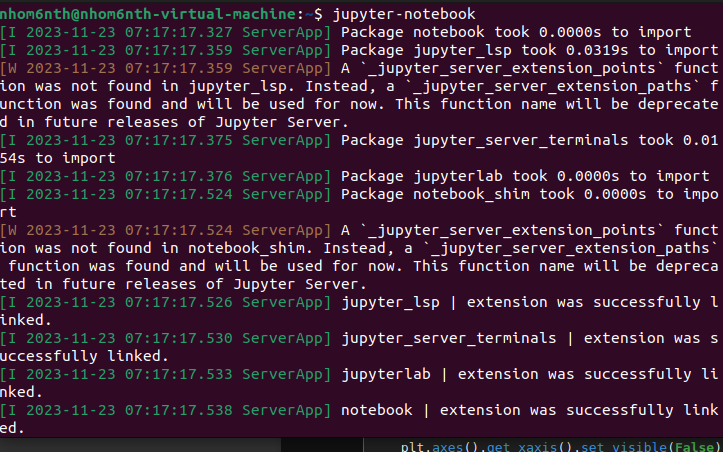
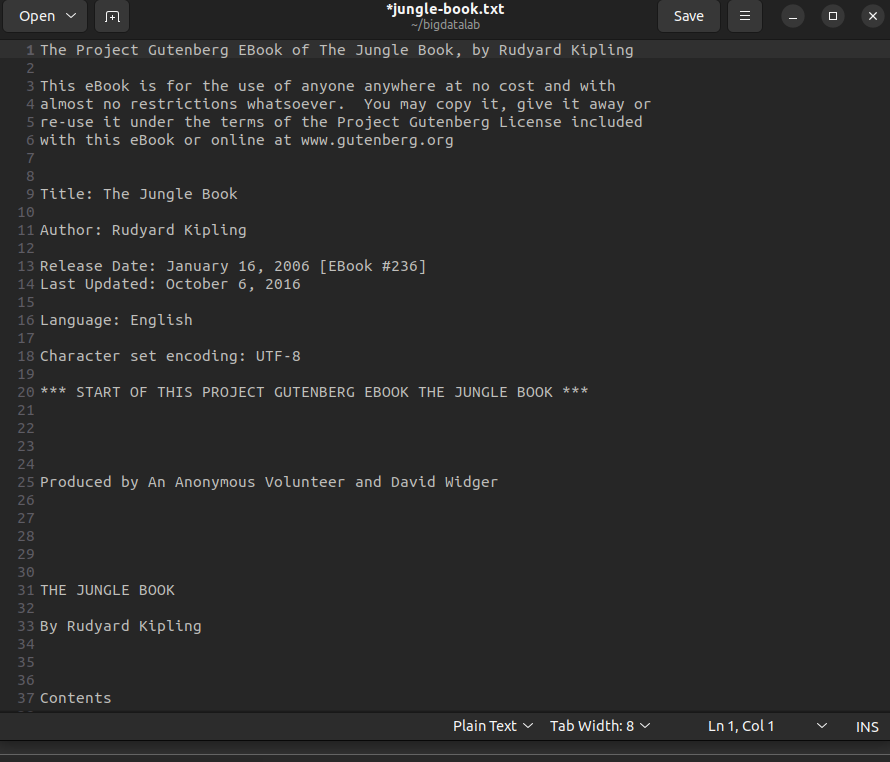
**Lab2**

**Lab6**

B1: Kết nối Jupyter



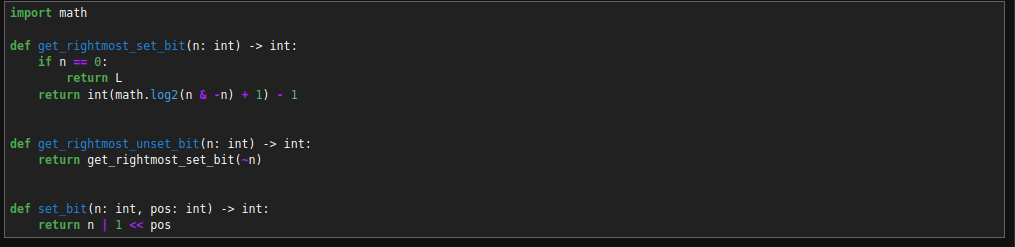
B2: tạo tệp jungle-book.txt



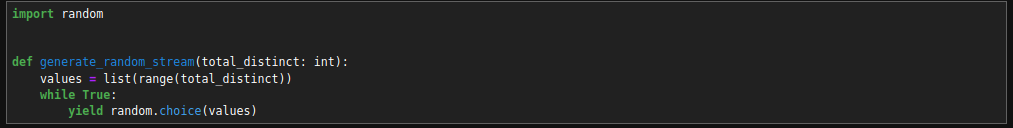
B3: định dạng cho chuỗi



B4: Sử dụng thuật toán của Flajolet-Martin để ước lượng độ phân biệt trong dữ liệu



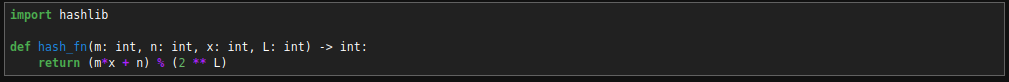
B5: Tạo ra một dãy số ngẫu nhiên từ một tập hợp giá trị có độ phân biệt



B6: Đếm số lượng giá trị phân biệt trong dãy số đầu vào



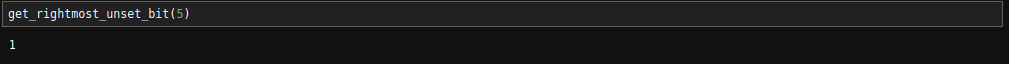
B7: chuyển đổi giá trị x thành một giá trị băm trong dãy bit có độ dài là L



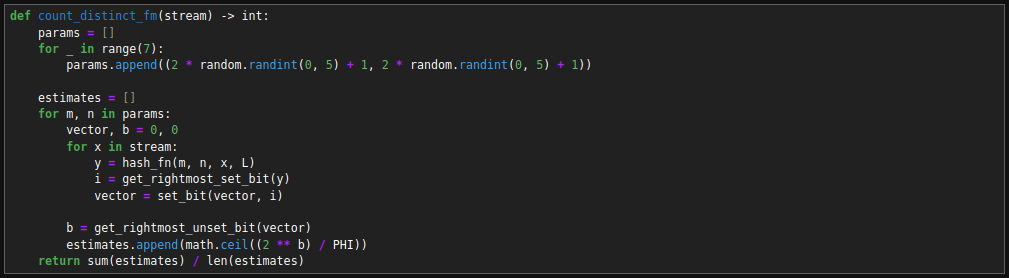
B8:



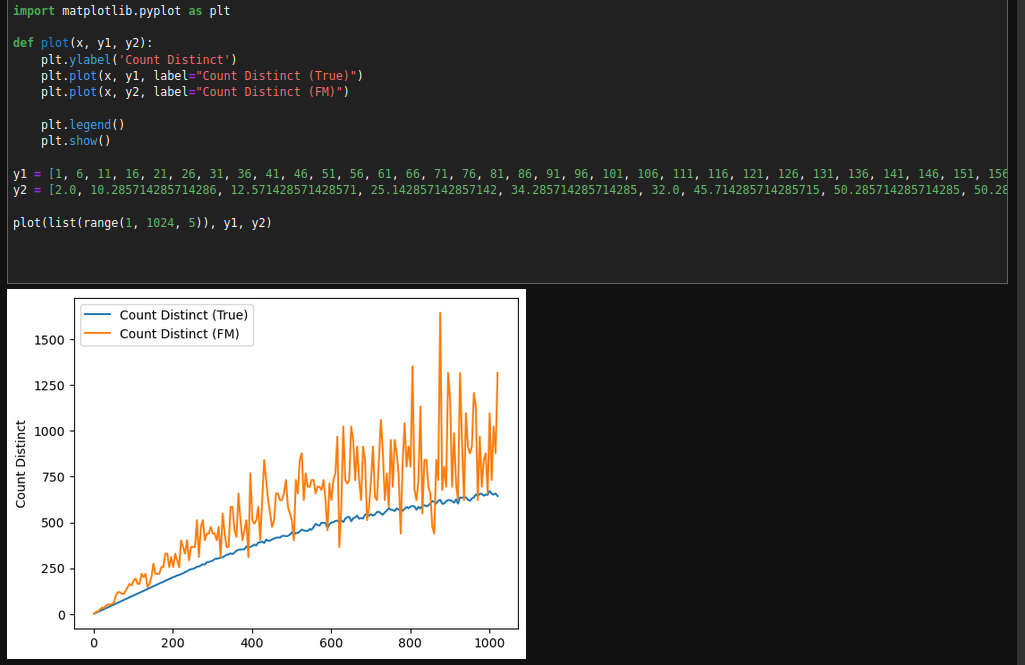
B9:



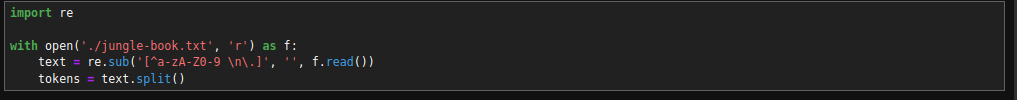
B10: Ước lượng độ phân biệt trong dãy số



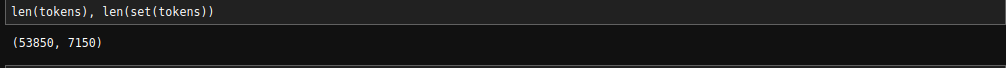
B11: Nhập dữ liệu và xuất ra màn hình biểu đồ



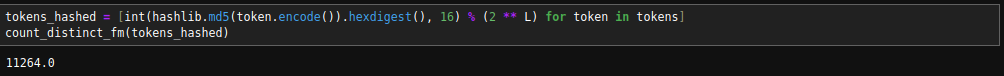
B12: mở file jungle-book.txt



B13: Đo độ dài file

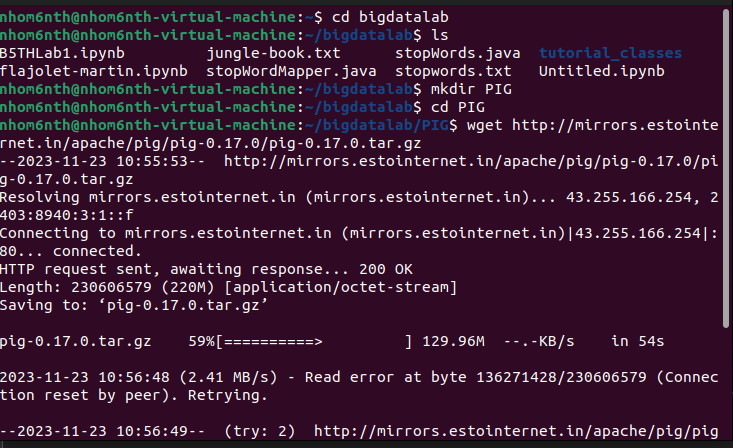


B14: ước lượng số lượng phần tử duy nhất trong danh sách tokens bằng cách sử dụng phương pháp Flajolet-Martin trên các giá trị băm của từng token

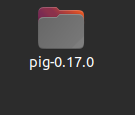


Lab7:

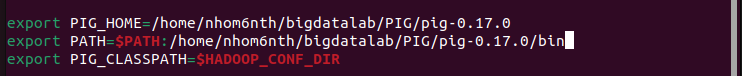
B1: Download the Pig Files from Apache



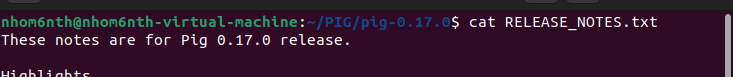
B2: Giải nén



B3: Đặt biến môi trường



B4: Kiểm tra version



B5: Khởi động Pig



