

## Progetto di laboratorio (SQL)

Studiare con attenzione la documentazione riguardante il database MusicBrainz ([http://musicbrainz.org/doc/MusicBrainz\\_Database](http://musicbrainz.org/doc/MusicBrainz_Database)) e in particolare lo schema del database ([http://musicbrainz.org/doc/MusicBrainz\\_Database/Schema](http://musicbrainz.org/doc/MusicBrainz_Database/Schema)).

Il database musicbrainz viene fornito in una versione ridotta e in una versione completa. Entrambe hanno lo stesso schema relazionale.

*Per usare la versione ridotta:*

Sulla pagina del corso è a disposizione una versione ridotta del database originale. Si vedano i file *musicbrainz\_schema.sql* e *musicbrainz\_data.sql* contenenti rispettivamente lo schema e i dati del database.

- a) importare lo schema del database eseguendo in pgAdmin lo script *musicbrainz\_schema.sql*
- b) importare i dati del database eseguendo in pgAdmin lo script *musicbrainz\_data.sql*

*Per usare la versione completa:*

Sul server PostgreSQL del corso di studi è installata la versione completa di musicbrainz. Per utilizzarla bisogna:

- a) collegarsi con pgAdmin tramite tunnel SSH oppure tramite l'interfaccia web dalle pagine riservate del corso di studi
- b) selezionare dall'elenco dei database *musicbrainz\_db*
- c) aprire il query tool di pgAdmin
- d) eseguire il comando `"set search_path = '$user', public, musicbrainz;"` (questo serve per evitare di qualificare con il prefisso musicbrainz ogni tabella delle query)

Infine risolvere le interrogazioni SQL riportate in questo documento.

**Per ogni interrogazione si richiede di:**

1. controllare nello schema relazionale quali sono le informazioni correlate a ciò che è richiesto;
2. scrivere quali informazioni e loro proprietà (chiave primaria o chiave esterna, valori nulli possibili, ecc.) si sono controllate al punto precedente;
3. scrivere i sottoproblemi attraverso cui si è arrivati a risolvere il problema posto dalla query. A ogni sottoproblema dovrebbe corrispondere una sottoquery della query finale.
4. se nel formulare una query si fa affidamento su assunzioni sui dati nel database (ad esempi valori non nulli, valori non duplicati), verificare sullo schema relazionale se tali assunzioni valgono effettivamente.
5. per le query nelle quali è indicato, formulare due versioni della query.
6. scrivere come si è controllato il risultato della query (ad esempio, riportare altre query SQL che considerate insieme permettono di verificare la correttezza del risultato).

**Avvertenze:**

- Verificare il funzionamento delle query usando il DBMS PostgreSQL sia sulla versione ridotta di musicbrainz che sulla versione completa disponibile sul server del corso di studi.
- Controllare nello schema relazionale la chiave primaria di ogni tabella usata.
- Utilizzare esclusivamente i costrutti SQL visti a lezione
- **Non usare** costrutti come OFFSET, LIMIT, ROWNUM, FETCH, TOP. **Non usare** viste.
- Non scrivere query inutilmente complicate. Nella valutazione del progetto si valuterà, oltre alla necessaria correttezza, anche la leggibilità e semplicità delle query.
- Ricordarsi che le query vanno scritte facendo riferimento allo schema della base di dati e non al suo contenuto; in altre parole, la query deve fornire il risultato corretto anche quando i dati (ma non lo schema) cambiano.

Consegnare un file di testo con estensione .sql che riporti nel nome i componenti del gruppo. Ad esempio, un gruppo composto dagli studenti di cognome A, B, C deve consegnare il file A\_B\_C.sql

Nel file consegnato ci dovrà essere, per ogni query:

- come commento il testo della domanda;
- come commento una discussione dei punti 1-6 riportati qui sopra;
- la/le query.

**Query 1:**

Contare il numero di lingue in cui le release contenute nel database sono scritte (il risultato deve contenere soltanto il numero delle lingue, rinominato "Numero\_Lingue").

**Query 2:**

Elencare gli artisti che hanno cantato canzoni in italiano (il risultato deve contenere il nome dell'artista e il nome della lingua).

**Query 3:**

Elencare le release di cui non si conosce la lingua (il risultato deve contenere soltanto il nome della release).

**Query 4:**

Elencare gli artisti il cui nome contiene tutte le vocali ed è composto da una sola parola (il risultato deve contenere soltanto il nome dell'artista).

**Query 5:**

Elencare tutti gli pseudonimi di Prince con il loro tipo, se disponibile (il risultato deve contenere lo pseudonimo dell'artista, il nome dell'artista (cioè Prince) e il tipo di pseudonimo (se disponibile)).

**Query 6:**

Elencare le release di gruppi inglesi ancora in attività (il risultato deve contenere il nome del gruppo e il nome della release e essere ordinato per nome del gruppo e nome della release)

**Query 7:**

Trovare le release in cui il nome dell'artista è diverso dal nome accreditato nella release (il risultato deve contenere il nome della release, il nome dell'artista accreditato (cioè artist\_credit.name) e il nome dell'artista (cioè artist.name))

**Query 8:**

Trovare gli artisti con meno di tre release (il risultato deve contenere il nome dell'artista ed il numero di release).

**Query 9:**

Trovare la registrazione più lunga di un'artista donna (il risultato deve contenere il nome della registrazione, la sua durata in minuti e il nome dell'artista; tenere conto che le durate sono memorizzate in millesimi di secondo) (scrivere due versioni della query con e senza operatore aggregato MAX).

**Query 10:**

Elencare le lingue cui non corrisponde nessuna release (il risultato deve contenere il nome della lingua, il numero di release in quella lingua, cioè 0, e essere ordinato per lingua) (scrivere due versioni della query).

**Query 11:**

Ricavare la seconda registrazione per lunghezza di un artista uomo (il risultato deve comprendere l'artista accreditato, il nome della registrazione e la sua lunghezza) (scrivere due versioni della query).

**Query 12:**

Per ogni stato esistente riportare la lunghezza totale delle registrazioni di artisti di quello stato (il risultato deve comprendere il nome dello stato e la lunghezza totale in minuti delle registrazioni (0 se lo stato non ha registrazioni) (scrivere due versioni della query).

**Query 13:**

Ricavare gli artisti britannici che hanno pubblicato almeno 10 release (il risultato deve contenere il nome dell'artista, il nome dello stato (cioè United Kingdom) e il numero di release) (scrivere due versioni della query).

**Query 14:**

Considerando il numero medio di tracce tra le release pubblicate su CD, ricavare gli artisti che hanno pubblicato esclusivamente release con più tracce della media (il risultato deve contenere il nome dell'artista e il numero di release ed essere ordinato per numero di release discendente) (scrivere due versioni della query).

**Query 15:**

Ricavare il primo artista morto dopo Louis Armstrong (il risultato deve contenere il nome dell'artista, la sua data di nascita e la sua data di morte) (scrivere due versioni della query).

**Query 16:**

Elencare le coppie di etichette discografiche che non hanno mai fatto uscire una release in comune ma hanno fatto uscire una release in collaborazione con una medesima terza etichetta (il risultato deve comprendere i nomi delle coppie di etichette discografiche) (scrivere due versioni della query).

**Query 17 (facoltativa):**

Trovare il nome e la lunghezza della traccia più lunga appartenente a una release rilasciata in almeno due paesi (il risultato deve contenere il nome della traccia e la sua lunghezza in secondi) (scrivere due versioni della query).