EIU
Đại Học Quốc Tế Miền Đông
Eastern International University

8weeksqlchallenge

Data Bank

Presented by

Hugo Nguyen

Subject: MIS 443

Lecturer: Dang Thai Doan

Quater: 2

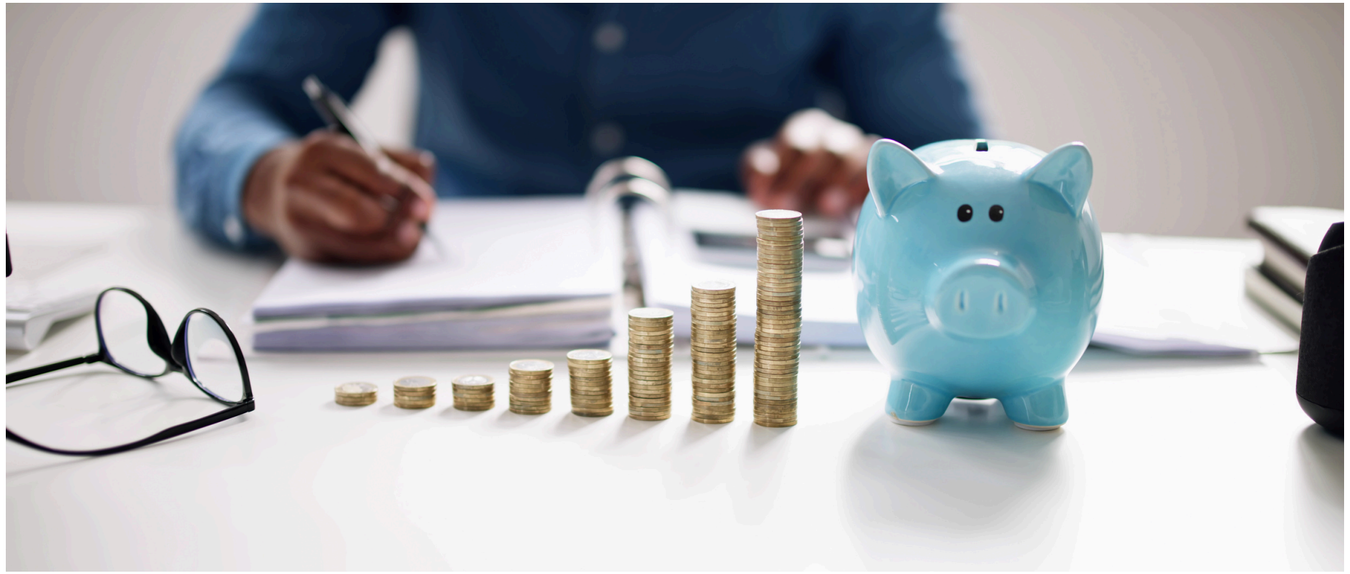
March 19, 2025

EIU
Đại Học Quốc Tế Miền Đông
Eastern International University

Table of Contents

EIU
Đại Học Quốc Tế Miền Đông
Eastern International University

Data Sources



1. Regions Table - Contains information about banking nodes worldwide.
2. Customer Nodes Table - Tracks customer allocations to banking nodes.
3. Customer Transactions Table - Stores deposits, withdrawals, and purchases.

Deliverables-A. Customer Nodes Exploration

How many unique nodes are there on the Data Bank system?

```
6  select
7      count(distinct node_id) as unique_nodes
8      from data_bank.customer_nodes
9  ;
```

	unique_nodes bigint
1	5

Insight: The system employs 5 global banking nodes, likely regional divisions, for data security and customer distribution, ensuring financial protection and balanced distribution across regions.

EIU
Đại Học Quốc Tế Miền Đông
Eastern International University

Deliverables-A. Customer Nodes Exploration

Đại Học Quốc Tế Miền Đông
Eastern International University

What is the number of nodes per region?

```
13  select regs.region_name,
14  count(distinct cns.node_id) as number_of_nodes
15  from data_bank.regions regs
16  inner join data_bank.customer_nodes cns
17  on regs.region_id=cns.region_id
18  group by regs.region_name
19  ;
```

	region_name character varying (9)	number_of_nodes bigint
1	Africa	5
2	America	5
3	Asia	5
4	Australia	5
5	Europe	5

Insight: The proposed infrastructure should ensure redundancy and security across all regions, but misnaming could cause issues in tracking financial activities, requiring correct correction for accurate data analysis.

Deliverables-A. Customer Nodes Exploration

Đại Học Quốc Tế Miền Đông
Eastern International University

How many days on average are customers reallocated to a different node?

```
23  select regs.region_name,
24  count(cns.customer_id) as customer_each_region
25  from data_bank.regions regs
26  inner join data_bank.customer_nodes cns
27  on regs.region_id=cns.region_id
28  group by regs.region_name
29  ;
```

	region_name character varying (9)	customer_each_region bigint
1	America	735
2	Australia	770
3	Africa	714
4	Asia	665
5	Europe	616

Insight: Australia's customer distribution is balanced, with 770 customers and 616 in Europe, suggesting potential for stronger engagement in Australia and targeted marketing strategies in Europe.

Deliverables-A. Customer Nodes Exploration

Đại Học Quốc Tế Miền Đông
Eastern International University

How many days on average are customers reallocated to a different node?

```sql
with node_days as(
select
customer_id,
node_id,
end_date - start_date as days_in_node
from data_bank.customer_nodes
where end_date != '9999-12-31'
group by customer_id, node_id, start_date, end_date
)
, total_node_days as (
select
customer_id,
node_id,
sum(days_in_node) as total_days_in_node
from node_days
group by customer_id, node_id
)

select round(avg(total_days_in_node)) as avg_node_reallocation_days
from total_node_days
;
```

	avg_node_reallocation_days numeric
1	24

Insight: Regular customer reallocation to different nodes, averaging 24 days, is crucial for security and system optimization, ensuring data protection but potentially causing minor disruptions for customers.

Deliverables-A. Customer Nodes Exploration

Đại Học Quốc Tế Miền Đông
Eastern International University

What is the median, 80th, and 95th percentile for reallocation days per region?

```sql
with node_days as (
select
cn.customer_id,
cn.region_id,
cn.node_id,
cn.end_date - cn.start_date as days_in_node
from data_bank.customer_nodes cn
where cn.end_date != '9999-12-31'
)
select
r.region_name,
nd.region_id,
percentile_cont(0.5) within group (order by nd.days_in_node) as median_days,
percentile_cont(0.8) within group (order by nd.days_in_node) as p80_days,
percentile_cont(0.95) within group (order by nd.days_in_node) as p95_days
from node_days nd
inner join data_bank.regions r
on nd.region_id = r.region_id
group by r.region_name, nd.region_id
order by nd.region_id
;
```

	region_name character varying (9)	region_id integer	median_days double precision	p80_days double precision	p95_days double precision
1	Australia	1	15	23	28
2	America	2	15	23	28
3	Africa	3	15	24	28
4	Asia	4	15	23	28
5	Europe	5	15	24	28

Insight: Customer reallocation takes 24 days for security and system optimization, moving customers to new nodes to prevent data storage in one location, potentially causing minor disruptions.

Deliverables - B. Customer Transactions Exploration

What is the unique count and total amount for each transaction type?

```sql
select
txn_type,
count(*) as unique_count,
sum(txn_amount) as total_amount
from data_bank.customer_transactions
group by txn_type
order by txn_type
;
```

	txn_type character varying (10)	unique_count bigint	total_amount bigint
1	deposit	2671	1359168
2	purchase	1617	806537
3	withdrawal	1580	793003

Insight: Data Bank's primary transaction type is deposits, suggesting customers use it as a savings or funding account, potentially indicating opportunities for increased spending activity through cashback promotions or rewards.

Deliverables - B. Customer Transactions Exploration

What is the average historical deposit count and amount for all customers?

```sql
select
round(avg(txn_count)) as avg_deposit_count,
round(avg(txn_amount)) as avg_deposit_amount
from (
select
customer_id,
count(customer_id) as txn_count,
avg(txn_amount) as txn_amount
from data_bank.customer_transactions
where txn_type='deposit'
group by customer_id
) as deposit
;
```

	avg_deposit_count numeric	avg_deposit_amount numeric
1	5	509

Insight: Data Bank's customers deposit 5 times with an average $509, likely aligning with payroll cycles. To increase deposit frequency, they could offer higher interest rates, automatic incentives, or tiered savings plans.

Deliverables - B. Customer Transactions Exploration

How many customers make more than one deposit and at least one purchase/withdrawal per month?

```sql
with monthly_transactions as (
select
customer_id,
date_part('month', txn_date) as mth,
sum(case when txn_type='deposit' then 1 else 0 end) as deposit_count,
sum(case when txn_type='purchase' then 1 else 0 end) as purchase_count,
sum(case when txn_type='withdrawal' then 1 else 0 end) as withdrawal_count
from data_bank.customer_transactions
group by customer_id, date_part('month', txn_date)
)

select
mth,
count(distinct customer_id) as customer_count
from monthly_transactions
where deposit_count>1
and (purchase_count >= 1 or withdrawal_count >= 1)
group by mth
order by mth
;
```

	mth double precision	customer_count bigint
1	1	168
2	2	181
3	3	192
4	4	70

Insight: The number of customers making multiple deposits and transactions increased in the first three months but declined in month 4, possibly due to seasonal trends, economic factors, or promotional incentives.

Sumary

1.The findings from this analysis help Data Bank understand:

- Customer behavior trends (deposits, spending, node reallocations)
- Security policies and banking habits
- Potential growth strategies (customer engagement, marketing efforts, and financial forecasting)

2. This report is structured for easy presentation, ensuring clarity in the business impact of SQL-based data analysis.

3. More focus should be placed on:

- Differences in regional policies
- Security concerns
- Customer preferences

4. Tailoring Data Bank's future improvements effectively based on these insights.