102 AAPL
102 ACP
102 ADAF
102 AHG
102 ALKO
102 AONC
102 COVA
102 CSO
103 AFL
103 AIG
103 AIZ
103 AJG
103 AKAM
103 ALB
103 ALGN
103 ALK
103 ALL
103 ALLE
103 AMAT
103 AMCR
103 AMD
103 AME
103 AMGN
103 AMP
103 AMT
103 AMZN
103 ANET
103 ANSS
103 AON
103 AOS
103 APA
103 APD
103 APH
103 APTV
103 ARE
103 ATO
103 ATVI
103 AVB
103 AVGO
103 AVY
103 AWK
103 AXON
103 AXP
103 AZO
103 BA
103 BAC
103 BALL
103 BAY
```

进一步分析

- 目前的代码在一个文件中，不易于维护和扩展。考虑在更复杂的情形下可以将Mapper、Reducer、主程序和配置参数分成不同的类和文件，以提高代码的可读性和维护性。

- 代码中没有足够的错误处理。在生产环境中，必须考虑更多的错误情况，如文件不存在、权限问题等。
- 输入路径的个数硬编码在代码中，限制了程序的可重用性。