PUNE INSTITUTE OF COMPUTER TECHNOLOGY

Subject: ADBMS (LP LAB)

Name: Aditya Kangune **Roll No. : 33323**

Batch: K11 **Academic Year: 2021-22**

Assignment 2
Map reduce and aggregate

Aim:

Implement Map reduces operation with suitable example on above MongoDB database and implement the following:

- Aggregation framework
- Create and drop different types of indexes and explain () to show the advantage of the indexes.

Objective:

- To understand the concept of Map-reduce in mongodb.
- To understand the concept of Aggregation in mongodb.
- To implement the concept of document-oriented databases.

Theory:

Map-reduce:

1. MapReduce is a programming model and an associated implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster.
2. A MapReduce program is composed of a map procedure, which performs filtering and sorting, and a reduce method, which performs a summary operation.
3. To perform map-reduce operations, MongoDB provides the mapReduce database command.
4. MapReduce functions are written in JavaScript.
5. Uses a "pipeline" approach where objects are transformed as they pass through a series of pipeline operators such as matching, projecting, sorting, and grouping. Pipeline operators need not produce one output document for every input document: operators may also generate new documents or filter out documents.

Map/Reduce involves two steps:

- First, map the data from the collection specified;
- Second, reduce the results.

Map Function:

- var mapFunction1 = function()
- { emit(this.cust_id, this.amount);};

Reduce Function:

- var reduceFunction1 = function(key, values)
- {return Array.sum(values); };

Aggregation:

Aggregation operations process multiple documents and return computed results. We can use aggregation operations to:

- Group values from multiple documents together.
- Perform operations on the grouped data to return a single result.
- Analyse data changes over time.

Indexing:

Indexes are special data structures that store a small portion of the collection's data set in an easy to traverse form.

The index stores the value of a specific field or set of fields, ordered by the value of the field.

The ordering of the index entries supports efficient equality matches and range-based query operations.

In addition, MongoDB can return sorted results by using the ordering in the index.

Output:

Current DB

```
db.Employee.find().pretty();
{
        "_id" : ObjectId("611f52deaf8bf9764e431af4"),
        "empName" : "Suresh",
        "uan" : 54356212,
        "dept" : {
                "location" : 412,
                "name" : "IT"
        },
        "designation" : "Teacher",
        "emailID" : "abc@gmail.com",
        "salary" : 20000
}
{
        "_id" : ObjectId("611f52f7af8bf9764e431af5"),
        "empName" : "Ramesh",
        "uan" : 34554222,
        "dept" : {
                "location" : 522,
```

```
            "name" : "IT"
        },
        "designation" : "Teacher",
        "emailID" : "cde@gmail.com",
        "salary" : 60000
}
{
        "_id" : ObjectId("611f59cfaf8bf9764e431af6"),
        "empName" : "Mangesh",
        "uan" : 362926328,
        "dept" : {
            "location" : 301,
            "name" : "IT"
        },
        "designation" : "HOD",
        "emailID" : "abcde@gmail.com",
        "salary" : 10000
}
{
        "_id" : ObjectId("611f657c76ba0a7a09d22ef5"),
        "empName" : "Rangesh",
        "uan" : 362926328,
```

```
        "dept" : {
                "location" : 513,
                "name" : "IT"
        },
        "designation" : "HOD",
        "emailID" : "abcdef@gmail.com",
        "salary" : 50000
}
{
        "_id" : ObjectId("611f659976ba0a7a09d22ef6"),
        "empName" : "Tungesh",
        "uan" : 362900328,
        "dept" : {
                "location" : 618,
                "name" : "Comp"
        },
        "designation" : "Teacher",
        "emailID" : "abcdefg@gmail.com",
        "salary" : 60000
}
{
        "_id" : ObjectId("611f65b676ba0a7a09d22ef7"),
```

```
        "empName" : "Sangesh",
        "uan" : 452900328,
        "dept" : {
                "location" : 102,
                "name" : "Comp"
        },
        "designation" : "Teacher",
        "emailID" : "abcdefgh@gmail.com",
        "salary" : 70000
}
{
        "_id" : ObjectId("611f65d176ba0a7a09d22ef8"),
        "empName" : "Vangesh",
        "uan" : 452900833,
        "dept" : {
                "location" : 213,
                "name" : "ENTC"
        },
        "designation" : "Teacher",
        "emailID" : "abcdefghi@gmail.com",
        "salary" : 80000
}
```

```
{
        "_id" : ObjectId("611f65e576ba0a7a09d22ef9"),
        "empName" : "Zangesh",
        "uan" : 722900833,
        "dept" : {
                "location" : 503,
                "name" : "ENTC"
        },
        "designation" : "Teacher",
        "emailID" : "abcdefghij@gmail.com",
        "salary" : 90000
}
{
        "_id" : ObjectId("611f65f876ba0a7a09d22efa"),
        "empName" : "Dangesh",
        "uan" : 7229002133,
        "dept" : {
                "location" : 321,
                "name" : "ENTC"
        },
        "designation" : "Teacher",
        "emailID" : "xyz@gmail.com",
```

```
        "salary" : 100000
}
{
        "_id" : ObjectId("611f662076ba0a7a09d22efb"),
        "empName" : "Mohan",
        "salary" : 140000,
        "uan" : 123441121,
        "dept" : {
                "location" : 542,
                "name" : "comp"
        },
        "designation" : "Teacher",
        "emailID" : "xyz@gmail.com",
        "teams" : [
                "technical"
        ]
}
{
        "_id" : ObjectId("61260b295e64af69224b448a"),
        "empName" : "Rohan",
        "salary" : 120000,
        "uan" : 123441121,
```

```
        "dept" : {
                "location" : 542,
                "name" : "comp"
        },
        "designation" : "Teacher",
        "emailID" : "abcd@gmail.com"
}
```


Aggregate

```
db.Employee.aggregate([{$group:{_id:"$dept"}}]);
{ "_id" : { "location" : 321, "name" : "ENTC" } }
{ "_id" : { "location" : 503, "name" : "ENTC" } }
{ "_id" : { "location" : 213, "name" : "ENTC" } }
{ "_id" : { "location" : 102, "name" : "Comp" } }
{ "_id" : { "location" : 618, "name" : "Comp" } }
{ "_id" : { "location" : 542, "name" : "comp" } }
{ "_id" : { "location" : 513, "name" : "IT" } }
{ "_id" : { "location" : 301, "name" : "IT" } }
{ "_id" : { "location" : 522, "name" : "IT" } }
{ "_id" : { "location" : 412, "name" : "IT" } }
```

```
> db.Employee.aggregate([{$group:{_id:"$dept.name"}}]);
{ "_id" : "comp" }
{ "_id" : "ENTC" }
{ "_id" : "Comp" }
{ "_id" : "IT" }

> db.Employee.aggregate([{$group:{_id:"$dept",totalSalary:{$sum:"$salary"}}}]);
{ "_id" : { "location" : 321, "name" : "ENTC" }, "totalSalary" : 100000 }
{ "_id" : { "location" : 503, "name" : "ENTC" }, "totalSalary" : 90000 }
{ "_id" : { "location" : 213, "name" : "ENTC" }, "totalSalary" : 80000 }
{ "_id" : { "location" : 102, "name" : "Comp" }, "totalSalary" : 70000 }
{ "_id" : { "location" : 618, "name" : "Comp" }, "totalSalary" : 60000 }
{ "_id" : { "location" : 542, "name" : "comp" }, "totalSalary" : 260000 }
{ "_id" : { "location" : 513, "name" : "IT" }, "totalSalary" : 50000 }
{ "_id" : { "location" : 301, "name" : "IT" }, "totalSalary" : 10000 }
{ "_id" : { "location" : 522, "name" : "IT" }, "totalSalary" : 60000 }
{ "_id" : { "location" : 412, "name" : "IT" }, "totalSalary" : 20000 }
```

>
db.Employee.aggregate([{$group:{_id:"$dept.name",totalSalary:{$sum:"$salary
"}}},{$match:{totalSalary:{$gte:150000}}}]);

{ "_id" : "comp", "totalSalary" : 260000 }

{ "_id" : "ENTC", "totalSalary" : 270000 }

>
db.Employee.aggregate([{$group:{_id:"$dept.name",totalSalary:{$sum:"$salary
"}}},{$match:{totalSalary:{$lte:150000}}}]);

{ "_id" : "Comp", "totalSalary" : 130000 }

{ "_id" : "IT", "totalSalary" : 140000 }

>
db.Employee.aggregate([{$group:{_id:"$dept.name",totalSalary:{$sum:"$salary
"}}},{$sort:{totalSalary:1}}]);

{ "_id" : "Comp", "totalSalary" : 130000 }

{ "_id" : "IT", "totalSalary" : 140000 }

{ "_id" : "comp", "totalSalary" : 260000 }

{ "_id" : "ENTC", "totalSalary" : 270000 }

>
db.Employee.aggregate([{$group:{_id:"$dept.name",totalSalary:{$sum:"$salary
"}}},{$sort:{totalSalary:-1}}]);

{ "_id" : "ENTC", "totalSalary" : 270000 }

{ "_id" : "comp", "totalSalary" : 260000 }

{ "_id" : "IT", "totalSalary" : 140000 }

{ "_id" : "Comp", "totalSalary" : 130000 }

>
db.Employee.aggregate([{$group:{_id:"$dept.name",avgSalary:{$avg:"$salary"}}
},{$sort:{totalSalary:-1}}]);

{ "_id" : "comp", "avgSalary" : 130000 }

{ "_id" : "ENTC", "avgSalary" : 90000 }

{ "_id" : "Comp", "avgSalary" : 65000 }

{ "_id" : "IT", "avgSalary" : 35000 }

>
db.Employee.aggregate([{$group:{_id:"$dept.name",avgSalary:{$avg:"$salary"}}
},{$limit:2}]);

{ "_id" : "comp", "avgSalary" : 130000 }

{ "_id" : "ENTC", "avgSalary" : 90000 }

>
db.Employee.aggregate([{$group:{_id:"$dept.name",avgSalary:{$avg:"$salary"}}
},{$limit:2},{$group:{_id:"$_id.dept.name",avgOfAll:{$avg:"$avgSalary"}}}]);

{ "_id" : null, "avgOfAll" : 110000 }

```
db.Employee.aggregate([{$group:{_id:"$dept.name",maxSalary:{$max:"$salary"
}}}]);

{ "_id" : "comp", "maxSalary" : 140000 }

{ "_id" : "ENTC", "maxSalary" : 100000 }

{ "_id" : "Comp", "maxSalary" : 70000 }

{ "_id" : "IT", "maxSalary" : 60000 }
```

```
db.Employee.aggregate([{$match:{salary:{$lt:100000}}},{$group:{_id:"$dept.na
me",maxSalary:{$max:"$salary"}}}]);

{ "_id" : "ENTC", "maxSalary" : 90000 }

{ "_id" : "Comp", "maxSalary" : 70000 }

{ "_id" : "IT", "maxSalary" : 60000 }
```

Map Reduce:

```
> var map1=function(){emit(this.empName,this.salary);};

> var reduce1=function(key,values){return Array.sum(values);};


db.Employee.mapReduce(map1,reduce1,{out:"total_salary"});

{
        "result" : "total_salary",

        "timeMillis" : 77,
```

```
        "counts" : {
                "input" : 11,
                "emit" : 11,
                "reduce" : 0,
                "output" : 11
        },
        "ok" : 1
}
> db.total_salary.find().pretty();
{ "_id" : "Dangesh", "value" : 100000 }
{ "_id" : "Mangesh", "value" : 10000 }
{ "_id" : "Mohan", "value" : 140000 }
{ "_id" : "Ramesh", "value" : 60000 }
{ "_id" : "Rangesh", "value" : 50000 }
{ "_id" : "Rohan", "value" : 120000 }
{ "_id" : "Sangesh", "value" : 70000 }
{ "_id" : "Suresh", "value" : 20000 }
{ "_id" : "Tungesh", "value" : 60000 }
{ "_id" : "Vangesh", "value" : 80000 }
{ "_id" : "Zangesh", "value" : 90000 }
>
```

```
db.Employee.mapReduce(map1,reduce1,{query:{salary:{$gte:100000}},out:"total_salary"});

{
        "result" : "total_salary",
        "timeMillis" : 62,
        "counts" : {
                "input" : 3,
                "emit" : 3,
                "reduce" : 0,
                "output" : 3
        },
        "ok" : 1
}

> db.total_salary.find().pretty();

{ "_id" : "Dangesh", "value" : 100000 }

{ "_id" : "Mohan", "value" : 140000 }

{ "_id" : "Rohan", "value" : 120000 }



db.Employee.mapReduce(map1,reduce1,{query:{salary:{$lt:100000}},out:"total_salary"});

{
        "result" : "total_salary",
```

```
        "timeMillis" : 72,
        "counts" : {
                "input" : 8,
                "emit" : 8,
                "reduce" : 0,
                "output" : 8
        },
        "ok" : 1
}
> db.total_salary.find().pretty();
{ "_id" : "Mangesh", "value" : 10000 }
{ "_id" : "Ramesh", "value" : 60000 }
{ "_id" : "Rangesh", "value" : 50000 }
{ "_id" : "Sangesh", "value" : 70000 }
{ "_id" : "Suresh", "value" : 20000 }
{ "_id" : "Tungesh", "value" : 60000 }
{ "_id" : "Vangesh", "value" : 80000 }
{ "_id" : "Zangesh", "value" : 90000 }



> var map1=function(){emit(this.dept.name,this.salary);};
```

```
> var reduce1=function(key,values){return Array.sum(values);};

> db.Employee.mapReduce(map1,reduce1,{query:{salary:{$lt:100000}},out:"total_salary"});

{
        "result" : "total_salary",
        "timeMillis" : 78,
        "counts" : {
                "input" : 8,
                "emit" : 8,
                "reduce" : 3,
                "output" : 3
        },
        "ok" : 1
}

> db.total_salary.find().pretty();

{ "_id" : "Comp", "value" : 130000 }

{ "_id" : "ENTC", "value" : 170000 }

{ "_id" : "IT", "value" : 140000 }
```

Conclusion:

- Map reduce operations were implemented on a database.
- The Aggregation framework was executed.
- Different types of indexes were created and dropped.