Министерство образования Республики Беларусь

Учреждение образования

Белорусский государственный университет  
информатики и радиоэлектроники

Кафедра экономической информатики

ОТЧЁТ  
по лабораторной работе №8  
на тему «NoSQL базы данных»

Выполнила:

студентка группы 772303

Желнерович Т.С.

Проверил:

Кунцевич А.А.

Минск 2020

Задание

Используя СУБД MongoDB, разработайте базу данных, предназначенную для хранения логов веб-сервера. Лог включает в себя следующие поля: адрес ресурса (URL), IP-адрес пользовательского компьютера, отметка времени начала просмотра ресурса, длительность просмотра ресурса.

Указания к заданию

1. Разработайте консольную утилиту для преобразования лога веб-сервера в формате CSV (Comma Separated Values), в формат JSON. Лог должен содержать поля со следующими названиями: URL, IP, timeStamp, timeSpent.

2. Разработайте запросы для загрузки полученных данных в формате JSON в СУБД MongoDB.

Разработка запросов в СУБД MongoDB

1. Разработайте следующие запросы, используя встроенные в СУБД MongoDB средствами выборки:

1) Выдать упорядоченный список URL ресурсов.

2) Выдать упорядоченный список IP-адресов пользователей, посетивших ресурс с заданным URL.

3) Выдать упорядоченный список URL ресурсов, посещенных в заданный временной период.

4) Выдать упорядоченный список URL ресурсов, посещенных пользовате-лем с заданным IP-адресом.

2. Разработайте следующие запросы, используя встроенные в СУБД MongoDB средства программирования на основе парадигмы MapReduce:

1) Выдать список URL ресурсов с указанием суммарной длительности по-сещения каждого ресурса, упорядоченный по убыванию.

2) Выдать список URL ресурсов с указанием количества посещений каждого ресурса в день за заданный период, упорядоченный URL ресурса и убыванию количества посещений.

3) Выдать список IP-адресов c указанием суммарного количества и суммар-ной длительности посещений ресурсов, упорядоченный по адресу, убыванию количества и убыванию длительности.

Ход работы

Создадим базу данных под именем Lab\_8, в ней создадим коллекцию logs и заполним БД данными:



Рисунок 1 – Заполение БД данными

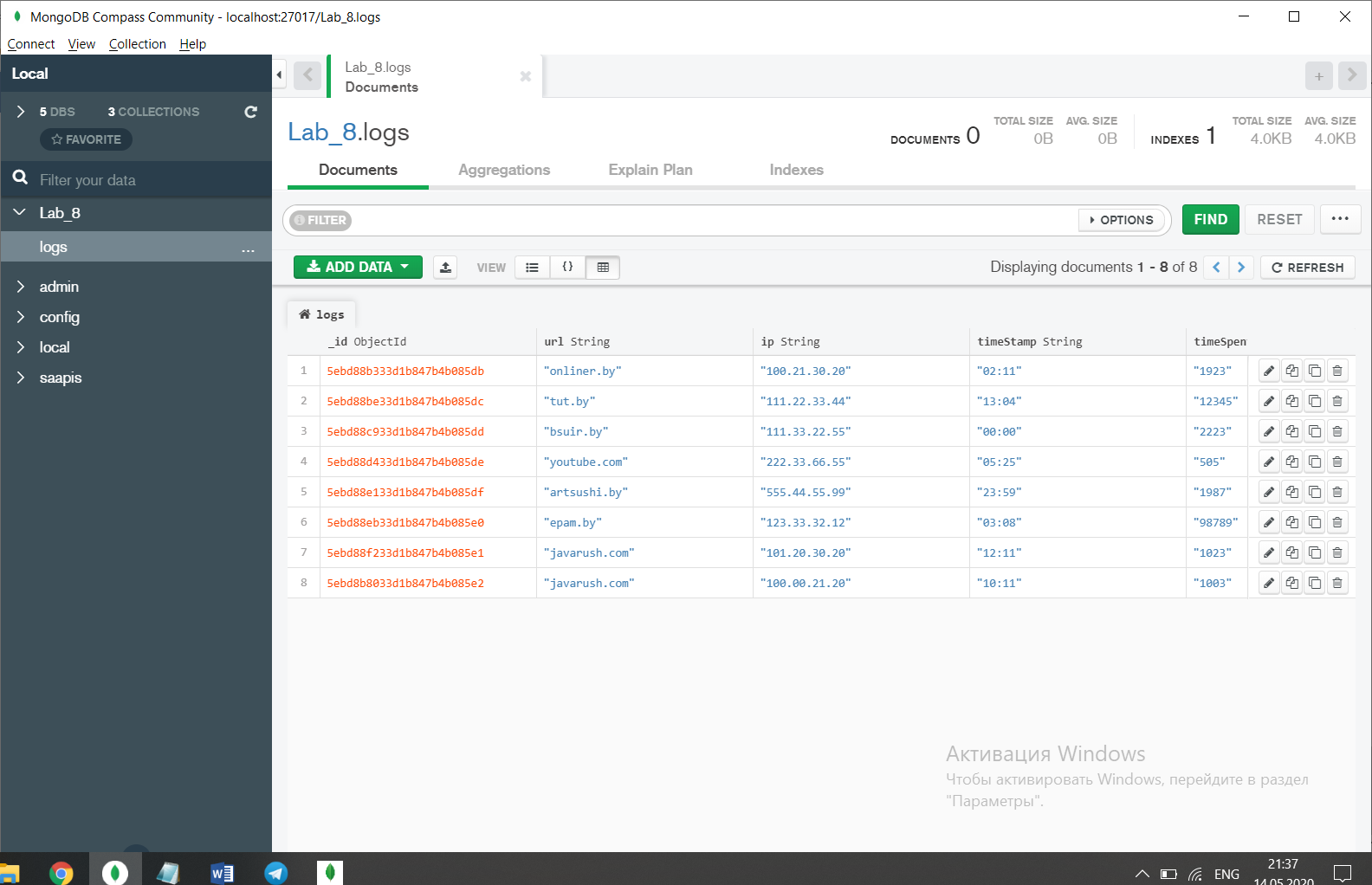


Рисунок 2 – Результат заполнения БД

Чтобы преобразовать данные из БД в файл формата json, воспользуемся следующей функцией:

mongoimport --db Lab\_8 --collection csvToLogs --file logs.csv --type csv - - headerline

mongoexport --db Lab\_8 --collection logs --out /home/user/logs.json

Чтобы наоборот – импортировать данные из файла json в БД, необходимо возспользоваться следующей командой:

mongoimport --db Lab\_8 --collection jsonToLogs --file logs.json --type json

Выведем упорядоченный список URL помощью команды:

db.logs.find({},{\_id:0,ip:0,timeStamp:0,timeSpent:0}).sort({url:1})

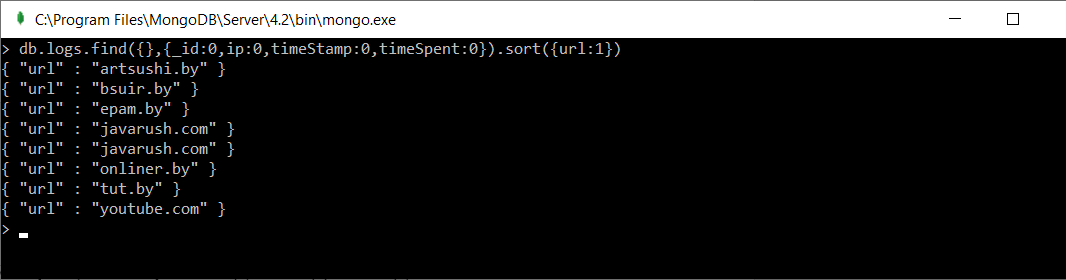


Рисунок 3 – Упорядоченный список URL

Выведем упорядоченный список IP-адресов пользователей, посетивших сай с заданным URL (офмфкгырюсщь) с помощью команды:

db.logs.find({url:"[javarush.com](https://yandex.com/)"},{\_id:0,url:0,timeStamp:0,timeSpent:0})

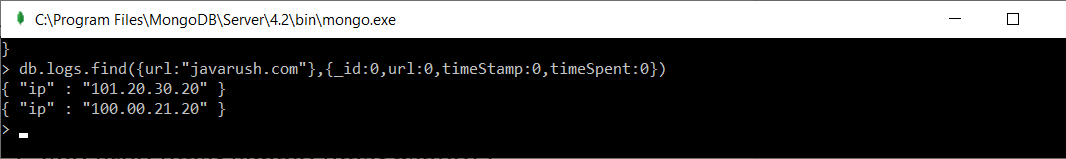


Рисунок 4 – Упорядоченный список IP-адресов пользователей, посетивших сай с заданным URL

Выведем упорядоченный список URL ресурсов, которые посещали пользователи с заданными IP с помощью команды:

db.logs.find({ip:"100.21.30.20"},{\_id:0,timeStamp:0,timeSpent:0,ip:0})

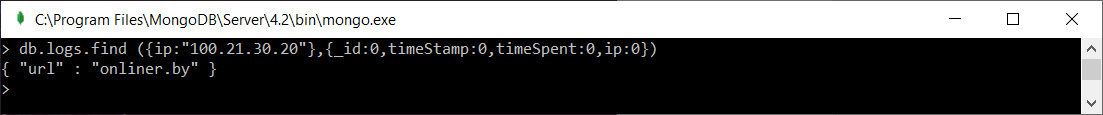


Рисунок 5 – Список URL ресурсов, посещенных пользователем с заданным IP



Рисунок 6 – Список URL ресурсов с указанием суммарной длительности посещения ресурса

Чтобы вывести список URLресурсов с указанием длительности суммарного сеанса, используем следующую команду:

function map() {

var urls = this.url;

var length = this.timeSpent;

emit(urls, length);

}

function reduce(key, values) {

var totalTimeSpent = 0;

for (var i = 0; i < values.length; i++) {

totalTimeSpent += values[i];

}

return totalTimeSpent;

}

db.logs.mapReduce(map,reduce,{out:"url"})

db.url.find().sort({value:-1})

Далее выведем список каждого URL с суммарным количеством посещений с помощью команды:

function map() {

var urls = this.url;

emit(urls, 1);

}

function reduce(key, values) {

var count = 0;

for (var i = 0; i < values.length; i++) {

count += values[i];

}

return count;

}

db.logs.mapReduce(map,reduce,{out:"visits"})

db.visits.find().sort({value:-1})



Рисунок 7 – Список URL ресурсов с указанием количества посещений ресурса

Выведем список URL ресурсов с указанием количества посещений каждого ресурса в день за заданный промежуток с помощью команды:

function map() {

var url = this.url;

var date = new Date(this.timeStamp \* 1000);

emit({url: url, date: date}, 1);

}

function reduce(key, values) {

var count = 0;

for (var i = 0; i < values.length; i++) {

count += values[i];

}

return count;

}

db.logs.mapReduce(map,reduce,{out:"day"})

db.day.find().sort({url:1,value:-1})



Рисунок 8 – Список URL ресурсов с указанием количества посещений каждого ресурса в заданный промежуток времени



Рисунок 9 – Список URL ресурсов с указанием количества посещений каждого ресурса и суммарной длительности