# Big Data Analytics

**ESSEC**

M. Ndaoud

Introduction Exercises

1. **Install Spark:** you can easily find tutorial on-line
   - e.g. for Ubuntu: https://otodiginet.com/software/how-to-install-apache-spark-on-ubuntu-20-04-lts/
   - If you use Python
     - You can install PySpark
     - Take a look on a tutorial how to use Spark with Jupyter
       * e.g. `https://www.sicara.fr/blog-technique/2017-05-02-get-started-pyspark-jupy`
       * `https://github.com/jadianes/spark-py-notebooks`
2. **Hash Functions**. Suppose hash-keys are drawn from the population of all non-negative integers that are multiples of some constant $c$, and hash function $h(x)$ is $x$ mod 15. For what values of $c$ will $h$ be a suitable hash function, i.e., a large random choice of hash-keys will be divided roughly equally into buckets?
3. **The Base of Natural Logarithms**.
   (a) In terms of $e$, give approximations to
   - $(1.01)^{500}$
   - $(1.05)^{1000}$
   - $(0.9)^{40}$

   (b) Use the Taylor expansion of $e^x$ to compute, to three decimal places:
   - $e^{1/10}$
   - $e^{-1/10}$
   - $e^2$