



Home



DataStorage



Caching



User Interface

Tweet

▶ Home



Twitter Search Application



TEAM 2

16:954:694:01 DBMS

Cohort Presentation Done 

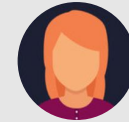
Team Members



Sai Adarsh
Kasula
sk2837



Krit Gupta
ksg124



Himani Hooda
hh660



Vanshika
Gurbani
vg460



Objective

- Dataset Format : JSON
- Total Documents: ~112,000
- Time Frame: April 2020 (Covid-19 period)
- Unique Users: ~90,000



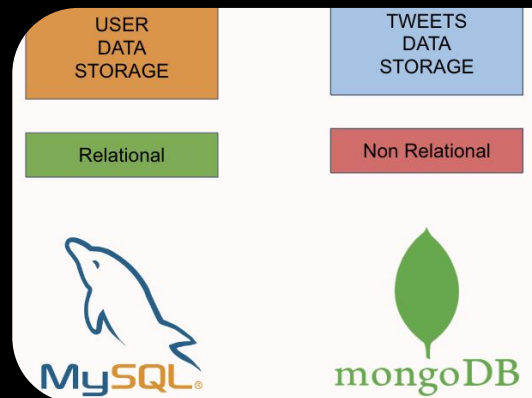
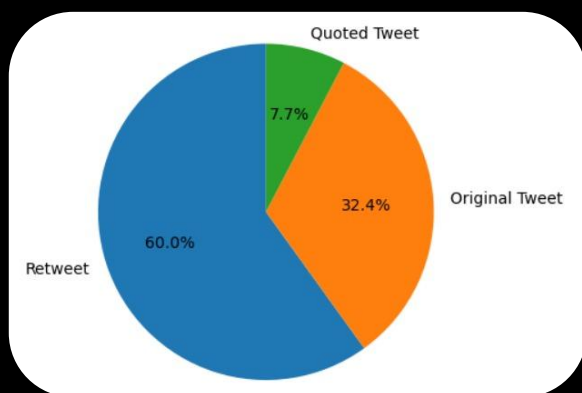
Tweet



DATA SUMMARY

...

Data Structure: Organized into separate datastores for clarity — one for user profiles and another for tweet contents. Tweets Distribution: Shown in the pie chart.



About:

Twitter, a leading social media platform globally, has millions of active users tweeting daily. The platform's search feature utilizes its extensive database, offering users a straightforward and effective way to swiftly locate pertinent information on any subject.

Tweet



ARCHITECTURE







Search Twitter



USER DATA STORAGE

Key libraries such as "mysql.connector" and "json" were installed. After establishing a connection to the MySQL server and setting up a cursor, a users table was created with fields like ID, name, location, and verification status. User data extraction involved two main functions: loadData() and UserInsert(). loadData() processes JSON line by line, identifying tweets and extracting user details. UserInsert() checks if the user exists in the database; if not, the user's details are inserted. After processing, unique users were added with attributes such as ID and name. For optimized search, indices were created on the ID and name columns, using a Binary Tree type for both.

Indexes in Table

Visible	Key	Type
<input checked="" type="checkbox"/>	 PRIMARY	BTREE
<input checked="" type="checkbox"/>	 idx_users_name	BTREE

Column	Type	Default Value	Nullable
id	varchar(255)		NO
name	varchar(255)		YES
screen_name	varchar(255)		YES
verified	tinyint(1)		YES
followers_count	bigint		YES
friends_count	bigint		YES
created_at	datetime		YES
location	varchar(255)		YES
tweets_count	bigint		YES
Description	text		YES





TWEET DATA STORAGE

Using Python and MongoDB client, along with libraries like pymongo, the database connection was established.. Functions such as insert_tweet() for adding tweets, find_tweet() to prevent duplicates, retweet_update() for updating retweet counts, and process_tweets() for reading and processing tweets were implemented.

The data was loaded in such a way that every retweet has a field to capture the source tweet thereby creating a link for any drill-through features of the search application

Indexes were also created on the "User_Id" and "Text" fields to enhance the search functionality, essential for efficient query performance in applications that combine results from different databases.

Tweets_collection

Storage size: 22.94 MB

Documents: 112 K

Avg. document size: 383.00 B

Indexes: 2

Total index size: 26.68 MB

`_id: ObjectId('662914d94d73db0d9c499964')`
`created_at: "2020-04-25 07:30:12"`
`Tweet_Id: "1253949413191344128"`
`Text: "India's war with Corona is ongoing.`
`Play your part and make sure no o..."`
`Hashtag: Array (empty)`
`User_Id: "207809313"`
`User_Name: "BJP"`
`Retweet_Count: 59`
`Likes_Count: 1870`

Tweets_collection

Documents 112.0K Aggregations Schema Indexes 2 Validation

Create Index Refresh

Name and Definition	Type	Size	Usage
<div><div>▼ _id_</div><div>_id ↑</div></div>	REGULAR ⓘ	2.6 MB	99 (since Wed Apr 24 2024)
<div><div>▼ Text_text</div><div>_fts (text) _ftsx ↑</div></div>	TEXT ⓘ	24.1 MB	16 (since Wed Apr 24 2024)



CACHING

Purpose: Efficiently store frequently accessed data to reduce database load and improve response times.

Speeding it up:

A Python dictionary-based cache with eviction strategy and periodic disk checkpoints is used.

Key Features:

- The size of the cache is around 10 GB.
- Eviction Strategy: Uses "least accessed" strategy to remove items when cache is full.
- Checkpoint Interval: The cache class includes a checkpoint interval of 5 minutes, which allows the cache to regularly save its contents to disk.



Tweet



Search Twitter



USER INTERFACE

TWITTER SEARCH APPLICATION : TEAM 2

Explore Twitter Data:

Trending Users

Trending Tweets

Trending Hashtags

Enter username, hashtag or part of a string:

Submit

Landing Page



Tweet



Search Twitter



USER INTERFACE

Trending Users

Name	Handle	Verified	Followers	Following	Location	Tweets	Bio
Barack Obama	BarackObama	Yes	116518121	607194	Washington, DC	0	Dad, husband, President, citizen.
Donald J. Trump	realDonaldTrump	Yes	78467254	46	Washington, DC	0	45th President of the United States of Americaus
CNN Breaking News	cnnbrk	Yes	57529057	120	Everywhere	0	Breaking news from CNN Digital. Now 56M strong. Check @cnn for all things CNN, breaking and more. Download the app for custom alerts: http://cnn.com/apps
Narendra Modi	narendramodi	Yes	55781248	2364	India	4	Prime Minister of India
Shakira	shakira	Yes	52250613	212	Barranquilla	0	👉 ME GUSTA Shakira & Anuel AA Nuevo Sencillo / New Single
CNN	CNN	Yes	47565193	1106	None	0	It's our job to #GoThere & tell the most difficult stories. Join us! For more breaking news updates follow @CNNBRK & Download our app http://cnn.com/apps
The New York Times	nytimes	Yes	46361159	904	New York City	1	News tips? Share them here: http://nyti.ms/2FVHq9v
BBC Breaking News	BBCBreaking	Yes	43014510	3	London, UK	0	Breaking news alerts and updates from the BBC. For news, features, analysis follow @BBCWorld (international) or @BBCNews (UK). Latest sport news @BBCSport.
Amitabh Bachchan	SrBachchan	Yes	41596464	1833	Mumbai, India	1	"तुमने हमें पूज पूज कर पत्थर कर डाला ; वे जो हमपर जुमले कसते हैं हमें ज़िंदा तो समझते हैं" ~ हरिवंश राय बच्चन
Salman Khan	BeingSalmanKhan	Yes	40094611	26	MUMBAI	0	Film actor, artist, painter, humanitarian

Trending Users



Tweet



Search Twitter



USER INTERFACE

Trending Tweets

User	Retweets	Likes	Created At	Text	Hashtags
Quirinale	798	9524	2020-04-25 09:23:42	#25Aprile, il Presidente #Mattarella si è voluto recare all'#AltaredellaPatria dove ha deposto una corona d'alloro... https://t.co/ev4b4DIgDf	['25Aprile', 'Mattarella', 'AltaredellaPatria']
dr. Shela Putri Sundawa	784	5498	2020-04-25 05:24:38	Ngomongin teori konspirasi corona sama orang yg percaya kalo bumi itu datar, I'm not a smart people, but please dud... https://t.co/aCIWBZg41m	[]
Quirinale	654	4062	2020-04-25 09:56:54	#25Aprile, nel 75° anniversario della #Liberazione il Presidente #Mattarella ha deposto una corona all'Altare della... https://t.co/vDcVcwEMQy	['25Aprile', 'Liberazione', 'Mattarella']
Brit Hume	1474	1831	2020-04-25 11:27:12	And where is the evidence that Covid 19 is easily spread outdoors? https://t.co/YPcJXU1uqw	[]
RU Only In Russia RU	586	1352	2020-04-25 11:26:44	Corona disinfecting in Russia. https://t.co/AEF5ccDmM3	[]
Şişli	445	1532	2020-04-23 19:58:13	Quando eu ligo a televisão e fica falando só de corona virus	[]
Ben Wikler	986	0	2020-04-25 12:51:03	Milwaukee's health commissioner has now tied 40 coronavirus infections to the April 7 election. https://t.co/fGlsLKzTkm	[]
Funda	732	0	2020-04-25 12:43:26	Gözün çıksın corona 🤔 Ülkece asabii olduk 🤔 muhtemel psikoloji ektedir 🤔🤔🤔🤔 #PideAlmayaDiyeÇıkıp https://t.co/KoGSbVAMxZ	['PideAlmayaDiyeÇıkıp']
Laura Ingraham	500	0	2020-04-25 13:53:37	A MUST READ...Coronavirus Restrictions: Government Bears the Burden of Proof Before Denying Freedoms National Rev... https://t.co/RcaAK9nwDs	[]
Gu Ru Thalaiva	422	0	2020-04-25 13:01:20	Thalapathy fans from Sivakasi Helped the Poor Family who are affected by this corona Crisis ! They have supplied th... https://t.co/ozoG6d9Ax8	[]

Trending Tweets



Tweet



Search Twitter



USER INTERFACE

Trending Hashtags

Hashtag	Tweets Count
Corona	5133
corona	1636
Mattarella	1516
25Aprile	1476
Covid_19	1049
COVID19	877
AltaredellaPatria	806
PideAlmayaDiyeÇikip	777
coronavirus	707
Liberazione	700

Trending Hashtags



Tweet



Search Twitter



USER INTERFACE

User Search Results

#	Name	Handle	Verified	Followers	Tweets	Description	Location	Creation Date
1	Donald J. Trump	realDonaldTrump	✓	76792576	0	45th President of the United States of Americaus	Washington, DC	2009-03-18 13:46:38
2	Karli Q ★★★★★ - Text TRUMP to 88022	KarluskaP	✗	55107	0	That Cackling Conservative 🦃 sings too! #Trump2020-Music- https://store.cdbaby.com/artist/karlibonne https://music.apple.com/in/artist/karli-bonne/29014046	None	2019-11-27 02:57:42
3	MAGA-Joanne us Text TRUMP to 88022	JoanneTarpon07	✗	49662	1	http://paypal.me/codeofvets #MAGA, TrumpTrain, #2A, #KAG Christian Conservation, Florida Patriot, God+Country+Family, Nationalist #NRA #seanhannity; #loudobbs	United States	2016-01-27 18:30:46
4	John Trumpfan	JohnTrumpFanKJV	✗	43817	3	Repentance toward God and Faith toward our Lord Jesus Christ. King James Bible Believer. Ye must be born again. Pro Life Pro Israel. God bless Pres Trump. #MAGA	Florida	2016-03-20 18:56:58
5	TWITMO INMATE ★★★★★ Text TRUMP to 88022	TWITMO_INMATE	✗	33385	1	THE MORE YOU RESEARCH THE CRAZIER YOU SOUND TO IGNORANT PEOPLE... WWG1WGA FB @GenFlynn and Epstein didn't kill himself	United States	2018-05-29 20:25:37

Enter the user number whose tweets you want to be displayed:

Submit

Search by
@username

- Search Results
were limited to 5.



Tweet



Search Twitter



USER INTERFACE

Tweets relevant to "Trump"

User	Retweets	Likes	Created At	Text	Hashtags
Fitri Iskandar	0	0	2020-04-25 14:48:35	RT @Dandhy_Laksono: Jokowi minta ventilator dari Trump. Trump ngetwit siap bantu. Dosen ITB:	[]
snoesjoe	0	0	2020-04-25 14:48:33	MIRACLES. MALARIA DRUGS. MOTHER NATURE. TRUMP'S DANGEROUS CORONA 'CURES' https://t.co/9tRAy1epYa	[]
mary c	0	0	2020-04-25 14:48:31	RT @JoyceWhiteVance: What a coincidence that this happened AND that Trump blocked a highly regarded former DOJ IG as corona-IG. https://t.co/9tRAy1epYa	[]
Nikhil Kumar	0	0	2020-04-25 14:48:22	#IndiaWithPMModi God save India from corona and incapable leaders of India. Trump-Modi ki masti Gujratiyon per padi... https://t.co/8pT4yvWpHF	['IndiaWithPMModi']
OurCountryNeedsEU	0	0	2020-04-25 14:48:22	RT @dennisdiclaudio: The elderly are being very unfair to Donald Trump.	[]

Search by String



Tweet



Search Twitter



USER INTERFACE

Top 5 Hashtags Matching Your Search for "#trump"

Hashtag	Tweets Count
Trump	139
Corona	73
corona	43
NamasteTrump	40
HeilenwieTrump	22

Enter the hashtag related to which the relevant tweets you want to be displayed:

Submit

Tweets of #Trump

Username	Retweets	Likes	Created At	Text	Hashtags
Hatice Bingöl	0	0	2020-04-25 14:48:10	RT @hakana: ABD Başkanı #Trump 25 Şubatta sadece 15 #corona vakası olduğunu ve hepsinin iyileşeceğini ilan etmişti. Bugün ülkede nerdeyse 1...	['Trump', 'corona']
Zekiye Özyürek ⚡	0	0	2020-04-25 14:46:10	RT @hakana: ABD Başkanı #Trump 25 Şubatta sadece 15 #corona vakası olduğunu ve hepsinin iyileşeceğini ilan etmişti. Bugün ülkede nerdeyse 1...	['Trump', 'corona']
Ayşe	0	0	2020-04-25 14:45:56	RT @hakana: ABD Başkanı #Trump 25 Şubatta sadece 15 #corona vakası olduğunu ve hepsinin iyileşeceğini ilan etmişti. Bugün ülkede nerdeyse 1...	['Trump', 'corona']

Search by
Hashtag #

Tweet



Search Twitter



THANK YOU



Tweet