



ELTE | IK
INFORMATIKAI KAR

BigData architektúrák és elemző módszerek

3. gyakorlat

Feladat: WordMean

- Adjuk meg, hogy a bemeneti fájlban mekkora a szavak átlagos hossza
- Példa bemenet: „Tedd, vagy ne tedd, de ne próbáld!”
- Példa kimenet: 3.57

WordMean átalakítás

- A counterek statisztikát készítenek a futás során.
- Pl: megszámlálják a map által létrehozott kulcs-érték párokat, stb.
- Használjunk counter-t a reducerben a szavak számának meghatározásához:

```
int word_num =  
(int)context.getCounter(TaskCounter.REDUCE_INPUT_RECORDS)  
.getValue()
```

- Saját counter-t is létre lehet hozni, ha szeretnénk bizonyos előfordulásokat megszámlálni
- Pl: grep esetén megszámlálni a talált sorokat, stb.

Feladat: MaxLength 1.

- Adjuk meg, hogy a bemeneti fájlban kezdőbetűnként milyen hosszú a leghosszabb szó.
- Példa bemenet: „Tedd, vagy ne tedd, de ne próbáld!”
- Példa kimenet:
 - T: 4
 - V: 4
 - D: 2
 - N: 2
 - P: 7

Feladat: MaxLength 2.

- Adjuk meg kezdőbetűnként a leghosszabb szavakat a bemeneti fájlból.
- Példa bemenet: „Tedd, vagy ne tudd, de ne próbáld!”
- Példa kimenet:
 - T: tedd tudd
 - V: vagy
 - D: de
 - N: ne
 - P: próbáld
- Plusz feladat: használj combinert a megoldáshoz!

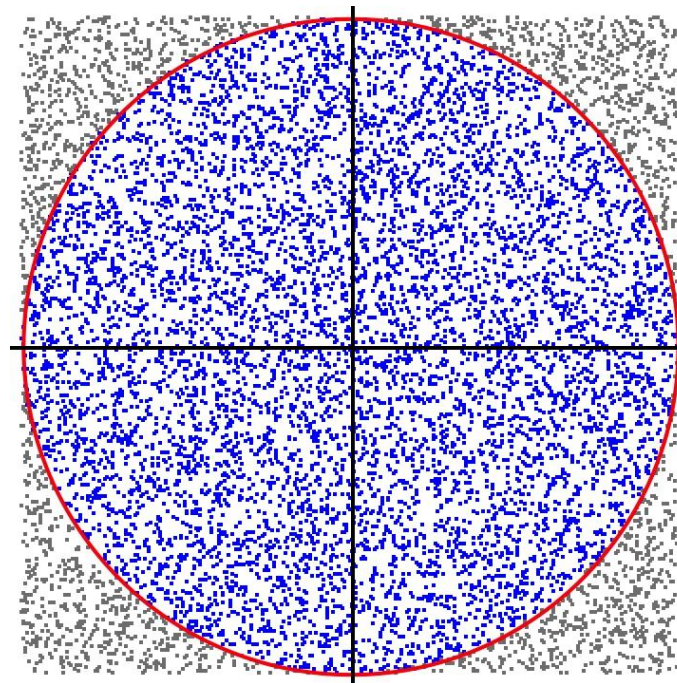
Feladat: Pí becslése Monte-Carlo módszerrel

- Az egységkör területe: πr^2
- A négyzet területe: $4r^2$

$$\frac{\text{kör területe}}{\text{négyzet területe}} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4}$$

$$\frac{\pi}{4} = \frac{\text{körbe eső pontok száma}}{\text{négyzetbe eső pontok száma}}$$

$$\pi = 4 \times \frac{\text{körbe eső pontok száma}}{\text{négyzetbe eső pontok száma}}$$



- Tipp: generáljunk random pontokat $[0,1]$ intervallumon és nézzük meg milyen messze esik az origótól. Ha közelebb van, mint 1, akkor a körbe esik.

Pí becslés átalakítása

- Használjunk a reducerben `cleanup()` függvényt.
- A `cleanup()` akkor fut le, ha a reducer véget ér.

```
public void cleanup(Context context) {...}
```

- Adjuk meg a pontok számát a driverben vagy akár argumentumként.
- Driverbe:

```
conf.setInt("num", Integer.parseInt(args[0]));
```

- Mapperbe:

```
Configuration conf = context.getConfiguration();  
int point_num = conf.getInt("num", 0);
```

Feladat: Maximális hőmérséklet

- Adjuk meg a naponkénti maximális hőmérsékletet.
- A megoldáshoz saját Writable osztályt hozunk létre.
- Az adatok formátuma:

```
1 location,Budapest
2 lat,47.49973
3 lon,19.01991
4 asl,158.301
5 variable,Temperature
6 unit,°C
7 level,2 m elevation corrected
8 resolution,hourly
9 aggregation,None
10 timestamp,Budapest Temperature [2 m elevation corrected]
11 20200914T0000,22.781458
12 20200914T0100,22.311459
13 20200914T0200,21.571459
14 20200914T0300,20.741459
15 20200914T0400,20.371458
16 20200914T0500,20.111458
17 20200914T0600,19.571459
18 20200914T0700,19.601458
19 20200914T0800,21.021458
20 20200914T0900,23.241459
```