UNIVERSITÀ DEGLI STUDI DI TRIESTE

ANNO ACCADEMICO 2021/2022

“CIRCUITO TENNISTICO”

Progetto Basi di Dati

Giovanni Zanin

IN0500810

Presentazione

Col presente progetto viene proposta una possibile organizzazione dei dati relativi a un circuito tennistico, ispirato a quello ATP (Association of Tennis Professionals).

Il fulcro dello schema sono i giocatori, che nel corso di un anno guadagnano punti in base ai risultati ottenuti nei tornei a cui partecipano. In particolare, ogni tennista ha un proprio punteggio per il singolare e uno per il doppio, equivalente ai punti ottenuti nelle due specialità negli ultimi 365 giorni (dopo un anno i punti "scadono" e vengono sottratti al giocatore).

Ogni torneo è ad eliminazione diretta ed ha un suo valore in punti (tipicamente 2000, 1000, 500, 250 o 120) che equivale a quelli assegnati al vincitore di ogni singola edizione. I punti assegnati agli altri turni equivalgono al 60% dei punti che il tennista avrebbe ottenuto venendo eliminato al turno successivo (arrotondando all'intero punteggi decimali). Ad esempio, per un torneo di valore 1000 devono essere assegnati 1000 punti al vincitore, 600 al finalista sconfitto, 360 agli eliminati in semifinale etc. I tornei hanno cadenza al più annuale e si svolgono in una nazione ciascuno. Un torneo di singolare è composto da sole partite di singolare e un torneo di doppio da sole partite di doppio.

Ogni partita dunque "fa parte" dell'edizione di uno specifico torneo, viene giocata in una certa arena (facente parte del circolo del torneo per cui vale la partita) e ai giocatori partecipanti vanno assegnati i punteggi che valore del torneo e turno della partita comportano (l'assegnazione punti può essere fatta ogni volta che un giocatore viene sconfitto, quindi eliminato, da un torneo oppure vince la finale).

Oltre ai tennisti le personalità principali del circuito sono gli arbitri e gli allenatori (qui verrà considerato che ogni allenatore possa essere coach di al più un tennista che a sua volta potrà collaborare con al più un allenatore).

Ciascuna persona è associata ad una nazione di appartenenza.

Un tennista può avere anche uno o più sponsor.

Alcune operazioni d'interesse

Aggiungere una partita	Interattiva	30/giorno
Aggiornare punteggi	Batch	60/giorno
Aggiungere un tennista	Interattiva	50/anno
Visualizzare il ranking (classifica dei tennisti in base al loro punteggio)	Batch	1000/giorno
Visualizzare informazioni coach	Interattiva	100/giorno
Creazione edizione di un torneo	Interattiva	200/anno
Visualizzazione informazioni tennista	Interattiva	200/giorno

N.B.: Lo scopo principale del database è quello di contenere e gestire dati relativi a tennisti, non quello di organizzare tornei; tuttavia questi possono essere regolarizzati in fase di costruzione delle tabelle con opportuni comandi CHECK o segnalazioni di errore (per controllare ad esempio che ogni edizione abbia quattro quarti di finale, due semifinali eccetera o che i giocatori vincenti in una semifinale giochino anche nella finale della stessa edizione).

Glossario Termini

Viene qui fornita una rapida descrizione delle entità e delle relazioni usate in un primo schema concettuale. Negli eventuali attributi sono stati sottolineati, per le entità, gli identificatori (alcuni identificatori potrebbero provenire da relazioni di cui l'entità fa parte).

Nome	Tipo	Descrizione	Attributi
ALLENATORE	E	Persona che può essere coach di un tennista	Nome, Cognome
ARBITRO	E	Persona che conduce una partita tra tennisti	Codice identificativo, nome, cognome
ARBITRAGGIO	R	Relazione tra una partita ed un arbitro	
ARENA	E	Arena in cui si possono svolgere partite	Nome, Circolo, capienza (numero spettatori ospitabili)
COACH	R	Collaborazione tennista-allenatore	Data di inizio collaborazione
COMPOSIZIONE	R	Relazione tra partita ed edizione di appartenenza (sinonimo: "parte di")	
CIRCOLO	R	Relazione tra torneo e arene in cui vengono giocate le sue partite	
EDIZIONE	E	Occorrenza di un torneo	Anno, Torneo, periodo di svolgimento
LUOGO	R	Relazione tra partite e arene in cui hanno luogo	
NAZIONALITÀ	R	Relazione tra persona e stato di appartenenza	
NAZIONE	E	Una delle 193 nazioni del mondo	Nome
OCCORRENZA	R	Relazione tra torneo e sue edizioni	
PARTITA	E	Sfida tra due o quattro giocatori (uno contro uno per le partite di singolare due contro due per quelle di doppio) valida per l'edizione di un certo torneo (sinonimo: incontro)	Numero, Edizione, punteggio (giochi e set), turno (fase), data
PERDENTE	R	Giocatore che perde una partita (sinonimo: sconfitto)	
SEDE	R	Nazione in cui un torneo ha sede	
SPONSOR	E	Ente o azienda che sponsorizza tennisti	Nome
SPONSORIZZAZIONE	R	Collaborazione tra sponsor e tennista	
TENNISTA	E	Partecipante a partite e tornei di tennis (sinonimo: giocatore)	Codice identificativo, nome, cognome, punti (di singolare e di doppio), titoli vinti (di singolare e di doppio)
TORNEO	E	Competizione tennistica con cadenza al più annuale	Codice identificativo, nome, valore, tipo di terreno su cui le partite del torneo vengono giocate
VINCITORE	R	Giocatore che vince una partita	

Schema Entity-Relationship

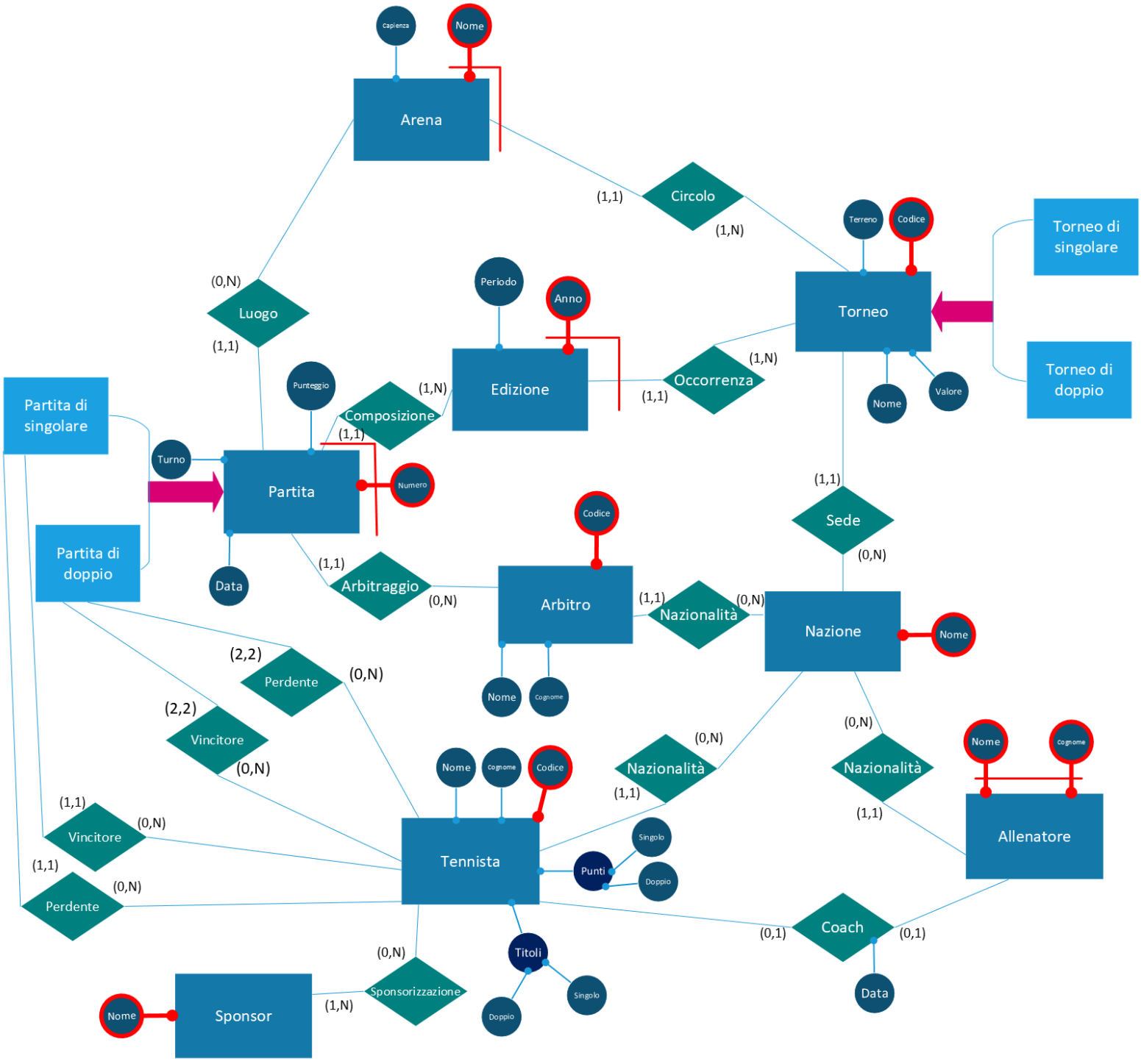


Tavola dei volumi

La seguente tabella indica gli approssimativi volumi delle entità e delle relazioni dello schema concettuale.

N.B.: per quanto riguarda il volume delle partite è presumibile che partite molto lontane nel passato, quindi di minor interesse, vengano rimosse dal database ed eventualmente memorizzate in un altro tipo di archivio.

Nome	Tipo	Volume
ALLENATORE	E	1000
ARBITRO	E	1000
ARBITRAGGIO	R	400000
ARENA	E	2000
COACH	R	1000
COMPOSIZIONE	R	400000
CIRCOLO	R	2000
EDIZIONE	E	8000
LUOGO	R	400000
NAZIONALITÀ ALL	R	1000
NAZIONALITÀ ARB	R	1000
NAZIONALITÀ TEN	R	1000
NAZIONE	E	193
OCCORRENZA	R	8000
PARTITA	E	400000
PERDENTE DOP	R	400000
PERDENTE SIN	R	200000
SEDE	R	200
SPONSOR	E	100
SPONSORIZZAZIONE	R	2000
TENNISTA	E	1000
TORNEO	E	200
VINCITORE DOP	R	400000
VINCITORE SIN	R	200000

Analisi delle ridondanze

Non sono presenti attributi direttamente derivabili da altri nella stessa tabella.

Il circolo di appartenenza di una data arena è un caso di ridondanza dovuta a ciclo, essendo ricavabile dal torneo a cui appartengono le partite aventi luogo in tale arena. La ridondanza non è stata eliminata, in modo da mantenere l'identificazione di un'arena basata su nome e circolo di appartenenza.

Per quanto riguarda gli attributi derivabili da altre entità si noti che per quanto riguarda l'entità "Edizione" gli attributi "Anno" e "Periodo" sono direttamente calcolabili dalla data delle partite giocate in quell'edizione del torneo. Entrambe le ridondanze saranno mantenute.

Per quanto riguarda "Anno" è stato scelto di mantenere l'attributo in quanto considerato fondamentale per l'identificazione di una data Edizione.

L'attributo "periodo" viene invece mantenuto in seguito a una breve analisi sul peso degli accessi con e senza ridondanza dato dalle azioni:

- A: aggiungere una partita (30 volte al giorno)
- B: stampare i dati di un'edizione, compreso il periodo (100 volte al giorno)

CON RIDONDANZA

Azione A

Numero accessi	Costrutto	Concetto	Tipo di accesso
1	Entità	Partita	Scrittura
1	Relazione	Composizione	Scrittura
1	Entità	Edizione	Lettura
1	Entità	Edizione	Scrittura

N.B.: gli accessi all'entità "edizione" sono necessari per verificare ed eventualmente aggiornare il periodo di svolgimento dell'edizione.

Azione B

1	Entità	Edizione	Lettura

Complessivamente 90 accessi al giorno di tipo scrittura e 130 di tipo lettura. Contando doppi quelli in scrittura 310 accessi giornalieri complessivi.

SENZA RIDONDANZA

Azione A

Numero accessi	Costrutto	Concetto	Tipo di accesso
1	Entità	Partita	Scrittura
1	Relazione	Composizione	Scrittura

Azione B

1	Entità	Edizione	Lettura
30	Relazione	Composizione	Lettura
30	Entità	Partita	Lettura

N.B.: è necessario accedere ad ogni partita dell'edizione per calcolarne il periodo di svolgimento.

Nel complesso più di 6000 accessi giornalieri, molti più del caso con ridondanza.

Altri casi di ridondanza sono gli attributi “Titoli” e “Punti” dell’entità “Tennista”, derivabili rispettivamente dalle partite di turno finale vinte e dalle partite generiche giocate dal tennista.

Anche in questo caso dall’analisi (con calcoli più complessi viste necessarie le distinzioni tra partite di singolare e di doppio e partite di fase finale e non, qui omessa) appare preferibile la presenza di ridondanza, visto l’alto numero giornaliero previsto di operazioni di visualizzazione dei dati di un tennista (in presenza di ridondanza gli attributi “Titoli” e “Punti” non andranno calcolati ogni volta).

Eliminazione delle generalizzazioni

La generalizzazione “Torneo” per “Torneo di singolare” e “Torneo di doppio” verrà eliminata integrando i figli nel padre, al quale verrà aggiunto un attributo “Tipo”, in quanto gli accessi sono contestuali.

La generalizzazione “Partita” verrà eliminata integrando il padre nei figli più che per una ragione legata agli accessi per il fatto che le partite di singolare e quelle di doppio richiedono un diverso numero di tennisti partecipanti, e quindi un diverso numero di relazioni (una soluzione meno elegante, in fase di costruzione delle tabelle, sarebbe potuta essere quella di accorpare i figli nel padre ed aggiungere gli attributi “Tipo partita”, “Vincitore1”, “Perdente1”, “Vincitore2” e “Perdente2”, mantenendo valore NULL per questi ultimi due attributi in caso di partita di singolare).

Accorpamenti e partizionamenti

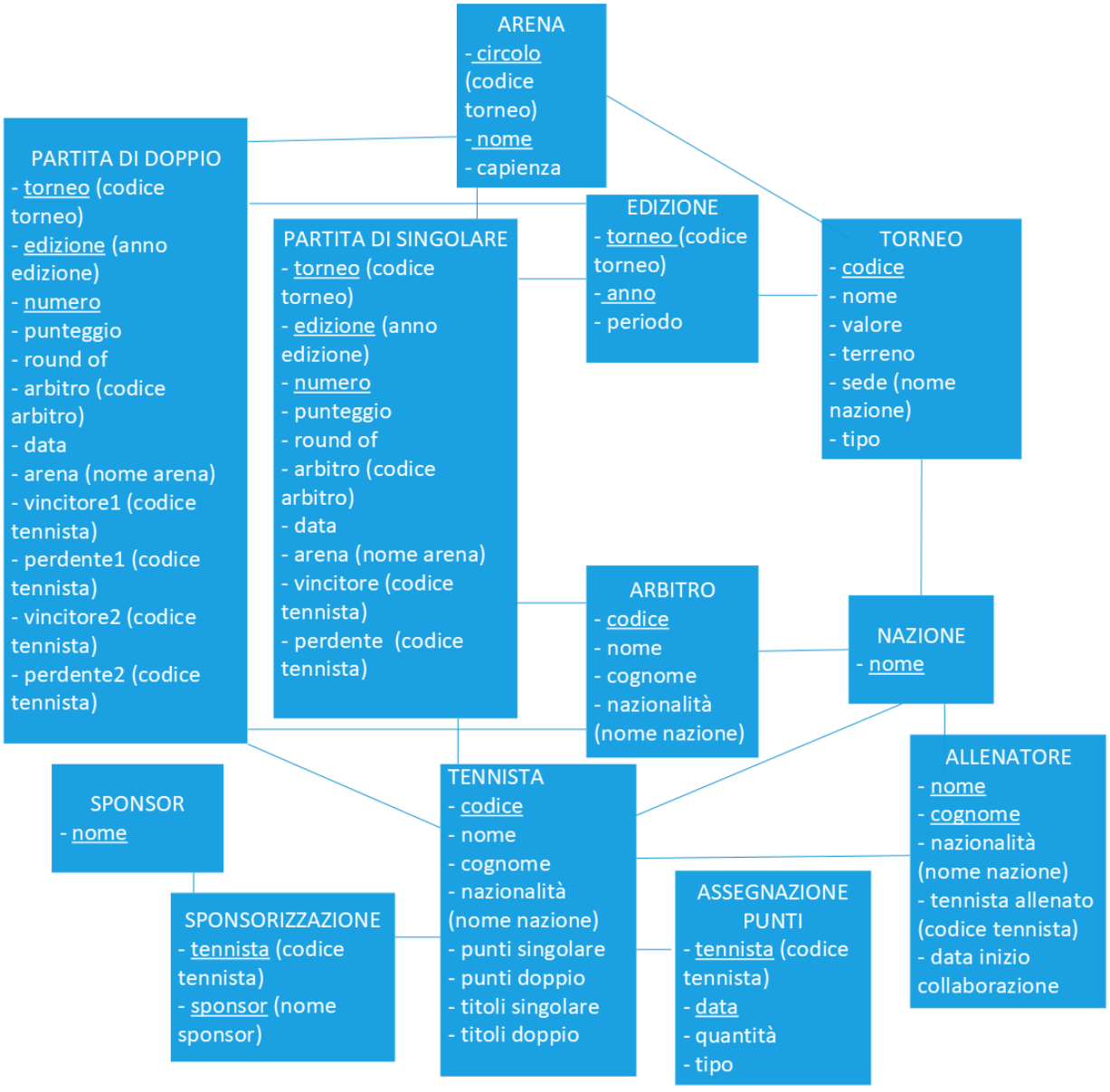
Si può valutare la possibilità di accorpare le entità “Allenatore” e “Tennista”. Qui è stato scelto di mantenerle separate per evitare di appesantire troppo l’entità “Tennista”, che ha una maggior frequenza di accesso.

Schema logico

Lo schema logico finale è ottenuto sulla base dei miglioramenti sopra detti dello schema concettuale ed eliminando relationship uno a molti e uno a uno traducendole in attributi (ad esempio tennisti perdenti e vincenti delle partite sono diventati attributi delle partite stesse).

Tra parentesi sono indicati i vincoli di integrità referenziale che alcuni attributi devono rispettare. Si noti che ogni partita ha come identificatore esterno l'edizione del torneo (assieme al numero della partita, che è invece interno). A sua volta l'edizione è identificata da anno e codice del torneo, dunque anche sulle partite ci sarà un vincolo di integrità referenziale legato al torneo ("torneo" deve essere tra i codici dell'entità torneo e "torneo"+"anno" deve comparire come edizione, qui il legame è stato rappresentato solo indirettamente attraverso "Edizione").

Per rispettare il requisito sulla durata della validità dei punti dei tennisti è stata creata una tabella di audit "Assegnazione Punti". Qui verrà tenuta traccia degli incrementi nei punti di singolare e doppio dei giocatori (con opportuni trigger), in modo che ad un anno di distanza questi possano essere rimossi.



Normalizzazione

Non sono presenti attributi multipli, dunque il database è in prima forma normale.

Gli attributi di ogni tabella descrivono un'unica entità e non sono presenti colonne calcolate direttamente da altre appartenenti alla stessa tabella, dunque sono presenti anche seconda e terza forme normali (gli attributi "Tennista allenato" e "Data inizio collaborazione" dell'entità "Allenatore" si riferiscono in realtà ad una relazione tra allenatore e tennista, tale relazione però, supponendo che ogni allenatore abbia in corso al più una collaborazione, è stata accorpata all'entità "Allenatore").

Progettazione fisica

Per quanto riguarda la scelta degli indici in generale si può scegliere di creare indici secondari sulle chiavi primarie.

Tuttavia, per le entità "Tennista", "Arbitro" e "Torneo", per cui gli identificatori sono dei codici, si può scegliere di costruire indici secondari sui nomi (o cognomi per arbitri e tennisti), che potrebbero ragionevolmente essere un attributo più usato per la ricerca.

Anche per le partite si potrebbe valutare di porre indici secondari sui tennisti partecipanti o sulla data, due attributi su cui la ricerca degli incontri può essere frequente. Ciononostante, si è ritenuto che, per quanto consueti, questi tipi di ricerche non saranno frequenti quanto quelle basate sull'edizione del torneo a cui appartengono le partite, quindi la scelta è stata quella di mantenere l'indice secondario sulle sole chiavi primarie.

Query SQL

Per creare la tabella partitaDiSingolare (le altre tabelle saranno create in modo analogo):

```
CREATE TABLE IF NOT EXISTS partitaDiSingolare (torneo char(5), edizione year,
numero smallint, punteggio varchar(20), roundOf smallint NOT NULL CHECK (roundOf
IN (2,4,8,16,32,64,128,256)), arbitro char (4), data date, arena varchar (30),
vincitore char (5) NOT NULL , perdente char (5) NOT NULL ,

PRIMARY KEY (torneo, edizione, numero),
FOREIGN KEY (torneo, edizione) REFERENCES edizione(torneo, anno),
FOREIGN KEY (torneo) REFERENCES torneo (codice),/*in realtà questo vincolo è
ridondante dato che già presente in "edizione"*/
FOREIGN KEY (arbitro) REFERENCES arbitro(codice),
FOREIGN KEY (arena, torneo) REFERENCES arena (nome, circolo),
FOREIGN KEY (vincitore) REFERENCES tennista (codice),
FOREIGN KEY (perdente) REFERENCES tennista (codice));
```

N.B.: il turno per cui vale la partita (quarti, semifinale etc.) è stato espresso come "round of n" dove n è la potenza di 2 che indica il numero di tennisti ancora in gara (ottavi= round of 16, sedicesimi= round of 32 etc.).

Per creare e mostrare la vista che visualizza il ranking del singolare:

```
CREATE VIEW rankingS AS
SELECT nome, cognome, puntiSingolare FROM tennista ORDER BY puntiSingolare;

SELECT * FROM rankingS;
```

Per creare ed eseguire la stored procedure per rimuovere giornalmente dal conteggio punti di singolare di ogni giocatore i punti scaduti (più vecchi di un anno):

```
DELIMITER $$

CREATE PROCEDURE rimozionePuntiScadutiS()
BEGIN

UPDATE tennista t
SET t.puntiSingolare=
t.puntiSingolare-
(SELECT quantità FROM assegnazionePunti a
WHERE (day(a.data)=day(NOW())
 AND month(a.data)=month(NOW())
 AND year(a.data)=year(NOW())-1
/*i punti risalgono a giorno e mese odierni un anno fa*/
 AND a.tennista=t.codice
 AND a.tipo='s'))

WHERE t.codice= (SELECT tennista FROM assegnazionePunti a
WHERE (day(a.data)=day(NOW())
 AND month(a.data)=month(NOW())
 AND year(a.data)=year(NOW())-1
 AND a.tipo='s') );

END $$

DELIMITER ;

CALL rimozionePuntiScadutiS();
```

Per creare il trigger che all'inserimento di ogni partita di singolare aggiorni i punti di singolare del tennista sconfitto (3/5 del valore del torneo in caso di eliminazione in finale, 9/25 in semifinale e così via) e, in caso di partita di finale, assegni un titolo e l'intero valore in punti del torneo al vincitore:

```
CREATE TRIGGER aggiornamentoPerPartiteS AFTER INSERT
ON partitaDiSingolare
FOR EACH ROW
BEGIN

update tennista g
SET
puntiSingolare=puntiSingolare+ROUND((POW((3/5),(log(2,NEW.roundOf))))*(SELECT
valore FROM torneo t WHERE NEW.torneo=t.codice))
WHERE g.codice=NEW.perdente;

INSERT INTO assegnazionePunti VALUES (NEW.perdente, NEW.data,
```

```
ROUND((POW((3/5),(log(2,NEW.roundOf))))*(SELECT valore FROM torneo t WHERE
NEW.torneo=t.codice)), 's');

IF NEW.roundOf=2 THEN /*in caso di partita di finale va aggiornato anche il
vincitore*/
update tennista g
SET puntiSingolare=puntiSingolare+(SELECT valore FROM torneo t WHERE
NEW.torneo=t.codice)
WHERE g.codice=NEW.vincitore;
update tennista g
SET titoliSingolare=titoliSingolare+1
WHERE g.codice=NEW.vincitore;

INSERT INTO assegnazionePunti VALUES (NEW.vincitore, NEW.data, (SELECT valore
FROM torneo t WHERE NEW.torneo=t.codice), 's');
end if;

    end;
```

Per creare trigger che segnali un errore in caso di tentativo di aggiunta di una partita di singolare in un torneo di doppio:

```
CREATE TRIGGER controlloFormulaS
BEFORE INSERT ON partitaDiSingolare
FOR EACH ROW
BEGIN

IF ((SELECT tipo FROM torneo WHERE codice=NEW.torneo)='d')
THEN signal sqlstate '45001'
SET message_text = 'Impossibile inserire partita di singolare in torneo di
doppio' ;
END IF ;

END;
```

Per creare una vista che mostri per ogni nazione il numero di tennisti nella top 100 del ranking del singolare:

```
CREATE VIEW tennistiInTop100 AS
    SELECT nazionalità, count(*) AS numeroTennisti FROM (SELECT * FROM tennista
ORDER BY puntiSingolare LIMIT 0,100) t GROUP BY t.nazionalità ORDER BY
numeroTennisti;
```