

Bright Assignment

SQL and Python (1 hr 30 min)

- Create 'Customer' table having following schema:

id	Integer Primary Key
Name	character varying(128)
Email	character varying(128)
customer_id	character varying(16)
address	jsonb <i>NULL</i>
created_time	timestampz

Indexes:

Primary	id
Index	customer_id
Index	Email
Unique	customer_id
Unique	Email

Example address field:

```
{
  "house_number": 10,
  "building_name": "Manyata",
  "street": "vivekanand street",
  "city": "Nagpur",
  "state": "Maharashtra",
  "country": "India",
  "Zipcode": 440001
}
```

Example row in Customer table:

id	Name	Email	customer_id	address	created_time
1	John Parker	john@gmail.com	b07yrrujg4578hfk	NULL	2020-01-22 19:57:23.512484+00

- Create 'Account' table having following schema:

id	Integer Primary Key
account_id	character varying(128)
account_no	character varying(16)
account_balance	double precision <i>NULL</i>
created_time	timestampz
user_id	Integer

account_balance field has the current account balance.

Indexes:

Primary	id
Index	user_id
Unique	account_id
Index	account_balance

Foreign Keys:

Source	Target
user_id	Customer (id)

Example row in Account Table:

id	account_id	account_no	account_balance	user_id	created_time
7	gtr567uhg9cvbnh7	12345678900987	1000.78	1	2020-01-22 19:57:23.512484+00

- Create 'Transaction' table with following schema:

id	Integer Primary Key
transaction_id	character varying(16)
amount	double precision
transaction_time	timestampz

type	character varying(8) (eg. 'DEBIT'/'CREDIT')
account_id	Integer

Indexes:

Primary	id
Index	transaction_id
Unique	transaction_id

Foreign Keys:

Source	Target
account_id	Account (id)

Example row in Transaction Table:

id	transaction_id	amount	transaction_time	type	account_id
7	hy764k8hfh679ksn	78.28	2019-12-29 19:30:28.973662+00	CREDIT	7

1) Create a sqlite file DB, and populate above 3 tables with enough data for the following queries.

2) Write SQL queries for the following tasks:

1. Fetch all state names and total of account balance for the users belonging to those states.
2. Write a query with following details(display all subfields as columns in single table):
 - a. For the state with the highest account balance total, get the daily total of account balances for the past 10 days based on transaction history from the 'Transaction' table.
 - b. If there is no transaction available for a date, consider sum of account balance at previous date as sum of account balance at current date.
 - c. Repeat a. and b. for 10 richest users belonging to each of the 5 poorest states.
 - d. Find the weekly sum of daily differences between c. and d.
3. Write a query with following Details:

. Top 5 users from each state based on sum of transaction amount in the last 2 weeks, sum of the transaction amount for each of these users and sum of transaction amount of these 5 users combined for that state.

a. State wise cumulative daily sum of transaction amount for users from each state who are present in a.

b. Fetch and display state wise account balance sum of all such users who are present in c.

Submit your queries and sqlite DB file.

3) Write python code to read all these tables in memory and perform the same operations in the pandas library.

Bonus Question

4) Let us Assume that:

- Customer table has 50,000 different customer entries.
- Account table has 300,000 entries account Entries.
- Transaction table has 6 million Transaction Entries.

Some of the tables above are intentionally designed in an inefficient way, identify the issues and propose a solution to optimize the tables keeping in mind that the queries mentioned above will be executed very frequently(hourly basis), also write the advantages and disadvantages of the proposed optimization techniques for the given scenario.

Case Study (30 mins)

To curb Corona in India, assuming you are the Health Minister of India, you need to find out which states to focus on. What data would you collect and how would you make the decision?