

# CS542 Project IV report

In this project, we aim to create a log file to redo the processing what we did on original data by following the instructor in course website.

First, we use the relation class created in project 3 to store data in a table and execute it with query.

Second, we need a Log class to record our actions on original data. In the log file, we use the (transaction name, pointer to the updated record, old data, new data) as the log. When user wants to create log file to record, he first needs to create a log file named by the data's file name followed by Log, i.e., CityLog. Each write process will call addLog() function to write corresponding log into log file. After finishing transaction, he can use pushLog() to push the log into disk. After that, if he wants to redo or read the log file, he could call readLog() to identify what he did before.

addLog(): Create a new record include transaction name, pointer to updated record, old value and new value.

pushLog(): After transaction, user needs to push the log file to disk.

readLog(): Read the log file from disk.

showLog(): Print the logs.

With these two classes: Log and Relation, we do our test following instructions in the course website. For city and country data file, we increase population of all records by 2%. At the same time, we also generate two undo/redo logs, respectively. Completing transactions, we push the logs into disk, and then we write updated city file and updated country file into disk. Then we read these logs to redo the transactions on backup files.

Finally, we check whether records in cs542.db are same with records in cs542A.db. We read cs542.db and cs542A.db into main memory and compare each pair of records contains same data. If every record in original file and backup file is element-wise, we return True, else we return false.

## Assumption:

1. The tests we run are pretty simple so we do not add checkpoints into log.
2. In this project, what we need to do is no more than updating population of each record in relation city and relation country. So in logs, we do not specify which attributes are updated.
3. One advantage of redo/undo logs is that users can write logs into disk before committing it. And system can decide whether undo or redo transaction based on whether it is committed or not. But since our task is simple so we push all logs at one time rather than writing logs into disk multiple times.

Here is the testing result:

Some logs in cityLog

```
start transaction UpdateCity
UpdateCity,3640,310100,316302
UpdateCity,2306,182142,185784
UpdateCity,3638,314800,321096
UpdateCity,2305,201569,205600
UpdateCity,3639,312000,318240
UpdateCity,2304,299118,305100
UpdateCity,2303,393030,400890
UpdateCity,2302,417517,425867
UpdateCity,2301,417810,426166
UpdateCity,2300,806475,822604
UpdateCity,3630,353400,360468
UpdateCity,3631,344500,351390
UpdateCity,3632,343300,350166
UpdateCity,3633,342000,348840
UpdateCity,3634,339300,346086
UpdateCity,2309,169841,173237
```

some logs in countryLog

```
start transaction UpdateCountry
UpdateCountry,NZL,3862000,3939240
UpdateCountry,FJI,817000,833340
UpdateCountry,PNG,4807000,4903140
UpdateCountry,GLP,456000,465120
UpdateCountry,STP,147000,149940
UpdateCountry,MHL,64000,65280
UpdateCountry,WLF,15000,15300
UpdateCountry,CUB,11201000,11425020
UpdateCountry,SDN,29490000,30079800
UpdateCountry,GMB,1305000,1331100
UpdateCountry,MYS,22244000,22688880
UpdateCountry,MYT,149000,151980
UpdateCountry,TWN,22256000,22701120
UpdateCountry,POL,38653600,39426672
UpdateCountry,OMN,2542000,2592840
UpdateCountry,SUR,417000,425340
```

Checking result:

```
city in cs542 and city in cs542A are identical or not?
the result is: true
country in cs542 and country in cs542A are identical or not?
the result is: true
```