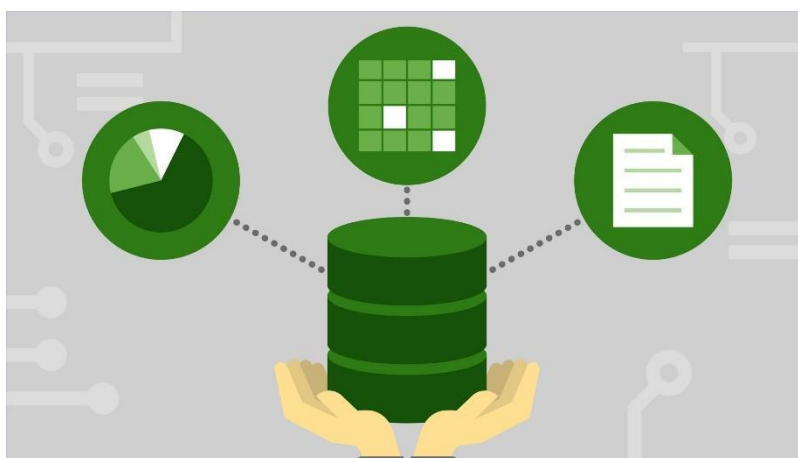


به نام خدا



دانشگاه تهران  
پردیس دانشکده‌های فنی  
دانشکده برق و کامپیوتر



آزمایشگاه پایگاه داده

دستور کار شماره 9

حمیدرضا خدادادی

810197499

بهمن ماه ۱۴۰۰

## گزارش دستور کار انجام شده

پس از مشاهده فیلم آموزشی و آشنایی با نحوه نصب و راه اندازی دروید و کلاستر کافکا، اقدام به پیاده سازی این گزارش کار کردم.

در ابتدا دروید را دانلود و نصب کردیم.

سپس از آن جایی که داکر را روی سیستم خود نداشتیم، آن را نصب کردیم.

سپس داکر کامپوز را نصب کردیم.

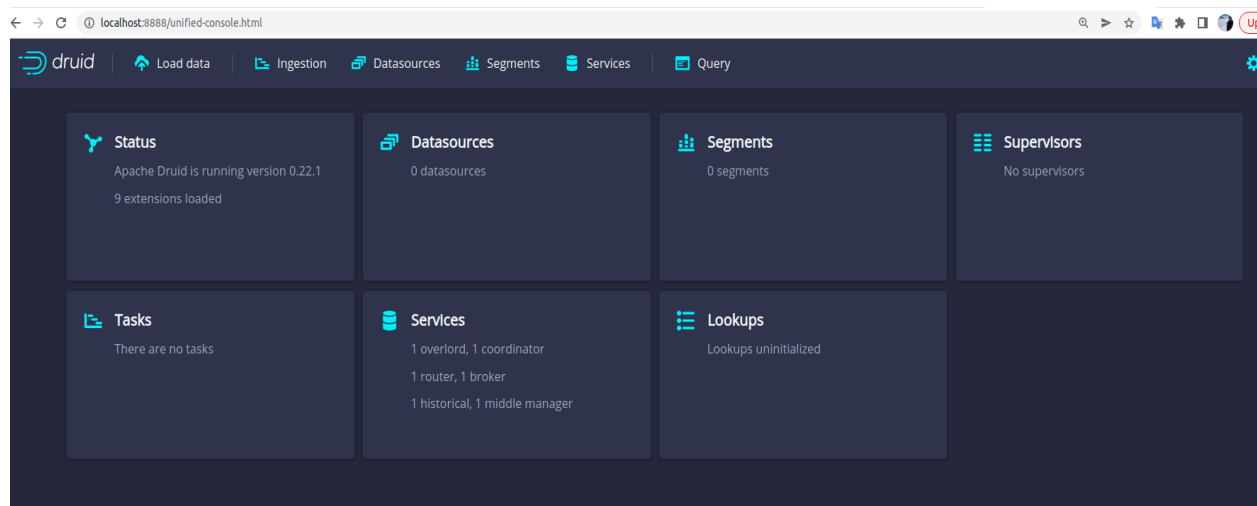
برای نصب دروید، در ابتدا verify-java را اجرا کردم و چون نسخه ای بالاتر از نسخه 8 جاوا داشتیم، کد `export DRUID_SKIP_JAVA_CHECK=1` را اجرا کردم. سپس دوباره verify-java را اجرا کردم که خروجی آن را در تصویر زیر می بینید.

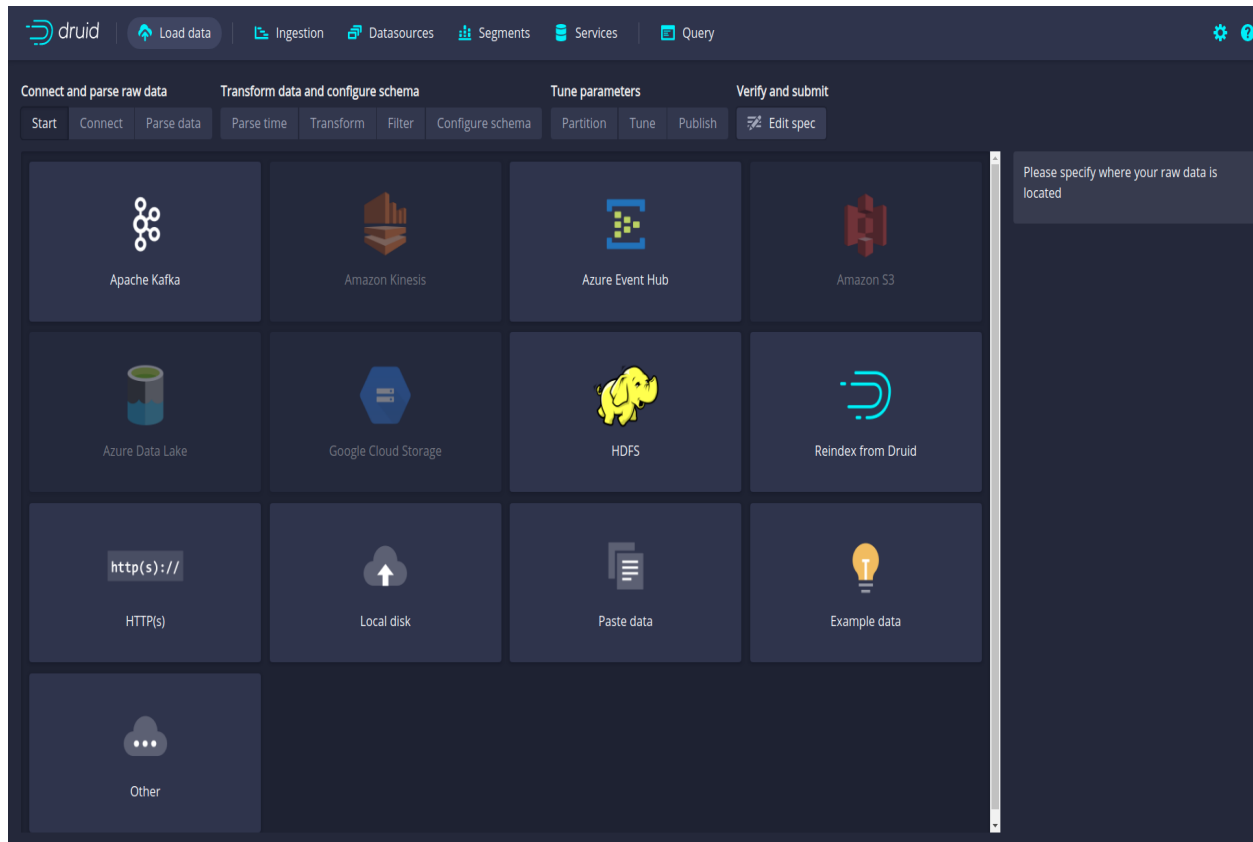
سپس verify-default-ports را اجرا کردم که خروجی آن را در تصویر زیر می بینید.

سپس از اطمینان از verify شدن، دروید را با کامند start-micro-quickstart بالا آوردم.

```
or get
* DRUID_JAVA_HOME
* JAVA_HOME
* java (installed on PATH)
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1$ export DRUID_SKIP_JAVA_CHECK=1
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1$ ./bin/verify-java
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1$ ./bin/verify-default-ports
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1$ ls
bin  conf  extensions  hadoop-dependencies  lib  LICENSE  licenses  NOTICE  quickstart  README
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1$ ls quickstart/tutorial
compaction-day-granularity.json  deletion-kill.json  transform-data.json  updates-data3.json  wikipedia-in
compaction-init-index.json      hadoop               transform-index.json  updates-data4.json  wikipedia-in
```

سپس پورت 8888 را بررسی کردم و دروید به خوبی بالا آمده بود.





سپس فایل docker compose را اجرا کردم که پس از نصب zookeeper و چند دیپندنسی دیگر، روی هاست 8000 رابط کاربری کافکا را بالا آورد.

```
" : dial tcp: lookup auth.docker.io on 216.146.35.35:53: read tcp 10.0.85.1:35710->216.146.35.35:53: i/o timeout
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1/workshops-main/1400-druid/Docker$ sudo docker-compose up
Pulling zookeeper (confluentinc/cp-zookeeper:6.2.0)...
6.2.0: Pulling from confluentinc/cp-zookeeper
c2c17d84f25a: Pull complete
46cdcde062b2: Pull complete
e73521032be7: Pull complete
77c96a9d0aa1: Pull complete
ff3e60059bc7: Pull complete
0b6fc8737347: Pull complete
1d8caf36535f: Pull complete
3b830c3d08d4: Pull complete
16d280dca34c: Pull complete
2308e5736c70: Pull complete
6855ebdfcdc5: Pull complete
Digest: sha256:9a69c03fd1757c3154e4f64450d0d27a6dec0dc3a1e401e8fc38e5cea881847
Status: Downloaded newer image for confluentinc/cp-zookeeper:6.2.0
Pulling kafka (confluentinc/cp-kafka:6.2.0)...
6.2.0: Pulling from confluentinc/cp-kafka
c2c17d84f25a: Already exists
46cdcde062b2: Already exists
e73521032be7: Already exists
77c96a9d0aa1: Already exists
ff3e60059bc7: Already exists
0b6fc8737347: Already exists
1d8caf36535f: Already exists
3b830c3d08d4: Already exists
16d280dca34c: Already exists
59e2e89eee9d: Pulling fs layer
890d6a70df7a: Pulling fs layer
```

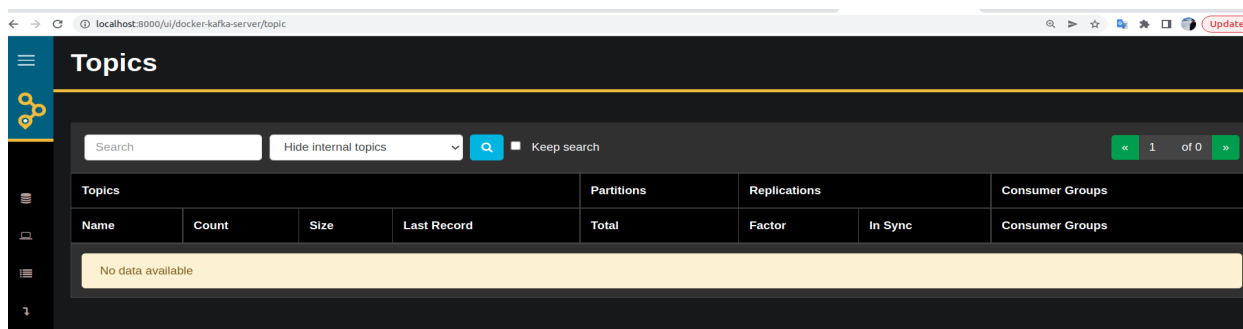
سپس کتابخانه ها و دیپندنسی های لازم آورده شده در فایل requirements.txt را با pip نصب کردم.

```
You should consider upgrading via the '/usr/bin/python3 -m pip install --upgrade pip' command.
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1/workshops-main/1400-druid/Codes$ pip install -r ./requirements.txt
WARNING: Value for scheme.platlib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.8/dist-packages
sysconfig: /usr/lib/python3.8/site-packages
WARNING: Value for scheme.purelib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.8/dist-packages
sysconfig: /usr/lib/python3.8/site-packages
WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.8/UNKNOWN
sysconfig: /usr/include/python3.8
WARNING: Value for scheme.scripts does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
WARNING: Value for scheme.data does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
WARNING: Additional context:
user = False
home = None
root = None
prefix = None
Defaulting to user installation because normal site-packages is not writeable
Collecting kafka-python
  Downloading kafka_python-2.0.2-py2.py3-none-any.whl (246 kB)
    | 246 kB 100 kB/s
Collecting faker
  Downloading Faker-11.3.0-py3-none-any.whl (1.2 MB)
    | 1.2 MB 136 kB/s
Requirement already satisfied: python-dateutil>=2.4 in /usr/lib/python3/dist-packages (from faker->-r ./requirements.txt (line 2)) (2.7.3)
Collecting text-unidecode==1.3
  Downloading text_unidecode-1.3-py2.py3-none-any.whl (78 kB)
    | 78 kB 340 kB/s
Installing collected packages: text-unidecode, kafka-python, faker
WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /home/hamid/.local/include/python3.8/UNKNOWN
sysconfig: /home/hamid/.local/include/python3.8
WARNING: Additional context:
user = True
home = None
root = None
prefix = None
Successfully installed faker-11.3.0 kafka-python-2.0.2 text-unidecode-1.3
WARNING: You are using pip version 21.1; however, version 22.0 is available.
You should consider upgrading via the '/usr/bin/python3 -m pip install --upgrade pip' command.
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1/workshops-main/1400-druid/Codes$
```

سپس هاست kafka را به لیست هاست های سیستم اضافه کردم تا داده های فیک تولیدی به رابط کاربری کافکا منتقل شود.

```
[sudo] password for hamid:
hamid@hamid-ZenBook-UX434FLC-UX434FL:~$ sudo -- sh -c "echo 127.0.0.1 kafka >> /etc/hosts"
[sudo] password for hamid:
hamid@hamid-ZenBook-UX434FLC-UX434FL:~$
```

سپس پورت 8000 را بررسی کردیم تا واسط کاربری AKHQ مربوط به کافکا بالا باشد.



سپس یک topic جدید به نام events\_topic را در رابط کاربری کافکا ایجاد می کنیم. و تعداد پارتیشن های آن را برابر 3 قرار می دهیم.

localhost:8000/ui/docker-kafka-server/topic/create

## Create a topic

Name	events_topic
Partition	3
Replicator Factor	1
Cleanup Policy	<input checked="" type="radio"/> Delete <input type="radio"/> Compact <input type="radio"/> Delete and Compact
Retention	86400000

سپس اسکریپت پایتون user\_event\_producer.py را اجرا می کنیم.

این اسکریپت، داده های فیک تولید می کند و روی هاستی که به سیستم اضافه کردیم، این داده ها را به واسط کاربری کافکا منتقل می کند.

```
hamid@hamid-ZenBook-UX434FLC-UX434FL:~/Desktop/apache-druid-0.22.1/workshops-main/1400-druid/Codes$ python3 user_event_producer.py
# 0 - Event Created : 6e5f2b4c-7030-45df-a223-1fd84a9493c2- Value : 3981.167
# 1 - Event Created : d8121422-8a1b-416c-b0ed-bf6292bdfcc1- Value : 6884.947
# 2 - Event Created : b492ac3d-e795-4110-ad88-91adfea01cc5- Value : 6941.949
# 3 - Event Created : d6c0cfed-ea54-4875-b856-53ff10b52ee9- Value : 1854.601
# 4 - Event Created : 83976f04-2519-4d06-aabd-2072017f41be- Value : 4882.441
# 5 - Event Created : 59a64568-ad1d-4c2d-926a-3ad0e3c3ce13- Value : 2426.078
# 6 - Event Created : 162372b7-bb1d-4d9e-9bfe-ff39b2a4a8d1- Value : 5986.388
# 7 - Event Created : 6749b2f2-9283-4056-8673-47cd842b6bce- Value : 8114.247
# 8 - Event Created : d38f8538-029e-4c2d-9895-095734f5b0e1- Value : 3499.505
# 9 - Event Created : a310ba90-54ef-449a-bfe4-e5458df85463- Value : 7525.503
# 10 - Event Created : b4b22c92-439e-4187-87ee-dbd806e0ace7- Value : 3757.74
# 11 - Event Created : 162543e8-7110-43e0-8705-6e31651001f6- Value : 337.836
```

در ادامه صحت اضافه شدن این داده های تولید شده به کافکا را بررسی می کنیم.

Search

Hide internal topics

Q

Keep search







Topics				Partitions	Replications		Consumer Groups
Name	Count	Size	Last Record	Total	Factor	In Sync	Consumer Groups
events_topic	≈ 34	7.958 KB	1 second ago	3	1	1	

## Topic: events\_topic

[Data](#)[Partitions](#)[Consumer Groups](#)[Configs](#)[ACLs](#)[Logs](#)

Sort: (Oldest) ▾ Partition: (All) ▾ Timestamp: ▾ Search: ▾ Offsets: ▾

≈ 7797

	Date	Partition	Offset	Headers	Schema	
 24c13e7d-2d55-4832-b4fb-ba0cb2bd6fb7	1 hour ago	0	0	0		
{ "event_id": "24c13e7d-2d55-4832-b4fb-ba0cb2bd6fb7", "event_datetime": "2021-10-15", "event_type": "buy", "product_id": "product04", "event_value": 6682.468, "user_name": "vasquezanne", "page": "blog/list" }						
52c8380c-d90c-45c2-b229-d4338abc0983	1 hour ago	0	1	0		
{ "event_id": "52c8380c-d90c-45c2-b229-d4338abc0983", "event_datetime": "2021-10-02", "event_type": "buy", "product_id": "product05", "event_value": 7099.164, "user_name": "tere satownsend", "page": "list" }						
98643c19-b1e3-43c8-ac63-84b21e6ae7a7	1 hour ago	0	2	0		
{ "event_id": "98643c19-b1e3-43c8-ac63-84b21e6ae7a7", "event_datetime": "2021-10-12", "event_type": "click", "product_id": "product09", "event_value": 8950.249, "user_name": "rebecca82", "page": "main" }						
581299c8-a657-4282-a577-8ee6f0e0624d	1 hour ago	0	3	0		
{ "event_id": "581299c8-a657-4282-a577-8ee6f0e0624d", "event_datetime": "2021-10-10", "event_type": "idle_5", "product_id": "product03", "event_value": 5632.934, "user_name": "omcneil", "page": "app/tag" }						
b6945143-8b34-4b2f-bcc9-ff3125e23bcf	1 hour ago	0	4	0		
{ "event_id": "b6945143-8b34-4b2f-bcc9-ff3125e23bcf", "event_datetime": "2021-12-04", "event_type": "idle_5", "product_id": "product04", "event_value": 5977.48, "user_name": "ca" }						

حال در گام بعدی، یک اینجسشن برای دریافت داده های موجود در تاپیک events\_topic کافکا ایجاد می کنیم. این کار را با لود کردن دیتای کافکا در درون دروید انجام می دهیم. بدین صورت داده های دروید با کافکا به صورت لحظه ای سینک می شود.

submit  
spec

Druid ingests raw data and converts it into a custom, **indexed format** that is optimized for analytic queries.

To get started, please specify what data you want to ingest.

[Learn more](#)

Bootstrap servers

localhost::9092

Topic

events\_topic

Consumer properties

```
{  
  "bootstrap.servers": "localhost::9092"  
}
```

Where should the data be sampled from?

Start of stream

Apply

Next: Parse data →

پس از apply کردن، مشاهده می کنید که نمونه ای از داده های اضافه شده نمایش داده می شود.

```
("event_id": "581299c0-a657-4282-a577-8ee6f0e0624d", "event_datetime": "2021-10-18", "event_type": "idle_5", "product_id": "product03", "event_value": 5632.934, "user_name": "omcneil",  
("event_id": "b6945143-8b34-4b2f-bcc9-ff3125e23bfc", "event_datetime": "2021-12-04", "event_type": "idle_5", "product_id": "product04", "event_value": 5977.48, "user_name": "cameron5",  
("event_id": "6dbfad2e-e96e-4431-af06-0f9dbb7aac0a", "event_datetime": "2021-11-15", "event_type": "click", "product_id": "product01", "event_value": 9118.787, "user_name": "mespinoz",  
("event_id": "a672e67e-f7ad-43b6-988b-4c0814e4ea1c", "event_datetime": "2021-10-21", "event_type": "buy", "product_id": "product01", "event_value": 2930.675, "user_name": "smaxwell",  
("event_id": "327d29de-ed52-41f9-b3fd-60a26a22c24d", "event_datetime": "2021-11-26", "event_type": "hover", "product_id": "product07", "event_value": 8549.992, "user_name": "scottphi",  
("event_id": "f309b5ea-8030-432b-af8b-4537c61ee05a", "event_datetime": "2021-11-03", "event_type": "buy", "product_id": "product09", "event_value": 4154.051, "user_name": "kimberly5",  
("event_id": "f8726bfd-6fcd-4ab7-b228-caabf2471fe3", "event_datetime": "2021-11-22", "event_type": "hover", "product_id": "product03", "event_value": 6597.824, "user_name": "twashingt",  
("event_id": "2841c7a5-824f-48f8-b223-17080f469040", "event_datetime": "2021-11-12", "event_type": "buy", "product_id": "product10", "event_value": 5123.844, "user_name": "bharrell", " ",  
("event_id": "d4cd049a-7a45-4d0b-bc9a-0d6a53c909fd", "event_datetime": "2021-12-05", "event_type": "click", "product_id": "product08", "event_value": 7516.406, "user_name": "vanessa6",  
("event_id": "53520779-db85-4a5c-9c51-572576036c7e", "event_datetime": "2021-10-21", "event_type": "buy", "product_id": "product01", "event_value": 244.563, "user_name": "lewisjulle",  
("event_id": "c5689efb-7e16-42d9-97a2-f28cb1e3e98c", "event_datetime": "2021-10-04", "event_type": "click", "product_id": "product02", "event_value": 7284.741, "user_name": "amanda12",  
("event_id": "5c0fc5d4-fc9d-43db-a4a5-9101997a3a36", "event_datetime": "2021-10-15", "event_type": "hover", "product_id": "product07", "event_value": 946.446, "user_name": "georgejor",  
("event_id": "93e26311-1184-41b2-9b9b-c190b6a6a06f", "event_datetime": "2021-11-17", "event_type": "idle_5", "product_id": "product02", "event_value": 4018.744, "user_name": "ttaylor",  
("event_id": "d6826841-ed66-4c95-b2c5-9134a879502a", "event_datetime": "2021-11-20", "event_type": "click", "product_id": "product06", "event_value": 2773.757, "user_name": "victor63",  
("event_id": "6e12e8f6-e809-4e20-b65a-2b935eff775a0", "event_datetime": "2021-10-18", "event_type": "buy", "product_id": "product02", "event_value": 4630.273, "user_name": "pughrober",  
("event_id": "543e3e27-ce4e-4fe2-89e6-0ab71373c648", "event_datetime": "2021-12-04", "event_type": "click", "product_id": "product03", "event_value": 213.264, "user_name": "reyesgary",  
("event_id": "c787a221-f403-4a32-bb97-1f34db519590", "event_datetime": "2021-11-11", "event_type": "buy", "product_id": "product07", "event_value": 8460.167, "user_name": "ssmith", " ",  
("event_id": "f5a85e8c-6f2f-4080-afb8-1d262e2f0fcf", "event_datetime": "2021-10-18", "event_type": "hover", "product_id": "product02", "event_value": 1265.739, "user_name": "walkerada",  
("event_id": "bbcb7f74-7c04-4c29-a8bb-f37ef676e201", "event_datetime": "2021-12-01", "event_type": "idle_5", "product_id": "product01", "event_value": 983.64, "user_name": "vincentcoff",  
("event_id": "297d7f9a-37dc-4a08-814b-0f87cb8689a3", "event_datetime": "2021-11-03", "event_type": "buy", "product_id": "product01", "event_value": 5435.019, "user_name": "tfletcher",  
("event_id": "9dc1d0c2-9968-487e-b8d0-7cfc54cb589", "event_datetime": "2021-10-30", "event_type": "buy", "product_id": "product05", "event_value": 9021.054, "user_name": "debbiesau",  
("event_id": "c9286a39-7846-492b-9478-5a6bd3951772", "event_datetime": "2021-11-21", "event_type": "click", "product_id": "product04", "event_value": 4051.954, "user_name": "hblackw",  
("event_id": "8d0aadd4-020f-49a3-bb56-c412c9390d54", "event_datetime": "2021-12-11", "event_type": "idle_5", "product_id": "product04", "event_value": 5589.682, "user_name": "samant",  
("event_id": "6c7d9600-cc05-4268-ba2a-7f3a6e20edde", "event_datetime": "2021-11-26", "event_type": "idle_5", "product_id": "product04", "event_value": 9519.028, "user_name": "stephan",  
("event_id": "a7e156b1-13e6-4a22-b760-ca1216b20c75", "event_datetime": "2021-10-13", "event_type": "click", "product_id": "product02", "event_value": 376.011, "user_name": "maria95",  
("event_id": "939ad088-c789-4303-9f3c-662880bb51f9", "event_datetime": "2021-10-02", "event_type": "idle_5", "product_id": "product08", "event_value": 7593.048, "user_name": "shepher",  
("event_id": "3fe3c633-e25e-46bd-9195-766a61c2e3e7", "event_datetime": "2021-10-04", "event_type": "buy", "product_id": "product10", "event_value": 9307.021, "user_name": "knappdav",  
("event_id": "f1102806-9a6e-457c-9af4-0440eb85bed1", "event_datetime": "2021-11-30", "event_type": "click", "product_id": "product06", "event_value": 4622.677, "user_name": "zhaynes",  
("event_id": "62852efc-67ea-45c8-8bac-a817f2a113e2", "event_datetime": "2021-11-07", "event_type": "hover", "product_id": "product09", "event_value": 8240.445, "user_name": "douglass",  
("event_id": "aafe37b0-63b7-4fec-be3a-a6d005928384", "event_datetime": "2021-11-11", "event_type": "idle_5", "product_id": "product02", "event_value": 7332.038, "user_name": "claudias",  
("event_id": "935f527b-8096-45d7-8a60-85154c4a247", "event_datetime": "2021-10-12", "event_type": "idle_5", "product_id": "product08", "event_value": 1641.423, "user_name": "joycebr",  
("event_id": "4b4a3d45-98c4-4147-8c7a-bba089fba309", "event_datetime": "2021-12-10", "event_type": "hover", "product_id": "product10", "event_value": 7929.306, "user_name": "roseach",
```

To get started, please specify what data you want to ingest.

[Learn more](#)

Bootstrap servers

localhost:9092

Topic

events\_topic

Consumer properties

```
{  
  "bootstrap.servers": "localhost:9092"  
}
```

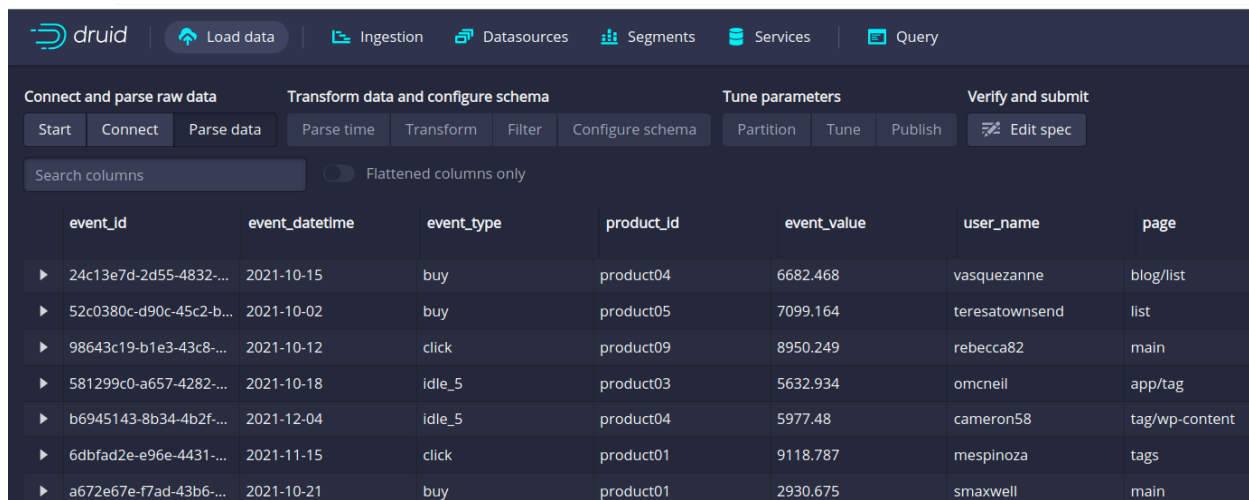
Where should the data be sampled from?

Start of stream

Apply

Next: Parse data →

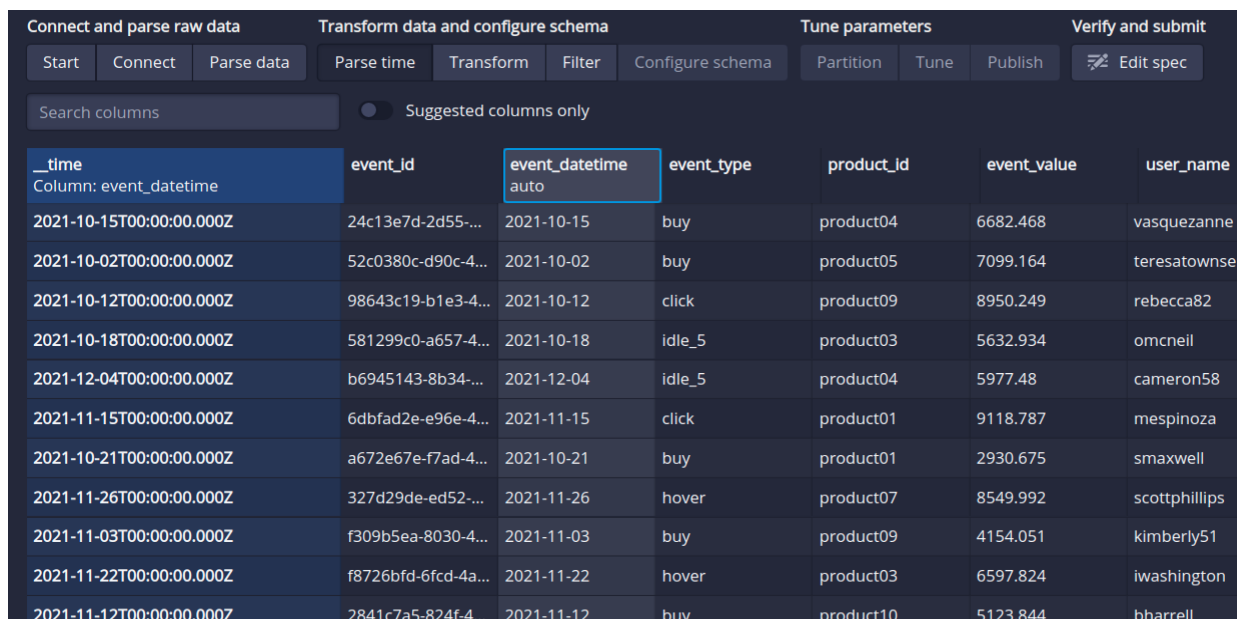
مشاهده می کنید که دروید به صورت اتوماتیک، داده ها را parse می کند.



The screenshot shows the Druid web console interface. The top navigation bar includes links for Load data, Ingestion, Datasources, Segments, Services, and Query. Below this, there are four main tabs: 'Connect and parse raw data', 'Transform data and configure schema', 'Tune parameters', and 'Verify and submit'. The 'Connect and parse raw data' tab is active, showing a 'Start' button and a 'Parse data' button. A search bar for columns is present, and a toggle for 'Flattened columns only' is visible. The main table displays a list of events with columns: event\_id, event\_datetime, event\_type, product\_id, event\_value, user\_name, and page. The data rows show various events such as 'buy', 'click', and 'idle\_5' with corresponding timestamps and user information.

event_id	event_datetime	event_type	product_id	event_value	user_name	page
24c13e7d-2d55-4832-...	2021-10-15	buy	product04	6682.468	vasquezanne	blog/list
52c0380c-d90c-45c2-b...	2021-10-02	buy	product05	7099.164	teresatownsend	list
98643c19-b1e3-43c8-...	2021-10-12	click	product09	8950.249	rebecca82	main
581299c0-a657-4282-...	2021-10-18	idle_5	product03	5632.934	omcneil	app/tag
b6945143-8b34-4b2f-...	2021-12-04	idle_5	product04	5977.48	cameron58	tag/wp-content
6dbfad2e-e96e-4431-...	2021-11-15	click	product01	9118.787	mespinoza	tags
a672e67e-f7ad-43b6-...	2021-10-21	buy	product01	2930.675	smaxwell	main

در قسمت parse time مشاهده می کنید که به صورت اتوماتیک، event\_datetime را به عنوان timestamp تشخیص داده است.



This screenshot shows the same Druid interface but with the 'event\_datetime' column highlighted in the 'Suggested columns only' section. The main table now includes an additional column, '\_time', which contains ISO 8601 timestamps. The 'event\_datetime' column is now labeled 'auto'. The data rows show the same events as before, but with the '\_time' column added, providing a more precise timestamp for each event.

_time Column: event_datetime	event_id	event_datetime auto	event_type	product_id	event_value	user_name
2021-10-15T00:00:00.000Z	24c13e7d-2d55-...	2021-10-15	buy	product04	6682.468	vasquezanne
2021-10-02T00:00:00.000Z	52c0380c-d90c-4...	2021-10-02	buy	product05	7099.164	teresatownse
2021-10-12T00:00:00.000Z	98643c19-b1e3-4...	2021-10-12	click	product09	8950.249	rebecca82
2021-10-18T00:00:00.000Z	581299c0-a657-4...	2021-10-18	idle_5	product03	5632.934	omcneil
2021-12-04T00:00:00.000Z	b6945143-8b34-...	2021-12-04	idle_5	product04	5977.48	cameron58
2021-11-15T00:00:00.000Z	6dbfad2e-e96e-4...	2021-11-15	click	product01	9118.787	mespinoza
2021-10-21T00:00:00.000Z	a672e67e-f7ad-4...	2021-10-21	buy	product01	2930.675	smaxwell
2021-11-26T00:00:00.000Z	327d29de-ed52-...	2021-11-26	hover	product07	8549.992	scottphillips
2021-11-03T00:00:00.000Z	f309b5ea-8030-4...	2021-11-03	buy	product09	4154.051	kimberly51
2021-11-22T00:00:00.000Z	f8726bfd-6fcd-4a...	2021-11-22	hover	product03	6597.824	iwashington
2021-11-12T00:00:00.000Z	2841c7a5-824f-4...	2021-11-12	buy	product10	5123.844	bharrell

برای داده ها، transform و filter ای مدنظر قرار نمی دهیم.



Connect and parse raw data

Transform data and configure schema

Tune parameters

Verify and submit

Start

Connect

Parse data

Parse time

Transform


Filter

Configure schema

Partition

Tune

Publish

 Edit spec

Search columns

☐ Transformed columns only

_time	event_id	event_type	product_id	event_value	user_name	page
2021-10-15T00:00:00.0...	24c13e7d-2d55-4832-b...	buy	product04	6682.468	vasquezanne	blog/list
2021-10-02T00:00:00.0...	52c0380c-d90c-45c2-b...	buy	product05	7099.164	teresatownsend	list
2021-10-12T00:00:00.0...	98643c19-b1e3-43c8-a...	click	product09	8950.249	rebecca82	main
2021-10-18T00:00:00.0...	581299c0-a657-4282-a...	idle_5	product03	5632.934	omcneil	app/tag
2021-12-04T00:00:00.0...	b6945143-8b34-4b2f-b...	idle_5	product04	5977.48	cameron58	tag/wp-content
2021-11-15T00:00:00.0...	6dbfad2e-e96e-4431-a...	click	product01	9118.787	mespinoza	tags
2021-10-21T00:00:00.0...	a672e67e-f7ad-43b6-9...	buy	product01	2930.675	smaxwell	main

Connect and parse raw data

Transform data and configure schema

Tune parameters

Verify and submit

Start

Connect

Parse data

Parse time

Transform


Filter

Configure schema

Partition

Tune

Publish

 Edit spec

Search columns

_time	event_id	event_type	product_id	event_value	user_name	page
2021-10-15T00:00:00.0...	24c13e7d-2d55-4832-b...	buy	product04	6682.468	vasquezanne	blog/list
2021-10-02T00:00:00.0...	52c0380c-d90c-45c2-b...	buy	product05	7099.164	teresatownsend	list
2021-10-12T00:00:00.0...	98643c19-b1e3-43c8-a...	click	product09	8950.249	rebecca82	main

برای input tuning هم فیلد use earliest offset را true کرده و task duration را برابر یک ساعت قرار می دهیم. بقیه تنظیمات دیفالت را تغییر نمی دهیم.

### Input tuning

Use earliest offset

False True

Task duration

PT1H

Task count

1

Replicas

1

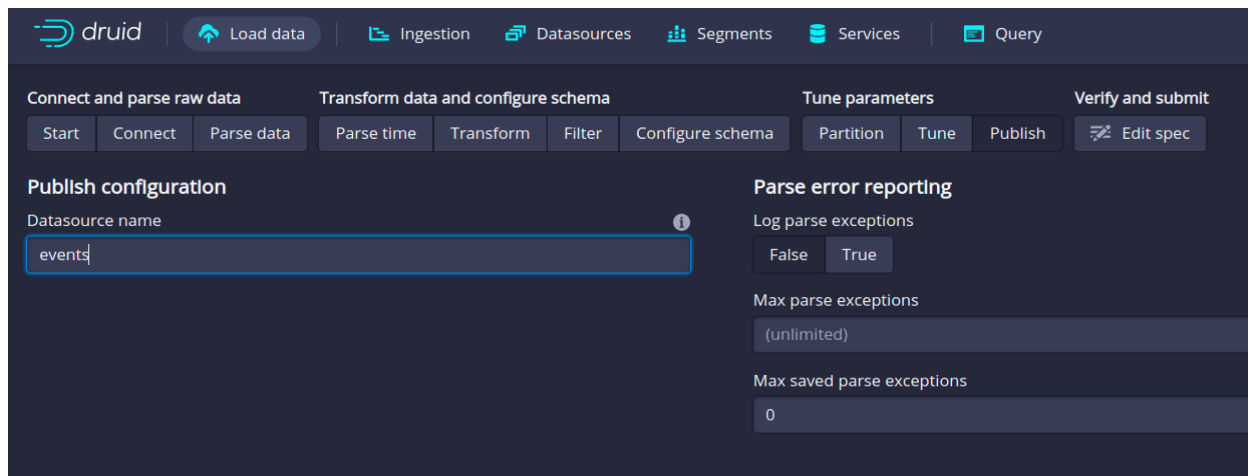
Completion timeout

PT30M

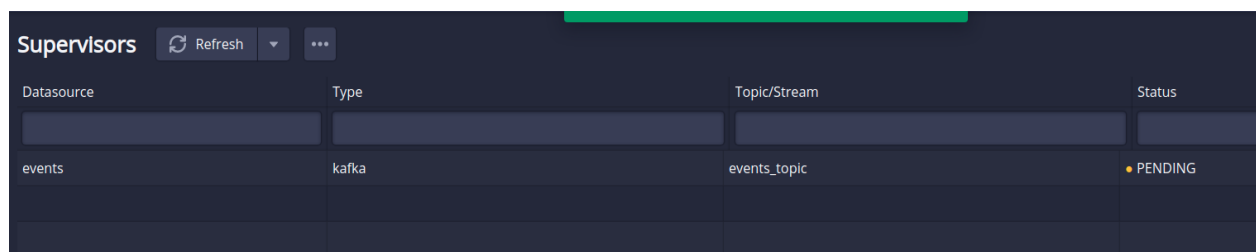
Poll timeout

100

اسم دیتاسورس را هم برابر events قرار می دهیم.




The image shows the 'Publish configuration' section of the Druid web console. It includes tabs for 'Connect and parse raw data', 'Transform data and configure schema', 'Tune parameters', and 'Verify and submit'. The 'Publish configuration' tab is active, showing a 'Datasource name' field with the value 'events'. To the right, there is a 'Parse error reporting' section with options for 'Log parse exceptions' (False/True), 'Max parse exceptions' (unlimited), and 'Max saved parse exceptions' (0).



The image shows the 'Supervisors' table in the Druid web console. It has columns for 'Datasource', 'Type', 'Topic/Stream', and 'Status'. The 'events' datasource is listed with type 'kafka' and topic 'events\_topic', with a status of 'PENDING'.

Datasource	Type	Topic/Stream	Status
events	kafka	events_topic	PENDING



The image shows the 'Datasources' table in the Druid web console. It has columns for 'Datasource name', 'Availability', 'Availability detail', 'Total data size', 'Segment size (rows)', 'Segment granularity', 'Total rows', 'Avg. row size (bytes)', 'Replicated size', and 'Compaction'. The 'events' datasource is listed with a status of 'Fully available (3 segments)' and a total data size of '0.00 B'.

Datasource name	Availability	Availability detail	Total data size	Segment size (rows)	Segment granularity	Total rows	Avg. row size (bytes)	Replicated size	Compaction
events	Fully available (3 segments)	No segment...	0.00 B	0 0 0	Month	8,560	0	0.00 B	Not enabled

این اینجسشن تجمیع نشده بود.

حال دو اینجسشن تجمیع شده، با rollup یک ساعته و یک روزه ایجاد می کنیم. Topic همین topic فعلی است. سه متریک count و min و max روی فیلد event\_value تعریف می کنیم و متریک sum هم به صورت دیفالت وجود داشت. ستون event\_id را به این علت حذف کردم که رول آپ با وجود این فیلد، به درستی کار نمی کرد و همه مقادیر این فیلد به صورت یکتا بودند. برای rollup یک ساعته، query

granularity را برابر یک ساعت و برای یک روزه، برابر یک روز قرار دادم. اگر تعداد داده ها بیشتر بود، تاثیر rollup بهتر نمایان می شد؛ ولی الان فیلد count فقط مقدار یک را شامل شده است.

## رول آپ یک ساعته:

	main	1	8950	8950
	app/tag	1	5632	5632
8	tag/wp-cont...	1	5977	5977
a	tags	1	9118	9118
	main	1	2930	2930
ps	category/po...	1	8549	8549
1	app	1	4154	4154
on	blog/list	1	6597	6597
	main/explore	1	5123	5123
0	category/tags	1	7516	7516
	main/tags/p...	1	244	244
2	posts/explore	1	7284	7284
dan	tag/blog	1	946	946

[Learn more](#)

☒ Explicitly specify dimension list
 ☒ Rollup

Query granularity
 

hour

Add dimension

Add metric

⬅

Search columns

__time	event_id	event_type	product_id	user_name	page	count	sum_event_value	min_event_value
long (time column)	string	string	string	string	string	count	longSum	longMin
2021-10-15T00:00:00.000Z	24c13e7d-2d...	buy	product04	vasquezanne	blog/list	1	6682	6682
2021-10-02T00:00:00.000Z	52c0380c-d9...	buy	product05	teresatowns...	list	1	7099	7099
2021-10-12T00:00:00.000Z	98643c19-b1...	click	product09	rebecca82	main	1	8950	8950
2021-10-18T00:00:00.000Z	581299c0-a6...	idle_5	product03	omcneil	app/tag	1	5632	5632
2021-12-04T00:00:00.000Z	b6945143-8...	idle_5	product04	cameron58	tag/wp-cont...	1	5977	5977
2021-11-15T00:00:00.000Z	6dbfad2e-e9...	click	product01	mespinoza	tags	1	9118	9118
2021-10-21T00:00:00.000Z	a672e67e-f7...	buy	product01	smaxwell	main	1	2930	2930
2021-11-26T00:00:00.000Z	327d29de-e...	hover	product07	scottphillips	category/po...	1	8549	8549
2021-11-03T00:00:00.000Z	f309b5ea-80...	buy	product09	kimberly51	app	1	4154	4154
2021-11-22T00:00:00.000Z	f8726bfd-6fc...	hover	product03	lwashington	blog/list	1	6597	6597
2021-11-12T00:00:00.000Z	2841c7a5-82...	buy	product10	bharrell	main/explore	1	5123	5123
2021-12-05T00:00:00.000Z	d4cd049a-7a...	click	product08	vanessa60	category/tags	1	7516	7516
2021-10-21T00:00:00.000Z	53520779-d...	buy	product01	lewisjulie	main/tags/p...	1	244	244
2021-10-04T00:00:00.000Z	c5689efb-7e...	click	product02	amanda12	posts/explore	1	7284	7284
2021-10-15T00:00:00.000Z	5c0fc5d4-fc9...	hover	product07	georgejordan	tag/blog	1	946	946
2021-11-17T00:00:00.000Z	93e26311-11...	idle_5	product02	ttaylor	tags/explore	1	4018	4018
2021-11-20T00:00:00.000Z	d6826841-e...	click	product06	victor63	search	1	2773	2773

Each column in Druid must have an assigned type (string, long, float, double, complex, etc).

Default primitive types have been automatically assigned to your columns. If you want to change the type, click on the column header.

[Learn more](#)

☒ Explicitly specify dimension list
 ☒ Rollup

Query granularity
 

hour

Add dimension

Add metric

⬅

Next: Partition ➔

## رول آپ یک روزه:

event\_value

min

max\_event\_value

longMax

6682

7099

8950

5632

5977

9118

2930

8549

4154

6597

5123

7516

244

7284

946

4018

Each column in Druid must have an assigned type (string, long, float, double, complex, etc).

Default primitive types have been automatically assigned to your columns. If you want to change the type, click on the column header.

[Learn more](#)

Explicitly specify dimension list

Rollup

Query granularity

day

Add dimension

Add metric

Connect and parse raw data

Transform data and configure schema

Tune parameters

Verify and submit

Start

Connect

Parse data

Parse time

Transform

Filter

Configure schema

Partition

Tune

Publish

Edit spec

Search columns

_time	event_id	event_type	product_id	user_name	page	count	sum_event_value	min_event_value	max_event_value
long (time column)	string	string	string	string	string	count	doubleSum	longMin	longMax
2021-10-15T00:00:00.000Z	24c13e7d-2d...	buy	product04	vasquezanne	blog/list	1	6682.468	6682	6682
2021-10-02T00:00:00.000Z	52c0380c-d9...	buy	product05	teresatown...	list	1	7099.164	7099	7099
2021-10-12T00:00:00.000Z	98643c19-b1...	click	product09	rebecca82	main	1	8950.249	8950	8950
2021-10-18T00:00:00.000Z	581299c0-a6...	idle_5	product03	omcneil	app/tag	1	5632.934	5632	5632
2021-12-04T00:00:00.000Z	b6945143-8...	idle_5	product04	cameron58	tag/wp-cont...	1	5977.48	5977	5977
2021-11-15T00:00:00.000Z	6dbfad2e-e9...	click	product01	mespinoza	tags	1	9118.787	9118	9118
2021-10-21T00:00:00.000Z	a672e67e-f7...	buy	product01	smaxwell	main	1	2930.675	2930	2930
2021-11-26T00:00:00.000Z	327d29de-e...	hover	product07	scottphillips	category/po...	1	8549.992	8549	8549
2021-11-03T00:00:00.000Z	f309b5ea-80...	buy	product09	kimberly51	app	1	4154.051	4154	4154
2021-11-22T00:00:00.000Z	f8726bfd-6f...	hover	product03	iwashington	blog/list	1	6597.824	6597	6597
2021-11-12T00:00:00.000Z	2841c7a5-82...	buy	product10	bharrell	main/explore	1	5123.844	5123	5123
2021-12-05T00:00:00.000Z	d4cd049a-7a...	click	product08	vanessa60	category/tags	1	7516.406	7516	7516
2021-10-21T00:00:00.000Z	53520779-d...	buy	product01	lewisjulia	main/tags/p...	1	244.563	244	244
2021-10-04T00:00:00.000Z	c5689efb-7e...	click	product02	amanda12	posts/explore	1	7284.741	7284	7284
2021-10-15T00:00:00.000Z	5c0fc5d4-fc9...	hover	product07	georgejordan	tag/blog	1	946.446	946	946
2021-11-17T00:00:00.000Z	93e26311-11...	idle_5	product02	ttaylor	tags/explore	1	4018.744	4018	4018

Each column in Druid must have an assigned type (string, long, float, double, complex, etc).

Default primitive types have been automatically assigned to your columns. If you want to change the type, click on the column header.

[Learn more](#)

Explicitly specify dimension list

Rollup

Query granularity

day

Add dimension

Add metric

در نهایت سه دیتاسورس events\_hour و events\_day و events ایجاد شد.

Supervisors		
	Refresh	...
Datasource	Type	Topic/Stream
events	kafka	events_topic
events_day	kafka	events_topic
events_hour	kafka	events_topic
Previous		Page 1 of 1 20 rows

پنج کوئری مختلف روی این سه دیتاسورس (شامل گروه بندی، بازه زمانی، تعیین ترتیب، شرط):

1. کوئری اول، نام کاربری و شناسه کالا را از جدول events به ازای همه کالاهایی که وضعیت خریداری شده دارند را پیدا می کند و نام کاربری و شناسه کالا را نمایش می دهد.

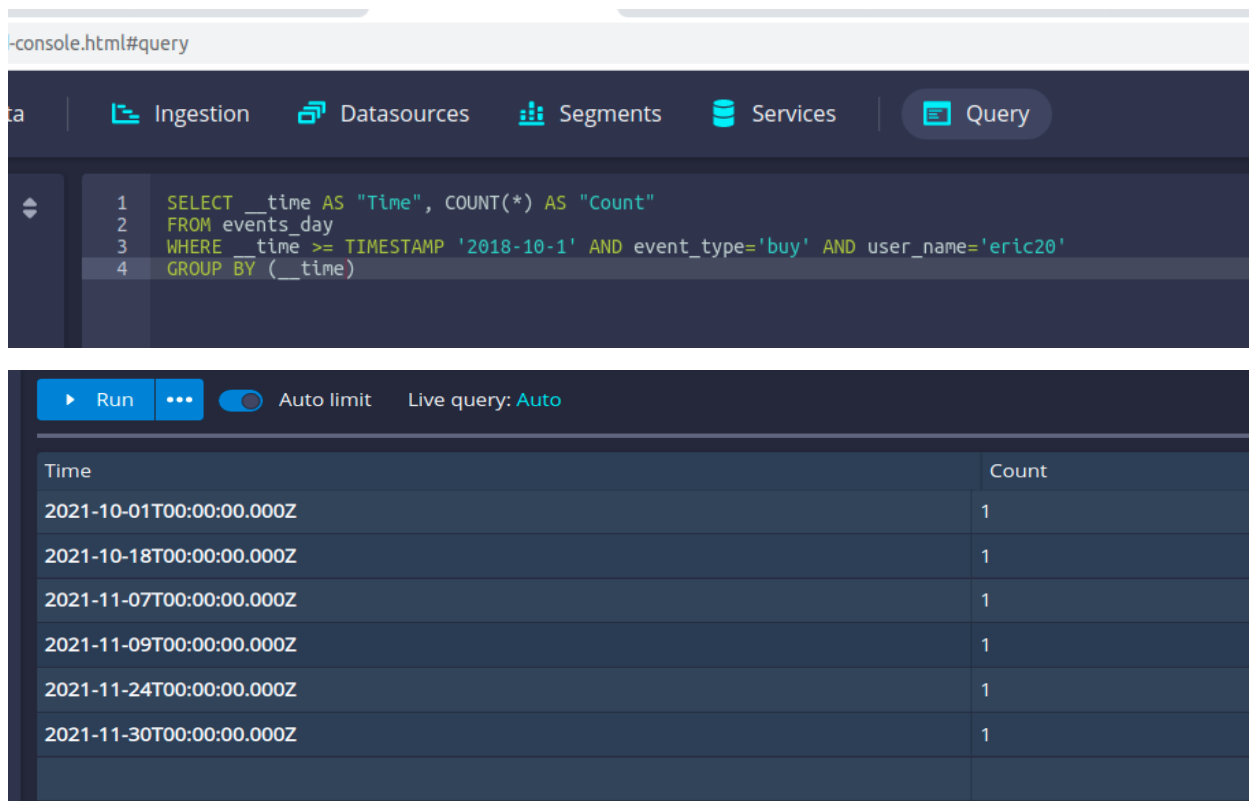
```

1 SELECT user_name, product_id
2 FROM events
3 WHERE event_type='buy'

```

Run ... Auto limit Live query: Auto	
user_name	product_id
peteredwards	product07
williamstomas	product06
bryanthony	product10
edwardmclaughlin	product10
bridget86	product04
eric20	product02
dlandry	product05
rebecca82	product02
jbaird	product10
ssantos	product10

2. کوئری دوم، از جدول events\_day برای کاربر مشخصی و بعد از تاریخ مشخصی، کالاهای خریداری شده اش را پیدا می کند و بر اساس زمان آن ها را گروه بندی می کند و زمان و تعداد آن گروه ها را نمایش می دهد.



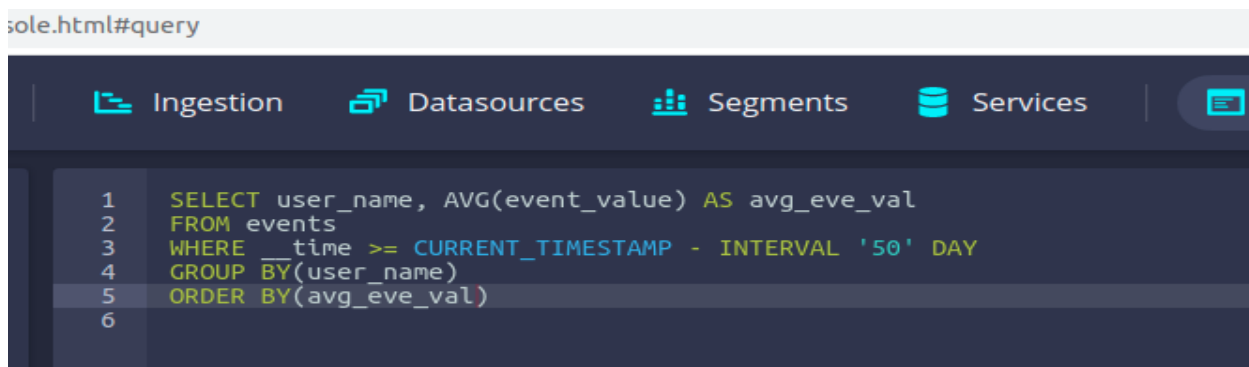
The screenshot shows a query editor interface with a dark theme. The top navigation bar includes 'Ingestion', 'Datasources', 'Segments', 'Services', and 'Query'. The query editor contains the following SQL code:

```
1 SELECT __time AS "Time", COUNT(*) AS "Count"
2 FROM events_day
3 WHERE __time >= TIMESTAMP '2018-10-1' AND event_type='buy' AND user_name='eric20'
4 GROUP BY (__time)
```

Below the query editor, there is a 'Run' button and a toggle for 'Auto limit'. The 'Live query' status is set to 'Auto'. The results are displayed in a table with two columns: 'Time' and 'Count'.

Time	Count
2021-10-01T00:00:00.000Z	1
2021-10-18T00:00:00.000Z	1
2021-11-07T00:00:00.000Z	1
2021-11-09T00:00:00.000Z	1
2021-11-24T00:00:00.000Z	1
2021-11-30T00:00:00.000Z	1

3. کوئری سوم، از جدول events برای هر کاربر در 50 روز اخیر، میانگین event\_value هایش را پیدا می کند و بر اساس این میانگین، سورت می کند. و نام کاربری و این مقدار میانگین را نمایش می دهد.



The screenshot shows a query editor interface with a dark theme. The top navigation bar includes 'Ingestion', 'Datasources', 'Segments', 'Services', and 'Query'. The query editor contains the following SQL code:

```
1 SELECT user_name, AVG(event_value) AS avg_eve_val
2 FROM events
3 WHERE __time >= CURRENT_TIMESTAMP - INTERVAL '50' DAY
4 GROUP BY(user_name)
5 ORDER BY(avg_eve_val)
6
```

Run	Auto limit	Live query: Auto
user_name	avg_eve_val	
jcarter	8.539	
timothy89	32.489	
sarahgraham	37.801	
powellisa	58.997	
amyking	73.487	
hsmith	101.166	
williamdavis	105.131	
briangreen	221.173	
adkinsmaurice	239.716	
rfernandez	265.889	
riveralisa	296.604	

4. کوئری چهارم، از جدول events\_hour برای کالای مشخصی، نام کاربر و زمانی که روی آن کالا کلیک کرده است را پیدا می کند و نمایش می دهد.

```
sole.html#query
```

Ingestion	Datasources	Segments	Services	Query
-----------	-------------	----------	----------	-------

```

1  SELECT user_name, __time
2  FROM events_hour
3  WHERE product_id='product07' and event_type='click'
```

Run ... Auto limit Live query: Auto	
user_name	_time
samanthajimenez	2021-10-01T00:00:00.000Z
mary56	2021-10-01T00:00:00.000Z
salinaslori	2021-10-01T00:00:00.000Z
jeremygordon	2021-10-01T00:00:00.000Z
aprilsanchez	2021-10-02T00:00:00.000Z
bharrell	2021-10-02T00:00:00.000Z
laurenkelly	2021-10-02T00:00:00.000Z
raymondparker	2021-10-02T00:00:00.000Z
andrewevans	2021-10-02T00:00:00.000Z
yhicks	2021-10-02T00:00:00.000Z
uparsons	2021-10-02T00:00:00.000Z
Previous Page 1 20 rows	

5. کوئری پنجم، از جدول events\_hour تعداد تکرار هر page را به صورت نزولی پیدا می کند و نام page و مقدار تکرارش را به صورت نزولی و سورت شده بر اساس تعداد تکرار نمایش می دهد.

```
unified-console.html#query
```

data | 
Ingestion
Datasources
Segments
Service

```

1 SELECT page, COUNT (*) AS "Count"
2 FROM events_hour
3 GROUP BY (page)
4 ORDER BY "Count" DESC

```



▶ Run

⋮

☒ Auto limit

Live query: [Auto](#)

page	Count
posts	480
app	479
list	450
blog	445
tag	443
tags	439
wp-content	439
main	432
search	426
explore	422
categories	409

Previous

Page 1

20 rows