

Kosmiczne zastosowania zaawansowanych technologii informatycznych - projekt 2019/2020

1. Należy dobrać się w pary
2. W parze wybrać osobę, która urodziła się poza Gdańskiem (jeśli taka jest) – określić miejsce urodzenia: M
3. Wybrać dzień i miesiąc urodzenia drugiej osoby: D
4. Otrzymujemy parę: MD = (M, D)

Część 1. (10 pkt.) – przegląd przykładowych danych satelitarnych (do 30.10.2019)

1. Sentinel-2 – projekt Copernicus
2. Odnaleźć dokumentację, przejrzeć jakie dane satelitarne są dostępne
3. Odnaleźć jeden przykładowy obraz satelitarny wraz z metadanymi tak aby był najbliższy do MD
4. Odpowiedzieć na pytania w formie pisemnej – sprawozdanie

Pytania dot. ściągniętego obrazu MD:

1. Opisać strukturę danych (wraz z obrazkiem)
2. Podać liczbę i długości plików
3. Wersja produktu – jakie są dostępne
4. Czy jest jedno czy więcej zdjęć, czym się różnią
5. Podać 3 przykładowe parametry opisu zdjęcia (np. % pokrycia chmurami) wraz z wartościami podanymi dla MD
6. Opisać kształt wielokąta opisującego zdjęcie
7. Sprawdzić liczbę i podać przykładową pozycję niepoprawnych pikseli na zdjęciu MD

Część 2. (10 pkt.) – podstawowe operacje na danych (do 11.12.2019)

Strona www: <https://bigdata.apl.task.gda.pl/>

1. Wygenerować RDD tekstowe: *names*, 64 elementy, każdy element ma składać się z imienia jednej z osób z zespołu, prawdopodobieństwo wystąpienia imienia: 70% osoba starsza, 30% osoba młodsza.
2. Stworzyć 2 RDD klucz-wartość:
 - a. *len*, w którym będą długości elementów *names* (liczby całkowite)
 - b. *sum*, w którym będą iloczyny wartości Unicode elementów *names* (liczby zmiennoprzecinkowe)
3. Obliczyć statystyki (liczba elementów, średnia, standardowa odchyłka, maksimum, minimum, wariancja, suma, histogram) dla kluczy *len* i *sum*.
4. Znaleźć unikatowe elementy we wszystkich RDD.
5. Zapisać wszystkie RDD na dysk.

Sprawozdanie ma zawierać:

1. Kody źródłowe programów w Scala'i

2. Wyniki działania programów
3. Listę elementów *names*
4. Histogramy dla wszystkich RDD w formie graficznej