

Big Data 1

by Gabor Jonas

FILE	994_GABOR_JONAS_BIG_DATA_1_226876_872843477.DOCX (93.29K)		
TIME SUBMITTED	27-MAR-2020 11:30PM (UTC+0000)	WORD COUNT	691
SUBMISSION ID	123000397	CHARACTER COUNT	4254



Big Data Assignment 1

Module: Big Data

Module code: ACCA6012

Student name: Gabor Jonas

Student ID: 1705136

Submission date: 27/03/2020

Table of Contents

1. Big Data.....	3
2. CAP Theorem.....	3
3. SQL vs. NoSQL	Error! Bookmark not defined.
4. Database recommendation to Zang Technologies	Error! Bookmark not defined.
5. MongoDB.....	Error! Bookmark not defined.
6. References	5

1. Big Data

Big data refers to a set of data files of such a large size and complexity that it is difficult to process with relational database management systems.

Characteristics of such data [1]:

Volume

The size of the data that has to be processed.

Velocity

The speed with which the data is created.

Variety

Data comes from variety of sources, mainly structured, semi-structured and unstructured

Veracity

It refers the quality of data that can be analysed

Value

Data has to be turned into value to be useful.

Big Data Databases are organised into four categories:

- Document oriented
- Column oriented
- Key-value
- Graph

2. CAP Theorem

The CAP theorem(as known as Brewer's theorem) states that in a distributed system, only two of the following three properties are available at a time [2]:

Consistency

A client read operation is guaranteed to return the record created by the last write.

Availability

All client requests are answered, but there is no guarantee that this will be the result of the latest writing.

Partitioning

The system remains functional even if the distributed system is divided into parts by the loss of certain nodes.

3. ACID vs BASE

In Relational databases ACID is required to protect the database from corruption.

ACID is an acronym for [3]:

Atomicity

Atomicity requires that any operation fails in a transaction, then the entire transaction fails.

Consistency

The transactions must always meet the rules defined by the system. All applicable rules (constraints, referential integrity keys, triggers, etc.) must be enforced.

Isolation

Transactions don't have access to other transactions that are in unfinished state.

The effects of an ongoing transaction are not visible from the other transactions.

Durability

Changes to executed transactions must be stored on a non-volatile data storage to retain software or hardware failure, power failure, or other malfunctions.

NoSQL relies on a model known BASE which is the acronym of:

Basically available

Soft state

Eventual consistency

4. References

- [1] I. Kalbandi and J. Anuradha, "A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology," *Procedia Computer Science*, vol. 48, pp. 319-324, December 2015.
- [2] F. D. Muñoz-Esco, J.-M. Rubén de, J.-R. García-Escrivá, M. J R González and B.-A. José, "CAP Theorem: Revision of Its Related Consistency Models," *The Computer Journal*, vol. 62, pp. 943-960, 2019.
- [3] A. G. Taylor, *SQL for dummies*, 7th ed., Indianapolis: Wiley Publishing, 2010, p. 307.

5. Appendix

```
<?php
require 'vendor/autoload.php';

//connection to mongodb server
$conn = new MongoClient('mongodb://localhost:27017');

//select database and collection if doesn't exists create one
$db = $conn->driver_tracker->location_log;

//loading 1 json object
$json = json_decode(file_get_contents('data_single.json'));
//inserting single item
$query = $db->insertOne($json);
//displaying the number of items inserted
echo 'Number of items inserted:'. $query->getInsertedCount() . '<br>';

//loading 50 json objects
$json = json_decode(file_get_contents('data_many.json'));
//Inserting 50 items
$query = $db->insertMany($json);
//displaying the number of items inserted
echo 'Number of items inserted:'. $query->getInsertedCount() . '<br>';

//finding driver using driver_id
$query = $db->findOne(["driver_id" => "23c28221-2020-44f5-b75e-"]
```

```

4c13c20f3b7f"]]);

echo 'Name:' . $query->driver_name . '<br>';
echo 'Last seen at:' . $query->last_seen_at . '<br>';
echo 'Latitude:' . $query->latitude . '<br>';
echo 'Longitude:' . $query->longitude . '<br>';

//find the drivers who were in London
$query = $collection->find(["last_location_city" => "London"]);

echo 'Drivers who were in London: <br>';
foreach ($query as $data) {
    echo 'Name: ' . $data['driver_name'] . '<br>';
    echo 'Driver id ' . $data['driver_id'] . '<br>';
    echo 'Car make: ' . $data['car_make'] . ' ';
    echo 'Car model: ' . $data['car_model'] . '<br>';
}

//updating car VIN where driver_id is 4588ef83-f70b-4ee1-87cf-
6744716a9cd2
$query = $collection->updateOne(['driver_id' => '4588ef83-f70b-4ee1-
87cf-6744716a9cd2'], ['$set' => ['car_vin' => '2HNYD184X2H991690']]);

echo 'Found ' . $query->getMatchedCount() . ' document <br>';
echo 'Updated ' . $query->getModifiedCount() . ' document <br>';

//delete the first document where the driver visited London
$query = $collection->deleteOne(['last_location_city' => 'London']);

echo 'Deleted ' . $query->getDeletedCount() . ' document <br>';

//delete all document where the driver visited London
$query = $collection->deleteMany(['last_location_city' => 'London']);

echo 'Deleted ' . $query->getDeletedCount() . ' document <br>';

```

Big Data 1

GRADEMARK REPORT

FINAL GRADE

30/100

GENERAL COMMENTS

Instructor

a (20): 6

b (20): 6

c (20): 4

di (20): 0

dii (20): 14

Disucssion:

Big data discussion is very terse, much more analysis was required.

Likewise, CAP Theorem.

I wanted an example that demonstrated the CAP theorem.

Mongodb using PHP was interesting given we used C# in lectures. However, the code was fairly short. This is a very poor attempt at this assignment.

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6