CCBD-Assignment

Directory Structure

- `images/` contains screenshots of satellite images from google maps. It contains images of each pincode.
- `generated/` contains cropped variants of each image in `images/`.
- `sector-images/` contains images of sectors (North, North-East, Central, etc.) of each pincode.
- `sector-masks/` contains masks for each sector in `sector-images/`.
- `masks/` contains masks for each image in `generated/`.

Files

- `crop.py` crops images from the `images/` directory and places cropped images in `generated/` directory.
- `percentage.py` this generates the sector images, masks and percentages of green from each sector. The sector images are placed in `sector-images/`, the sector masks in `sector-masks/` and the list of percentages of each sector is written into `data.txt`.
- `masks.py` this generates mask for each image in the `generated/` directory and places them in the `masks/` directory.
- `run.sh` this is a simple shell script to run all the python scripts in the correct order and generate the final `data.txt`
- `mapreduce.py` a mapreduce program to list out the areas with percentage of green above a certain value. This output is saved in `oytput.txt`
- `data.txt` contains comma seperated values of sectors with corresponding green percentage.
- `output.txt` contains output of mapreduce program.

How to execute?

- Execute `run.sh` from a terminal. This will give `data.txt` if everything executes correctly.
- Execute `mapreduce.py` from a hadoop cluster with a hadoop streaming file. Use `data.txt` as input.

Dependencies

- Python 3 or higher with following dependencies:
    - OpenCV
    - Numpy
    - OS
    - Math
    - csv
    - copy

- Hadoop cluster with appropriate Hadoop-Streaming.jar and MRJob installed.