

```
import hashlib
```

```
def flajolet_martin(data):
```

```
    max_rightmost_zero_bit_position = -1
```

```
    for element in data:
```

```
        # Tính giá trị hash của phần tử
```

```
        hash_value = hashlib.md5(element.encode()).hexdigest()
```

```
        # Tính vị trí bit 0 cuối cùng
```

```
        rightmost_zero_bit_position = len(hash_value) - len(hash_value.rstrip('0'))
```

```
        # Cập nhật vị trí bit 0 cuối cùng lớn nhất
```

```
        if rightmost_zero_bit_position > max_rightmost_zero_bit_position:
```

```
            max_rightmost_zero_bit_position = rightmost_zero_bit_position
```

```
    # Tính ước lượng số phần tử riêng biệt
```

```
    estimated_number_of_distinct_elements = 2 ** max_rightmost_zero_bit_position
```

```
    return estimated_number_of_distinct_elements
```

```
data = "ab23 Big data is a field that treats ways to analyze from or otherwise deal with data  
Current usage of the term big data tends to predictive analytics user behavior, analytics or  
advanced data analytics Data sets grow rapidly in part because they are increasingly  
Internet of things devices such as mobile devices. Xz33. Small Small Small Small Small  
Small Small Small Small Small Small Small Small Small Small Small Small"
```

```
# Chia dữ liệu thành các phần tử riêng biệt
```

```
elements = data.split()
```

```
# Đếm số tên của các phần tử riêng biệt
```

```
count = flajolet_martin(elements)
```

```
print("Số tên của các phần tử riêng biệt là:", count)
```

B1: Thêm thư viện

```
import hashlib
```

B2: Cài đặt code

```
def flajolet_martin(data):
    max_rightmost_zero_bit_position = -1

    for element in data:
        # Tính giá trị hash của phần tử
        hash_value = hashlib.md5(element.encode()).hexdigest()

        # Tính vị trí bit 0 cuối cùng
        rightmost_zero_bit_position = len(hash_value) - len(hash_value.rstrip('0'))

        # Cập nhật vị trí bit 0 cuối cùng lớn nhất
        if rightmost_zero_bit_position > max_rightmost_zero_bit_position:
            max_rightmost_zero_bit_position = rightmost_zero_bit_position

    # Tính ước lượng số phần tử riêng biệt
    estimated_number_of_distinct_elements = 2 ** max_rightmost_zero_bit_position

    return estimated_number_of_distinct_elements

data = "ab23 Big data is a field that treats ways to analyze from or otherwise deal with data Current usage of the term big data tends to predictive analytics user behavior, analytics or advanced data analytics Data sets grow rapidly in part bec
# Chia dữ liệu thành các phần tử riêng biệt
elements = data.split()
```

B3:: Chụp hình kết quả

Unzbleed1.pyynb

Python 3 (ipykernel)

```
[2]: import hashlib

def flajolet_martin(data):
    max_rightmost_zero_bit_position = -1

    for element in data:
        # Tính giá trị hash của phần tử
        hash_value = hashlib.md5(element.encode()).hexdigest()

        # Tính vị trí bit 0 cuối cùng
        rightmost_zero_bit_position = len(hash_value) - len(hash_value.rstrip('0'))

        # Cập nhật vị trí bit 0 cuối cùng lớn nhất
        if rightmost_zero_bit_position > max_rightmost_zero_bit_position:
            max_rightmost_zero_bit_position = rightmost_zero_bit_position

    # Tính ước lượng số phần tử riêng biệt
    estimated_number_of_distinct_elements = 2 ** max_rightmost_zero_bit_position

    return estimated_number_of_distinct_elements

data = "ab23 Big data is a field that treats ways to analyze from or otherwise deal with data Current usage of the term big data tends to predictive analytics user behavior, analytics or advanced data analytics Data sets grow rapidly in part bec
# Chia dữ liệu thành các phần tử riêng biệt
elements = data.split()

# Đếm số tên của các phần tử riêng biệt
count = flajolet_martin(elements)

print("Số tên của các phần tử riêng biệt là:", count)
Số tên của các phần tử riêng biệt là: 5
```