

### What you need to do:

The objective of the assignment is to test your proficiency in handling data and Python coding skills. The assignment has 3 parts.

- Section 1: Install MySQL database and load Northwind data sets.
- Section 2: Analyse the relationships in the Northwind data sets using an ER diagram and answer some business questions about the data by writing SQL statements.
- Section 3: Analyse the Stackoverflow data set and answer the business questions using Python Pandas library.

### Section 1: Environment setup and loading Northwind data.

1. Install MySQL (server and client) on your laptop:
2. Load the Northwind dataset into the cluster
  - a. Northwind database is a set of data sets that is shipped with Microsoft Access and is used in learning SQL. There are several resources available online to learn more about Northwind. (<https://theaccessbuddy.wordpress.com/2011/07/03/northwind-database-explained/>)
  - b. We have sampled down the files and create a zipped file. You can download the files from [here](#).
  - c. The ER diagram is also present along with the files. Please review the data model and understand the relationship between the entities.
  - d. Create DDL statements using the ERD diagram and create those files in MySQL. Load the files into those tables. The table names, column names and data types should match with what is provided. If your schema doesn't match with ours, we will not be able to test your code.
3. Deliverables:
  - a. DDL statements
  - b. Loading scripts for populating data into the tables

### Section 2: Working with Northwind data and SQL.

- a. For this task you must write the SQL statements that will provide the necessary results.
  - i. Write a query to get Product list (name, unit price) of ten most expensive products.
  - ii. Number of units in stock by Category(Category Name) and supplier country?
  - iii. Find out the quartely orders for each product for the year 1997?
  - iv. Give the name of employees and the city where they live for employees who have sold to customers in the same city?
  - v. Give the names of employees who are strictly older than any employee who lives in London.
- b. Please refer to the ER diagram provided in the zipped files to understand the relationship between the tables. Pay attention when to use outer joins vs. inner joins.
- c. Deliverables:
  - i. Queries / SQL statements