

# NBA & NCAA



L'impatto dei college sugli aspetti del gioco professionistico

Marco Faretra

Gabriele Marini

# Obiettivi: Domande di partenza



Nel vasto mondo dei college, esiste una gerarchia di valori tra di essi?

Un ragazzo di 17 anni a quale college dovrebbe iscriversi per poter diventare un professionista?

L'esperienza all'estero può essere migliore?

Quale college hanno scelto i grandi campioni?

Esiste un college migliore di un altro, se voglio specializzarmi?

# Sorgenti

Non abbiamo trovato dataset già pronti allo scopo, allora si lavora di “raschietto”...



**WIKIPEDIA**  
The Free Encyclopedia



# Dati: Qualche numero...

Ogni record (giocatore) contiene mediamente **90K** statistiche atomiche.

L'intero dataset contiene all'incirca **407M** statistiche.

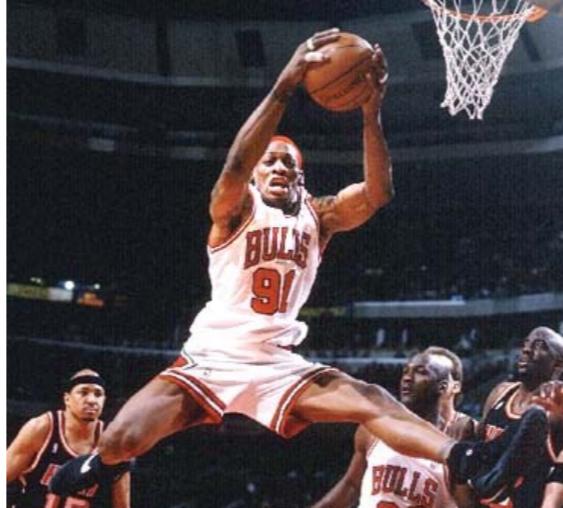
Abbiamo analizzato fino a **1,2B** statistiche, **~47M** di partite.

Abbiamo stimato i tempi di calcolo con un dataset contenente **10,8B** statistiche, ovvero circa **mezzo miliardo** di partite.

# Obiettivi: Categorie



Tiratori



Rimbalzisti



All around



+/- guys



Difensori

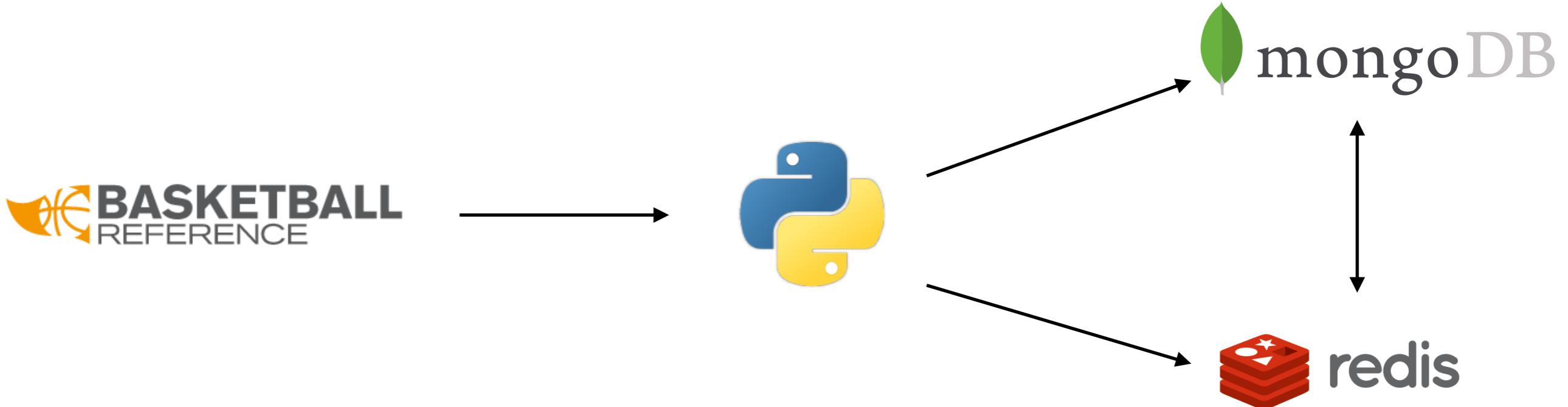


Attaccanti

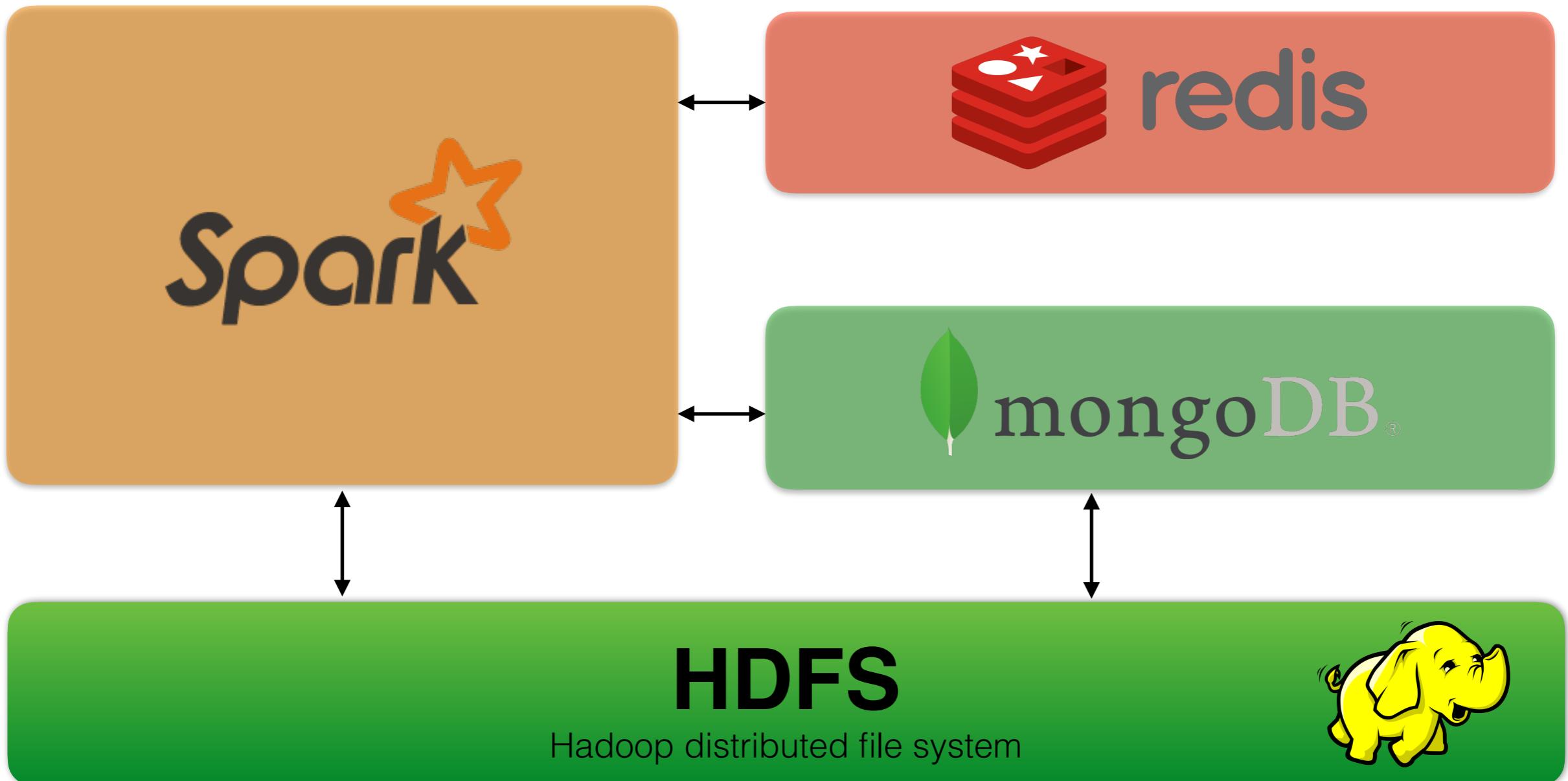
# Persistenza dei dati

Dopo diverse operazioni di pulitura dei dati,  
tutti i dati estratti sono stati riversati su  
MongoDB e Redis.

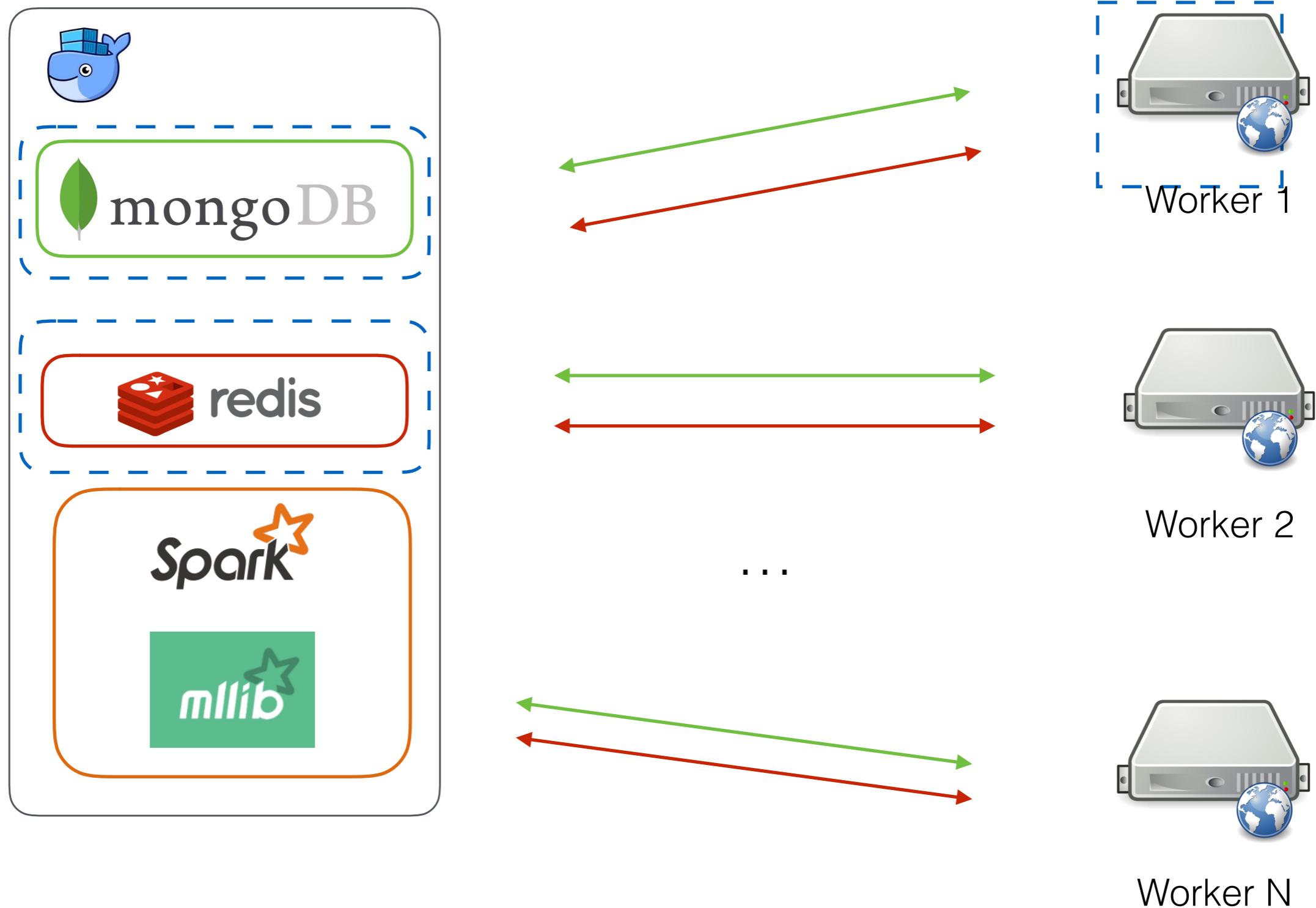
Con le dovute differenze di modellazione!



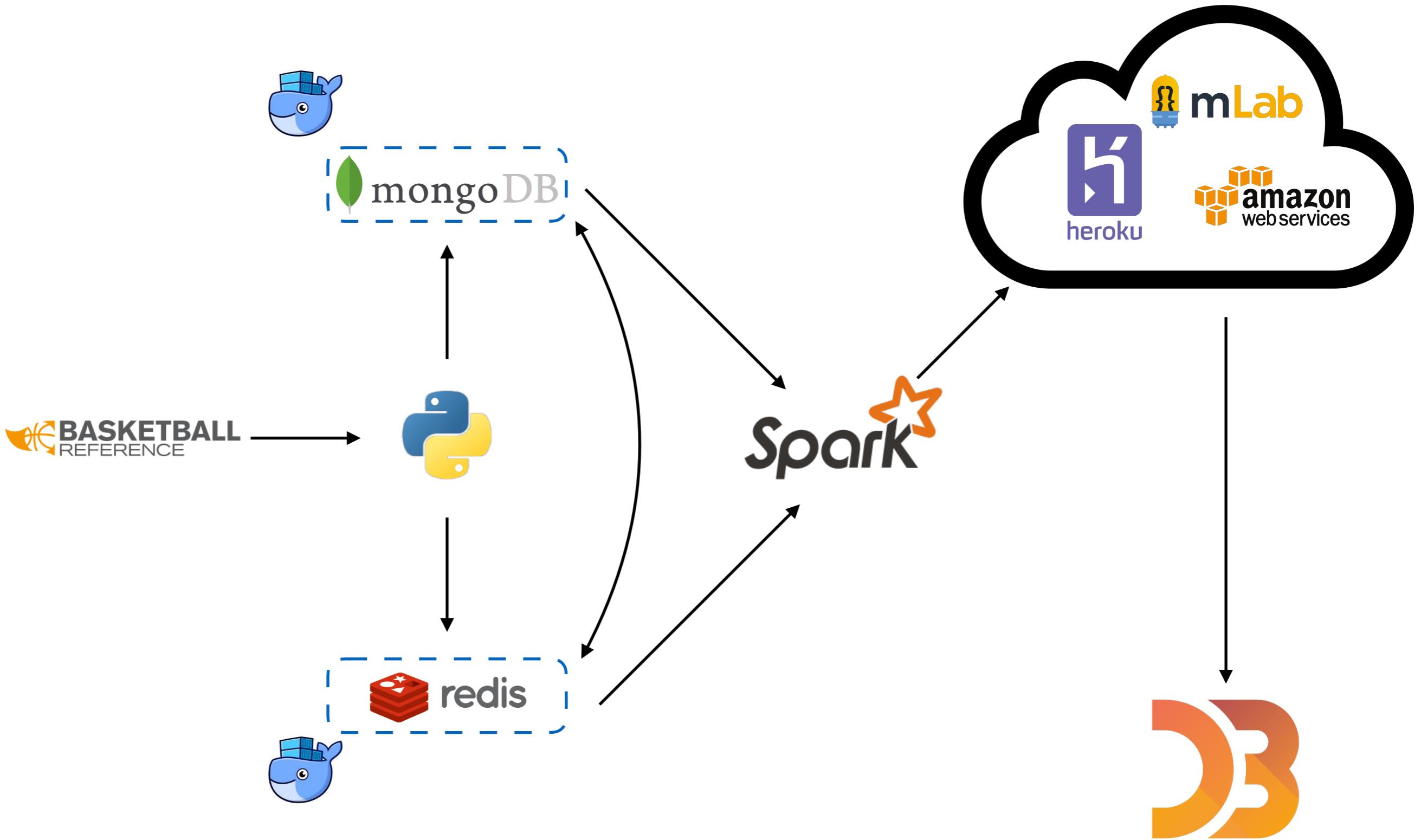
# Architettura



# Architettura



# Architettura



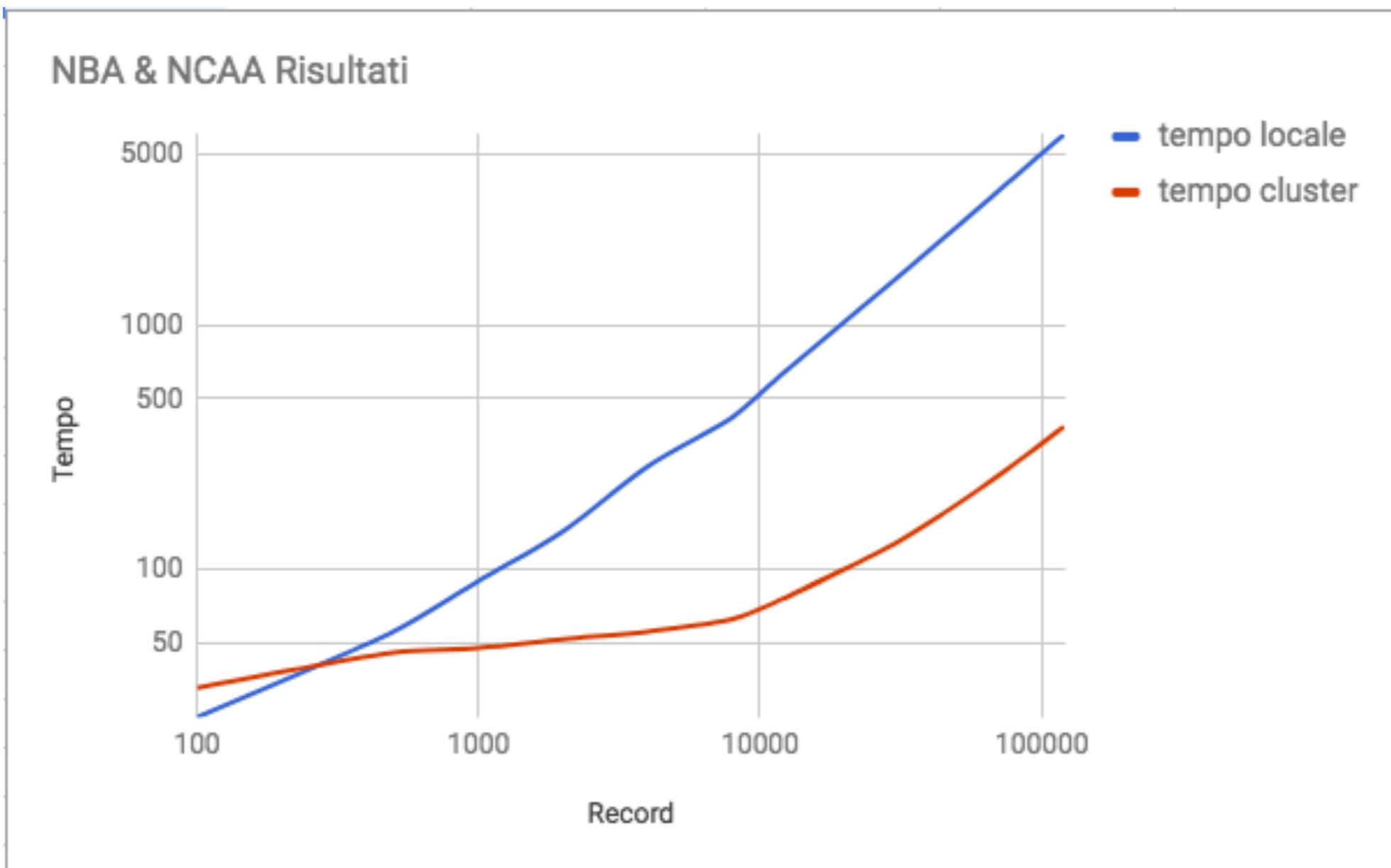
# Risultati: Tempi

Analisi temporale sia in locale, su una sola macchina, che su cluster EMR di AWS, composto da 20 nodi.

Facendo delle stime siamo riusciti a prevedere che con un dataset da 120.000 record, in locale il tempo di esecuzione è di 1 ora e 30 minuti, mentre su cluster il tempo di esecuzione è di 6 minuti, una riduzione del tempo di esecuzione pari a 15 volte!

In realtà le prestazioni in locale dovrebbero degradare molto più rapidamente, dando vita a dei tempi maggiori. Le prestazioni del cluster invece si dovrebbero degradare molto meno rapidamente, il guadagno dunque può anche essere maggiore di quello previsto.

# Risultati



# Risultati

I college alti nel ranking di una categoria, sono alti anche per altre categorie: i college migliori sono sempre stati i migliori, quelli peggiori sono sempre stati i peggiori.

Non esistono college veramente specializzati: un college cura in egual modo ogni aspetto. Specializzarsi non paga.

I giocatori tendono a scegliere comunque i college più blasonati, anche se questi sono molto distanti dal loro stato di origine.

Nella top10, 8 giocatori su 10 hanno frequentato un college. Non frequentare un college è una possibilità, tuttavia nei primi 100 giocatori, solo 3 non hanno frequentato il college.

# Risultati: College



**Complessivo**



**2-point**



**3-point**



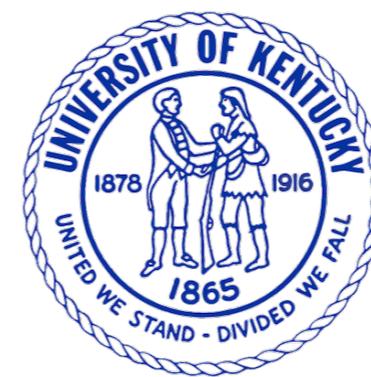
**All-around**



**+/-**



**Difensori**

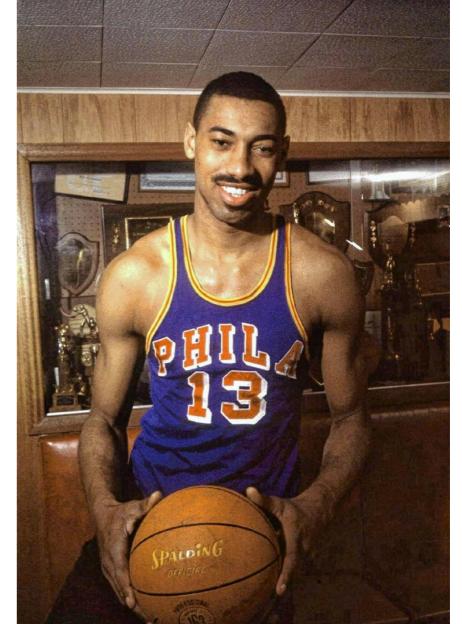
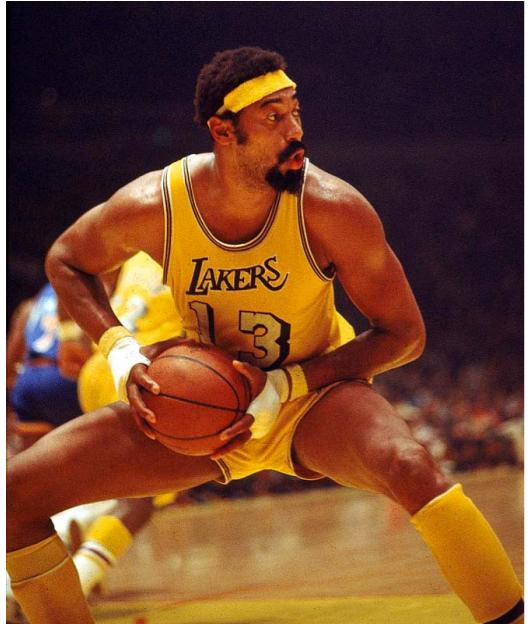


**Attaccanti**



**Rimbalzisti**

# Risultati: Giocatori



Chamberlain

**Complessivo**

Barkley

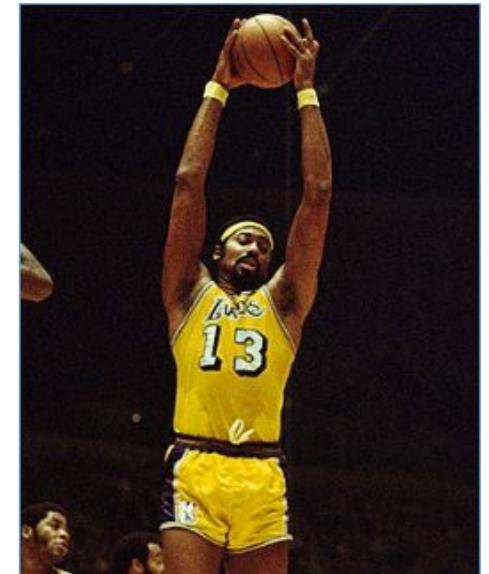
**2-point**

Kapono

**3-point**

Chamberlain

**All-around**



Green

**+/-**

Walton

**Difensori**

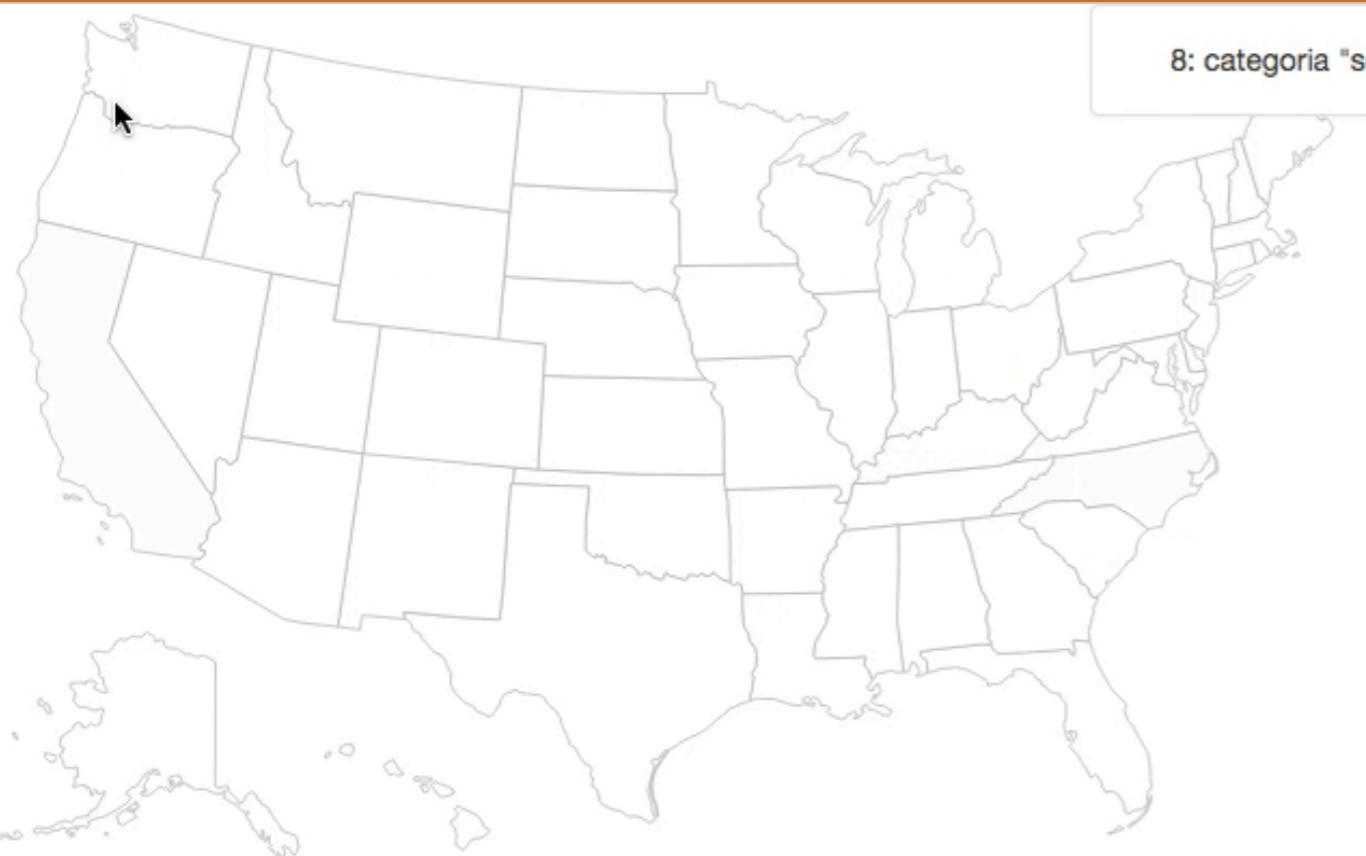
Chamberlain

**Attaccanti**

Chamberlain

**Rimbalzisti**

# Visualizzazione: Demo



8: categoria "score generale"



Grazie a tutti per l'attenzione.