**Phạm Đức Phát**

**1050080151**

**10CNTT4**

**Big Data thực hành tuần 3**

**Lab – Install Apache Kafka (Ubuntu)**

**Bước 1:** Tải Apache Kafka

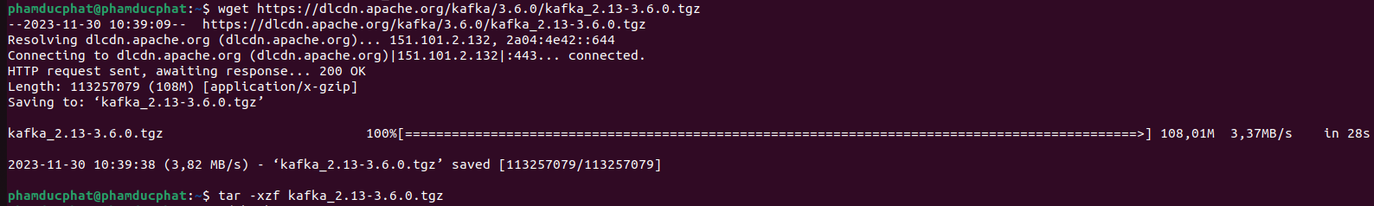
Tải tệp:

wget <https://dlcdn.apache.org/kafka/3.6.0/kafka_2.13-3.6.0.tgz>



Giải nén:

tar -xzf kafka\_2.13-3.6.0.tgz



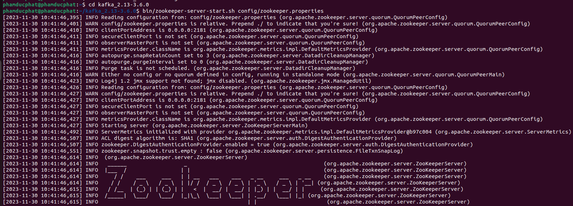
**Bước 2:** Khởi động môi trường Apache Kafka

Truy cập vào kafka\_2.13-3.6.0:

cd kafka\_2.13-3.6.0

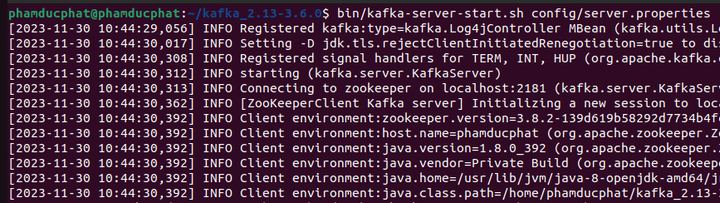
Khởi động Zookeeper:

bin/zookeeper-server-start.sh config/zookeeper.properties



Khởi động Kafka broker:

bin/kafka-server-start.sh config/server.properties



Kafka với Kraft:

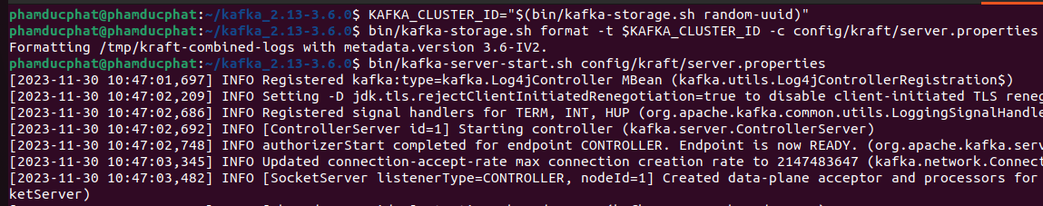
KAFKA\_CLUSTER\_ID="$(bin/kafka-storage.sh random-uuid)"

Định dạng thư mục nhật ký:

bin/kafka-storage.sh format -t $KAFKA\_CLUSTER\_ID -c config/kraft/server.properties

Khởi động máy chủ Kafka:

bin/kafka-server-start.sh config/kraft/server.properties

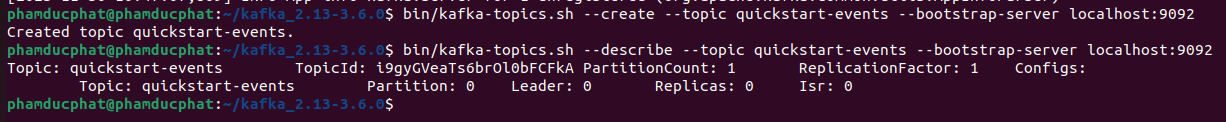


**Bước 3:** Tạo chủ đề và lưu trữ lại

bin/kafka-topics.sh --create --topic quickstart-events --bootstrap-server localhost:9092

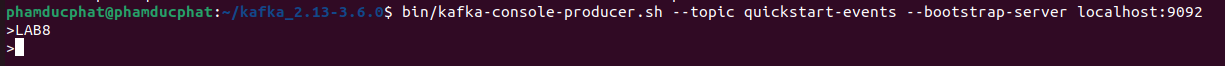
bin/kafka-topics.sh --describe --topic quickstart-events --bootstrap-server localhost:9092

bin/kafka-console-producer.sh --topic quickstart-events --bootstrap-server localhost:9092



**Bước 4:** Viết một số nội dung vào chủ đề

Lab 8



**Bước 5:** Đọc chủ đề

Lab 8



**Lab 8 Install Apache Hive**

Không có mô tả.

**Bước 1:** Tải Apache Hive

Tải tệp:

wget <http://archive.apache.org/dist/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz>



Giải nén và đổi tên thư mục

tar -xzf apache-hive-3.1.2-bin.tar.gz

mv apache-hive-3.1.2-bin hive



**Bước 2:** Cài biến môi trường Hive

Vào file .bashrc bằng trình chỉnh sửa nano

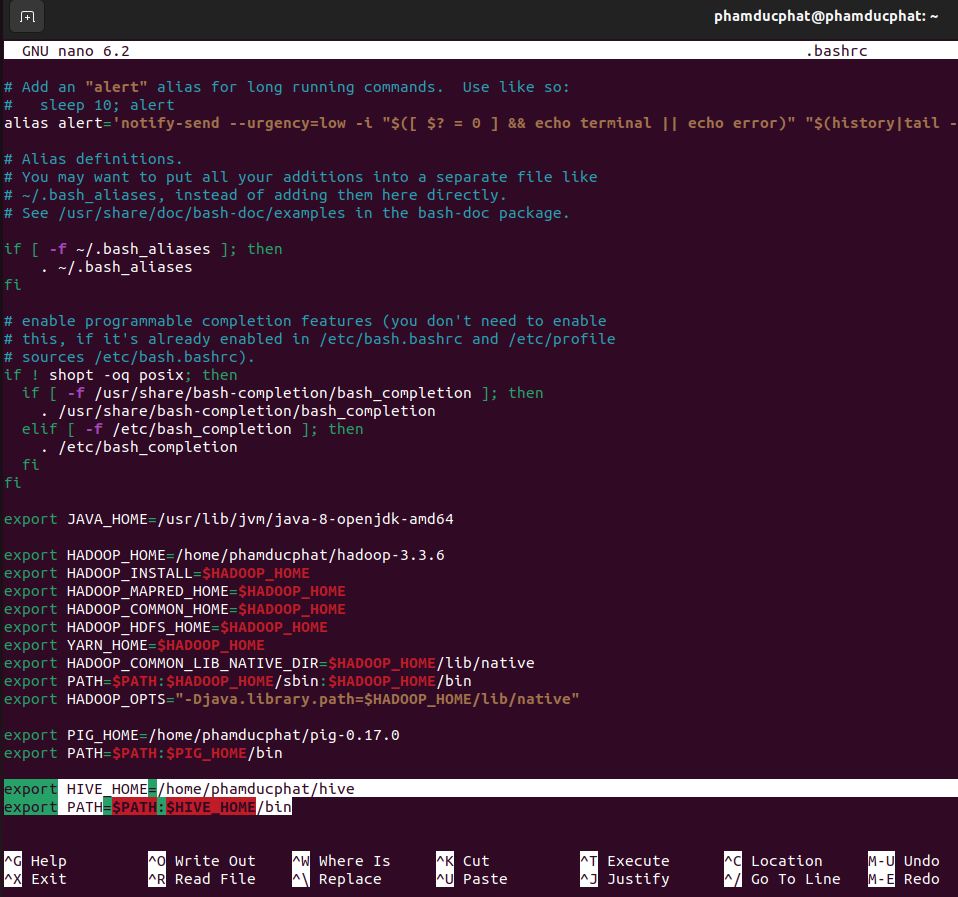
nano .bashrc



Chèn đường dẫn

export HIVE\_HOME=/ home/phamducphat/hive

export PATH=$PATH:$HIVE\_HOME/bin



Lưu (Ctrl S) và thoát (Ctrl X)

Thực thi môi trường

source ~/.bashrc



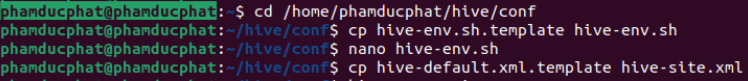
**Bước 3:** Thêm đường dẫn Hadoop vào tệp hive-env.sh

Tới vị trí /home/phamducphat/hive/conf

cd /home/phamducphat/hive/conf

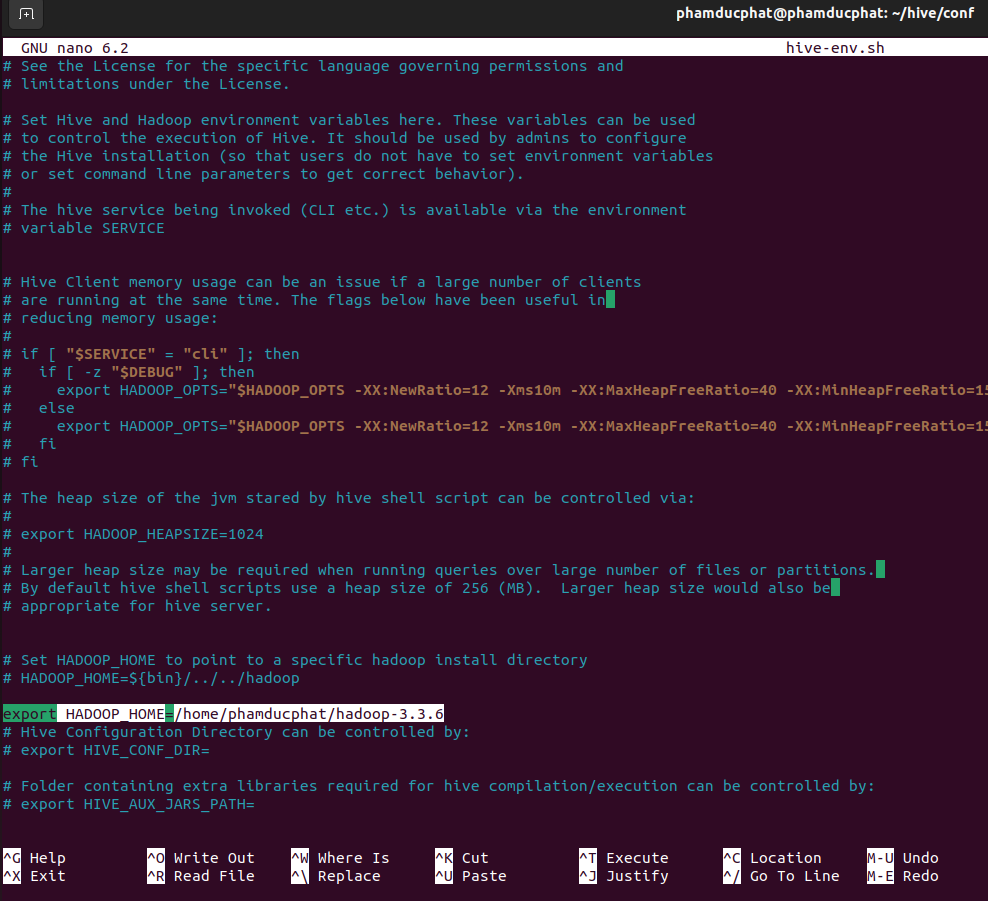
cp hive-env.sh.template hive-env.sh

nano hive-env.sh



Thêm đường dẫn

HADOOP\_HOME=/home/phamducphat/hadoop



Lưu (Ctrl S) và thoát (Ctrl X)

**Bước 4:** Tạo tệp cấu hình Hive

Tới vị trí cấu hình Hive

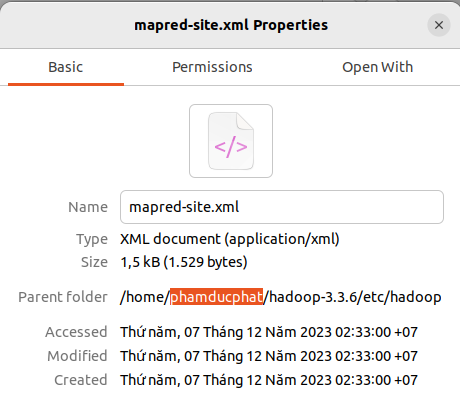
cd /home/phamducphat/hive/conf

Tạo tệp hive-site.xml

cp hive-default.xml.template hive-site.xml



**Bước 5:** Bổ sung tham số cho tệp mapred-site.xml



Chèn đoạn mã sau

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

<property>

<name>yarn.app.mapreduce.am.env</name> <value>HADOOP\_MAPRED\_HOME=${HADOOP\_HOME}</value>

</property>

<property>

<name>mapreduce.map.env</name> <value>HADOOP\_MAPRED\_HOME=${HADOOP\_HOME}</value>

</property>

<property>

<name>mapreduce.reduce.env</name> <value>HADOOP\_MAPRED\_HOME=${HADOOP\_HOME}</value>

</property>

<property>

<name>mapreduce.map.memory.mb</name>

<value>2048</value>

</property>

<property>

<name>mapreduce.reduce.memory.mb</name>

<value>4096</value>

</property>

<property>

<name>mapreduce.map.java.opts</name>

<value>-Xmx1638m</value>

</property>

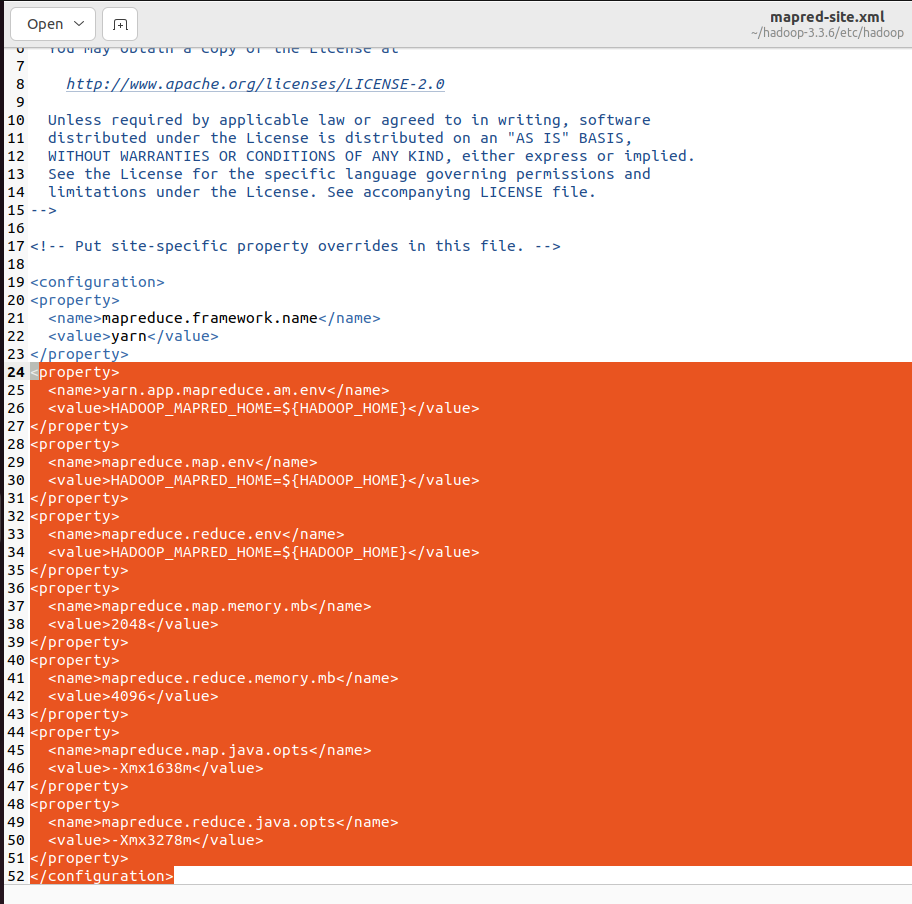
<property>

<name>mapreduce.reduce.java.opts</name>

<value>-Xmx3278m</value>

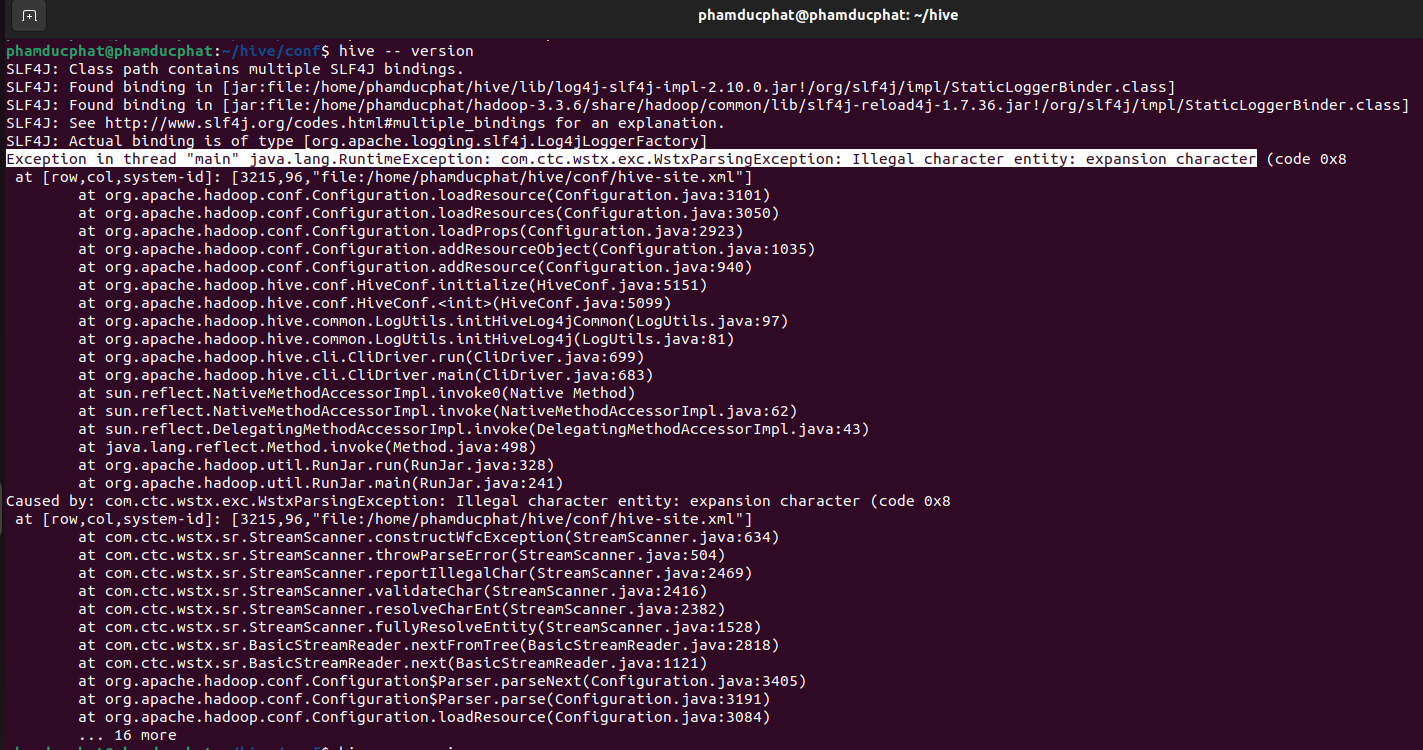
</property>

Lưu



**Bước 6:** Kiểm tra Hive

hive – version



Gặp lỗi với ngoại lệ

Exception in thread "main" java.lang.RuntimeException: com.ctc.wstx.exc.WstxParsingException: Illegal character entity: expansion character

Mở tệp hive-site.xml bằng trình chỉnh sửa nano

cd /home/phamducphat/hive/conf

nano hive-site.xml

Không có mô tả.

Tìm (Ctrl W) “Ensures commands with OVERWRITE” và xóa “&#8” ở cùng dòng

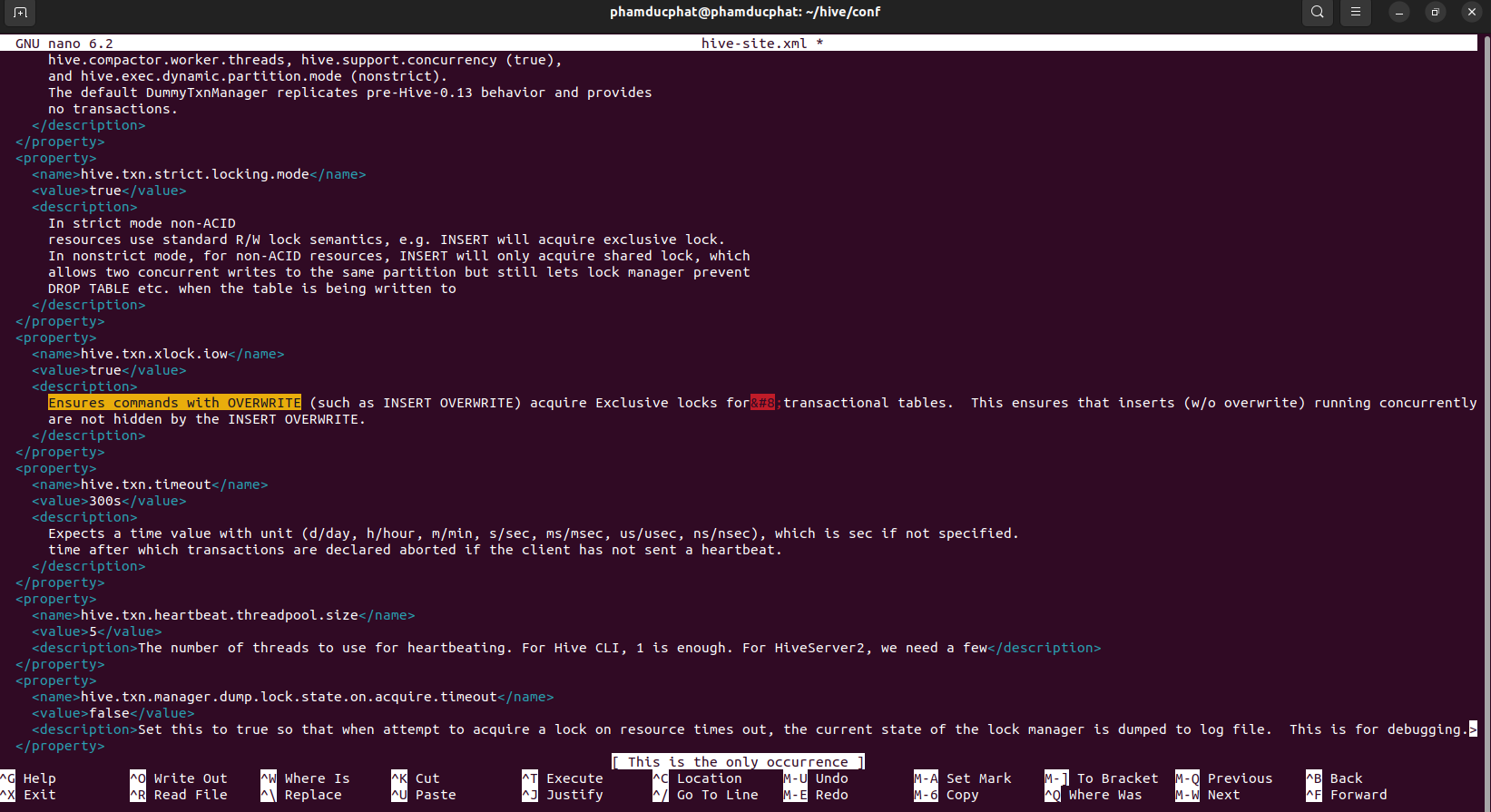
<description>

Ensures commands with OVERWRITE (such as INSERT OVERWRITE) acquire Exclusive locks for&#8;

transactional tables. This ensures that inserts (w/o overwrite) running concurrently

are not hidden by the INSERT OVERWRITE.

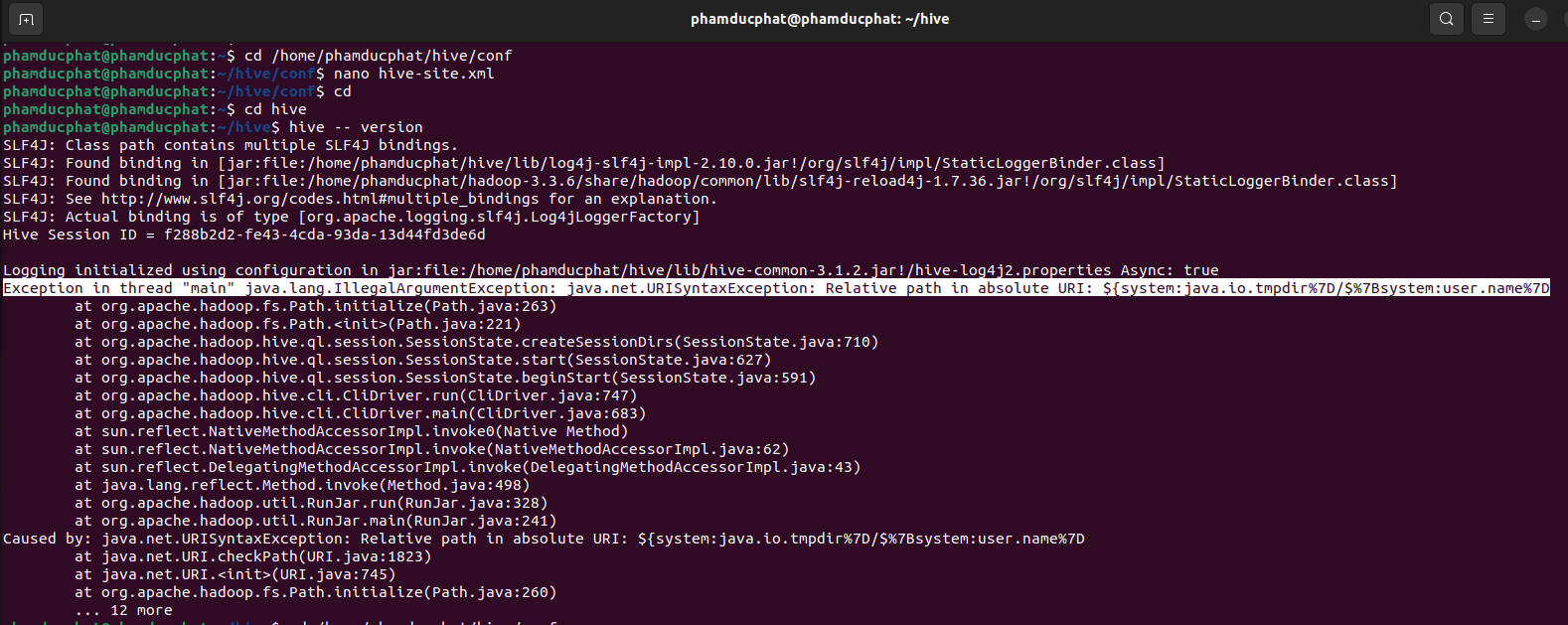
</description>



Lưu (Ctrl S)

**Bước 7:** Kiểm tra Hive lần 2

hive –- version



Gặp lỗi ngoại lệ khác

Exception in thread "main" java.lang.IllegalArgumentException: java.net.URISyntaxException: Relative path in absolute URI: ${system:java.io.tmpdir/Bsystem:user.name%7D

Mở tệp hive-site.xml bằng trình chỉnh sửa nano

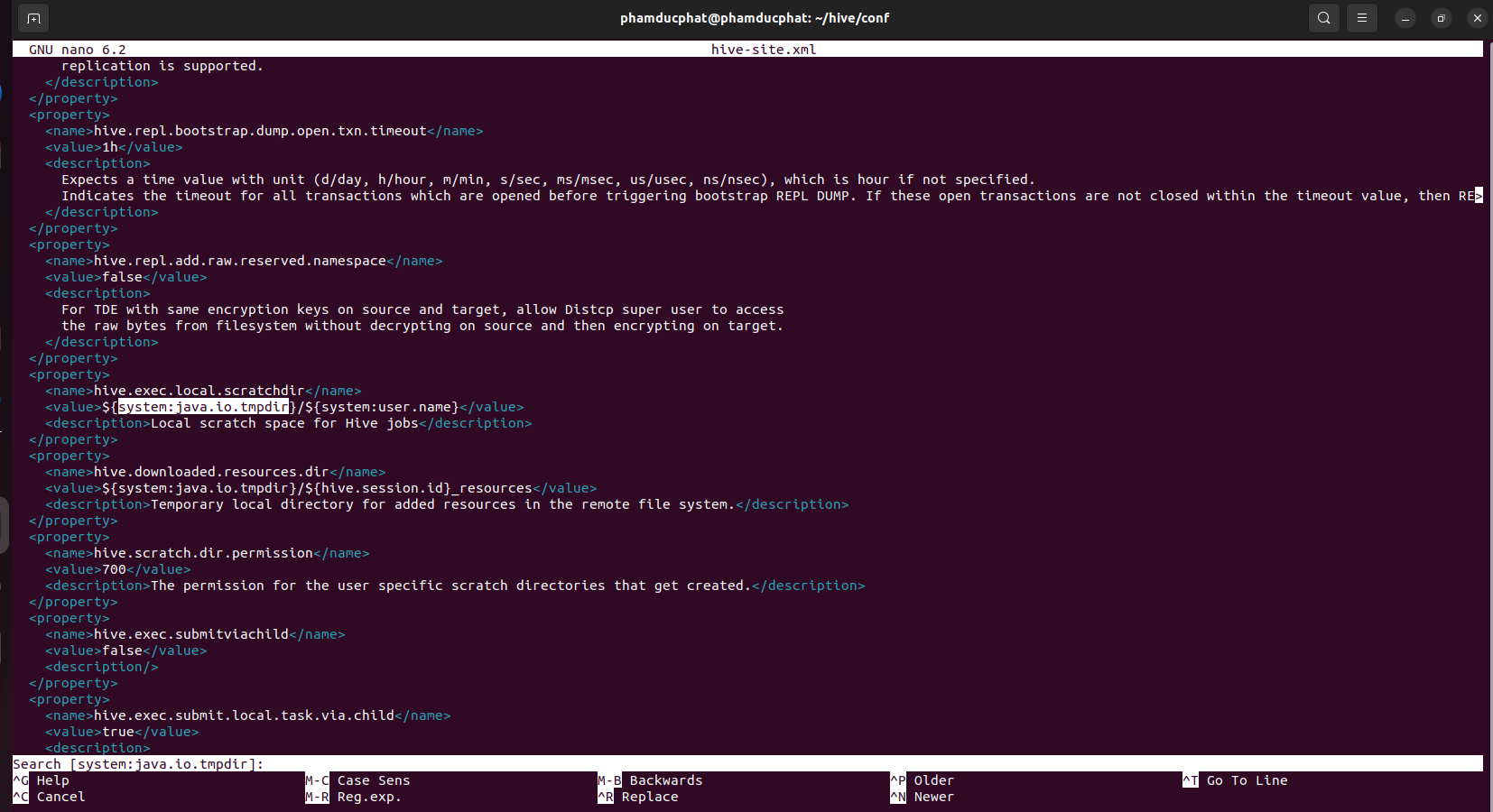
cd /home/phamducphat/hive/conf

nano hive-site.xml

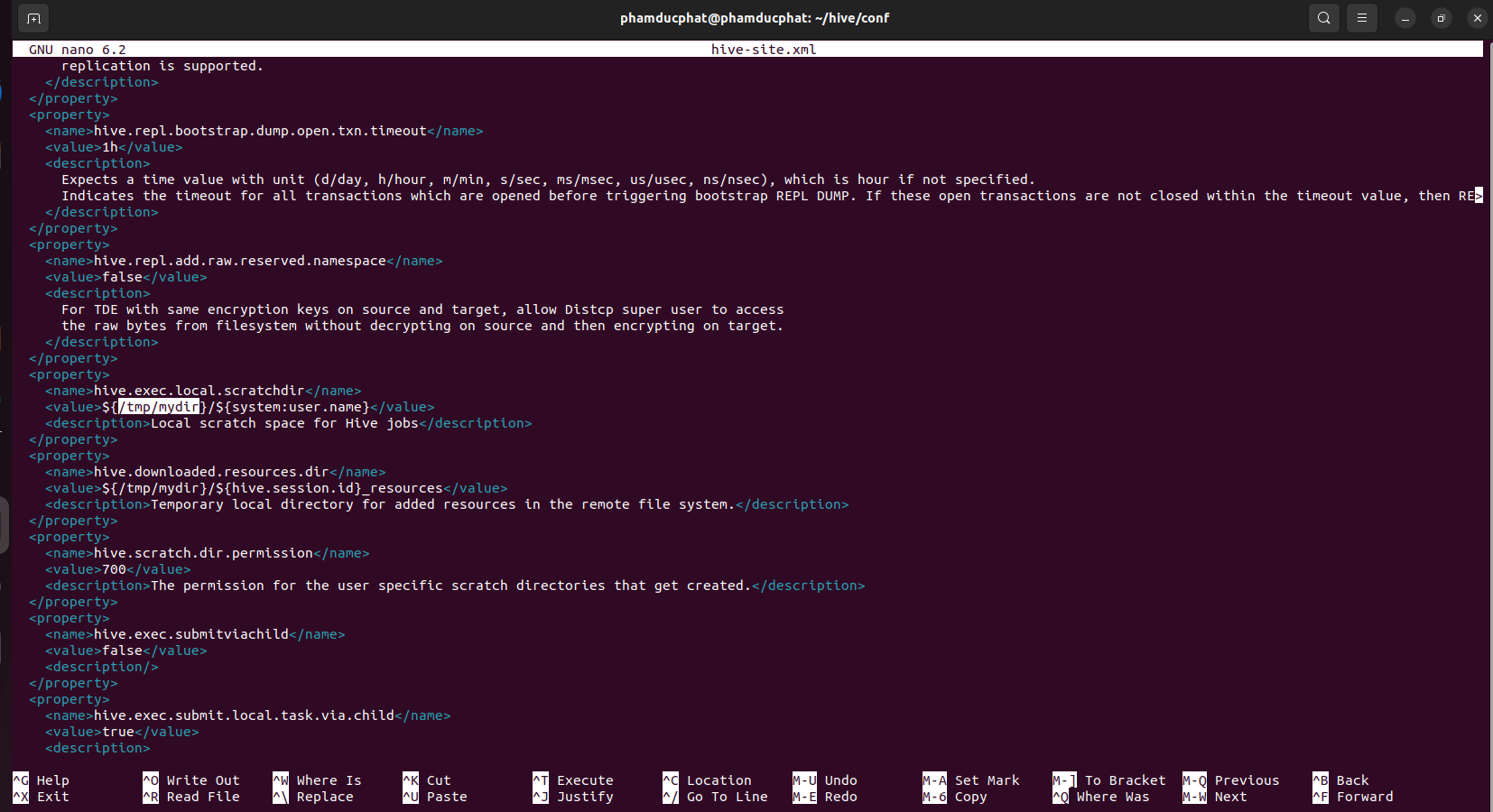
Không có mô tả.

Tìm (Ctrl W) “system:java.io.tmpdir” nhiều lần (4) và thay thế bằng “/tmp/mydir””

Trước:



Sau:

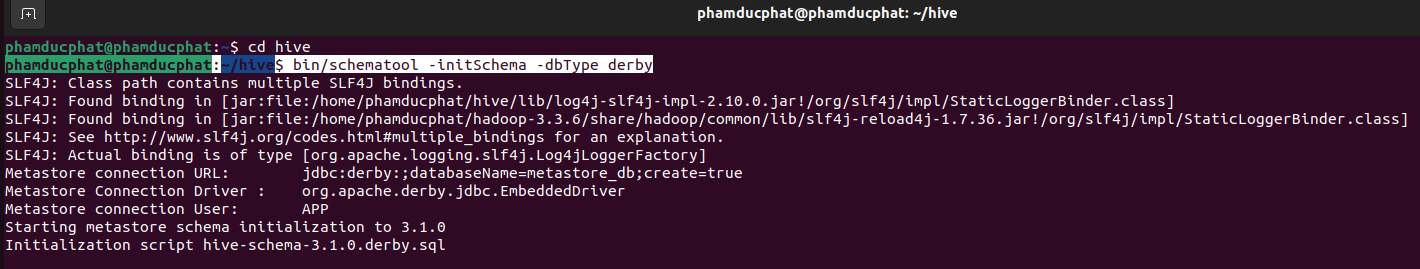


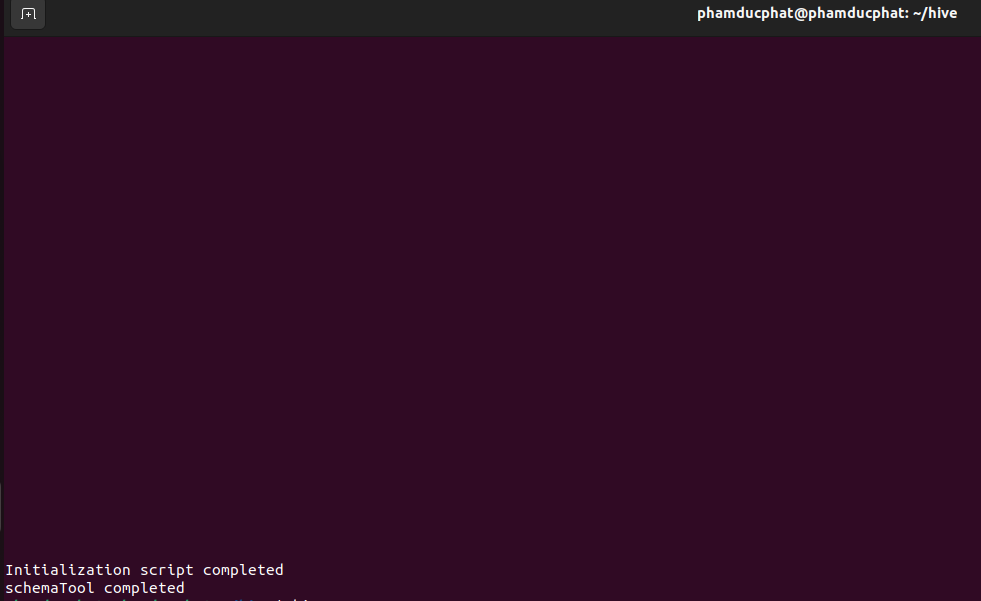
**Bước 8:** Khở tạo CSDL mặc định (Derby) cho Apache Hive

Tạo tại /home/phamducphat/hive

cd hive

bin/schematool -initSchema -dbType derby





**Bước 9: Khởi động Hive shell**

hive

