Q_1

(1) ABC

(2) A

(3) AD

(4) ABCD

(5) B C

(6) AD

Q_2

(1) T

(2) T

(3) T

(4) F

(5) F

3.

(a)

```
rddPartitioned = x.partitionBy(2)
```

table 0
(2, 'a')
(4, 'c')

table 1
(1, 'a')
(3, 'a')
(1, 'b')

(b)

```
rddInverted = rddPartitioned.map(lambda x : (x[1], x[0]))
```

(c)

```
sum = rddInverted(lambda value : (Value, 1),
                  lambda x, value: (x[0]+Value, x[1]+1),
                  lambda x, y : (x[0]+y[0], x[1]+y[1]))
perkeyAverage = sum.map(lambda,(sum_value, count)) :
                  (label, sum_value/count))
```

No. this process doesn't have shuffle partition

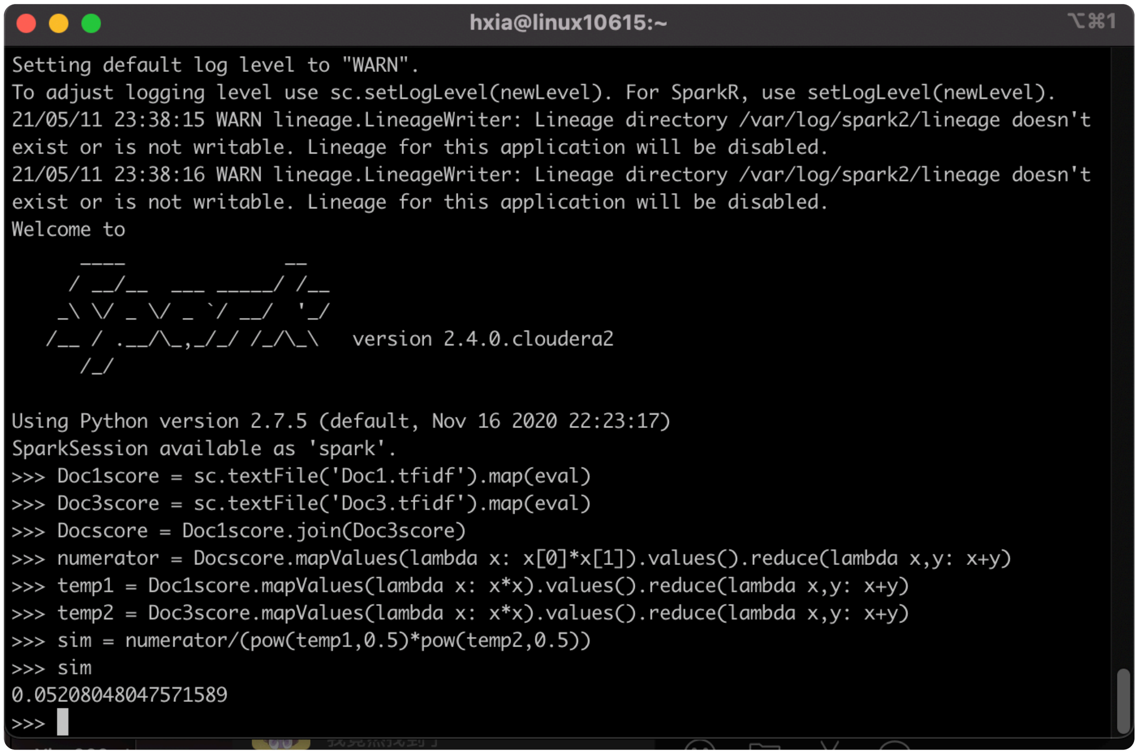
4. (1)

	Doc1	Doc2	Doc3	Doc4
approach	0	0	1	0
breakthrough	1	0	0	0
drug	1	1	0	0
for	1	0	1	1
hopes	0	0	0	1
new	0	1	1	1
of	0	0	1	0
patients	0	0	0	1
schizophrenia	1	1	1	1
treatment	0	0	1	0

(2)

	Doc1	Doc2	Doc3	Doc4
approach	0	0	0.693	0
breakthrough	0.693	0	0	0
drug	0.288	0.288	0	0
for	0.0	0	0.0	0.0
hopes	0	0	0	0.693
new	0	0.0	0.0	0.0
of	0	0	0.693	0
patients	0	0	0	0.693
schizophenia	-0.223	-0.223	-0.223	-0.223
treatment	0	0	0.693	0

(3)



```
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
21/05/11 23:38:15 WARN lineage.LineageWriter: Lineage directory /var/log/spark2/lineage doesn't
exist or is not writable. Lineage for this application will be disabled.
21/05/11 23:38:16 WARN lineage.LineageWriter: Lineage directory /var/log/spark2/lineage doesn't
exist or is not writable. Lineage for this application will be disabled.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.4.0.cloudera2
      /_/

Using Python version 2.7.5 (default, Nov 16 2020 22:23:17)
SparkSession available as 'spark'.
>>> Doc1score = sc.textFile('Doc1.tfidf').map(eval)
>>> Doc3score = sc.textFile('Doc3.tfidf').map(eval)
>>> Docscore = Doc1score.join(Doc3score)
>>> numerator = Docscore.mapValues(lambda x: x[0]*x[1]).values().reduce(lambda x,y: x+y)
>>> temp1 = Doc1score.mapValues(lambda x: x*x).values().reduce(lambda x,y: x+y)
>>> temp2 = Doc3score.mapValues(lambda x: x*x).values().reduce(lambda x,y: x+y)
>>> sim = numerator/(pow(temp1,0.5)*pow(temp2,0.5))
>>> sim
0.05208048047571589
>>> 
```

5.

(a)

$$M = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$r_y = \frac{r_y}{3} + \frac{r_a}{2}$$

$$r_a = \frac{r_y}{3} + \frac{r_m}{2}$$

$$r_m = \frac{r_y}{3} + \frac{r_a}{2} + \frac{r_m}{2}$$

(b) The initial vector V_0 has 3 components, each $\frac{1}{3}$

$$\begin{array}{c} r_y \\ r_a \\ r_m \end{array} = \left(\begin{array}{cccc} \frac{1}{3} & \frac{5}{18} & \frac{25}{108}(0.23) & \frac{19}{81}(0.23) \\ \frac{1}{3} & \frac{5}{18} & \frac{17}{54}(0.31) & \frac{197}{648}(0.30) \\ \frac{1}{3} & \frac{4}{9} & \frac{49}{108}(0.45) & \frac{299}{648}(0.46) \end{array} \right.$$

$$| r^{(t+1)} - r^{(t)} |_1 < \varepsilon (0.01)$$

(C) Google's matrix A:

$\beta = 0.8$

$$0.8 \overset{M}{\begin{bmatrix} \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}} + 0.2 \overset{\left[\frac{1}{N}\right]_{N\times N}}{\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}}$$

$$= \begin{bmatrix} \frac{4}{15} & \frac{2}{5} & 0 \\ \frac{4}{15} & 0 & \frac{2}{5} \\ \frac{4}{15} & \frac{2}{5} & \frac{2}{5} \end{bmatrix} + \begin{bmatrix} \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \frac{1}{3} & \frac{7}{15} & \frac{1}{15} \\ \frac{1}{3} & \frac{1}{15} & \frac{7}{15} \\ \frac{1}{3} & \frac{7}{15} & \frac{7}{15} \end{bmatrix}}_{A} .$$

(d)

$$\begin{array}{c} y \\ a \\ m \end{array} = \begin{array}{cccc} \frac{1}{3} & \frac{13}{45} & \frac{7}{27}(0.26) & \frac{2641}{10125}(0.26) \\ \frac{1}{3} & \frac{13}{45} & \frac{211}{675}(0.31) & \frac{3109}{10125}(0.31) \\ \frac{1}{3} & \frac{19}{45} & \frac{289}{675}(0.43) & \frac{35}{81}(0.43) \end{array}$$

$$|r^{(t+1)} - r^{(t)}|_1 < \varepsilon(0.01)$$