

## **Descrizione dei file e delle cartelle**

- La cartella Sporting\_odds contiene le seguenti sottocartelle:
  - La cartella dataset contiene i file csv relativi ai dataset sorgente
  - La cartella etl contiene gli script e i file utilizzati per i processi di etl volti all'alimentazione di:
    - ODS
    - Datawarehouse.
- La cartella so\_workspace contiene il workspace Eclipse.
- La cartella dump contiene i dump dell'ods (sottocartella ods) e del datawarehouse (sottocartella ss).
- Il file Sporting\_odds.xml rappresenta il modello Pentaho per l'analisi.

## **Guida all'utilizzo per l'esecuzione del processo di ETL**

1. Copiare la directory Sporting\_odds all'interno della cartella "Data" di MySQL Server
2. Se il path dalla directory relativa al punto 1 è diverso da:  
"C:/ProgramData/MySQL/MySQL Server 5.7/Data/Sporting\_odds/"  
modificare il contenuto del file "basedir.txt" presente all'interno del workspace Eclipse nella sottocartella "data"
3. Eseguire gli script contenuti nella cartella Sporting\_odds/etl nell'ordine riportato.  
Gli script da eseguire solo per il primo popolamento del datawarehouse hanno, nel nome del file, una C dopo il numero che rappresenta l'ordine di esecuzione.  
I file con numeri 2 e 5 (nel cui nome del file, si trova una J dopo il numero), non sono da eseguire direttamente. Rappresentano delle esecuzioni di codice Java. I file indicano i metodi main da eseguire.

Dopo ogni esecuzione del processo di ETL cancellare i seguenti file:

- Sporting\_odds/etl/csv/created/participants\_esdb.csv
- Sporting\_odds/etl/csv/created/competitions\_esdb.csv
- Sporting\_odds/etl/csv/created/max\_date.csv