**CSC 555 Assignment 1**

**Yiyang Yang**

**Part 1**

a)

212 = 26 \* 26 = 64 \* 64 = 4096

45 = 16 \* 64 = 1024

45 = 16 \* 64 = 1024

12 \* 101 + 9 = 1221, then 1221 MOD 12 = 9

13 \* 4 + 9 = 61, then 61 MOD 13 = 9

7 \* 6 + 1 = 43, then 43 MOD 7 = 1

b)

V1 + V2 = (3, 2, 5)

V1 – V2 = (1, 0, 1)

|V1| = sqrt(22 + 12 + 32) = sqrt(14) = 3.7

|V2| = sqrt(12 + 12 + 22) = sqrt(6) = 2.4

M \* V1 = =

M \* M = =

M4 = M2 \* M2 = =

c)

P(HHT) = 0.6 \* 0.6 \* 0.4 = 0.144

P(THHT) = 0.4 \* 0.6 \* 0.6 \* 0.4 = 0.0576

P = \* 0.62 \* 0.4 = 0.432

P = \* 0.62 \* 0.4 = 0.432

d)

i.

SELECT Name

FROM Employee

WHERE Name LIKE ‘Jane%’;

ii.

SELECT COUNT(DISTINCT CertName)

AS NumberOfCertName

FROM Certificates;

iii.

SELECT ID, Name, COUNT(DISTINCT CertName)

FROM Employee e, Certificates c

WHERE e.ID = c.EID

GROUP BY Name

HAVING COUNT(DISTINCT CertName) BETWEEN 0 AND 4;

e).

I think since the function of hash is h(x) = x mod 15, which means h(x) can provide 15 buckets numbered from 0 to 14. Co-prime numbers of 15 should be good for c, such as 1, 2, 4, 7, 8, 11, 13, 14 and so on.

f).

i.

Mapper: Filename is the key, number is the value

Reducer: Use the key and associated value to sort, then output the smallest number.

ii.

Mapper: All the number between 100 and 1000 are the key, and its value is the value

Reducer: Use the key and value to compare, then get the unique numbers.

iii.

Mapper: ID is the key, First and Last are the values

Reducer: Use the values to compare, then get the amount of the same name students.

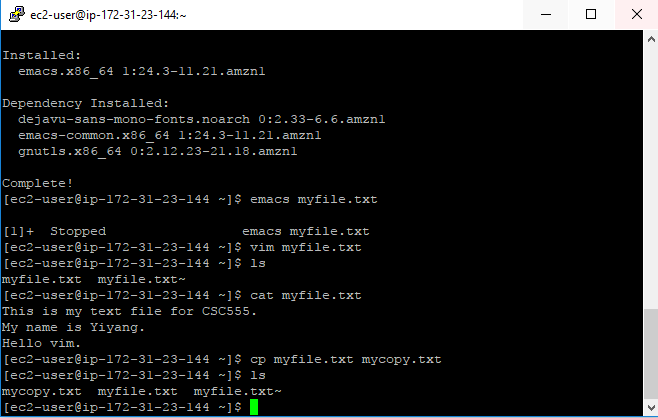
iv.

Mapper: ID is the key, grade is the value

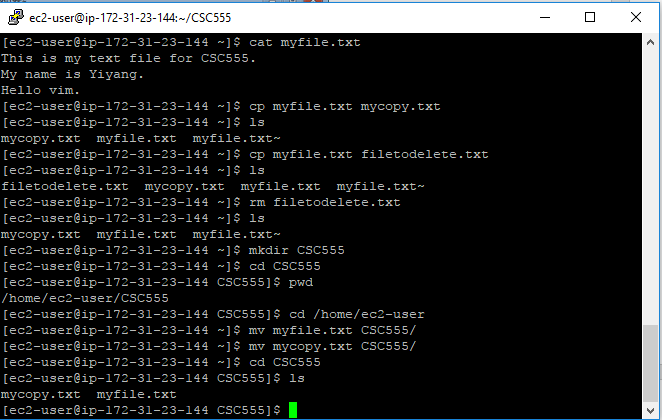
Reducer: Use the grade to calculate GPA.

**Part 2**

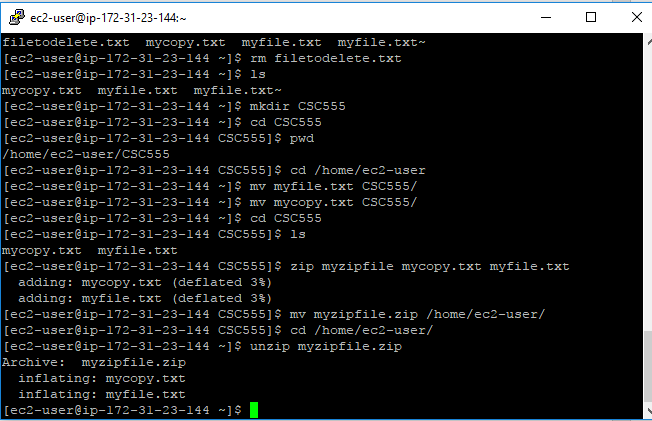
2. Copy file.



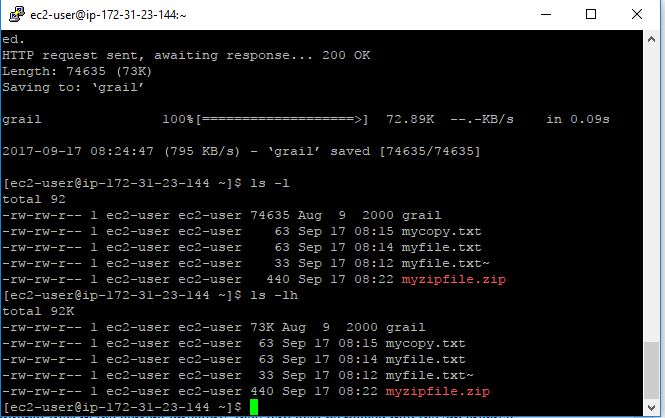
5. Move



6. Zip and unzip



9. ls format

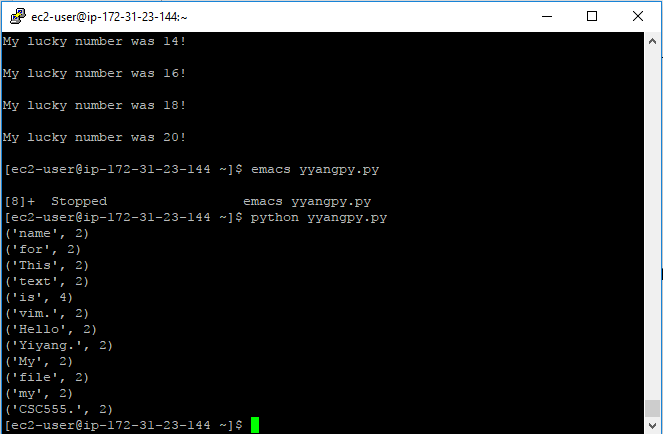


11.Change access



12. Python

Output:

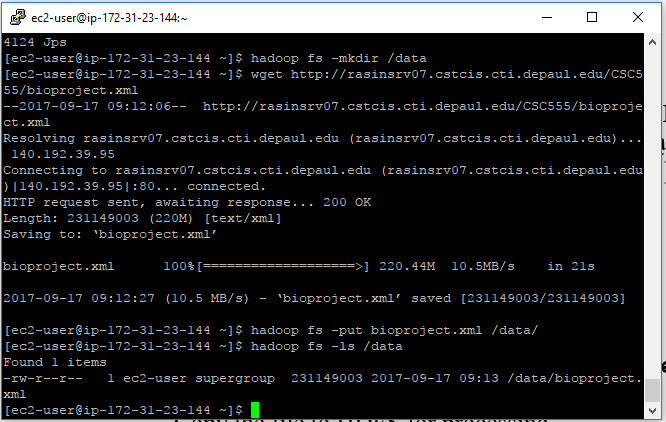


Source Code:



**Part 3**

Move bioproject.xml to Hadoop



Runtime

