# CCBD Assignment
# Task 1 and 3

Abhishek Das
PES1201800177

Bhargav SNV
PES1201800308

N Sanketh Reddy
PES1201800389

Bengaluru's Greenest Part

# Task3

# Dataset Creation

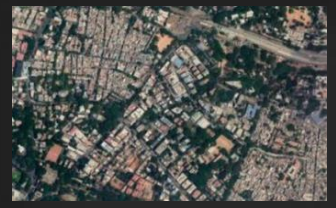
1. Image Capture
2. Image Cropping
3. Image Splitting

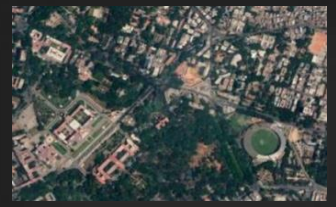
- Google Maps was used to get images.
- Images were captured for each pin code in the City. (Screenshots were taken)
- The screenshots were then processed to give only the required area of the map.
- Each of the resulting images were further cropped on basis of locality. E.g.: Central, North, South-West, etc.

# Image Processing Workflow

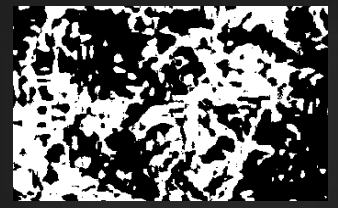

Screenshot

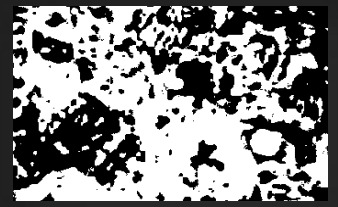
Cropped Image

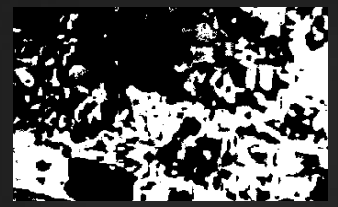




Cropped Sector



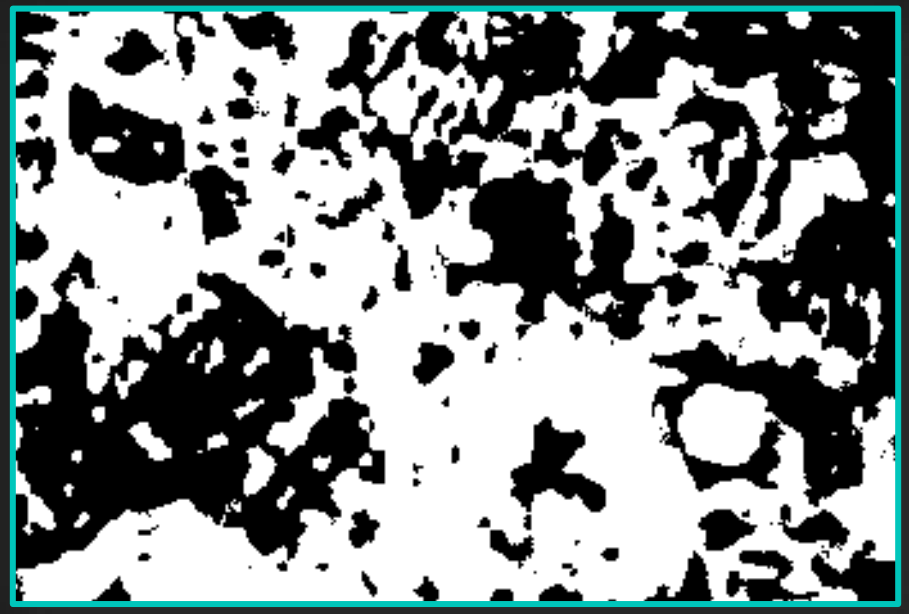




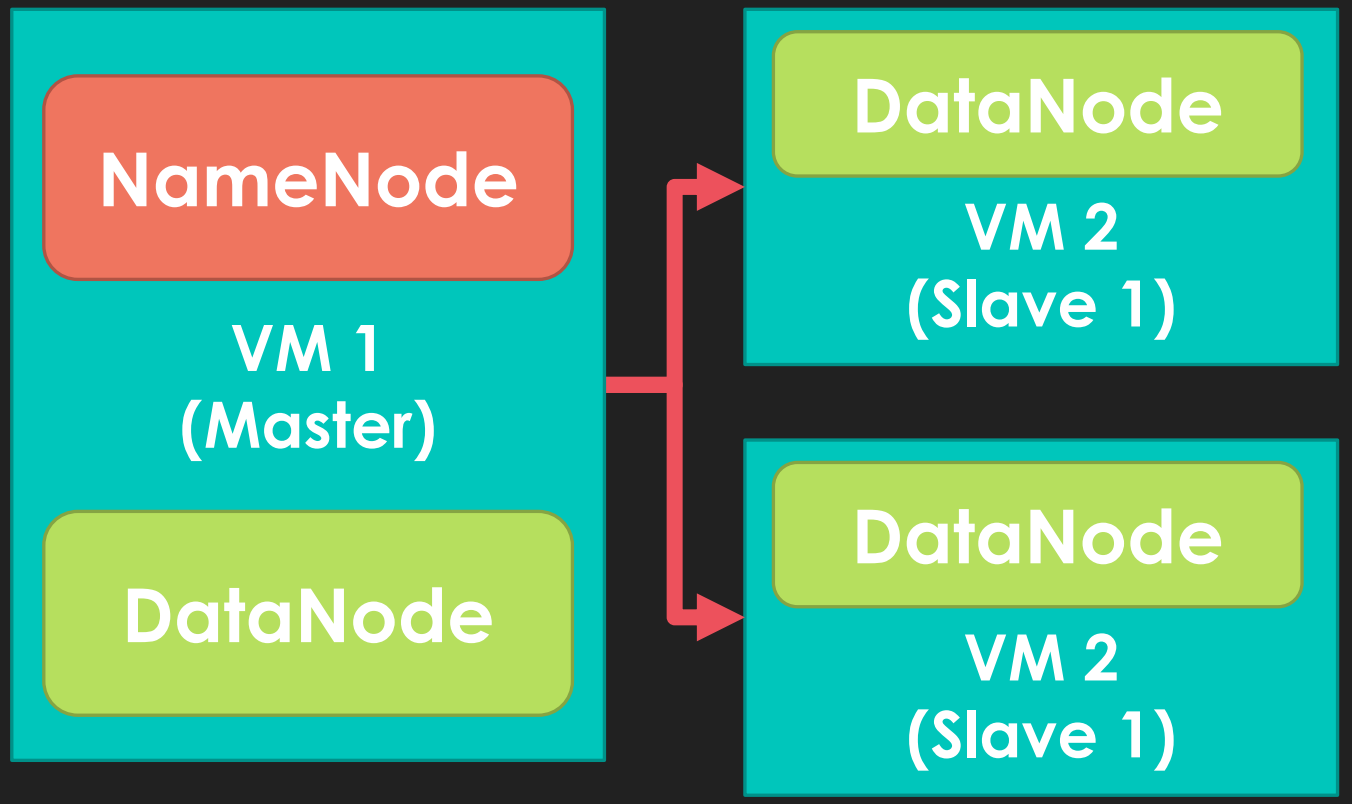
Sector Mask

# Image Processing Implementation

- A green mask is acquired for each sector using a threshold HSV value (Specific to green).
- This mask results in a black and white image where all green pixels are white and pixels of any other colour are black.
- The percentage of green pixels is found and stored into a file.



# Hadoop Cluster setup and MapReduce




- A 3-node Hadoop Cluster was setup using Vagrant. (3 Virtual Machines running Hadoop)
- MapReduce was run over the image processing output.

# Results

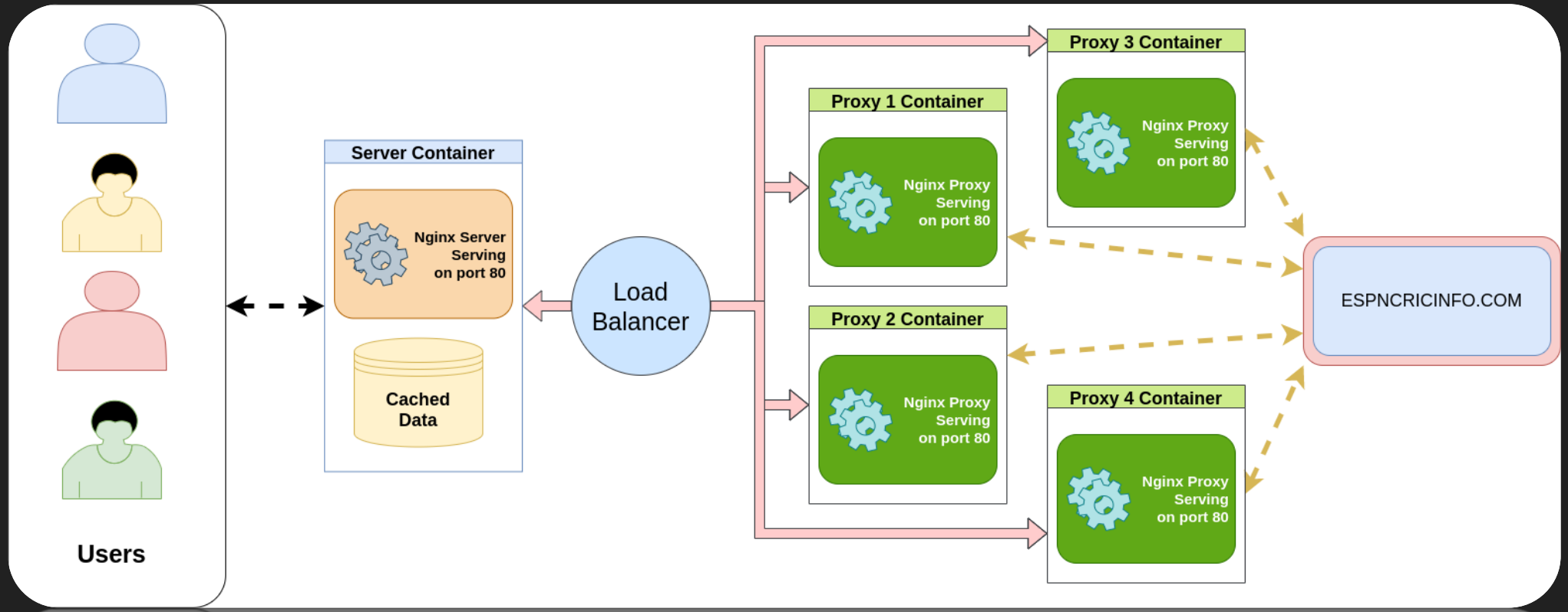
```
Task-3 ❯ cat output.txt                    master
"560004-SW"     60.21
"560006-NW"     70.84
"560014-E"      60.23
"560017-E"      70.02
"560033-C"      62.18
"560054-N"      60.86
"560066-E"      61.62
"560066-S"      65.45
"560083-E"      68.62
"560094-NW"     67.48
"560094-SW"     79.66
Task-3 ❯                                   master
```

- The picture lists places above Bangalore that have above 60% Greenery.
- Only one of these places have greenery above 75%.

Who will share my Load?

# Task1

# Network Layout



# Server setups

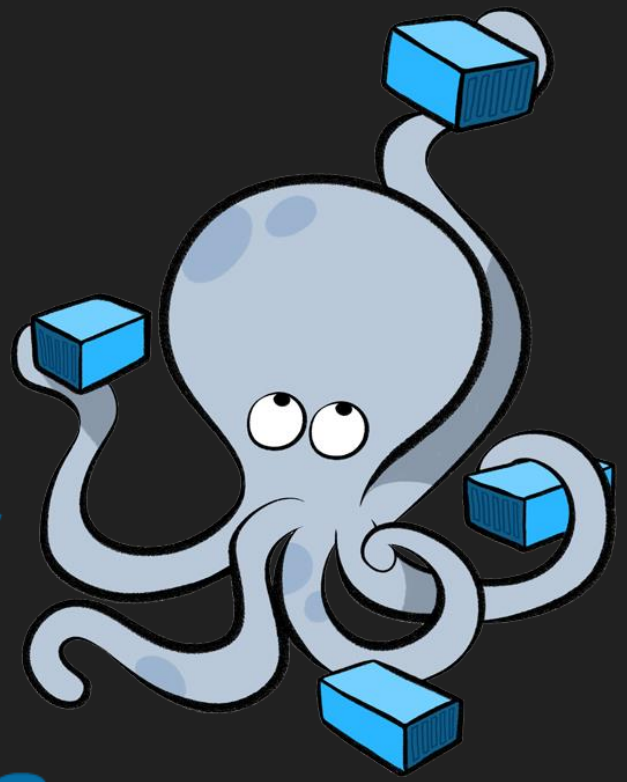
Web Server

- This is a container running an NGINX Web Server.
- It takes requests from users and forwards them to the Proxy Servers. These requests are cached and returned on subsequent requests.
- It Handles load balancing of requests to the proxy server using round robin algorithm

Proxy Server

- This server takes makes requests to espncricinfo.com and returns it to the main server.
- It's sole purpose is to make requests to espncricinfo.com.

# Server Deployment



- Each server (Web and Proxy) is deployed within a container.
- Containers are brought up using docker and docker compose.
- If needed, more proxy servers can be deployed to maintain heavy loads.

# Thank You