Project:
- Team Number: 10
- Project Title: Document Search Engine
- Project Description: 'Document Search Engine' is a platform which aims to return relevant documents for a given input query. It is a distributed data-intensive computing platform which quickly returns related documents from the dataset while balancing the load at the back end. The platform also updates the dataset in real time.
- Team members:
    ○ Deepak Gupta - dgupta22
    ○ Laxmikant Kishor Mokadam - lmokada
    ○ Rohan Girish Chandavarkar - rgchanda
	Deliverables:
	1. A docker image which contains all the dependencies of the system and will automatically setup distributed system with 8 VCL images with single command.
	2. Source code
	3. Sample data set
	4. Readme explaining how to use the project.

Dependencies:
1. 8 VCL image (for 3 months)
2. AWS/Google cloud for testing/demo
3. Apache Spark
4. Python
5. Apache Streaming
6. OpenCV
7. Tensorflow
8. Wikipedia Dataset
9. NIPS paper dataset
10. Hacker News Dataset
	Issues:
	1. Infrastructure Issues:
	    a. Interconnection between VCL images for distributed network setup
	    b. VCL image timeout problem
	2. Software issues: Compatibility between different softwares/libraries and their versions