

Cassandra

Command can be run on the Cassandra Web at `localhost:3000` or in the docker container with:

```
sudo docker exec -it cass1 cqlsh
```

Note the `MATERIALIZED VIEWS` and `CUSTOM INDEX` are already created in the [startup script](#) and don't need to be run again.

Queries

All queries can also be found [here](#).

1. Listing all the posts made by an account.

```
SELECT content FROM twitter.tweets where author_id = 233248636;
```

Rows
content
Me and @grantmickelson are coloring. http://t.co/XtBtM8F
Ahhhh!! Story of Us just came on the radio! I'm dancing around my apartment!!
The day @selenagomez and I learned a valuable lesson from a pillow. http://t.co/dj59rgJ
We're at rehearsals watching videos of the tornadoes in Tuscaloosa. Our love and hopes and thoughts and prayers go out to everyone affected.
It's @grantmickelson's birthday- Abraham Lincoln cardboard cutout: check. http://t.co/mhUuvvV
Headed back to Nashville after the MET Ball in NYC. Wondrous, magical, fantastic gowns EVERYWHERE. I had a really wonderful time.
...And I got to tell Jimmy Fallon how awesome his 'Whip My Hair' sketch with Bruce Springsteen was.
"Ya gotta keep your head up OHHH, but you can let your down EHHH" @andygrammer on the radio- Singing along on the way to rehearsal. Smiling!"
Friends marathon after a long day of rehearsals. This time tomorrow night, the 'Mean' video will be out!!
Excited about tonight's tv line up! Gonna watch the debut of the Mean video at 10pm onrnCMT. Then @iamstevenT and @thecivilwars are on Leno!

2. Find the 100 accounts with the most followers

```
# 2. Find the 100 accounts with the most followers
rows = session.execute('SELECT user_id,follower_len FROM twitter.most_fc')
sorted_rows = dict(sorted(dict(rows).items(), key=lambda item: item[1]))
most_follows = list(sorted_rows.keys())[-100:]
print("Top 100 most followed accounts:",most_follows)
```

```
Top 100 most followed accounts: [817268, 173732041, 14506809,
158804228, 358775055, 45416789, 302282272, 65913144, 261001122,
14246001, 7702232, 101204352, 280365428, 15439395, 26929220,
46537966, 32774989, 2367911, 7429892, 88097807, 274153775, 41147062,
127973392, 10350, 12127832, 18666844, 279787626, 14511951, 116036694,
225784456, 270449528, 63796828, 22784458, 364917755, 15693493,
```

```
17346342, 28933226, 7872262, 14180231, 100581193, 88323281,
131029775, 25952851, 14692604, 21681252, 309366491, 25376226,
22705686, 188108667, 18715024, 3829151, 41172837, 92319025, 24081780,
15102849, 83943787, 15680204, 7846, 184097849, 7377812, 204317520,
24641194, 14691709, 112939321, 7081402, 153226312, 14269220,
94414805, 108811740, 1065921, 221829166, 116952434, 15222083,
6608332, 17759701, 16098603, 7860742, 20471349, 6519522, 19040580,
14536491, 197504076, 12611642, 36198161, 3840, 22841103, 440963134,
20273398, 17092592, 13348, 17093617, 22679419, 208132323, 18776017,
15846407, 18581803, 5442012, 813286, 3359851, 59804598]
```

3. Finding the 100 accounts that follow the most of the accounts found in 2).

```
followed_accs = session.execute(f'SELECT follower_id FROM twitter.follow
_list = list()
for row in followed_accs:
    _list.append(row[0])

followed_top_100 = Counter(_list).most_common(100)
print("100 accounts that follow the most of the accounts found in 2)",fc
```

```
100 accounts that follow the most of the accounts found in 2)
[(3359851, 47), (7860742, 47), (15913, 45), (7861312, 45), (16098603,
43), (18776017, 41), (10350, 39), (48485771, 39), (22679419, 38),
(3443591, 37), (18581803, 37), (18927441, 37), (26281970, 37),
(7872262, 36), (24742040, 36), (5442012, 36), (10671602, 35),
(11928542, 35), (14922225, 35), (16453996, 35), (21681252, 35),
(14589257, 34), (15853668, 34), (9451052, 34), (65913144, 34),
(14269220, 34), (15234657, 34), (17092592, 34), (59804598, 34),
(87764480, 34), (40981798, 34), (16475194, 34), (16464746, 33),
(93905958, 33), (20152005, 33), (42361118, 33), (36629388, 33),
(18666844, 32), (19413393, 32), (12127832, 32), (14180231, 32),
(14983833, 32), (25026165, 32), (18742444, 32), (29758446, 31),
(20880546, 31), (116952434, 31), (30207757, 31), (21195122, 31),
(43003845, 31), (22784458, 30), (19040580, 30), (14691709, 30),
(24004172, 30), (1065921, 30), (26280712, 30), (43170475, 30),
(22462180, 30), (7846, 29), (1183041, 29), (3040621, 29), (7377812,
29), (34428380, 29), (20273398, 28), (14230524, 28), (15222083, 28),
(24641194, 28), (83943787, 28), (36198161, 28), (14536491, 28),
(813286, 27), (12611642, 27), (29514951, 27), (31353077, 27),
(18996905, 27), (14471778, 27), (15838599, 27), (16112634, 27),
(17526132, 27), (17759701, 27), (31331740, 27), (25952851, 27),
(84043660, 27), (9431932, 27), (20397258, 27), (127973392, 27),
(16674726, 27), (116036694, 27), (972651, 26), (13687132, 26),
(4068821, 26), (7702232, 26), (6080022, 26), (17224642, 26),
(20935355, 26), (24081780, 26), (26033920, 26), (28933226, 26),
(43933017, 26), (101633415, 26)]
```

4. Listing the information for the personal home page of any account (best try with the accounts found in 2); the start page should contain the following (implement as separate queries):

- the number of followers && the number of followed accounts

```
SELECT follower_len, follows_len FROM twitter.user_stats WHERE user_id = 233248636; //cheack user_id can be changed
```

Rows

follower_len	follows_len
60	20

Total: 1 results

- either the 25 newest or the 25 most popular posts of the followed accounts (via DB query)

25 newest

```
CREATE MATERIALIZED VIEW twitter.start_view_new AS
SELECT
user_id_x, follower_id, number_of_likes, number_of_shares, date_time
FROM twitter.user
WHERE user_id_x IS NOT NULL AND follower_id IS NOT NULL AND
date_time IS NOT NULL AND id IS NOT NULL
PRIMARY KEY ((user_id_x), date_time, follower_id, id);

SELECT * FROM twitter.start_view_new WHERE user_id_x =
172883064 ORDER BY date_time DESC LIMIT 25;
```

Rows

user_id_x	date_time	follower_id	id	author	content	name	number_of_likes	number_of_sh
172883064	2017-01-24 02:30:00 +0000	299293996	8.2372e+17	YouTube	You must be at least 1 ½ inches tall to ride.mrnLego Night Ride: https://t.co/kgurndAo1D https://t.co/5xDNB04aeR	null	719	107
172883064	2017-01-24 01:30:00 +0000	299293996	8.23704e+17	YouTube	New Year's resolutions got us like ... 🚩💎🚩 https://t.co/xGIRLZje2t https://t.co/s8n4WeOtw3	null	646	79
172883064	2017-01-24 00:30:00 +0000	299293996	8.23689e+17	YouTube	Feeling so energized after the #Sundance panel for #ThisIsEverythingDoc today. Emotional, thoughtful, empowering. 🙌 https://t.co/QB7ablts4y	null	522	57
172883064	2017-01-23 23:30:00 +0000	299293996	8.23674e+17	YouTube	Forget Throwback Thursday. This Monday, we're going prehistoric. https://t.co/CzokOmzsXU https://t.co/eTPMzxWmrp	null	542	49

25 most popular

```

CREATE MATERIALIZED VIEW twitter.start_view_like AS
SELECT
user_id_x,follower_id,number_of_likes,number_of_shares,date_time,aut
FROM twitter.user
WHERE user_id_x IS NOT NULL AND follower_id IS NOT NULL AND
number_of_likes IS NOT NULL AND id IS NOT NULL
PRIMARY KEY ((user_id_x),number_of_likes,follower_id,id);

SELECT * FROM twitter.start_view_like WHERE user_id_x =
172883064 ORDER BY number_of_likes DESC LIMIT 25;

```

Rows								
user_id_x	number_of_likes	follower_id	id	author	content	date_time	nan	
172883064	371255	233248636	7.54895e+17	taylorswift13	That moment when Kanye West secretly records your phone call, then Kim posts it on the Internet. https://t.co/4GJqdykQu	2016-07-18 04:24:00 +0000		null
172883064	288738	125120339	6.91627e+17	rihanna	listening to ANTI 📍 https://t.co/AgqlhU0GEz	2016-01-25 14:22:00 +0000		null
172883064	196155	233248636	4.27648e+17	taylorswift13	And you know... We're on each other's team. #LORDE #CLEANINGUP #GRAMMYS http://t.co/jRwHkPYfPE	2014-01-27 03:45:00 +0000		null
172883064	187367	233248636	7.00891e+17	taylorswift13	I'm very happy to say the next single from 1989 will be 'New Romantics'.	2016-02-20 03:52:00 +0000		null
172883064	183459	233248636	6.17398e+17	taylorswift13	When @edsheeran shows up for the 4th of July in a red coat because he just can't let it go. http://t.co/ulTbJhUt4	2015-07-04 18:21:00 +0000		null
172883064	182169	233248636	5.66746e+17	taylorswift13	Sending my love to all the lonely Starbucks lovers out there this Valentine's Day.....even though that is not the correct lyric.	2015-02-14 23:50:00 +0000		null
172883064	181316	233248636	6.80099e+17	taylorswift13	You used to call me on my elf phone. https://t.co/nMgUI0lgp6	2015-12-24 18:54:00 +0000		null

5. Caching of the posts for the home page (cf. 4) requires a so-called fan-out in the cache of each follower when writing a new post

```

CREATE MATERIALIZED VIEW twitter.start_view_user1 AS
SELECT
follower_id,number_of_likes,date_time,author,name,content,id FROM
twitter.user
WHERE user_id_x IS NOT NULL AND user_id_x=172883064 AND
follower_id IS NOT NULL AND number_of_likes IS NOT NULL AND id IS
NOT NULL
PRIMARY KEY ((user_id_x),number_of_likes,follower_id,id);
// # order by need partion key in WHERE
SELECT * from twitter.start_view_taylor WHERE
user_id_x=172883064 ORDER BY number_of_likes DESC LIMIT 25;
// INSERT new tweet
INSERT INTO twitter.user
(user_id_x,follower_id,name,author ,content ,country ,date_time
,id ,language ,latitude ,longitude ,number_of_likes
,number_of_shares ,user_id_y)
VALUES
(172883064, 233248636, 'NoName','taylorswift12', 'Hallo there
BDEA','DE', dateof(now()), 'NoID', 'de', 48, 48, 10000000, 0, 0);

INSERT INTO twitter.tweets(author, content, country, date_time,
id, language, latitude, longitude, number_of_likes,
number_of_shares ,author_id)
VALUES
('taylorswift12', 'Hallo there BDEA','DE', dateof(now()),
'NoID', 'de', 48, 48, 10000000, 0, 0 '233248636');

```

Rows								
user_id_x	number_of_likes	follower_id	id	author	content	date_time	nan	
172883064	371255	233248636	7.54895e+17	taylorswift13	That moment when Kanye West secretly records your phone call, then Kim posts it on the Internet. https://t.co/4GJqdykQu	2016-07-18 04:24:00 +0000	null	
172883064	288738	125120339	6.91627e+17	rihanna	listening to ANTI 📍 https://t.co/AgqlhU0GEz	2016-01-25 14:22:00 +0000	null	
172883064	196155	233248636	4.27648e+17	taylorswift13	And you know... We're on each other's team. #LORDE #CLEANINGUP #GRAMMYS http://t.co/jRwHkPYfPE	2014-01-27 03:45:00 +0000	null	
172883064	187367	233248636	7.00891e+17	taylorswift13	I'm very happy to say the next single from 1989 will be 'New Romantics'.	2016-02-20 03:52:00 +0000	null	
172883064	183459	233248636	6.17398e+17	taylorswift13	When @edsheeran shows up for the 4th of July in a red coat because he just can't let it go. http://t.co/ulTbJhUt4	2015-07-04 18:21:00 +0000	null	
172883064	182169	233248636	5.66746e+17	taylorswift13	Sending my love to all the lonely Starbucks lovers out there this Valentine's Day.....even though that is not the correct lyric.	2015-02-14 23:50:00 +0000	null	
172883064	181316	233248636	6.80099e+17	taylorswift13	You used to call me on my elf phone. https://t.co/nMgUI0lgp6	2015-12-24 18:54:00 +0000	null	

6. List of the 25 most popular posts that contain a given word (if possible also with AND linking several words)

```
// To order by the 25 most number_of_like is at our knowledge
not possible with the current dataschema
CREATE CUSTOM INDEX search_in ON twitter.tweets (content) USING
'org.apache.cassandra.index.sasi.SASIIndex'
WITH OPTIONS = { 'mode': 'CONTAINS', 'analyzer_class':
'org.apache.cassandra.index.sasi.analyzer.NonTokenizingAnalyzer',
'case_sensitive': 'false' };
SELECT * from twitter.tweets WHERE content LIKE '%world%' limit
25;
```

Rows

author_id	date_time	id	author	content	country
6502892	2015-06-08 14:57:00 +0000	6.07924e+17	KimKardashian	Happy Birthday 2my best friend in the entire world! U R the most amazing husband &dad! U inspire me every single day http://t.co/XbsHW1RpFv	null
6502892	2015-06-23 06:41:00 +0000	6.13235e+17	KimKardashian	So happy I ran back inside to grab these! Most comfy shoes in the world! Yeezy Boosts 350 http://t.co/4JXWgm00Ud	null
6502892	2015-06-27 14:57:00 +0000	6.1481e+17	KimKardashian	Happy Birthday 2 the best sister in the world! I love u so much! Words can't even describe how beautiful your soul is http://t.co/cr6XYfMDBI	null
6502892	2015-07-21 23:49:00 +0000	6.23641e+17	KimKardashian	Wait wait I'm in Paris it's the middle of the night & I'm posting my Vogue Spain pics	null