

LF08 – C# – Projekt

BSE Web Log ETL

1. Projekt-Beschreibung

Das aus dem zurückliegenden LF05-Projekt zur Analyse einer Webserver-log-Datei¹ entstandene Programm bewla.exe konnte lediglich eine log-Datei analysieren. Um den Zugriff auf einen Webserver über einen längeren Zeitraum analysieren zu können, ist es jedoch notwendig, viele dieser log-Dateien in einer Analyse zu berücksichtigen.

Dieses LF08-Projekt hat daher zum Inhalt, eine Software zu schreiben, die es erlaubt, log-Dateien in einem sog. ETL-Prozess² in einer SQLite-Datenbank zusammenzuführen und diese bereitgestellten Daten zu analysieren.

Die Software soll als Windows-Desktop-Client in C# realisiert werden, wobei entweder WinForms oder WPF verwendet wird.

2. Logfile

Für dieses Projekt kommt die bereits bekannte log-Datei aus dem LF05-Projekt zu Einsatz. Hier noch einmal der Ausschnitt eines logfiles in der hier verwendeten Form:

```
48.76.177.1 - - [2019-01-22 00:16:26.819379] "GET /cds/snippets/themes/customizr/assets/shared/fonts/fa/webfonts/fa-solid-900.woff2 HTTP/1.1" 200 36772
48.76.177.1 - - [2019-01-22 00:16:26.868053] "GET /cds/includes/js/emoji-release.min.js?ver=4.9.8 HTTP/1.1" 200 4382
48.76.177.1 - - [2019-01-22 00:16:26.862946] "GET /cds/snippets/themes/customizr/assets/shared/fonts/fa/webfonts/fa-brands-400.woff2 HTTP/1.1" 200 53928
48.76.177.1 - - [2019-01-22 00:16:27.115903] "GET /cds/snippets/themes/customizr/inc/assets/img/glyphicons-halflings.png HTTP/1.1" 200 12799
48.76.177.1 - - [2019-01-22 00:16:26.629916] "GET /cds/snippets/uploads/2013/10/nms3.jpg HTTP/1.1" 200 57443
48.76.177.1 - - [2019-01-22 00:16:26.627159] "GET /cds/snippets/uploads/2013/09/skyline5.jpg HTTP/1.1" 200 137901
48.76.177.1 - - [2019-01-22 00:16:26.619373] "GET /cds/snippets/uploads/2013/09/wall51.jpg HTTP/1.1" 200 277813
213.148.60.1 - - [2019-01-22 00:16:44.654435] "GET /cds/ HTTP/1.1" 200 31332
109.60.1.1 - - [2019-01-22 00:19:24.291334] "GET /cds/feed/ HTTP/1.1" 304 -
70.163.125.1 - - [2019-01-22 00:19:26.218422] "HEAD /cds/ HTTP/1.1" 200 -
213.148.60.1 - - [2019-01-22 00:21:44.809110] "GET /cds/ HTTP/1.1" 200 31332
41.78.189.1 - - [2019-01-22 00:22:23.498156] "GET /cds/snippets/themes/customizr/assets/shared/fonts/fa/webfonts/fa-solid-900.ttf HTTP/1.1" 200 97644
41.78.190.1 - - [2019-01-22 00:22:24.492875] "GET /cds/snippets/themes/customizr/assets/shared/fonts/fa/webfonts/fa-regular-400.ttf HTTP/1.1" 200 30900
41.78.189.1 - - [2019-01-22 00:22:25.238243] "GET /cds/snippets/themes/customizr/assets/shared/fonts/fa/webfonts/fa-brands-400.ttf HTTP/1.1" 200 97348
70.163.125.1 - - [2019-01-22 00:24:26.099570] "HEAD /cds/ HTTP/1.1" 200 -
83.103.20.1 - - [2019-01-22 00:25:22.442329] "GET /cds/tutorials/example-guide-9_setting-up-a-reverse-proxy/ HTTP/1.1" 200 23480
6.102.117.1 - - [2019-01-22 00:25:24.232343] "GET /cds/tutorials/example-guide-9_setting-up-a-reverse-proxy/ HTTP/1.1" 200 77936
186.87.165.1 - - [2019-01-22 00:25:57.203575] "GET /cds/login.do HTTP/1.1" 200 2853
186.87.165.1 - - [2019-01-22 00:25:57.323673] "POST /cds/login.do HTTP/1.1" 200 3636
213.148.60.1 - - [2019-01-22 00:26:44.967592] "GET /cds/ HTTP/1.1" 200 31332
78.76.79.1 - - [2019-01-22 00:26:55.490270] "GET /robots.txt HTTP/1.1" 200 24
78.76.79.1 - - [2019-01-22 00:26:55.862005] "GET /cds/?q=example-guide-4-ssl-server-konfigurieren HTTP/1.1" 200 8107
180.62.135.1 - - [2019-01-22 00:27:27.834638] "GET /cds/tutorials/example-guide-6-einbinden/ HTTP/1.0" 200 31678
6.102.117.1 - - [2019-01-22 00:28:07.920633] "GET /cds/tutorials/example-guide-9_setting-up-a-reverse-proxy/ HTTP/1.1" 200 77936
70.163.125.1 - - [2019-01-22 00:29:26.254711] "HEAD /cds/ HTTP/1.1" 200 -
202.156.195.1 - - [2019-01-22 00:30:54.932121] "GET /cds/tutorials/example-guide-9_setting-up-a-reverse-proxy/ HTTP/1.1" 200 23480
202.156.195.1 - - [2019-01-22 00:30:55.813548] "GET /cds/snippets/themes/customizr/assets/shared/fonts/fa/css/fontawesome-all.min.css?ver=4.1.12 HTTP/1.1" 200 7439
202.156.195.1 - - [2019-01-22 00:30:56.236505] "GET /cds/snippets/themes/customizr-child/inc/assets/css/neagreen.min.css?ver=4.1.12 HTTP/1.1" 200 2458
202.156.195.1 - - [2019-01-22 00:30:56.250365] "GET /cds/snippets/themes/customizr-child/style.css?ver=4.1.12 HTTP/1.1" 200 483
202.156.195.1 - - [2019-01-22 00:30:56.445271] "GET /cds/includes/js/jquery/jquery-migrate.min.js?ver=1.4.1 HTTP/1.1" 200 4014
202.156.195.1 - - [2019-01-22 00:30:56.243483] "GET /cds/snippets/themes/customizr/inc/assets/css/tc_common.min.css?ver=4.1.12 HTTP/1.1" 200 28225
202.156.195.1 - - [2019-01-22 00:30:56.654019] "GET /cds/includes/js/comment-reply.min.js?ver=4.9.8 HTTP/1.1" 200 589
202.156.195.1 - - [2019-01-22 00:30:56.661015] "GET /cds/includes/js/jquery/ui/core.min.js?ver=1.11.4 HTTP/1.1" 200 1821
202.156.195.1 - - [2019-01-22 00:30:56.667534] "GET /cds/snippets/themes/customizr/assets/front/js/libs/retina.min.js?ver=4.1.12 HTTP/1.1" 200 1127
202.156.195.1 - - [2019-01-22 00:30:56.667991] "GET /cds/snippets/themes/customizr/assets/front/js/libs/modernizr.min.js?ver=4.1.12 HTTP/1.1" 200 5926
202.156.195.1 - - [2019-01-22 00:30:56.458703] "GET /cds/snippets/themes/customizr/inc/assets/js/tc-scripts.min.js?ver=4.1.12 HTTP/1.1" 200 36038
```

¹ http://httpd.apache.org/docs/current/mod/mod_log_config.html#formats

² <https://de.wikipedia.org/wiki/ETL-Prozess>

3. Anforderungsbeschreibung

Es versteht sich von selbst, dass das Programm ohne Abstürze läuft und im Fehlerfall geeignete Fehlermeldungen angezeigt.

Für die gesamte Anwendung ist die .NET-SQLite-Bibliothek Microsoft.Data.Sqlite zu verwenden. Andere Bibliotheken oder das Entity-Framework sind nicht erlaubt.

Die folgenden Anforderungsbereiche ordnen die Funktionalität der Anwendung.

3.1 Anforderungsbereich GUI

Es gibt für den Aufbau des GUI keine Vorgaben, sondern ist von jeder Projektgruppe selbst zu entwickeln. Es gilt lediglich, dass die folgenden Funktionalitäten vorhanden sein müssen:

- Der Benutzer muss die log-Dateien auswählen können, die in die Datenbank geladen werden sollen. Die Dateiendung ist .log.
- Die in die Datenbank bereits importierten log-Dateien müssen angezeigt werden.
- Das erneute Importieren von bereits importierten log-Dateien muss verhindert werden.
- Der Benutzer muss die unten aufgeführten Analysen durchführen können.

3.2 Anforderungsbereich Datenbank

Es ist eine SQLite-Datenbank zu verwenden.

Die Struktur der Datenbank ist komplett selbst zu entwickeln.

3.3 Anforderungsbereich Analyse

Dieser Anforderungsbereich muss komplett durch an die Datenbank gerichtete SQL-Abfragen realisiert werden.

Alle angegebenen IPs werden im Format IPv4 erwartet und müssen vom prinzipiellen Aufbau her korrekt sein. Beispielsweise wäre eine IP 123.12.12.967 nicht korrekt. Eine Unterscheidung bzgl. öffentlicher oder privater IP-Bereiche ist nicht nötig.

Die folgenden Analysemöglichkeiten soll die Software bieten:

Analyse 1	Anzeigen von Log-Datei-Einträgen
Beschreibung	Es werden die Anzahl sowie die Inhalte aller log-Datei-Einträge dem Benutzer angezeigt. Da das evtl. sehr viele sein können, muss dem Benutzer die Möglichkeit zum Filtern gegeben werden. Die im Folgenden angegebenen Filter können weggelassen oder auch kombiniert werden.
Zeitraum-Filter	Der Benutzer kann einen Zeitraum angeben, aus dem er die Inhalte angezeigt bekommt. Zeiträume sind hierbei nicht auf Tage beschränkt.
IP-Filter	Der Benutzer kann eine oder mehrere IPs angeben, für die er die Inhalte angezeigt bekommt.
Attribut-Filter	Der Benutzer kann die Attribute der log-Datei-Einträge auswählen, die er angezeigt bekommt.

Analyse 2	Analyse der Anzahl an Zugriffen pro IP
Beschreibung	Ermittelt die Anzahl an Einträgen für jede vorkommende IP und gibt diese aus. Da das evtl. sehr viele sein können, muss dem Benutzer die Möglichkeit zum Filtern gegeben werden. Die im Folgenden angegebenen Filter können weggelassen oder auch kombiniert werden.
Zeitraum-Filter	Der Benutzer kann einen Zeitraum angeben, aus dem er die Inhalte angezeigt bekommt. Zeiträume sind hierbei nicht auf Tage beschränkt.
IP-Filter	Der Benutzer kann eine oder mehrere IPs angeben, für die er die Inhalte angezeigt bekommt.

Analyse 3	Analyse der Anzahl an Einträgen mit den request-methods GET, POST und HEAD.
Beschreibung	Ermittelt die Anzahl an Einträgen für jede vorkommende request-method und gibt diese aus. Es soll auch hier dem Benutzer die Möglichkeit zum Filtern gegeben werden. Die im Folgenden angegebenen Filter können weggelassen oder auch kombiniert werden.
Method-Filter	Der Benutzer kann eine oder mehrere der request-methods angeben.
Zeitraum-Filter	Der Benutzer kann einen Zeitraum angeben, aus dem er die Inhalte angezeigt bekommt. Zeiträume sind hierbei nicht auf Tage beschränkt.
IP-Filter	Der Benutzer kann eine oder mehrere IPs angeben, für die er die Inhalte angezeigt bekommt.

Analyse 4	Analyse der Anzahl an Einträgen mit den verschiedenen error-codes.
Beschreibung	Ermittelt die Anzahl an Einträgen für jeden vorkommenden error-code und gibt diese aus. Es soll auch hier dem Benutzer die Möglichkeit zum Filtern gegeben werden. Die im Folgenden angegebenen Filter können weggelassen oder auch kombiniert werden.
code-Filter	Der Benutzer kann einen oder mehrere der error-codes angeben.
Zeitraum-Filter	Der Benutzer kann einen Zeitraum angeben, aus dem er die Inhalte angezeigt bekommt. Zeiträume sind hierbei nicht auf Tage beschränkt.
IP-Filter	Der Benutzer kann eine oder mehrere IPs angeben, für die er die Inhalte angezeigt bekommt.

4 Präsentation – nach der Abgabe zählt für das zweite Halbjahr

Jede Projektgruppe muss die entwickelte Lösung in einer kurzen Präsentation vorführen. Hierbei ist neben dem GUI auch auf die dahinterliegende Architektur der Anwendung einzugehen und ein Quellcode-Fragment exemplarisch zu erläutern.

5 Abgabe

Das Projekt ist als (auf den Schulrechnern!) lauffähiges Programm abzugeben. Zusätzlich dazu ist das VisualStudio-Projekt abzugeben.

Das Projekt kann zu zweit oder zu dritt bearbeitet werden.

Die Abgabe erfolgt über den moodle-Kurs am: 13.01.2022
zum Ende der Unterrichtsstunde um: 11:20 Uhr.