

Moez Skanji. Janvier 2021	SPARK DATAFRAME Walmart Stock	URCA Master SEP, CS
------------------------------	----------------------------------	------------------------

## **Exercice du projet d'analyse des données Spark Walmart**

Let's get some quick practice with your new Spark DataFrame skills, you will be asked some basic questions about some stock market data, in this case Walmart Stock from the years 2012-2017. This exercise will just ask a bunch of questions, unlike the future machine learning exercises, which will be a little looser and be in the form of "Consulting Projects", but more on that later! For now, just answer the questions and complete the tasks below.

- 1- Start a simple Spark Session
- 2-Load the Walmart Stock CSV File, have Spark infer the data types.
- 3-What are the column names?
- 4-What does the Schema look like?
- 5-Create a new dataframe with a column called HV\_Ratio that is the ratio of the High Price versus volume of stock traded for a day
- 6-What day had the Peak High in Price?
- 7-What is the mean of the Close column?
- 8-What is the max and min of the Volume column?
- 9-How many days was the Close lower than 60 dollars?
- 10-What percentage of the time was the High greater than 80 dollars ?(In other words, (Number of Days High>80)/(Total Days in the dataset))
- 11-What is the max High per year?
- 12-What is the average Close for each Calendar Month? In other words, across all the years, what is the average Close price for Jan, Feb, Mar, etc... Your result will have a value for each of these months.