

عنوان: تست کوثری برای دیتابیس postgres

مقدمه:

این گزارش با استفاده از کوثری‌های SQL برای تجزیه و تحلیل داده‌های جدول "REF_CBS_SMS2"، سوالات مطرح شده را بررسی می‌کند. در ابتدا باید به این موضوع اشاره کرد که نوع ذخیره سازی داده ها در جدول از اهمیت بالایی برخوردار است، بنابراین در ابتدا به بررسی داده ها و نحوه ذخیره سازی آنها می‌پردازیم.

ذخیره سازی داده‌ها:

در ابتدا با استفاده از پایتون داده‌های مورد نظر را از فایل حاوی داده‌ها یعنی REF_CBS_SMS2.CSV می‌خوانیم. سپس آنها را در پایگاه داده ذخیره می‌کنیم.

```
columns = {  
    "ROAMSTATE_519": "SMALLINT",  
    "CUST_LOCAL_START_DATE_15": "BIGINT",  
    "CDR_ID_1": "BIGINT",  
    "CDR_SUB_ID_2": "SMALLINT",  
    "CDR_TYPE_3": "CHAR",  
    "SPLIT_CDR_REASON_4": "CHAR",  
    "RECORD_DATE": "TIMESTAMP",  
    "PAYTYPE_515": "SMALLINT",  
    "DEBIT_AMOUNT_42": "BIGINT",  
    "SERVICEFLOW_498": "SMALLINT",  
    "EVENTSOURCE_CATE_17": "CHAR",  
    "USAGE_SERVICE_TYPE_19": "SMALLINT",  
    "SPECIALNUMBERINDICATOR_534": "SMALLINT",  
    "BE_ID_30": "SMALLINT",  
    "CALLEDPARTYIMSI_495": "VARCHAR(255) DEFAULT NULL",  
    "CALLINGPARTYIMSI_494": "BIGINT",  
}
```

نوع ذخیره سازی داده‌ها در شکل بالا قابل مشاهده می‌باشد.

، ۱ و ۰ حاوی PAYTYPE_515، فقط عدد CDR_SUB_ID_2 می‌باشد، ۲ و ۰ فقط حاوی اعداد ROAMSTATE_519، 28 تا 21 حاوی اعداد دو رقمی در بازه تقریبی USAGE_SERVICE_TYPE_19، ۱۰ حاوی SERVICEFLOW_498، 3 و 2 و 0 حاوی SPECIALNUMBERINDICATOR_534، BE_ID_30، 10101 نیز حاوی عدد 30، بود که حجم کمتری را اشغال می‌کند. SMALLINT سازی این داده‌ها بهترین انتخاب نوع

اما فیلدهای CUST_LOCAL_START_DATE_15، CDR_ID_1، DEBIT_AMOUNT_42، CALLINGPARTYIMSI_494 به دلیل دارا بودن اعداد طولانی و بزرگ BIGINT برای آنها انتخاب شده است.

CDR_TYPE_3 و SPLIT_CDR_REASON_4 و EVENTSOURCE_CATE_17 به دلیل اینکه فقط حاوی یک حرف هستند به صورت char و CALLEDPARTYIMSI_495 نیز به دلیل خالی بودن و عدم اطلاع از محتویات آن به صورت varchar انتخاب شده که به صورت پیش‌فرض به صورت null ذخیره می‌شود.

کوئری‌ها:

کوئری 1: میزان درآمد روزانه به تومان چه قدر می‌باشد.

• کوئری:

```
-- this query took 296ms to complete.
SELECT
    date_trunc('day','RECORD DATE') as date,
    SUM("DEBIT AMOUNT_42"*POWER(10,-4))::int AS total_income
FROM "REF_CBS_SMS2" rcs
GROUP BY date
ORDER BY date ASC;
```

Results 1

	date	total income
1	2020-08-30 00:00:00.000	9,386,862

این کوئری میزان درآمد را بر اساس روز پیدا و میلی ریال را به تومان تبدیل می‌کند. ابتدا از فیلد زمان بازه های زمانی را به روز تبدیل و اقسیم بندی می‌کنیم تا راحت‌تر بتوان دسته‌بندی کرد و سپس با استفاده از تابع توان مبلغ موجود را در ضریب 10000 تقسیم می‌کنیم تا از میلی ریال به تومان برسیم و در ادامه مبلغ را به int تبدیل می‌کنیم تا اعشاری نمایش داده نشود.

کوئری 2: میزان درآمد با ریزدانی 15 دقیقه به ازای هر paytype چه قدر می‌باشد.

برای این سوال دو کوئری پیاده سازی شد سپس بر اساس زمان اجرا کوئری مناسب انتخاب و برای بقیه مراحل از آن استفاده کردیم.

• کوئری:

```
-- this query took 267ms to complete.
SELECT
    date_bin('15 minutes','RECORD DATE' , TIMESTAMP '2001-01-01') as date_time,
    "PAYTYPE 515" as payment,
    sum("DEBIT AMOUNT_42"*power(10,-4))::int as revenue
from "REF_CBS_SMS2" rcs
group by date_time,payment;
```

CBS_SMS2 1

date time	payment	revenue
2020-08-30 00:00:00.000	0	1,722,947
2020-08-30 00:00:00.000	1	652,891
2020-08-30 00:15:00.000	0	1,661,709
2020-08-30 00:15:00.000	1	620,119
2020-08-30 00:30:00.000	0	1,772,425
2020-08-30 00:30:00.000	1	621,323
2020-08-30 00:45:00.000	0	1,709,129
2020-08-30 00:45:00.000	1	626,318

در این کوئری با تقسیم بندی رکورد زمانی به بازه های 15 دقیقه با استفاده از تابع date_bin و تعیین کردن timestamp ، زمان‌ها در بازه های 15 دقیقه ای تقسیم شدند که بر اساس نوع پرداخت و زمان پرداخت نمایش داده شده‌اند.

کوئری 3: میزان حداکثر حداقل درآمد با ریزدانی 15 دقیقه به ازای هر paytype چه قدر میباشد.

- کوئری:

در این کوئری با تنها با اضافه کردن توابع max ,min میزان حداقل و حداکثر درآمد را محاسبه کردیم.

```
-- this query runs in 292ms.
SELECT
  date bin('15 minutes', "RECORD_DATE", TIMESTAMP '2001-01-01') AS date_time,
  "PAYTYPE_515" AS payment,
  SUM("DEBIT_AMOUNT_42" * POWER(10, -4))::INT AS total_revenue,
  MAX("DEBIT_AMOUNT_42" * POWER(10, -4))::INT AS max_revenue,
  MIN("DEBIT_AMOUNT_42" * POWER(10, -4))::INT AS min_revenue
FROM "REF_CBS_SMS2" rcs
GROUP BY date_time, payment;
```

date time	payment	total revenue	max revenue	min revenue
2020-08-30 00:00:00.000	0	1,722,947	1,643	-545
2020-08-30 00:00:00.000	1	652,891	2,180	0
2020-08-30 00:15:00.000	0	1,661,709	1,098	-545
2020-08-30 00:15:00.000	1	620,119	2,180	0
2020-08-30 00:30:00.000	0	1,772,425	545	-545
2020-08-30 00:30:00.000	1	621,323	7,739	0
2020-08-30 00:45:00.000	0	1,709,129	545	-545
2020-08-30 00:45:00.000	1	626,318	5,450	0

کوئری 4: میزان درآمد و تعداد رکورد با ریزدانی 15 دقیقه به ازای هر paytype چه قدر میباشد. با جویین زدن با فایل رفرنس مشخص شود.

- کوئری:

در این کوئری با join زدن دو جدول بر اساس paytype میتوان نوع پرداخت را که مشاهده کرد سپس میتوان با استفاده از تابع count تعداد رکوردهای هر نوع پرداخت را نیز محاسبه کرد.

```
SELECT
  date bin('15 minutes', "RECORD_DATE", TIMESTAMP '2001-01-01') AS date_time,
  pt.value AS paytype_value,
  SUM("DEBIT_AMOUNT_42" * POWER(10, -4))::INT AS total_revenue,
  COUNT(*) AS record_count
FROM "REF_CBS_SMS2" rcs
INNER JOIN paytype pt ON rcs."PAYTYPE_515" = pt.pay_type
GROUP BY date_time, rcs."PAYTYPE_515", pt.value;
```

date time	paytype value	total revenue	record count
2020-08-30 00:00:00.000	Prepaid	1,722,947	135,930
2020-08-30 00:00:00.000	Postpaid	652,891	53,094
2020-08-30 00:15:00.000	Prepaid	1,661,709	133,479
2020-08-30 00:15:00.000	Postpaid	620,119	51,286
2020-08-30 00:30:00.000	Prepaid	1,772,425	139,158
2020-08-30 00:30:00.000	Postpaid	621,323	51,128
2020-08-30 00:45:00.000	Prepaid	1,709,129	135,189
2020-08-30 00:45:00.000	Postpaid	626,318	50,149