Week 1 Quiz

Bookmark this page

Access Locality

1/1 point (graded)

Which of the following statements are true considering the below mentioned access patterns for an array "A" in a program?

(1) A[1], A[2], A[3]

(2) A[0][1], A[0][2], A[0][3]

(3) A[1], A[5000], A[1], A[1], A[5000], A[30], A[30], A[5000], A[1], A[5000]

There could be more than one option correct.

- [x] (1) demonstrates spatial locality
- [ ] (1) demonstrates temporal locality
- [x] (2) demonstrates spatial locality
- [ ] (2) demonstrates temporal locality
- [ ] (3) demonstrates spatial locality
- [x] (3) demonstrates temporal locality

Submit | You have used 3 of 5 attempts | Show Answer

Cluster Computing

1/1 point (graded)

Which of the following are advantages of data computation on a cluster?

- ( ) The ability to store data that can not otherwise fit in a single node's memory.
- ( ) The ability to localize computations on nodes that actually store the data, thus saving on communication cost and time.
- ( ) The ability to parallelize computations.
- (•) All of the above.

Submit | You have used 1 of 5 attempts | Show Answer

Finding Average

1/1 point (graded)

Given a list L of integers, which of the following expressions compute the average of the elements in the list?

- (•) reduce(lambda x,y: x+y, l) / reduce(lambda x,y: x+y, map(lambda x: 1, l))
- ( ) reduce(lambda x,y: x+y, l) / reduce(lambda x,y: x+y, map(lambda x: x, l))
- ( ) reduce(lambda x,y: x+y, l) / map(lambda x: x, reduce(lambda x,y: x+y, l))
- ( ) reduce(lambda x,y: x+y, l) / reduce(lambda x,y: 1+1, l)

Submit | You have used 2 of 5 attempts | Show Answer

Replication Factor

1.0/1.0 point (graded)

Replication factor on a HDFS cluster defines the number of redundant copies of each chunk. The idea is that if a server crashes, the information can be recovered from the other copies. Hence, if the replication factor is *N*, and a file has a size *x*, then it effectively occupies a space of *N*x* on the cluster. A replication factor of *N* ensures recovery from up to *N-1* servers failing simultanously.

You are given a cluster of 5 servers, each with 500GB of storage. You run HDFS on these 5 servers with a replication factor of 4. What is the effective amount of storage space available on the HDFS cluster?

625

625

Submit | You have used 1 of 5 attempts | Show Answer

reduceByKey

1.0/1.0 point (graded)

`a = sc.parallelize([(1,2), (1,−5), (7,30), (9,6), (1,3), (7,−2), (1,−3), (7,−1), (1,4)])` `a = a.reduceByKey(lambda x,y: x+y)`

What will a.collect() give?

- ( ) `[(1, 32), (9, 6), (7, 32)]`
- ( ) `[(1, 180), (9, 6), (7, 60)]`
- ( ) `set((1, 1), (9, 6), (7, 27))`
- (•) `[(1, 1), (9, 6), (7, 27)]`

Submit | You have used 1 of 5 attempts | Show Answer

GroupByKey

1/1 point (graded)

```
a=sc.parallelize([(1,2), (1,−5), (7,30), (9,6), (1,3), (7,−2), (1,−3), (7,−1), (1,4)])
a.groupByKey().map(lambda k : (k[0], [x for x in k[1]]))
```

What does a.collect() give?

- ( ) `[(1, 1), (9, 6), (7, 27)]`
- ( ) `[(1, [2, −5, −3, 4)), (9, [6]), (7, [30, −1]))]`
- (•) `[(1, [2, −5, 3, −3, 4)), (9, [6]), (7, [30, −2, −1]))]`
- ( ) `[(1, [2, −5, −3, 4)), (9, [6]), (7, [30, −2, −1]))]`

Submit | You have used 1 of 5 attempts | Show Answer

Map Reduce

1/1 point (graded)

What does the following code yield?

rdd = sc.parallelize([1,2,3,4,5])

rdd.map(lambda x: x+1).reduce(lambda x,y: x*y)

720

720

Submit | You have used 1 of 5 attempts | Show Answer