mongoDB®

Bigdata Lab 8

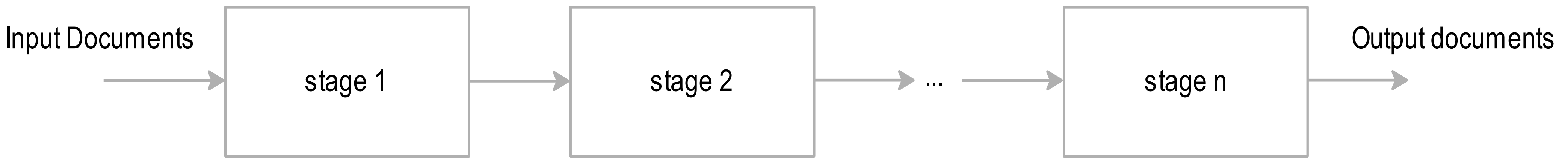
MongoDB

Aggregation in MongoDB

- Aggregation **is the process of selecting data from a collection in MongoDB**. It processes **multiple documents and returns computed** results.
- Use aggregation to **group values from multiple documents**, or perform **operations on the grouped data to return a single result**.
- Typically, you use aggregation operations to **group documents by specific field values and perform aggregations on the grouped documents** to return computed results.
- Aggregation operations can be performed in two ways:
  - Using Aggregation Pipeline.
  - Using single purpose aggregation methods: **db.collection.estimatedDocumentCount()**, **db.collection.count()** and **db.collection.distinct().**

Aggregation Pipelines

- The aggregation pipeline is an array of one or more stages passed in the db.aggregate() or db.collection.aggregate() method.



- **db.collection.aggregate([{stage1}, {stage2}, {stage3}...])**

- Each stage in the aggregation pipeline performs an operation on the input documents and returns the output documents. The output documents are then passed to the next stage. The final stage returns the calculated result.

Aggregation Pipelines

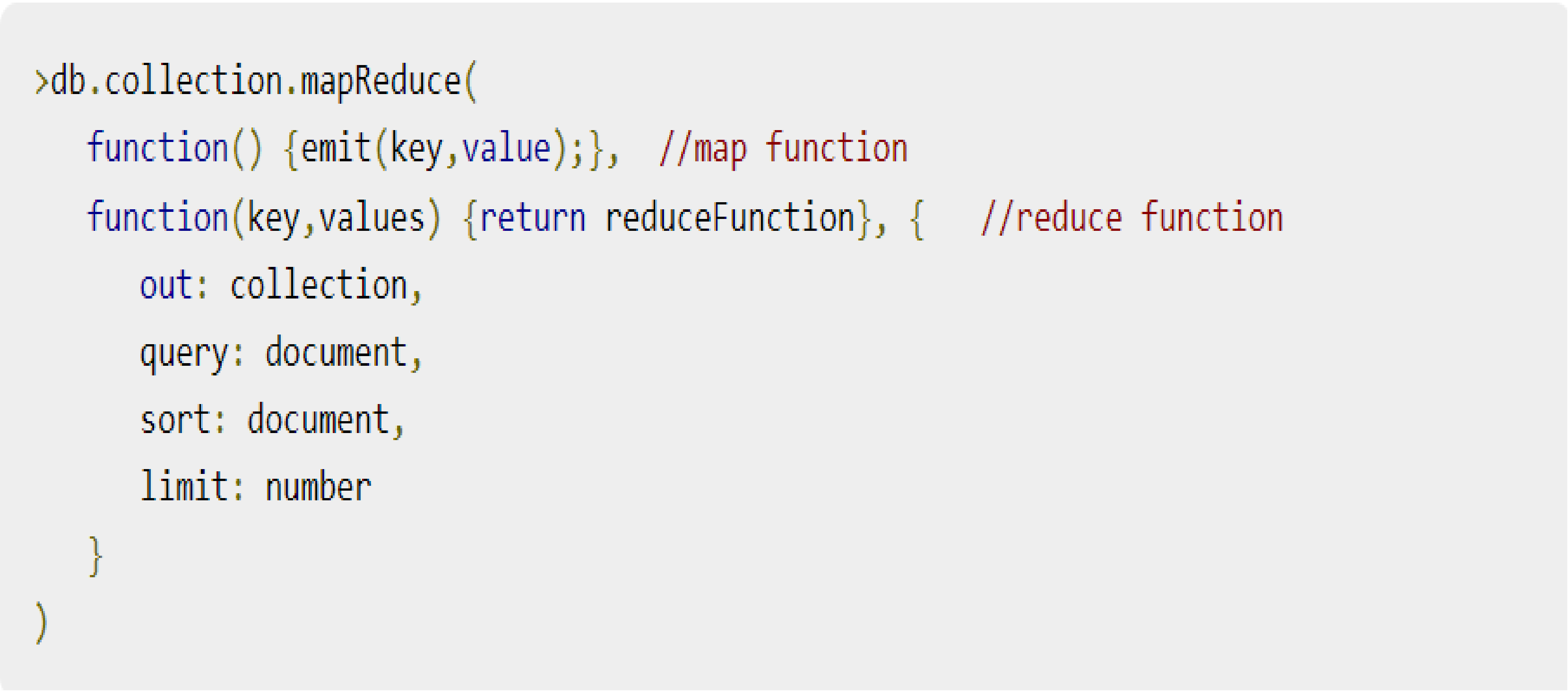
- The operations on each stage can be one of the following:
  - **$project** – select fields for the output documents.
  - **$match** – select documents to be processed.
  - **$limit** – limit the number of documents to be passed to the next stage.
  - **$sort** – sort documents.
  - **$group** – group documents by a specified key.
  - **$unwind** – breaks the embedded documents.

Let’s do Some Practice

Map-Reduce

- In MongoDB, map-reduce is a data processing programming model that helps to perform operations on large data sets and produce aggregated results.
- MongoDB provides the mapReduce() function to perform the map-reduce operations.
- This function has two main functions, i.e., map function and reduce function. The map function is used to group all the data based on the key-value and the reduce function is used to perform operations on the mapped data.
- So, the data is independently mapped and reduced in different spaces and then combined together in the function and the result will save to the specified new collection.

Map-Reduce Function



```
>db.collection.mapReduce(
   function() {emit(key,value);},  //map function
   function(key,values) {return reduceFunction}, {   //reduce function
      out: collection,
      query: document,
      sort: document,
      limit: number
   }
)
```

Thanks