Assignment 1 (Date 14-06-2022)

Q.Python

A:

Input:

a = [5,10,9,['Saif',12,'Sachin'],5,5,100]

Output:

[5, 10,9,['Saif',12,'Sachin'],100]

Ans:

```python
a = [5,10,9,['Saif',12,'Sachin','Sachin'],['Saif',12,'Sachin'],5,5,100]
b = list()

for val in a:

    if type(val) == int:
        if val in b:
            pass
        else:
            b.append(val)

    if type(val) == str:
        if val in b:
            pass
        else:
            b.append(val)
    if type(val) == list:
        if val in b:
            pass
        else:

            li = list()
            for element in val:

                if type(element) == int:
                    if element in li:
                        pass
                    else:
                        li.append(element)

                if type(element) == str:
                    if element in li:
                        pass
                    else:
                        li.append(element)
            b.append(li)

    print(b)
```

Output:

```
[5, 10, 9, ['Saif', 12, 'Sachin'], 100]
```

Q. *****Informatica Assignment*****

Input:

Col1   Col2   Col3

A   B   C

Output:

C

B

A

Solution:

output:

C:\Training\01_Sources\TGT\assignment_tgt1.out - Notepad++ [Administrator]

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

PROJECT_5_OUTPUT.out Increment_salary_project_5.txt SCD_TYPE_01_SRC.txt new 6 assignment1.txt assignment_tgt1.out

```
1    C
2    B
3    A
4
```
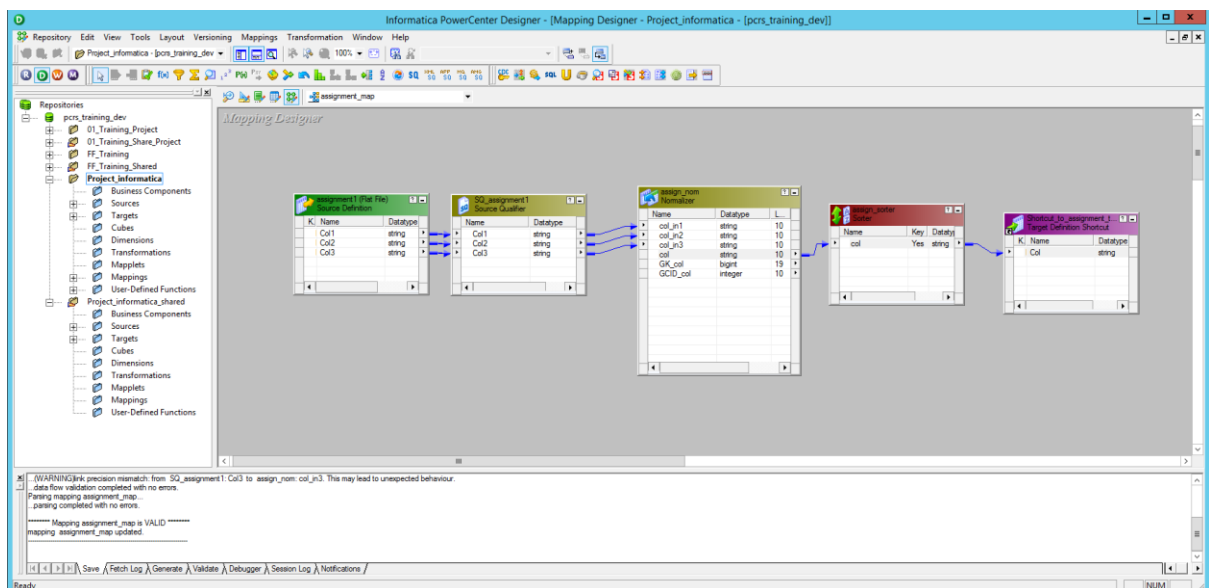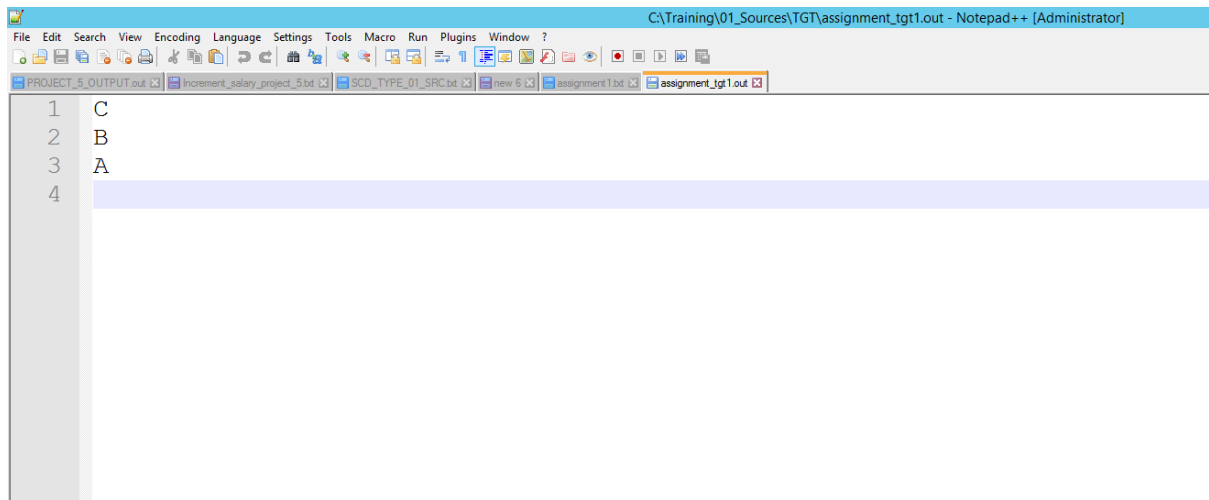
Q. *****Sqoop Assignment*****

1) Import below column data from retail_db to HDFS file system under directory ORDER_PRODUCT_PRICE.

Preview the count from the part files.

Output Record count: 172198.

```
+------------+------------------------------------------+----------------------+-------------------------+

| product_id | product_name                             | order_item_product_id | order_item_product_price |

+------------+------------------------------------------+----------------------+-------------------------+

|     957 | Diamondback Women's Serene Classic Comfort Bi |           957 |           299.98 |

|    1073 | Pelican Sunstream 100 Kayak              |          1073 |           199.99 |

|     502 | Nike Men's Dri-FIT Victory Golf Polo     |           502 |            50 |

|     403 | Nike Men's CJ Elite 2 TD Football Cleat  |           403 |           129.99 |

|     897 | Team Golf New England Patriots Putter Grip  |           897 |            24.99 |

+------------+------------------------------------------+----------------------+-------------------------+
```

Note: Above Sqoop Assignment needs to be automated using Shell Script with Validations, logging, Error Handling,

Success Failure message.

Solution:

Script:

```bash
#!/bin/bash


# sourcing the parameter file

. /home/saif/cohort_F11/env/sqp.prm


# validating command line arguments

if [ $# -eq 0 ]

then

        echo "No command line arguments are passes"

        echo "Please try below command"

        echo "bash [file_name] [space_seperated_arguments]"

else


LOG_DIR=/home/saif/cohort_F11/logs/

FILE_NAME=`basename $0`

DATE_STAMP=`date '+%Y%m%d_%H:%M:%S'`


LOG_FILE_NAME=${LOG_DIR}${FILE_NAME}_${DATE_STAMP}.log


#checking if directory exist or not

if hadoop fs -test -d ${OP_DIR}${1}

    then
```

```bash
				# deletes directory if exists

		hadoop fs -rm -r ${OP_DIR}${1}

fi



# runs sqoop command importing data from RDBMS to the HDFS

sqoop import \

		--connect jdbc:mysql://${HOST}:${PORT}/${DB_NAME}?useSSL=False \

		--username ${USERNAME} \

		--password-file ${PASSWORD_FILE} \

		--query 'select p.product_id product_id,  p.product_name, o.order_item_product_id,
o.order_item_product_price from order_items o right join products p on p.product_id = o.order_item_product_id where
o.order_item_product_id is not null and $CONDITIONS' \

		-m 1 \

		--target-dir ${OP_DIR}${1}


echo "Data Ingestion of ${1} completed sucessfully." > ${LOG_FILE_NAME}



fi
```

```
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=8955679
2022-06-15 23:55:27,465 INFO mapreduce.ImportJobBase: Transferred 8.5408 MB in 40.0602 seconds (218.316 KB/sec)
2022-06-15 23:55:27,468 INFO mapreduce.ImportJobBase: Retrieved 172198 records.
saif@smidsy-technologies:~/cohort_F11/scripts$
```

```
saif@smidsy-technologies:~/cohort_F11/scripts$ hdfs dfs -ls HFS/Output/
2022-06-15 23:58:42,968 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builti
n-java classes where applicable
Found 8 items
drwxr-xr-x   - saif supergroup          0 2022-06-14 16:21 HFS/Output/Inc_Imp_Dt_tmstmp
drwxr-xr-x   - saif supergroup          0 2022-06-15 23:58 HFS/Output/ORDER_PRODUCTS_PRICE
drwxr-xr-x   - saif supergroup          0 2022-06-14 11:35 HFS/Output/all_tables
drwxr-xr-x   - saif supergroup          0 2022-06-14 11:42 HFS/Output/all_tables1
drwxr-xr-x   - saif supergroup          0 2022-06-14 11:50 HFS/Output/all_tables2
drwxr-xr-x   - saif supergroup          0 2022-06-14 11:17 HFS/Output/cohort_f11
drwxr-xr-x   - saif supergroup          0 2022-06-14 11:55 HFS/Output/orders
drwxr-xr-x   - saif supergroup          0 2022-06-15 23:55 HFS/Output/products
saif@smidsy-technologies:~/cohort_F11/scripts$
```

output:

```
saif@smidsy-technologies:~/cohort_F11/scripts$ hdfs dfs -ls HFS/Output/ORDER_PRODUCTS_PRICE
2022-06-15 23:59:27,912 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builti
n-java classes where applicable
Found 2 items
-rw-r--r--   1 saif supergroup          0 2022-06-15 23:58 HFS/Output/ORDER_PRODUCTS_PRICE/_SUCCESS
-rw-r--r--   1 saif supergroup    8955679 2022-06-15 23:58 HFS/Output/ORDER_PRODUCTS_PRICE/part-m-00000
saif@smidsy-technologies:~/cohort_F11/scripts$
```