

Задание №2

Давайте обратимся к базе данных ClickHouse для извлечения данных и переноса их в DataFrame. Вот пример кода:

```
from clickhouse_driver import Client
import pandas as pd
import json

client = Client(host='localhost', port=9000, user='default', password='pivanet')

query_2 = 'SELECT parameters,timestamp, user, communication_number, communication_id, script_id, script_name, mrf, client_mrf, script_owner, c
result_2 = client.execute(query_2)

df_all = pd.DataFrame(result_2, columns=['parameters','timestamp', 'user', 'communication_number', 'communication_id', 'script_id', 'script_na
```

Список колонок:

```
df_all.columns
✓ 0.0s

Index(['parameters', 'timestamp', 'user', 'communication_number',
      'communication_id', 'script_id', 'script_name', 'mrf', 'client_mrf',
      'script_owner', 'current_script_owner', 'script_responsible',
      'current_script_responsible', 'crm_departament'],
      dtype='object')
```

Функция `extract_fields` конвертирует JSON-строку в словарь Python и извлекает специфические поля. С помощью `np.vectorize` создается векторизованная версия функции. Это позволяет применить функцию ко всему массиву данных одновременно, ускоряя выполнение. Эта векторизованная функция применяется к столбцу `parameters` DataFrame. Результат, массив словарей, преобразуется в DataFrame. Столбец `parameters` удаляется из исходного DataFrame, а полученный DataFrame добавляется в него. Затем выводятся первые строки обновленного DataFrame для проверки. Вот сравнение по скорости применения `.apply` с методом `np.vectorize`.

`apply`

```
pandas apply: 0.010205268859863281
```

np.vectorize

Vectorize time: 0.008959531784057617

Таким образом, np.vectorize работает быстрее

Мы получили итоговый датафрейм со всеми необходимыми колонками.

```
Index(['timestamp', 'user', 'communication_number', 'communication_id',  
      'script_id', 'script_name', 'mrf', 'client_mrf', 'script_owner',  
      'current_script_owner', 'script_responsible',  
      'current_script_responsible', 'crm_departament', 'ACCOUNT_NUMBER',  
      'CALLER_ID', 'COMMUNICATION_THEME', 'COMMUNICATION_DETAIL',  
      'COMMUNICATION_RESULT'],  
      dtype='object')
```

Затем мы создали таблицу `dataset_newww` с нашими колонками и загрузили в неё наш датафрейм.

dataset_newww | Введите SQL выражение чтобы отфильтровать результаты

	timestamp	user	communication_number	communication_id	script_id	script_name	mrf
1	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
2	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
3	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
4	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
5	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
6	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
7	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
8	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
9	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
10	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
11	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
12	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
13	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
14	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
15	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
16	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
17	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
18	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
19	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
20	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
21	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
22	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
23	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
24	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
25	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
26	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
27	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
28	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
29	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга
30	2024-04-09	user_03	344175369	53150993.0	17424b97-bf0a-45f9-9ce2-062eb39f3c37	Проблема с интернетом	Волга

Обновить Save Cancel Экспорт данных ... 200 760 760 строк получено - 0,025s (0,007s получ.)

