

Привет! Спасибо за интерес к нашей компании. В этом тестовом задании вы разработаете инструмент для парсинга логов.

Business Solutions Department разрабатывает высоконагруженную систему аналитики и маркетинга, обрабатывающую запросы со всех игровых и рекламных серверов компании в режиме онлайн. Поэтому качество инструментов и алгоритмов, которые проектирует и развивает департамент, крайне важно.

Задание

Разработать комплекс приложений для парсинга логов web-сервера Apache (apache_access_log). Результаты необходимо сохранить в базе данных и представить в виде JSON-файла.

В качестве тестового набора данных можно взять любой из этих файлов:

- ita.ee.lbl.gov/traces/NASA_access_log_Jul95.gz;
- <http://www.monitorware.com/en/logsamples/download/apache-samples.rar>.

Спецификация

1. При парсинге необходимо получить из лога такой набор данных:
 - дату и время запроса;
 - IP-адрес или имя хоста, который выполнял запрос;
 - роут, к которому был реализован запрос;
 - дополнительные параметры, с которыми происходил запрос (URL Query Parameters);
 - результат выполнения запроса;
 - размер в байтах, переданный клиенту.
2. Запросы на получение картинок, CSS-файлов и скриптов обрабатывать не нужно.
3. Набор данных, полученных из парсинга логов, необходимо расширить полем «Геолокация» и при помощи доступных онлайн-сервисов заполнить его страной клиента.
4. Результаты парсинга хранятся в реляционной базе данных (MS SQL / MySQL).
5. Приложение парсера выполняется как отдельный проект — консольное или GUI-приложение в Windows (Windows Forms / WPF).
6. Приложение для отображения результатов парсинга — ASP.NET WebAPI с набором роутов для отображения результатов парсинга в формате JSON.
7. Роуты, которые необходимо создать в этом приложении:

| Описание метода | Параметры |
|--|---|
| получить топ-N хостов, отсортированных по убыванию количества запросов за указанный промежуток времени | n — количество хостов (по умолчанию — 10) start — дата начала промежутка времени (по умолчанию — не фильтровать) end — дата окончания промежутка времени (по умолчанию — не фильтровать) |
| получить топ-N роутов, отсортированных по убыванию количества запросов за указанный промежуток времени | n — количество роутов (по умолчанию — 10) start — дата начала промежутка времени (по умолчанию — не фильтровать) end — дата окончания промежутка времени (по умолчанию — не фильтровать) |
| получить полный лог запросов, отсортированный по увеличению времени за указанный интервал | start — дата начала промежутка времени (по умолчанию — не фильтровать) end — дата окончания промежутка времени (по умолчанию — не фильтровать) offset — индекс первой выводимой сущности в результирующей коллекции limit — количество выводимых сущностей (по умолчанию — 10) |

Технические требования

- Проект должен быть написан на языке C# версии 4+.
- Может быть использован ADO.NET или произвольные ORM для организации взаимодействия с базой данных.
- Готовая работа содержит разумное количество комментариев.
- Проект парсера должен работать с учетом того, что исходный лог может быть большим.
- Проект реализован с использованием одного из принятых стандартов кодирования.
- Готовая работа содержит:
 1. исходные файлы, необходимые для компиляции и запуска приложений;
 2. пример исходного лога, с которым приложение работает;
 3. SQL-скрипт создания структуры базы данных.

Дополнительная информация

Мы ценим ваше время и сделали тестовое задание лаконичным настолько, насколько это возможно, чтобы оценить ваш профессиональный уровень. Обычно на выполнение уходит до 7 дней.

Мы свяжемся с вами в течение 5 рабочих дней после получения готовых материалов. Если у вас возникли какие-либо вопросы, задавайте их — мы обязательно ответим. Удачи!