

The TopN extension: Maintaining top 10 lists at scale

- Furkan Şahin
- Software Engineer
- Citus Data @Microsoft
- Author of [postgresql-topn](#) extension



Presentation Flow

- TopN lists and their usages
- How to calculate TopN lists in PostgreSQL?
- Why do we need an extension?
- How does postgresql-topn work?
- How do we use postgresql-topn?
- Demo
- Summary



@sahinffurkan



@furkansahin

TopN Lists and Their Usages

How to calculate TopN lists in PostgreSQL?

Why do we need an extension?

How does postgresql-topn work?

How do we use postgresql-topn?

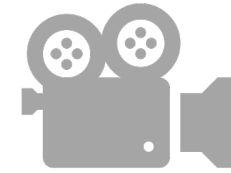
Demo

Summary

TopN Lists and Their Usages



Spotify: [Your Top Songs 2018](#)



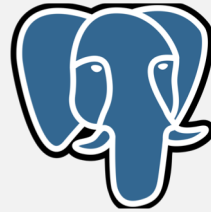
IMDB: Top Rated Movies



E-commerce: Most selling items



How to calculate TopN lists in PostgreSQL



```
SELECT item, count(*)  
FROM table  
GROUP BY 1  
ORDER BY 2 DESC  
LIMIT 10;
```



```
SELECT  
    topn(topn_add_agg(item),10)  
FROM  
    table;
```



TopN Lists and Their Usages

How to calculate TopN lists in PostgreSQL

Why do we need an extension?

How does postgresql-topn work?

How do we use postgresql-topn?

Demo

Summary

Why do we need an extension?

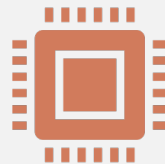


Continuous calculations



Memory usage

Scanning



CPU usage

Sorting



@sahinffurkan



@furkansahin

Why needed a TopN extension?



High load times



Solution?



An extension that uses
count-min sketch

- Low memory footprint
- Low CPU usage
- Facilitate pre-processing



@sahinffurkan



@furkansahin

TopN Lists and Their Usages

How to calculate TopN lists in PostgreSQL

Why do we need an extension?

How does postgresql-topn work?

How do we use postgresql-topn?

Demo

Summary

How does postgresql-topn work?



Count-min sketch




Limited number of counters



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Heavy fuel	1

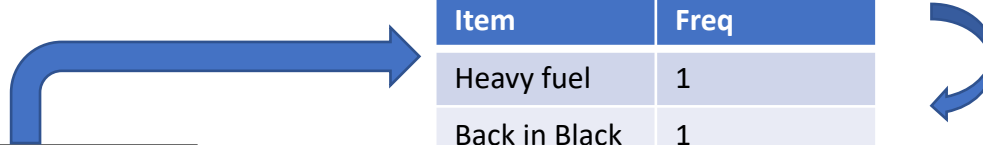
```
SELECT
    topn(topn_add_agg(Music),1)
FROM
    my_music_history;
```



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Heavy fuel	1
Back in Black	1

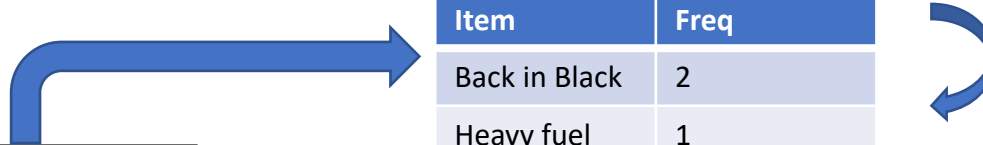
```
SELECT
    topn(topn_add_agg(Music),1)
FROM
    my_music_history;
```



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Back in Black	2
Heavy fuel	1

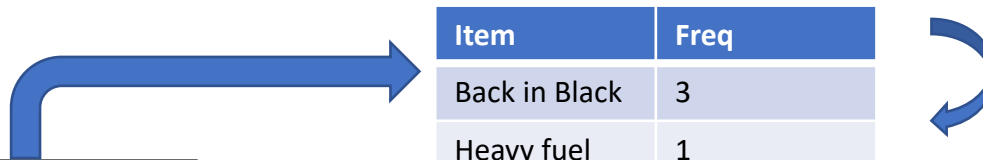
```
SELECT
    topn(topn_add_agg(Music),1)
FROM
    my_music_history;
```



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Back in Black	3
Heavy fuel	1

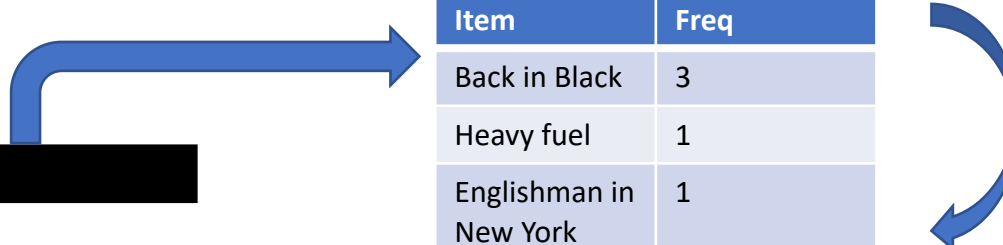
```
SELECT
    topn(topn_add_agg(Music),1)
FROM
    my_music_history;
```



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Back in Black	3
Heavy fuel	1
Englishman in New York	1

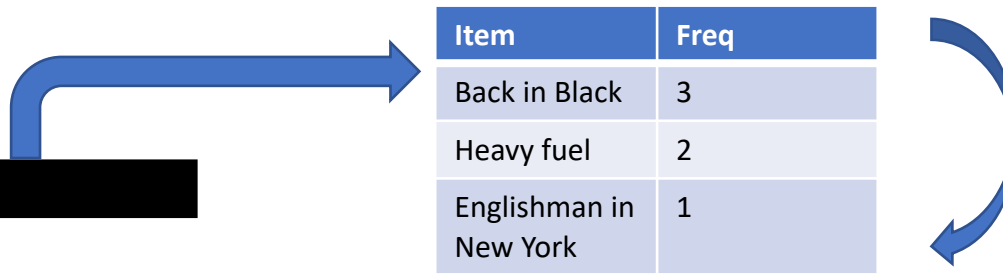
```
SELECT
    topn(topn_add_agg(Music),1)
FROM
    my_music_history;
```



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Back in Black	3
Heavy fuel	2
Englishman in New York	1

```
SELECT
    topn(topn_add_agg(Music),1)
FROM
    my_music_history;
```



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Back in Black	3
Heavy fuel	2
Englishman in New York	1
Breakfast in America	1

Wipe the bottom half

```
SELECT
    topn(topn_add_agg(Music), 1)
FROM
    my_music_history;
```



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Back in Black	3
Heavy fuel	2

No more data


```
SELECT
    topn(topn_add_agg(Music),1)
FROM
    my_music_history;
```



Example: Calculating the Top 1 song

topn.number_of_counters = 1

Date	Music
10-10-2019	Heavy Fuel
11-10-2019	Back in Black
12-10-2019	Back in Black
13-10-2019	Back in Black
13-10-2019	Englishman in New York
14-10-2019	Heavy Fuel
14-10-2019	Breakfast in America



Item	Freq
Back in Black	3

```
SELECT
    topn(topn_add_agg(Music),1)
FROM
    my_music_history;
```



How does postgresql-topn work?

`topn(column::JSONB, n::int) => returns
Tuple Set(item, frequency)`

`topn_add_agg(column::text) => returns
JSONB`

`topn_union_agg(column::JSONB) => returns
JSONB`

`topn_add(column::JSONB, data::text) =>
returns JSONB`

`topn_union(column_1::JSONB,
column_2::JSONB) => returns JSONB`



Why JSONB?

Human readable

Portable

Easy to process



@sahinffurkan



@furkansahin

TopN Lists and Their Usages

How to calculate TopN lists in PostgreSQL

Why do we need an extension?

How does postgresql-topn work?

How do we use postgresql-topn?

Demo

Summary

How do we use
postgresql-topn?

Pre-aggregation

Materialization of TopNs

TopN Lists and Their Usages

How to calculate TopN lists in PostgreSQL

Why do we need an extension?

How does postgresql-topn work?

How do we use postgresql-topn?

Demo

Summary

Summary



Requires pre-processing for performance gain



Low memory footprint



Approximates the results



Open source



Production ready

Resources

- <https://github.com/citusdata/postgresql-topn>
- TopN for your Postgres database
<https://www.citusdata.com/blog/2018/03/27/topn-for-your-postgres-database/>



@sahinffurkan



@furkansahin

Thank you

 @sahinffurkan  @furkansahin

