

Laboratorio 6. (3° giorno - pomeriggio)

Obiettivi

- Esperimenti con il cluster.
- Programmi in Spark: redazione, compilazione ed esecuzione.

Riferimenti

Esecuzione (spark-submit): <https://spark.apache.org/docs/latest/submitting-applications.html>

La Spark UI: <https://spark.apache.org/docs/latest/monitoring.html>
<https://jaceklaskowski.gitbooks.io/mastering-apache-spark/spark-webui.html>

Per compilare usiamo Maven: <https://maven.apache.org/>

La struttura generale di un programma Spark:
<https://spark.apache.org/docs/latest/quick-start.html#self-contained-applications>

```
import org.apache.spark.sql.SparkSession
object MioProgramma {
  def main(args: Array[String]) {
    val spark = SparkSession.builder
      .appName("Mio Programma")
      .getOrCreate()

    // ....
  }
}
```

Trovate due esempi completi dentro *corso_tim_spark/projects*

Esercizi

Esercizio.1

Andate nel folder *corso_tim_spark/projects/collatz*. Aprite con un editor di testo (*gedit* oppure *Eclipse*) il file *src/main/scala/collatz.scala* ed esaminate il contenuto. Se volete potete fare delle modifiche (ad es. cambiare il testo delle *println* oppure modificare il valore di N).

Provate a compilare con il comando:

```
mvn package
```

Se la compilazione ha successo viene creato il file *target/collatz-1.0-SNAPSHOT.jar*

Lanciate l'esecuzione (in modalità locale) con il comando *spark-submit* .

Esercizio.2

Confrontate i tempi di esecuzione con e senza thread: (*--master local[4]* vs. *--master local*)

Esercizio.3

Assicuratevi di avere un N tale che l'esecuzione richieda parecchi secondi. Durante l'esecuzione aprite il browser e andate all'indirizzo *localhost:4040*. Esplorate l'interfaccia grafica di Spark.

Esercizio.4

Modificate il programma e cambiate il numero di partizioni nella *parallelize* iniziale (linea 36). Cambia il tempo di esecuzione? Per quali valori riusciamo ad ottenere il tempo minimo?

Esercizio.5

Provate ad esplorare l'altro programma: *square_triangular_number*

Esercizio.6

Provate a scrivere un programma nuovo, ad es. riprendendo uno degli esercizi dei giorni scorsi. Copiate e modificate il *pom.xml* dei due programmi già presenti. Provate a compilare ed eseguire