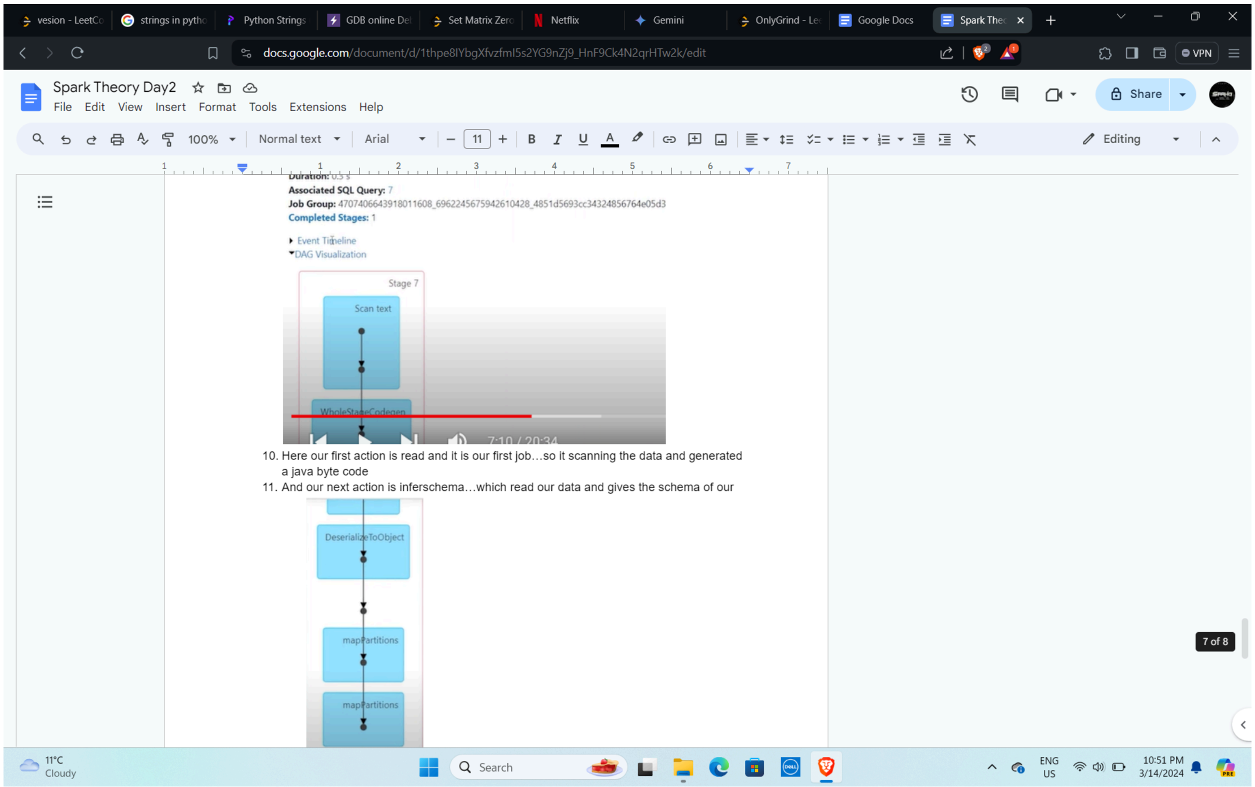
Day38 - March 14th 2024

1. Started my day at 6 am
2. Cooked food for lunch and packed it in box
3. Headed to library at 8:20am
4. Started solving leetcode questions to improve my logical building skills
5. Learning spark theory from manish kumar



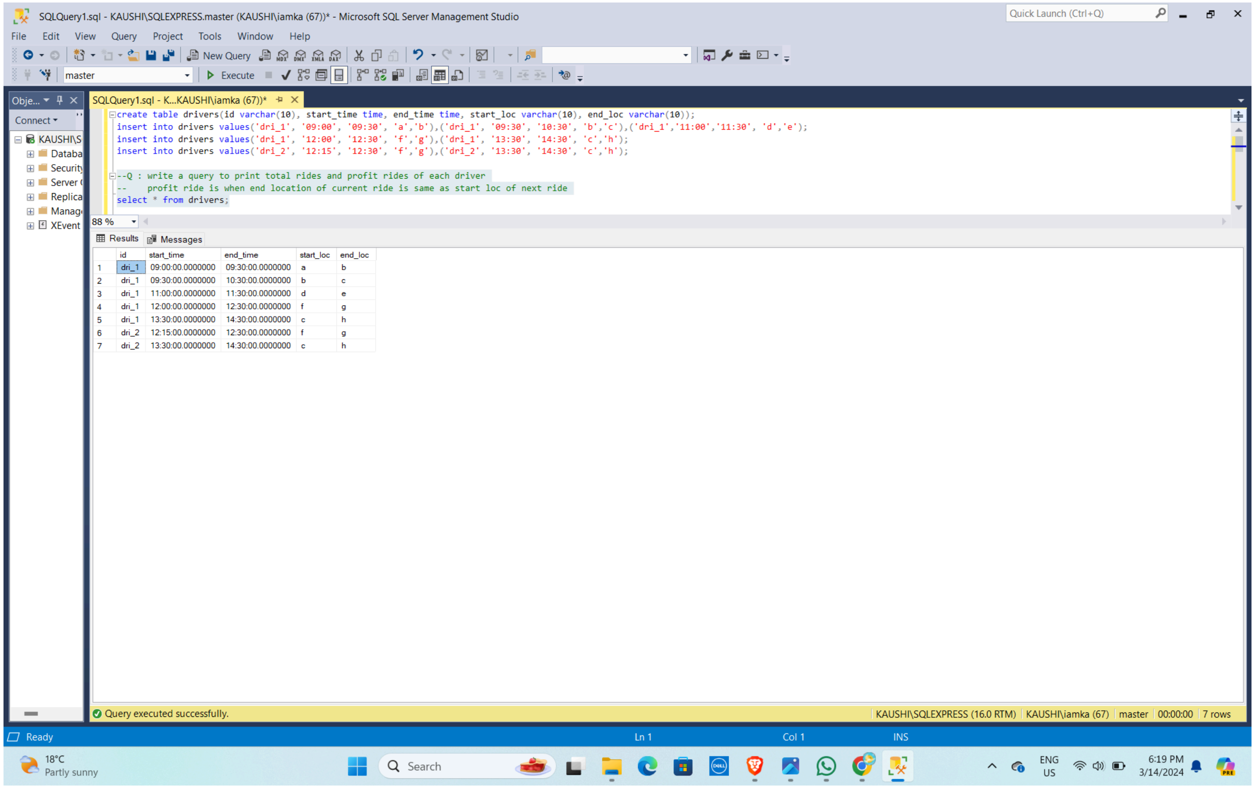
Spark Theory Day2

Duration: 0.5 s
Associated SQL Query: 7
Job Group: 4707406643918011608_6962245675942610428_4851d5693cc34324856764e05d3
Completed Stages: 1

10. Here our first action is read and it is our first job…so it scanning the data and generated a java byte code
11. And our next action is inferschema…which read our data and gives the schema of our

6. Spark theory day 2 docs : https://docs.google.com/document/d/1thpe8IYbgXfvzfmI5s2YG9nZj9_HnF9Ck4N2qrHTw2k/edit?usp=sharing

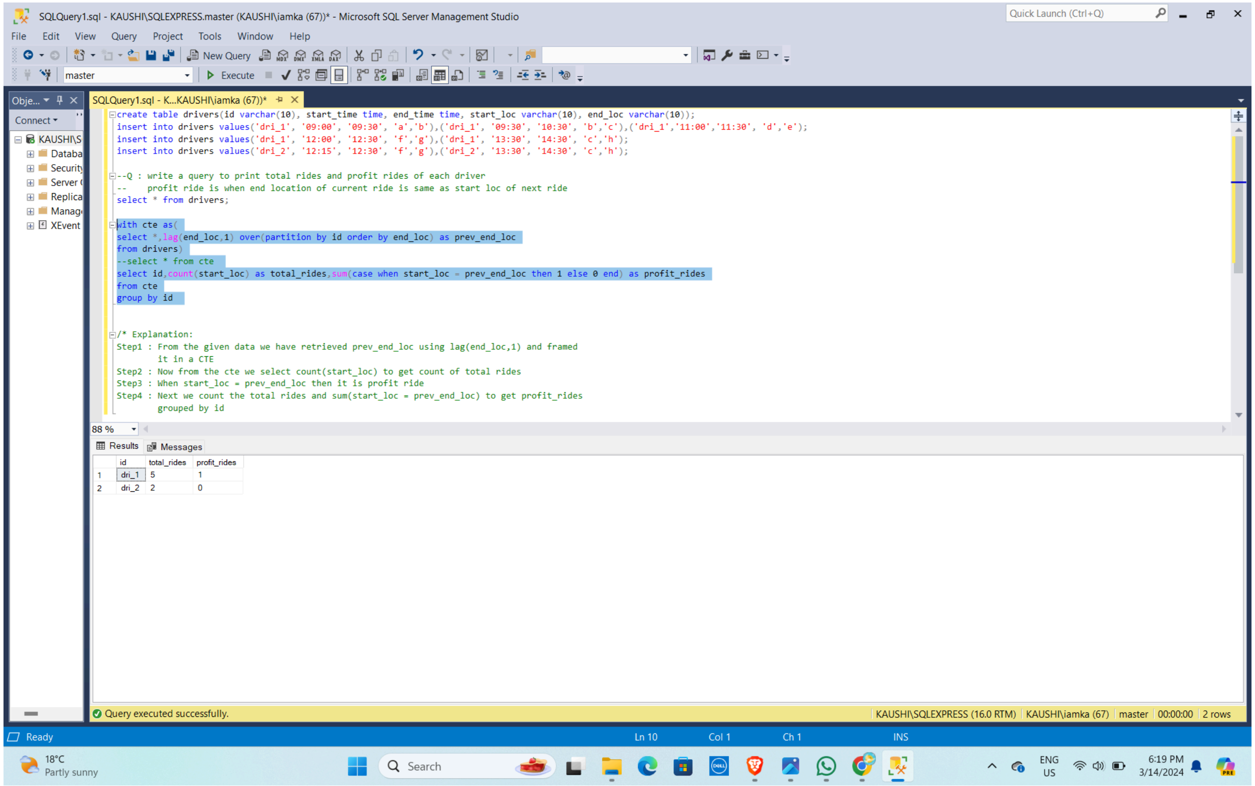
7. Ended my day by solving a complex SQL from Ankit'YT



```sql
create table drivers(id varchar(10), start_time time, end_time time, start_loc varchar(10), end_loc varchar(10));
insert into drivers values('dri_1', '09:00', '09:30', 'a','b'),('dri_1', '09:30', '10:30', 'b','c'),('dri_1','11:00','11:30', 'd','e');
insert into drivers values('dri_1', '12:00', '12:30', 'f','g'),('dri_1', '13:30', '14:30', 'c','h');
insert into drivers values('dri_2', '12:15', '12:30', 'f','g'),('dri_2', '13:30', '14:30', 'c','h');

--Q : write a query to print total rides and profit rides of each driver
--    profit ride is when end location of current ride is same as start loc of next ride
select * from drivers;
```

	id	start_time	end_time	start_loc	end_loc
1	dri_1	09:00:00.0000000	09:30:00.0000000	a	b
2	dri_1	09:30:00.0000000	10:30:00.0000000	b	c
3	dri_1	11:00:00.0000000	11:30:00.0000000	d	e
4	dri_1	12:00:00.0000000	12:30:00.0000000	f	g
5	dri_1	13:30:00.0000000	14:30:00.0000000	c	h
6	dri_2	12:15:00.0000000	12:30:00.0000000	f	g
7	dri_2	13:30:00.0000000	14:30:00.0000000	c	h



```sql
create table drivers(id varchar(10), start_time time, end_time time, start_loc varchar(10), end_loc varchar(10));
insert into drivers values('dri_1', '09:00', '09:30', 'a','b'),('dri_1', '09:30', '10:30', 'b','c'),('dri_1','11:00','11:30', 'd','e');
insert into drivers values('dri_1', '12:00', '12:30', 'f','g'),('dri_1', '13:30', '14:30', 'c','h');
insert into drivers values('dri_2', '12:15', '12:30', 'f','g'),('dri_2', '13:30', '14:30', 'c','h');

--Q : write a query to print total rides and profit rides of each driver
--    profit ride is when end location of current ride is same as start loc of next ride
select * from drivers;

with cte as(
select *,lag(end_loc,1) over(partition by id order by end_loc) as prev_end_loc
from drivers)
--select * from cte
select id,count(start_loc) as total_rides,sum(case when start_loc = prev_end_loc then 1 else 0 end) as profit_rides
from cte
group by id


/* Explanation:
Step1 : From the given data we have retrieved prev_end_loc using lag(end_loc,1) and framed
        it in a CTE
Step2 : Now from the cte we select count(start_loc) to get count of total rides
Step3 : When start_loc = prev_end_loc then it is profit ride
Step4 : Next we count the total rides and sum(start_loc = prev_end_loc) to get profit_rides
        grouped by id
```

	id	total_rides	profit_rides
1	dri_1	5	1
2	dri_2	2	0