



**МІНІСТЕРСТВО ОСВІТИ, НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КІЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**

**Лабораторна робота №5
Аналіз даних**

Підготував:

студент 4 курсу

групи ФІ-84

Коломієць Андрій Юрійович

E-mail: andrew.kolomiets.work@gmail.com

Київ – 2021

Лабораторна робота №5

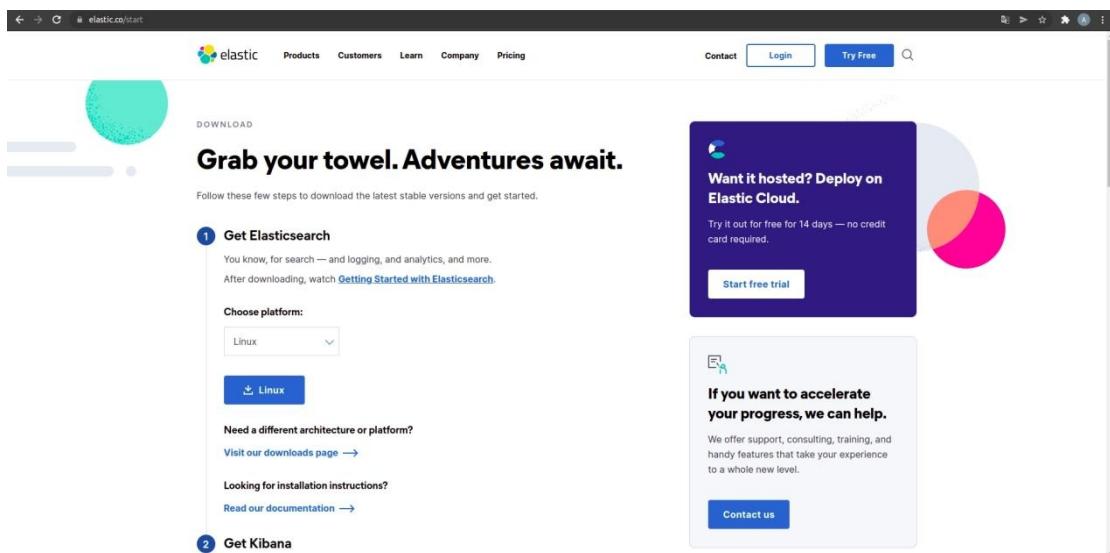
Аналіз даних

Завдання на самостійну роботу

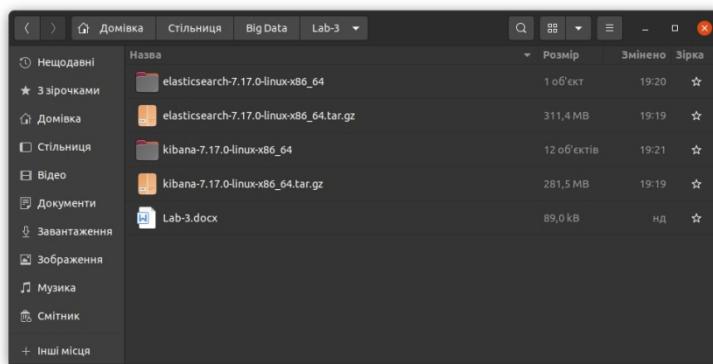
1. Забезпечити функціонування на комп'ютері системи **Kibana**.
2. Детально ознайомитись із можливостями візуалізації в системі **Kibana**.
3. Побудувати часові діаграми кількості публікацій із різних **RSS-джерел**.

Виконання завдання

Переходимо на сайт скачування продукту та завантажуємо **Elasticsearch** та **Kibana**.



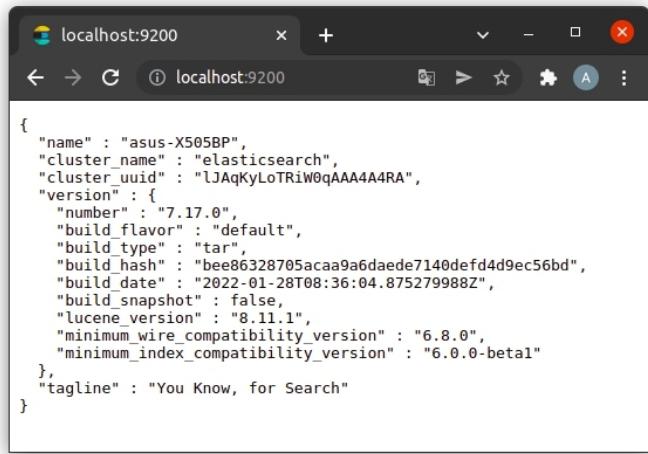
В результаті буде завантажено два архіви, впевнімося в цьому.



Запускаємо відповідно до інструкцій на сайті два вебінтерфейси.

Перевіряємо коректність запуску програми.

-використання браузера:

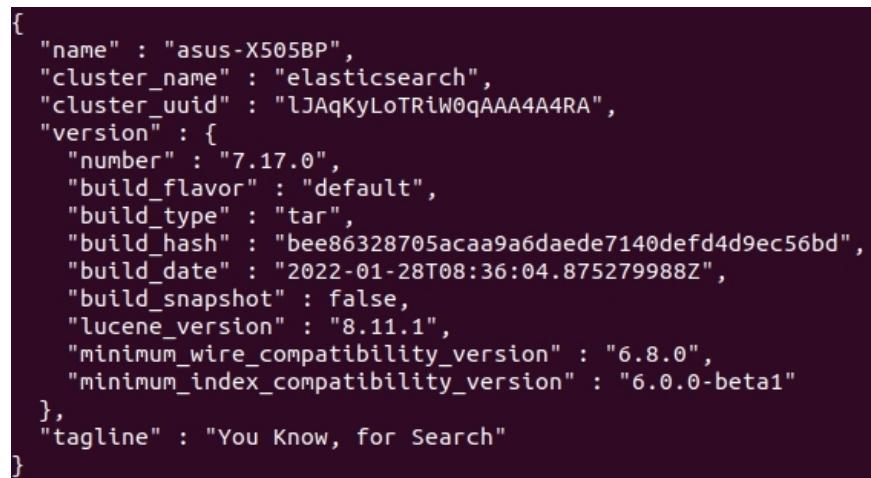


A screenshot of a web browser window titled "localhost:9200". The address bar also shows "localhost:9200". The page content is a JSON object representing the Elasticsearch version:

```
{  
  "name" : "asus-X505BP",  
  "cluster_name" : "elasticsearch",  
  "cluster_uuid" : "lJAqKyLoTRiW0qAAA4A4RA",  
  "version" : {  
    "number" : "7.17.0",  
    "build_flavor" : "default",  
    "build_type" : "tar",  
    "build_hash" : "bee86328705acaa9a6daede7140defd4d9ec56bd",  
    "build_date" : "2022-01-28T08:36:04.875279988Z",  
    "build_snapshot" : false,  
    "lucene_version" : "8.11.1",  
    "minimum_wire_compatibility_version" : "6.8.0",  
    "minimum_index_compatibility_version" : "6.0.0-beta1"  
  },  
  "tagline" : "You Know, for Search"  
}
```

-використання терміналу:

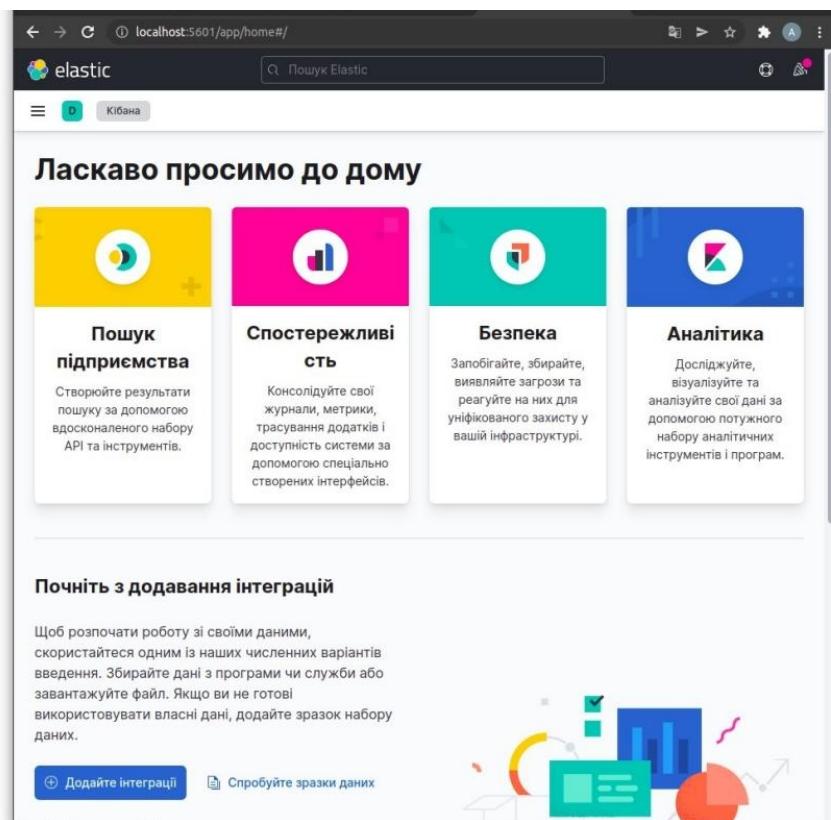
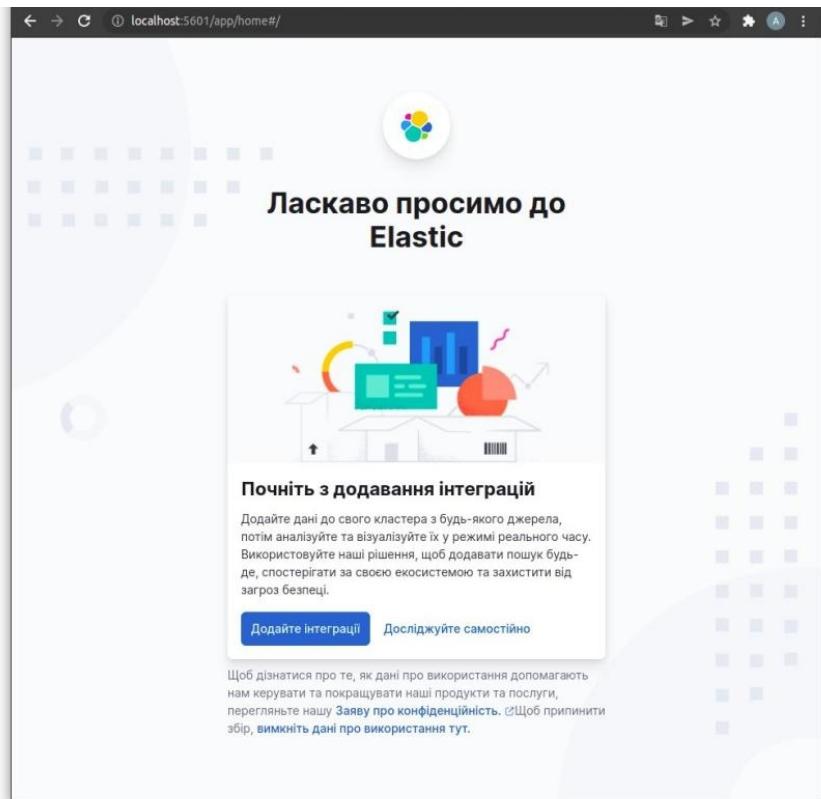
```
curl -X GET "localhost:9200/?pretty"
```



A screenshot of a terminal window displaying the output of a curl command. The command is "curl -X GET "localhost:9200/?pretty"". The output is identical to the one shown in the browser screenshot, representing the Elasticsearch version information as a JSON object.

```
{  
  "name" : "asus-X505BP",  
  "cluster_name" : "elasticsearch",  
  "cluster_uuid" : "lJAqKyLoTRiW0qAAA4A4RA",  
  "version" : {  
    "number" : "7.17.0",  
    "build_flavor" : "default",  
    "build_type" : "tar",  
    "build_hash" : "bee86328705acaa9a6daede7140defd4d9ec56bd",  
    "build_date" : "2022-01-28T08:36:04.875279988Z",  
    "build_snapshot" : false,  
    "lucene_version" : "8.11.1",  
    "minimum_wire_compatibility_version" : "6.8.0",  
    "minimum_index_compatibility_version" : "6.0.0-beta1"  
  },  
  "tagline" : "You Know, for Search"  
}
```

Запустимо Kibana.



Завантаження даних з використанням терміналу Linux

Створюємо скрипт

Зauważення: команди не вводилися вручну! Використовуючи **Visual Studio Code** можна замінити кожне входження певного символу на команду відповідну. В даній лабораторній роботі для створення скрипту було здійснено заміну симовола:

```
{
```

на

```
curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d
'{
```

та

```
}
```

на

```
'
```

Додамо нове розширення до файлу `#!/bin/bash` та зробимо файл виконуваним `$chmod a+x script.sh`.

Здійснимо таку процедуру відповідно кожне входження легко заміняється.

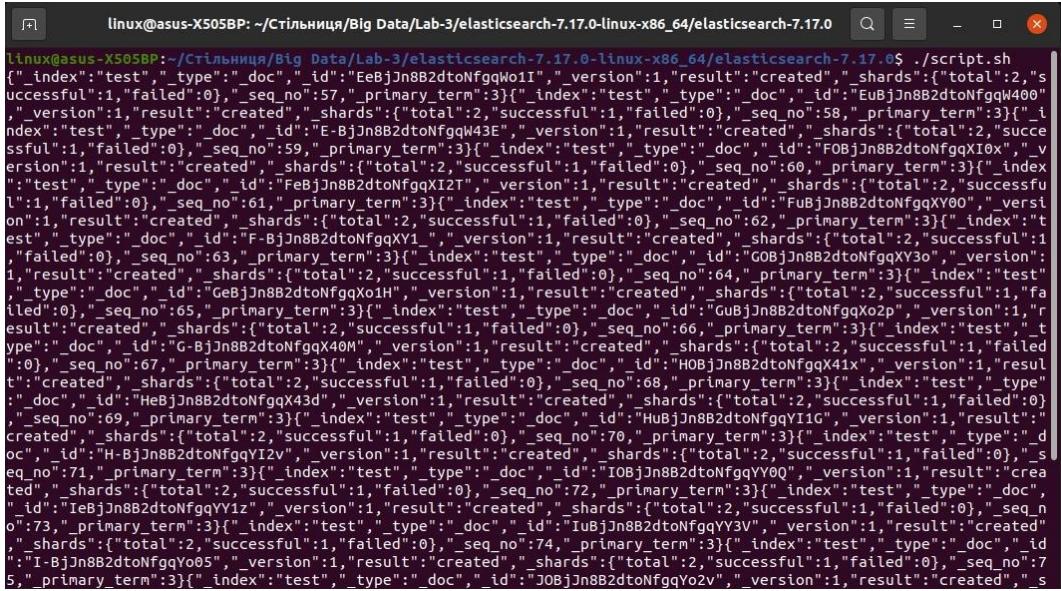
```
#!/bin/bash
curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d
{
  "title": "НБУ не втримав долар у межах 29 гривень",
  "textBody": "Торги на міжбанківському валютному ринку станом на 12:40 відбувалися на рівні 28 гривень 28–32 копійки за долар",
  "source": "НБУ не втримав долар у межах 29 гривень",
  "PubDate": "2022-02-23T13:06:00Z",
  "URL": "https://www.radiosvoboda.org/a/news-hryvnya-dolar-kurs-nbu/31718337.html"
}

curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d
{
  "title": "Засідання РНБО лишається відкритим, ніхто Київ не покидає – Данілов",
  "textBody": "У разі необхідності РНБО збиратиметься у дуже короткий проміжок часу",
  "source": "НБУ не втримав долар у межах 29 гривень",
  "PubDate": "2022-02-23T13:06:00Z",
  "URL": "https://www.radiosvoboda.org/a/news-rnbo-zasidannia-danilov/31718332.html"
}

curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d
{
  "title": "Визнання Росією «ДНР». Які це несе загрози для прифронтового Маріуполя",
  "textBody": "У Маріуполі сподіваються, що рішення Путіна спричинить швидші та жорсткіші санкції проти його",
  "source": "НБУ не втримав долар у межах 29 гривень",
  "PubDate": "2022-02-23T13:06:00Z",
  "URL": "https://www.radiosvoboda.org/a/novyny-pryazovya-rosiya-dnr-mariupol/31718220.html"
}

curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d
{
  "title": "В більшості областей України запровадять надзвичайний стан – секретар РНБО",
  "textBody": "Заходи можуть передбачати посилення охорони громадського порядку, обмеження певного руху транспорту, перевірку документів у громадян",
  "source": "НБУ не втримав долар у межах 29 гривень",
  "PubDate": "2022-02-23T13:06:00Z",
  "URL": "https://www.radiosvoboda.org/a/news-nadzvychainyi-stan-rnbo/31718296.html"
}
```

Використовуємо `./script.sh` для запуску файлу та отримуємо.



```
linux@asus-X505BP:~/Стільниця/Big Data/Lab-3/elasticsearch-7.17.0-linux-x86_64/elasticsearch-7.17.0$ ./script.sh
{
  "_index": "test", "_type": "doc", "_id": "EeBjJn8B2dtoNfgqXoI", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 57, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "EeBjJn8B2dtoNfgqW400", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 58, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "E-BjJn8B2dtoNfgqW43E", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 59, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "FOBjJn8B2dtoNfgqX10x", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 60, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "FeBjJn8B2dtoNfgqX12T", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 61, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "FuBjJn8B2dtoNfgqY00", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 62, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "F-BjJn8B2dtoNfgqXY1", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 63, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "G0BjJn8B2dtoNfgqXY3o", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 64, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "GeBjJn8B2dtoNfgqXoIh", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 65, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "GuBjJn8B2dtoNfgqXo2p", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 66, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "G-BjJn8B2dtoNfgqX40M", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 67, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "H0BjJn8B2dtoNfgqX41x", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 68, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "HeBjJn8B2dtoNfgqX43d", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 69, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "HuBjJn8B2dtoNfgqY11G", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 70, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "H-BjJn8B2dtoNfgqY12v", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 71, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "I0BjJn8B2dtoNfgqYYQ0", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 72, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "IeBjJn8B2dtoNfgqYY1z", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 73, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "IuBjJn8B2dtoNfgqYY3v", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 74, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "I-BjJn8B2dtoNfgqYo05", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 75, "_primary_term": 3} {"_index": "test", "_type": "doc", "_id": "JOBjJn8B2dtoNfgqYo2v", "_version": 1, "result": "created", "_shards": {"total": 2, "successful": 1, "failed": 0}, "_seq_no": 76, "_primary_term": 3}
```

Перевірка результатів завантаження даних.

```
linux@asus-X505BP:~/Стільниця/Big Data/Lab-3/elasticsearch-7.17.0-linux-x86_64/elasticsearch-7.17.0$ curl -XGET 'http://localhost:9200/test/_doc/_count'
```

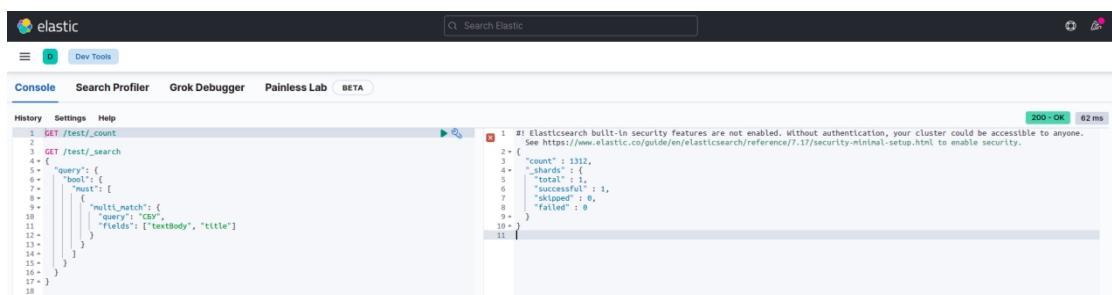
```
{"count":815, "_shards": {"total":1,"successful":1,"skipped":0,"failed":0}}
```

Створення пошукових запитів в Kibana аналогічних діям в терміналі з попередніх лабораторних робіт

Перевіримо котрі індекси наявні в базі даних Elasticsearch.

```
linux@linux-X505BP:~/Стільниця/Big Data/Lab-3/elasticsearch-7.17.0-linux-x86_64/elasticsearch-7.17.0$ curl 'localhost:9200/_cat/indices?v'
health status index          uuid                               pri  rep docs.count docs.deleted store.size pri.store.size
green  open   .geoip_databases Yekg6k2G5Je0TARC7d3nw  1   0      44           39        45mb       45mb
yellow open    test          H9RLM_WnTDGELVpkEyHgQ  1   1     1312            0      1.3mb      1.3mb
```

Перевіримо чисельність даних за допомогою Console в Kibana.



```
1 GET /test/_count
2
3 GET /test/_search
4 {
5   "query": {
6     "bool": {
7       "must": [
8         {
9           "multi_match": {
10             "query": "CBD",
11             "fields": ["textBody", "title"]
12           }
13         }
14       ]
15     }
16   }
17 }
```

```
1 #! Elasticsearch built-in security features are not enabled. Without authentication, your cluster could be accessible to anyone. See https://www.elastic.co/guide/en/elasticsearch/reference/7.17/security-minimal-setup.html to enable security.
2
3 {"error": true, "status": 401, "index": "test", "type": "search", "id": null, "score": null, "index_score": null}
4 {
5   "total": 1,
6   "successful": 1,
7   "skip": 0,
8   "failed": 0
9 }
10
11 }
```

Створимо простий пошуковий запит. На привеликий жаль, поточна версія локального вебсерверу не має фільтрації даних з відповідними полями котрі були подані в методичних рекомендаціях.

Але фільтрація виконується успішно.

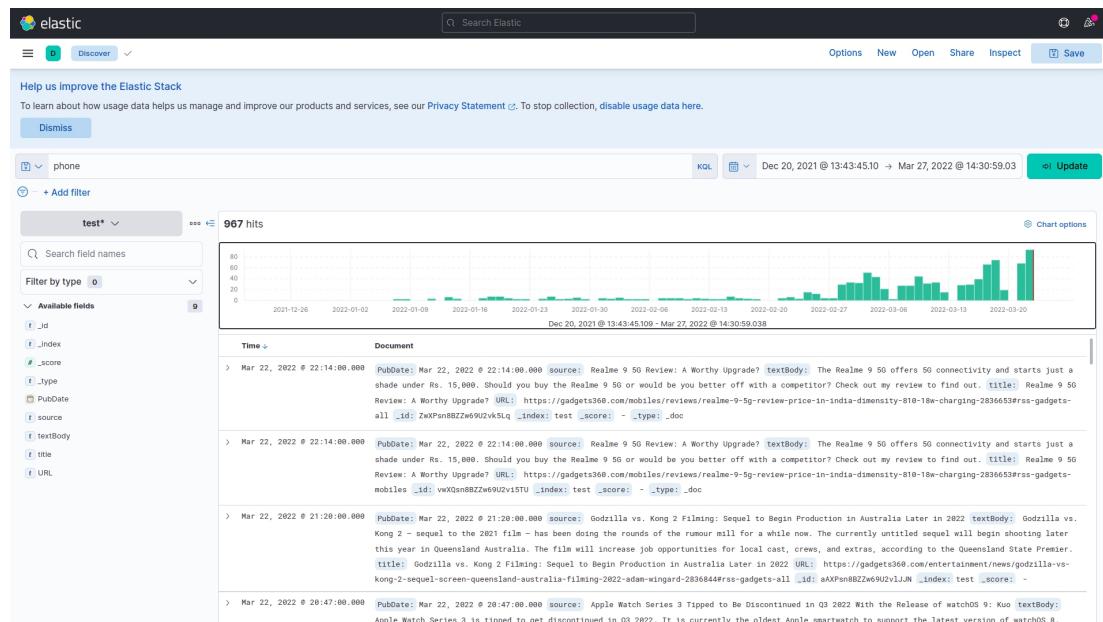
```

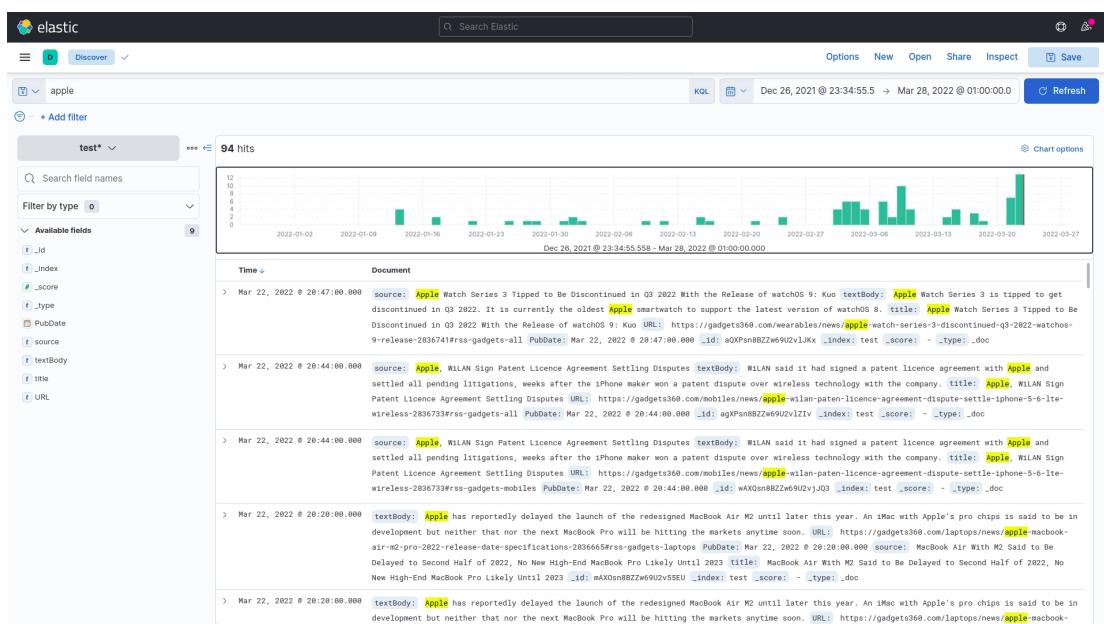
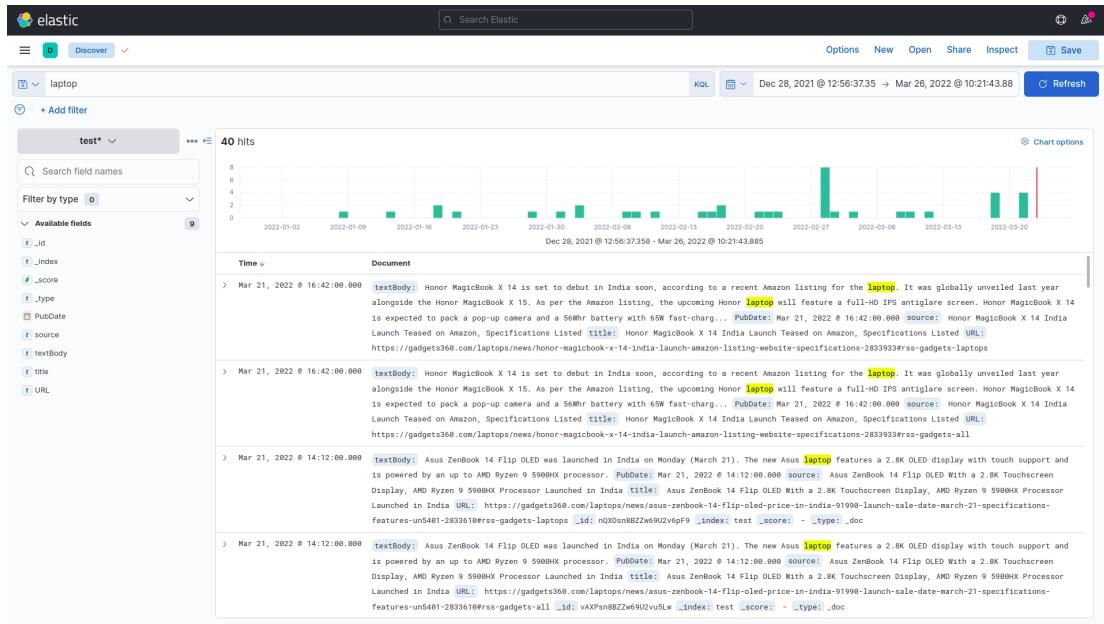
1 #! flancetsearch built-in security features are not enabled. Without authentication, your cluster could be accessible to anyone.
2 See https://www.elastic.co/guide/en/elasticsearch/reference/7.17/security-minimal-setup.html to enable security.
3
4 {
5   "query": {
6     "bool": {
7       "must": [
8         {
9           "multi_match": {
10             "query": "CD",
11             "fields": ["textBody", "title"]
12           }
13         }
14       ]
15     }
16   }
17 }
18
19
20
21
22
23
24
25
26
27
28

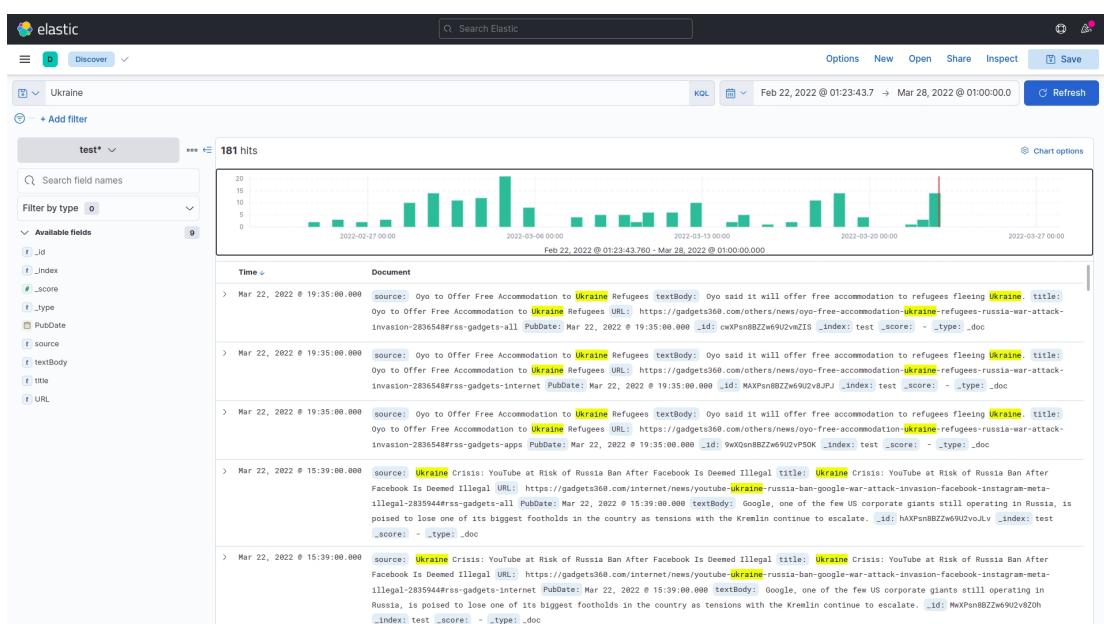
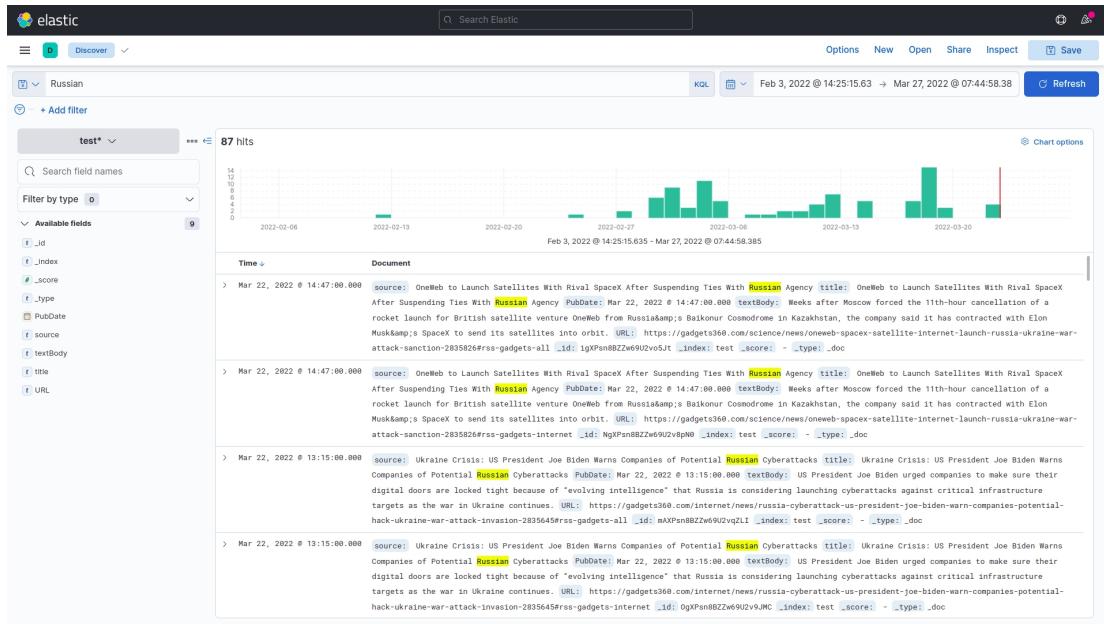
```

The screenshot shows the Elasticsearch Dev Tools interface with the 'Console' tab selected. A search query is entered and executed successfully. The response shows a single hit from a document with ID 'VhmpM8XKZr9sq7VdVd0'. The hit includes fields like '_index', '_id', '_score', '_type', '_source', and '_text'. The '_text' field contains the full text of the document, which is a news article about a CD being sold for 150 rubles.

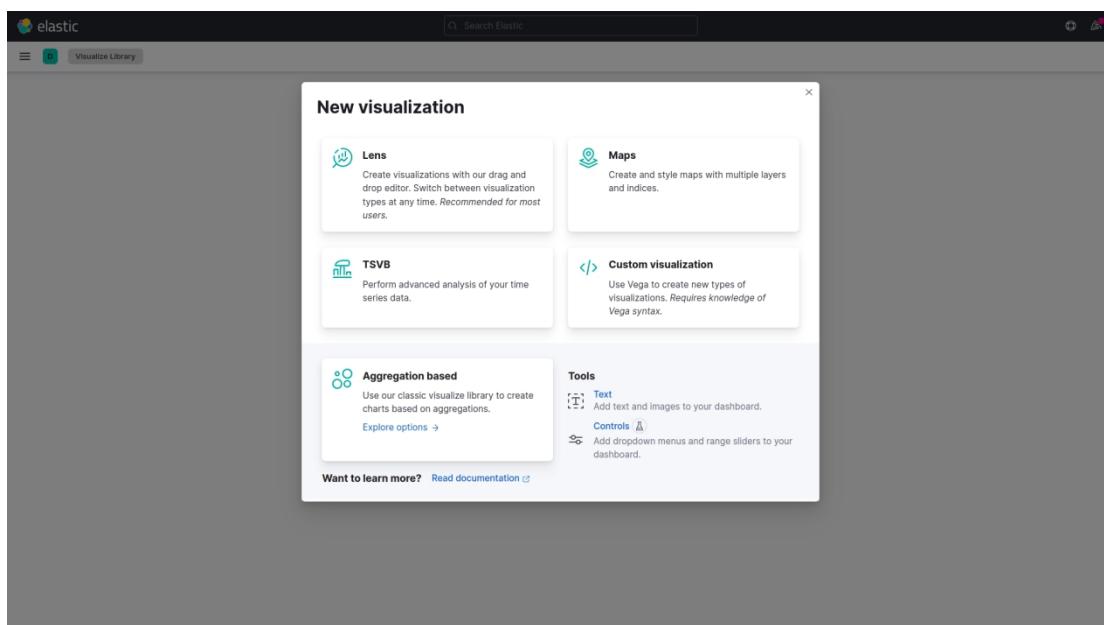
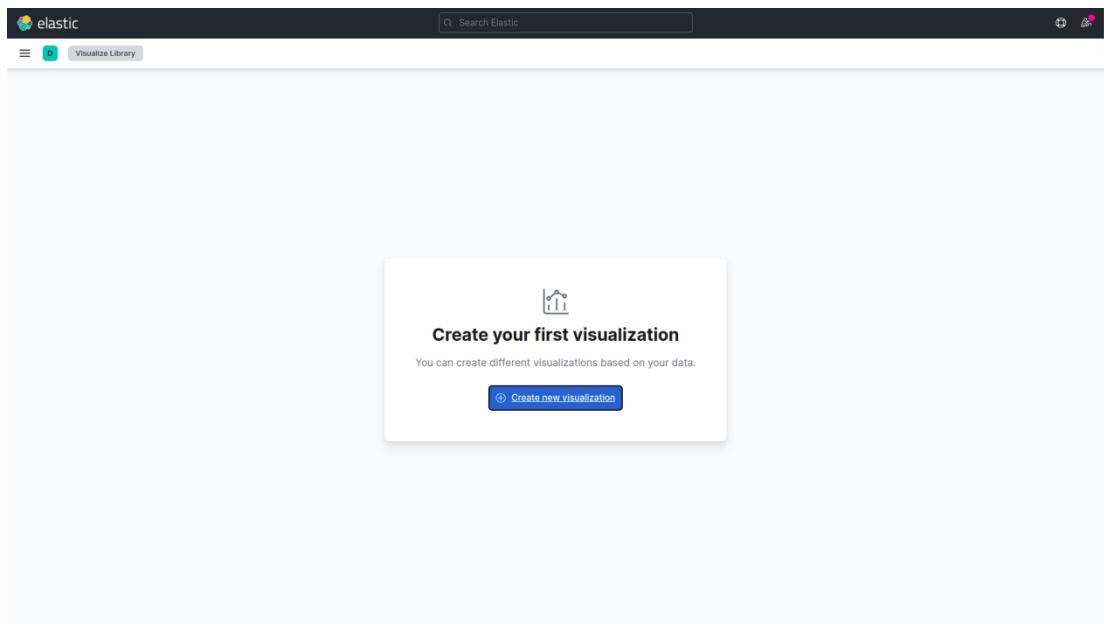
Робота із інструментом Discover

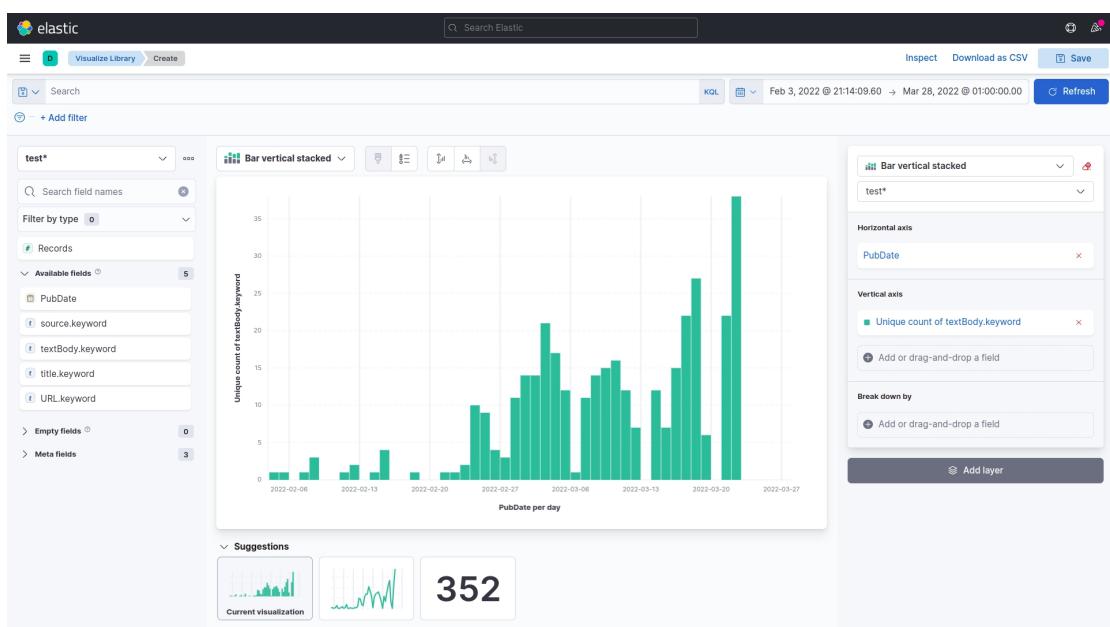
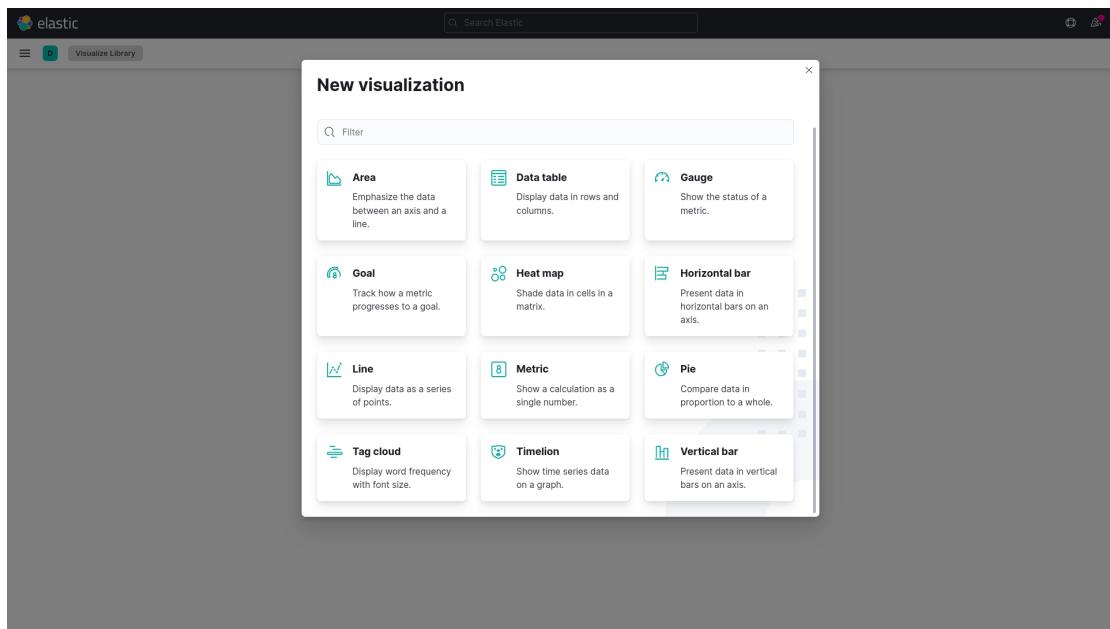


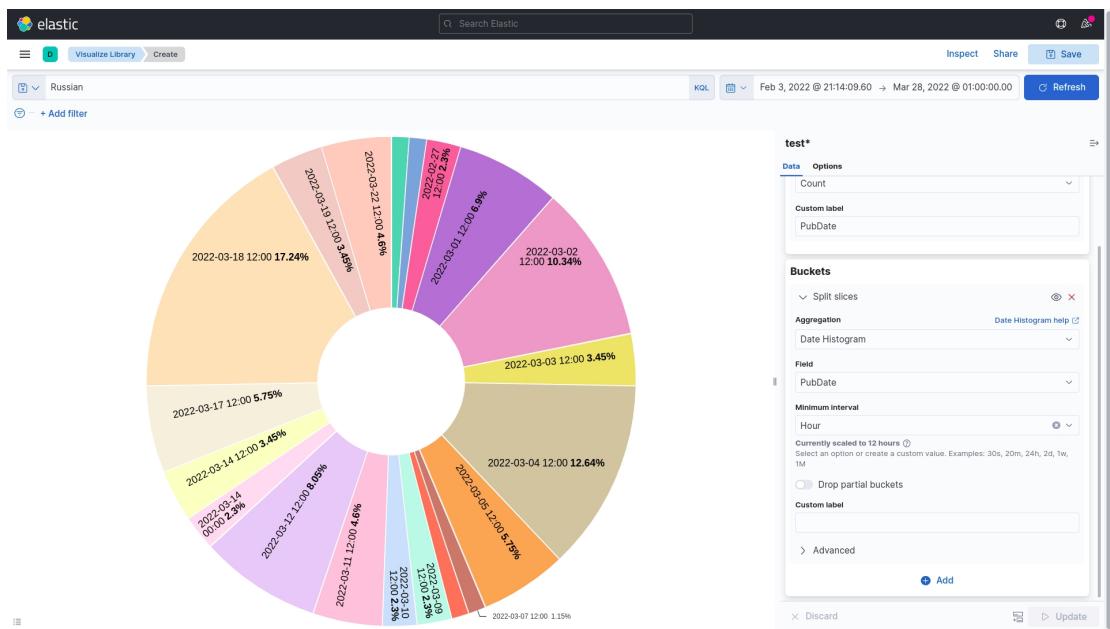
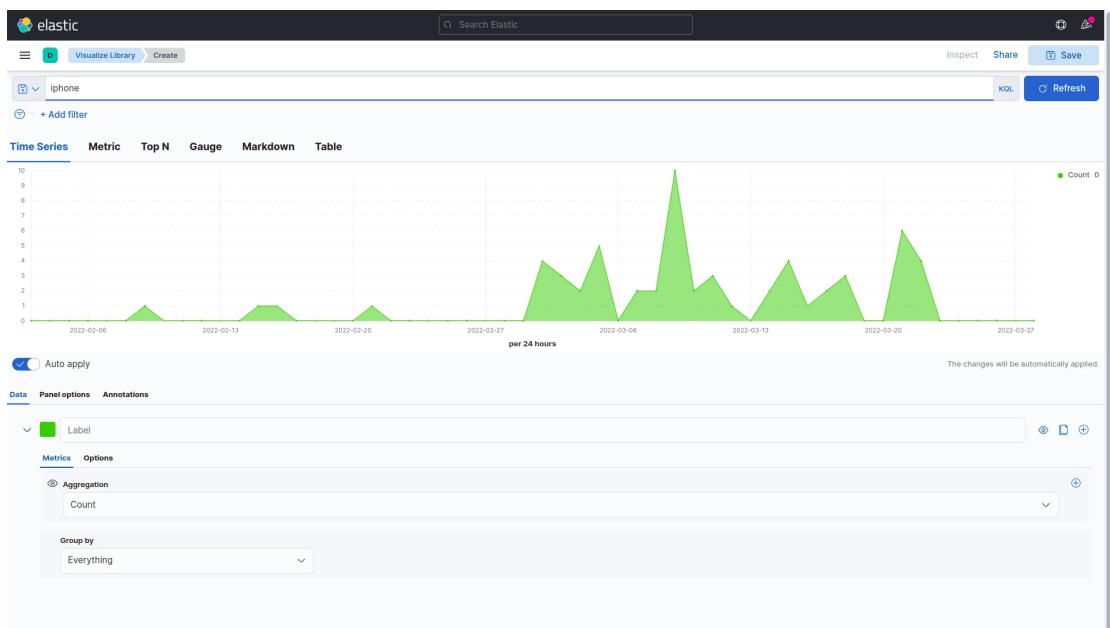


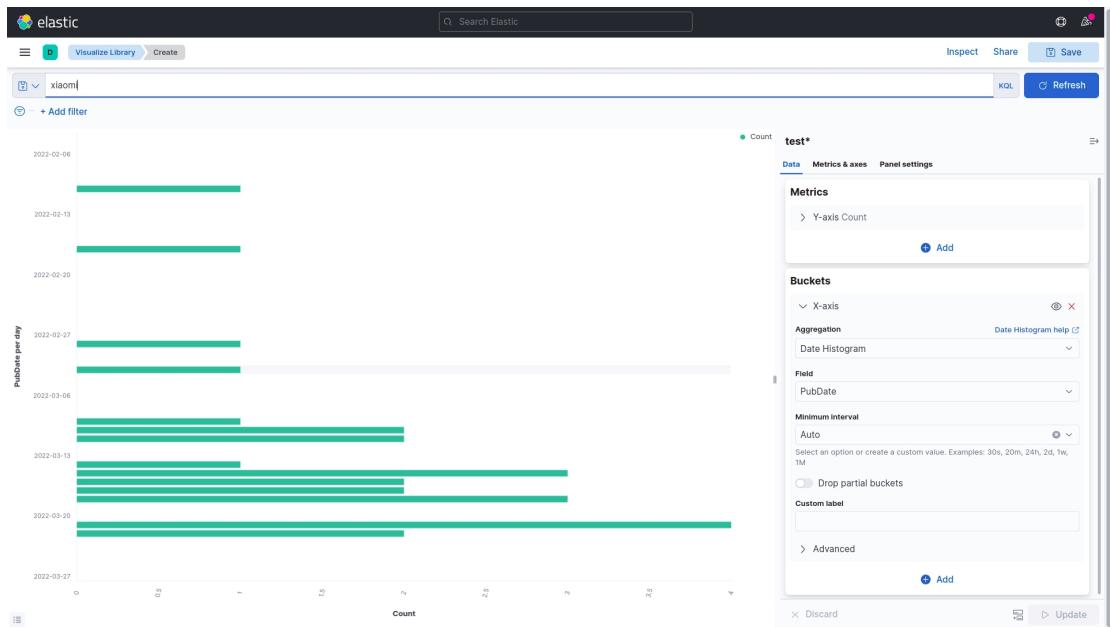
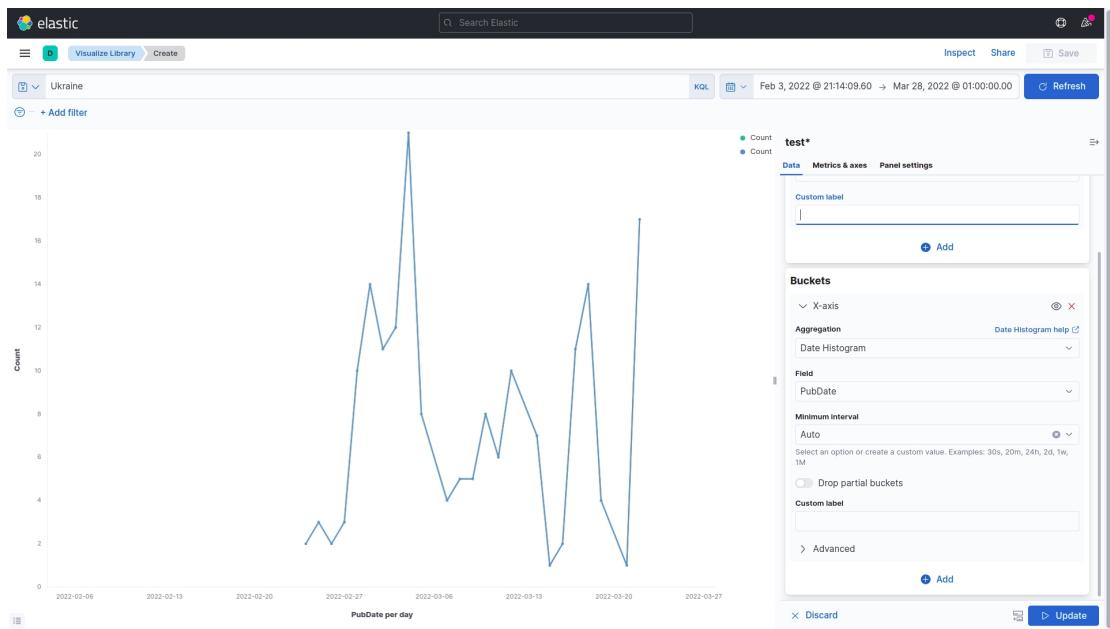


Робота із інструментом Visualizations









Питання до практичної роботи

1. Що таке система **Kibana** та яка її роль в обробці даних системою **Elasticsearch**?

Kibana — це plug-in з відкритим кодом для візуалізації даних з **Elasticsearch**. Він забезпечує можливості візуалізації змісту проіндексованого кластером **Elasticsearch**. Користувачі можуть створювати гістограми, лінійні і точкові діаграми, або діаграми і карти на великих обсягах даних.

2. Що таке інструмент **Console** та її застосування при роботі з **Elasticsearch**?

Kibana Console представляє собою зручний редактор, який підтримує функцію автозаповнення та форматування запитів під час їх написання. В **Elasticsearch** такий редактор значно полегшує роботу, та заміняє роботу з терміналом, його можна порівнювати з **IDE**.

3. Що таке інструмент **Discover** і його можливості при візуалізації результатів роботи **Elasticsearch**?

Інструмент **Kibana Discover** (Дослідження) призначено для інтерактивного аналізу даних. Користувач за допомогою цього інструменту може виконувати запити пошуку, фільтрації, переглядати дані документів. Можна також зберігати запити на пошук або критерії фільтрації для повторного використання або створення візуалізацій на базі відфільтрованих результатів.

4. Що таке інструмент **Visualizations** і його можливості при візуалізації результатів роботи **Elasticsearch**?

Інструмент **Kibana Visualizations** призначений для створення візуальних образів на основі даних (візуалізацій). Існують можливості сформувати різні варіанти оформлення: гістограми, лінійні діаграми, карти, хмари тегів тощо. Користувач може підібрати необхідні візуалізації для полегшення аналізу даних.