

Laboratorio 7. (4° giorno - mattina)

Obiettivi

- Familiarizzarsi con l'uso di packages esterni
- Laboratorio su GraphFrames

Riferimenti

I packages: <https://spark-packages.org/>

In particolare, *GraphFrames*: <https://spark-packages.org/package/graphframes/graphframes>

Sorgente e documentazione:

<https://github.com/graphframes/graphframes>

https://graphframes.github.io/graphframes/docs/_site/index.html

Per utilizzare GraphFrames con Zeppelin:

Editare il file `/opt/bigdata/zeppelin-0.8.1/conf/zeppelin-env.sh`. Aggiungere la riga:

```
export SPARK_SUBMIT_OPTIONS="--packages graphframes:graphframes:0.7.0-spark2.4-s_2.11"
```

Far ripartire zeppelin

Esercizi

Esercizio.1

Provate a cercare il package GraphFrames a partire da *spark-packages.org*. Esplorate i packages presenti. Ci sono cose interessanti.

Esercizio.2

Lanciate zeppelin e provate a dare il comando

```
import org.graphframes._
```

Dovrebbe dare errore.

Fate la modifica indicata sopra (nel file *zeppelin-env.sh*) e fate ripartire zeppelin. Adesso il comando dovrebbe avere successo.

Esercizio.3

Importate e caricate la nota *Spark/Lab7/1.Es*. Esaminate i vari paragrafi provando a rieseguirli e facendo piccole modifiche.

Esercizio.4

Aprite una nuova nota. Caricate il file *spark/keep/partnerships.csv*. (attenzione! Il separatore non è la virgola!). Create un grafo considerando legate due istituzioni che hanno fatto almeno un progetto insieme.

Qual è l'istituzione che ha avuto più partner?