Задача

- Необходимо реализовать ETL процесс получения данных из внешнего API, трансформации и сохранение в хранилище.
- Процесс загрузки должен запускаться выбранным вами оркестратором с заданной частотой.
- Загруженные данные поступают в очередь, откуда вычитываются отдельным процессом, трансформируются и записываются в хранилище
- Результат выполнения должен быть оформлен в виде единого `git` репозитория и выложен на `github`. Обязательно должно присутствовать сам код, описание того как это запустить.

Источник

- https://randomuser.me/api/?results=5
- Структура входной записи:


```
{
  "results": [
    {
      "gender": "male",
      "name": {
        "title": "Mr",
        "first": "Nicolas",
        "last": "Lavigne"
      },
      "location": {
        "street": {
          "number": 1933,
          "name": "3rd St"
        },
        "city": "Summerside",
        "state": "Northwest Territories",
        "country": "Canada",
        "postcode": "Y2A 6S0",
        "coordinates": {
          "latitude": "-64.5793",
          "longitude": "-73.8569"
        },
        "timezone": {
          "offset": "+11:00",
          "description": "Magadan, Solomon Islands, New Caledonia"
        }
      },
      "email": "nicolas.lavigne@example.com",
      "login": {
        "uuid": "a8f31d60-af92-4f63-8cd2-3c42c4f08348",
        "username": "ticklishdog291",
        "password": "sergio",
        "salt": "P0Li8AoU",
        "md5": "740e39c5d841545b31a1e25c36514f2d",
        "sha1": "47fde4aafd0b28f3e4c17a55c4e35987cdd121ca",
        "sha256": "1dc75a50eebb7ce8ae7b2ab870b98a37534b10140ea03cb388c108b07db3b820"
      },
```

```
      "dob": {
        "date": "1990-04-21T07:25:46.953Z",
        "age": 33
      },
      "registered": {
        "date": "2007-08-25T04:49:29.906Z",
        "age": 16
      },
      "phone": "U29 A52-2599",
      "cell": "Q35 O08-6954",
      "id": {
        "name": "SIN",
        "value": "095191763"
      },
      "picture": {
        "large": "https://randomuser.me/api/portraits/men/13.jpg",
        "medium": "https://randomuser.me/api/portraits/med/men/13.jpg",
        "thumbnail": "https://randomuser.me/api/portraits/thumb/men/13.jpg"
      },
      "nat": "CA"
    }
  ],
  "info": {
    "seed": "0b52d5c09293331a",
    "results": 1,
    "page": 1,
    "version": "1.4"
  }
}
```


Хранилище

Выбор хранилища предоставляется исполнителю. Для простоты можно использовать локальный диск. Будет большим плюсом использование `s3-like` хранилище или БД.

- Структура выходной записи:


```
{
  "results": [
    {
      "gender": "male",
      "name": {
        "title": "Mr",
        "first": "Nicolas",
        "last": "Lavigne"
      },
      "location": {
        "city": "Summerside",
        "state": "Northwest Territories",
        "country": "Canada",
        "postcode": "Y2A 6S0"
      },
      "email": "nicolas.lavigne@example.com",
```

```
      "dob": {
        "date": "1990-04-21T07:25:46.953Z",
        "age": 33
      },
      "registered": {
        "date": "2007-08-25T04:49:29.906Z",
        "age": 16
      },
      "extracted": {
        "date": "${CURRENT_DATE}"
      },
      "phone": "U29 A52-2599",
      "cell": "Q35 O08-6954",
      "id": {
        "name": "SIN",
        "value": "095191763"
      }
    }
  ]
}
```

Формат выходных файлов

Требуется выбрать наиболее подходящий формат для представленных данных. Это может быть `AVRO`, `PARQUET`, `ORC` или другой. Необходимо обосновать свой выбор.

Примечание

Требуется чтобы весь пайплайн запускался локально. Для этих целей мы предлагаем использовать `docker compose`. Также необходимо использовать очередь сообщений, какую именно выбор за исполнителем. Для `TL` шага предпочтительно использовать `SPARK`

Дополнительно

В данной задаче необходимо показать свое знание технологий современного `Data Engineering` и общий технический кругозор. Оцениваться будет ход выполнения задания, выбор подходящих технологий и правильность их использования.