Project Setup GCP

```
Follow below steps if you want to use Hadoop and pull the data using Jupyter notebook.
```

Instance Setup

- Region : Australia-southeast1 (Sydeny)
- Default Zone
- MAchine Configuration : N1
- Machine-type : N1-Standard-2
- Boot Disk : OS(Ubuntu), Size(50GB)
- Allow : HTTP, HTTPS trafic
- Create the instance

Steps for SSH shell

- sudo apt-get update
- sudo apt install docker.io
- sudo apt install docker-compose

Follow Practical 8 : Advance RDD programming

- mkdir $HOME/prac8 && cd $HOME/prac8
- curl -L -o MLlib.zip https://www.dropbox.com/s/388xpkjkcv5bwyv/MLlib.zip?dl=0
- sudo apt install unzip
- unzip MLlib.zip
- sudo docker-compose -f docker-compose_hdfs_spark.yml up -d
- Go to firewall > allow default http > expose TCP port : 80,8000,8080,9000,8082,8888,4040
- Upload csv file to using ssh command line

AFTER UPLOADING :

- sudo docker ps
- mv loan_defaulters.parquet.gzip /home/prajwalgowda2101997/prac8/nbs
- sudo docker exec -it container_name(this is just an example :3238129fb9ec) bash
    - hdfs dfs -put /home/nbs/* /
- exit

- cd prac8/
- Giving permission: sudo chmod -R 777 nbs/