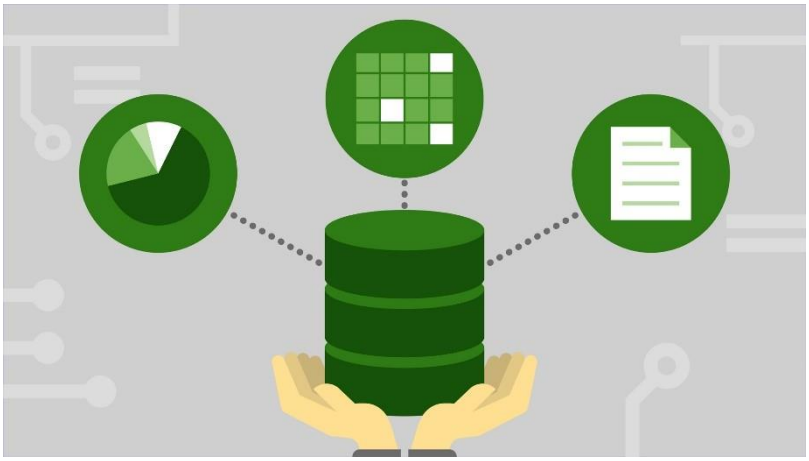
به نام خدا

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



آزمایشگاه پایگاه داده

دستور کار شماره ۹

سجاد علی‌زاده
۸۱۰۱۹۷۵۴۷

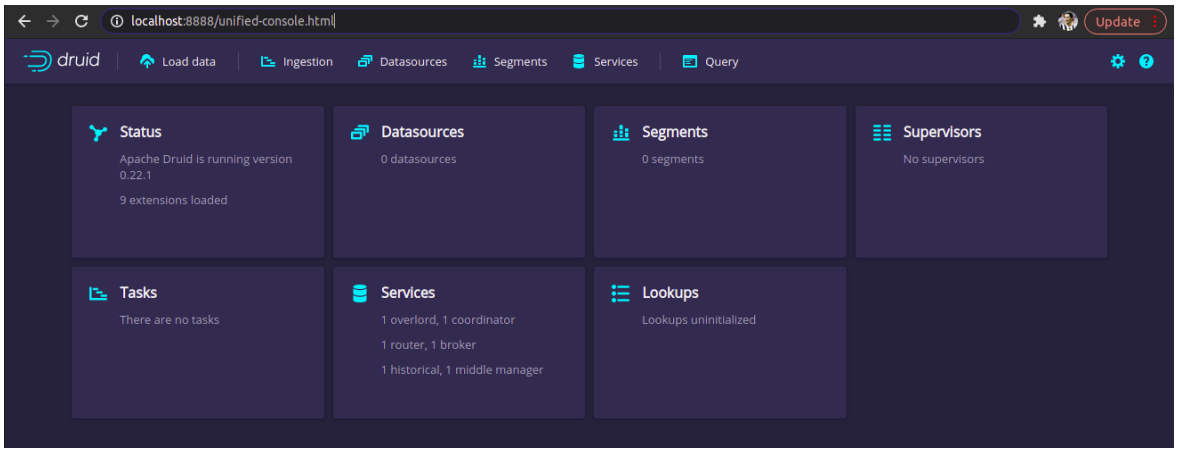
بهمن ماه ۱۴۰۰

گزارش دستورکار انجام شده

در ابتدا به نصب و راه اندازی دروید پرداختیم. نسخه جاوا و پورت مربوطه را وریفای میکنیم و سپس آن را اجرا میکنیم:

```
→ apache-druid-0.22.1 export DRUID_SKIP_JAVA_CHECK=1
→ apache-druid-0.22.1 ./bin/verify-java
→ apache-druid-0.22.1 ./bin/verify-default-ports
→ apache-druid-0.22.1 ./bin/start-micro-quickstart
[Mon Jan 24 15:26:03 2022] Running command[zk], logging to[/home/sajjad/DB/apache-druid-0.22.1/var/sv/zk.log]: bin/run-zk conf
[Mon Jan 24 15:26:03 2022] Running command[coordinator-overlord], logging to[/home/sajjad/DB/apache-druid-0.22.1/var/sv/coordinator-overlord.log
]: bin/run-druid coordinator-overlord conf/druid/single-server/micro-quickstart
[Mon Jan 24 15:26:03 2022] Running command[broker], logging to[/home/sajjad/DB/apache-druid-0.22.1/var/sv/broker.log]: bin/run-druid broker conf
/druid/single-server/micro-quickstart
[Mon Jan 24 15:26:03 2022] Running command[router], logging to[/home/sajjad/DB/apache-druid-0.22.1/var/sv/router.log]: bin/run-druid router conf
/druid/single-server/micro-quickstart
[Mon Jan 24 15:26:03 2022] Running command[historical], logging to[/home/sajjad/DB/apache-druid-0.22.1/var/sv/historical.log]: bin/run-druid his
torical conf/druid/single-server/micro-quickstart
[Mon Jan 24 15:26:03 2022] Running command[middleManager], logging to[/home/sajjad/DB/apache-druid-0.22.1/var/sv/middleManager.log]: bin/run-dru
id middleManager conf/druid/single-server/micro-quickstart
```

مشاهده میکنیم بدون مشکل وریفای شد و اجرا شد. برای چک کردن آن به آدرس localhost:8000 مراجعه میکنیم:

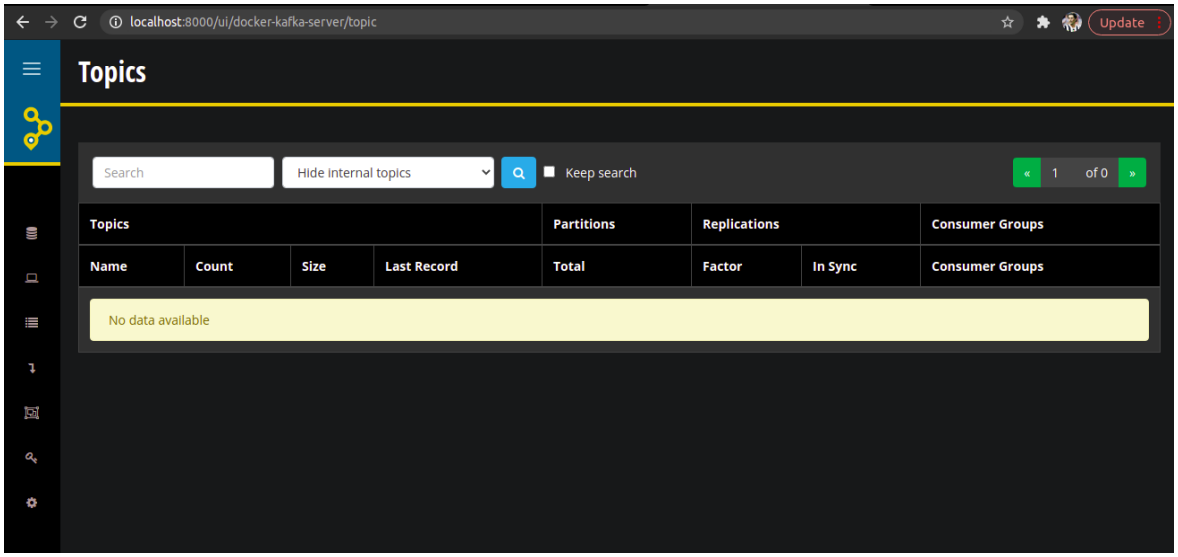


مشاهده میشود همه چیز بدون مشکل پیش رفته است.

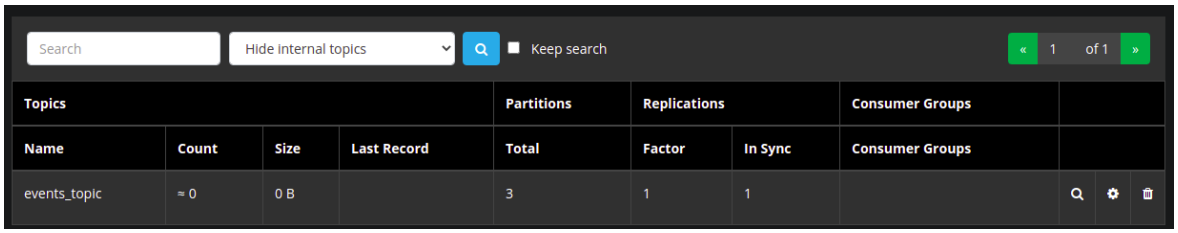
حال داکر کامپوز مربوطه را اجرا میکنیم:

```
   [Date: 2022-01-24T21:51:25.057958Z] [Duration: 1 ms] [Url: GET /ui/static/med
ia/logo.45903e1f.svg] [Status: 200] [Ip: /172.18.0.1] [User: -]
akhq_1        | 2022-01-24 21:51:25,205 INFO  r-thread-2 org.akhq.log.access
   [Date: 2022-01-24T21:51:25.188478Z] [Duration: 15 ms] [Url: GET /api/docker-k
afka-server/ui-options] [Status: 200] [Ip: /172.18.0.1] [User: -]
akhq_1        | 2022-01-24 21:51:25,339 INFO  pGroup-1-6 org.akhq.log.access
   [Date: 2022-01-24T21:51:25.334981Z] [Duration: 4 ms] [Url: GET /ui/static/med
ia/icon.648ce9c8.svg] [Status: 200] [Ip: /172.18.0.1] [User: -]
akhq_1        | 2022-01-24 21:51:25,483 INFO  pGroup-1-4 org.akhq.log.access
   [Date: 2022-01-24T21:51:25.482602Z] [Duration: 0 ms] [Url: GET /ui/static/med
ia/fontawesome-webfont.af7ae505.woff2] [Status: 200] [Ip: /172.18.0.1] [User: -]
akhq_1        | 2022-01-24 21:51:26,290 INFO  r-thread-2 org.akhq.log.access
   [Date: 2022-01-24T21:51:25.451825Z] [Duration: 838 ms] [Url: GET /api/docker-
kafka-server/topic] [Status: 200] [Ip: /172.18.0.1] [User: -]
```

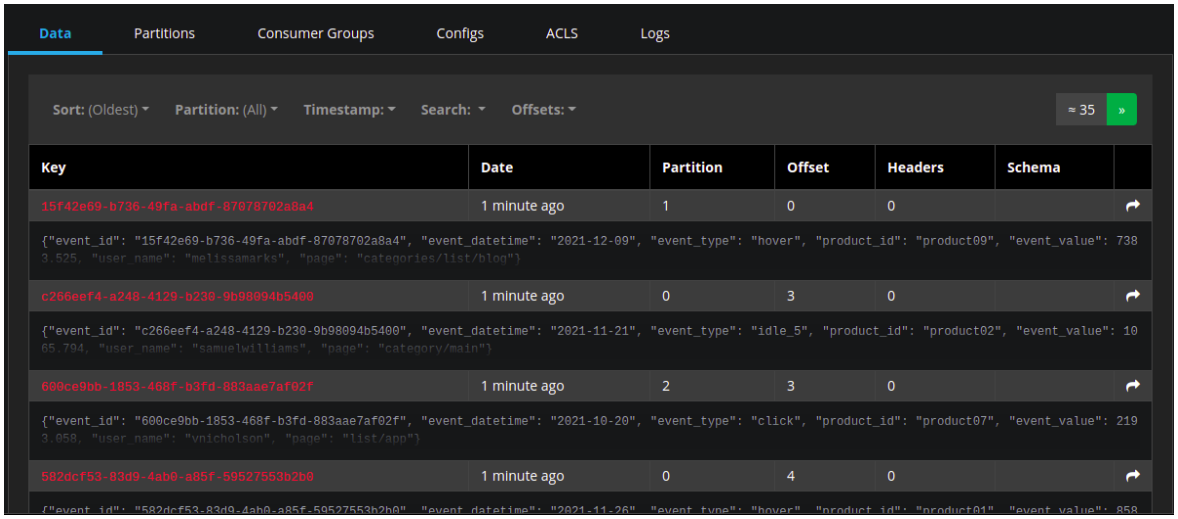
همچنین با مشاهده localhost:8000 متوجه میشویم همه چیز بدون مشکل اجرا شده است.



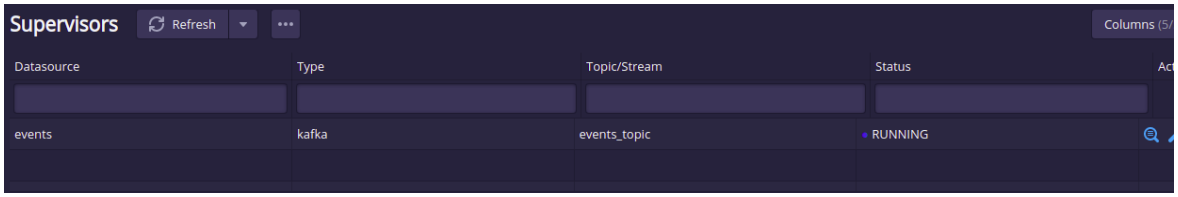
حال یک تاپیک با سه پارتیشن را تولید میکنیم:



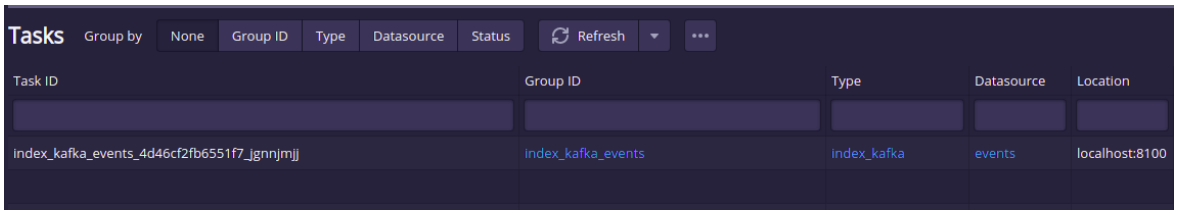
با اجرای کد پایتون داده شده مشاهده میکنیم داده ها به تاپیک تولید شده اضافه میشود.



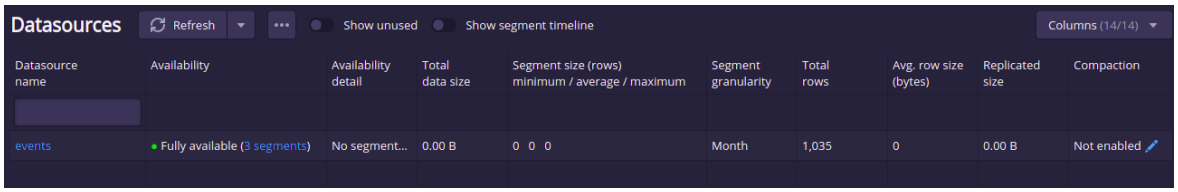
یک اینجسشن تولید میکنیم. مراحل آن را دقیقا مانند ویدیو پیش میبریم. در نهایت موفقیت آمیز بودن آن را مشاهده میکنیم:



همانطورکه مشاهده میشود وضعیت آن رانینگ است. تسک مربوط به آن نیز به شکل زیر است:

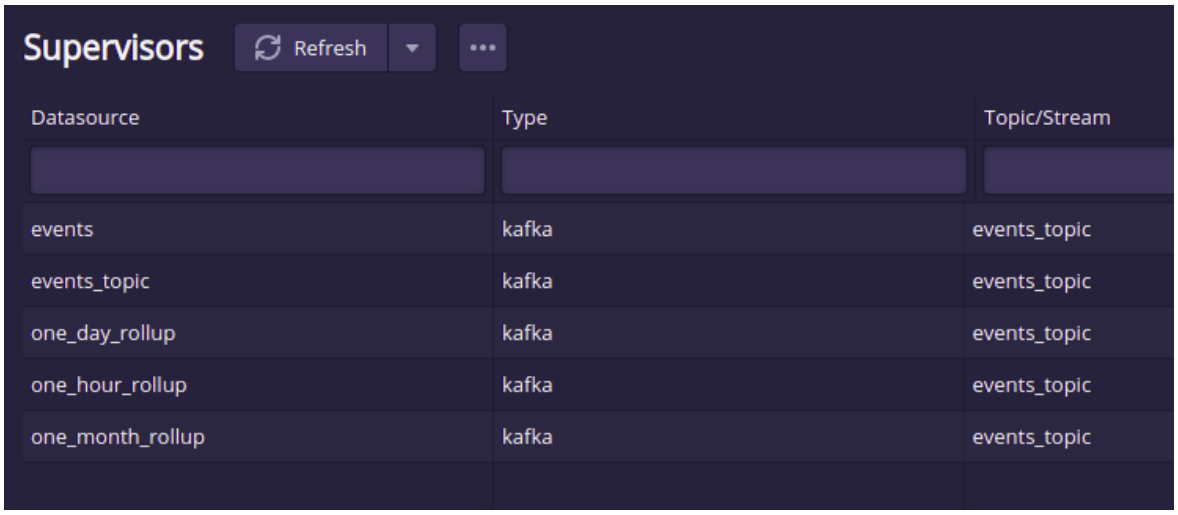


همچنین دیتاسورس مربوطه را نیز چک میکنیم تا داده ها اضافه شده باشند:

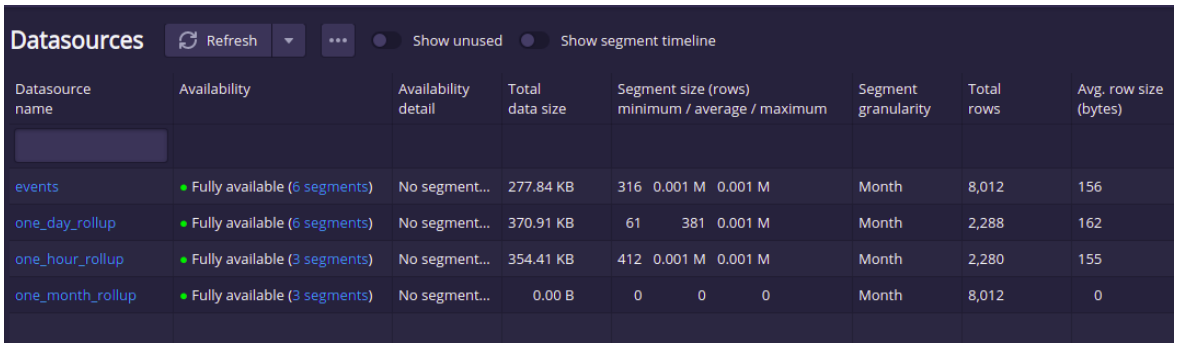


مشاهده میشود این دادهها بدون مشکل اضافه شده‌اند.

سه اینجسشن دیگر نیز تولید میکنیم که عملیات تجمیع را یک ساعت یکبار، یک روز یکبار و یک ماه یکبار انجام دهد:



دیتاسورس‌های آن نیز به شکل زیر می‌شوند:



همانطور که مشاهده میشود دیتاسورس غیرتجمیعی بیشترین داده را به خود اختصاص داده است.

کوئری اول:

```
SELECT user_name, COUNT(*) AS number_of_clicks
FROM events
WHERE __time >= CURRENT_TIMESTAMP - INTERVAL '6' MONTH AND event_type = 'click'
GROUP BY user_name
ORDER BY number_of_clicks DESC
```

در شش ماه گذشته یوزرهایی که event آنها از نوع کلیک بوده را استخراج میکنیم و بر اساس تعداد به صورت نزولی مرتب میکنیم. بدین ترتیب در ابتدا یوزرهایی می‌آیند که بیشترین event از نوع کلیک را داشتند.

نتیجه روی دیتاسورس events:

user_name	number_of_clicks
vfrank	6
amanda18	5
crystal34	5
donald07	5
ebailey	5
edward63	5
hjackson	5

نتیجه روی دیتاسورس one_day_rollup:

user_name	number_of_clicks
jordanjulie	3
ronniemcdonald	3
snovak	3
xbarber	3
yrogers	3
amycarr	2
arthurwalker	2

نتیجه روی دیتاسورس one_month_rollup:

user_name	number_of_clicks
vfrank	6
amanda18	5
crystal34	5
donald07	5
ebailey	5
edward63	5
hjackson	5

مشاهده میکنیم نتیجه دو کوئری روی دیتاسورس‌های one_month_rollup و events تقریبا مشابه است زیرا دیتاسورس one_month_rollup درشت‌دانه‌تر است و چیزهایی که در event است در آن هم هست.

کوئری دوم:

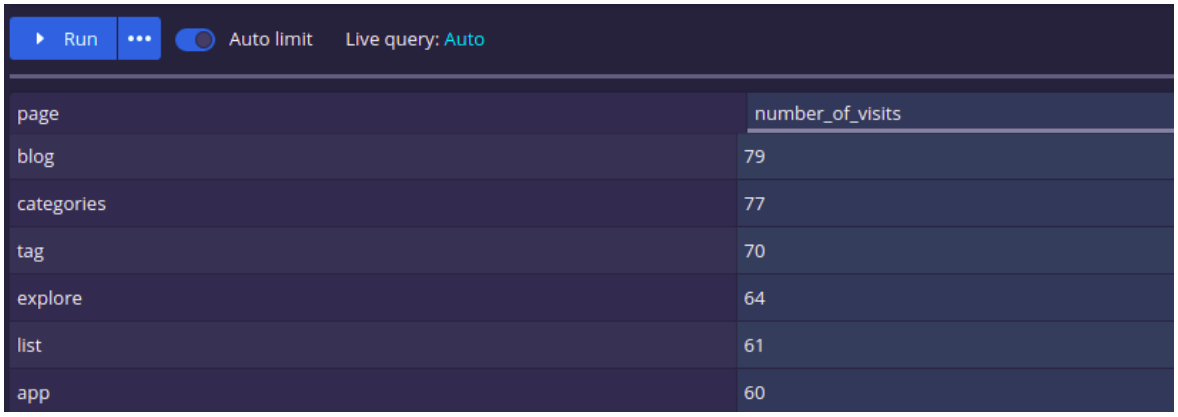
```
SELECT page, COUNT(*) as number_of_visits
FROM events
WHERE __time >= CURRENT_TIMESTAMP - INTERVAL '6' MONTH
GROUP BY page
ORDER BY number_of_visits DESC
```

در این کوئری به ازای هر صفحه تعداد ویزیت‌هایی که در شش ماه گذشته از آن شده است را می‌شماریم. در نهایت این تعداد را به صورت نزولی مرتب میکنیم. (فرض کرده‌ایم به ازای هر ویزیت یک رکورد در دیتاسورس ساخته می‌شود)

نتیجه روی دیتاسورس events:

page	number_of_visits
app	243
tag	237
blog	236
list	234
tags	226
categories	218
posts	217

نتیجه روی دیتاسورس one_day_rollup:



page	number_of_visits
blog	79
categories	77
tag	70
explore	64
list	61
app	60

نتیجه روی جدول one_month_rollup:

page	number_of_visits
app	243
tag	237
blog	236
list	234
tags	226
categories	218
posts	217

کوئری سوم:

```
SELECT product_id, page, COUNT(*) as num
FROM events
WHERE __time >= CURRENT_TIMESTAMP - INTERVAL '6' MONTH
GROUP BY product_id, page
ORDER BY num DESC
```

میخواهیم بفهمیم کدام محصول در کدام صفحه چه میزان تکرار شده است.

نتیجه در دیتاسورس events:

product_id	page	num
product07	app	39
product10	blog	37
product10	tag	35
product04	category	32
product04	blog	31
product04	tag	31

نتیجه در دیتاسورس one_day_rollup:

product_id	page	num
product10	tag	13
product07	list	12
product01	categories	11
product02	blog	11
product04	blog	11
product06	wp-content	11
product08	explore	11

نتیجه در دیتاسورس one_month_rollup:

product_id	page	num
product07	app	39
product10	blog	37
product10	tag	35
product04	category	32
product04	blog	31
product04	tag	31
product09	app	31
product09	tags	31

کوئری چهارم:

```
SELECT user_name, event_type, COUNT(*) as num
FROM events
WHERE __time >= CURRENT_TIMESTAMP - INTERVAL '6' MONTH
GROUP BY user_name, event_type
ORDER BY num DESC
```

میخواهیم ببینیم هر کاربر از هر event چه تعداد داشته است.

نتیجه در دیتاسورس events:

user_name	event_type	num
kyle25	buy	7
portergrace	hover	7
jonathan07	buy	6
kellysimmons	buy	6
kimberly06	hover	6
martinmary	hover	6
martinmary	idle_5	6

نتیجه در دیتاسورس one_day_rollup:

user_name	event_type	num
ericfloyd	idle_5	3
fharris	idle_5	3
jordanjulie	click	3
nscott	buy	3
rebecca09	idle_5	3
ronniemcdonald	click	3
silvakatie	idle_5	3

نتیجه در دیتاسورس one_month_rollup:

user_name	event_type	num
kyle25	buy	7
portergrace	hover	7
jonathan07	buy	6
kellysimmons	buy	6
kimberly06	hover	6
martinmary	hover	6
martinmary	idle_5	6

کوئری پنجم:

```
SELECT page, event_type, COUNT(*) as num
FROM events
WHERE __time >= CURRENT_TIMESTAMP - INTERVAL '6' MONTH AND user_name='angelayoung'
GROUP BY page, event_type
ORDER BY num DESC
```

میخواهیم ببینیم یک کاربر خاص در چه صفحاتی چه کارهایی کرده است. و تعداد آن را بشماریم.

نتیجه روی دیتاسورس events:

page	event_type	num
explore	buy	2
category	hover	1
explore	click	1
search/category	hover	1
tag	click	1
tags/tag	click	1

نتیجه روی دیتاسورس one_day_rollup:

page	event_type	num

Query returned no data

نتیجه روی دیتاسورس one_month_rollup:

page	event_type	num
explore	buy	2
category	hover	1
explore	click	1
search/category	hover	1
tag	click	1
tags/tag	click	1

مشکلات و توضیحات تکمیلی

هنگام راه‌اندازی روی wsl مشکلاتی داشتم که به دلیل آن مجبور شدم کارها را روی اوبونتوی واقعی انجام دهم.

آنچه آموختم / پیشنهادات

در این بخش که البته مانند بخش قبل، اختیاری است مهم‌ترین مطلبی که از دستورکار جاری یاد گرفته اید را می توانید ذکر کنید.

این مساله باعث میشود که فیدبک مناسبی از کیفیت و کارآیی دستورالعمل ها داشته باشیم. اگر پیشنهادی هم برای ارائه بهتر این بخش از آزمایشگاه داده در سالیان آتی دارید، در این قسمت آنرا ذکر نمایید.