Implementing Billing Services in Distributed Systems with Hadoop

CS230

Julius Aguma
Mowhebat Bazargani
Elaine Thai

Project Introduction (Motivation + Goals)

- In our project, we are trying to simulate the billing services in a distributed system.
- Our goal is to split the workload between different machines in the Hadoop cluster to improve the performance and availability of the billing services when we have a high load.

Project Design and Implementation Specifics

Server:

- Runs on multi node cluster
- Uses sockets and ssh for communication
- Has a local master, local slave and aws cloud node
- Uses opencsv library for manipulating csv files

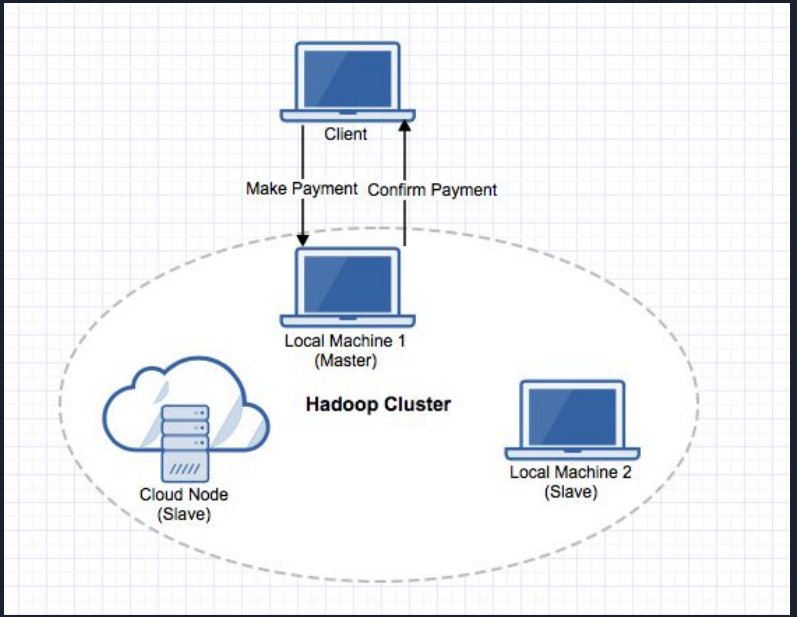
Client:

- Runs on a local machine
- Randomly selects test student
- Uses sockets to communicate with server

Project Architecture

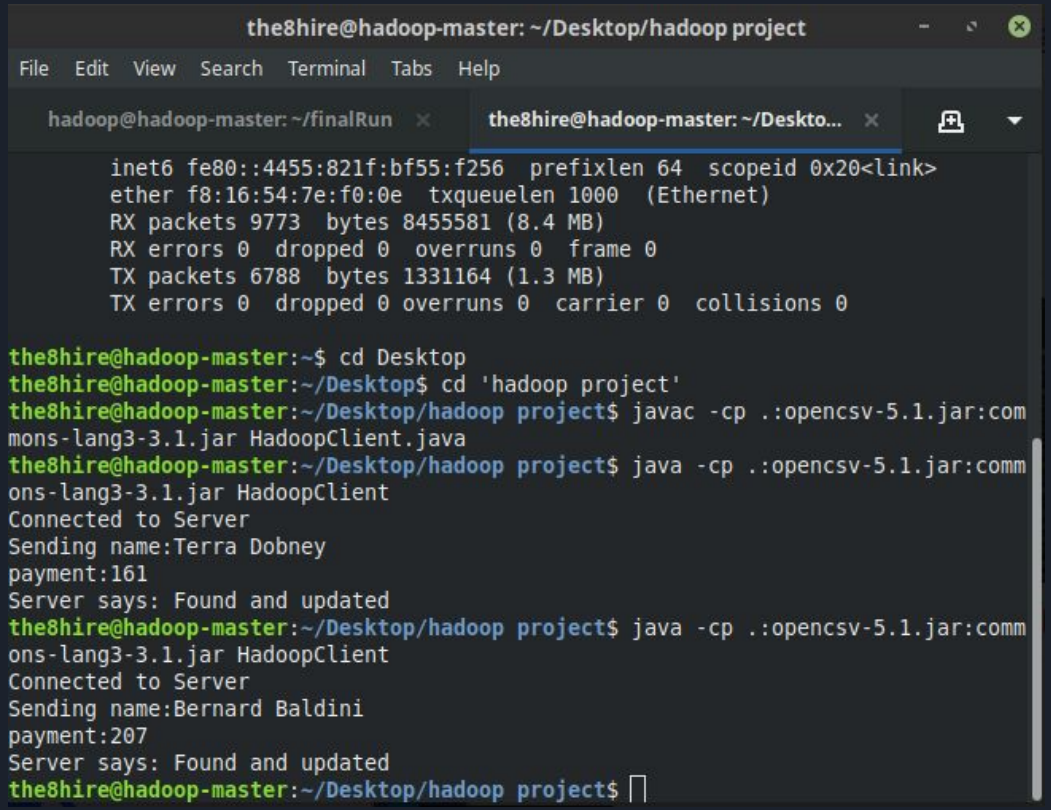
We will have one Hadoop cluster with one master node and two slave nodes. One local machine will act as the master while another and a cloud node will act as slaves.

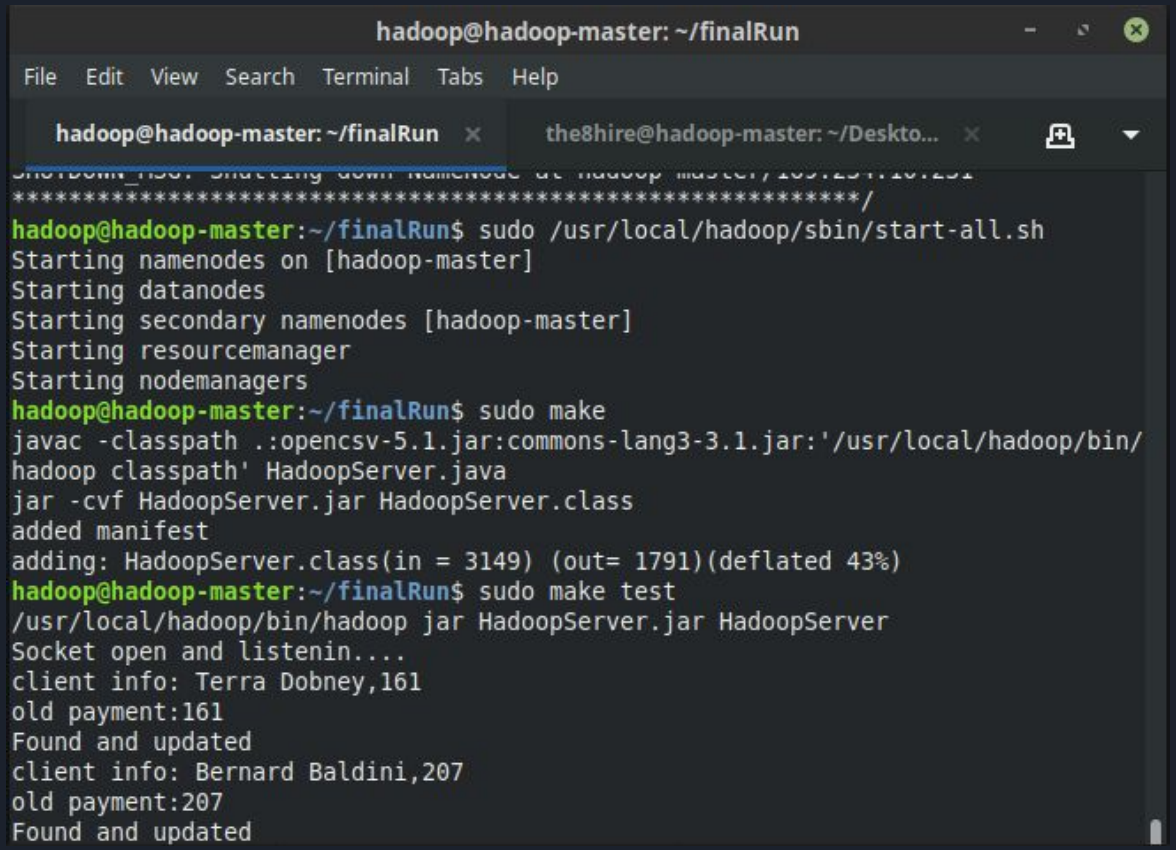
The database is stored in the master node of the Hadoop cluster, including students' name and the amount of money they owe. When requests are received from the client, the Hadoop cluster splits the works between the nodes.



Project Testing & Evaluation

Running client code:





Running server code

Contribution

Elaine Thai:
Elaine implemented early cloud cluster instances, wrote up the project proposal, worked on the project slides and wrote up the first draft of the project report. We intended to run the client on Elaine's local machine but firewall permissions prevented that.

Mowhebat Bazargani:

Mowhebat implemented one of the local slave nodes, worked on the project slides and contributed to the project proposal and project report. We intended to run the hadoop slave 1 on Mowhebat's local machine but firewalls and network filter blocks prevented that.

Julius Ceasar Aguma:
Ceasar was responsible for setting up the final cloud instance, wrote the server code and final draft of the project report. We intended to use Ceasar's laptop as the hadoop master but ended up having to run both server and client on it due to obstacles with multi-machine connection using ip addresses.

Conclusion

- Limitations:

  Firewalls, and IP blocking

  Long configuration process