

Big Data architektúrák és elemző módszerek Gyakorlat

Gombos Gergő

Demo (Pi számítás)

- Nézzük meg a PiCalc kódot!
 - Saját InputFormat
 - Annyi Splitet csinál, ahány map task kell
 - Saját RecordReader
 - Annyi recorodot csinál, amennyi egy taskba kell
 - DummyInputSplit
 - PiReducer
 - cleanup() fv.
 - Minden Reducer saját destruktora.
 - Ez nem globális!

Demo (Elosztott cache)

- Nézzük meg a CacheSetup kódot!
 - Driver
 - `job.addCacheFile(new URI("stopWords.txt"));`
 - egy cache file-t allitunk be, amelyet minden task el tud érni
 - Mapper
 - `setup()` fv
 - minden Mapper saját konstruktora.
 - Ez nem globális!

Feladat (Hadoop beadandó)

- Hadoop beadandó elkészítése!