Problem Statement:

You have been given a dataset of bank transactions. The file contains the transaction amount, account Id, transaction Id and few other attributes. The aim is to find out the fraudulent transactions. Here, instead of going for a machine learning model and much more advanced stuff, we will follow a simple algorithm. Generally, the fraudsters first draw a tiny amount of money from the account and then another with a huge amount.

We will simulate the real-world scenario, where every transaction comes to the data pipeline as it occurs. The transactions will be received from Kafka, then the possible candidate for fraudulent transactions will be written to the sink. In this case, the sink will be a hdfs file. In the real world, the analytics engineer can query the data in hive to display in the report.

Input & output criteria:

The dataset contains a single file: transactions.csv. The file contains these columns:

```
transactionId, time, type, amount, sourceAccountId, destinationAccountId,
sourceOldBalance, sourceNewBalance, destinationOldBalance, destinationNewBalance
```

You need to send the file's content to your kafka topic. You can use the following command for sending the file's content into a kafka topic. Update the bootstrap server & file path accordingly.

```
bin/kafka-console-producer.sh --bootstrap-server localhost:9092 --topic test <
../../DataSet/transaction.csv
```

If you do not have the topic created, you may get errors while running the above command. Please go ahead and create the topic in that case.

Once the transaction data has been processed, the output needs to be written to HDFS, as and when the fraudulent transaction is detected. The format should be the same as the input.

Program argument:

The flink program needs to be provided with the input & output details. A sample program argument will be like this:

```
--bootstrap.servers localhost:9092 --group.id test --topic test --output
hdfs:///user/root/transaction_op.txt
```

Update the bootstrap server, topic & output file path as required.

Good Luck!