

PART - 6

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 code in provided in the file)

Once the data are inserted into MongoDB, do the followings using MapReduce:

- Number of times any webpage was visited by the same IP address.

```
mongosh mongodb://127.0.0.1:27020
mydb> var mapFunction1 = function() {
...   emit(this.webpage, this.ip_address);
... };
mydb> var reduceFunction1 = function(key, values) {
...   return Array.from(new Set(values)).length;
... };
mydb> db.access.mapReduce(
...   mapFunction1,
...   reduceFunction1,
...   { out: "webpage_visits_by_ip" }
... );
DeprecationWarning: Collection.mapReduce() is deprecated. Use an aggregation instead.
See https://docs.mongodb.com/manual/core/map-reduce for details.
{ result: 'webpage_visits_by_ip', ok: 1 }
mydb> db.webpage_visits_by_ip.find()
[
  { _id: '/xampp/phpMyAdmin-2.6.1-rc1/main.php', value: 1 },
  { _id: '/php-my-admin/index.php', value: 21 },
  { _id: '/webdav/plugins/access.ssh/checkInstall.php?destServer=47CK7Cecho%2069212', value: 1 },
  { _id: '/www/lamp/main.php', value: 1 },
  { _id: '/horde-3.0.9/README', value: 3 },
  { _id: '/britishlanguageservices/plugins/access.ssh/checkInstall.php?destServer=47CK7Cecho%2068605', value: 1 },
  { _id: '/phpcollab/www/main.php', value: 1 },
  { _id: '/myadmin/mysql-admin/main.php', value: 1 },
  { _id: '/source/websites/main.php', value: 1 },
  { _id: '/phpMyAdmin-2.8.0-4/scripts/setup.phpmain.php', value: 1 },
  { _id: '/p/webdav/xmltools/minidom/xml/sax/saxutils/os/popen2?cmd=perl%20/tmp/x.txt', value: 1 },
  { _id: '/source/tools/main.php', value: 1 },
  { _id: '/web/phpMyAdmin/setup.php', value: 2 },
  { _id: '/db/phpMyAdmin-2.10.0-rc1/main.php', value: 1 },
  { _id: '/db/p/a/main.php', value: 9 },
  { _id: '/mysql/phpMyAdmin-2.5.5-pl1/main.php', value: 10 },
  { _id: '/phpMyAdmin/', value: 11 },
  { _id: '/administrator/phpMyAdmin-2.9.0/main.php', value: 1 },
  { _id: '/phpMyAdmin-2.4.7-pl3/main.php', value: 1 },
  { _id: '/web/source/main.php', value: 1 }
]
Type "it" for more
mydb>
```

- Number of times any webpage was visited each month. In value:{'9':1} '9' is month & 1 is times visited

```
mongosh mongodb://127.0.0.1:27020
mydb> var mapFunction2 = function() {
...   var dateParts = this.date.split("/");
...   var month = parseInt(dateParts[1]);
...   emit(this.webpage, month);
... };
mydb> var reduceFunction2 = function(key, values) {
...   var counts = {};
...   values.forEach(function(value) {
...     if (counts[value]) {
...       counts[value]++;
...     } else {
...       counts[value] = 1;
...     }
...   });
...   return counts;
... };
mydb> db.access.mapReduce(
...   mapFunction2,
...   reduceFunction2,
...   { out: "webpage_visits_by_month" }
... );
{ result: 'webpage_visits_by_month', ok: 1 }
mydb> db.webpage_visits_by_month.find()
[
  { _id: '/databank/phpMyAdmin-2.8.0.1/main.php', value: { '9': 1 } },
  { _id: '/databank/phpMyAdmin-2.11.2/main.php', value: { '9': 1 } },
  { _id: '/db/funcubes/main.php', value: { '9': 1 } },
  { _id: '/test/sqladmin/main.php', value: { '9': 1 } },
  { _id: '/yusuf/images/ell.jpg', value: { '3': 19 } },
  { _id: '/_sql/mysqlmain/main.php', value: { '9': 2 } },
  { _id: '/mysql/pMA/main.php', value: { '1': 3, '2': 2, '4': 1, '8': 1, '9': 1, '11': 3 } },
  { _id: '/yusuf/bigopen', value: { '3': 8, '4': 1 } },
  { _id: '/support/phpmyadmin-2.8.0.1', value: { '9': 1 } },
  { _id: '/db/phpMyAdmin-2.11.0-rc2/main.php', value: { '9': 1 } },
  { _id: '/admin/phpMyAdmin-2.11.7-rc1-english/main.php', value: { '9': 1 } },
  { _id: '/admin/mysql/scripts/setup.php', value: { '3': 1 } },
  { _id: '/includes/mysql/main.php', value: { '9': 1 } },
  { _id: '/db/phpMyAdmin-2.8.0.1', value: { '5': 1 } },
  { _id: '/support/sql/main.php', value: { '9': 1 } },
  { _id: '/jma/labtools/main.php', value: { '9': 1 } },
  { _id: '/phpMyAdmin-2.2.7-13/main.php', value: { '9': 1 } },
  { _id: '/sqlserver/', value: { '8': 5 } },
  { _id: '/mysql/phpMyAdmin-2.8.0/main.php', value: { '1': 3, '3': 1, '4': 1, '8': 1, '9': 1, '11': 3 } },
  { _id: '/mysql/phpMyAdmin-2.7.0-pl1/main.php', value: { '1': 3, '3': 2, '4': 1, '8': 1, '9': 1, '11': 3 } }
]
Type "it" for more
mydb>
```

PART – 7

Redo Part-6 using Aggregation Pipeline.

- Number of times any webpage was visited by the same IP address.

```
mongosh mongodb://127.0.0.1:27000
mydb> db.access.aggregate([
...   {
...     $group: {
...       _id: { webpage: "$webpage", ipAddress: "$ip_address" },
...       count: { $sum: 1 }
...     }
...   },
...   {
...     $group: {
...       _id: "$_id.webpage",
...       visitsByIP: { $sum: 1 }
...     }
...   },
...   {
...     $sort: {
...       visitsByIP: -1
...     }
...   },
...   {
...     $limit: 5
...   }
... ])
...   {
...     _id: "/favicon.ico"
...     visitsByIP: 409
...   },
...   {
...     _id: "/yiran/finalP2.zip"
...     visitsByIP: 130
...   },
...   {
...     _id: "/mysql/backup.pdf"
...     visitsByIP: 80
...   },
...   {
...     _id: "/phpMyAdmin/scripts/setup.php"
...     visitsByIP: 67
...   },
... ]
mydb>
```

- Number of times any webpage was visited each month.

```
mongosh mongodb://127.0.0.1:27000
mydb> db.access.aggregate([
...   {
...     $project: {
...       webpage: 1,
...       month: { $dateFromStrings: { dateString: "$date", format: "%d/%m/%Y" } }
...     }
...   },
...   {
...     $group: {
...       _id: { webpage: "$webpage", month: "$month" },
...       count: { $sum: 1 }
...     }
...   },
...   {
...     $group: {
...       _id: "$_id.webpage",
...       visitsByMonth: { $push: { month: "$_id.month", count: "$count" } }
...     }
...   },
...   {
...     $project: {
...       _id: 0,
...       webpage: "$_id",
...       visitsByMonth: {
...         $slice: [ "$visitsByMonth", 5 ] // Limit to top 5
...       }
...     }
...   },
...   {
...     $limit: 5 // Limit the output to 5 documents
...   }
... ])
...   {
...     webpage: "/sql/phpMyAdmin-2.10.1-rc1/main.php"
...     month: 9
...     visitsByMonth: [ { month: 9, count: 1 } ]
...   },
...   {
...     webpage: "/db/phpMyAdmin-2.11.0/main.php"
...     month: 9
...     visitsByMonth: [ { month: 9, count: 1 } ]
...   },
...   {
...     webpage: "/support/mysql/main.php"
...     month: 9
...     visitsByMonth: [ { month: 9, count: 1 } ]
...   },
...   {
...     webpage: "/mysql/phpmyadmin3/main.php"
...     month: 9
...     visitsByMonth: [ { month: 9, count: 1 } ]
...   },
...   {
...     webpage: "/phpMyAdmin-3.1.3.0-english/scripts/setup.php"
...     month: 11
...     visitsByMonth: [ { month: 11, count: 2 } ]
...   },
... ]
mydb>
```