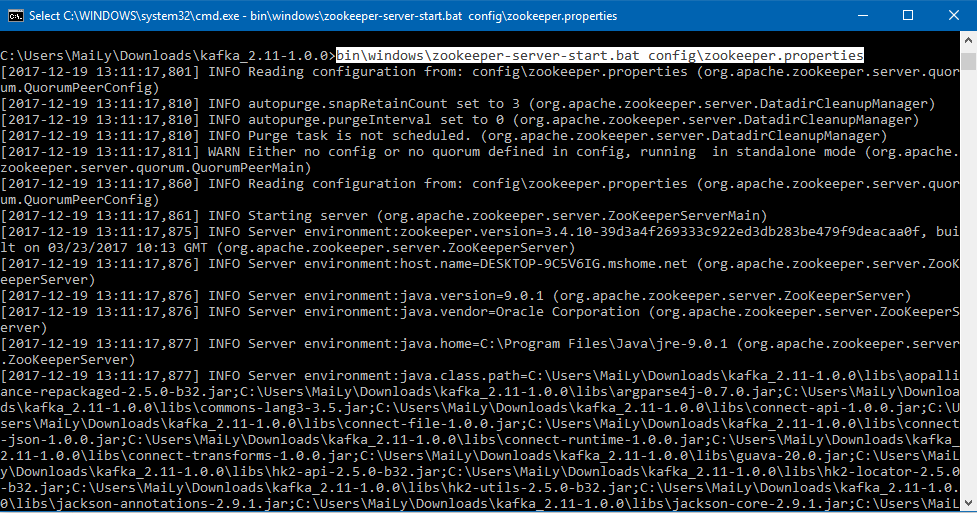
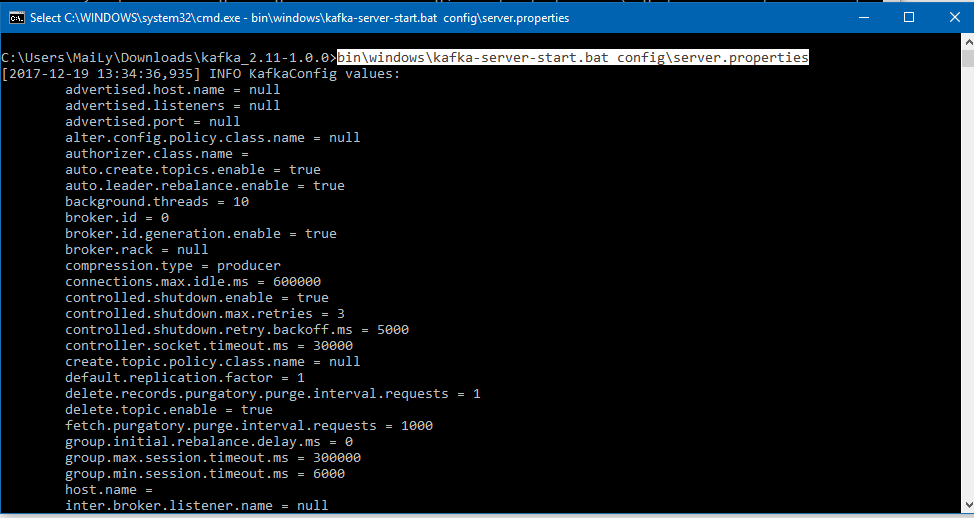
**Apache Kafka – Basic Operations**

Mở command line và trỏ đường dẫn tới folder kafka tải về, sau đó thực hiện các lệnh sau:

1. Start ZooKeeper server
   * ***bin\windows\zookeeper-server-start.bat config\zookeeper.properties***

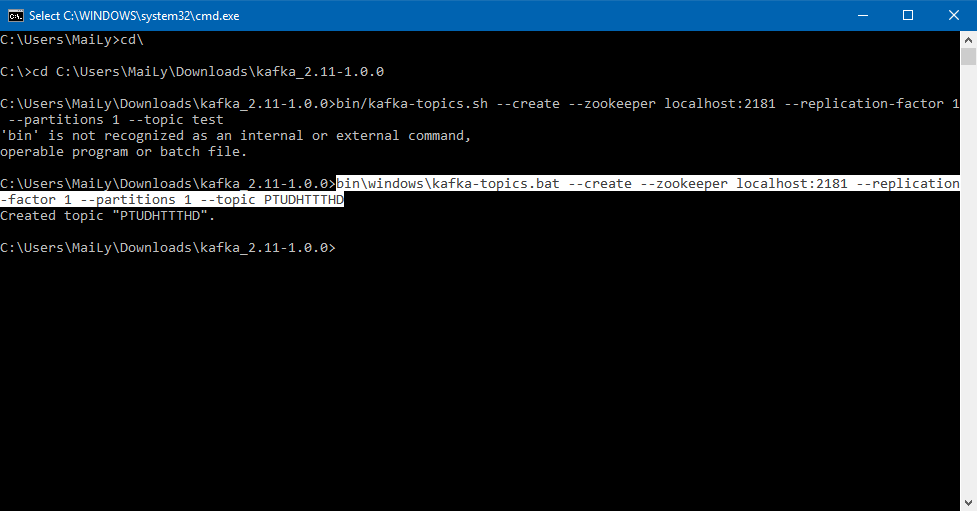


1. Start Kafka server
   * ***bin\windows\kafka-server-start.bat config\server.properties***



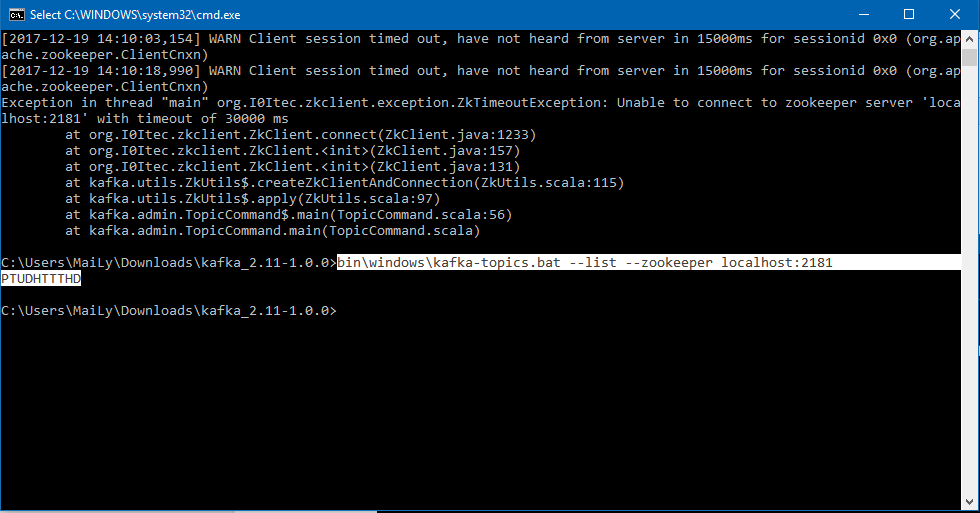
1. Create a Topic
   * Tạo 1 topic có tên “PTUDHTTTHD” với 1 partition và 1 replication factor:

***bin\windows\kafka-topics.bat --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic PTUDHTTTHD***

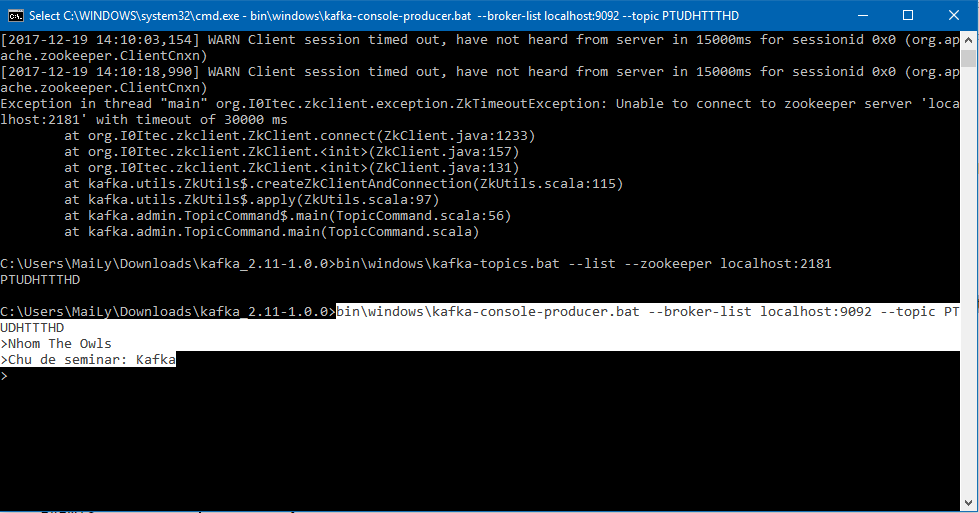


* + Xem topic:

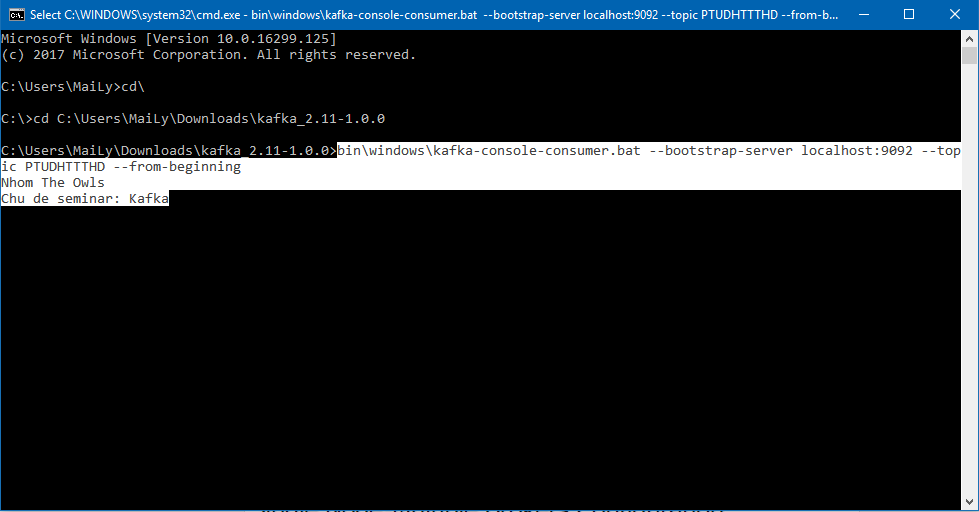
***bin\windows\kafka-topics.bat --list --zookeeper localhost:2181***



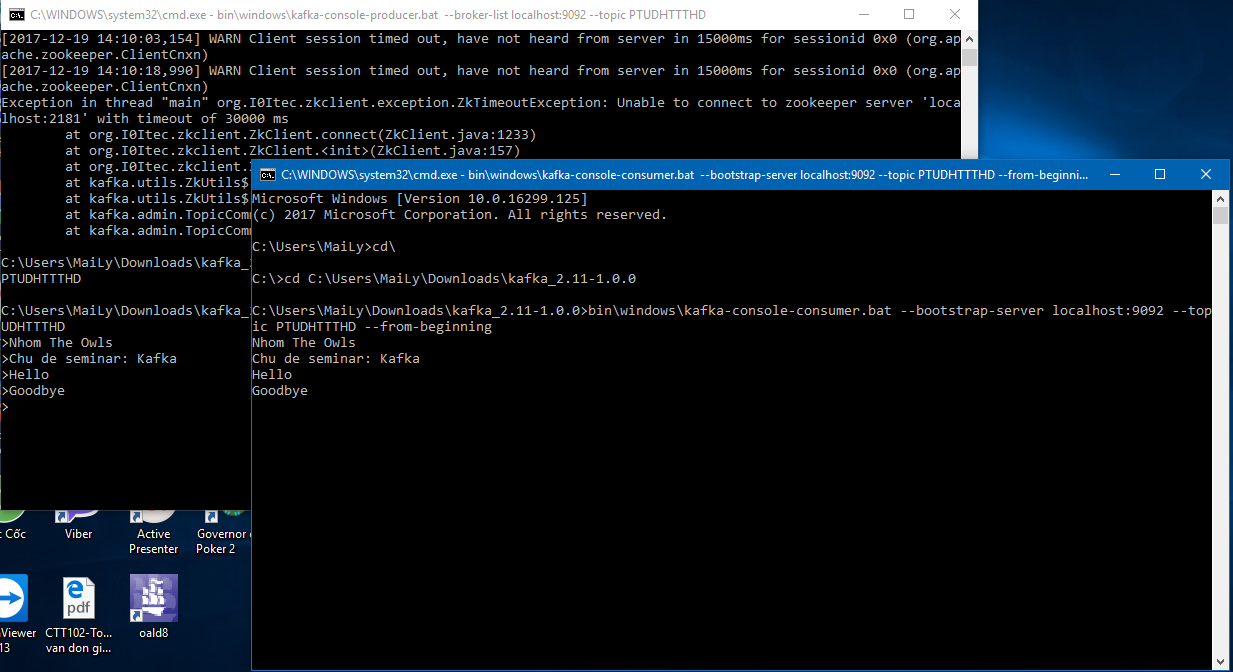
1. Start Producer to Send Messages
   * ***bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic PTUDHTTTHD***
   * Producer sẽ chờ input từ stdin và publish đến Kafka Cluster. Mỗi dòng input sẽ xuất ra dưới dạng một message mới. Trong câu lệnh trên thì dữ liệu được truyền thuộc topic “PTUDHTTTHD”, dữ liệu này sec được đưa đến partition của topic “PTUDHTTTHD” lưu trữ trên broker.



1. Start Consumer to Receive Messages
   * ***bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic PTUDHTTTHD --from-beginning***



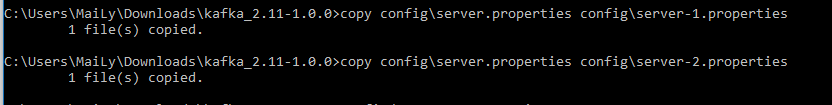
* Nếu producer và consumer ở trên vừa chạy trên 2 command line khác nhau thì khi tiếp tục gõ messages bên producer cmd thì messages sẽ xuất hiện bên consumer cmd ngay sau đó mà không cần chạy lại câu lệnh.



1. Setting up a Multi-Broker Cluster
   * Ta đã có sẵn một file server.properties cho 1 broker. Tiếp theo cần cấu hình thêm file cho mỗi broker cần tạo thêm. Copy file server.properties đã có thành file mới và đổi tên sau đó chỉnh sửa các thuộc tính của file:

***copy config\server.properties config\server-1.properties***

***copy config\server.properties config\server-2.properties***



**config/server-1.properties:**

broker.id=1

    listeners=PLAINTEXT://:9093

    log.dir=/tmp/kafka-logs-1

**config/server-2.properties:**

    broker.id=2

    listeners=PLAINTEXT://:9094

    log.dir=/tmp/kafka-logs-2

* Start Multi-Broker:

***bin\windows\kafka-server-start.bat config/server-1.properties***

***bin\windows\kafka-server-start.bat config/server-2.properties***

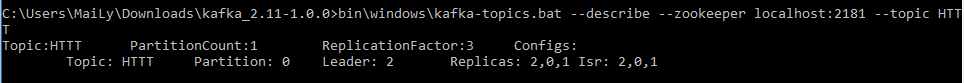
1. Create a Topic
   * Tạo một Topic mới với 3 replication factor vì có 3 broker khác nhau đang chạy:

***bin\windows\kafka-topics.bat --create --zookeeper localhost:2181 --replication-factor 3 --partitions 1 --topic HTTT***



* Để biết được broker đang làm những gì thì chạy lệnh “describe topics”:

***bin\windows\kafka-topics.bat --describe --zookeeper localhost:2181 --topic HTTT***

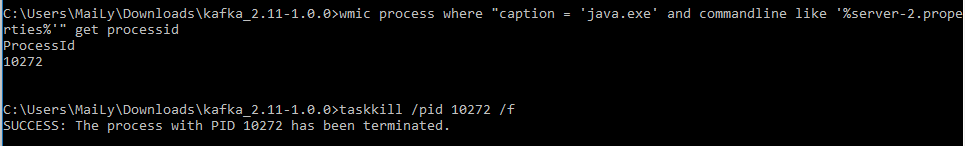


Với output từ lệnh trên thì dòng đầu tiên mô tả thông tin chung của topic với tổng số partition, số replication factor. Mỗi dòng con ở dưới mô tả thông tin về mỗi partition, vì topic này chỉ có một partition nên chỉ có một dòng mô tả partition.

* + “Leader”: chịu trách nhiệm cho tất cả các lần đọc và ghi cho partition đó. Broker Leader được lựa chọn ngẫu nhiên trong partition.
  + “Replicas”: danh sách các broker bản sao của topic.
  + “Isr”: bộ các bản sao “in-sync”, đây là bộ con của “Replicas” mà đang còn sống và “caught-up” với Leader.

1. Kill a Broker
   * Hiện tại thì broker 2 đang giữ vai trò Leader nên ta sẽ thử kill nó bằng lệnh:

***taskkill /pid 10272 /f***



1. Modify a Topic
   * Ta đã có topic tên”PTUDHTTTHD” với 1 partition, sử dụng lệnh sau để thay đổi số partiton thành 2:

***bin\windows\kafka-topics.bat --zookeeper localhost:2181 --alter --topic PTUDHTTTHD --partitions 2***

* + Xóa topic:

***bin\windows\kafka-topics.bat --zookeeper localhost:2181 --delete --topic HTTT***

1. Use Kafka Connect to import/export data
   * Kafka Connect là một công cụ kèm theo Kafka để nhập và xuất dữ liệu cho Kafka. Ta sẽ dùng Kafka Connect với các trình kết nối đơn giản nhập dữ liệu từ một file text vào một Topic và xuất dữ liệu từ Topic đó vào một file khác.
   * Tạo một số dữ liệu vào file test.txt:

***echo foo> test.txt***

***echo bar>> test.txt***

* + Cấu hình cho quá trình Kafka Connect gồm các cấu hình chung như broker để kết nối và định dạng serialization cho dữ liệu,…

***bin\windows\connect-standalone.bat config/connect-standalone.properties config/connect-file-source.properties config/connect-file-sink.properties***

* + Các file cấu hình này sử dụng cấu hình cluster cục bộ mặc định đã bắt đầu trước đó và tạo ra hai connectors: đầu tiên là connector đọc các dòng từ file input và tạo ra một topic và thứ hai là connector chìm đọc các messages từ topic và tạo ra ghi vào trong một file output.
  + Khi quá trình Kafka Connect bắt đầu, connector nguồn sẽ bắt đầu đọc các dòng từ test.txt và gửi tới topic connect-topic, và connector chìm bắt đầu đọc messages từ connect-topic và ghi vào file test.sink.txt, có thể xem dữ liệu ở file test.sink.txt như sau:

