

## Homework Assignment - SQL

**SQL:** Please refer to the tutorial slides.

**Database Description:** Please refer to the tutorial slides.

**Database Installation:** Please refer to “MySQL\_Installation.pdf”.

**Database Creation:** Please refer to “Tennis\_Database\_Creation.pdf”.

**Please do not modify the database after its complete creation, use queries only for data fetching.**

The database you will be writing queries about contains information concerning the Association of Tennis Professionals. It contains information about players, matches, and tournaments concerning the ATP. If you are unfamiliar with tennis terminology (especially concerning scoring), a quick description can be found on Wikipedia at

<https://en.wikipedia.org/wiki/Tennis#Manner>

**Several assumptions about the database are included below:**

- The source of the data used for this database only contained information about the ATP (men’s tennis) and not the WTA (women’s tennis). All players involved are male; we have no information pertaining to women’s singles or doubles, or mixed doubles.
- While it should be clear that Player’s PID and Tournament’s TID field should be unique, this is also true for Match’s MID field; no IDs are re-used throughout different tournaments. Thus, a Match’s MID field uniquely represents it across all Matches known to exist. This avoids having to compare Tournament IDs as well as Match IDs for several of the queries.
- When a player participates in a tournament, he first has to register. For the purposes of our database, this includes receiving a registration number for the tournament. Because the ATP manages all of the tournaments in our database, we assume that a registration number is unique throughout all of the tournaments. This means that given only a registration number, we can determine the tournament it corresponds to 1.

- A Player's ID (Player. PID) is only used in our database for registration in tournaments. In all other locations, the Player is referred to by the RegistrNum (notably in the Winner attribute of MatchResults).
- Two players participating in doubles tennis as a team share the same registration number. Note that several players register for both singles and double tennis in the same tournament, players in this category get two separate registration numbers.
- Both a player registered for singles play or two people registered for doubles play under the same number (i.e., are playing with each other) can both be referred to as a team: a team of one or a team of two.
- In most tournaments, the top fraction of the players registering are seeded, to help spread out the players expected to perform well (this makes the later matches more interesting!). Seed values are given on a per-tournament basis. Players who do not receive the seed in a particular tournament have a NULL value instead.
- The Tournament relation stores the number of rounds the tournament has as an attribute, which is an integer. The Round attribute of the Match relation is similar. A value of 1 corresponds to the first round, 2 to the second, and so on. Match. Round does not say "Quarterfinals," "Semifinals," or "Finals." You will have to figure out how to compute what rounds correspond to these. Don't try to hard-code values in (i.e., if you look at the database and realize that the US open singles had a total of 7 rounds, so the Semifinals is round 6). Rather, use a nested query to help.
- Each entity in the Set relation contains information about a particular set of some match. The attributes involving the "Winner" refer to the winner of the match, not necessarily the set (similarly for the "Loser" attributes). In cases where the set was determined by a tie-breaker, the TieBreaker relation has WinnerTB and LoserTB, which contain the points won in the tie-breaker game. Again, the "Winner" column of the TieBreaker refers to the winner of the match, not necessarily the set.
- Sometimes players will need to retire from a match, which counts as a forfeit. If this happens, we keep records for all completed sets of the match. For the set in which the player retired, the WinnerGames and LoserGames attributes hold the number of games won and lost. The RetiredMatch relation holds all Match IDs where the loser retired.

**Solve the following SQL queries (Database used- tennis)**

1. Find out all the tournament names with tournament type, surface type, and the number of rounds it has. Sort the results in ascending order by the number of rounds.
2. List the names of all the players from Australia in the database.
3. Find out all the tournaments having playing surfaces mentioned as 'Hard' and rounds more than 5. Print the name of the tournament, the tournament type, the start and end dates, the length of the tournament in days, and the number of rounds.
4. Find out the total number of matches played in each tournament. List the Tournament name, Tournament type, corresponding year, and the total number of matches played in that tournament.
5. Find out the names, tournament types, and lengths (in days) of all tournaments that were longer than two weeks.
6. List the names of all players who were assigned to seed for any tournament (doubles or singles).
7. List the names of all male Australian players who were registered for the 2007 French Open Doubles.
8. List the players who played doubles tennis at Wimbledon in 2007.
9. List the players who played singles tennis at the Australian Open in 2007.
10. List the players who played doubles tennis on clay surfaces.

11. List all the Match IDs of the US Open in 2007 and the winner's name of each match.
12. Find the name of the player who has registered for the maximum number of tournaments in a year. Get the number of tournaments he has registered for. [Registration for both singles and doubles in the same tournament counts as two tournaments.]
13. List the names of all partners of a doubles tennis player who was also seeded in a singles tournament at some point. This means that we want all players x who have partners y where partner y was seeded in men's singles.
14. For all singles quarterfinal, semifinal, and final round matches that only took 3 sets, list the Tournament Name, Year, Winner of the Match, and the Match score (For the match score, print two columns per set, displaying the number of games each player won. Call these columns something meaningful, and ignore any tiebreaker results).
15. List the names of all players who have reached it to at least the quarter-finals (singles) of the tournaments in 2007. They could have made it to the finals for each tournament that is considered.