

## Map reduce on AccessLog file attached

**Write a Java (could be a console app - will only run once to import the data into MongoDB) program to read the access.log file (attached), and insert into access collection.**

### JAVA API:

```
package mongoapi;

import com.mongodb.client.*;
import com.mongodb.client.MongoDatabase;
import java.io.File;
import java.io.IOException;
import java.util.ArrayList;
import java.util.List;
import java.util.Scanner;
import org.bson.Document;

public class AccessLog {

    public static ArrayList access = new ArrayList<String>();

    public static void main(String[] args) throws IOException {
        System.out.println("before read");
        Scanner sc = new Scanner(new File("C:\\Users\\SayaliGirish\\Desktop\\access.csv"));
        System.out.println("after read");
        while (sc.hasNextLine()) {
            access.add(sc.nextLine());
        }
        //      System.out.println(sc.nextLine());
        //  }

        MongoClient connection = MongoClient.create();
        // connect to database
        MongoDatabase db = connection.getDatabase("AccessLog");
        MongoCollection<Document> collection1 = db.getCollection("accesslog");
        List<Document> documents = new ArrayList<Document>();

        for (int i = 0; i < access.size(); i++) {
            String s = access.get(i).toString();
            String[] arrayOfStrings = s.split("\\\\");
            documents.add(new Document("IPAddress", arrayOfStrings[0])
                .append("Date", arrayOfStrings[1])
                .append("Website", arrayOfStrings[3]));
        }
        collection1.insertMany(documents);
    }
}
```

# Map reduce on inserted data

## 1] Number of times any webpage was visited by the same IP address.

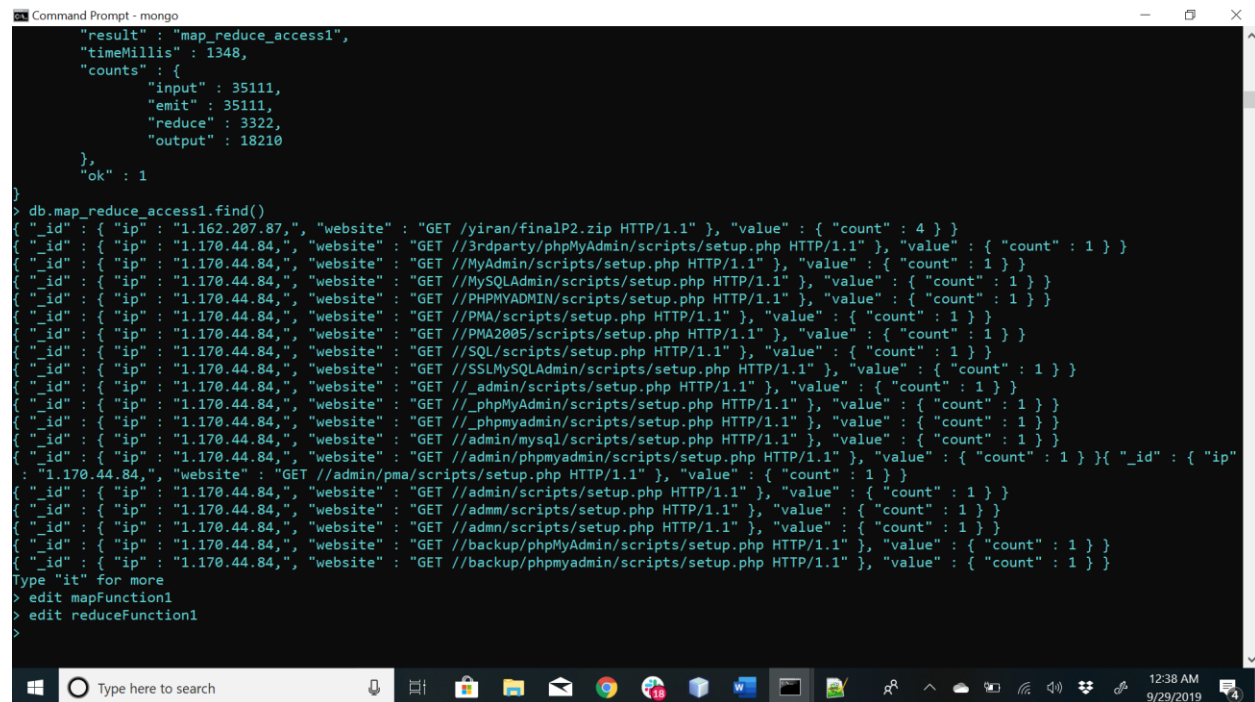
### Map function-

```
function() {  
  var value = {count:1};  
  emit({"ip":this.IPAddress,"website":this.Website}, value);  
}
```

### Reduce function-

```
function(key, values)  
{  
  counterVal = { count:0};  
  for (var id = 0; id < values.length; id++) {  
    counterVal.count += values[id].count ;  
  }  
  return counterVal; }
```

## Output-



```
Command Prompt - mongo  
{"result" : "map_reduce_access1",  
  "timeMillis" : 1348,  
  "counts" : {  
    "input" : 35111,  
    "emit" : 35111,  
    "reduce" : 3322,  
    "output" : 18210  
  },  
  "ok" : 1  
}  
> db.map_reduce_access1.find()  
{ "_id" : { "ip" : "1.162.207.87", "website" : "GET /yiran/finalP2.zip HTTP/1.1" }, "value" : { "count" : 4 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //3rdparty/phpMyAdmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //MyAdmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //MySQLAdmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //PHPMYADMIN/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //PMA/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //PMA2005/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //SQL/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //SSLMySQLAdmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //admin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //phpMyAdmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //phpmyadmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //admin/mysql/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //admin/phpmyadmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //admin/pma/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //admin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //admm/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //admn/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //backup/phpMyAdmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "ip" : "1.170.44.84", "website" : "GET //backup/phpmyadmin/scripts/setup.php HTTP/1.1" }, "value" : { "count" : 1 } }  
Type "it" for more  
> edit mapFunction1  
> edit reduceFunction1  
>
```

## 2] Number of times any webpage was visited each month

### Map Function-

```
function() {  
  var value = {count:1};  
  var month= this.Date.split("/");  
  emit({"month":month[1],"website":this.Website}, value);  
}
```

### Reduce function-

```
function(key, values)  
{  
  counterVal = { count:0};  
  for (var id = 0; id < values.length; id++) {  
    counterVal.count += values[id].count ; }  
  return counterVal; }
```

## Output-

```
Command Prompt - mongo  
{ "_id" : { "month" : "Apr", "website" : "GET http://www.sciencedirect.com/science?_ob=MiamiSearchURL&method=requestForm&btn=Y&_zone=TopNavBar&_origin=home&_acct=C000061991&_version=1&_urlVersion=1&_userid=3987678&md5=70a3f8cc2206cbf27815c684f03124fd HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "month" : "Apr", "website" : "GET http://www.sina.com.cn/ HTTP/1.1" }, "value" : { "count" : 2 } }  
{ "_id" : { "month" : "Apr", "website" : "GET http://www.springerlink.com/ HTTP/1.1" }, "value" : { "count" : 2 } }  
{ "_id" : { "month" : "Apr", "website" : "HEAD / HTTP/1.0" }, "value" : { "count" : 35 } }  
{ "_id" : { "month" : "Apr", "website" : "HEAD / HTTP/1.1" }, "value" : { "count" : 16 } }  
{ "_id" : { "month" : "Apr", "website" : "HEAD /Belarc.html HTTP/1.1" }, "value" : { "count" : 2 } }  
{ "_id" : { "month" : "Apr", "website" : "HEAD /MM62 HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "month" : "Apr", "website" : "HEAD /manager/html HTTP/1.0" }, "value" : { "count" : 5 } }  
{ "_id" : { "month" : "Apr", "website" : "HEAD /manager/status HTTP/1.1" }, "value" : { "count" : 17 } }  
Type "it" for more  
> it  
{ "_id" : { "month" : "Apr", "website" : "POST /ReportServer_SQLEXPRESS/ReportService2005.asmx HTTP/1.1" }, "value" : { "count" : 2 } }  
{ "_id" : { "month" : "Apr", "website" : "POST /ReportServer_SQLEXPRESS/ReportService2006.asmx HTTP/1.1" }, "value" : { "count" : 2 } }  
{ "_id" : { "month" : "Apr", "website" : "POST /ReportServer_SQLEXPRESS/ReportService2010.asmx HTTP/1.1" }, "value" : { "count" : 2 } }  
{ "_id" : { "month" : "Apr", "website" : "POST /ReportServer_SQLEXPRESS/_vti_bin/ReportServer/ReportServiceAuthentication.asmx HTTP/1.1" }, "value" : { "count" : 4 } }  
{ "_id" : { "month" : "Apr", "website" : "POST /phpmyadmin/import.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "month" : "Apr", "website" : "POST /phpmyadmin/index.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "month" : "Apr", "website" : "POST /phpmyadmin/tbl_addfield.php HTTP/1.1" }, "value" : { "count" : 2 } }  
{ "_id" : { "month" : "Apr", "website" : "POST /phpmyadmin4/index.php HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "month" : "Apr", "website" : "PUT /webdav/2105006110761779629 HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "month" : "Apr", "website" : "PUT /webdav/7975640416900352643 HTTP/1.1" }, "value" : { "count" : 1 } }  
{ "_id" : { "month" : "Apr", "website" : "GET / HTTP/1.1" }, "value" : { "count" : 10 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //admin/index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //admin/phpmyadmin/index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //admin/pma/index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //db/index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //dbadmin/index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //myadmin/index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //mysql/index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
{ "_id" : { "month" : "Aug", "website" : "GET //mysqldadmin/index.php HTTP/1.1" }, "value" : { "count" : 3 } }  
Type "it" for more  
> it
```

