

# PO12: Przetwarzanie Big Data 22/23-Z

Rozpoczęto	niedziela, 16 października 2022, 19:33
Stan	Ukończone
Ukończono	niedziela, 16 października 2022, 19:42
Wykorzystany czas	8 min. 50 sek.
Punkty	29,00/34,00
Ocena	8,53 pkt. na 10,00 pkt. możliwych do uzyskania (85,29%)

09

Nasze pliki nie posiadają zbyt dużych rozmiarów. HDFS wykorzystuje z reguły bloki o rozmiarach 64-256MB. Nawet nasz największy plik nie posiada takiej wielkości. Zobaczmy na ile bloków został on podzielony.  

```
hdfs fsck /user/$USER/input/trip2.csv -files -blocks -locations
```

a. Ile bloków zostało wykorzystanych?

1

✓

b. Jaki jest domyślny stopień replikacji?

2

✓

c. Jaki port został wykorzystany do połączenia się z serwerem nazw (namenode) systemu HDFS?

9866

✗

Wszystkie terminy użyte w powyższych pytaniach zostaną omówione w późniejszym czasie.

Pytanie 1

Częściowo poprawnie

Punkty: 2,00 z 3,00

Oflaguj pytanie

13

Zanim przeanalizujemy uzyskane wyniki odpowiedz na następujące pytania – skorzystaj ze

- statystyk wyświetlonych po wykonaniu zadania
- analizy kodu programu
- parametrów jego uruchomienia

a. Co było na wejściu do zadania MapReduce w naszym przypadku (jakie słowa on zliczał)?

Pliki ze wskazanego rozproszonego systemu plików HDFS

✓

b. Jak wyglądają pary klucz-wartość generowane w wyniku użycia funkcji map?

zawartość linii z pliku, słowo

✗

c. Ile rekordów wygenerowanych zostało na wyjściu funkcji map?

13108471

✓

d. Czy zastosowana została funkcja łącząca (agregator łączący - combiner) pomiędzy funkcją map, a reduce ograniczającą liczbę danych przesyłanych do funkcji reduce.

nie

✓

e. Ile rekordów wygenerowała funkcja reduce?

239093

✓

Pytanie 2

Częściowo poprawnie

Punkty: 4,00 z 5,00

Oflaguj pytanie

28

Zacznijmy od klasy Mapper. Znajdź jego definicję ponownie w naszym programie. W Javie implementuje interfejs Mapper i jest klasą generyczną z czterema parametrami formalnymi, które określają typy:

- klucz wejściowy – w naszym przypadku: przesunięcie linii w pliku
- wartość wejściowa – w naszym przypadku: wiersz tekstu
- klucz pośredni – w naszym przypadku: powinien to być rok instalacji stacji – jest to tekst, ponieważ używamy go tylko jako klucza
- wartość pośrednia – w naszym przypadku: jest to wielkość stacji – liczba.

Przeanalizuj definicję tej klasy i odpowiedz na następujące pytania:

a. Jak nazywa się klasa, która implementuje funkcjonalność Mappera w naszym programie?

AvgSizeStationMapper

✓

b. Jakiego typu jest klucz wejściowy?

LongWritable

✓

c. Jakiego typu jest klucz pośredni?

Text

✓

d. Jakiego typu jest wartość pośrednia?

IntWritable

✓

e. Do której zmiennej tej klasy przypisana jest wartość pośrednia?

size

✓

f. Który parametr metody map zawiera zawartość kolejnych wierszy z pliku źródłowego?

lineText

✓

g. Jaki jest cel warunku: if (offset.get() != 0)

pominięcie pierwszej linii (zawierającej nagłówek)

✓

Pytanie 3

Poprawnie

Punkty: 7,00 z 7,00

Oflaguj pytanie

31

Przejdź do klasy Reduktora. Znajdź jego definicję ponownie w naszym programie. W Javie implementuje on interfejs Reducer, i jest to również klasa generyczna z czterema parametrami formalnymi, które określają typy:

- klucz pośredni – klucz pośredni zwracany przez mapera lub przez... (wiesz co?)
- wartość pośrednia – wartość pośrednia zwrócona przez mapera lub przez ... (wiesz co?)
- klucz wyjściowy – w naszym przypadku będzie to tekst zawierający informację o roku
- wartość wyjściowa funkcji redukcyjnej – w naszym przypadku średnia wielkość stacji – liczba.

Przeanalizuj definicję tej klasy i odpowiedz na następujące pytania:

a. Jak nazywa się klasa, która implementuje funkcjonalność reduktora w naszym programie?

AvgStationReducer

✗

b. Jaki jest typ klucza pośredniego?

Text

✓

c. Jaki jest typ klucza wyjściowego?

Text

✓

d. Jaki jest typ wartości wyjściowej?

DoubleWritable

✓

e. Jaką wartość zawiera parametr key metody reduce za każdym razem, gdy jest wywoływana?

Rok instalacji stacji

✓

f. Jaką wartość zawiera parametr value metody reduce za każdym razem, gdy jest wywoływana?

średnią wielkość stacji

✗

Pytanie 4

Częściowo poprawnie

Punkty: 4,00 z 6,00

Oflaguj pytanie

34

Pozostał ostatni element TODO – znajdziesz go w definicji sterownika. Znajdź miejsca konfiguracji zadania MapReduce.

a. Jak nazywa się metoda, w której skonfigurowano zadanie MapReduce?

run

✓

b. Jak możemy określić dane wejściowe dla zadania MapReduce?

jest to pierwszy parametr uruchomienia zadania

✓

c. Gdzie będą zapisywane dane wyjściowe MapReduce?

w katalogu wskazanym przez drugi parametr uruchomienia zadania

✓

d. Która klasa domyślnie będzie używana jako klasa Mapera?

Mapper

✓

e. Jaka klasa domyślnie będzie używana jako klasa Reduktora?

Reducer

✓

Pytanie 5

Poprawnie

Punkty: 5,00 z 5,00

Oflaguj pytanie

47

Zagladnij do statystyk wyświetlanych po uruchomieniu zadania MapReduce.

a. Ile stacji było w pliku wejściowym (nie licz nagłówek)?

58

✓

b. Ile par pośrednich zostało wysłanych między maperami a reduktorami?

3

✗

c. Ile jednostek zadań mapowania zostało uruchomionych?

1

✓

Pytanie 6

Częściowo poprawnie

Punkty: 2,00 z 3,00

Oflaguj pytanie

49

Zobacz do statystyk wyświetlanych po uruchomieniu zadania MapReduce.

a. Ile par pośrednich zostało tym razem wysłanych między maperem (combinerami) a reduktorami?

3

✓

b. Czy ta liczba mogłaby się zmienić, gdyby uruchomiono więcej maperów (np. z powodu większej liczby plików wejściowych)?

tak

✓

Pytanie 7

Poprawnie

Punkty: 2,00 z 2,00

Oflaguj pytanie

54

Czy zadanie MapReduce zostało pomyślnie uruchomione przy użyciu usługi Hadoop Streaming? Jeśli tak, przyjmij gratulacje.

a. Spójrz na wyniki. Czy są poprawne?

nie

✓

b. Brakuje jednego polecenia w jednym z powyższych programów.

Wprowadź numer linii w której brakujące polecenie powinno być dodane: 19

✓

W którym programie?

reduktora

✓

Pytanie 8

Poprawnie

Punkty: 3,00 z 3,00

Oflaguj pytanie

Nawigacja w teście

1

2

3

4

5

6

7

8

Pokaż jedną stronę na raz

Zakończ przegląd

Zakończ przegląd

Jesteś zalogowany(a) jako Andrzej Kapczyński (Wylóguj)

WiIT-Inf-st-I-sem7-PrzBigDat-22/23-Z