**License Info : -**

This project is licensed under the terms of the Apache License 2.0 as <https://tinyurl.com/y85g3ct7>

**Project Scope**

1. It could be used to anticipate the performance of Udemy application for providing online courses to a huge sector in industry.
2. It is valid for descriptive analysis and sneak peak of prescriptive analysis.
3. It is valid for particular set of compatible attributes.

**Tools and Working Environment :**

Project is made on Cloudera in VMware(Centos v6.5). It includes data manipulation and libraries such as :

1. **Linux Commands**
2. **Hadoop**
3. **Sqoop**
4. **MySql**
5. **Hive**
6. **Pig**

**Queries**

1. Which courses are available at affordable prices(price<200)?
2. Which courses are the best rated courses(subscribers>5000)?
3. What is the highest price for a course(max(price))?
4. What is the average price from all paid courses?
5. How many course entries are there in present data?
6. Which courses are available at free of cost(ispaid=FALSE)?
7. Group the Udemy courses by each levels?
8. How many courses are there in each level(All,Beginner,Intermediate,Expert)?
9. Select courses according to ascending order of lectures?
10. Partition the courses by subject?
11. Filter the Udemy Courses acc. to lectures (lectures = 26)?
12. Group the Udemy Courses having same price?
13. For each courses generate specified data-(id,coursetitle,lectures)?

**Big data project implemented by Aniket**