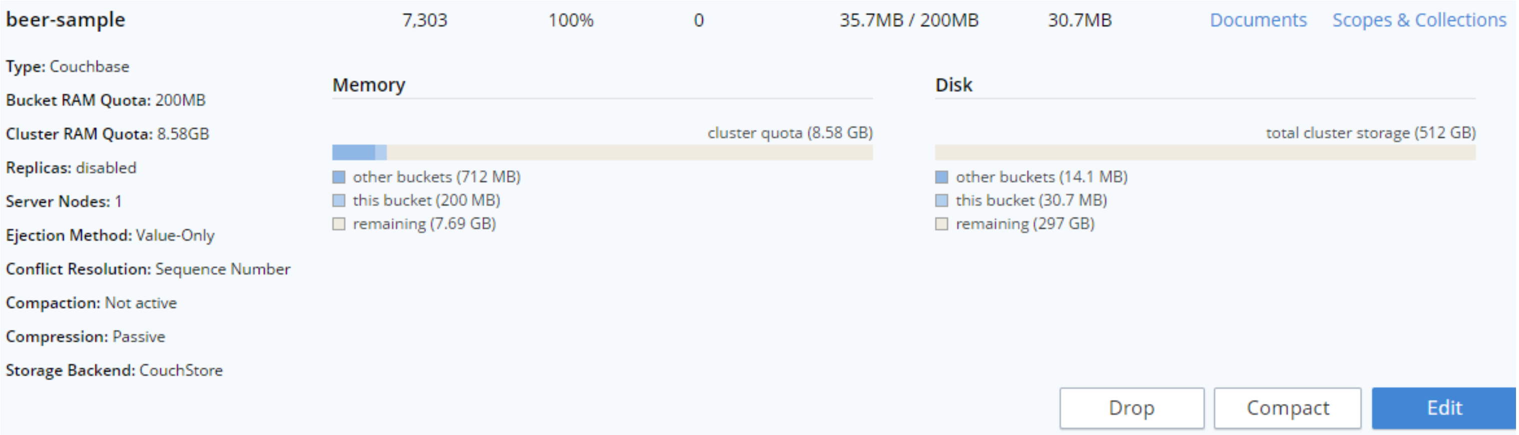
Kishan Mishra Email : kishan.mishra@accolitedigital.com

AU 2021 - NoSQL Basics &amp; Fundamentals

1: Import `beer-sample` bucket



beer-sample 7,303 100% 0 35.7MB / 200MB 30.7MB

Type: Couchbase
Bucket RAM Quota: 200MB
Cluster RAM Quota: 8.58GB
Replicas: disabled
Server Nodes: 1
Ejection Method: Value-Only
Conflict Resolution: Sequence Number
Compaction: Not active
Compression: Passive
Storage Backend: CouchStore

Memory

cluster quota (8.58 GB)

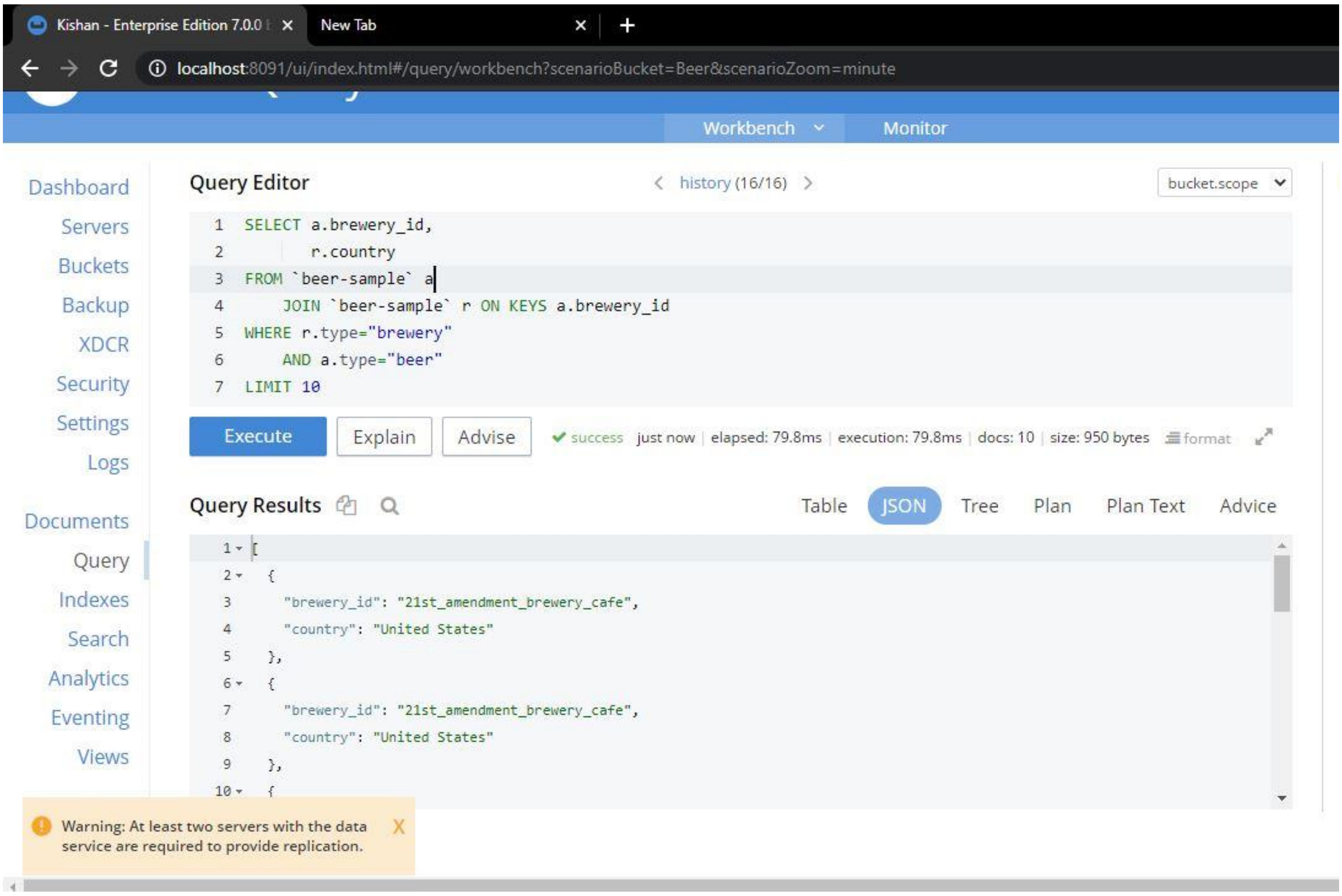
🟦 other buckets (712 MB)
🟦 this bucket (200 MB)
□ remaining (7.69 GB)

Disk

total cluster storage (512 GB)

🟦 other buckets (14.1 MB)
🟦 this bucket (30.7 MB)
□ remaining (297 GB)

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



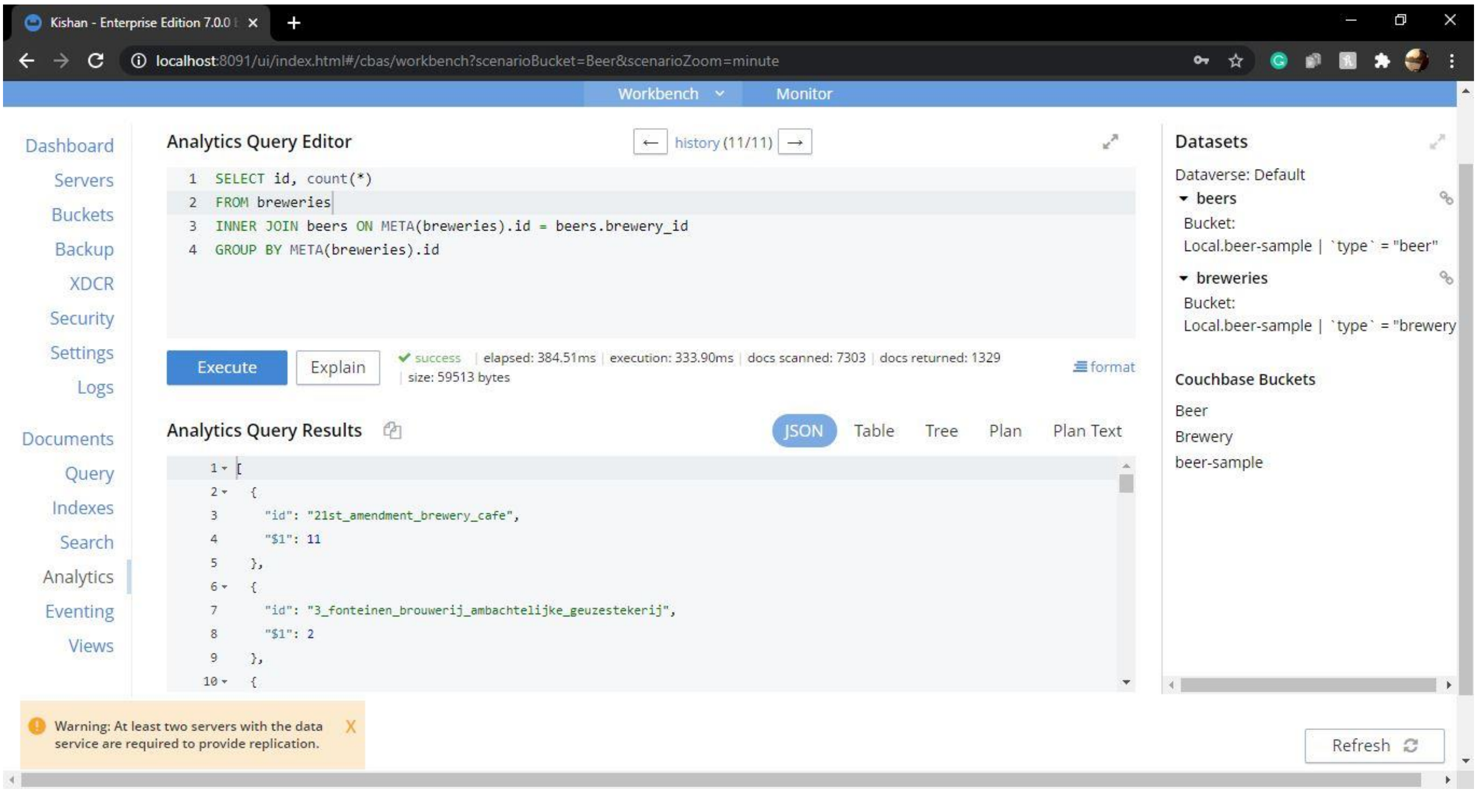
```
SELECT a.brewery_id,
       r.country
FROM `beer-sample` a
    JOIN `beer-sample` r ON KEYS a.brewery_id
WHERE r.type="brewery"
    AND a.type="beer"
LIMIT 10
```

✔ success | just now | elapsed: 79.8ms | execution: 79.8ms | docs: 10 | size: 950 bytes

```
[
  {
    "brewery_id": "21st_amendment_brewery_cafe",
    "country": "United States"
  },
  {
    "brewery_id": "21st_amendment_brewery_cafe",
    "country": "United States"
  },
  {
```

Warning: At least two servers with the data service are required to provide replication.

Basic join query to return brewery id(type = beer) and country(type = brewery)



```
SELECT id, count(*)
FROM breweries
INNER JOIN beers ON META(breweries).id = beers.brewery_id
GROUP BY META(breweries).id
```

✔ success | elapsed: 384.51ms | execution: 333.90ms | docs scanned: 7303 | docs returned: 1329 | size: 59513 bytes

```
[
  {
    "id": "21st_amendment_brewery_cafe",
    "$1": 11
  },
  {
    "id": "3_fonteinen_brouwerij_ambachtelijke_geuzestekerij",
    "$1": 2
  },
  {
```

Dataverse: Default
- beers
  Bucket:
  Local.beer-sample | `type` = "beer"
- breweries
  Bucket:
  Local.beer-sample | `type` = "brewery

Couchbase Buckets

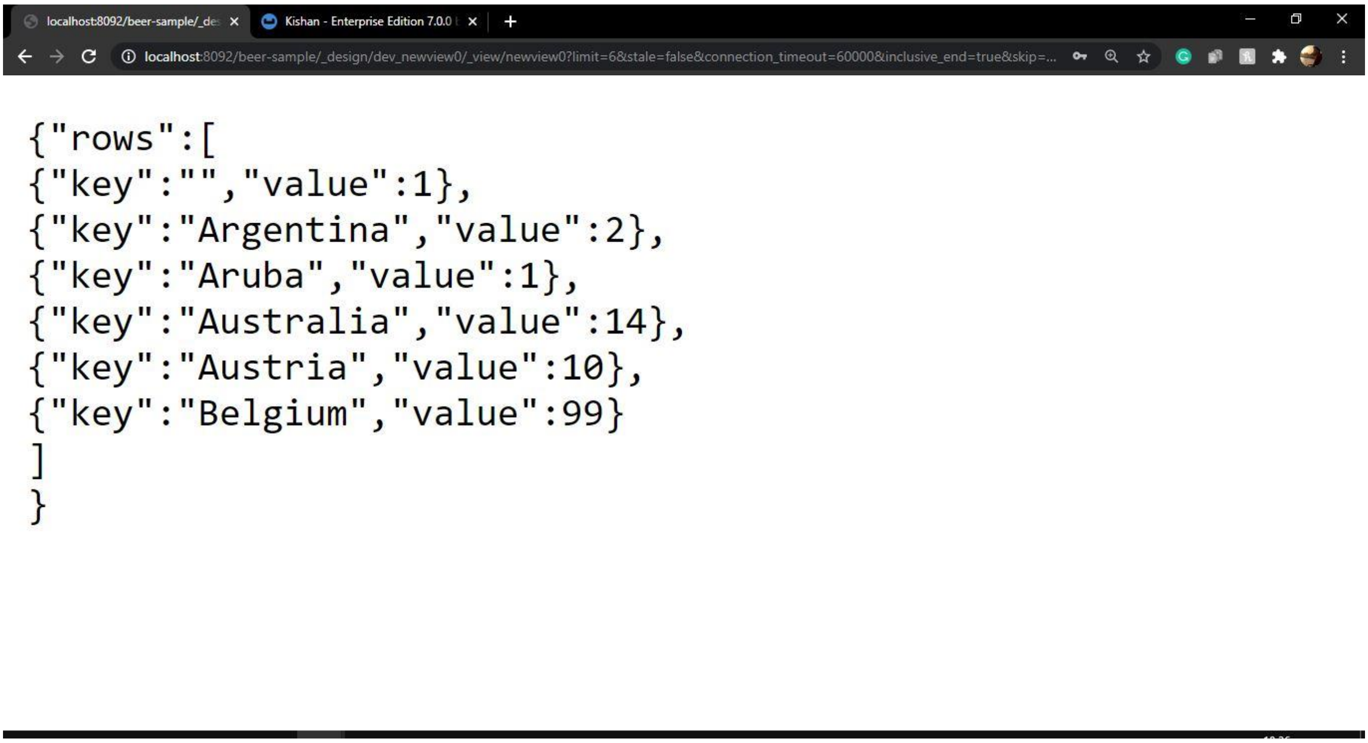
Beer
Brewery
beer-sample

Warning: At least two servers with the data service are required to provide replication.

Analytics query: First created two separate datasets: beers and breweries. For providing count of the variety of beers produced by a brewery along with it’s id

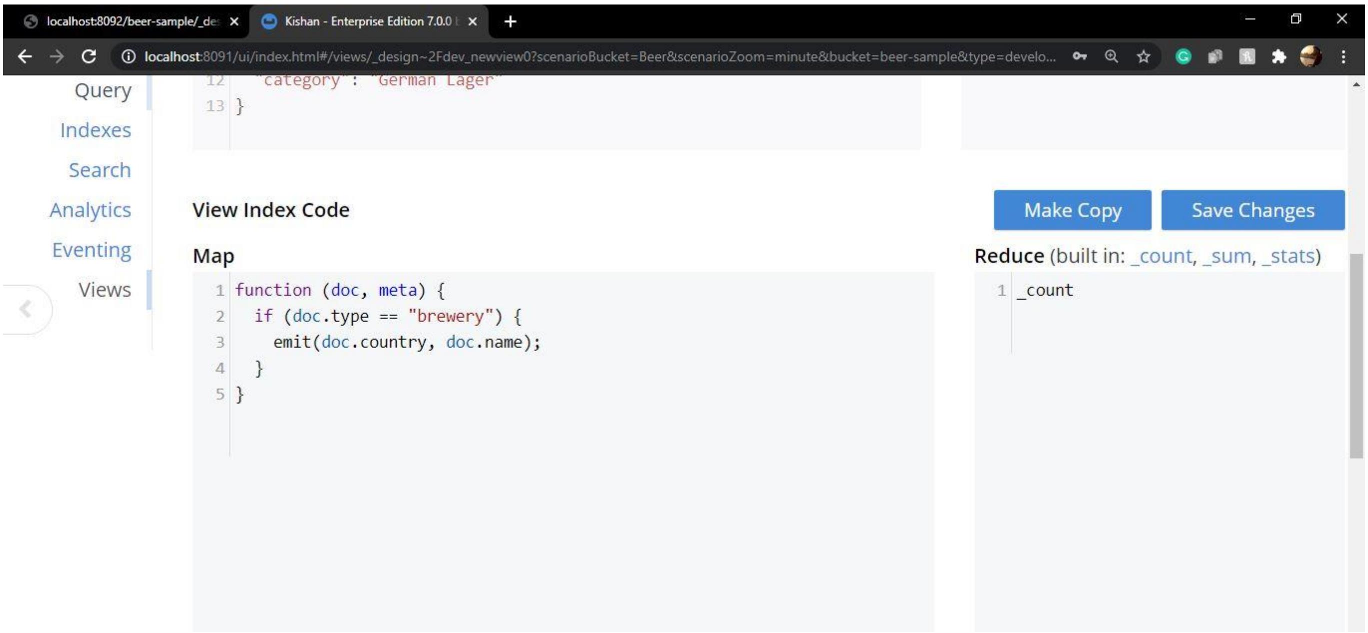
```
Breweries = CREATE DATASET breweries ON `beer-sample` WHERE `type` = "brewery";
Beers = CREATE DATASET beers ON `beer-sample` WHERE `type` = "beer";
```

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


localhost:8092/beer-sample/_des
Kishan - Enterprise Edition 7.0.0
localhost:8092/beer-sample/_design/dev_newview0/_view/newview0?limit=6&stale=false&connection_timeout=60000&inclusive_end=true&skip=...

```
{"rows":[
{"key":"","value":1},
{"key":"Argentina","value":2},
{"key":"Aruba","value":1},
{"key":"Australia","value":14},
{"key":"Austria","value":10},
{"key":"Belgium","value":99}
]
}
```

JSON output of MapReduce Query (number of breweries based on country)


localhost:8092/beer-sample/_de
Kishan - Enterprise Edition 7.0.0
localhost:8091/ui/index.html#/views/_design~2Fdev_newview0?scenarioBucket=Beer&scenarioZoom=minute&bucket=beer-sample&type=develo...
Query
Indexes
Search
Analytics
Eventing
Views
Make Copy
Save Changes

```
  "category": "German Lager"
}
```

View Index Code

Map

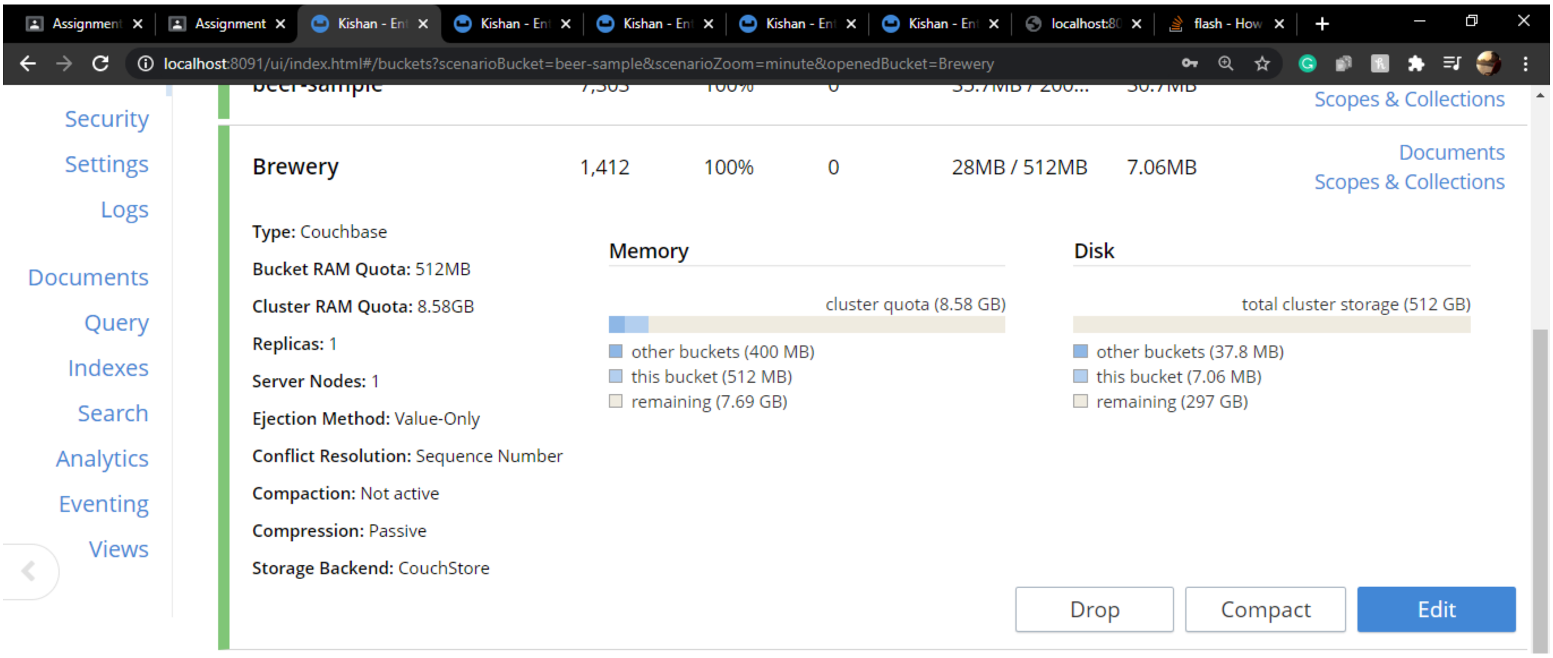
```
function (doc, meta) {
  if (doc.type == "brewery") {
    emit(doc.country, doc.name);
  }
}
```

Reduce (built in: _count, _sum, _stats)

```
_count
```

Map Reduce Code

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



Brewery	1,412	100%	0	28MB / 512MB	7.06MB

Type: Couchbase
Bucket RAM Quota: 512MB
Cluster RAM Quota: 8.58GB
Replicas: 1
Server Nodes: 1
Ejection Method: Value-Only
Conflict Resolution: Sequence Number
Compaction: Not active
Compression: Passive
Storage Backend: CouchStore

Memory

cluster quota (8.58 GB)

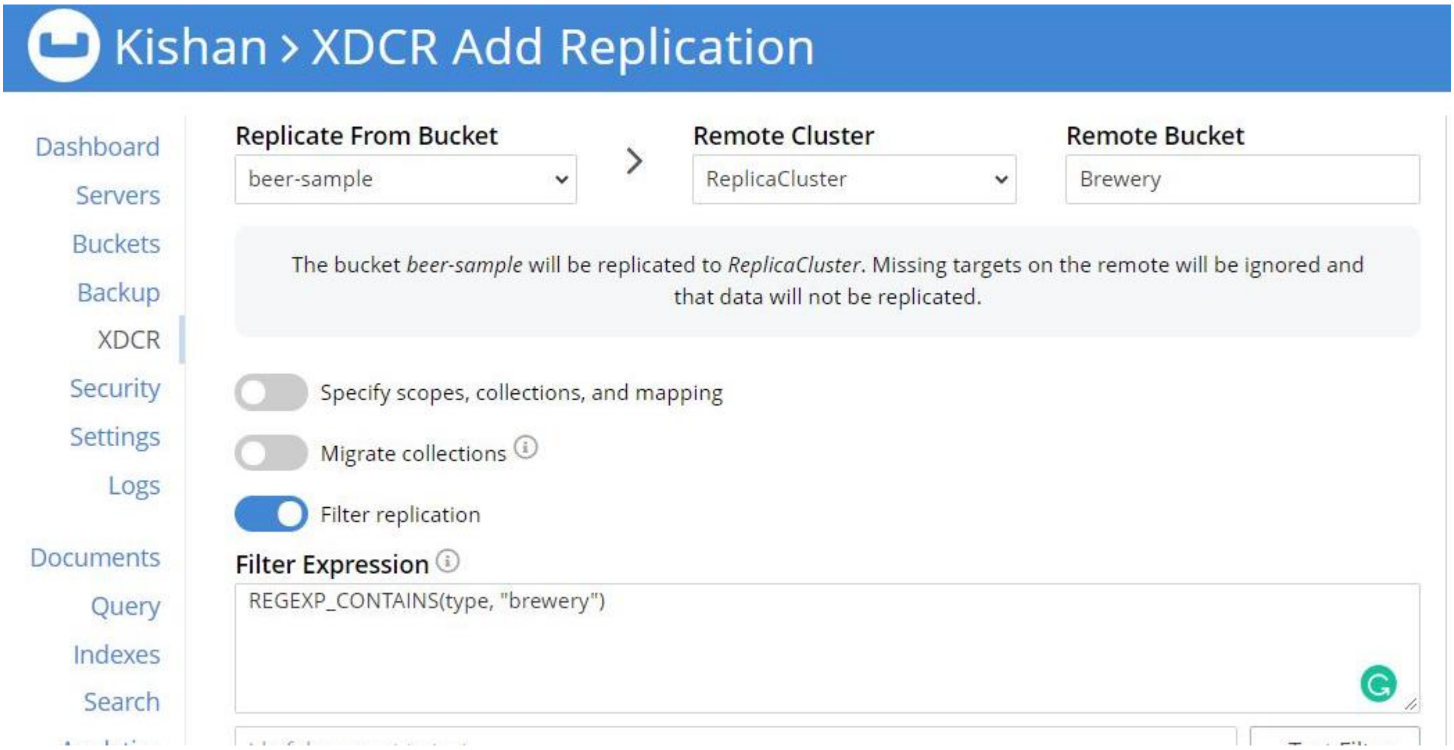
- ▨ other buckets (400 MB)
- ▨ this bucket (512 MB)
- □ remaining (7.69 GB)

Disk

total cluster storage (512 GB)

- ▨ other buckets (37.8 MB)
- ▨ this bucket (7.06 MB)
- □ remaining (297 GB)

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



Replicate From Bucket: beer-sample

Remote Cluster: ReplicaCluster

Remote Bucket: Brewery

The bucket *beer-sample* will be replicated to *ReplicaCluster*. Missing targets on the remote will be ignored and that data will not be replicated.

○ Specify scopes, collections, and mapping

○ Migrate collections

● Filter replication

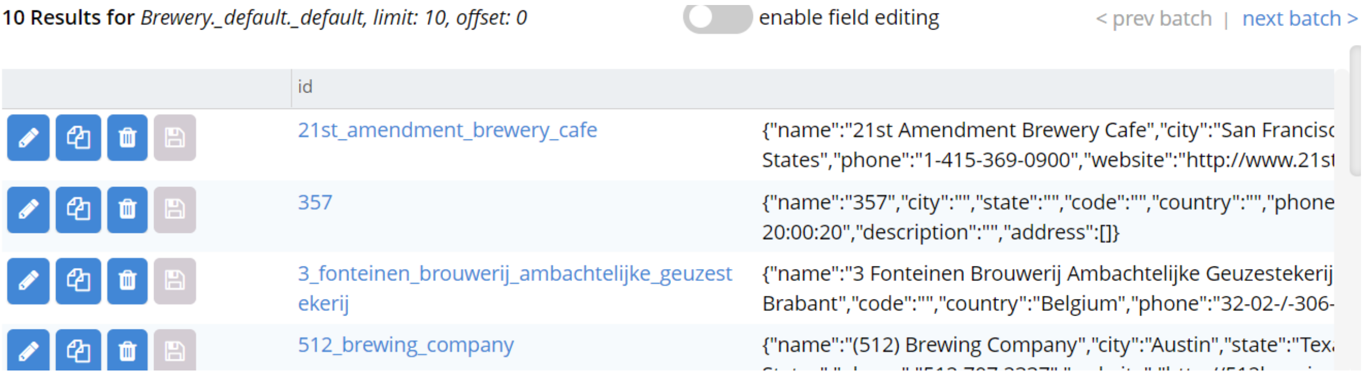
Filter Expression

```
REGEXP_CONTAINS(type, "brewery")
```

Created a new replica cluster named "ReplicaCluster", and started replication from `beer-sample` bucket to "Brewery" as Remote Bucket using "ReplicaCluster" as Remote Cluster.

Filter Expression used = REGEXP_CONTAINS(type, "brewery")

This replicates entities in Brewery bucket from beer-sample bucket, only the one's which have type = brewery.



10 Results for *Brewery._default._default, limit: 10, offset: 0*

○ enable field editing

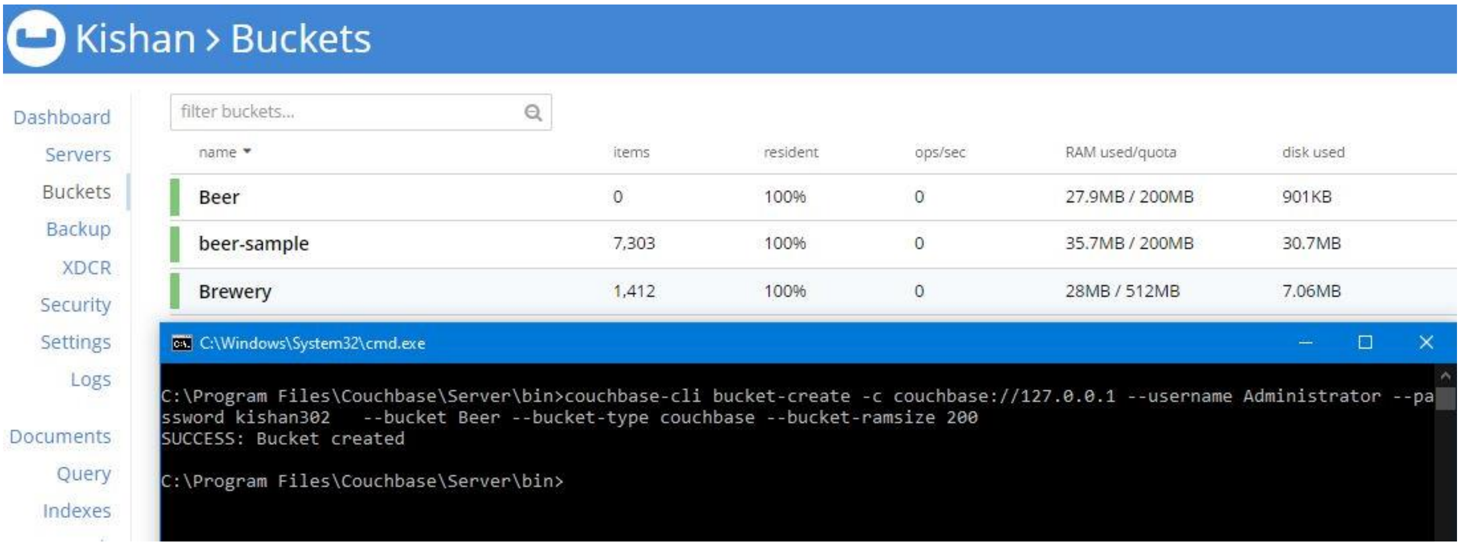
< prev batch | next batch >

id	
21st_amendment_brewery_cafe	{"name":"21st Amendment Brewery Cafe","city":"San Francisc States","phone":"1-415-369-0900","website":"http://www.21st
357	{"name":"357","city":"","state":"","code":"","country":"","phone 20:00:20","description":"","address":[]}
3_fonteinen_brouwerij_ambachtelijke_geuzestekerij	{"name":"3 Fonteinen Brouwerij Ambachtelijke Geuzestekerij Brabant","code":"","country":"Belgium","phone":"32-02-/-306-
512_brewing_company	{"name":"(512) Brewing Company","city":"Austin","state":"Tex.

Brewery table after replication

5) CLI:

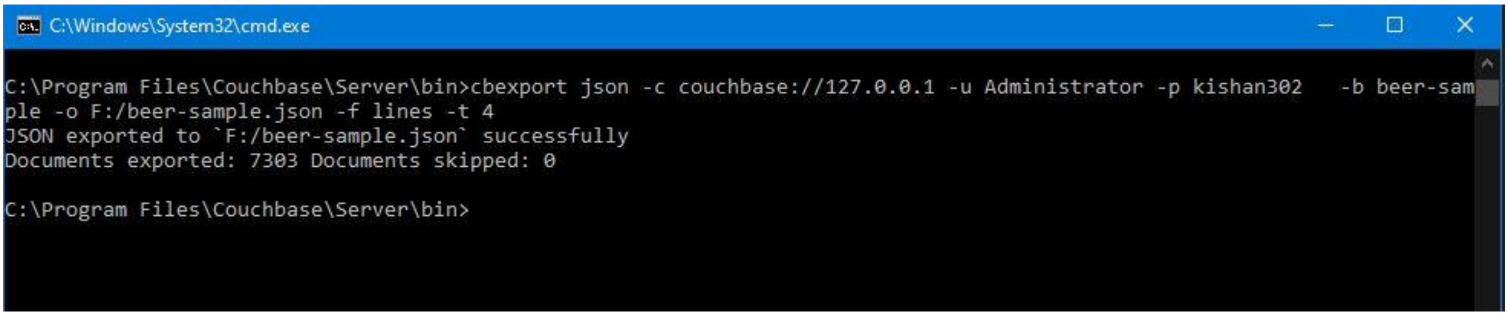
a) Add a new bucket “Beer”.



name	items	resident	ops/sec	RAM used/quota	disk used
Beer	0	100%	0	27.9MB / 200MB	901KB
beer-sample	7,303	100%	0	35.7MB / 200MB	30.7MB
Brewery	1,412	100%	0	28MB / 512MB	7.06MB

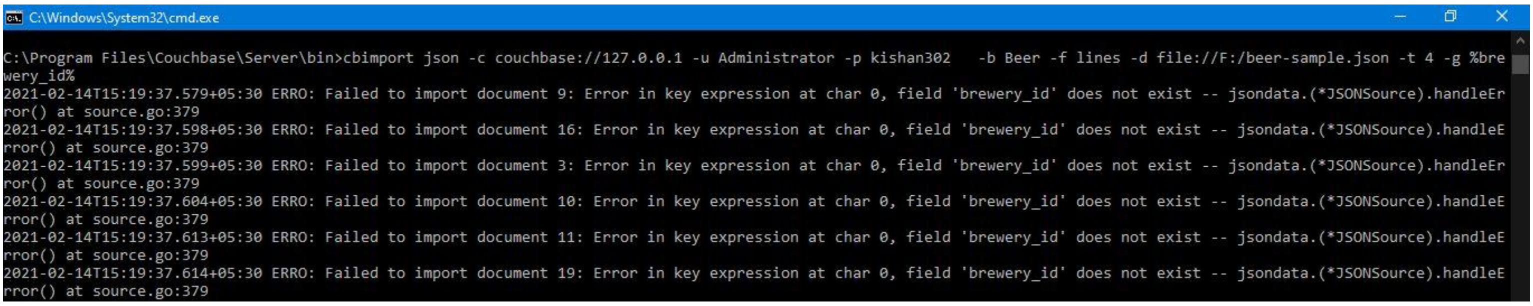
```
C:\Program Files\Couchbase\Server\bin>couchbase-cli bucket-create -c couchbase://127.0.0.1 --username Administrator --pa
ssword kishan302   --bucket Beer --bucket-type couchbase --bucket-ramsize 200
SUCCESS: Bucket created

C:\Program Files\Couchbase\Server\bin>
```

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



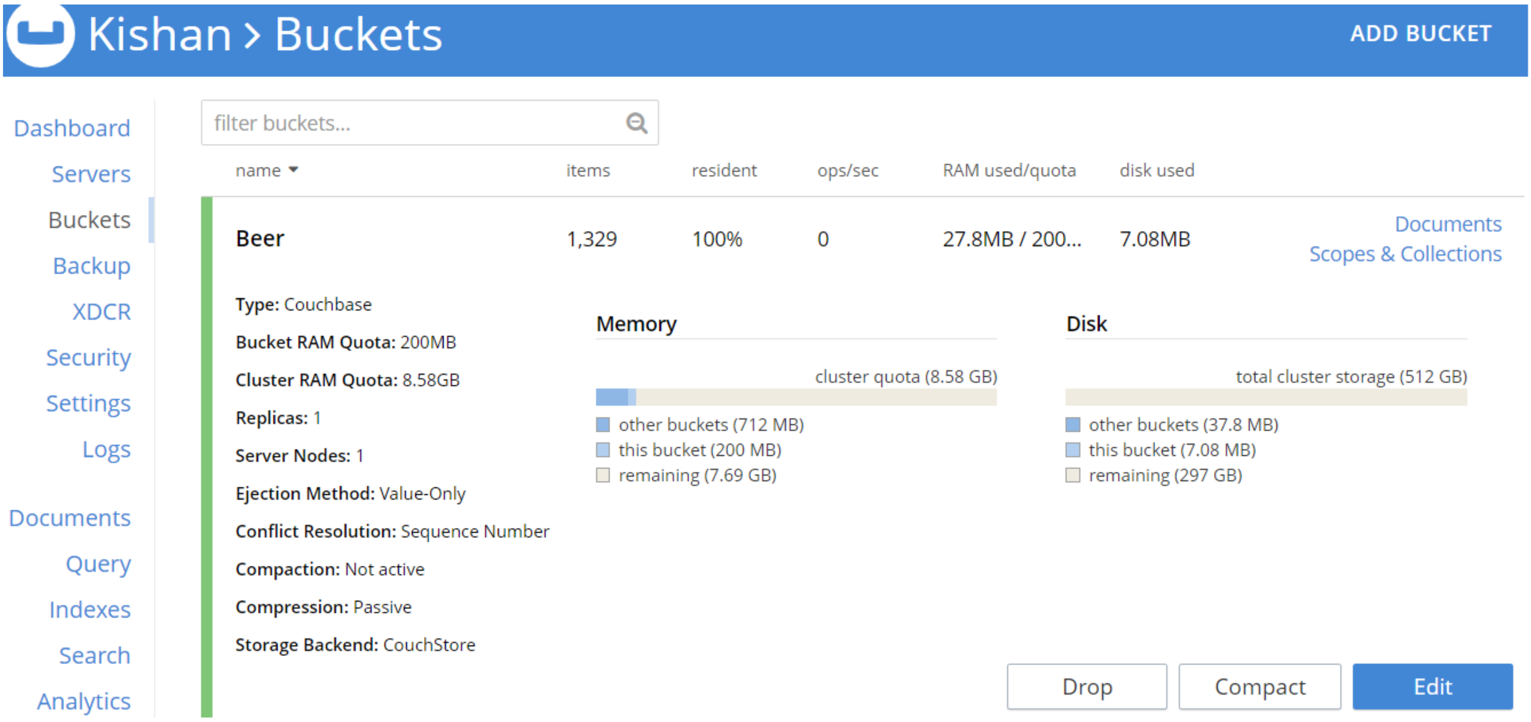
```
C:\Program Files\Couchbase\Server\bin>cbexport json -c couchbase://127.0.0.1 -u Administrator -p kishan302   -b beer-sam
ple -o F:/beer-sample.json -f lines -t 4
JSON exported to `F:/beer-sample.json` successfully
Documents exported: 7303 Documents skipped: 0

C:\Program Files\Couchbase\Server\bin>
```

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



```
C:\Program Files\Couchbase\Server\bin>cbimport json -c couchbase://127.0.0.1 -u Administrator -p kishan302   -b Beer -f lines -d file://F:/beer-sample.json -t 4 -g %bre
wery_id%
2021-02-14T15:19:37.579+05:30 ERRO: Failed to import document 9: Error in key expression at char 0, field 'brewery_id' does not exist -- jsondata.(*JSONSource).handleEr
ror() at source.go:379
2021-02-14T15:19:37.598+05:30 ERRO: Failed to import document 16: Error in key expression at char 0, field 'brewery_id' does not exist -- jsondata.(*JSONSource).handleE
rror() at source.go:379
2021-02-14T15:19:37.599+05:30 ERRO: Failed to import document 3: Error in key expression at char 0, field 'brewery_id' does not exist -- jsondata.(*JSONSource).handleEr
ror() at source.go:379
2021-02-14T15:19:37.604+05:30 ERRO: Failed to import document 10: Error in key expression at char 0, field 'brewery_id' does not exist -- jsondata.(*JSONSource).handleE
rror() at source.go:379
2021-02-14T15:19:37.613+05:30 ERRO: Failed to import document 11: Error in key expression at char 0, field 'brewery_id' does not exist -- jsondata.(*JSONSource).handleE
rror() at source.go:379
2021-02-14T15:19:37.614+05:30 ERRO: Failed to import document 19: Error in key expression at char 0, field 'brewery_id' does not exist -- jsondata.(*JSONSource).handleE
rror() at source.go:379
```

It imports only those documents which contain “brewery_id”, Hence successfully imports 1329 documents which contain “brewery_id”, and fails to import 5974 documents, because they don’t contain a field called “brewery_id”.



name ▾	items	resident	ops/sec	RAM used/quota	disk used
Beer	1,329	100%	0	27.8MB / 200...	7.08MB

Type: Couchbase
Bucket RAM Quota: 200MB
Cluster RAM Quota: 8.58GB
Replicas: 1
Server Nodes: 1
Ejection Method: Value-Only
Conflict Resolution: Sequence Number
Compaction: Not active
Compression: Passive
Storage Backend: CouchStore

Memory

cluster quota (8.58 GB)

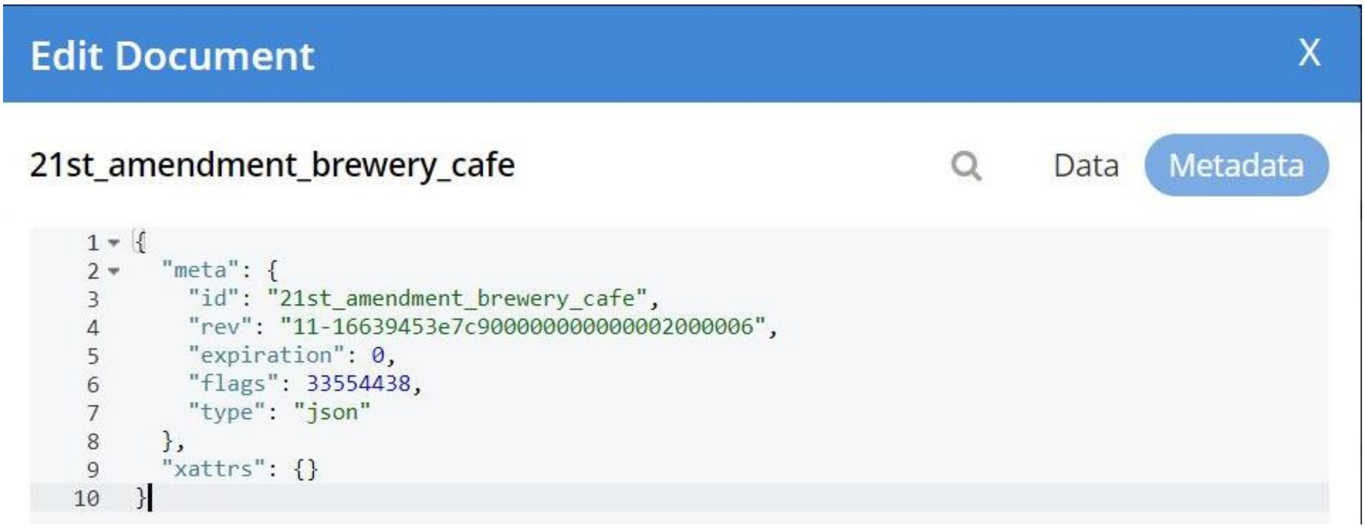
- ▨ other buckets (712 MB)
- ▨ this bucket (200 MB)
- □ remaining (7.69 GB)

Disk

total cluster storage (512 GB)

- ▨ other buckets (37.8 MB)
- ▨ this bucket (7.08 MB)
- □ remaining (297 GB)

Beer bucket after cbimport (contains 1329 documents)



```
{
  "meta": {
    "id": "21st_amendment_brewery_cafe",
    "rev": "11-16639453e7c90000000000000002000006",
    "expiration": 0,
    "flags": 33554438,
    "type": "json"
  },
  "xattrs": {}
}
```

Beer bucket’s meta().id = brewery name