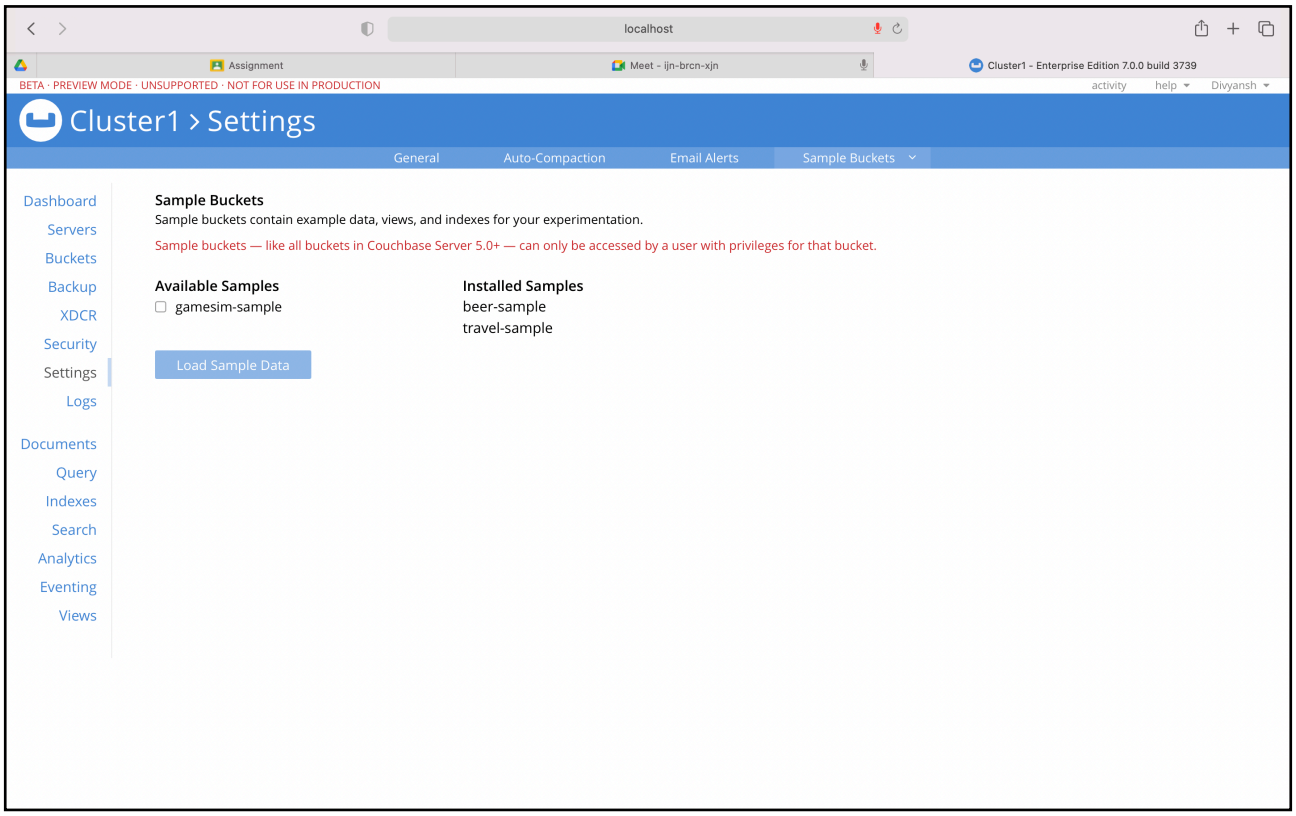
AU 2021 - NoSQL Basics & Fundamentals Assignment

Divyansh khatri

1. Import `beer-sample` bucket.

Solution.
We can use the Couchbase web interface itself to import the beer-sample bucket.
Steps.
- Go to Settings.
- Go to Sample Buckets.
- Choose beer-sample from Available Samples list.
- Click on Load Sample Data



2. Write a join query to fetch the Top 10 brewery(type=”beer”) and their country(type=”brewery”) which produces more varieties of beers.

Solution.

1. We first have to create two Datasets first namely,

- beers

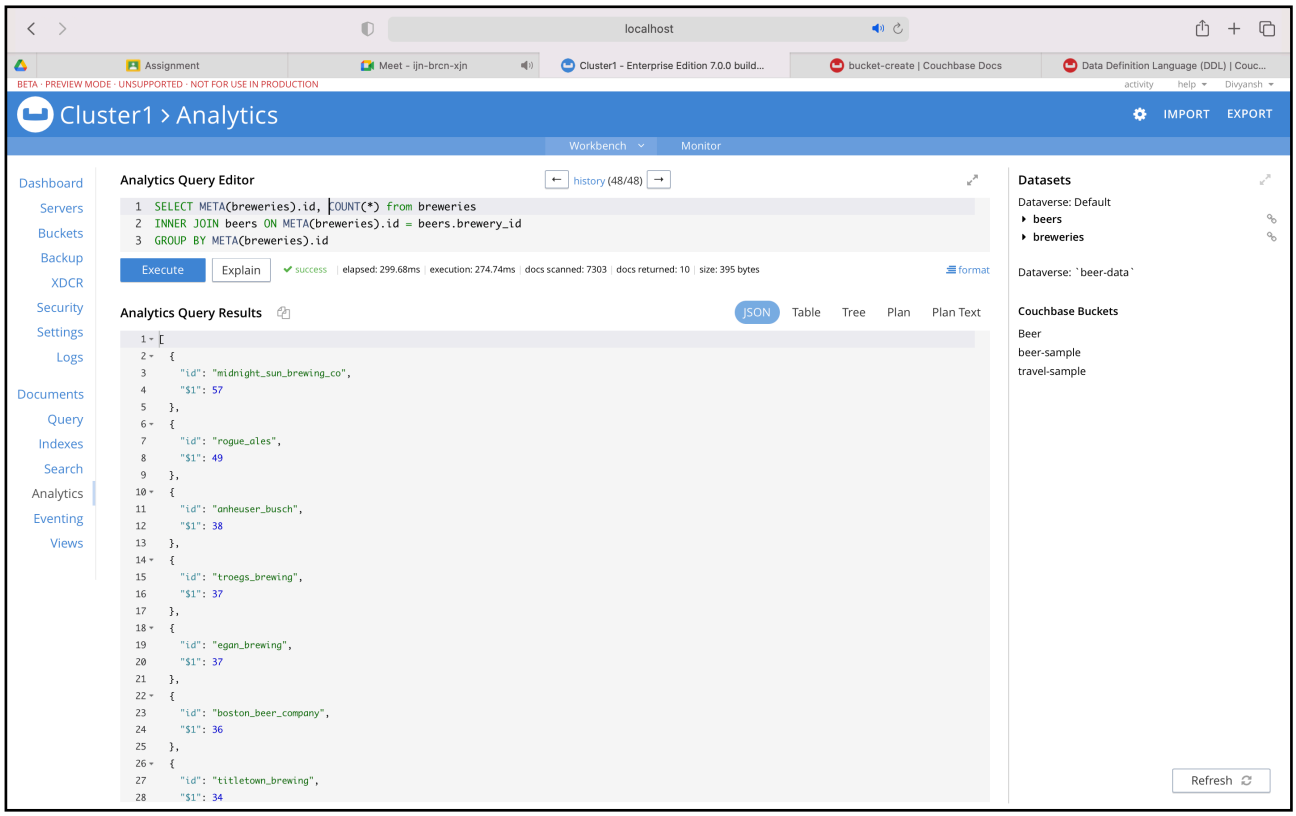
2. CREATE DATASET beers ON `beer-sample` WHERE `type` = “beer";

- breweries

CREATE DATASET breweries ON `beer-sample` WHERE `type` = “brewery”;

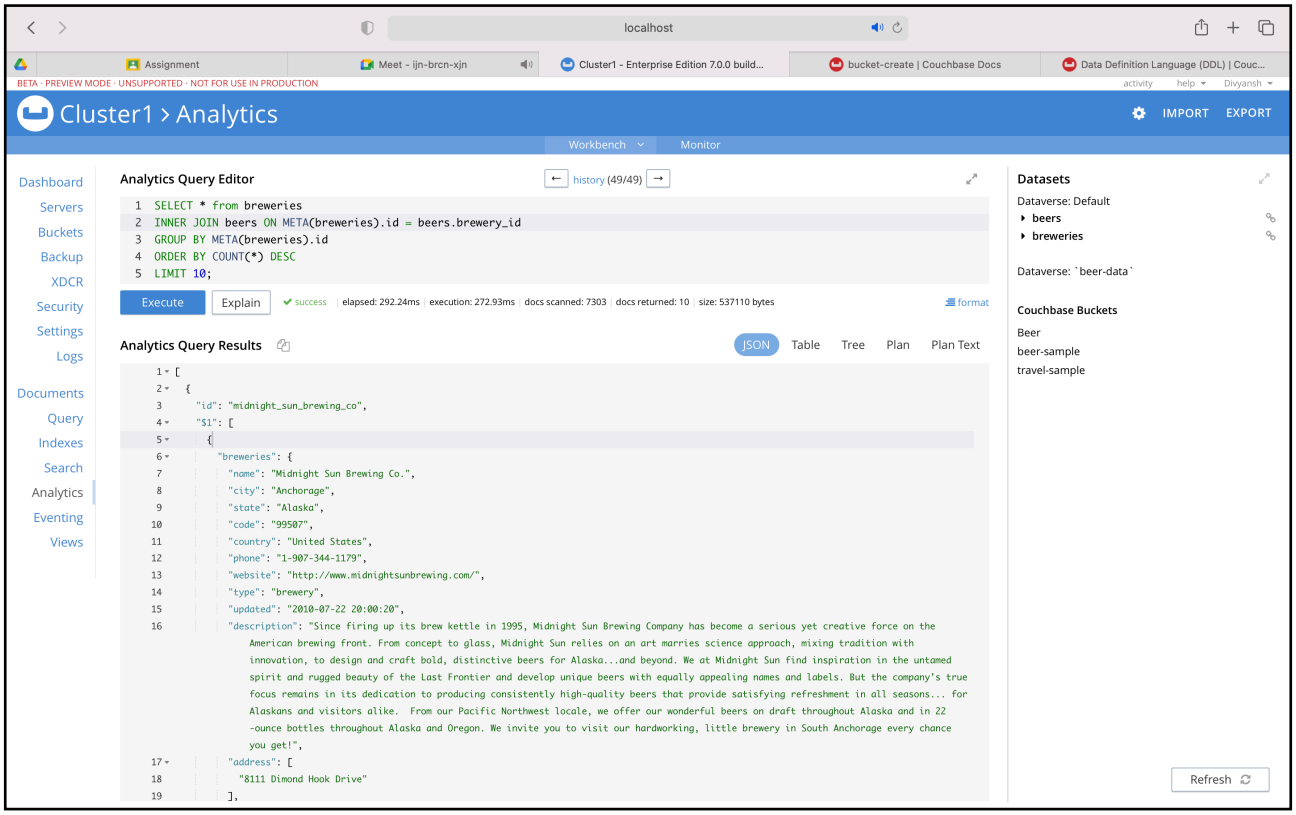
2.1 The following command can be used to get the top 10 breweries and the number of types of beer they produce.

```
SELECT META(breweries).id, COUNT(*) from breweries
INNER JOIN beers ON META(breweries).id = beers.brewery_id
GROUP BY META(breweries).id
ORDER BY COUNT(*) DESC
LIMIT 10;
```



2.2 The following command can be used to get the details of the top 10 breweries and all the types of beers with the countries that they produce.

```
SELECT * from breweries
INNER JOIN beers ON META(breweries).id = beers.brewery_id
GROUP BY META(breweries).id
ORDER BY COUNT(*) DESC
LIMIT 10;
```



3. Write a mapreduce to get the number of breweries based on country. Please attach the mapreduce code and json output screenshot.

Solution.

The map query for the following query will be -

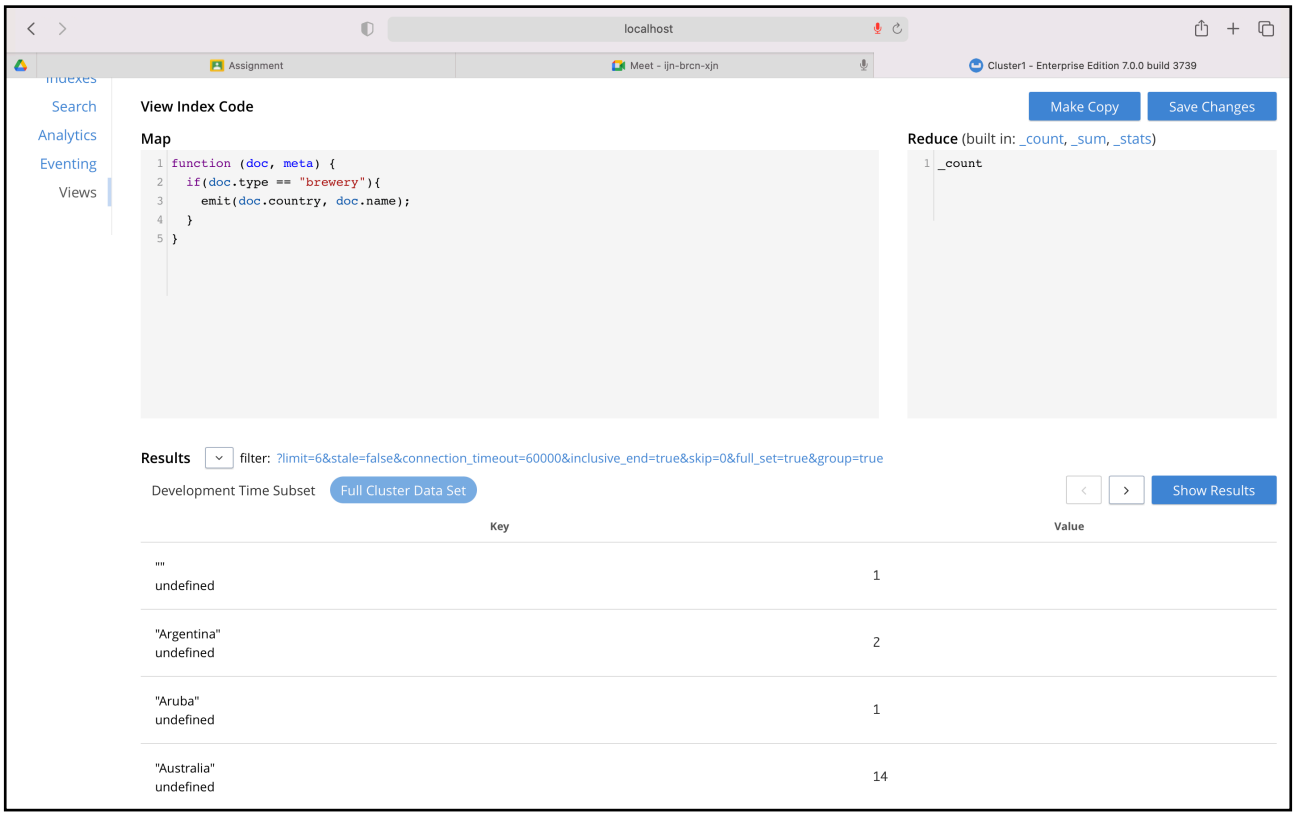
_MAP QUERY

```
function (doc, meta) {
  if(doc.type == "brewery"){
        emit(doc.country, doc.name);
  }
}
```

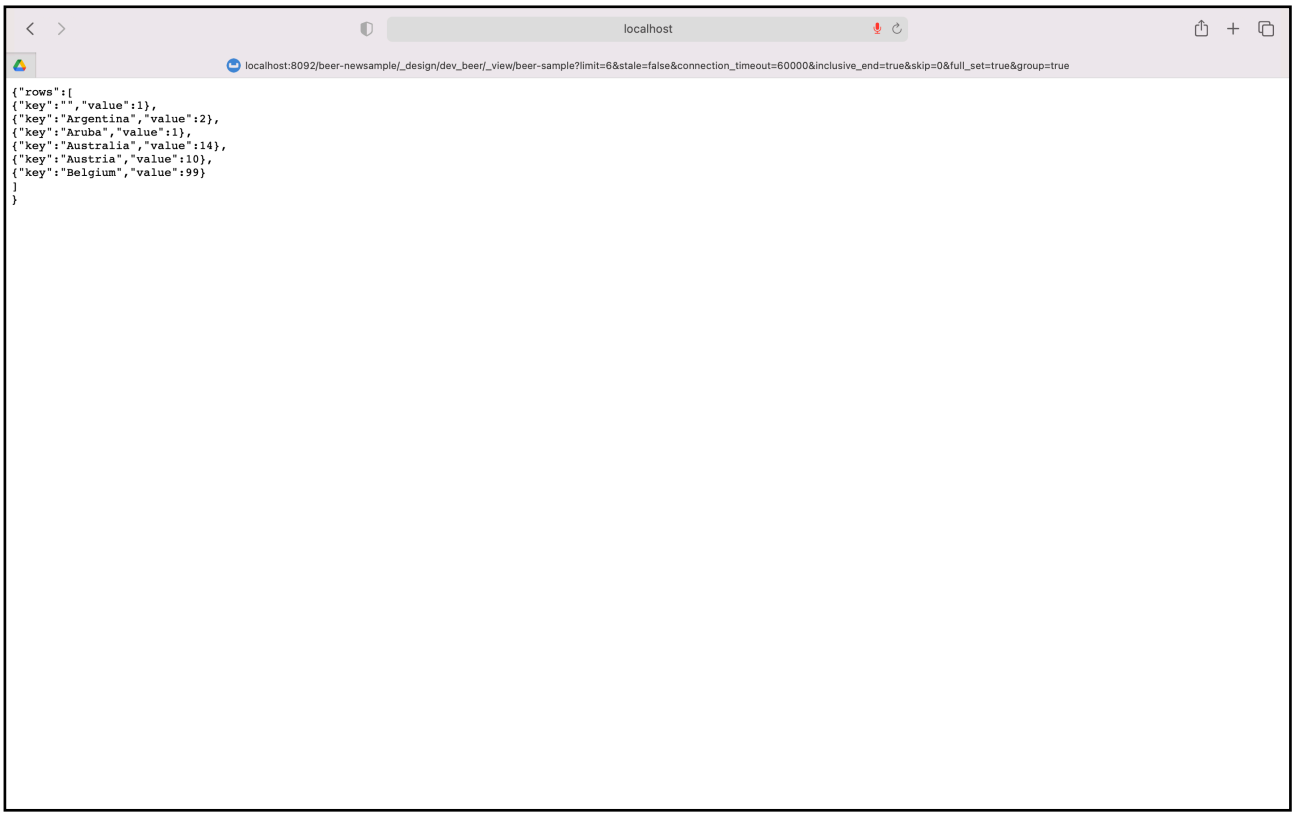
REDUCE QUERY

_count

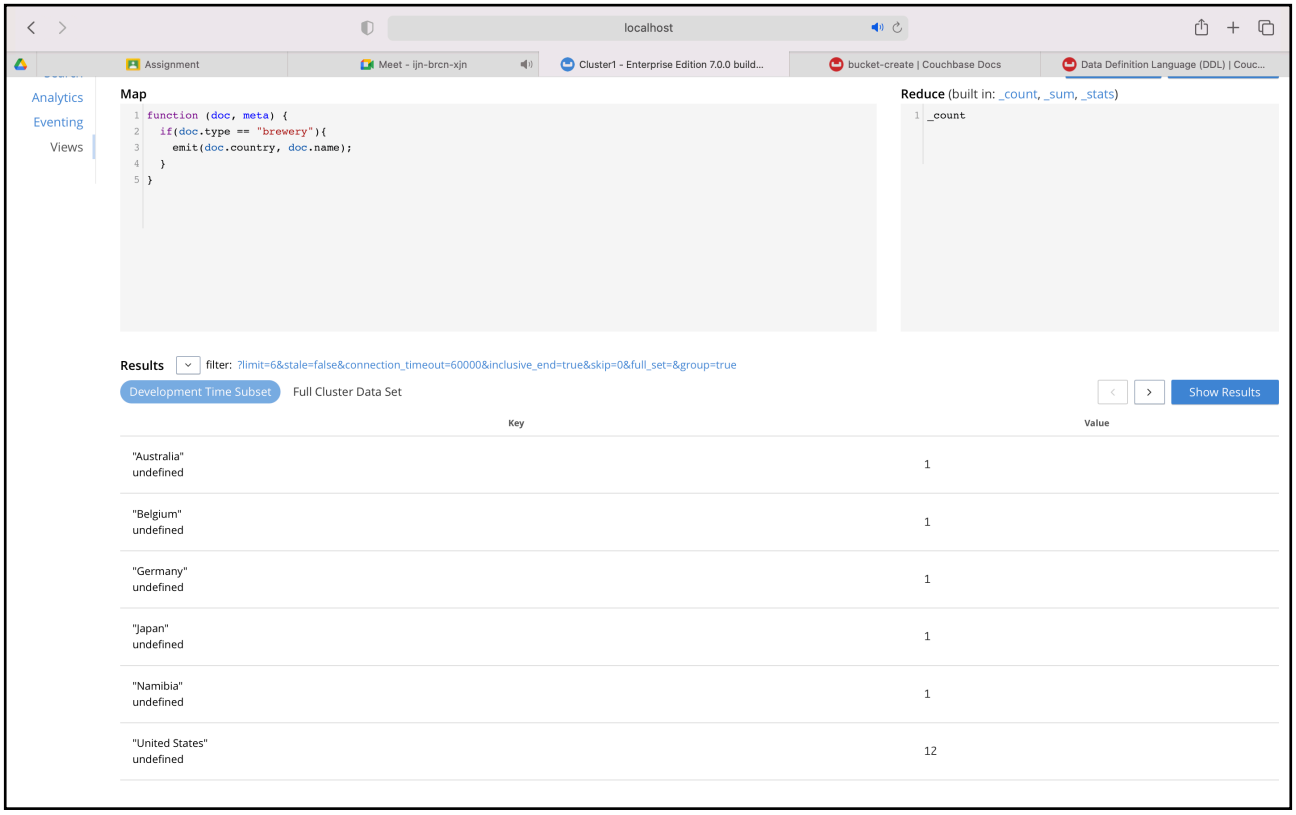
1.1. Output for Full Cluster Data Set



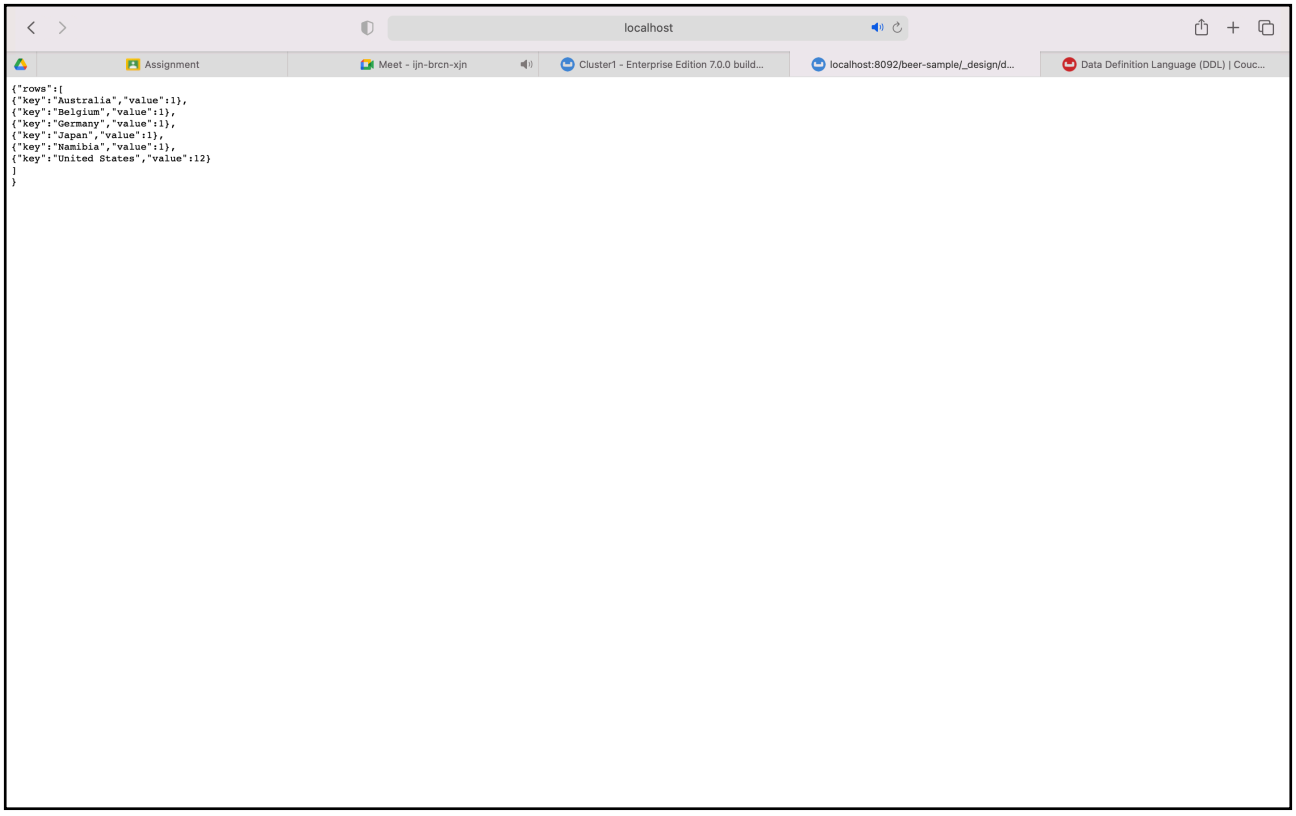
1.2. Output json for the Full Cluster Data Set which shows the country name and the number of breweries each country has.



2.1 Output for Development Time Subset

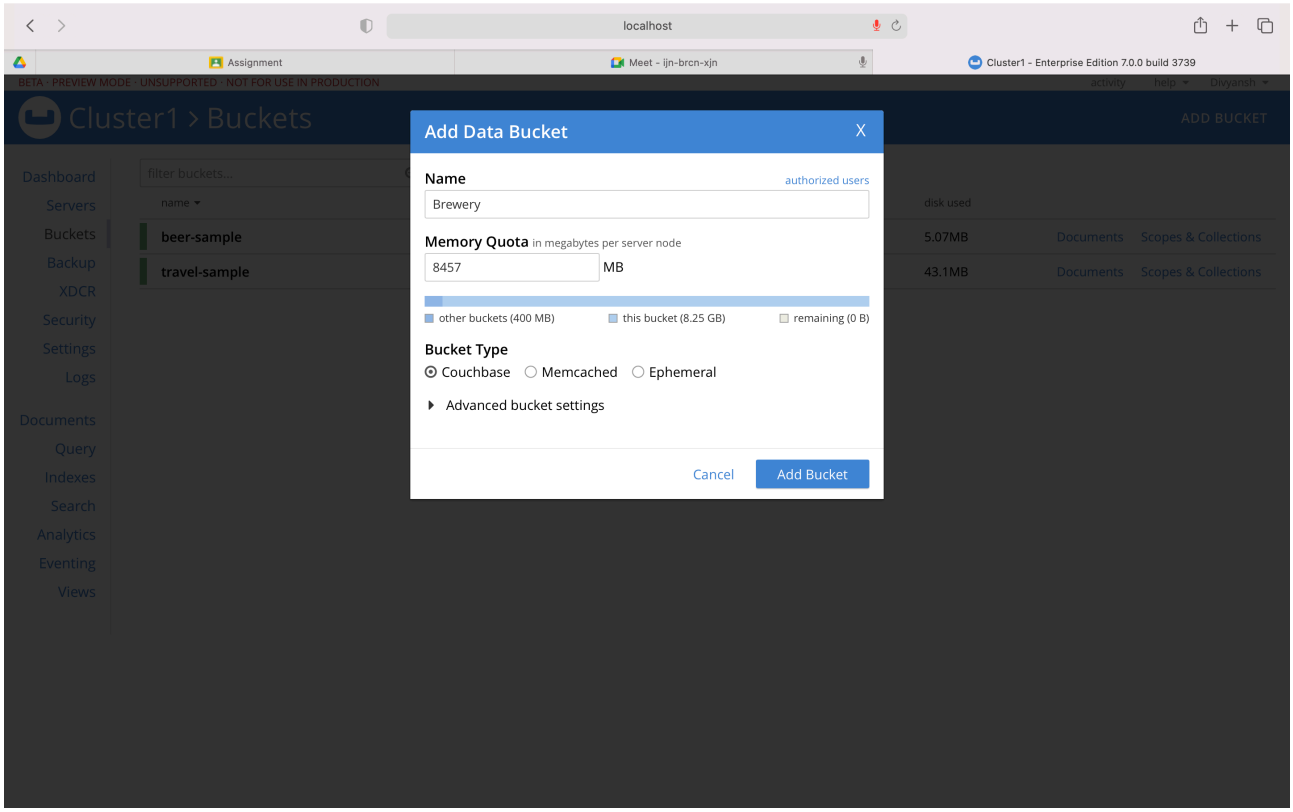


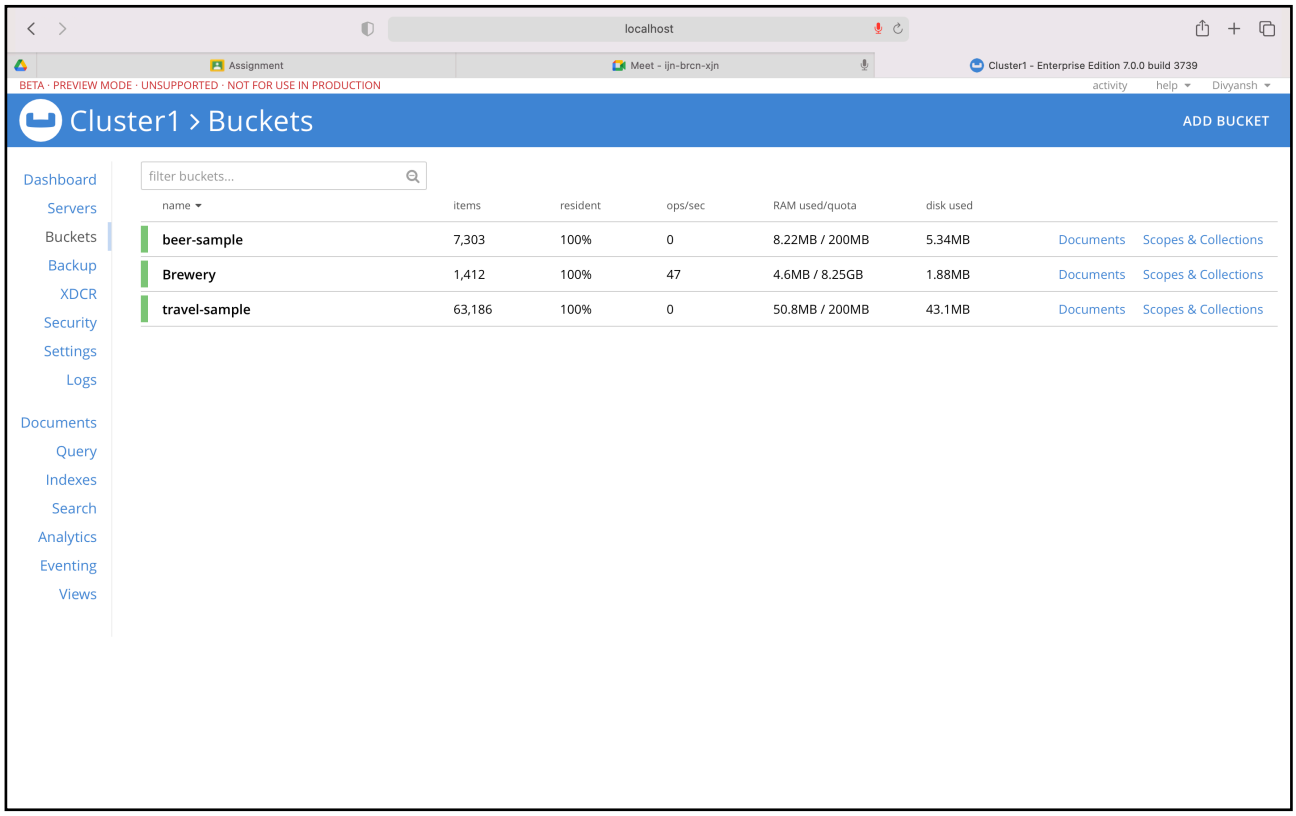
2.2 Output json for the Development Time Subset which shows the country name and the number of breweries each country has.



4. XDCR:
a) Add a new bucket “Brewery”.

We can create a new bucket by going to buckets and then clicking on ADD BUCKET on the top right corner.



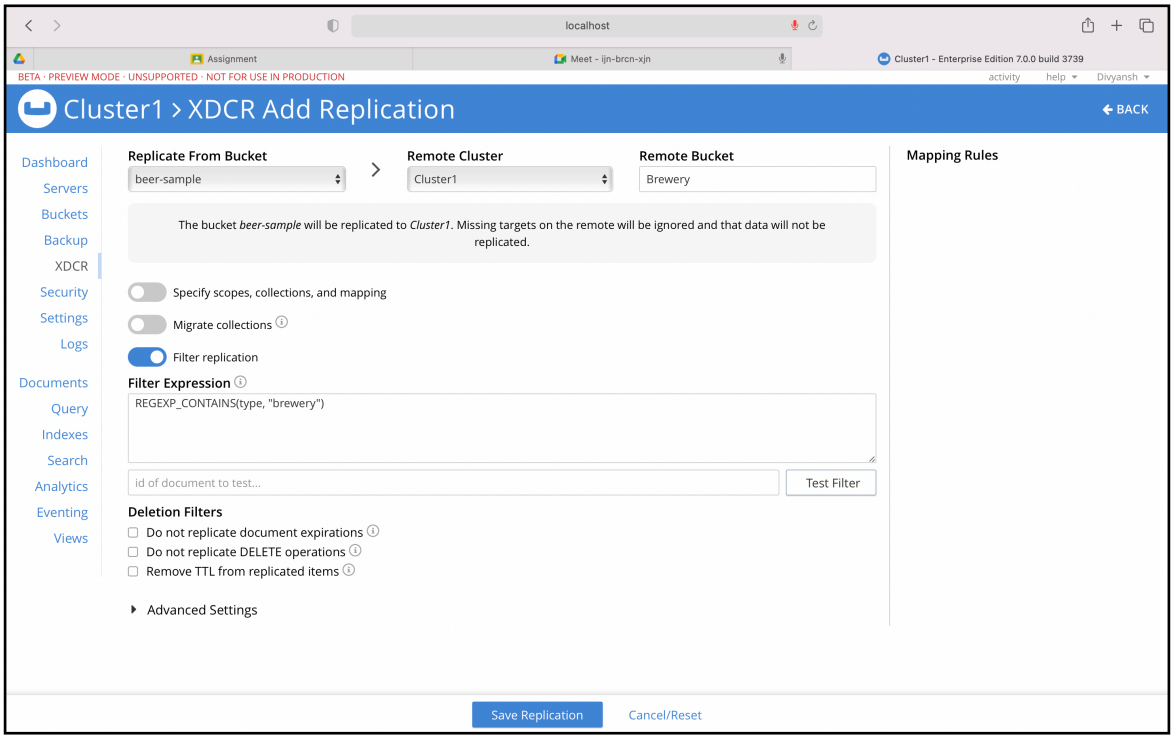


b) Create a XDCR with a filter(type='brewery') to replicate only the brewery entity from `beer-sample` bucket.

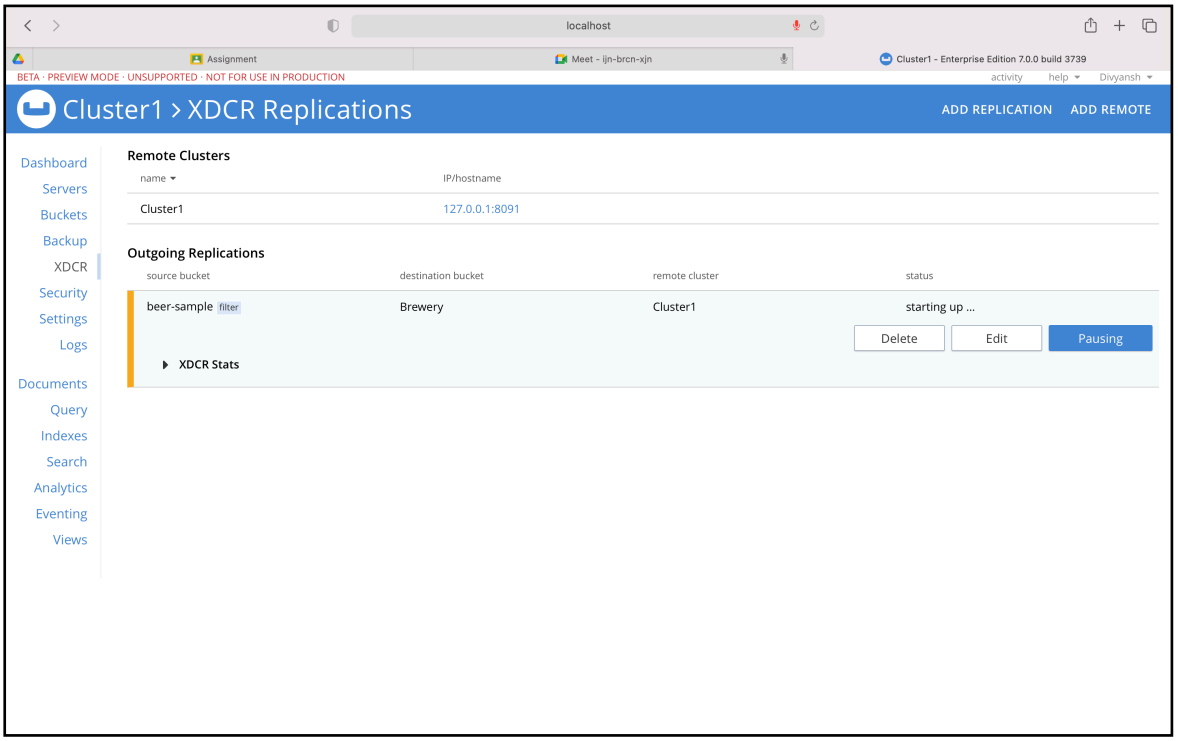
Solution. To replicate the beer-sample bucket using XDCR, we will follow the following steps-

- Click on XDCR and select ADD REPLICATION on the top right corner.
- Select FIlter Replication from the options and type the filter expression -

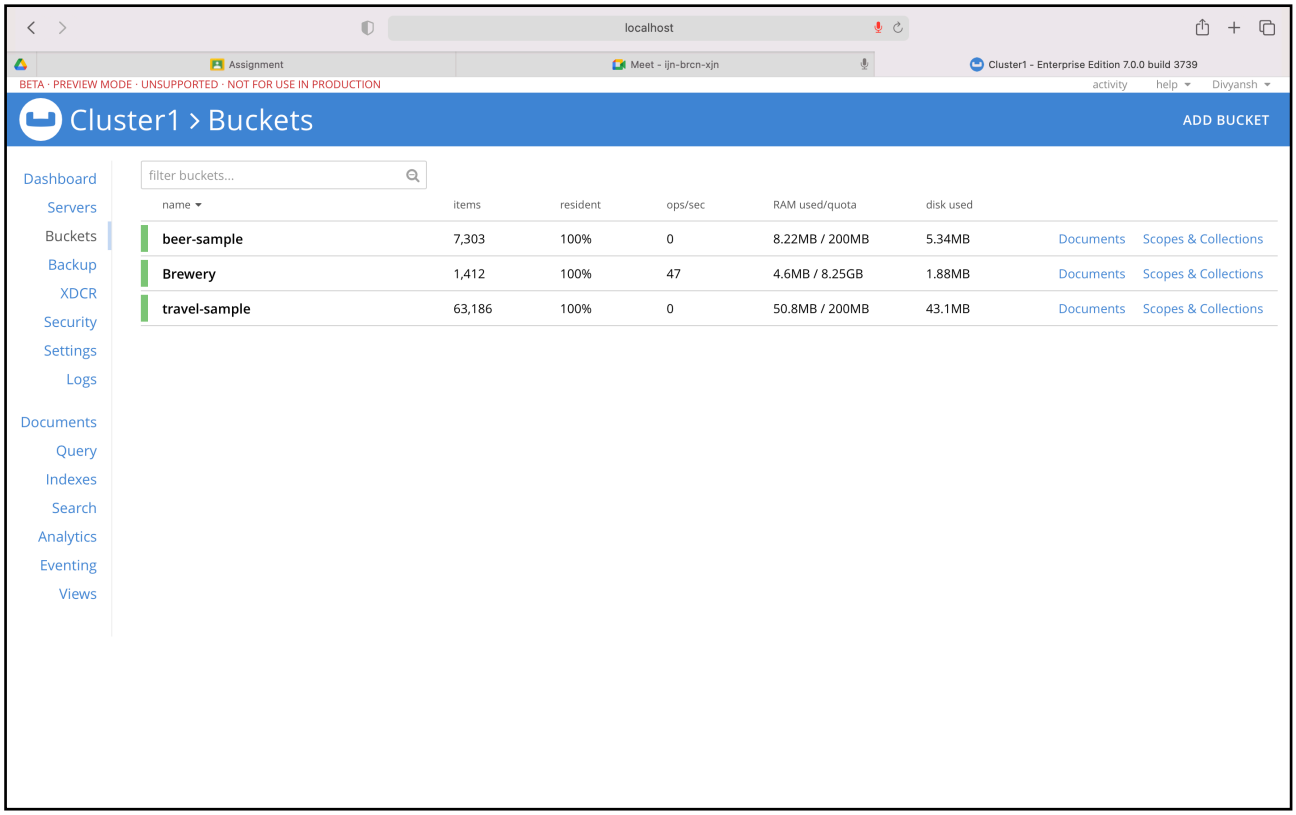
REGEXP_CONTAINS(type, "brewery")



- Click on Save replication.



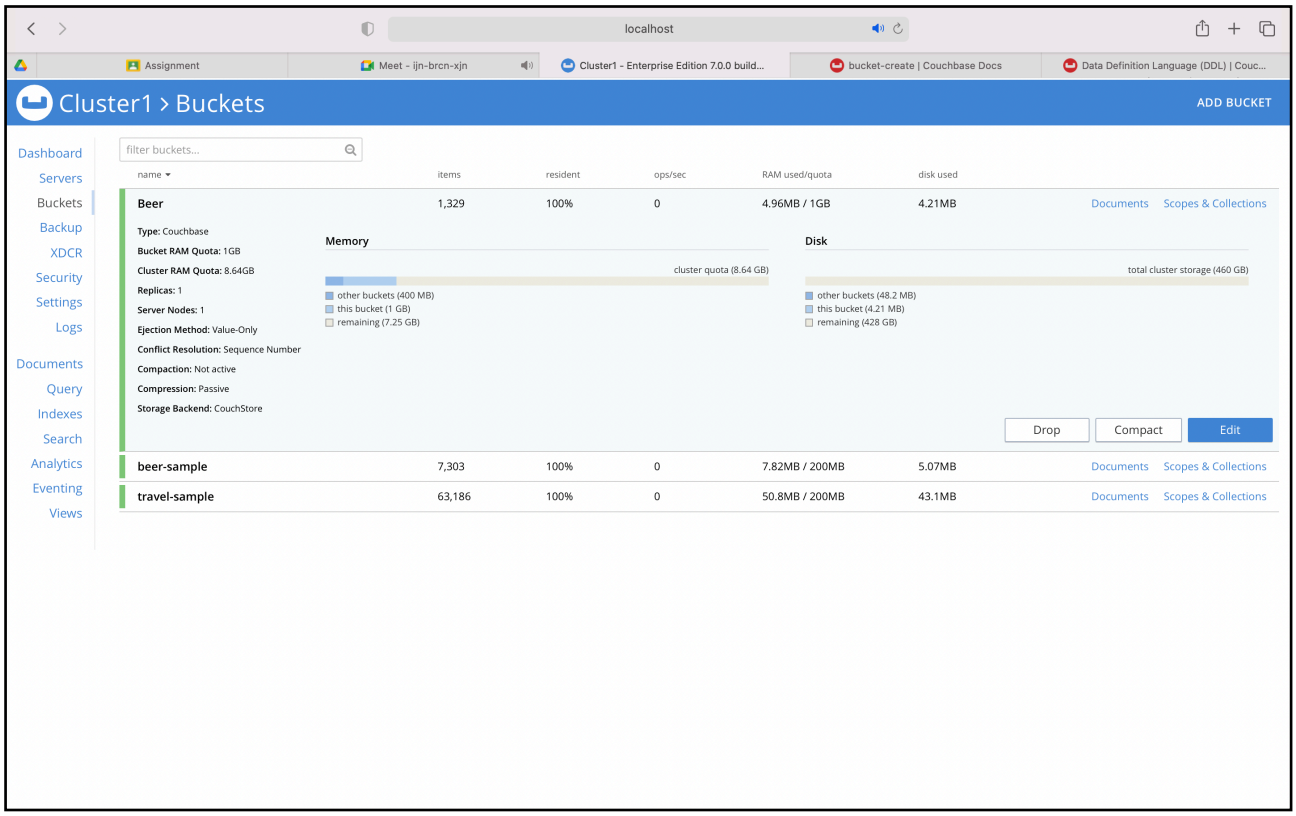
This is the result of replication into the Brewery bucket.



5. CLI:
a) Add a new bucket “Beer”.

We will use this command to create a bucket named beer -

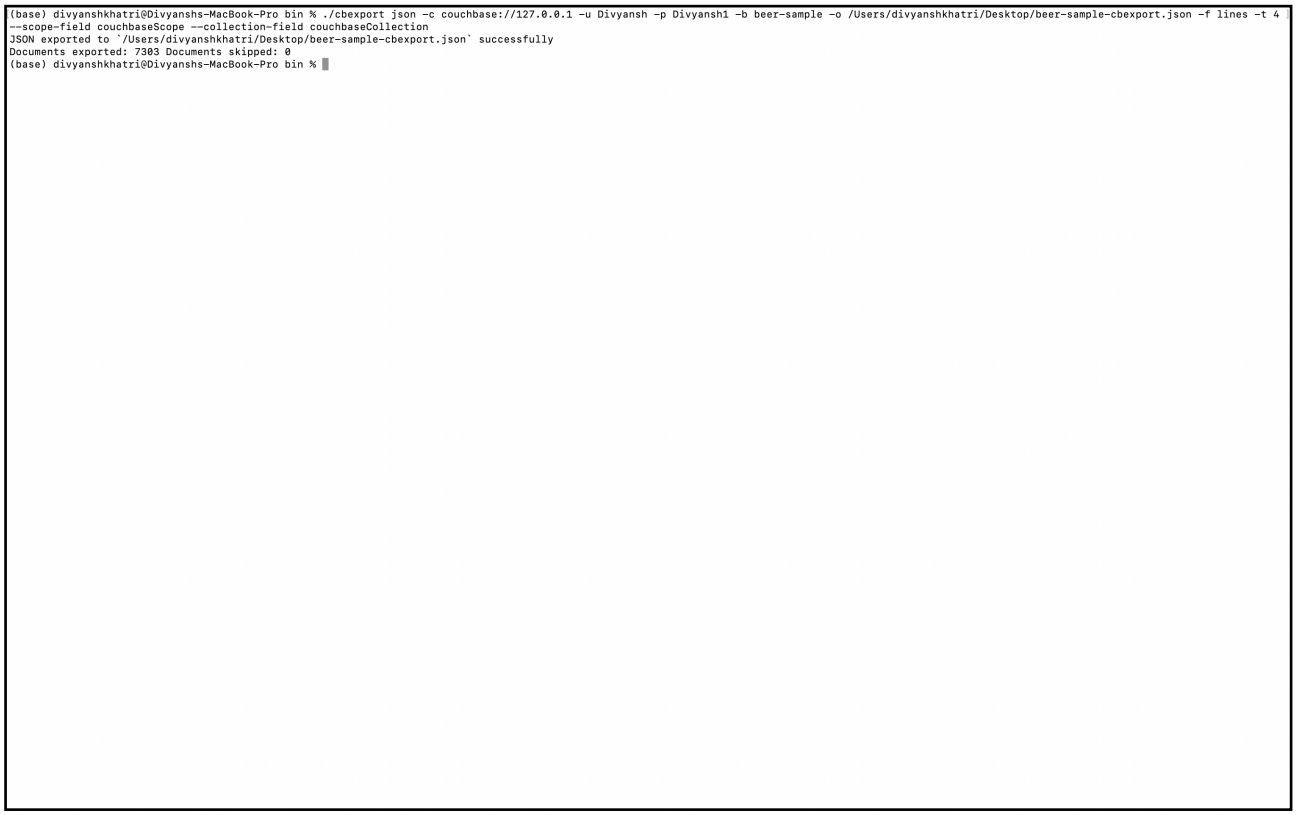
./couchbase-cli bucket-create -c couchbase://127.0.0.1 --username Divyansh \ --password Divyansh1 --bucket Beer --bucket-type couchbase \ --bucket-ramsize 1024

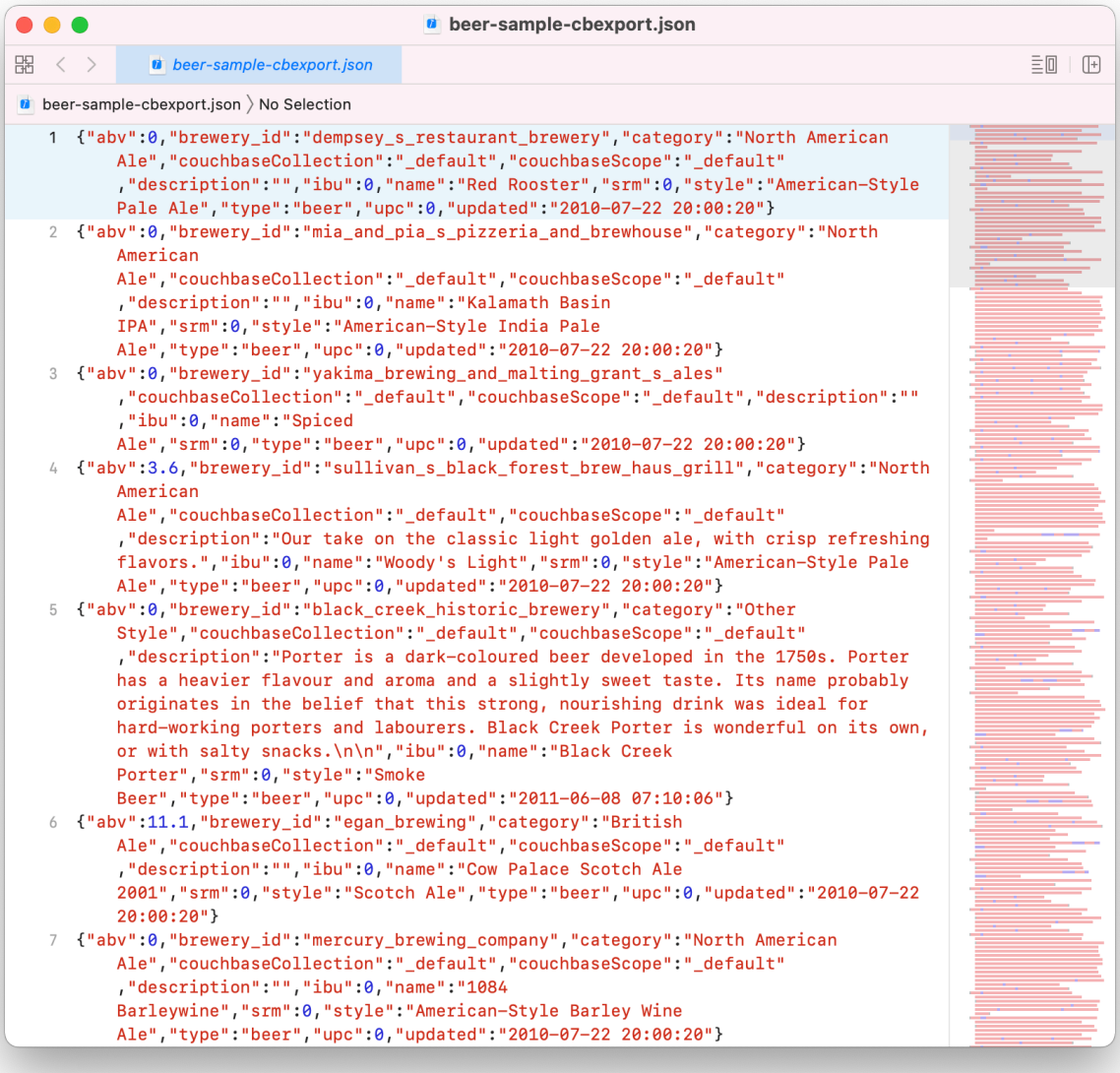


b) Using CLI - do a cbexport of the entire `beer-sample`

We can use the following command for cbexport-

./cbexport json -c couchbase://127.0.0.1 -u Divyansh -p Divyansh1 -b beer-sample -o /Users/divyanshkhatri/Desktop/beer-sample-cbexport.json -f lines -t 4 --scope-field couchbaseScope --collection-field couchbaseCollection

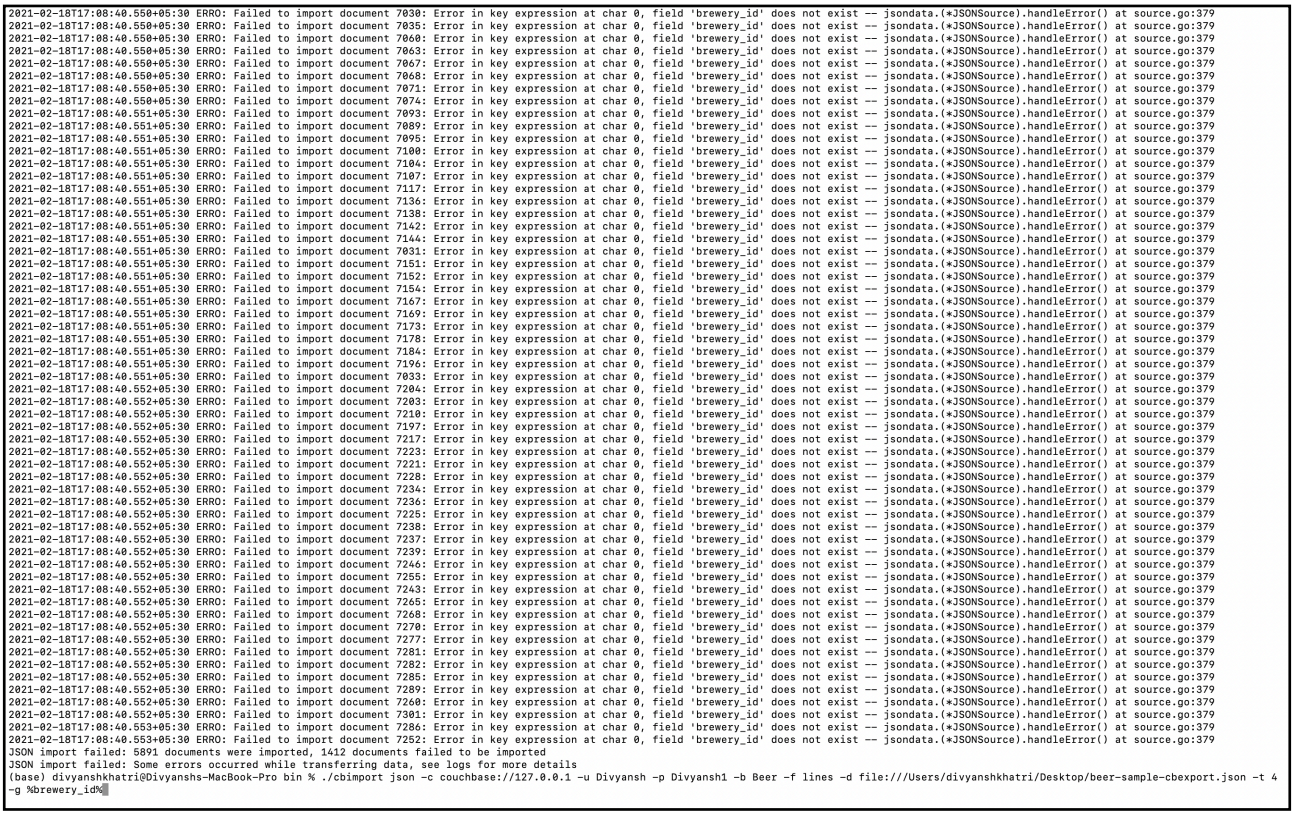




c) And do a cbimport with “brewery_id” as primary key. As a result, in the new bucket - only “beer” documents will be imported with their respective brewery name as meta().id
Solution.

We can use the following command for cbimport-

./cbimport json -c couchbase://127.0.0.1 -u Divyansh -p Divyansh1 -b Beer -f lines -d file://Users/divyanshkhatri/Desktop/beer-sample-cbexport.json -t 4 -g %brewery_id%



d) Share your observation on counts with screenshot.

Solution. As the cbimport had to include only the documents which had brewery_id in it, so it had excluded the ones which did not have the brewery_id parameter with them as we cannot have a null primary key.