

COMP9313 2017s2 Assignment

z5045582

Yunhe Zhang

Question 1. MapReduce

```
class mapper {
    map (key, record) {
        split_record  $\Leftarrow$  split record with '\s+'
        product  $\Leftarrow$  split_record[1]
        price  $\Leftarrow$  split_record[2]
        EMIT(product, price)
    }
}

class reducer {
    HashMap< String, Integer> myMap
    reducer (key, values) {
        for all value  $\in$  values
            myMap.put(key, record)
            break
    }
    cleanup () {
        Map sortedMap  $\Leftarrow$  sort by values in decreasing order (myMap)
        counter  $\Leftarrow$  0
        for all entry  $\in$  sortedMap {
            if counter == 5
                break
            counter++
            EMIT(null, entry's key)
        }
    }
}
```

Question 2. MinHash

Row	C_1	C_2
0	0	1
1	1	0
2	0	1
3	0	0
4	1	1
5	1	1
6	1	0

$$h1(n) = (3n + 2) \bmod 7$$

$$h2(n) = (2n - 1) \bmod 7$$

	$Sig1$	$Sig2$	
	∞	∞	
$h1(0) = 2$	∞	2	update Sig2
$h2(0) = 6$	∞	6	update Sig2
$h1(1) = 5$	5	2	update Sig1
$h2(1) = 1$	1	6	update Sig1
$h1(2) = 1$	5	1	update Sig2
$h2(2) = 3$	1	3	update Sig2
$h1(3) = 4$	5	1	no change
$h2(3) = 5$	1	3	no change
$h1(4) = 0$	0	0	update Sig1 Sig2
$h2(4) = 0$	0	0	update Sig1 Sig2
$h1(5) = 3$	0	0	no change
$h2(5) = 2$	0	0	no change
$h1(6) = 6$	0	0	no change
$h2(6) = 4$	0	0	no change

Result

	$Sig1$	$Sig2$
$h1(n)$	0	0
$h2(n)$	0	0

Question 3. Streaming Data

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (1, 197) (1, 200)

Input from 200 to 210: 0101010101

201 Input: 0

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (1, 197) (1, 200)

202 Input: 1

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (1, 197) (1, 200) (1, 202)

Combine (1, 197) (1, 200)

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (2, 200) (1, 202)

203 Input: 0

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (2, 200) (1, 202)

204 Input: 1

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (2, 200) (1, 202) (1, 204)

205 Input: 0

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (2, 200) (1, 202) (1, 204)

206 Input: 1

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (2, 200) (1, 202) (1, 204) (1, 206)

Combine (1, 202) (1, 204)

(16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (2, 200) (2, 204) (1, 206)

Combine (2, 192) (2, 200)

(16, 148) (8, 162) (8, 177) (4, 183) (4, 200) (2, 204) (1, 206)

207 Input: 0

(16, 148) (8, 162) (8, 177) (4, 183) (4, 200) (2, 204) (1, 206)

208: Input: 1

(16, 148) (8, 162) (8, 177) (4, 183) (4, 200) (2, 204) (1, 206) (1, 208)

209: Input: 0

(16, 148) (8, 162) (8, 177) (4, 183) (4, 200) (2, 204) (1, 206) (1, 208)

210 Input: 1

(16, 148) (8, 162) (8, 177) (4, 183) (4, 200) (2, 204) (1, 206) (1, 208) (1, 210)

Combine (1, 206) (1, 208)
(16, 148) (8, 162) (8, 177) (4, 183) (4, 200) (2, 204) (2, 208) (1, 210)

The result is
(16, 148) (8, 162) (8, 177) (4, 183) (4, 200) (2, 204) (2, 208) (1, 210)

Question 4. Collaborative Filtering

(a)

	m1	m2	m3
u1	2		3
u2	5	2	
u3	3	3	1
u4		2	2

$$\mathbf{sim}(x, y) = \frac{\sum_i \mathbf{r}_{xi} \cdot \mathbf{r}_{yi}}{\sqrt{\sum_i \mathbf{r}_{xi}^2} \cdot \sqrt{\sum_i \mathbf{r}_{yi}^2}}$$

$$\text{sim}(u1, u2) = 0.515$$

$$\text{sim}(u1, u3) = 0.573$$

$$\text{sim}(u1, u4) = 0.588$$

$$\begin{aligned} \text{Predict } u1 \text{ to } m2 &= (0.515 \cdot 2 + 0.573 \cdot 3 + 0.588 \cdot 2) / (0.515 + 0.573 + 0.588) \\ &= 2.34 \end{aligned}$$

(b)

	u1	u2	u3	u4
m1	2	5	3	
m2		2	3	2
m3	3		1	2

$$\mathbf{sim}(x, y) = \frac{\sum_i \mathbf{r}_{xi} \cdot \mathbf{r}_{yi}}{\sqrt{\sum_i \mathbf{r}_{xi}^2} \cdot \sqrt{\sum_i \mathbf{r}_{yi}^2}}$$

$$\text{sim}(m2, m1) = 0.748$$

$$\text{sim}(m2, m3) = 0.454$$

$$\begin{aligned} \text{Predict } u1 \text{ to } m2 &= (0.748 \cdot 2 + 0.454 \cdot 3) / (0.748 + 0.454) \\ &= 2.38 \end{aligned}$$