Parser

The goal is to write a parser in Java that parses a web server access log file, loads the log to MySQL and checks if a given IP makes more than a certain number of requests for the given duration.

Java
---

1. Create a java tool that can parse and load the given log file to MySQL.

The delimiter of the log file is pipe (|)

2. The tool takes "startDate", "duration" and "threshold" as command line arguments. "startDate" is of "yyyy-MM-dd.HH:mm:ss" format, "duration" can take only "hourly", "daily" as inputs and "threshold" can be an integer.

3. This is how the tool works:

```
java -cp "parser.jar" com.ef.Parser --startDate=2017-01-01.13:00:00 --duration=hourly
--threshold=100
```

The tool will find any IPs that made more than 100 requests starting from 2017-01-01.13:00:00 to 2017-01-01.14:00:00 (one hour) and print them to console AND also load them to another MySQL table with comments on why it's blocked.

```
java -cp "parser.jar" com.ef.Parser --startDate=2017-01-01.13:00:00 --duration=daily
--threshold=250
```

The tool will find any IPs that made more than 250 requests starting from 2017-01-01.13:00:00 to 2017-01-02.13:00:00 (24 hours) and print them to console AND also load them to another MySQL table with comments on why it's blocked.

SQL
---

1. Write MySQL query to find IPs that mode more than a certain number of requests for a given time period.

E.g. Write SQL to find IPs that made more than 100 requests starting from 2017-01-01.13:00:00 to 2017-01-01.14:00:00.

2. Write MySQL query to find requests made by a given IP.

LOG Format

Date, IP, Request, Status, User Agent (pipe delimited, open the example file in text editor)

Date Format: "yyyy-MM-dd HH:mm:ss.SSS"

Also, please find an attached a log file for your reference.

The log file assumes 200 as hourly limit and 500 as daily limit, meaning that:

1. When you run your parser against this file with the following parameters

```
java -cp "parser.jar" com.ef.Parser --startDate=2017-01-01.15:00:00 --duration=hourly
--threshold=200
```

The output will have 192.168.11.231.

If you open the log file, 192.168.11.231 has 200 or more requests between 2017-01-01.15:00:00 and 2017-01-01.15:59:59

2. When you run your parser against this file with the following parameters

```
java -cp "parser.jar" com.ef.Parser --startDate=2017-01-01.00:00:00 --duration=daily
--threshold=500
```

The output will have 192.168.102.136. If you open the log file, 192.168.102.136 has 500 or more requests between 2017-01-01.00:00:00 and 2017-01-01.23:59:59

Deliverables
---

1. Java program that can be run from command line

   ```
   java -cp "parser.jar" com.ef.Parser --accesslog=/path/to/file
   --startDate=2017-01-01.13:00:00 --duration=hourly --threshold=100
   ```

2. Source Code for the Java program

3. MySQL schema used for the log data

4. SQL queries for SQL test

EXECUTION
---

1. You will find an executable jar file named parser.jar.

2. You could run -

```
java -cp "parser.jar" com.ef.Parser --accesslog=/path/to/file --startDate=2017-01-01.13:00:00
--duration=hourly -- threshold=100
```

3. For Example:

```
java -cp "parser.jar" com.ef.Parser X:/access.log 2017-01-01.15:00:00 daily 200
```

4. Then it asks if you want to store the data to DataBase. If you press ‘Y’,

You will be asked for the MYSQL URL.
Enter - jdbc:mysql://localhost:3306.
Hoping MYSQL is installed on your local machine.

5. Then it asks for 'username' and 'password.
Enter them. The program will create the database for you and store the records.
You need not create the Schema.

6. If you Press 'N', No data is stored to the database.

7. All the Matching IP’s are printed in both the cases.

8. You can also Type/Enter:

'R' or 'L' or 'Q'

'R' = readingEntireFileWithoutLoop
'L' = loadTheFile
'Q' = runQueries

ADDITIONAL NOTES
---

a. Start Mysql via: net start MySQL.

or if you are using XAMMP point to location e.g. E:\Program Files\xampp\mysql\bin>mysql -u root

b. Start XAMPP.

c. Create a test database called parser.

d. Create two test tables.

e. Here is one row of the log file.

2017-01-01 23:59:59.608|192.168.122.135|"GET / HTTP/1.1"|200|"Mozilla/5.0 (Windows NT 6.3; Win64; x64; rv:53.0) Gecko/20100101 Firefox/53.0"

f. Start eclipse and do a project named parser.

For this project I used the following:

    java version "1.8.0_111"
    Java(TM) SE Runtime Environment (build 1.8.0_111-b14)
    Java HotSpot(TM) Client VM (build 25.111-b14, mixed mode)

    Eclipse Neon

    mysql version is:

    mysql Ver 14.14 Distrib 5.1.41, for Win32 (ia32)

    XAMPP Control Version 2.5.8 (ApacheFriends Edition)

    phpMyAdmin 3.2.4

If you are going to play around with the source I encourage you to

make sure that you place the two critical jars files at:

E.g. - X:\Java\jdk1.8.0_111\jre\lib\ext

They are: mysql-connector.jar and mysql-connector-java-8.0.12.jar

Note due to the fact that my version of mysql is a little old,

I could not employ certain functionality like:

a. the function ‘INET6_ATON’

b. the function ‘INET6_NTOA’ for back conversion of the IP Address

What I used instead was:

a. inet_ntoa(ip)

With respect to running the application from the dos prompt

I suggest using this:

java -cp "parser.jar" com.ef.Parser X:/access.log 2017-01-01.15:00:00 daily 200

That does depend on if you place the log file e.g. at X:/access.log

The log file is fairly large and contains 116484 log records in it.

That is one of the reasons I created a tiny text file containing one row in it.

The preferred order of running is:

Y, or N,

R, or L,

Q should be the last one.

Instructions for the WalletHub Test
