REPORT PROGRESS 1

Main task: Set up real-time data streaming from database

Tool: Redis, Spark

Task completed:

- Successfully installed Redis
- Successfully uploaded data to Redis

Tasks not completed:

- Set up real-time data streaming

Task execution process

1. **Install and set up Redis on Ubuntu**

Step 1: Update the system

Open terminal and run the following command to update the software package list:

```
sudo apt update && sudo apt upgrade -y
```

Step 2: Install Redis

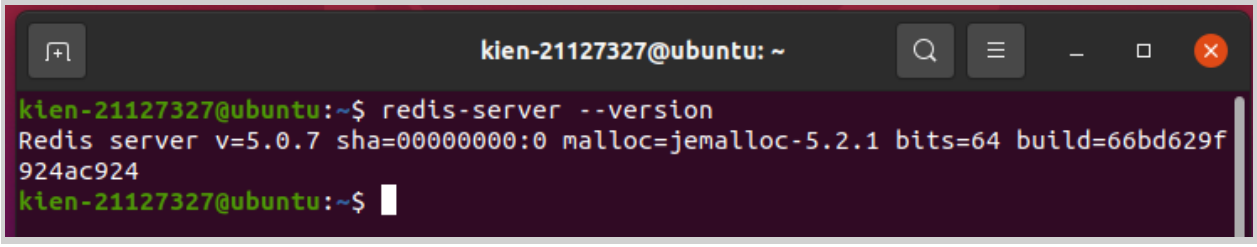
Ubuntu already has Redis available in the official repository, you can install it with the command:

```
sudo apt install redis-server -y
```

Step 3: Check Redis version

After installation, check Redis version to make sure the installation was successful:

```
redis-server–version
```



Step 4: Configure Redis

Open Redis config file for editing:

```
sudo nano /etc/redis/redis.conf
```

Find line:

```
supervised no
```

And replace with:

```
supervised system
```

This makes Redis run as a systemd service.

Also, if you want Redis to listen on all IPs (for example to connect from another machine), find:

```
bind 127.0.0.1 ::1
```

Replace with:

```
bind 0.0.0.0
```

Once done, press CTRL + X, select Y and press Enter to save.

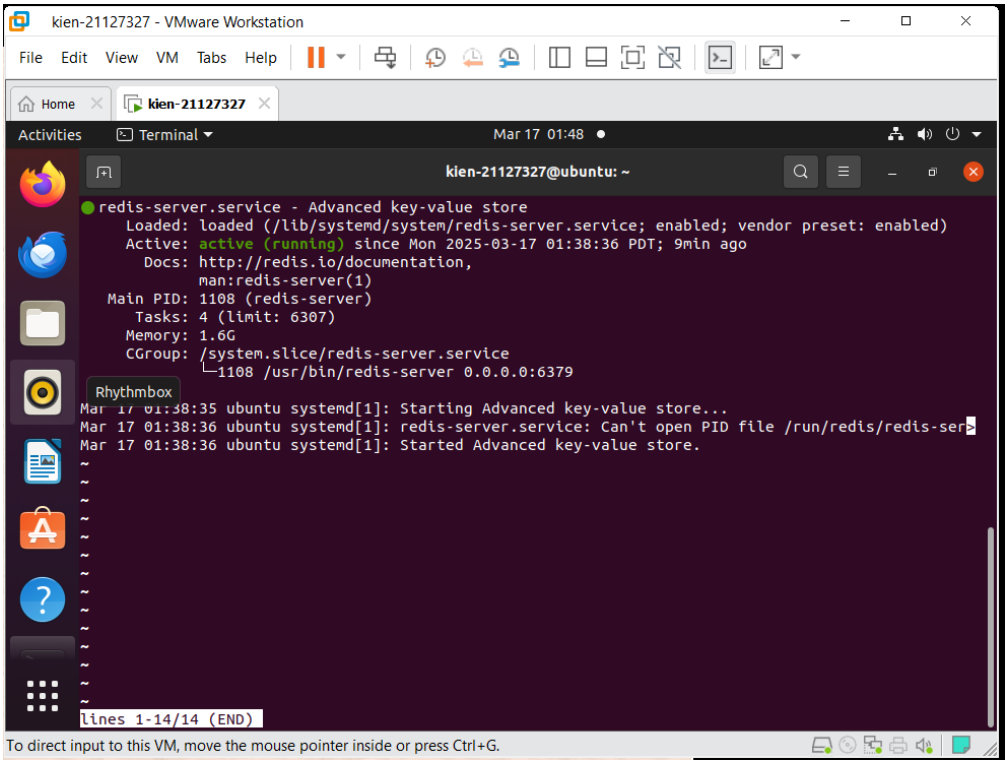
Step 5: Restart Redis

After changing the configuration, restart Redis to apply::

```
sudo systemctl restart redis
```

Check Redis status:

```
sudo systemctl status redis
```



If the line is found to be **active** (running), then Redis has run successfully.

Step 7: Set Redis to run automatically on startup

```
sudo systemctl enable redis
```

2. Upload data to Redis

There are many ways to do this, but in this article, we will use Python as the main language and the necessary libraries to perform this process.

The algorithm will be written in the file **Read_Data.py**.

Step 1: Install the necessary libraries

Before running the script, make sure Redis is installed on Ubuntu (see installation instructions above) and you need to install the **Redis-Py** and **Pandas** libraries:

```
pip install redis pandas
```

You can also use the **gc** library for purposes such as reducing memory consumption, increasing performance and preventing memory leaks.

Step 2: Prepare the data

The script uses three CSV files to upload data to Redis:

- **anime-dataset-2023.csv**: Contains information about anime series.
- **user-details-2023.csv**: Contains user personal information.
- **user-score-2023.csv**: Contains user anime rating information.

Step 3: Configure Redis Connection

Open the ReadData.py file and check the Redis connection information:

```
redis_host = "192.168.126.131" # Redis server IP address
redis_port = 6379 # Redis default configuration
```

If Redis runs on localhost, edit **redis_host = "127.0.0.1"**.

Step 4: Run the script to upload data

Run the **ReadData.py** file in Python to proceed with Uploading data:

Save anime information to Redis:

- Read the anime-dataset-2023.csv file in chunks (chunk_size=50000) to save memory.
- The data is saved as **Redis Hash**, each anime has a key of the type anime:{anime_id}.

```
r.hset(f"anime:{anime_id}", mapping=row)
```

Save user information to Redis:

- Đọc file user-details-2023.csv theo từng phần (chunk_size=50000).
- Mỗi người dùng có một loại khóa user:{user_id} được lưu dưới dạng Redis Hash.

```
r.hset(f"user:{user_id}", mapping=row)
```

Save user's ratings to Redis:

- Read users-score-2023.csv file (chunk_size=100000) in chunks.
- Save ratings as user_id_anime_id pairs as Redis Strings.

```
r.set(f"{user_id}_{anime_id}", rating)
```

Step 5: Check the data in Redis

After running the script, you can check Redis using Redis CLI.

Check the number of Keys currently in the Database

```
redis-cli

dbsize
```

```
127.0.0.1:6379> dbsize
(integer) 25081386
```

Check information of an anime

```
redis-cli

hgetall anime:1  # Kiểm tra thông tin anime có id = 1
```

```
11) "Aired"
12) "Apr 3, 1998 to Apr 24, 1999"
13) "Synopsis"
14) "Crime is timeless. By the year 2071, humanity has expanded across the galaxy, filling the sur
face of other planets with settlements like those on Earth. These new societies are plagued by mur
der, drug use, and theft, and intergalactic outlaws are hunted by a growing number of tough bounty
 hunters.\n\nSpike Spiegel and Jet Black pursue criminals throughout space to make a humble living
. Beneath his goofy and aloof demeanor, Spike is haunted by the weight of his violent past. Meanwh
ile, Jet manages his own troubled memories while taking care of Spike and the Bebop, their ship. T
he duo is joined by the beautiful con artist Faye Valentine, odd child Edward Wong Hau Pepelu Tivr
usky IV, and Ein, a bioengineered Welsh Corgi.\n\nWhile developing bonds and working to catch a co
lorful cast of criminals, the Bebop crew's lives are disrupted by a menace from Spike's past. As a
 rival's maniacal plot continues to unravel, Spike must choose between life with his newfound fami
ly or revenge for his old wounds."
15) "Studios"
16) "Sunrise"
17) "Members"
18) "1771505"
19) "Rank"
20) "41.0"
21) "Rating"
22) "R - 17+ (violence & profanity)"
23) "Episodes"
24) "26.0"
25) "Licensors"
26) "Funimation, Bandai Entertainment"
27) "anime_id"
28) "1"
29) "Popularity"
```

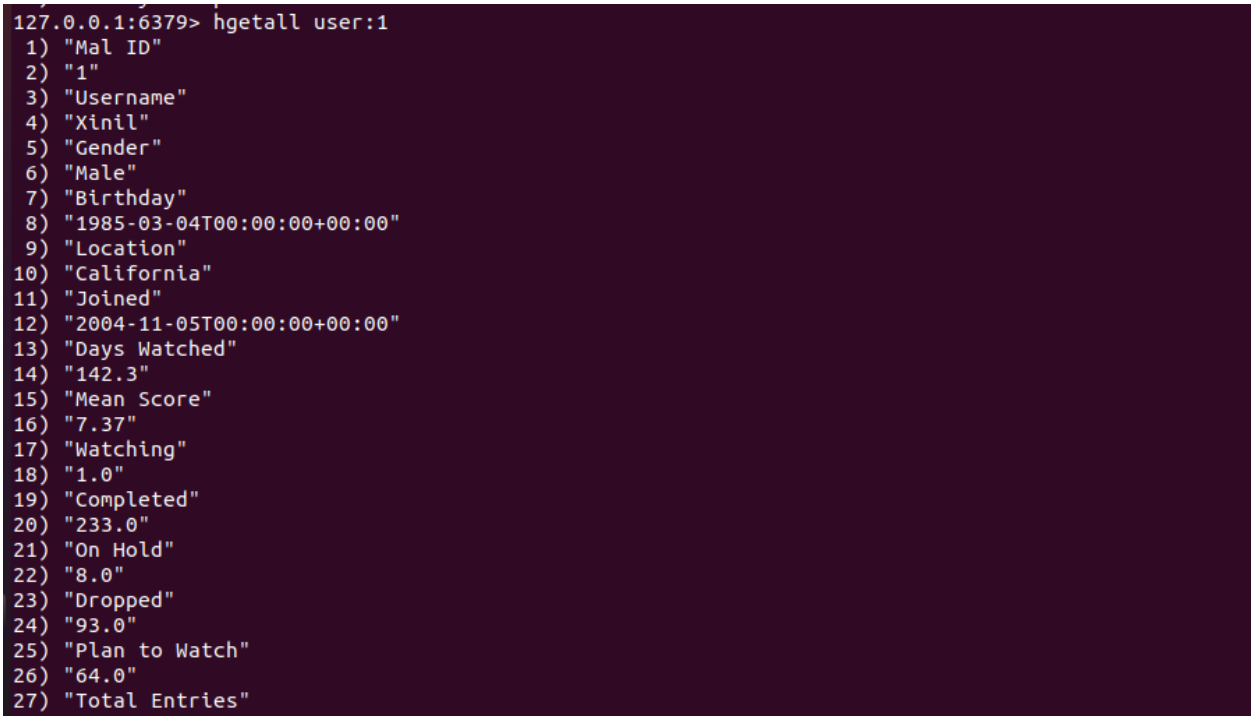
Check a user's rating

```
get 1_49  # Kiểm tra điểm đánh giá của user 1 cho anime có anime_id là 49
```

```
127.0.0.1:6379> GET 1_49
"8"
```

Check a user's information

```
hgetall user:1  # Kiểm tra thông tin user có id = 1
```



References

https://redis.io/docs/latest/operate/oss_and_stack/install/install-redis/install-redis-on-linux/

https://blog.vietnamlab.vn/khai-quat-ve-redis/

https://viblo.asia/p/redis-co-ban-57rVRq5OR4bP

https://chatgpt.com/