



الإجابة النموذجية لامتحان الدورة العادية في مقياس تحليل البيانات الضخمة

العلامة	السؤال الاول
1	الإجابات الصحيحة: 1: د
1	2: أ ج د
1	3: أ
1	4: ب
1	5: أ ج
1	6: ج
1	7: د
07	المجموع

النقاط	السؤال الثاني
	<p>توضيح آلية عمل الـ HDFS في تخزين الملفات الثلاثة بالاستعانة بالمخطط البياني:</p> <p>The diagram illustrates the HDFS architecture. On the left, three directories are shown: /workersdata/2022 (129MB) containing files B1 (128MB) and B2 (1MB); /workersdata/2023 (379MB) containing files B3 (128MB), B4 (128MB), and B5 (123MB); and /workersdata/2024 (33MB) containing file B6 (33MB). On the right, the HDFS Cluster consists of six Data Nodes (A-F). Data Node A contains B1, B2, B3; Data Node B contains B2, B3, B4; Data Node C contains B3, B4, B5; Data Node D contains B4, B5, B6; Data Node E contains B5, B6, B1; and Data Node F contains B6, B1, B2. A red '2' is next to the 2023 directory, and a red '1.5' is next to the HDFS Cluster. Below the cluster is the Name Node containing meta data: /workersdata/2022: B1, B2; /workersdata/2023: B3, B4, B5; /workersdata/2024: B6; B1: A, E, F; B2: A, B, F; B3: A, B, C; B4: B, C, D; B5: C, D, E; B6: D, E, F. A red '1.5' is next to the Name Node.</p>
05	المجموع

السؤال الثالث

النقاط

```

%% file wordcount.py
from mrjob.job import MRJob
class MRWordCount(MRJob):
    def mapper(self, _, line):
        for word in line.split('\t'):
            yield(word, 1)
    def reducer(self, word, counts):
        yield(word, sum(counts))
if __name__ == '__main__':
    MRWordCount.run()

```

1

إعادة كتابة الكود مع تصحيح كل الأخطاء الموجودة:

```

%%file wordcount.py
from mrjob.job import MRJob
class MRWordCount(MRJob):
    def mapper(self, _, line):
        for word in line.split('\t'):
            yield(word, 1)
    def reducer(self, word, counts):
        yield(word, sum(counts))
if __name__ == '__main__':
    MRWordCount.run()

```

5

كتابة الأمر اللازم لتنفيذ الكود السابق لحساب عدد تكرار الكلمات في الملف:

1

! python wordcount.py D:/data/inputdata.txt

2

كتابة الأمر اللازم لتنفيذ الكود السابق لحساب عدد تكرار الكلمات في الملف بافتراض أن الملف موجود في HDFS:

3

2

! python wordcount.py -r hadoop hdfs:///tpdata/inputdata.txt

08

المجموع