

# MapReduce, Hadoop Streaming, Hive

## Zestaw 6 – fifa-players

---

### Program MapReduce (2)

Działając na zbiorze `datasource1` należy dla każdej ligi wyznaczyć średnie zarobki oraz średni wiek grających w niej piłkarzy oraz liczbę zawodników. W obliczeniach należy pominąć piłkarzy mających powyżej 100 kg wagi.

W wynikowym zbiorze (3) powinny znaleźć się atrybuty:

- identyfikator ligi
- średnie zarobki piłkarzy w lidze
- średni wiek piłkarzy w lidze
- liczba zawodników w lidze

### Program Hive (5)

Działając na wyniku zadania MapReduce oraz zbiorze danych `datasource4` należy wyznaczyć oddzielnie dla każdego poziomu (`league_level`), trzy ligi z piłkarzami o największych średnich zarobkach.

Ostateczny wynik (6) powinien zawierać następujące atrybuty:

- `league_id` – identyfikator ligi
- `league_name` – nazwa ligi
- `league_level` – poziom ligi
- `avg_wage` – średnie zarobki piłkarzy w lidze
- `avg_age` – średni wiek piłkarzy w lidze
- `count_players` – liczba zawodników

Cyfry w nawiasach odnoszą się do cyfr wykorzystanych na graficznej reprezentacji projektu – patrz opis.