

Dear Candidate,

Below are a few sample questions for your reference. Please spend some quality time in preparation and then appear for the interview.

=====

1. What is Spark, Spark Architecture, Why we use Spark over MapReduce
2. Spark optimisation Techniques, Different optimization techniques used
3. Different spark Transformations
4. Spark Deployment
5. Dynamic allocation of executors in spark
6. How the stages and task are formed in spark
7. Why we are using spark if we have map reduce --> other than in memory processing and lazy evaluation
8. Explain dynamic resource allocation in spark
9. When to cache, Explain Cache in spark
10. Higher order functions in Scala
11. Broadcasting
12. How to handle data skew
13. Group connected edges from a graph
14. Main difference between recursion and tail recursion
15. Explain accumulator
16. Explain broadcast variable
17. Difference between readtextfile and wholetextfile
18. What is the difference between calling a function from reduce method and creating anonymous function in reduce itself
19. Difference between flatMap and Map with an example
20. Transformations and narrow transformations
21. RDD vs Dataframes vs Datasets
22. What are the different ways to address data skewness

=====

**Programming Questions:**

1. Merging the tuple using scala

List((10,15),(12,20),(18,28),(30,40),(40,50),(55,60))

---

- 2. Given a list of dependencies tuples, separate them into different stages - Tasks get executed after all dependent tasks get completed.**

Input -

Tasks - 1,2,3,7,8

Dependency list - [(1,2),(2,3),(7,8)]

Output -

Stage 1 - 3,8

Stage 2 - 2,7

Stage 3 - 1

---

**3. List of tuples**

Group the tuples if difference between two tuples is less than 2 and provide the output of groups with tuples

---

- 4. Find if the given word is present in a file or not without reading entire data (after the first instance of that word we should no longer scan the file)**
- 

- 5. Average of numbers in a given file using reduce function also by using reduce as higher order function (File consists of one number in each row)**
- 

- 6. Write a logic to find word exist or not in a file using spark**
- 

- 7. Let's assume you have one textfile with numbers**

**Find a avg of number using reduce method**

**OR**

**Calculate average of 2 numbers in spark using reduce and few more questions on reduce function**

---

- 8. Explain foldleft with example**

aaabbbcca -> 3a3b2c1a

-----  
**9. Write a program to convert the string from aaabbbccdda to 3a3b2c2d1a**  
-----

10. q1. given a = List(1,2,3,4,5)

add this using recursion in scala

q2. solve above problem using tail recursion

q3. check if a string is palindrome

q4. there is a user and collage scala classes with coordinates x,y

User(userid , x , y)

User(collage\_name , x , y)

x,y in range of 0 - 100

find all college for users based on proximity of 10

=====