Big Data Analytics

**Lab Practical and date** – Practical 9, Monday 26th October 2020

**Name and Roll Number**- Het Shah, 17BIT103

**Practical Objective**-Extend MongoDB functionality for MapReduce on document collection

**Steps Involved-**

We coded the mapper and reducer functions in javascript and applied the same on the Restaurant dataset using mongoDB

**Background**

**MongoDB**

**MongoDB** is a [cross-platform](https://en.wikipedia.org/wiki/Cross-platform) [document-oriented database](https://en.wikipedia.org/wiki/Document-oriented_database) program. Classified as a [NoSQL](https://en.wikipedia.org/wiki/NoSQL) database program, MongoDB uses [JSON](https://en.wikipedia.org/wiki/JSON)-like documents with optional [schemas](https://en.wikipedia.org/wiki/Database_schema). MongoDB is developed by [MongoDB Inc.](https://en.wikipedia.org/wiki/MongoDB_Inc.) and licensed under the Server Side Public License (SSPL).

**Restaurant dataset**

This dataset contains a list of restaurants within Baltimore City.

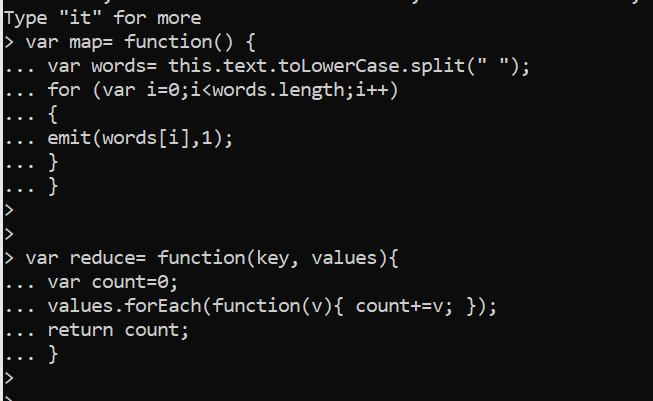
**MapReduce**

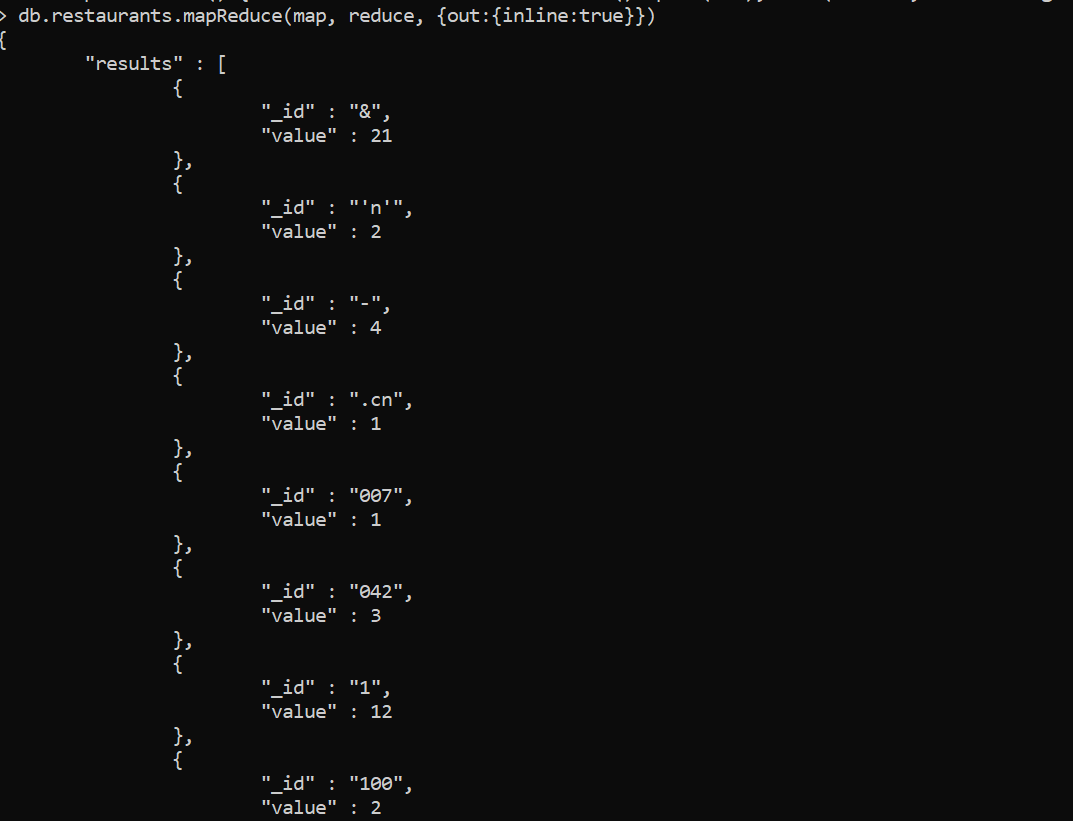
**MapReduce** is a processing technique and a program model for distributed computing based on java. The **MapReduce** algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs).

**Steps**

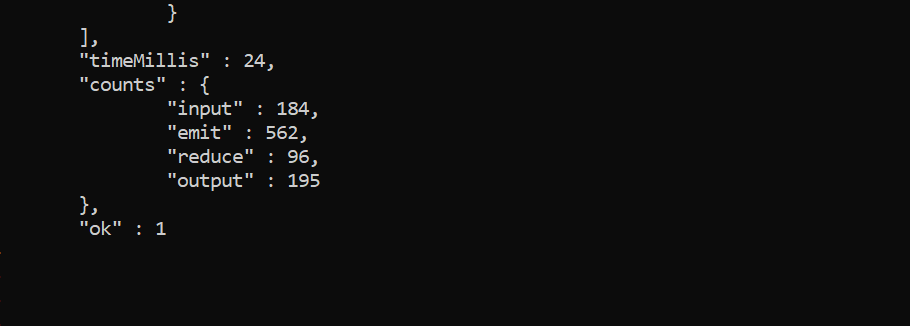
In this Practical I have coded **3 mappers and 3 reducers to perform 3 different tasks** on the dataset

1. **WordCount in the Names Column of the dataset**



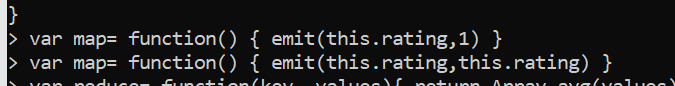
Mapper and Reducer code in Javascript for the wordcount problem

MapReduce Output

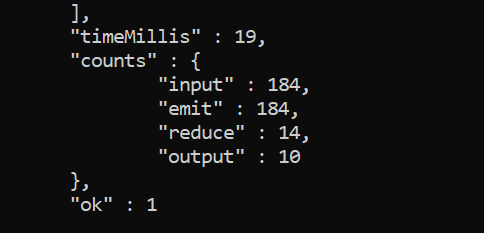


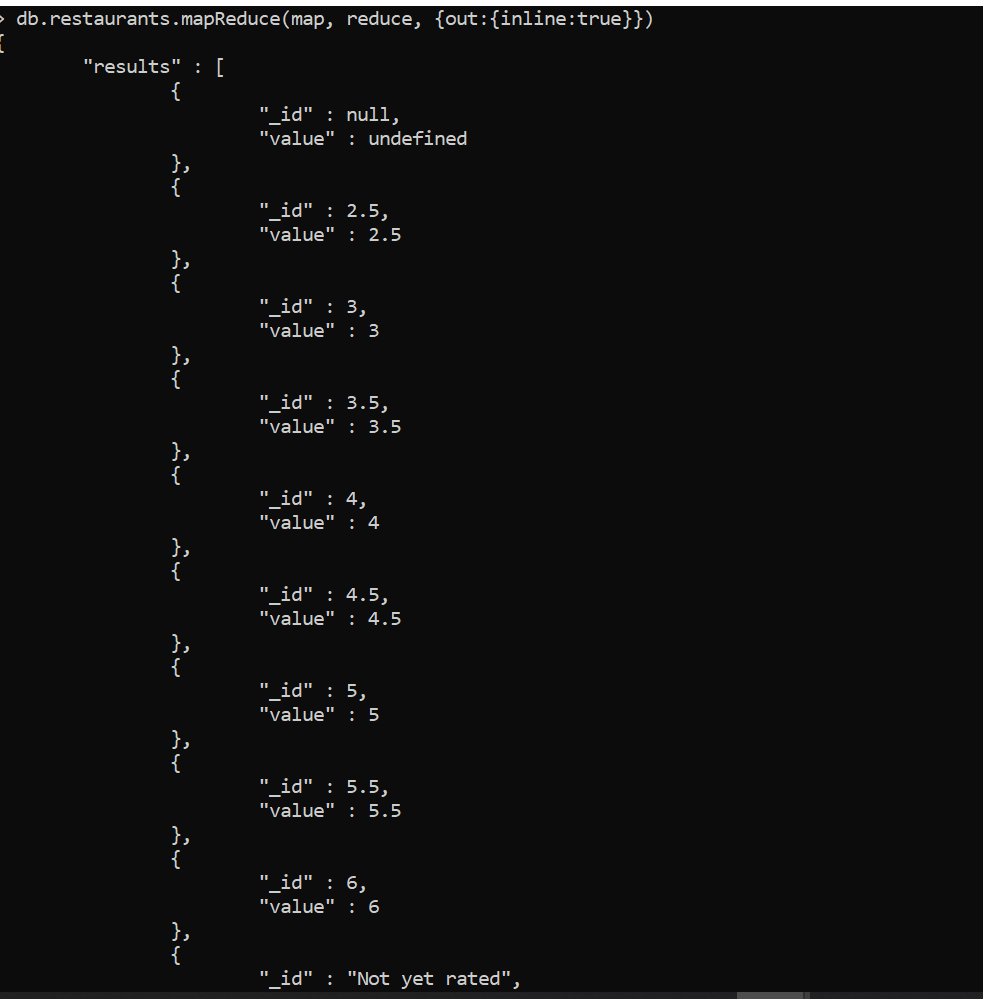
The job statistics, time taken and how many times mapper emitted and how many time the reducer was called

1. **Counting the Number of Times Rating Comes**



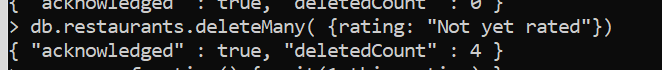
Mapper and Reducer for the class

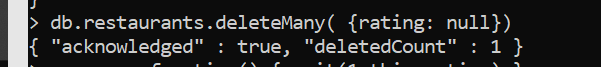




As we can see we have non-numerical values as well such as null and “not yet rated”

We will delete them

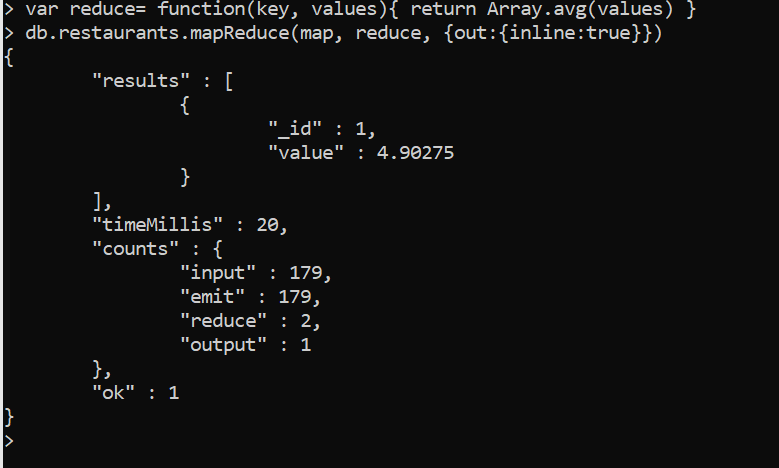




1. **Finding the average rating from the dataset**



For the reducer, we used the inbuild avg function of javascript



As we can see, the average rating is 4.90275

**Conclusion**

In this Practical we learned about the MapReduce Programming paradym in MongoDB and ran a few jobs on the restaurant collection