# NO SQL Exercise.

In this exercise you will learn how to make a DynamoDB application.

https://aws.amazon.com/dynamodb/

DynamoDB is a key-value store.

## Before we start

1. Start a cluster on EMR. (make it cheep and small this time)
2. Start a DynamoDB key value store.

## How to start a DynamoDB

1. On AWS, click the AWS icon in the top left.
2. Use the search bar and search for DynamoDB, and click on the DynamoDB option
3. Once inside, click create table
4. Give it a name, in the examples further down it is "testDB"
5. In PartitionKey, chose string, and set it to "value" without the ". (it needs to be value based on the table header name committed to DynamoDB.)
6. Click create table.

## Exercise

This exercise has two programs.

- WordStreamProducer, that supplies a stream of words, like last exercise, but unlike last time we stream these words onto HDFS.
- StreamWordCount, that takes the stream and puts it into a DynamoDB.

To run the program 1. Compile locally with 'sbt assembly' 2. Move the jar file to the cluster 3. Execute the program as follows:

3. 1. Execute a consumer that write the output to a DynamoDB instance, if started correctly it will never stop.

```
spark-submit \
--master yarn \
--deploy-mode cluster \
--class dk.itu.BIDMT.Exercises.Exercise13.StreamWordCount \
Exercise13-assembly-0.1.jar \
data us-east-1 testDB
```

4. 2. Execute a producer, in this the last line is the arguments, you can change these.

```
spark-submit \
--master yarn \
--deploy-mode cluster \
--class dk.itu.BIDMT.Exercises.Exercise13.WordStreamProducer \
Exercise13-assembly-0.1.jar \
data 100 100 10 1000
```

## To stop the consumer

In the console that you started the consumer, you can exit your client program using ctrl+c. It is also possible to login on another console and do the same commands.

1. Get the application list of current applications running on spark.

   ```
   yarn application -list
   ```

The result of this looks something like:

```
Total number of applications (application-types: [] and states: [SUBMITTED, ACCEPTED, RUNNIN
                Application-Id      Application-Name        Application-Type          User
application_1574844428287_0003  dk.itu.BIDMT.Exercises.Exercise13.StreamWordCount
```

2. To stop the application you send the Kill command using yarn as follows:

   ```
   yarn application -kill <applicationID>
   ```

In my case the command ended up being:

```
yarn application -kill application_1574844428287_0003
```