# YAH: YET ANOTHER HADOOP

*Team Members:*
BHAVANA R: PES2UG19CS089
DEVIKA S P: PES2UG19CS109
ESHA H: PES2UG19CS122
KRUTHIKA HEBBAR K K: PES2UG19CS194

*Design Details*:

We have implemented a system that mimic the Hadoop.

*Brief information about implementation details of each unit:*

Datanodes:

Each datanode is implemented using directories and the file inputted by the user is split into blocks and stored into each of these blocks based on the hashing function.

Namenodes:

Namenodes is implemented using a namenode.py file which is always running in the background. If any of the blocks in the datanodes are deleted due to some error then the namenode makes sure to replicate the deleted namenodes.

Datanode log:

The datanode log will just have information of when the datanode is updated. Each datanode has its own datanode log file

Namenode log:

The namenode log file stores information about each datanode in a dictionary format. The dictionary contains the datanode number as the key and the values are the number of blocks in each datanode along with block names.

Filesystem:

The filesystem is implemented using a directory as well and every new folder and file created is stored in the filesystem directory.

*Reason behind design decisions:*

We chose this particular design because a distributed file system is a very efficient way to manage and store large files and is an easy way to run different operations on these files using the concept of Hadoop.

*Take Away from the project*:

We learnt how to work with namenode and datadone by creating directory.

But with this project helps us how to distribute files to each datanode, how to run the namenode in the background and checking for any failures in our datanodes.

*Thank you*