

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



آزمایشگاه پایگاه داده

دستور کار شماره ۹

آتیه آرمین

۸۱۰۱۹۷۶۴۸

۱۴۰۰ مهرماه

گزارش دستورکار انجام شده

ابتدا دقیقاً مانند تمام مراحلی که در فیلم گفته شده بود، دروید و کافکا را راه اندازی کرده و بالا آوردم. سپس کد پایتون را اجرا کرده و داده هایی که هر لحظه به صورت تصادفی اجرا می شدند را به صورت زیر واسط کاربری کافکا در آدرس localhost:8000 دیدم.

حال در ادامه طبق ویدئو آموزشی، یک اینجشن برای دریافت داده های موجود در تاپیک events_topic کافکا ایجاد کردم. مراحل انجام شده و دیتابسورس ایجاد شده در ادامه قابل مشاهده است.

The screenshot shows the Druid UI's configuration interface for transforming data. It includes a table of event data and various configuration tabs like 'Parse time', 'Transform', 'Filter', and 'Configure schema'.

The screenshot shows the Druid UI's query interface, displaying a SELECT query and its results. The results table includes columns: _time, event_id, event_type, event_value, page, product_id, and user_name. The results show multiple rows of event data.

حال در ادامه یک roll up یک ساعته و یک roll up یک روزه با توجه به دینتای تولید شده ایجاد کردم.

The screenshot shows the Druid UI's supervisor and task management interface. The 'Supervisors' section lists data sources and their types. The 'Tasks' section lists tasks, their group IDs, types, data sources, locations, creation times, and statuses.

The screenshot shows the Druid Data Sources interface. The top navigation bar includes 'druid', 'Load data', 'Ingestion', 'Datasources' (selected), 'Segments', 'Services', and 'Query'. Below the navigation is a search bar and filter buttons for 'Show unused' and 'Show segment timeline'. A table lists data sources with columns: Datasource name, Availability, Availability detail, Total data size, Segment size (rows) minimum / average / maximum, Segment granularity, Total rows, Avg. row size (bytes), Replicated size, Compaction, and % Compacted bytes / segme. One entry, 'events_one_hour', is selected and highlighted in green, indicating it is 'Fully available (75 segments)'. The table has 14 columns and 1 row. At the bottom, there are navigation buttons for 'Previous', 'Page 1 of 1', '50 rows', and 'Next'.

The screenshot shows the Druid Segments interface. The top navigation bar includes 'druid', 'Load data', 'Ingestion', 'Datasources', 'Segments' (selected), 'Services', and 'Query'. Below the navigation is a search bar and filter buttons for 'Group by', 'None', 'Interval', and 'Show segment timeline'. A table lists segments with columns: Segment ID, Datasource, Start, End, Version, Time span, Partitioning, Partition..., Size, and Num rows. The 'events_one_hour' datasource is selected. The table has 16 columns and 75 rows. At the bottom, there are navigation buttons for 'Previous', 'Page 1', '25 rows', and 'Next'.

البته در اینجا پس اضافه کردن roll up یک روزه، چون تعداد اسلات ها که ۲ عدد بودند پر شده بود، مجبور شدم تسك اولی را متوقف کنم تا تسك roll up مذکور انجام شود.

The screenshot shows the Druid Supervisors and Tasks interfaces. The top navigation bar includes 'druid', 'Load data', 'Ingestion', 'Datasources', 'Supervisors' (selected), and 'Tasks'. A green message box says 'Supervisor submitted successfully. Going to task view...'. The 'Supervisors' section shows three entries: 'events' (kafka, events_topic, RUNNING), 'events_one_day' (kafka, events_topic, PENDING), and 'events_one_hour' (kafka, events_topic, RUNNING). The 'Tasks' section shows two tasks: 'index.kafka_events_one_hour_a07f4e4918c1a_afmpjiae' (index.kafka, events_one_hour, RUNNING) and 'index.kafka_events_99d86c52a8deda_eapglaj' (index.kafka, events, RUNNING). Both tasks were submitted at 2022-01-25T21:02:08.280Z. At the bottom, there are navigation buttons for 'Previous', 'Page 1 of 1', '20 rows', and 'Next'.

حال در ادامه به پیاده سازی ۵ کوئری مبپردازیم که به ترتیب به صورت زیر هستند:

- تمام event های یک روزه‌ای که در آخرین سال هستند.
- تعداد event هایی که در هر یک روز اتفاق افتاده است.
- تعداد event هایی که به ازای هر product و به ازای هر page اتفاق افتاده است.
- تعداد باری که هر product در کل event ها تکرار شده است به صورت نزولی (بر اساس تعداد تکرار).
- تمام event هایی که event_value آنها بیش از ۹۰۰۰ است.

The screenshot shows the Druid UI interface with a sidebar containing a tree view of data sources and tables. The main area displays a query results table.

```

SELECT
  2   TIME_FLOOR(_time, 'PT1H') AS "Time",
  3   COUNT(*) AS "Count"
  4 FROM events_one_day
  5 GROUP BY 1
  6 ORDER BY 1 ASC
  
```

Time

| Time | Count |
|--------------------------|-------|
| 2021-10-01T00:00:00.000Z | 22 |
| 2021-10-02T00:00:00.000Z | 29 |
| 2021-10-03T00:00:00.000Z | 57 |
| 2021-10-04T00:00:00.000Z | 35 |
| 2021-10-05T00:00:00.000Z | 32 |
| 2021-10-06T00:00:00.000Z | 35 |
| 2021-10-07T00:00:00.000Z | 26 |
| 2021-10-08T00:00:00.000Z | 23 |
| 2021-10-09T00:00:00.000Z | 33 |

Previous Page 1 of 4 Next

75 results in 0.06s

The screenshot shows the Druid UI interface with a sidebar containing a tree view of data sources and tables. The main area displays a query results table.

```

SELECT
  2   product_id,
  3   page,
  4   COUNT(*) AS "Count"
  5 FROM events_one_hour
  6 GROUP BY 1, 2
  7 ORDER BY 3 DESC
  
```

| product_id | page | Count |
|------------|------------|-------|
| product10 | tag | 17 |
| product06 | tags | 15 |
| product09 | tag | 14 |
| product07 | category | 13 |
| product01 | tags | 12 |
| product02 | list | 12 |
| product03 | explore | 12 |
| product04 | list | 12 |
| product04 | wp-content | 12 |

Previous Page 1 of 4 Next

100+ results in 0.35s

The screenshot shows the Druid UI interface with a sidebar containing a tree view of data sources and tables. The main area displays a query results table.

```

SELECT
  2   product_id,
  3   COUNT(*) AS "Count"
  4 FROM events
  5 GROUP BY 1
  6 ORDER BY 2 DESC
  
```

| product_id | Count |
|------------|-------|
| product04 | 204 |
| product10 | 200 |
| product03 | 196 |
| product07 | 196 |
| product06 | 195 |
| product09 | 195 |
| product01 | 193 |
| product05 | 185 |
| product02 | 177 |

Previous Page 1 of 1 Next

10 results in 0.04s

The screenshot shows the Druid UI interface. On the left, there's a sidebar with a tree view of a database named 'druid'. Under 'druid', there are several tables: 'events', 'events_one_day', 'events_one_hour', and 'user_name'. The 'events' table is expanded, showing columns: '_time', 'event_id', 'event_type', 'event_value', 'page', 'product_id', and 'user_name'. A sample query is displayed in the main area:

```

1 SELECT *
2 FROM events
3 WHERE event_value > 9000
4
    
```

Below the query, there are buttons for 'Run' and 'Auto limit'. The status bar indicates 'Live query: Auto' and '100+ results in 0.04s'. The results table shows 10 rows of data. At the bottom, there are navigation buttons for 'Previous', 'Page 1', 'Next', and a dropdown for '20 rows'.

| _time | event_id | event_type | event_value | page | product_id | user_name |
|--------------------------|--------------------------|------------|-------------|------------------|------------|-----------------|
| 2021-10-01T00:00:00.000Z | 3c91796f91ca45c9-a169- | hover | 9196.017 | posts | product04 | carrolldaniel |
| 2021-10-02T00:00:00.000Z | 432830ae-460b-4fe3-9d2d | idle_5 | 9543.011 | search | product07 | morganorozco |
| 2021-10-02T00:00:00.000Z | 7e9e6869-3911-4b3d-a91c | hover | 9913.014 | search/search | product02 | othomas |
| 2021-10-03T00:00:00.000Z | 04c3b7c7-5674-4f3e-a9f7- | click | 9506.979 | tag | product07 | smithmichael |
| 2021-10-03T00:00:00.000Z | 6bc4c3f3-a7d7-4780-8d1a | idle_5 | 9361.522 | tag/app | product09 | wcole |
| 2021-10-03T00:00:00.000Z | a0694318-6660-4243-b261 | hover | 9904.749 | posts/search/app | product05 | mwalters |
| 2021-10-03T00:00:00.000Z | c17fea86-1fab-4e15-bd21- | hover | 9761.528 | posts | product03 | lawrencecandice |
| 2021-10-03T00:00:00.000Z | caa71995-5b53-4eb8-c81c8 | buy | 9663.835 | explore | product08 | thompsonlaura |
| 2021-10-03T00:00:00.000Z | d3cedda8-b645-4a8f-a959 | hover | 9757.63 | tags | product01 | smithmichael |