DATA MANAGEMENT FOR DATA SCIENCE HW3

FRANCESCO PEZONE 1913202 & SABRIYE ELA ESME 1912580

Collection Name	#Docs	#Fields	Explanation
atp_matches_2018	2604	43	Single(1vs1) matches of multiple tournaments on 2018.
atp_matches_2019	2635	43	Single(1vs1) matches of multiple tournaments on 2019.
atp_matches_doubles_2018	1183	57	Double(2vs2) matches of multiple tournaments on 2018.
atp_matches_doubles_2019	1087	57	Double(2vs2) matches of multiple tournaments on 2019
atp_matches_futures_2018	16291	20	Futures matches of 2018 that allow for players to win career titles and improve their rankings.
atp_matches_futures_2019	15487	20	Futures matches of 2019 that allow for players to win career titles and improve their rankings.
atp_matches_qual_chall_2018	8428	43	ATP Challenger tour matches of 2018.
atp_matches_qual_chall_2019	7689	43	ATP Challenger tour matches of 2019.
atp_players	43542	6	ID, nationality, hand, first name, last name and birthdays of players of all of these matches
atp_rankings_current	17080	4	Rankings of the players on different ranking dates.

ATP: Association of Tennis Professionals

Mostly used fields:
atp_players collection:

- player_id, country_code, hand, first_name, last_name:

On different collections that shows matches:

- winner_id, loser_id, tourney_name, winner_age, loser_age, winner_ioc(country), loser_ioc(country), minutes

**for doubles matches there are two winner and loser fields

	CHARACTERISTICS
NOSQL	• Flexibility(unstructured) • Preferred when data need a flexible schema • For developers: complex expresiveness • Use when it's more important to have fast data than correct data • Base (Basically Available, Soft state, Eventually Consistent) is a model of many NoSQL systems
SQL	• Non-flexible(structured) • Preferred when you want to use joins and execute complex queries • For developers: simple expressiveness • It should be used when data validity is super important. • ACID(Atomicity, Consistency, Isolation, and Durability) is a standard for RDBMS