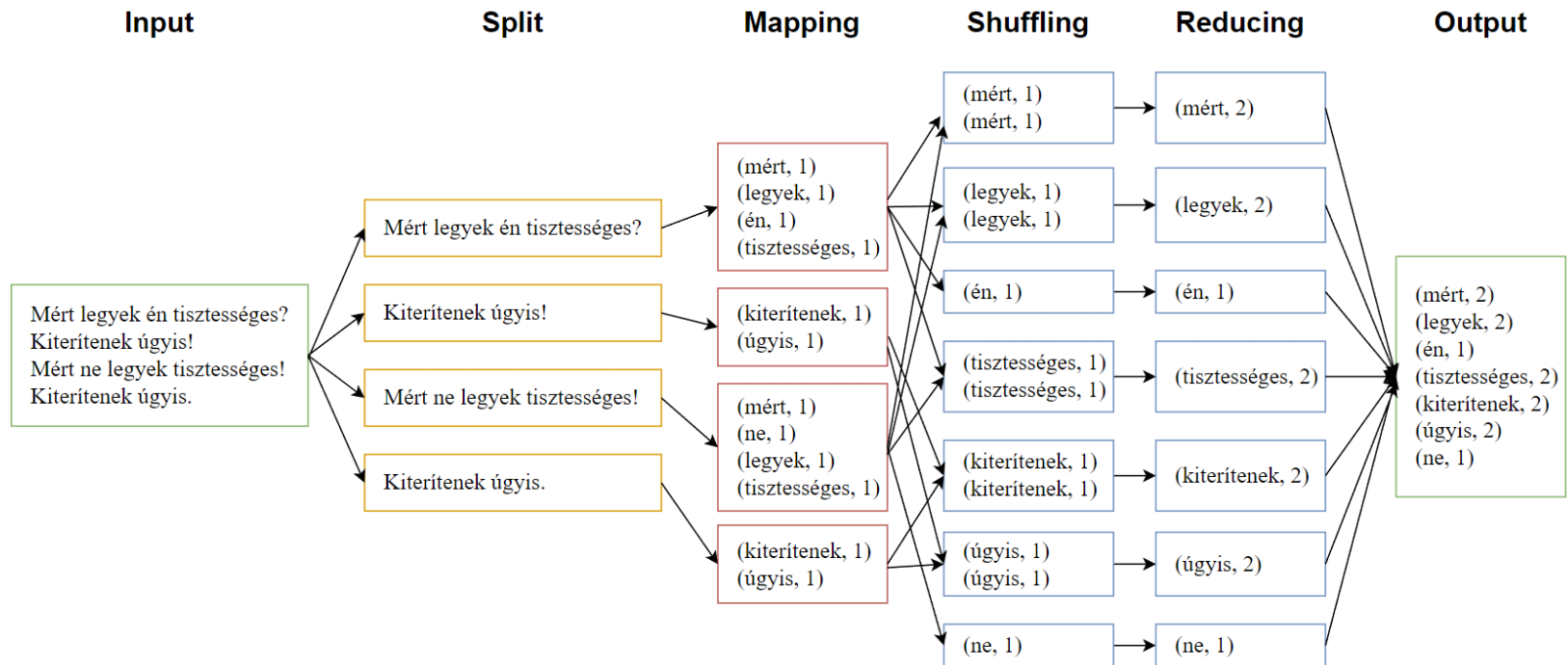


# Big Data architektúrák és elemző módszerek Gyakorlat

Gombos Gergő

# Feladat: WordCount



Szöveg: József Attila - Két hexameter

# WordCount átalakítása

---

- Ha nem szeretnénk a kimeneti mappát minden futtatásnál kézzel törölni írjuk az alábbi kódrészletet a driver-be.
- Az outputPath az a kimeneti útvonalunk legyen.

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

# WordCount átalakítása

---

- Ha szeretnénk combiner segítségével lokálisan aggregálni a szavak számát használjuk a reducerünket combinerként.
- Ehhez állítsuk be a megírt reducer osztályt combinernek a driverben:

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

# Feladat: Grep

---

- Készítsünk egy MapReduce programot, amely egy karaktersorozatot keres a bemeneti fájlban. Írjuk ki azokat a sorokat a bementi fájlból, amelyek tartalmazzák a keresett karaktersorozatot.
- Pl:
  - Keresett karaktersorozat: „sütsz”
  - Bemeneti fájl tartalma:  
*Mit sütsz kis szűcs,  
tán sós húst sütsz  
kis szűcs?*
  - Kimeneti fájl elvárt tartalma:  
*Mit sütsz kis szűcs,  
tán sós húst sütsz*

# Feladat: Grep

---

- A grep feladat esetébe a reducer akár el is hagyható, az alábbi sor driver-be írásával:

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