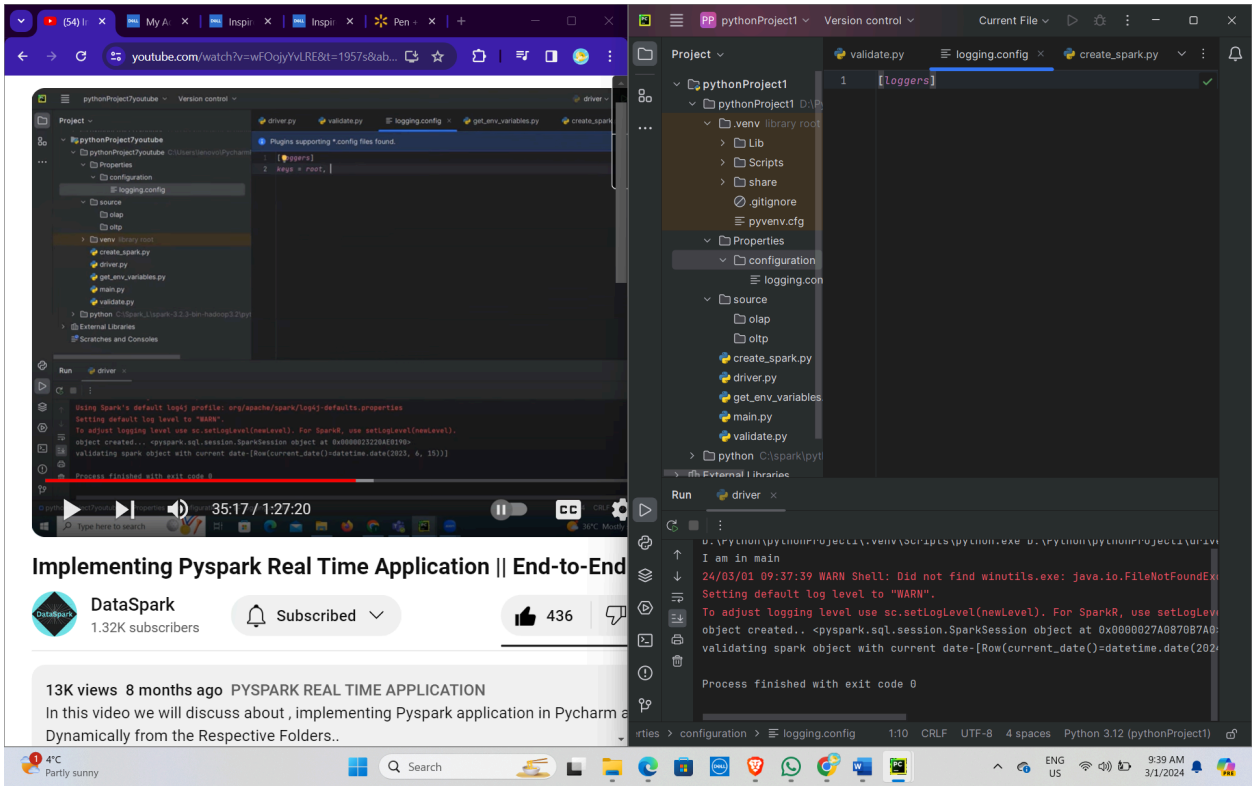
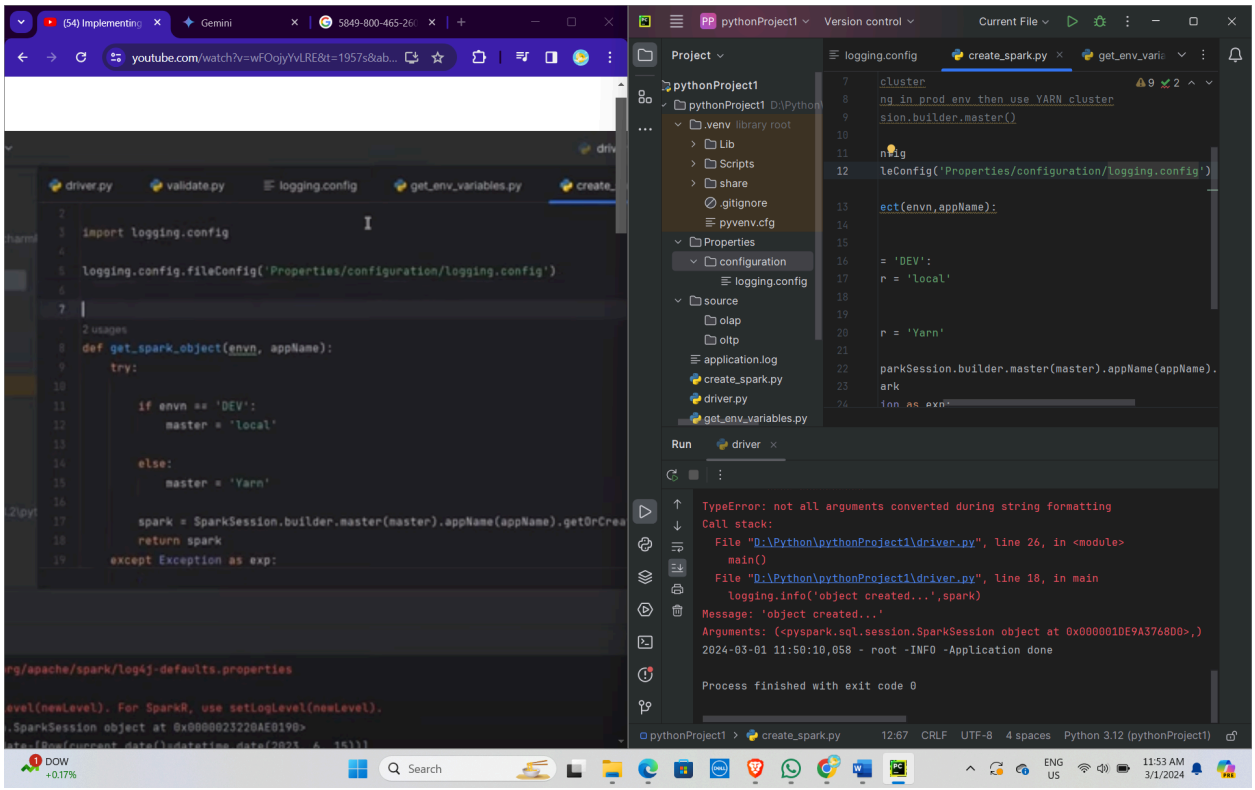
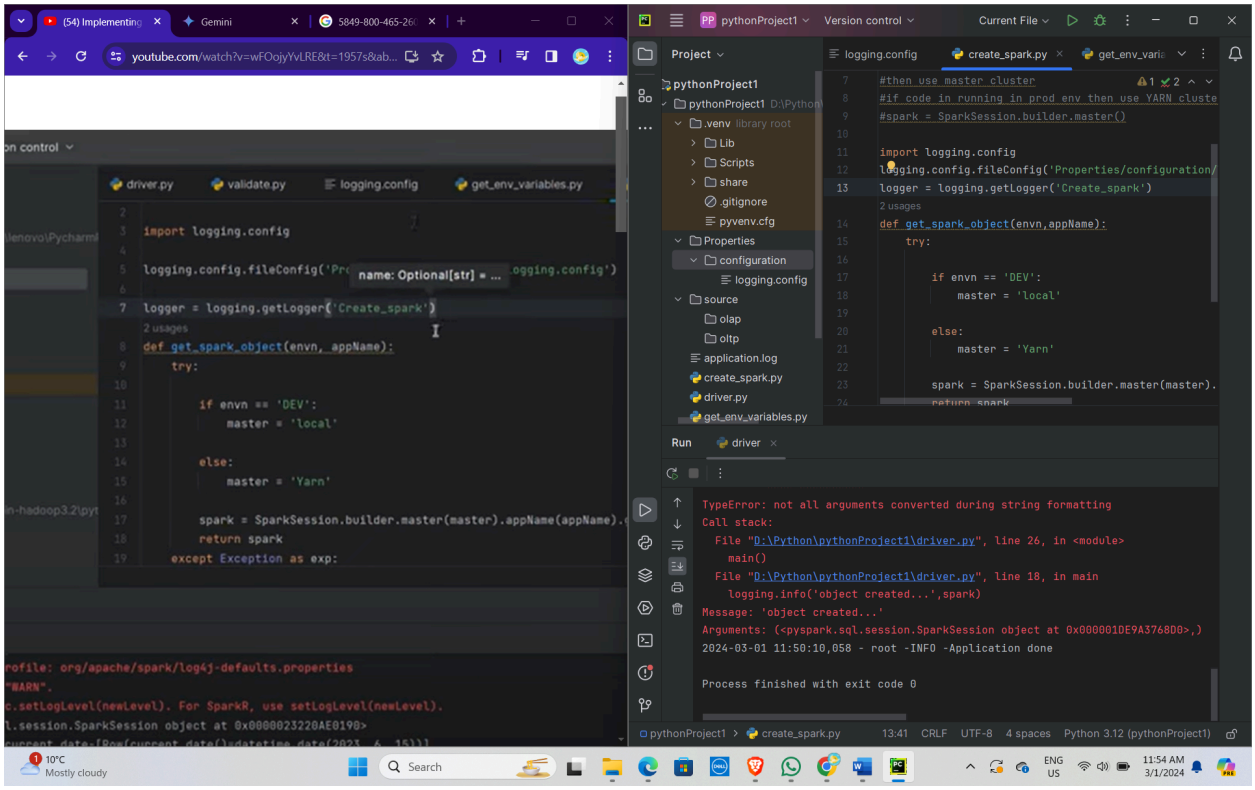
Day25 - March1st 2024

1. Woked up early and started my day
2. Cooked food for mrng and for afternoon
3. Solved one medium leetcode problem of binary search
4. A realtime pyspark project is in progress
5. Faced with some errors solved it on my own

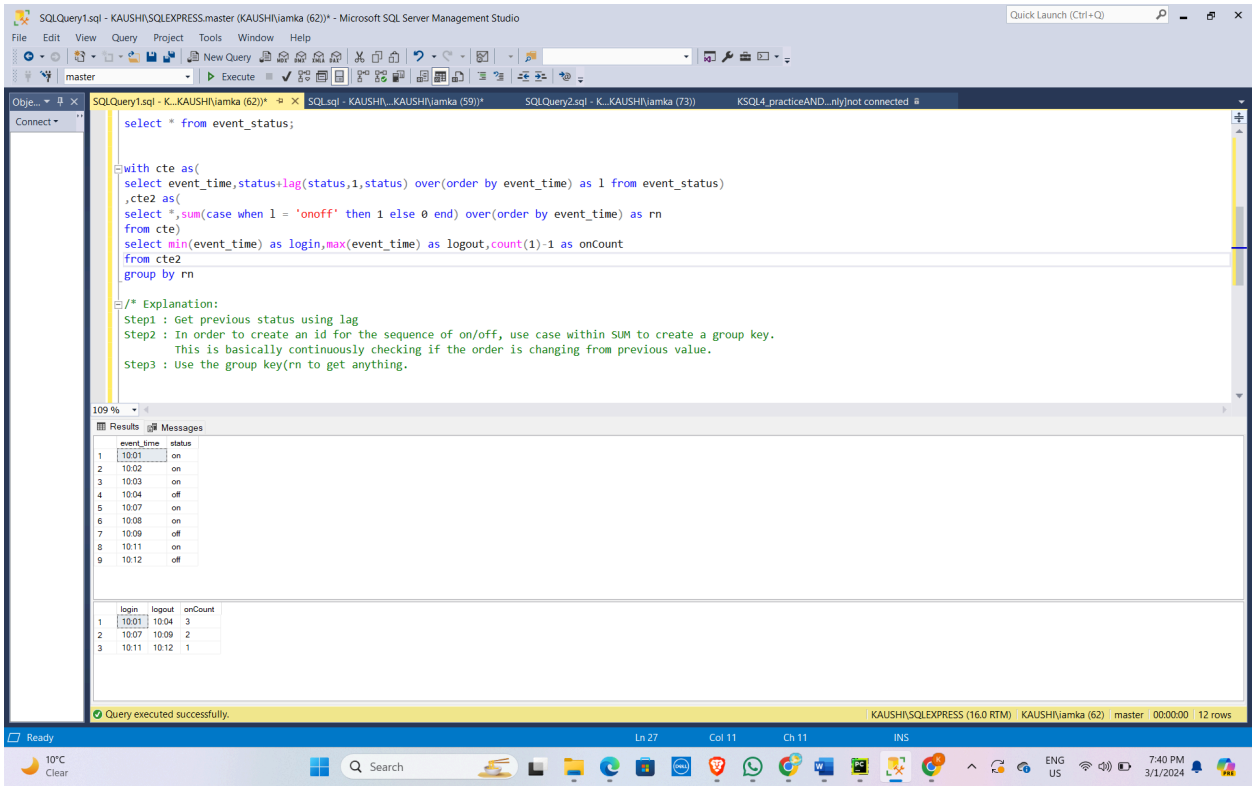
6.

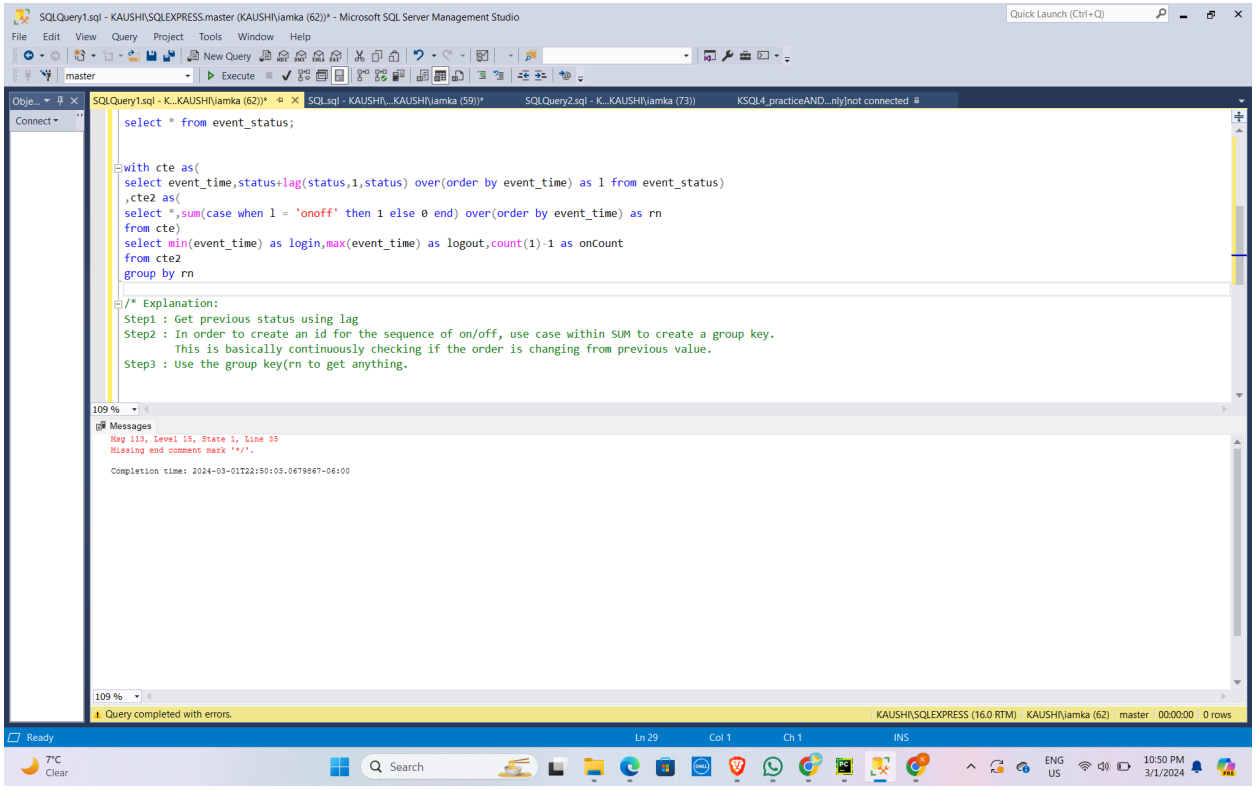

youtube.com/watch?v=wFOojyYvLRE&t=1957s&ab...
35:17 / 1:27:20
Implementing Pyspark Real Time Application || End-to-End
DataSpark
1.32K subscribers
Subscribed
436
13K views 8 months ago PYSPARK REAL TIME APPLICATION
In this video we will discuss about , implementing Pyspark application in Pycharm
Dynamically from the Respective Folders..
pythonProject1
Version control
Current File
Project
validate.py
logging.config
create_spark.py
[loggers]
Run
driver
I am in main
24/03/01 09:37:39 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundEx
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLev
object created.. <pyspark.sql.session.SparkSession object at 0x0000027A0870B7A0
validating spark object with current date-[Row(current_date()=datetime.date(202
Process finished with exit code 0
4°C
Partly sunny
Search
ENG
US
9:39 AM
3/1/2024

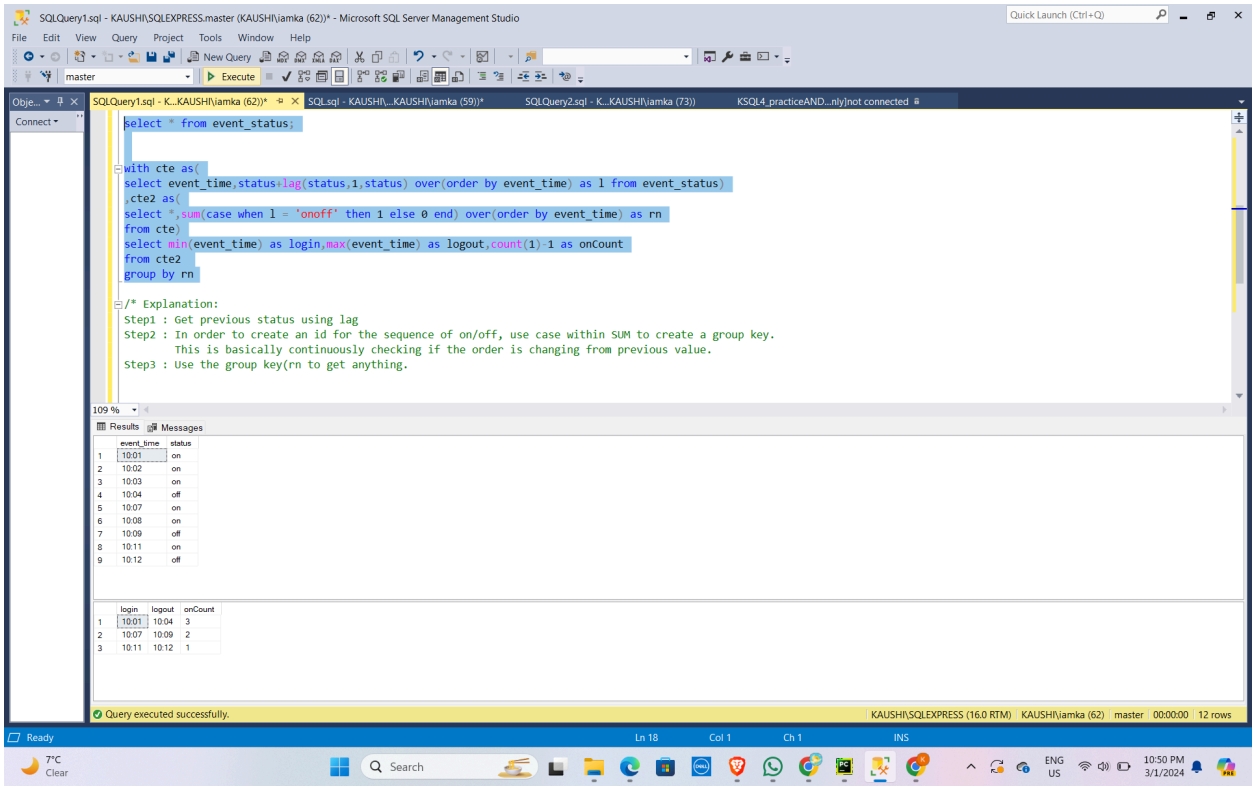

youtube.com/watch?v=wFOojyYvLRE&t=1957s&ab...
driver.py
validate.py
logging.config
get_env_variables.py
import logging.config
logging.config.fileConfig('Properties/configuration/logging.config')
def get_spark_object(envn, appName):
try:
if envn == 'DEV':
master = 'local'
else:
master = 'Yarn'
spark = SparkSession.builder.master(master).appName(appName).getOrCrea
return spark
except Exception as exp:
pythonProject1
Version control
Current File
Project
logging.config
create_spark.py
get_env_vari
Run
driver
TypeError: not all arguments converted during string formatting
Call stack:
File "D:\Python\pythonProject1\driver.py", line 26, in <module>
main()
File "D:\Python\pythonProject1\driver.py", line 18, in main
logging.info('object created...',spark)
Message: 'object created...'
Arguments: (<pyspark.sql.session.SparkSession object at 0x000001DE9A3768D0>,)
2024-03-01 11:50:10,058 - root -INFO -Application done
Process finished with exit code 0
DOW
+0.17%
Search
ENG
US
11:53 AM
3/1/2024

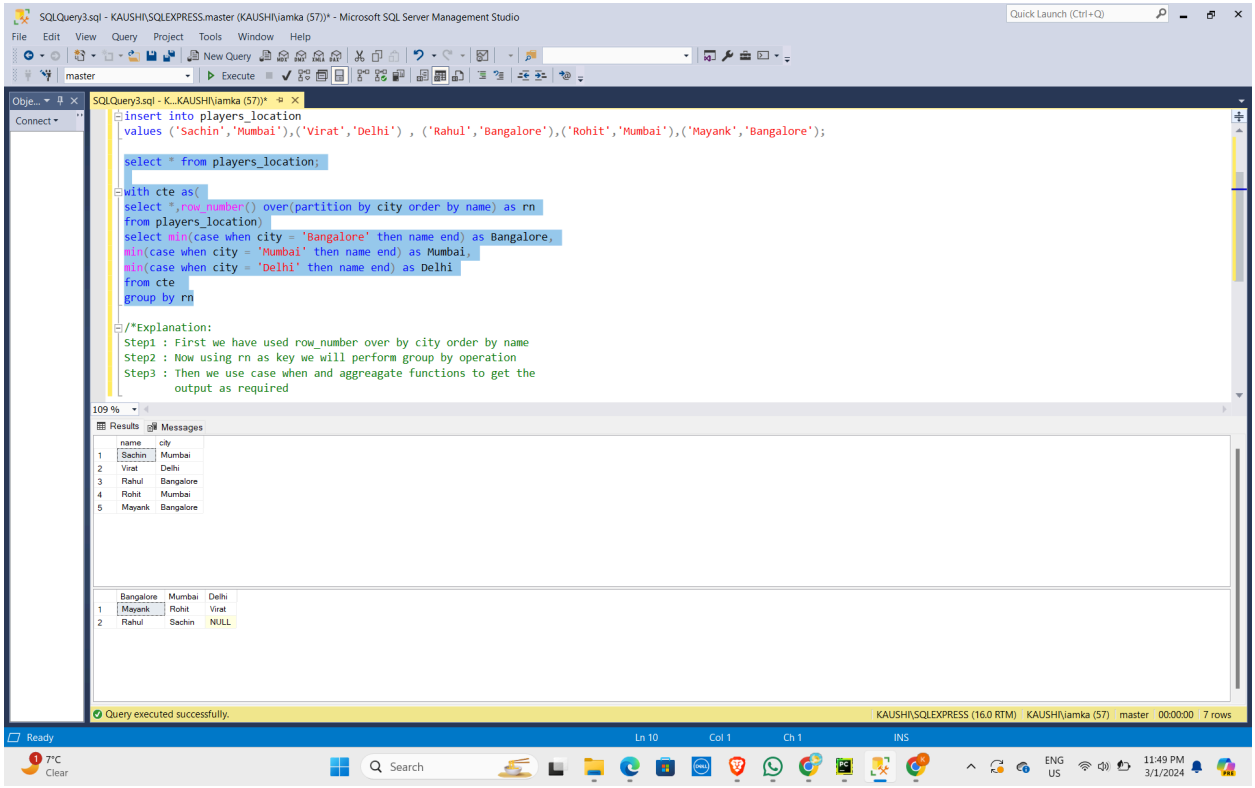


7. Ended my day by solving a real world SQL problem from Ankit's YT


SQLQuery1.sql - KAUSHI\SQLEXPRESS.master (KAUSHI\iamka (62))* - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
New Query
Execute
master
SQLQuery1.sql - K...KAUSHI\iamka (62))*
SQL.sql - KAUSHI\...KAUSHI\iamka (59))*
SQLQuery2.sql - K...KAUSHI\iamka (73))
KSQL4_practiceAND...nly]not connected
select * from event_status;
with cte as(
select event_time,status+lag(status,1,status) over(order by event_time) as l from event_status)
,cte2 as(
select *,sum(case when l = 'onoff' then 1 else 0 end) over(order by event_time) as rn
from cte)
select min(event_time) as login,max(event_time) as logout,count(1)-1 as onCount
from cte2
group by rn
/* Explanation:
Step1 : Get previous status using lag
Step2 : In order to create an id for the sequence of on/off, use case within SUM to create a group key.
This is basically continuously checking if the order is changing from previous value.
Step3 : Use the group key(rn to get anything.
109 %
Results
Messages
event_time status
1 10:01 on
2 10:02 on
3 10:03 on
4 10:04 off
5 10:07 on
6 10:08 on
7 10:09 off
8 10:11 on
9 10:12 off
login logout onCount
1 10:01 10:04 3
2 10:07 10:09 2
3 10:11 10:12 1
Query executed successfully.
KAUSHI\SQLEXPRESS (16.0 RTM) KAUSHI\iamka (62) master 00:00:00 12 rows
Ready
Ln 27 Col 11 Ch 11 INS
10°C Clear
Search
ENG US
7:40 PM 3/1/2024


SQLQuery1.sql - KAUSHI\SQLEXPRESS.master (KAUSHI\iamka (62))* - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
New Query
Execute
master
SQLQuery1.sql - K...KAUSHI\iamka (62))*
SQL.sql - KAUSHI\...KAUSHI\iamka (59))*
SQLQuery2.sql - K...KAUSHI\iamka (73))
KSQL4_practiceAND...nly]not connected
select * from event_status;
with cte as(
select event_time,status+lag(status,1,status) over(order by event_time) as l from event_status)
,cte2 as(
select *,sum(case when l = 'onoff' then 1 else 0 end) over(order by event_time) as rn
from cte)
select min(event_time) as login,max(event_time) as logout,count(1)-1 as onCount
from cte2
group by rn
/* Explanation:
Step1 : Get previous status using lag
Step2 : In order to create an id for the sequence of on/off, use case within SUM to create a group key.
This is basically continuously checking if the order is changing from previous value.
Step3 : Use the group key(rn to get anything.
109 %
Messages
Msg 113, Level 15, State 1, Line 35
Missing end comment mark '*/'.
Completion time: 2024-03-01T22:50:03.0679867-06:00
Query completed with errors.
KAUSHI\SQLEXPRESS (16.0 RTM) KAUSHI\iamka (62) master 00:00:00 0 rows
Ready
Ln 29 Col 1 Ch 1 INS
7°C Clear
Search
ENG US
10:50 PM 3/1/2024


SQLQuery1.sql - KAUSHI\SQLEXPRESS.master (KAUSHI\iamka (62))* - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
New Query
Execute
master
SQLQuery1.sql - K...KAUSHI\iamka (62))*
SQLQuery2.sql - K...KAUSHI\iamka (73))
select * from event_status;
with cte as(
select event_time,status+lag(status,1,status) over(order by event_time) as l from event_status)
,cte2 as(
select *,sum(case when l = 'onoff' then 1 else 0 end) over(order by event_time) as rn
from cte)
select min(event_time) as login,max(event_time) as logout,count(1)-1 as onCount
from cte2
group by rn
/* Explanation:
Step1 : Get previous status using lag
Step2 : In order to create an id for the sequence of on/off, use case within SUM to create a group key.
This is basically continuously checking if the order is changing from previous value.
Step3 : Use the group key(rn to get anything.
109 %
Results
Messages
event_time status
1 10:01 on
2 10:02 on
3 10:03 on
4 10:04 off
5 10:07 on
6 10:08 on
7 10:09 off
8 10:11 on
9 10:12 off
login logout onCount
1 10:01 10:04 3
2 10:07 10:09 2
3 10:11 10:12 1
Query executed successfully.
KAUSHI\SQLEXPRESS (16.0 RTM) KAUSHI\iamka (62) master 00:00:00 12 rows
Ready
Ln 18 Col 1 Ch 1 INS
7°C Clear
Search
ENG US
10:50 PM 3/1/2024


SQLQuery3.sql - KAUSHI\SQLEXPRESS.master (KAUSHI\iamka (57))* - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
New Query
Execute
master
SQLQuery3.sql - K...KAUSHI\iamka (57))*
insert into players_location
values ('Sachin','Mumbai'),('Virat','Delhi') , ('Rahul','Bangalore'),('Rohit','Mumbai'),('Mayank','Bangalore');
select * from players_location;
with cte as(
select *,row_number() over(partition by city order by name) as rn
from players_location)
select min(case when city = 'Bangalore' then name end) as Bangalore,
min(case when city = 'Mumbai' then name end) as Mumbai,
min(case when city = 'Delhi' then name end) as Delhi
from cte
group by rn
/*Explanation:
Step1 : First we have used row_number over by city order by name
Step2 : Now using rn as key we will perform group by operation
Step3 : Then we use case when and aggreagate functions to get the
output as required
109 %
Results
Messages
name city
1 Sachin Mumbai
2 Virat Delhi
3 Rahul Bangalore
4 Rohit Mumbai
5 Mayank Bangalore
Bangalore Mumbai Delhi
1 Mayank Rohit Virat
2 Rahul Sachin NULL
Query executed successfully.
KAUSHI\SQLEXPRESS (16.0 RTM) KAUSHI\iamka (57) master 00:00:00 7 rows
Ready
Ln 10 Col 1 Ch 1 INS
7°C Clear
Search
ENG US
11:49 PM 3/1/2024