

**Team Name:** BD2\_039\_134\_247

**Team Members:**

1. Ananya M Gupta PES2UG19CS039
2. G P Prerna PES2UG19CS134
3. Namitha Nayak PES2UG19CS247

**Project title:** Yet another Hadoop

**Design Details:** Our project Yet Another Hadoop is like a mini-HDFS setup on our system with the architectural structure consisting of data node and name node and replication of data across nodes.

**Surface level implementation about each unit:**

Using this, we possess finer control over data blocks and their replication. It allows us to create a distributed file system as per our needs. It also allows us to run Hadoop, for example, jobs to breakdown computationally heavy tasks into smaller distributed tasks.

The units are:

1. DFS Setup config
2. FS Path
3. Data Nodes
4. Block size replication factor
5. Name Nodes
6. Sync period
7. Mapper and reducer

**Reason behind design decisions:** Our code for the project is implemented in the python programming language. We chose this because:

1. High-level programming language, easy to learn and write
2. Dynamically typed
3. Free and open source
4. Vast libraries support
5. Portability
6. Interpreted Language

**Take away from the project:**

This project helped us with the in-depth understanding of Hadoop and its underlying architecture. It helped us observe its inner workings.

Helped us understand the commands better and how to schedule Hadoop jobs.