

Assignment 5: Normalization of Twitter Schema

SQL commands to make the tables for Tweets Data:

```
CREATE TABLE users(  
    uname VARCHAR PRIMARY KEY,  
    nickname VARCHAR,  
    bio VARCHAR,  
    picture VARCHAR,  
    followers BIGINT,  
    following BIGINT  
);  
  
CREATE TABLE countries(  
    countryId VARCHAR PRIMARY KEY,  
    country VARCHAR  
);  
  
CREATE TABLE places(  
    placeId BIGINT PRIMARY KEY,  
    place VARCHAR  
);  
  
CREATE TABLE locations(  
    locationId BIGINT PRIMARY KEY,  
    latitude DOUBLE PRECISION,  
    longitude DOUBLE PRECISION,  
    country BIGINT REFERENCES countries(countryId),  
    place BIGINT REFERENCES places(placeId)  
);  
  
CREATE TABLE tweets (  
    id BIGINT PRIMARY KEY,  
    date DATE,  
    hour TIME,  
    message VARCHAR,  
    favs BIGINT,  
    rts BIGINT,  
    lang VARCHAR,  
    url VARCHAR,  
    listed BIGINT,  
    uname VARCHAR REFERENCES users(uname),  
    location BIGINT REFERENCES locations(locationId)  
);
```

For better understanding on my solution, see next page.

TABLES: *The different color of the column header represents the table referencing by the foreign key.*

users

uname	nickname	bio	picture	followers	following
MALUMA	maluma	PARA CONTRATACIONES E INFORMACION contacto@malumamusik.com http://t.co/kX2IUvZCUP https://t.co/nIA6IEoOIY	Pic01552	2525716	190
Tanya Burr	TanyaBurr	British Blogger & YouTuber, Ambassador for Global Goals https://t.co/yWzpRD7iQo SNAPCHAT: tanyaburr	Pic57562	1813146	577
Yo Gotti	YoGottiKOM	The Greatest Story Ever Told Thru Music #TheArtofHustle Is Available Now! https://t.co/goCuZEG8kz #GottiNation	Pic25675	1405159	643

countries

countryId	country
US	United States
UK	United Kingdom
CA	Canada

The **countries** and **places** has its own table since it could have more added attributes based on the data set here:

<http://followthehashtag.com/datasets/free-twitter-dataset-usa-200000-free-usa-tweets/>

places

placeId	place
47	Ottawa
54	Cumberland
4658	Florida

id	population	total	Real Cover	Relative Interest
United Sta	310232863	164839	531.33958	67.88577719941658
Mexico	112468855	18110	161.02235	-1.809242495664371
Canada	33679000	5414	160.75299	-1.859937842442982
France	64768389	2	0.0308792	-32.1084112137608
United Kin	62348447	2	0.0320777	-32.1081856475484

locations

locationId	latitude	longitude	country	place
4524	28.35272725	-81.60341115999999	US	4658
2657	45.5088889	-73.5541667	CA	47
1549	37.2243853	-77.4006182	UK	54

tweets

id	date	hour	message	favs	rts	lang	url	listed	uname	location
12594	2016-04-16	02:46:24	Y DELE ... Viernes por la noche sin excusa alguna!! @ Mexico City, Mexico	2089	449	es	https://t.co/ /lpwfZqAtiu	120	MALUMA	2657
45867	2016-04-15	17:20:47	Let's Talk Money & OWNERSHIP... @ Manhattan, New York	83	74	en	https://t.co/ /hek7XpyC dr	45	Yo Gotti	4524

The original table tweet complies already the first two forms since it has only one single primary key (id) – 1NF and every relation that is 1NF with a single attribute key is automatically in the 2NF. To make the original table tweet complies to 3NF, it has to break into separate tables to remove the transitive dependency as shown above.