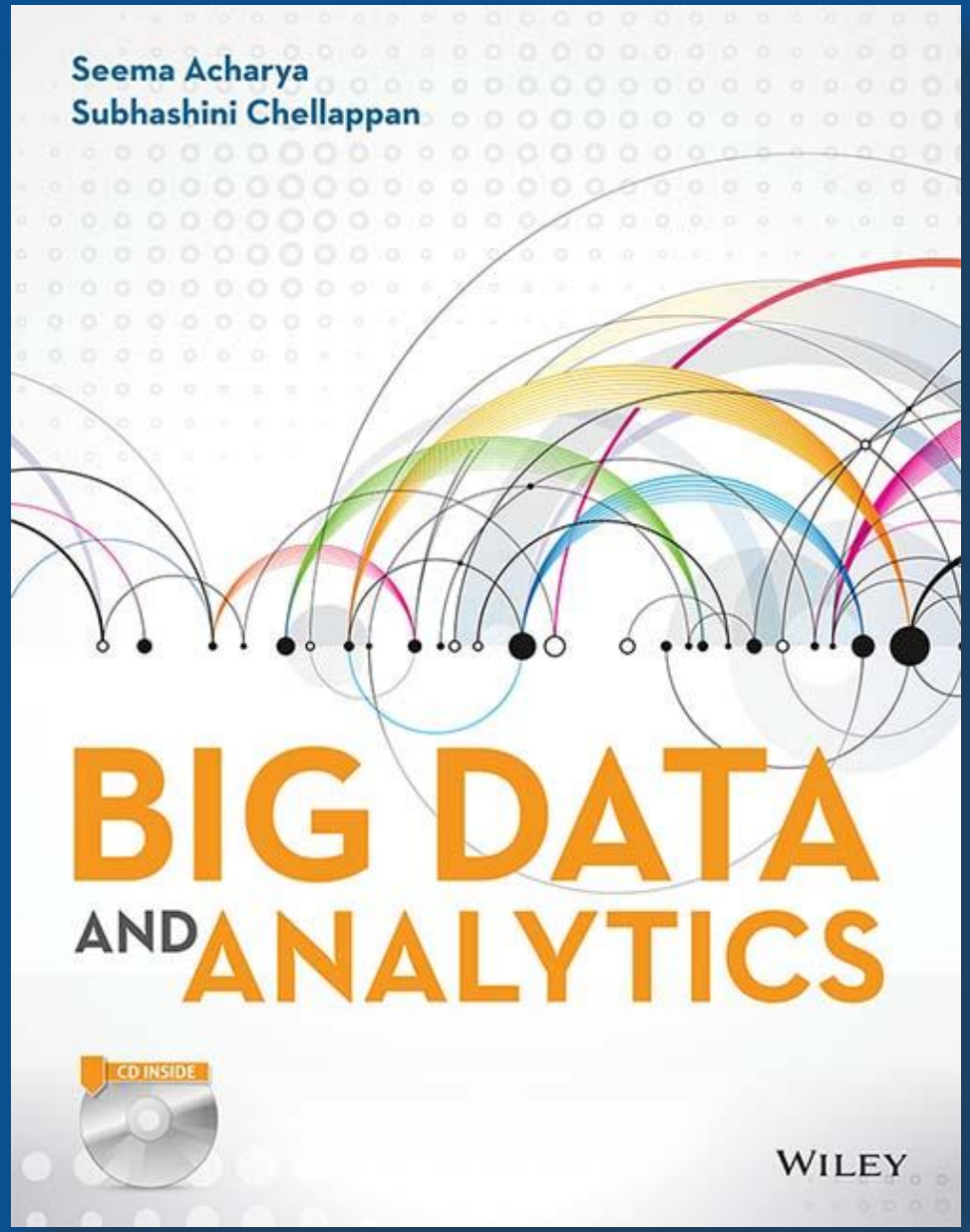


Big Data And Analytics

Seema Acharya
Subhashini Chellappan

Chapter 6

Introduction to MongoDB

Learning Objectives and Learning Outcomes

Learning Objectives	Learning Outcomes
Introduction to MongoDB 1. To study the features of MongoDB. 2. To learn how to perform CRUD operations. 3. To study aggregation. 4. To study the MapReduce Framework. 5. To import from and export to CSV format.	a) To comprehend the reasons behind the popularity of NoSQL database. b) To be able to perform CRUD operations. c) To comprehend MapReduce framework. d) To understand the aggregation. e) To be able to successfully import from and export to CSV.

Session Plan

Lecture time	90 to 120 minutes
Q/A	15 minutes

Agenda

- What is MongoDB?
- Why MongoDB?
  - Using JSON
  - Creating or Generating a Unique Key
  - Support for Dynamic Queries
  - Storing Binary Data
  - Replication
  - Sharding
  - Terms used in RDBMS and MongoDB
- Data Types in MongoDB
- CRUD (Insert(), Update(), Save(), Remove(), Find())
  - MapReduce Functions
  - Aggregation
  - Java Scripting
  - MongoImport
  - MongoExport

MongoDB- An Introduction

What is MongoDB?

MongoDB is:

1. Cross-platform.
2. Open source.
3. Non-relational.
4. Distributed.
5. NoSQL.
6. Document-oriented data store.

Why MongoDB?

Why MongoDB?

- Open Source
- Distributed
- Fast In-Place Updates
- Replication
- Full Index Support
- Rich Query Language
- Easy Scalability
- Auto sharding

JSON

JSON (Java Script Object Notation)

Sample JSON Document

{
FirstName: John,
LastName: Mathews,
ContactNo: [+123 4567 8900, +123 4444 5555]
}

Unique Identifier

Each JSON document should have a unique identifier. It is the _id key.

0	1	2	3	4	5	6	7	8	9	10	11
Timestamp				Machine ID			Process ID		Counter		

Support for Dynamic Queries

MongoDB has extensive support for dynamic queries.

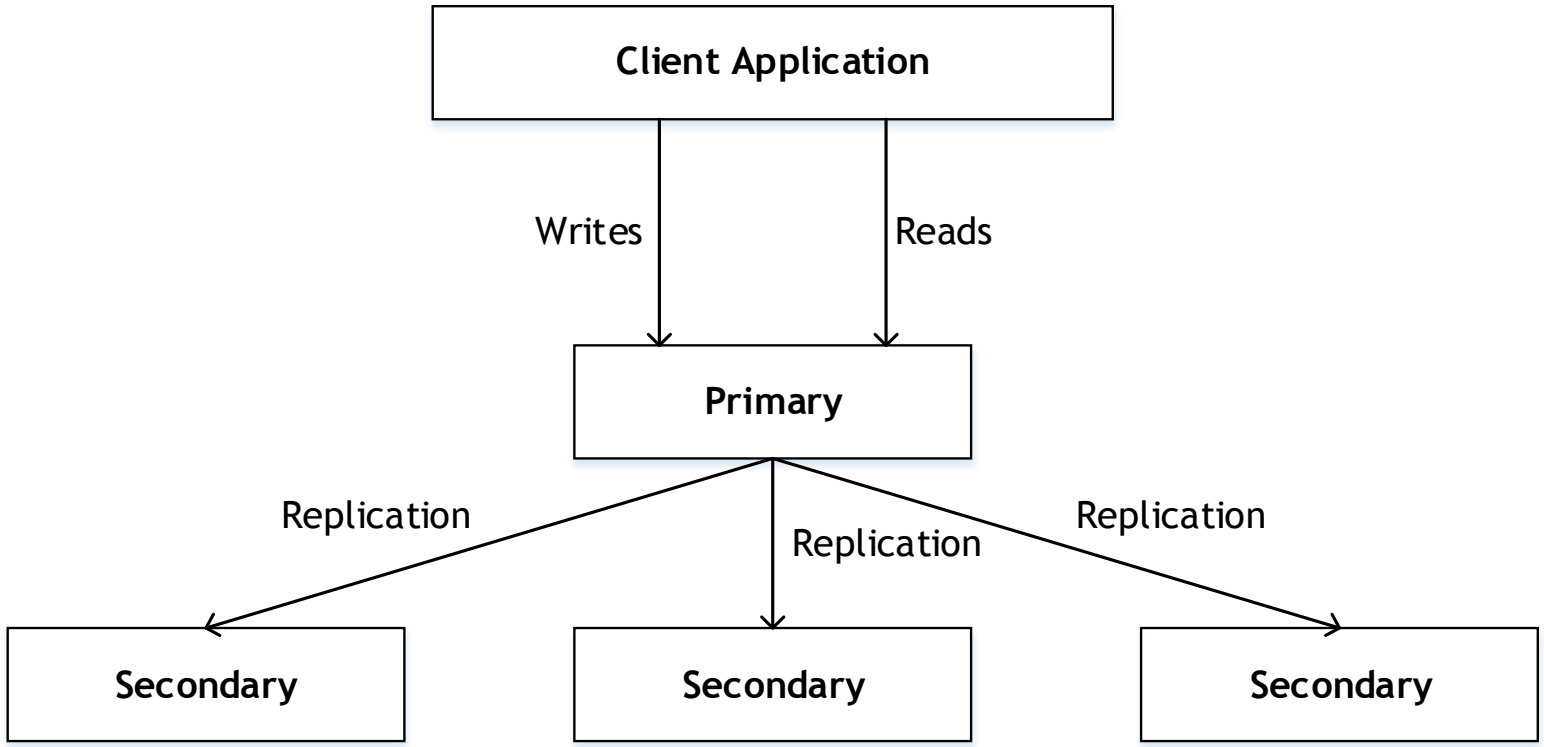
This is in keeping with traditional RDBMS wherein we have static data and dynamic queries.

Storing Binary Data

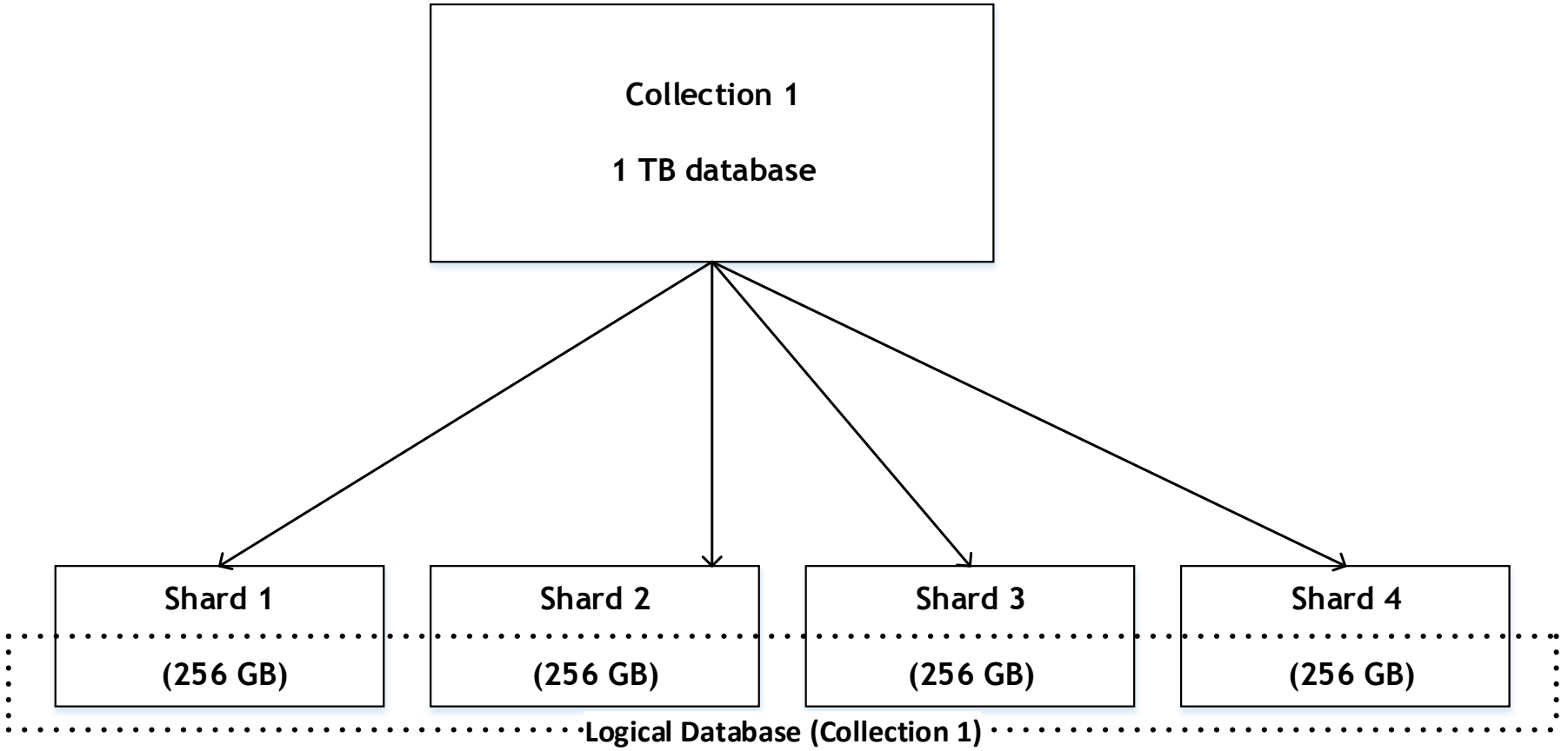
MongoDB provides GridFS to support the storage of binary data.

It can store up to 4 MB of data.

Replication in MongoDB



Sharding in MongoDB



Terms Used in RDBMS and MongoDB

Terms Used in RDBMS and MongoDB

RDBMS	MongoDB
Database	Database
Table	Collection
Record	Document
Columns	Fields / Key Value pairs
Index	Index
Joins	Embedded documents
Primary Key	Primary key (_id is a identifier)

Data Types in MongoDB

Data Types in MongoDB

String	Must be UTF-8 valid. Most commonly used data type.
Integer	Can be 32-bit or 64-bit (depends on the server).
Boolean	To store a true/false value.
Double	To store floating point (real values).
Min/Max keys	To compare a value against the lowest or highest BSON elements.
Arrays	To store arrays or list or multiple values into one key.
Timestamp	To record when a document has been modified or added.
Null	To store a NULL value. A NULL is a missing or unknown value.
Date	To store the current date or time in Unix time format. One can create object of date and pass day, month and year to it.
Object ID	To store the document's id.
Binary data	To store binary data (images, binaries, etc.).
Code	To store javascript code into the document.
Regular expression	To store regular expression.

CRUD in MongoDB

Collections

To create a collection by the name “Person”. Let us take a look at the collection list prior to the creation of the new collection “Person”.

```
db.createCollection(“Person”);
```

Collections

To drop a collection by the name “food”.

```
db.food.drop();
```

Insert Method

Create a collection by the name “Students” and store the following data in it.

db.Students.insert({_id:1, StudName:"Michelle Jacintha", Grade: "VII", Hobbies: "Internet Surfing"});

Update Method

Insert the document for “Aryan David” into the Students collection only if it does not already exist in the collection. However, if it is already present in the collection, then update the document with new values. (Update his Hobbies from “Skating” to “Chess”.) Use “Update else insert” (if there is an existing document, it will attempt to update it, if there is no existing document then it will insert it).

```
db.Students.update({_id:3, StudName:"Aryan David", Grade: "VII"},{$set:{Hobbies: "Skating"}},{upsert:true});
```

Find Method

To search for documents from the “Students” collection based on certain search criteria.

db.Students.find({StudName:"Aryan David"});

Find Method

To display only the StudName and Grade from all the documents of the Students collection. The identifier _id should be suppressed and NOT displayed.

db.Students.find({},{StudName:1,Grade:1,_id:0});

Find Method

To find those documents where the Grade is set to ‘VII’

db.Students.find({Grade:{$eq:'VII'}}).pretty();

Find Method

To find those documents from the Students collection where the Hobbies is set to either ‘Chess’ or is set to ‘Skating’.

db.Students.find ({Hobbies :{ $in: ['Chess','Skating']}}).pretty ();

Find Method

To find documents from the Students collection where the StudName begins with “M”.

db.Students.find({StudName:/^M/}).pretty();

Find Method

To find documents from the Students collection where the StudName has an “e” in any position.

db.Students.find({StudName:/e/}).pretty();

Find Method

To find the number of documents in the Students collection.

db.Students.count();

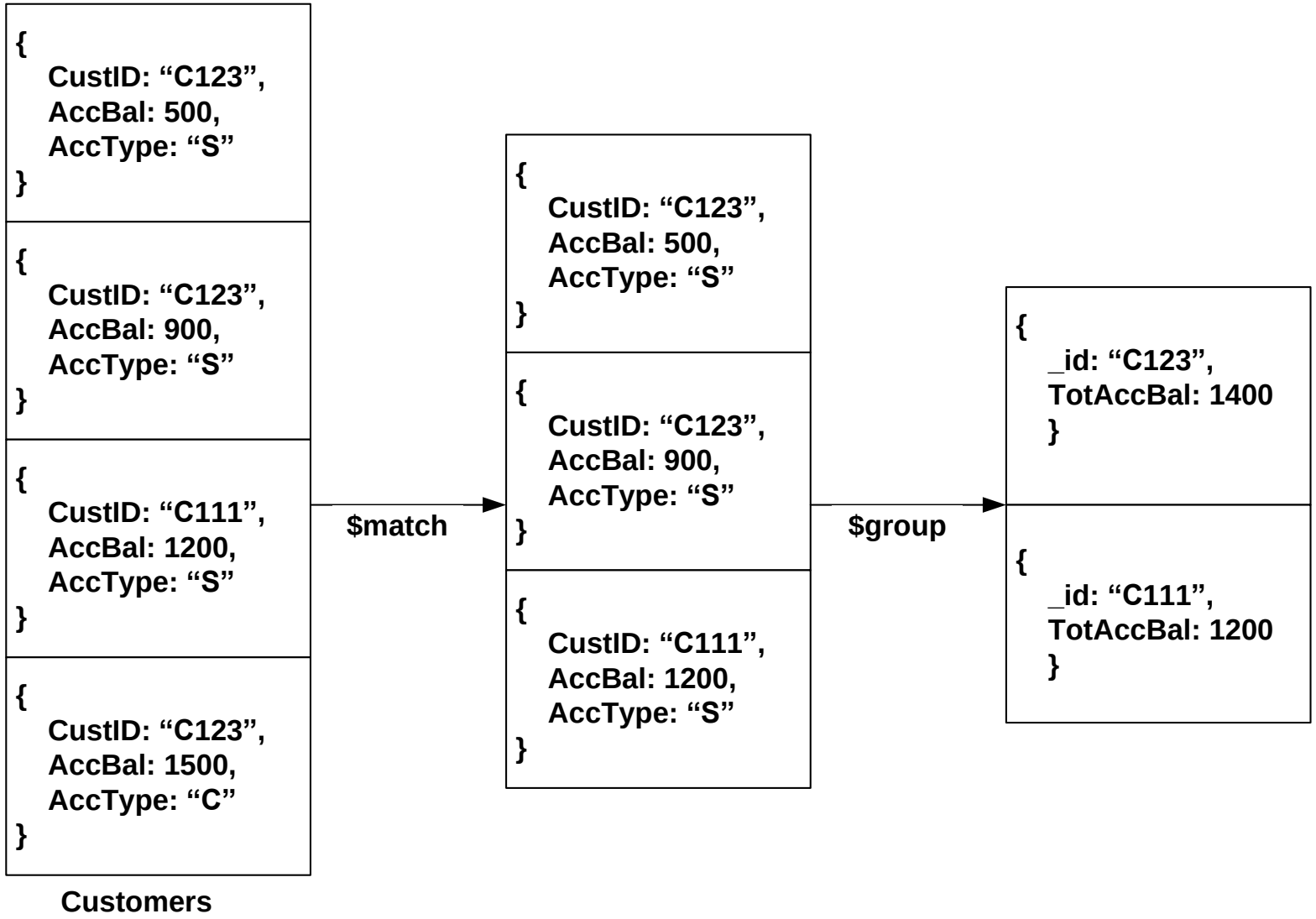
Find Method

To sort the documents from the Students collection in the descending order of StudName.

db.Students.find().sort({StudName:-1}).pretty();

Aggregate Function

Aggregate Function



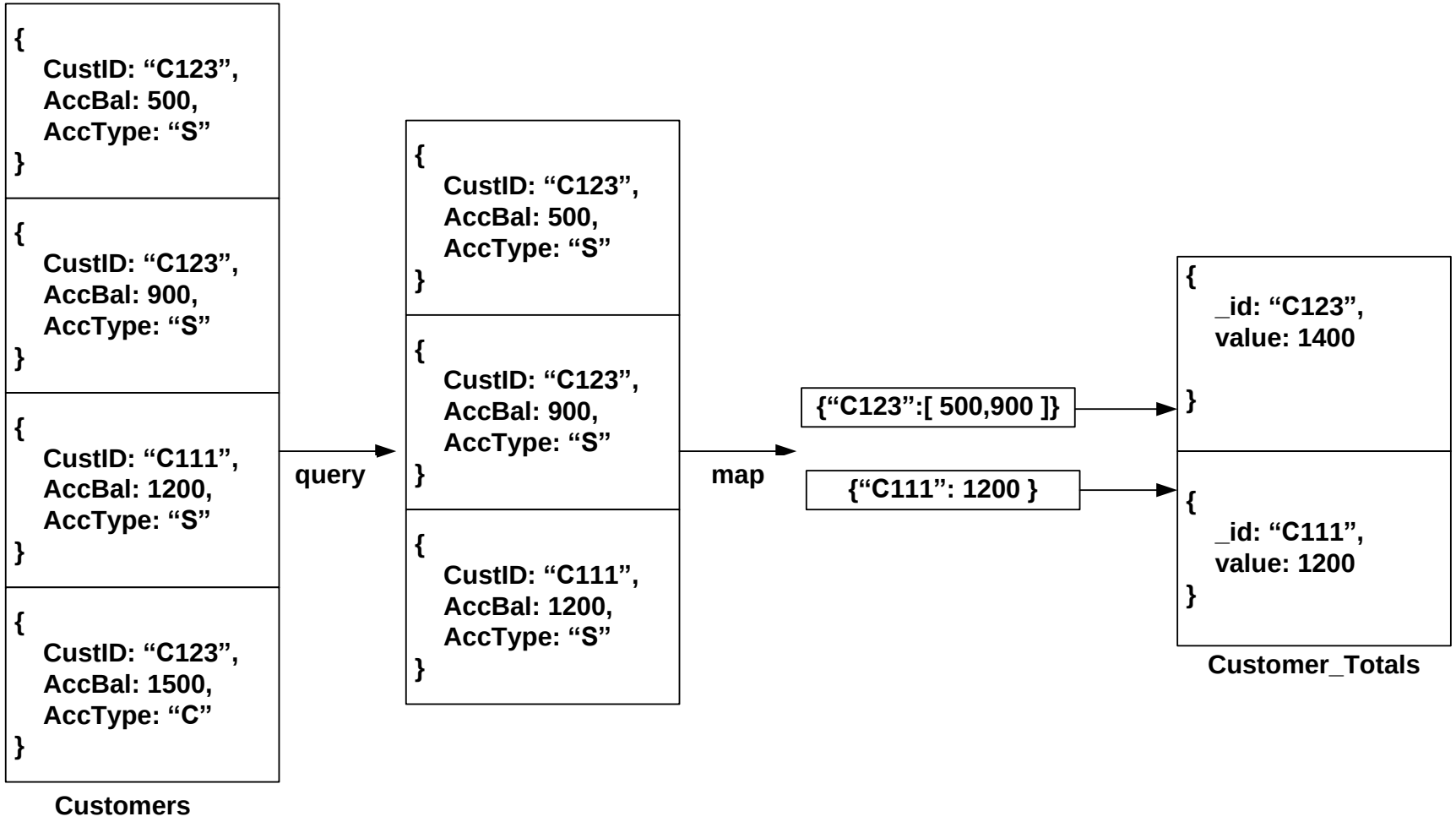
Aggregate Function

First filter on “AccType:S” and then group it on “CustID” and then compute the sum of “AccBal” and then filter those documents wherein the “TotAccBal” is greater than 1200, use the below syntax:

```
db.Customers.aggregate( { $match : {AccType : "S" } },
{ $group : { _id : "$CustID",TotAccBal : { $sum : "$AccBal" } } },
{ $match : {TotAccBal : { $gt : 1200 } }});
```

MapReduce Framework

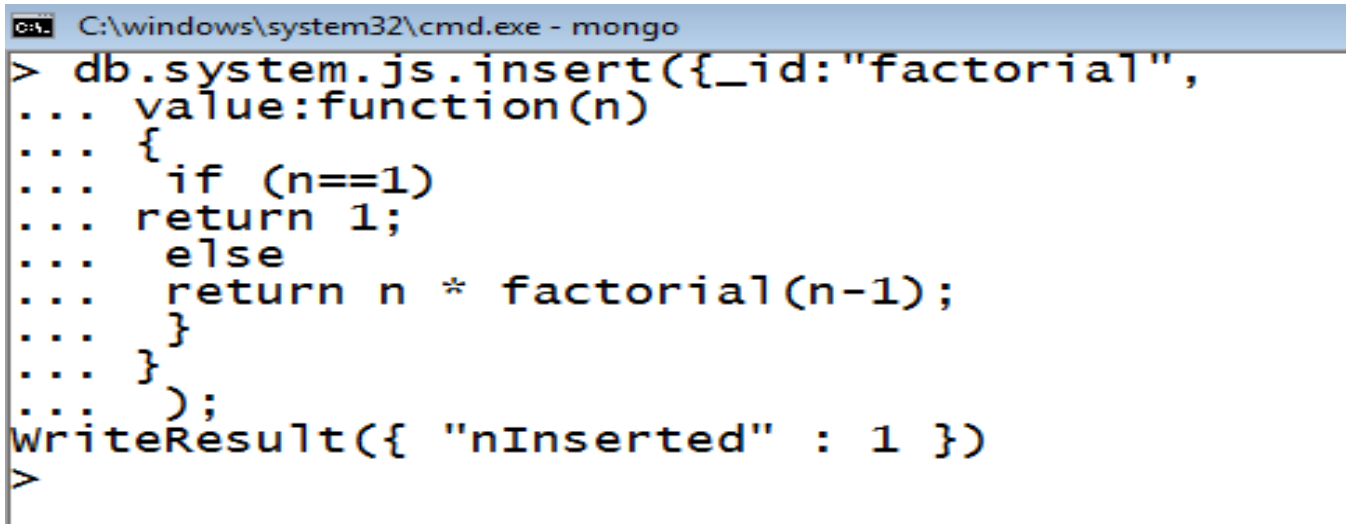
MapReduce Framework

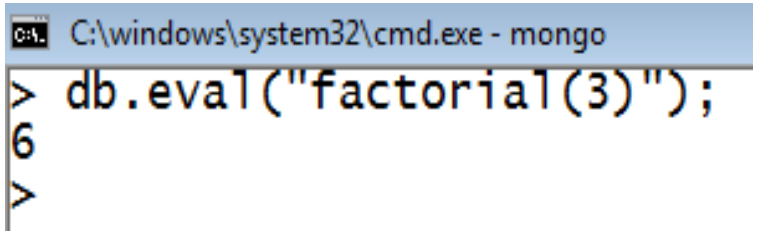


Java Script Programming

Java Script Programming

To compute the factorial of a given positive number. The user is required to create a function by the name “factorial” and insert it into the “system.js” collection.





MongoImport

Import data from a CSV file

Given a CSV file “sample.txt” in the D: drive, import the file into the MongoDB collection, “SampleJSON”. The collection is in the database “test”.

Mongoimport --db test --collection SampleJSON --type csv --headerline --file d:\sample.txt

MongoExport

Export data to a CSV file

This command used at the command prompt exports MongoDB JSON documents from “Customers” collection in the “test” database into a CSV file “Output.txt” in the D: drive.

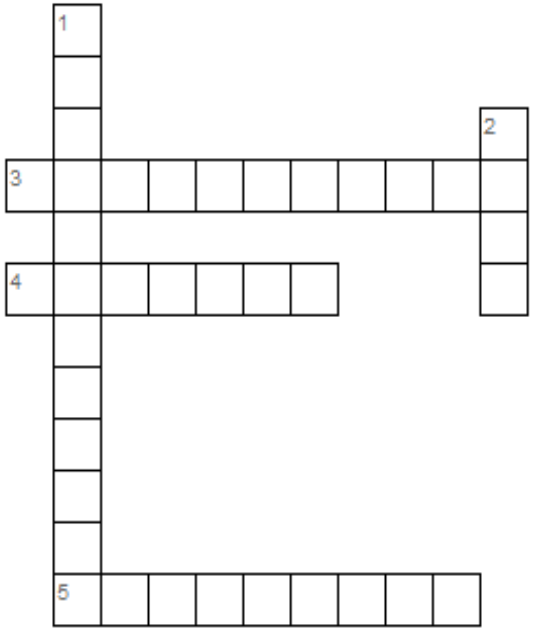
Mongoexport --db test --collection Customers --csv --fieldFile d:\fields.txt --out d:\output.txt

Answer a few quick questions …

Crossword

MongoDB

Basic puzzle on MongoDB



Across

3 MongoDB database stores its data in .....

4 MongoDB uses ---------- schemas.

5 A collection holds one or more -- ----------.

Down

1 MongoDB uses ---------- files.

2 MongoDB uses ---------, a binary object format similar to, but more expensive than JSON.

Answer Me

- What is MongoDB?
- Comment on Auto-sharding in MongoDB.
- What are collections and documents?
- What is JSON?
- Explain your understanding of Update In-Place.

Summary please…

Ask a few participants of the learning program to summarize the lecture.

References …

Further Readings

http://www.mongodb.org/

https://university.mongodb.com/

http://www.tutorialspoint.com/mongodb/

Thank you