Remaining Time: 28 minutes, 56 seconds.

Question Completion Status:

QUESTION 47

Suppose you have a **movies** dataframe as below:

movieId	title
1	Toy Story
2	Jumanji
3	Grumpier Old Men
...	...

and a **ratings** dataframe as below:

userId	movieId	rating
1	1	4.8
1	3	4.0
...	...	...

You would like to find out which userId has rated the most movies. Which of the following accomplishes this? Which of the following accomplishes this: (You can assume that the relevant libraries have been imported)

- ratings.groupBy("rating").orderBy(desc("count"))
- ratings.groupBy("rating").orderBy(desc("userId"))
- ratings.groupBy("userId").count().orderBy(desc($"count"))
- ratings.groupBy("userId").count().orderBy(desc("count"))

QUESTION 48

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Remaining Time: 28 minutes, 49 seconds.

Question Completion Status:

QUESTION 48

Which of the following are true about narrow transformations for RDDs?

- [x] Map and Filter operations are examples of narrow transformation
- [x] The partition in a RDD has one-to-one relationship with its parent RDD's partition
- [ ] Join is an example of narrow transformation.
- [ ] count is an example of narrow transformation

QUESTION 49

Suppose I have 5 documents as shown below:
Doc1: the weather is great today
Doc2: summer weather in dallas is nice
Doc3: dallas is a wonderful city to live in
Doc4: summer is a great time for walking
Doc5: walking is wonderful for health

What will be the tf-idf for the word "great" in Doc4.
Note: You do not have to perform any sort of normalization. You can just use the basic definition of tf-idf

- ○ tf-idf("great", Doc4) = 1 x log(5/1)
- ○ tf-idf("great", Doc4) = 1 x log(5/2)
- ◉ tf-idf("great", Doc4) = 2 x log(5/2)
- ○ None of the above

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

- tf-idf("great", Doc4) = 1 x log(5/2)
- (•) tf-idf("great", Doc4) = 2 x log(5/2)
- None of the above

QUESTION 50

Suppose I have 5 documents as shown below:
Doc1: the weather is great today
Doc2: summer weather in dallas is nice
Doc3: dallas is a wonderful city to live in
Doc4: summer is a great time for walking
Doc5: walking is wonderful for health

What will be the cosine similarity between Doc3 and Doc5
Note: You can use the same method that you used in assignment 1

- 2/(sqrt(11) * sqrt(5))
- 1/(sqrt(8) * sqrt(5))
- (•) 2/(sqrt(8) * sqrt(5))
- None of the above

Click Save and Submit to save and submit. Click Save All Answers to save all answers.