**IDC Big Data Assignment 3: DynamoDB (NoSQL)**

In this assignment you are required to create a DynamoDB database and implement CRUD operations (Create, Read, Update, Delete).

**Details**:

1. Before doing the assignment you should learn how to work with DynamoDB with Python and the Boto3 library. Boto3 is the AWS Software Development Kit (SDK) for Python. It is highly recommended that you read the AWS Guide [Getting Started Developing with Python and DynamoDB](https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GettingStarted.Python.html) and follow all the steps (1-4) and also the extra step at the end about [Improving Data Access with Secondary Indexes](https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/SecondaryIndexes.html).
2. The input file **nyt2.json** is based on the New York Times best Sellers list, and is located in Moodle.
3. See **Ex3 Load Data - Jupyter Notebook** in Moodle for how to read the json data into a DynamoDB table

**Assignment**

1. Copy the data file to an EC2 instance, e.g. using wget.
2. Write Python code in a Jupyter notebook to do the following:
   1. Create a dynamodb table. **Pay attention to the keys and attributes you define**! Make sure the values you assign to ReadCapacityUnits and WriteCapacityUnits will impact the time it will take to perform the following operations (and also your bill!). The default for both is 10. **Do not go over 100**!
   2. Read the data into the table. Use the JSON loads function to load the entire file. Also make sure to explicitly give a type to each value (either str or int).
   3. Show all the items where the author is John Sandford. (107 items)
   4. Show the same items with only title, author and publication date, one item per line.
   5. Show books titles written by Alex Berenson with their price and rank, if the title begins with "THE" (case sensitive). Note that rank is a reserved word. (14 items)
   6. Print the number of titles where the title begins with "THE" (case sensitive). (3524 items)
   7. Print the number of titles if this is the 1st week the book is on the list and its rank is in top 15 (1954 items). The output should include the Count, ScannedCount, ConsumedCapacity (add ReturnConsumedCapacity = 'INDEXES') and also the running time (use %%time).
   8. Create a global index (GSI) on weeks\_on\_list and rank.
   9. Repeat the last query (count titles where this is the 1st week the book is on the list and its rank is in top 15). Again the output should include the Count, ScannedCount, ConsumedCapacity and running time.
   10. Update the price of the book “BELIEVING THE LIE” written by Elizabeth George (from 28) to 15.
   11. Remove the publisher attribute from the book “BELIEVING THE LIE” written by Elizabeth George.
   12. Delete one book (item) named “THIS BODY OF DEATH “ and written by Elizabeth George.
3. When you finish the assignment, submit your notebook to Moodle.
4. **Delete** all [the DynamoDB table(s) and indexes you created](https://console.aws.amazon.com/dynamodb/home?region=us-east-1#tables:)! Otherwise you will keep paying for them.

**Tips**

1. You should create an EC2 instance on AWS cloud and work with Jupyter lab on it.