License plate detection system with Hadoop

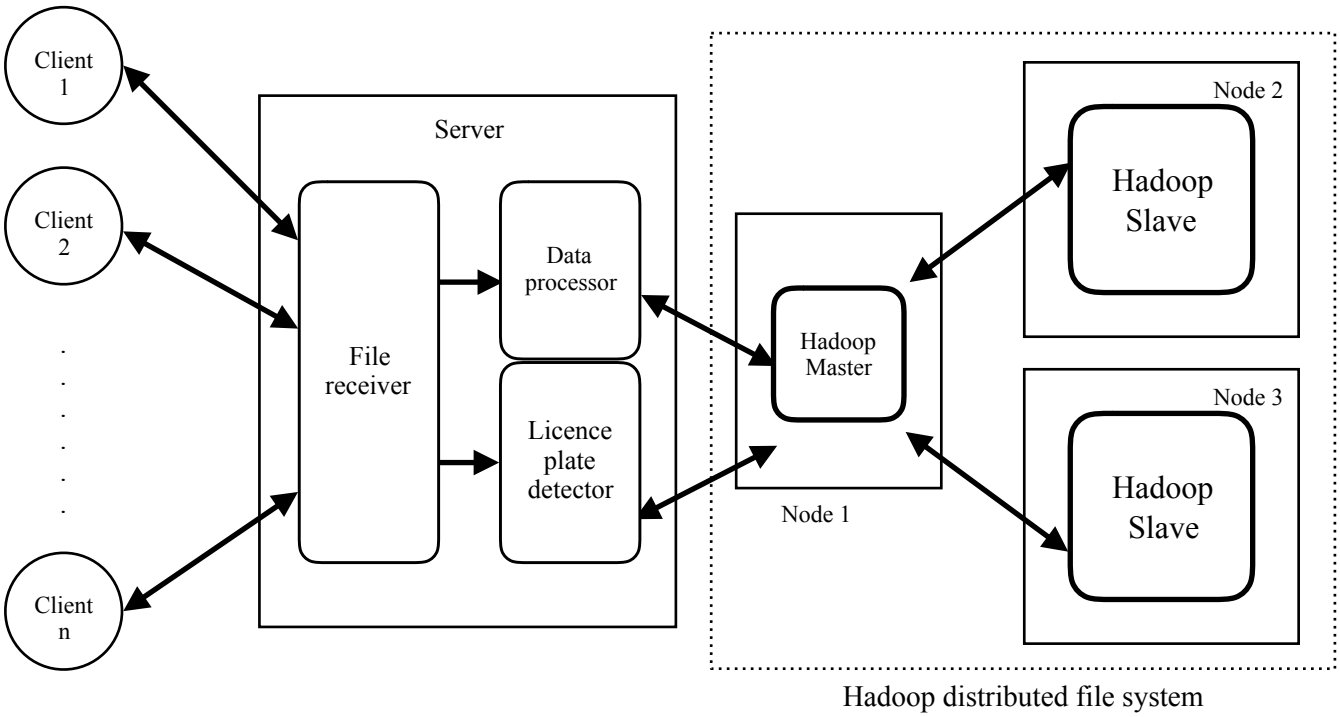
Overview

Apache Hadoop is a collection of open-source software utilities that facilitate using a network of many computers to solve problems involving massive amounts of data and computation.

In this homework, we are going to setup a Hadoop distributed file system with a real time server to handle the multiple data streaming.

Specification

- Components of server:
  - **File receiver**
    - Receive video files and names from multiple clients at the same time.
    - The maximum number of the clients is 25.
    - The size of the file is about 200MB.
  - **Data processor**
    - Save the data to the HDFS while file receiver receiving the data.
  - **Licence plate detector**
    - Check if there is any license plate in every frame received from the clients.
    - Scan every frame from every file, if there is a license plate in this frame, record the number of this frame.
    - Each video will have a file to record the number of those frame with license plate.
    - The file name should be <video name>.txt and the format of the file will be provided in the attach file.
    - These files should also save in HDFS.
- Component of HDFS:
  - Your HDFS will have one name node(master) and two data node(slave).
  - The specification of the instance is limited. When launch the instances,
    - In step 2, please select the “Free tier eligible” one
      - Family : General purpose
      - Type : t2.micro
      - vCPUs : 1
      - Memory(GiB) : 1
      - Instance storage(GB) : EBS only
    - In step 4, please select
      - Size(GiB) : 30
      - Volume type : General Purpose SSD (gp2)



Grades

During the demo, you need to create 25 clients and send the camera videos to your server.

- Basic
  - If the accomplish all the requirements, you will get 50**%**.
- Efficiency
  - All of the students will be divided to five groups by the execution time.
  - The fastest **10** students will get **25%**.
  - 11th to 20th will get **20%**.
  - 21th to 30th will get **15%**.
  - 31th to 40th will get **10%**.
  - 41th to 50% will get **5%**.
- Accuracy
  - All of the students will be divided to five groups by the error rate.
  - The lowest **10** students will get **25%**.
  - 11th to 20th will get **20%**.
  - 21th to 30th will get **15%**.
  - 31th to 40th will get **10%**.
  - 41th to 50% will get **5%**.

P.S. if clients do not send the files at the same time, **you will not get any score**.

Note

- We have no limitation on the programming language.
- Server, clients and Hadoop distributed file system should setup on AWS instances.

- Server and Hadoop master node should run on different AWS instances.
- The server should be terminated after it finish all the works.
- During the demo, TA will use **our client** to send the video file.