COL733 Quiz 1

Viraj Agashe

TOTAL POINTS

6 / 10

QUESTION 1

Question 1 10 pts

1.1 Question 1.1 **0 / 2**

✓ **+ 0 pts** *Incorrect*

+ 1 pts Estimates Pi

+ 1 pts f returns 1 if random coordinates are inside the unit circle

1.2 Question 1.2 **2 / 2**

+ 0 pts Incorrect

✓ **+ 1 pts** *r -> m -> count (Nodes are RDDs)*

✓ **+ 1 pts** *Edges are transformations*

1.3 Question 1.3 **2 / 2**

+ 0 pts Incorrect

✓ **+ 1 pts** *N map tasks are narrow*

✓ **+ 1 pts** *One reduce task is wide*

1.4 Question 1.4 **2 / 4**

+ 0 pts Incorrect. Or without justification

✓ **+ 2 pts** *Generates correct outputs after crashes*

+ 2 pts After restart on a new worker, function still estimates pi

Name: VIRAJ AGASHE Entry number: 2020CS10567

Let us say that we are given the following Spark program:

```
1.  import sys
2.  from random import random
3.  from operator import add
4.
5.  from pyspark.sql import SparkSession
6.
7.  if __name__ == "__main__":
8.    # Usage: pi [partitions]
9.    spark = SparkSession\
10.        .builder\
11.        .appName("PythonPi")\
12.        .getOrCreate()
13.
14.   p = int(sys.argv[1]) if len(sys.argv) > 1 else 2
15.   n = 100000 * p
16.
17.   def f(_: int) -> float:
18.     x = random() * 2 - 1
19.     y = random() * 2 - 1
20.     return 1 if x ** 2 + y ** 2 <= 1 else 0
21.
22.   r = spark.sparkContext.parallelize(range(1, n + 1), p)
23.   m = r.map(f)
24.   count = m.reduce(add)
25.   print("Pi is roughly %f" % (4.0 * count / n))
26.   spark.stop()
```

[2 marks] Guess and explain what this program is computing between lines 14 to 25.

{Assuming random returns $\in [0,1]$.}

Then note that $2x-1 \rightarrow \begin{cases} x=0 \Rightarrow -1 \\ x=1 \Rightarrow 1 \end{cases}$

$\therefore$ $f(-)$: Always returns 0.

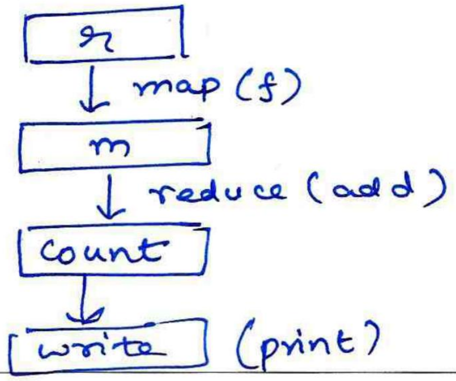
(Since x^2, y^2 both are $=1$. always)

$\therefore$ All values map to 0. (count $= n$, $n=n$).

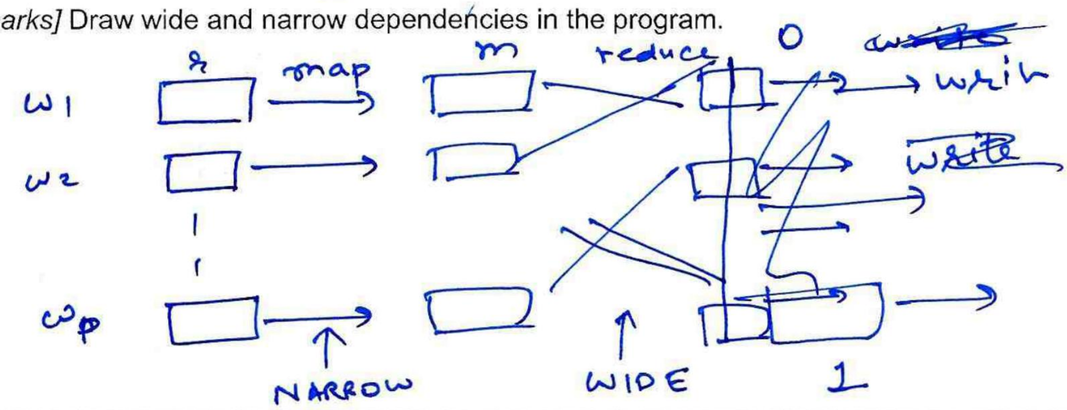
$\therefore$ The program is computing the value $\boxed{4}$ always.

Name: VIRAJ AGASHE Entry number: 2020CS10567.

[2 marks] Draw the lineage graph of the program.



[2 marks] Draw wide and narrow dependencies in the program.



[4 marks] Note that the function `f` is non-deterministic. Due to this non-determinism, can the program generate "incorrect" outputs in case workers crash? Justify your answer.

Although the function is non-deterministic, the output is always deterministic (1).

∴ Even if workers crash & recomputation is done, we will always return the value '0' from the function. (whenever it runs)

So, it is independent of crashes.