

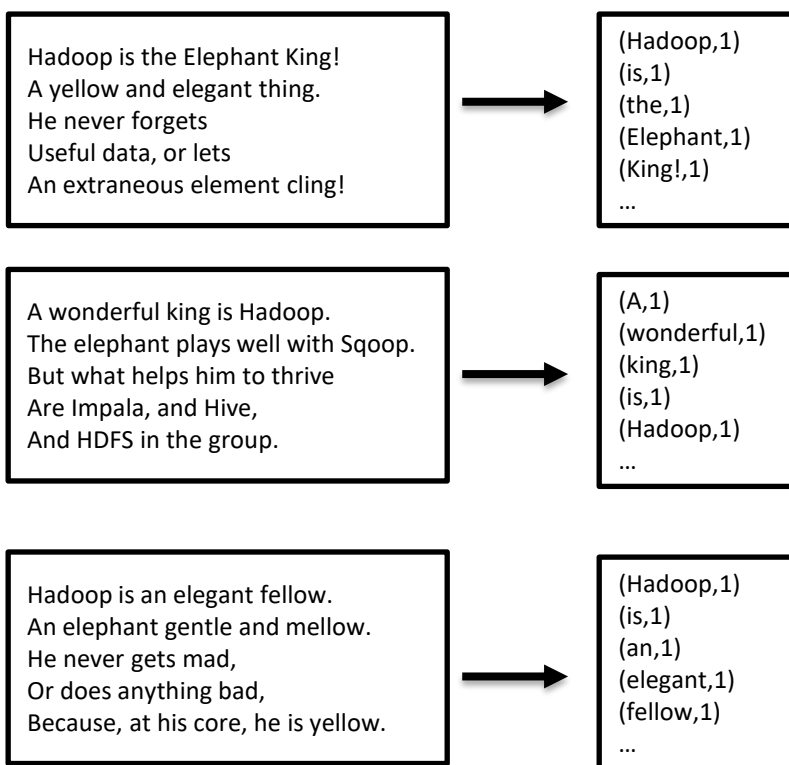
Big Data architektúrák és elemző módszerek

Gyakorlat

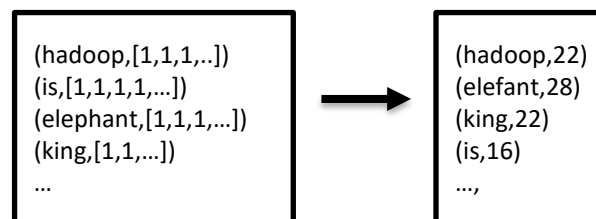
Gombos Gergő

WordCount példa

Mapper



Reducer



Feladat1 - Mapper

- Írjuk meg a Mapper osztályt!
- A mapperek
 - szedjenek le speciális karaktereket (',','.', '!')
 - kisbetűsége alakítsák a szavakat
- Állítsuk be hogy ne legyenek reducer folyamatok! Ezzel meg tudjuk nézni a map folyamatok kimenetelét!

```
job.setNumReduceTasks(0);
```

Feladat2 - Reducer

- Írjuk meg a Reducer osztályt!
- Driverből vegyük ki hogy ne legyen reducer!
- A Reducer egy összegzést végezzen a kapott értékeken!

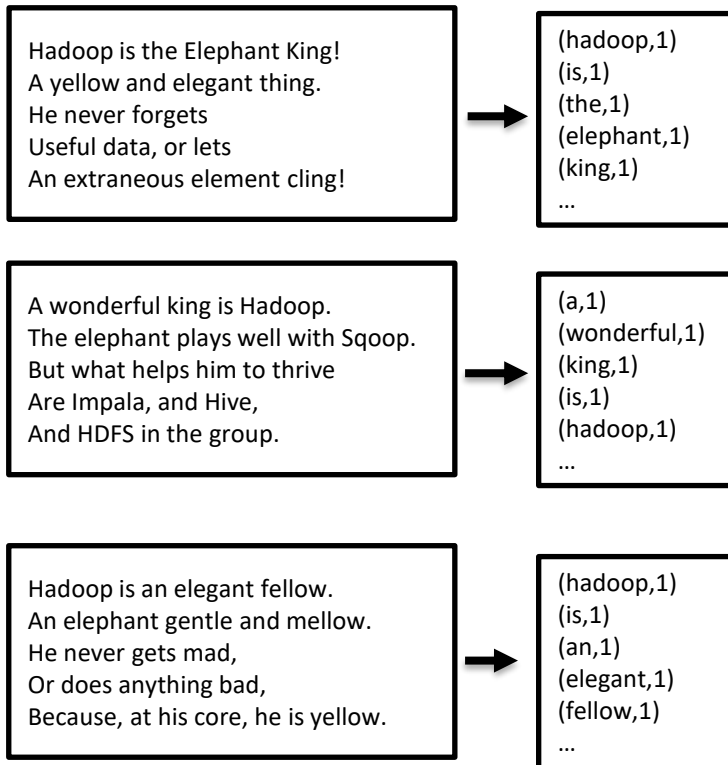
Driver

- Ha nem szeretnénk minden futás előtt törölni a kimeneti mappát, akkor az alábbi kód elvégzi nekünk.

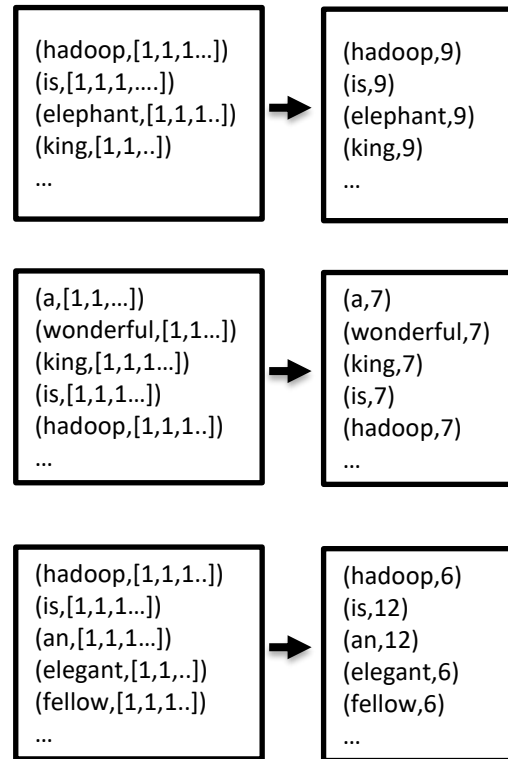
```
FileSystem fs;  
fs = FileSystem.get(conf);  
if(fs.exists(outPath)) {  
    fs.delete(outPath, true);  
}
```

WordCount példa

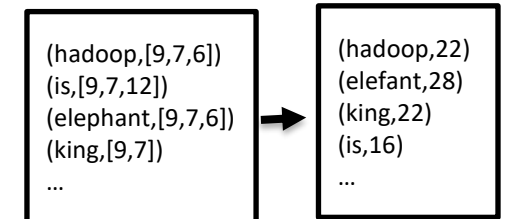
Mapper



Combiner



Reducer



Feladat3 - Combiner

- Készítsünk egy Combiner osztályt!
 - összegezze az eddigi részeredményeket
 - szűrje ki az 'a' és 'an' szavakat!
- A Combiner egy Reducer osztály!
- Driveben kell beállítani:

```
job.setCombinerClass(WordCountCombiner.class);
```

Feladat4 - Grep

- Készítsünk egy grep funkcionalitást elvégző MapReduce Jobot!
- Grep feladata:
 - Visszaadni azokat a sorokat, amelyben a keresett karaktersorozat megtalálható
 - Keresett szó: "elephant"
 - Eredmény:
 - An **elephant** gentle and mellow.
 - Hadoop is the **Elephant** King!
 - The **elephant** plays well with Sqoop.

Feladat gyakorlásnak

- Számítsuk ki az átlagos szóhosszt MapReduce algoritimussal!

Köszönöm a Figyelmet!