

Day41 - March 17th 2024

1. Solved one leetcode medium problem
2. Started learning spark from youtube

The image shows a dual-screen setup. The left screen displays a YouTube video titled "spark fundamental (Theory)" by MANISH KUMAR. The video content includes a diagram of a driver with 1GB memory and 750MB of data, broadcasting 50MB of data to three executors (df2, df3, and another). The text "5 df broadcast" and "50 x 5 = 250 MB" is visible. The right screen shows a Google Docs document titled "Spark Theory Day5". It contains a list of notes and a diagram similar to the one in the video. The notes include:

- 23. If our broadcast join size is greater than driver memory size..then we may get OOM
- 24. (Diagram showing driver 1GB, 750MB data, and 50MB broadcast to 5 df)
- 25. Here if we want to join two dataframe and broadcast it to executors...then if these two df size is less than driver memory..then we don't get any issue
- 26. But if the df's size exceeds the driver memory..then we get OOM
- 27. And also if by mistake we configured overhead memory less than 10%..then we may get OOM error
- 28. Common reasons explained : <https://q.co/gemini/share/94b1d7105aa8>
- 29.

The diagram on the right screen is identical to the one in the video, showing a driver with 1GB memory, 750MB of data, and 50MB of broadcast data being sent to 5 dataframes (df2, df3, and three others). The text "5 df broadcast" and "50 x 5 = 250 MB" is also present.



4. Ended my day by solving SQL problems

SQLQuery1.sql - KAUSHI\SQLEXPRESS.master (KAUSHI\jamka (53)) - Microsoft SQL Server Management Studio

Object Explorer

- KAUSHI\SQLEXPRESS (SQL Server 16.0.1000)
  - Databases
  - Security
  - Server Objects
  - Replication
  - Management
  - XEvent Profiler

SQLQuery1.sql - K...KAUSHI\jamka (53) \*

```
create table SALE (merchant varchar(100), amount int, pay_mode varchar(100));

insert into SALE values('merchant_1',150,'cash');
insert into SALE values('merchant_1',500,'online');
insert into SALE values('merchant_2',450,'online');
insert into SALE values('merchant_1',100,'cash');
insert into SALE values('merchant_3',600,'cash');
insert into SALE values('merchant_5',200,'online');
insert into SALE values('merchant_2',100,'online');

--Q: Write a SQL query to find total amount received by each merchant
-- via CASH & Online mode
select * from SALE
```

Results

	merchant	amount	pay_mode
1	merchant_1	150	cash
2	merchant_1	500	online
3	merchant_2	450	online
4	merchant_1	100	cash
5	merchant_3	600	cash
6	merchant_5	200	online
7	merchant_2	100	online

Query executed successfully.

KAUSHI\SQLEXPRESS (16.0 RTM) KAUSHI\jamka (53) master 00:00:00 7 rows

SQLQuery1.sql - KAUSHI\SQLEXPRESS.master (KAUSHI\jamka (53)) - Microsoft SQL Server Management Studio

Object Explorer

- KAUSHI\SQLEXPRESS (SQL Server 16.0.1000)
  - Databases
  - Security
  - Server Objects
  - Replication
  - Management
  - XEvent Profiler

SQLQuery1.sql - K...KAUSHI\jamka (53) \*

```
select * from SALE

select merchant,coalesce(sum(case when pay_mode = 'cash' then amount end),0) as cash_amount,
coalesce(sum(case when pay_mode = 'online' then amount end),0) as online_amount
from sale
group by merchant

/* Explanation:
Step1 : There's no big logic here..see the code and understand
```

Results

	merchant	cash_amount	online_amount
1	merchant_1	250	500
2	merchant_2	0	550
3	merchant_3	600	0
4	merchant_5	0	200

Query executed successfully.

KAUSHI\SQLEXPRESS (16.0 RTM) KAUSHI\jamka (53) master 00:00:00 4 rows

Spark TlOnlyGrrMinimuNetfli(6) execdf.colleGeminiMembe(6) ShiriAmazon(6) IReplaceStra xYouTub+

platform.stratascratch.com/coding/10046-top-5-states-with-5-star-businesses?code\_type=3

stratascratchCoding QuestionsNon-coding QuestionsData ProjectsGuidesBlogPricingProfile

Back to QuestionsMySQLTop 5 States With 5 Star Businesses

QuestionYour SubmissionsSolutionSolutions from UsersDiscussionResources00:00

### Top 5 States With 5 Star Businesses

Interview Question Date: March 2020

YelpHardID 1004691

Data EngineerData ScientistBI AnalystData AnalystML Engineer

Find the top 5 states with the most 5 star businesses. Output the state name along with the number of 5-star businesses and order records by the number of 5-star businesses in descending order. In case there are ties in the number of businesses, return all the unique states. If two states have the same result, sort them in alphabetical order.

Table: yelp\_business

HintsExpected Output

All required columns and the first 5 rows of the solution are shown

state	five_star_counts
AZ	10
ON	5
NV	4
IL	3
OH	3
WI	3

```
1 select state,sum(case when stars = 5 then 1 end) as five_star_counts
2 from yelp_business
3 group by state
4 order by five_star_counts desc,state
5 limit 6
```

#### Your Solution:

state	five_star_counts
AZ	10
ON	5
NV	4
IL	3
OH	3
WI	3

2°C ClearSearchENG US10:59 PM 3/17/2024