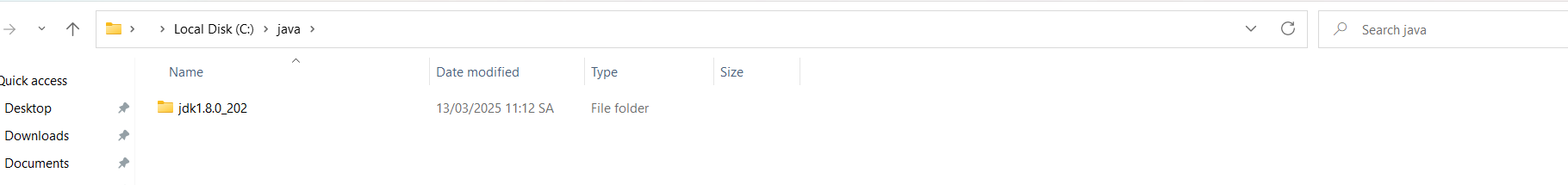
**CÀI ĐẶT HBASE TRÊN WINDOW**

1. **Cài đặt JDK bản 1.8 (bắt buộc)**

Link tải: [Java Archive Downloads - Java SE 8](https://www.oracle.com/java/technologies/javase/javase8-archive-downloads.html)



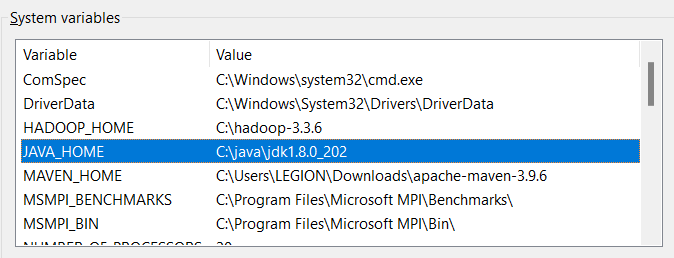
* Lưu vào ổ đĩa C



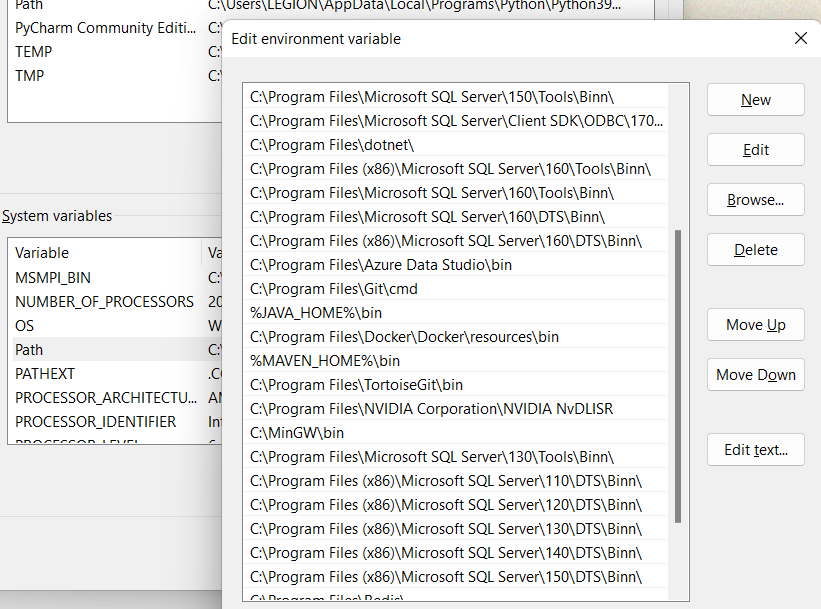
* Cài đặt thành công

1. **Thiết lập biến môi trường cho Java jdk**

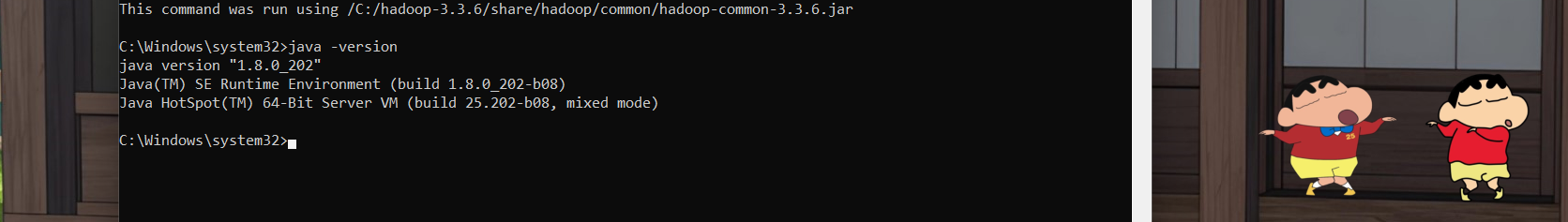
* **CẤU HÌNH JAVA\_HOME trong phần cài đặt biến môi trường**



* **CẤU HÌNH PATH**



* Thiết lập thành công

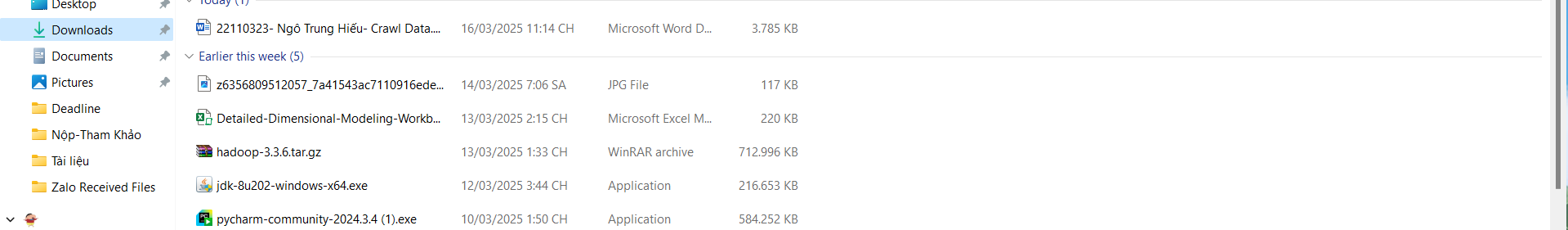


* Kiểm tra phiên bản Java đã cài đặt

1. **Tải Hadoop và cài đặt**

Link tải: [Apache Hadoop](https://hadoop.apache.org/releases.html)

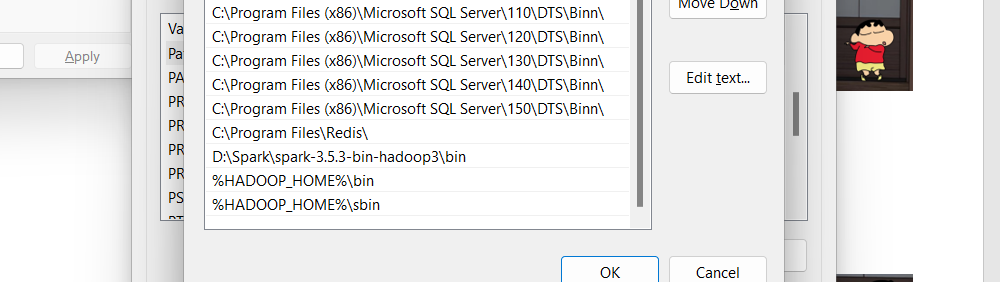
* Tải về và lưu vào ổ C cho dễ quản lý và sử dụng



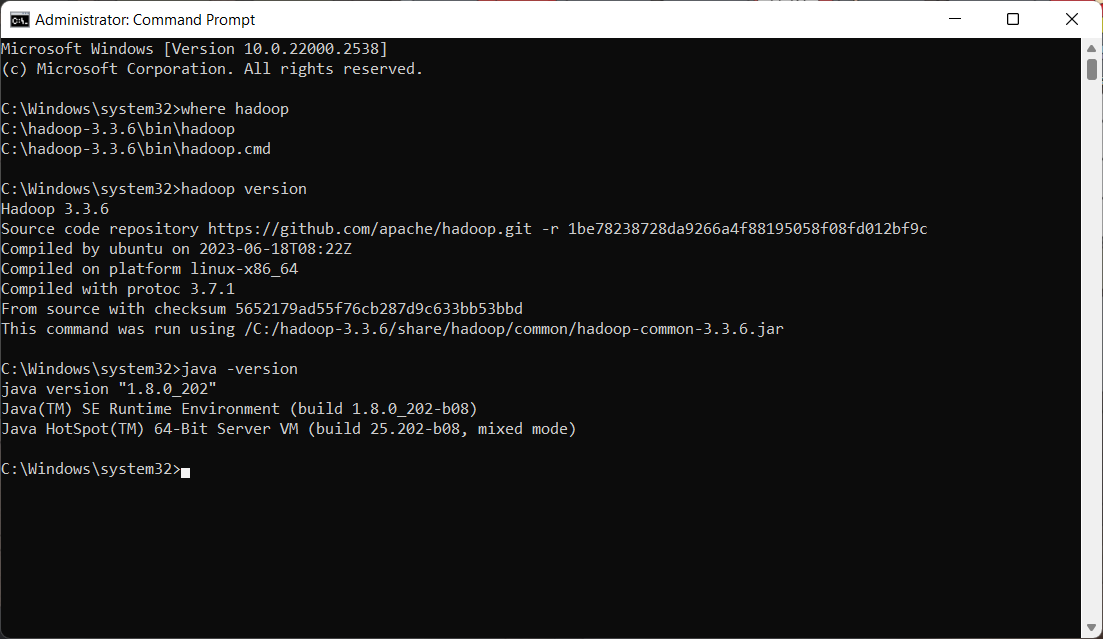
* **CẤU HÌNH BIẾN MÔI TRƯỜNG CHO HADOOP**



* **CẤU HÌNH PATH**



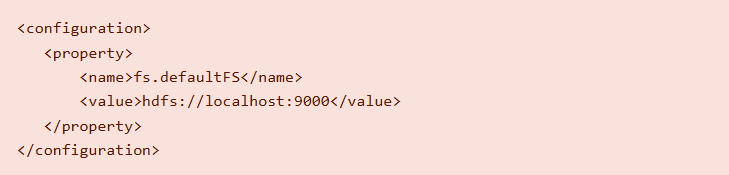
* **Mở cmd và test lại**



1. **Cấu hình các tập tin cấu hình Hadoop**

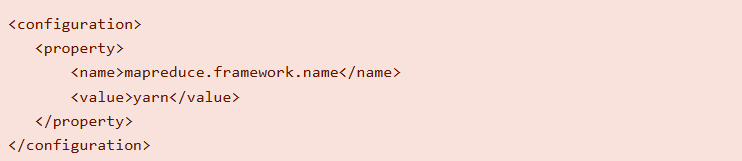
Trong thư mục **C:/hadoop-3.3.6/etc/hadoop** lần lượt chỉnh sửa các file:

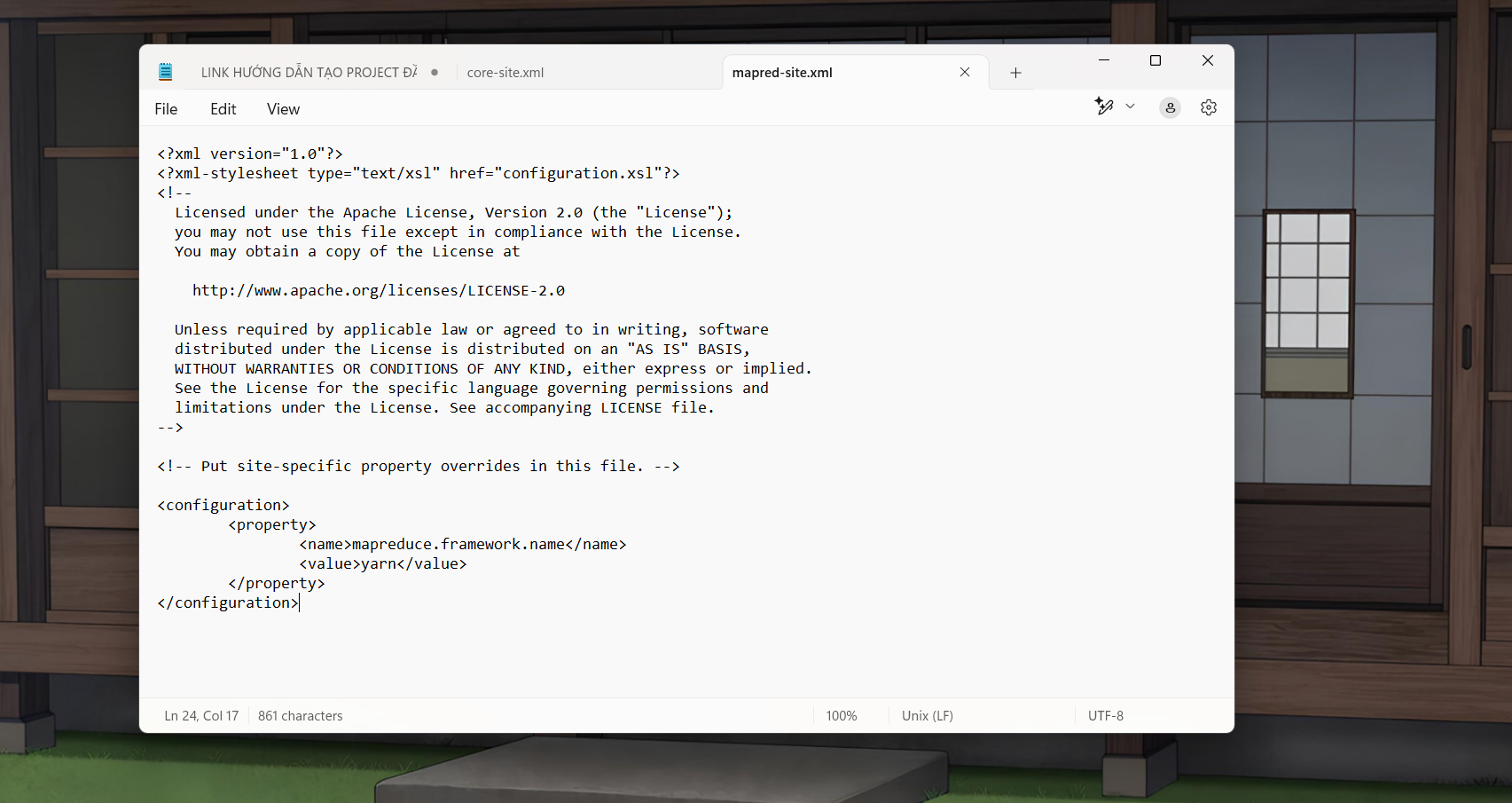
* core-site.xml
* mapred-site.xml
* hdfs-site.xml
* yarn-site.xml
* hadoop-env.cmd
* **Chỉnh cấu hình core-site.xml**

****

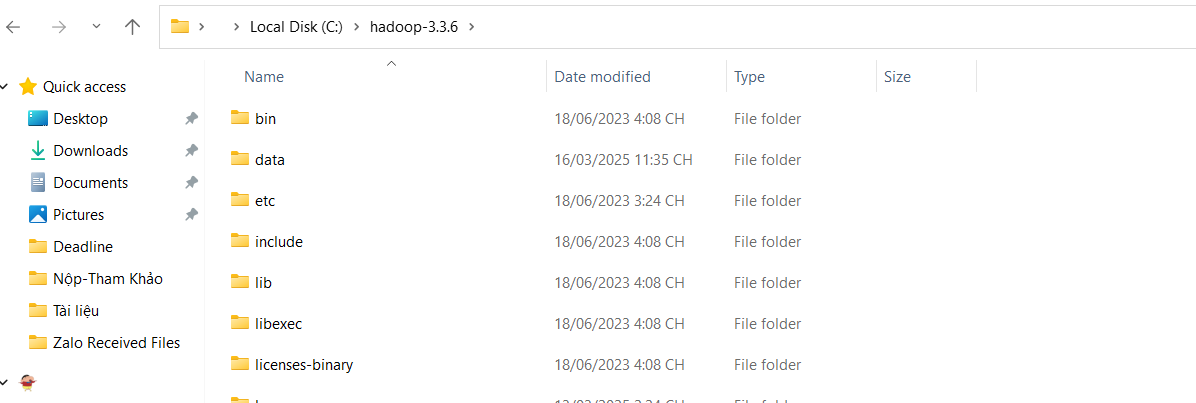
****

* **Chỉnh cấu hình mapred-site.xml**

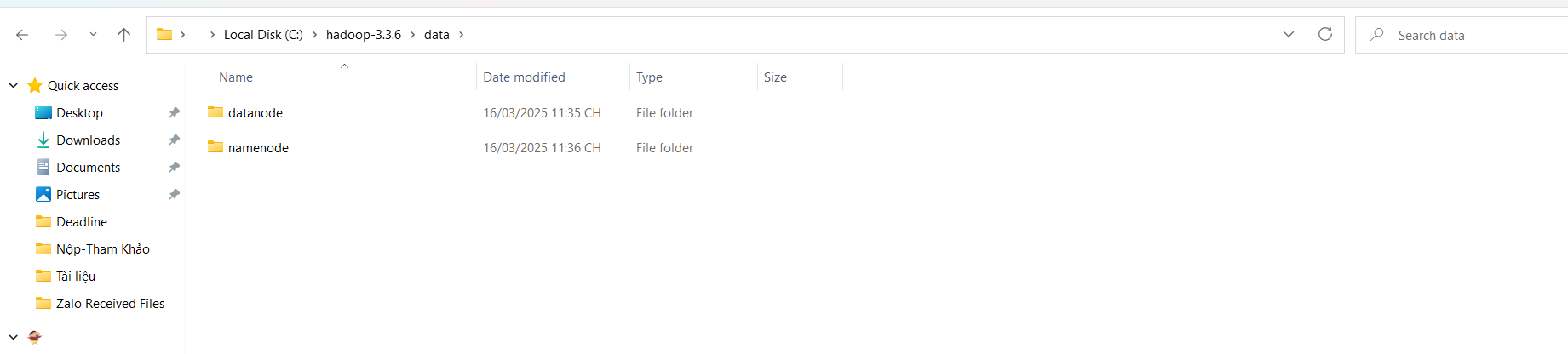
****

****

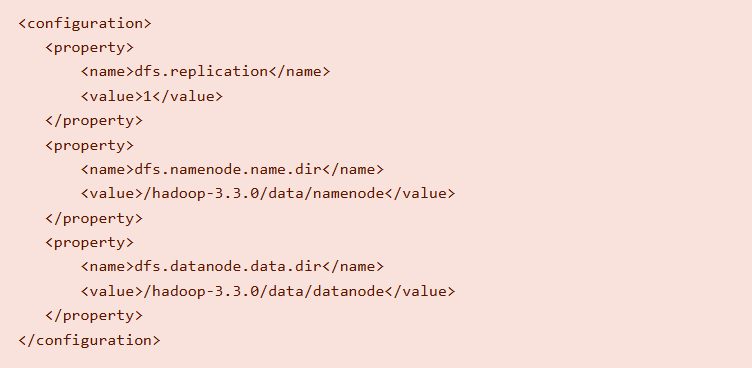
* **Chỉnh cấu hình hdfs-site.xml**
* Tạo thư mục  “data” trong “C:/hadoop-3.3.6”
* Tạo thư mục con “datanode” trong “C:/hadoop-3.3.6/data”
* Tạo thư mục con “namenode” trong “C:/hadoop-3.3.6/data”

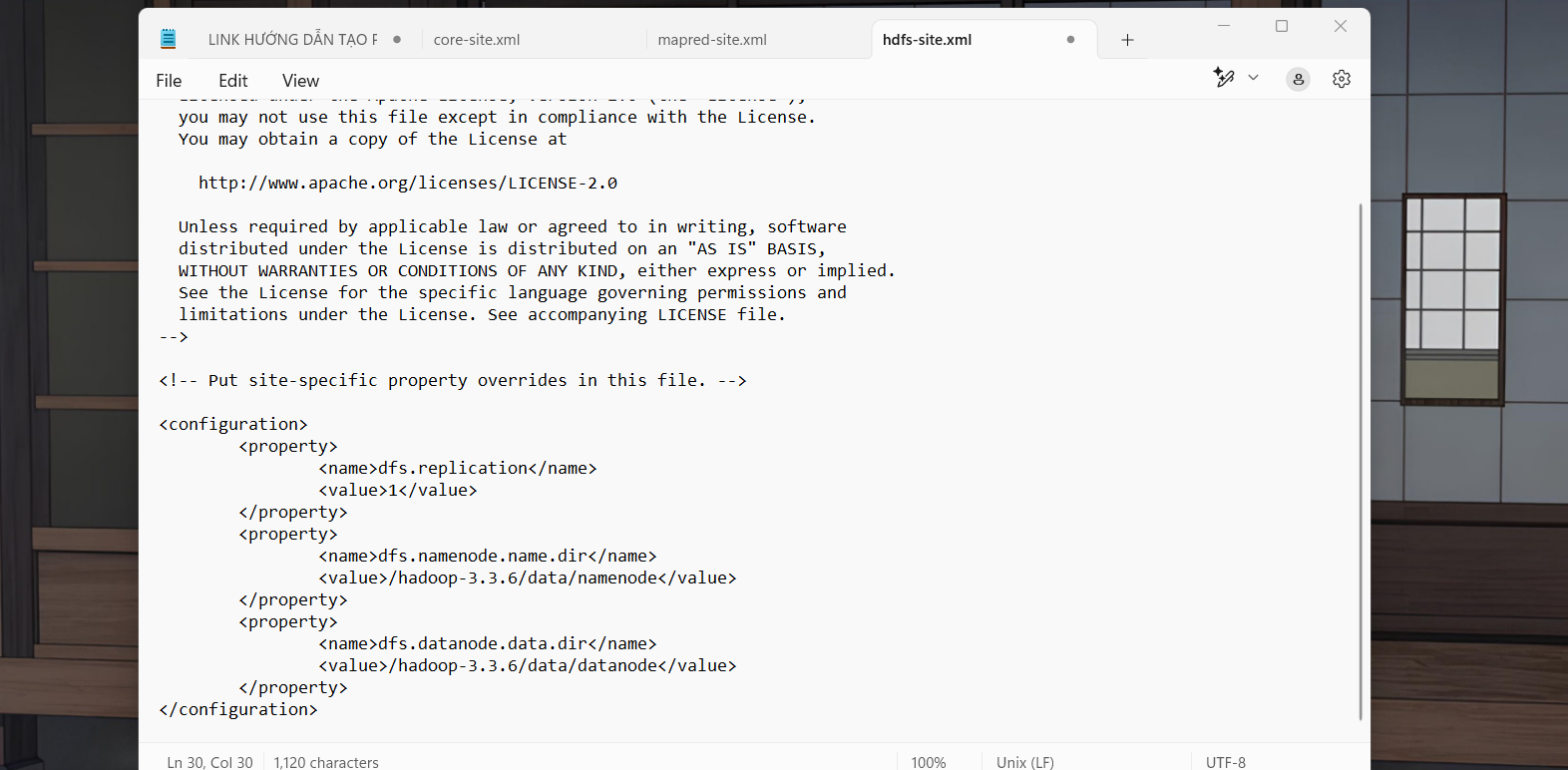
****



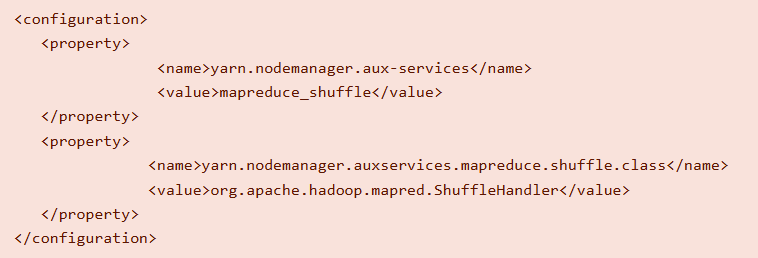
****

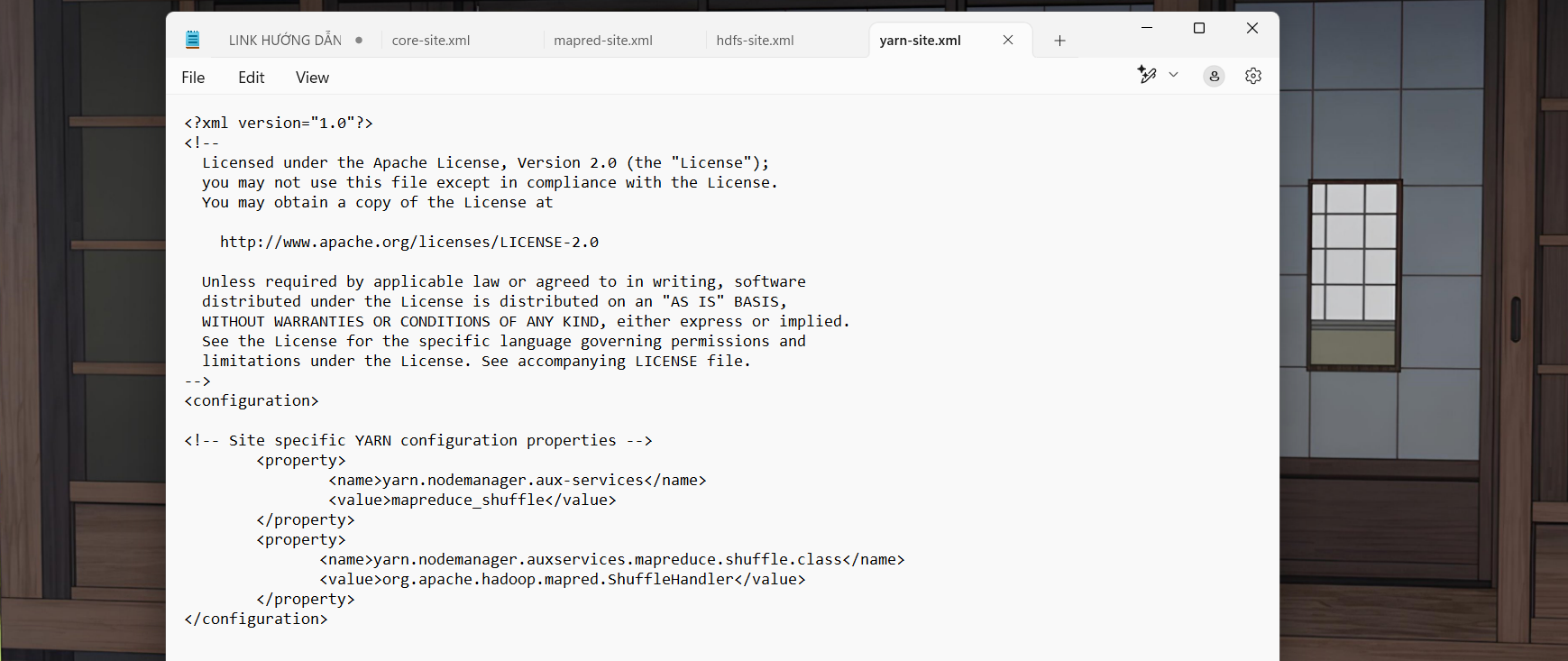


****

****

* **Chỉnh cấu hình yarn-site.xml**

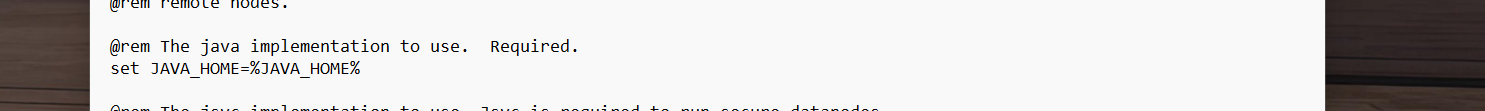
****

****

* **Chỉnh cấu hình hadoop-env.cmd**

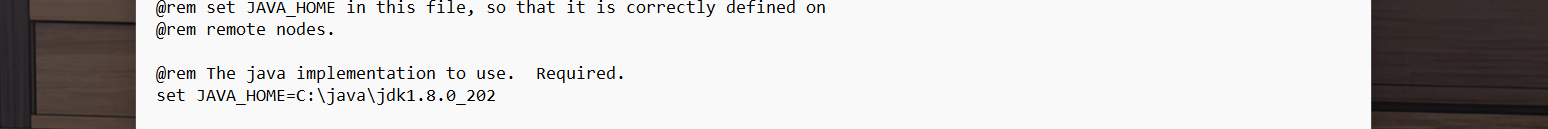
Mở file này lên và tìm tới lệnh:

set JAVA\_HOME=%JAVA\_HOME%



Sửa %JAVA\_HOME% thành đường dẫn cài JDK trong ổ C:

set JAVA\_HOME= C:/java/jdk1.8.0\_202



1. **Cập nhật các Hadoop Configurations**

Link tải: <https://github.com/s911415/apache-hadoop-3.1.0-winutils>

  
Tải về giải nén ra thấy thư mục bin ở bên trong

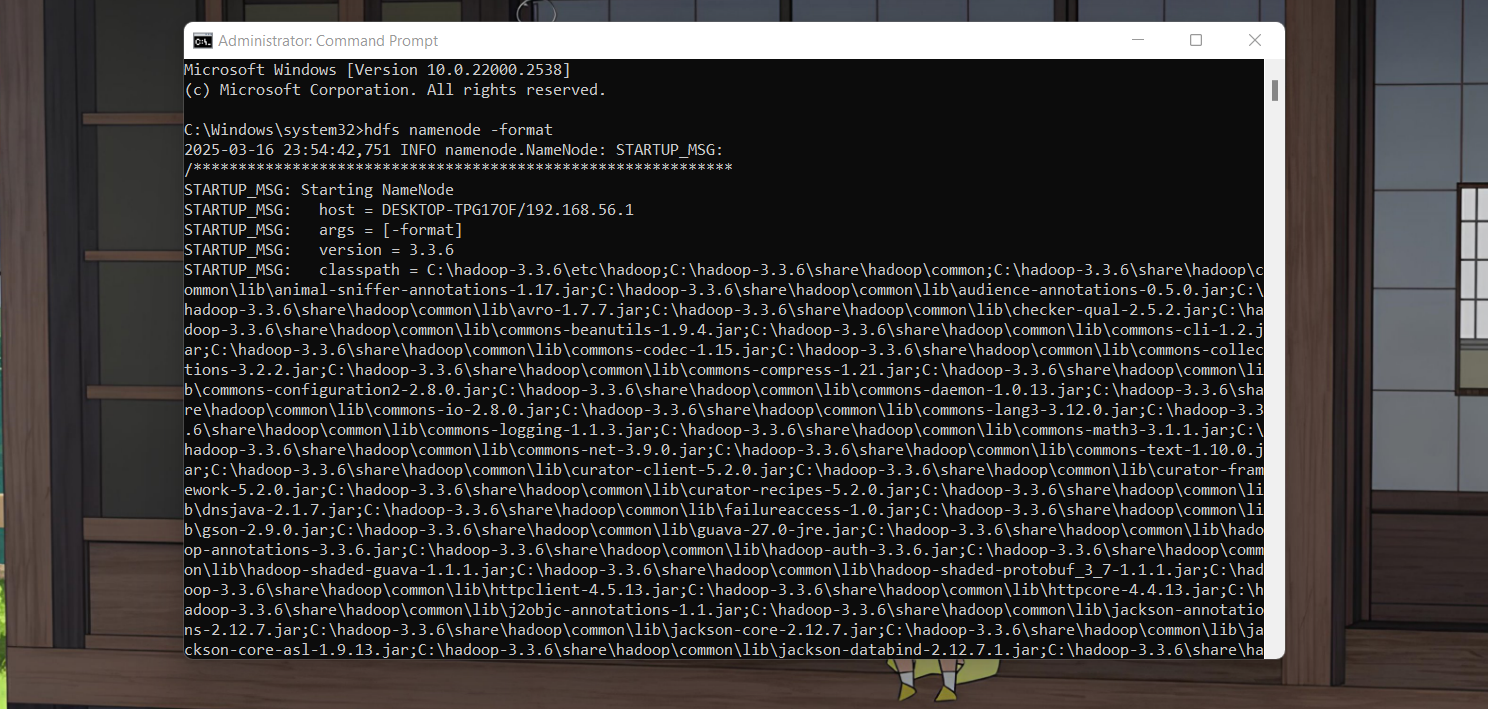


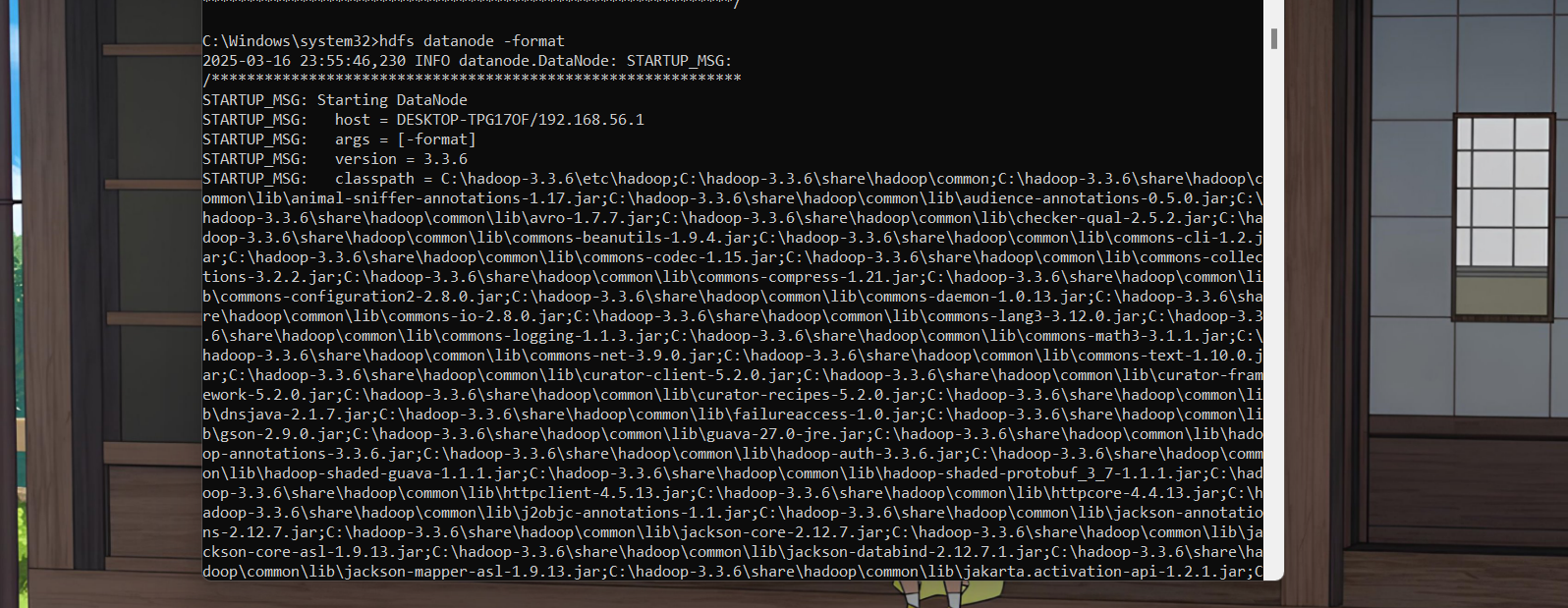
Chép đè thư mục bin này trong thư mục bin của C:\hadoop-3.3.6\bin

Sau đó format lại namenode và datanode: mở command line lên, gõ lệnh sau:

$ hdfs namenode –format

$ hdfs datanode -format

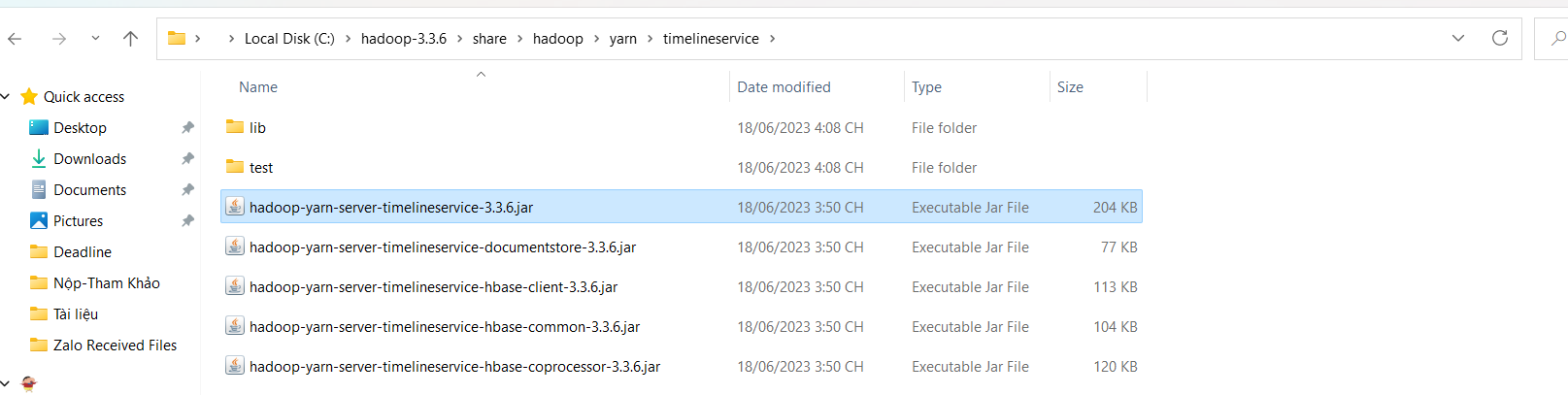


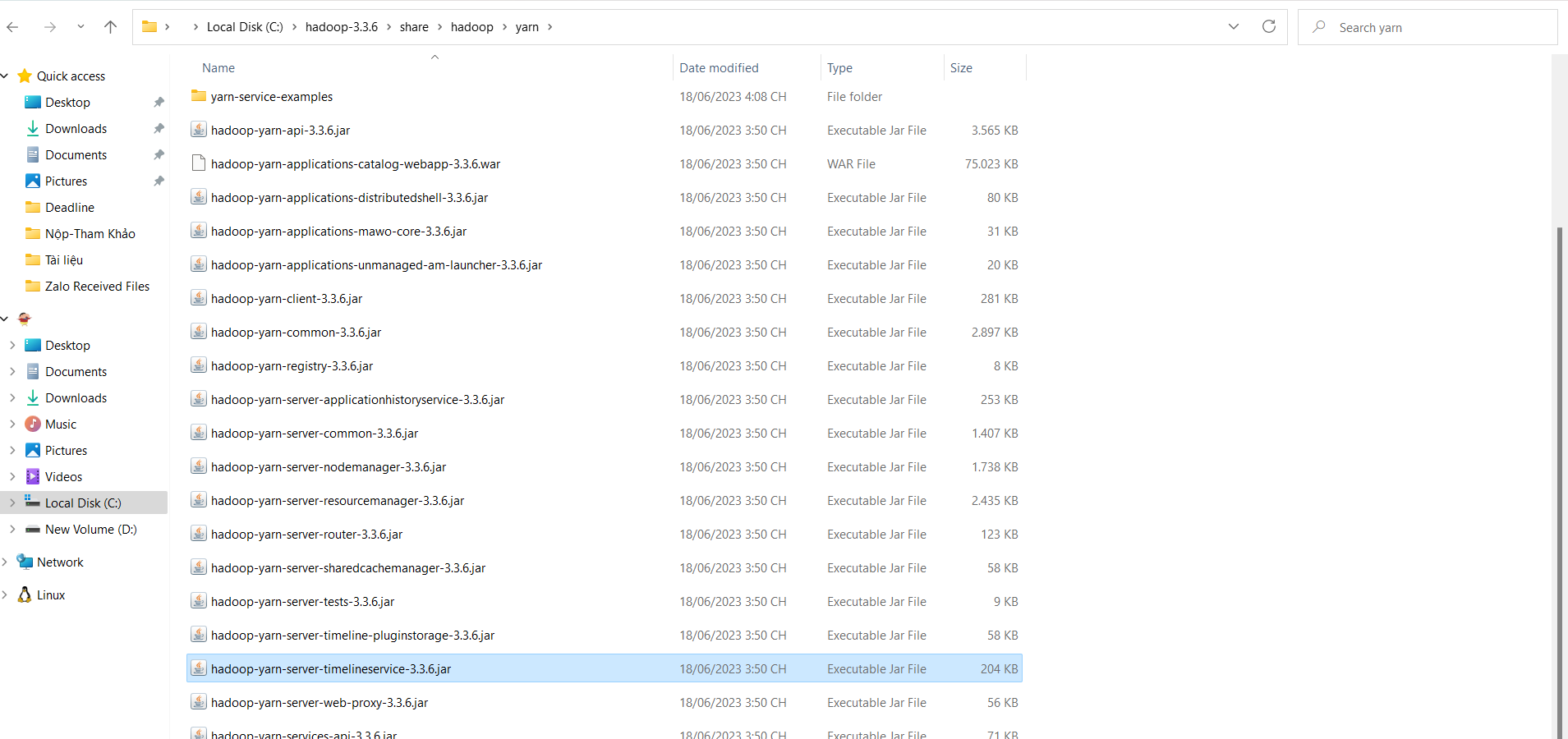


Tiếp theo sao chép file:

“C:/hadoop-3.3.6/share/hadoop/yarn/timelineservice/ hadoop-yarn-server-timelineservice-3.3.6.jar”

vào “C:/hadoop-3.3.6/share/hadoop/yarn/hadoop-yarn-server-timelineservice-3.3.6.jar”

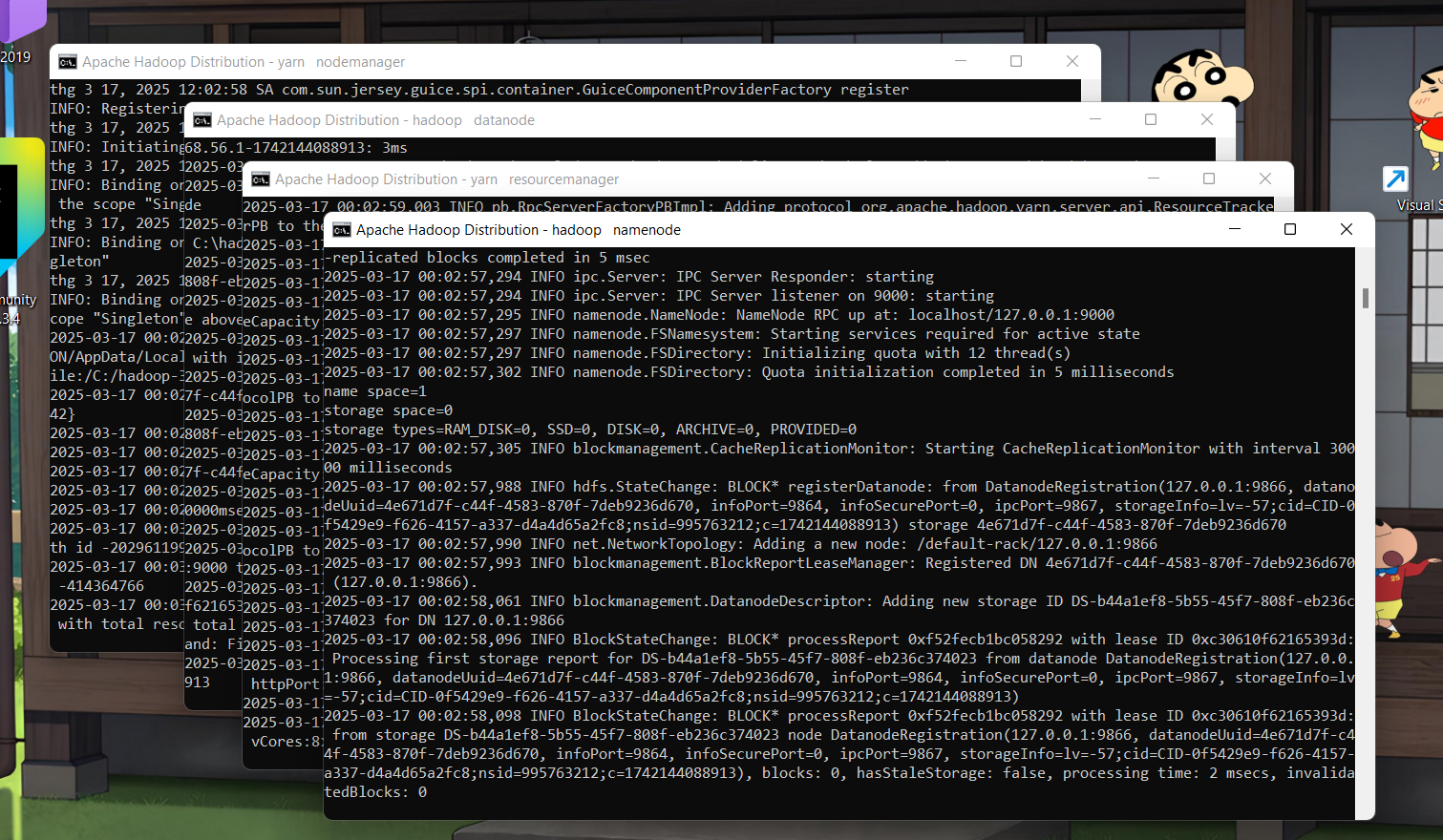


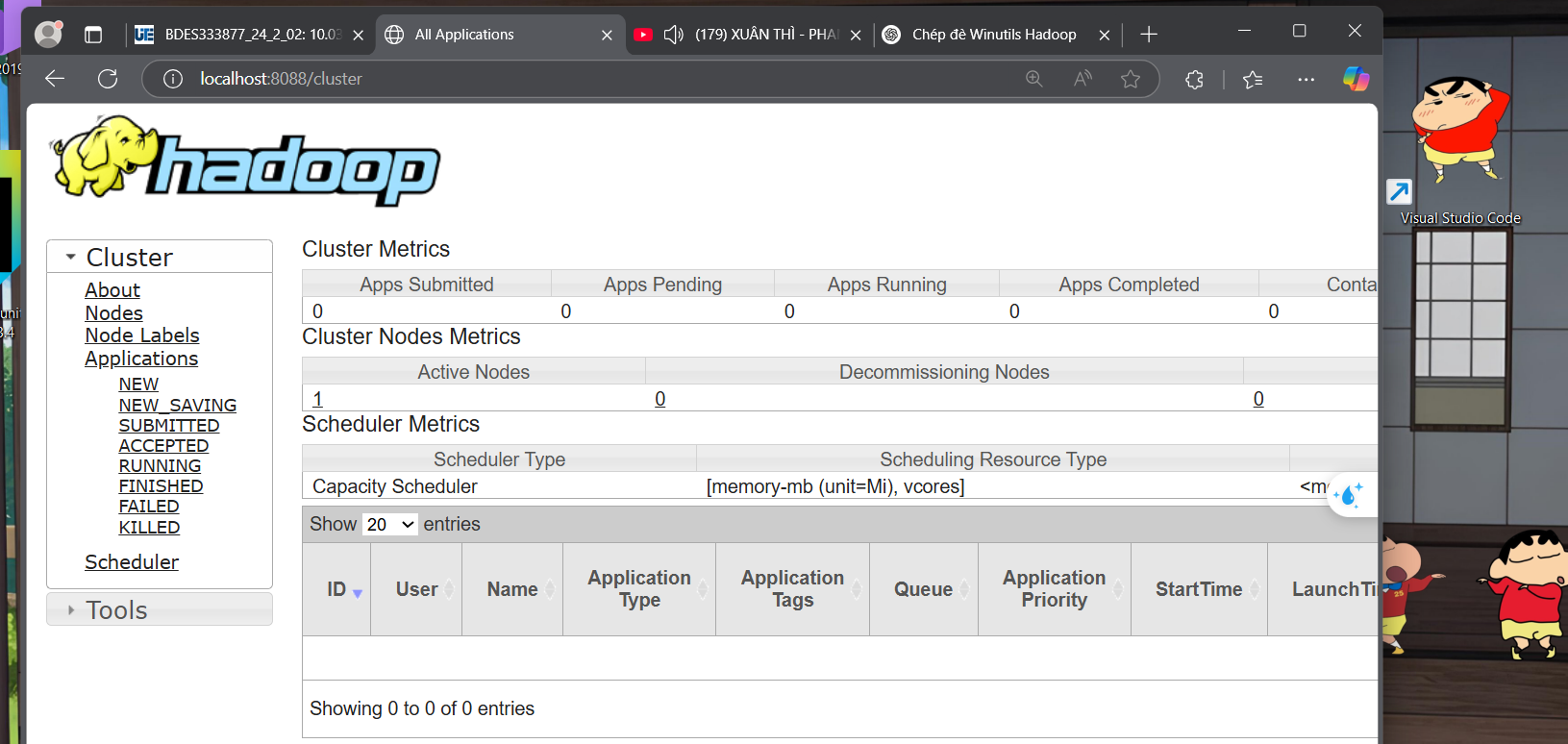


1. **Hoàn thành cài đặt Hadoop và test thử nghiệm với start-all.cmd**

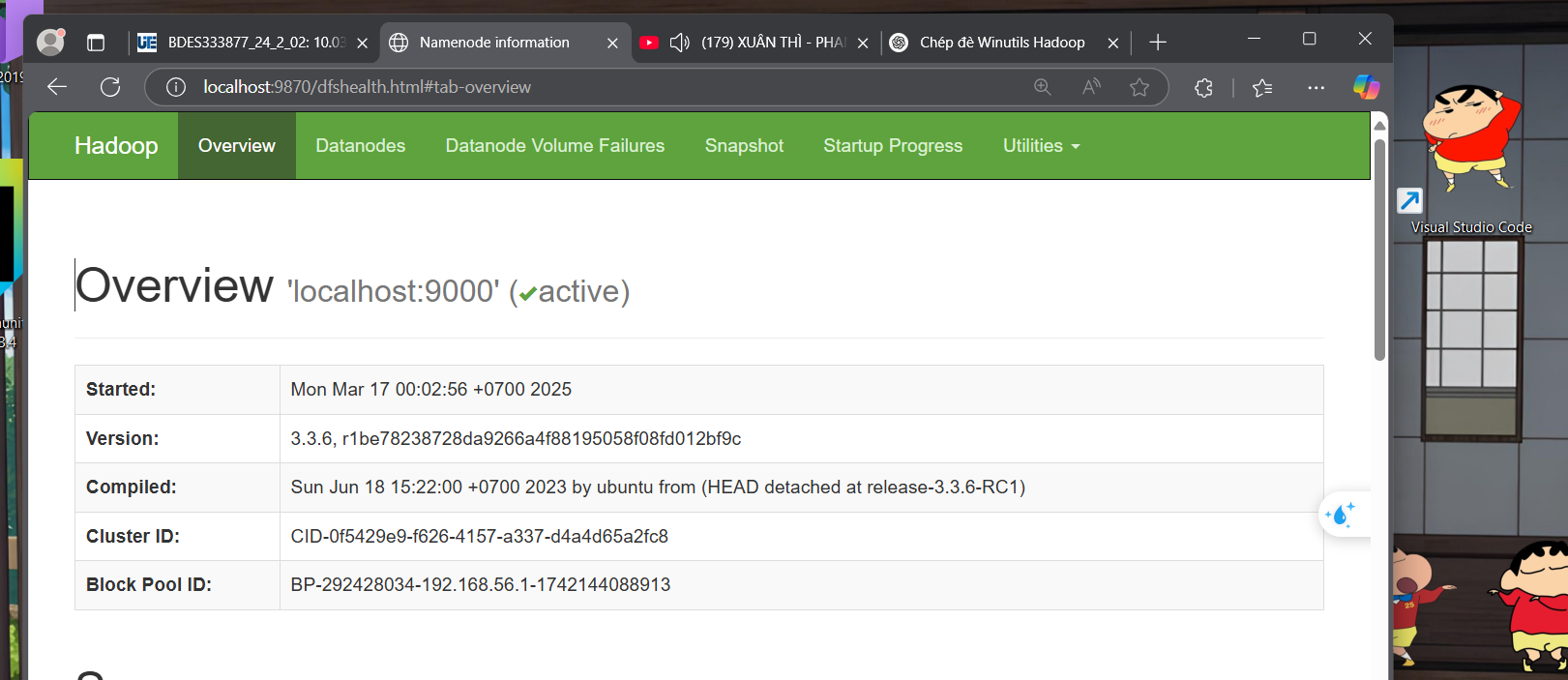
Để test Hadoop, ta mở command line và di chuyển tới thư mục C:/hadoop-3.3.0/sbin

Sau đó gõ lệnh: $ start-all.cmd









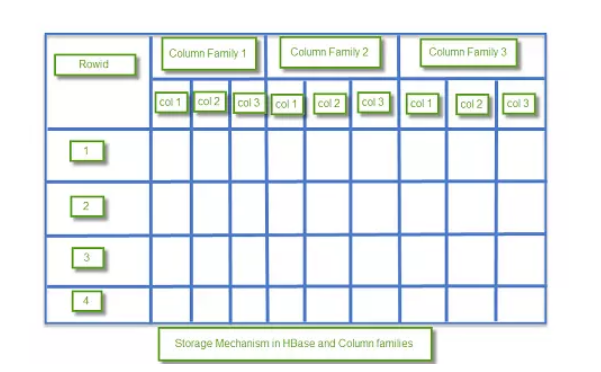




* Test Hadoop thành công

1. **Cài đặt Hbase**

* **Cơ chế lưu trữ trong Hbase**

****HBase là một cơ sở dữ liệu theo cột và dữ liệu được lưu trữ trong bảng. Các bảng được sắp xếp bởi RowId. Như được trình bày dưới đây, HBase có RowId, collect một vài Column Family có trong bảng:

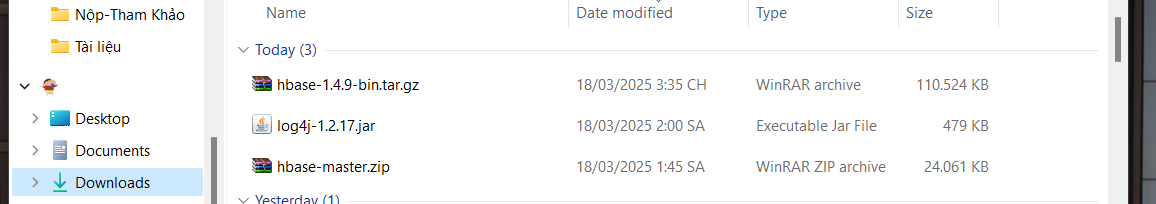
Các Column Family có trong lược đồ là các cặp key-value. Column Family có nhiều cột (col). Giá trị col được lưu trữ trong bộ nhớ đĩa. Mỗi ô của bảng có dữ liệu Meta riêng như time stamp và các thông tin khác. Các thuật ngữ:

* **Table**: Collect tất cả các row
* **Row**: Collect tất cả các column family
* **Column** **Family**: Collect tất cả các cột (col)
* **Column**: Collect các cặp key-value
* **Namespace**: Logical grouping of tables.
* **Cell**: Bộ {row, column, version} xác định một cell trong Hbase
* **Tầm quan trọng của cơ sở dữ liệu NoSQL trong Hadoop**

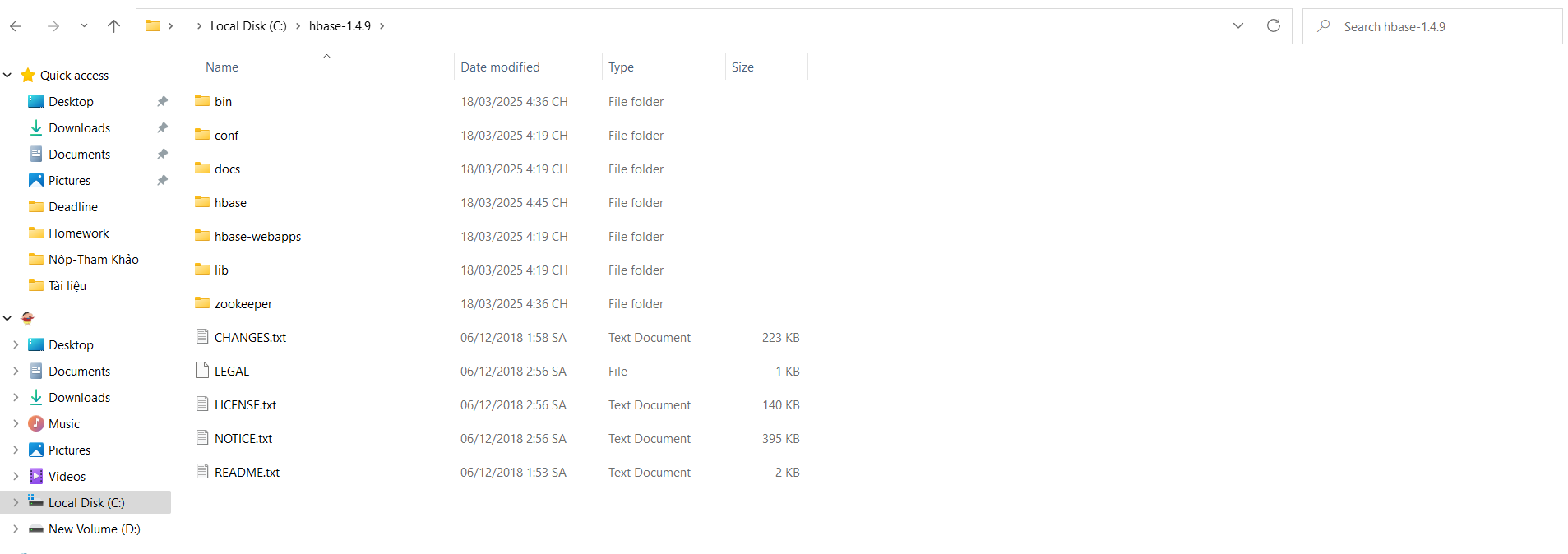
Trong phân tích dữ liệu lớn, Hadoop đóng một vai trò quan trọng trong việc giải quyết các vấn đề business điển hình bằng cách quản lý bộ dữ liệu lớn và đưa ra các giải pháp tốt nhất. Trong hệ sinh thái Hadoop, mỗi thành phần đóng vai trò đặc biệt trong việc: Xử lí dữ liệu, xác nhận dữ liệu, lưu trữ dữ liệu

Link tham khảo Hbase: [HBase: Overview, Architecture và Data Flow](https://viblo.asia/p/hbase-overview-architecture-va-data-flow-63vKj6J6K2R)

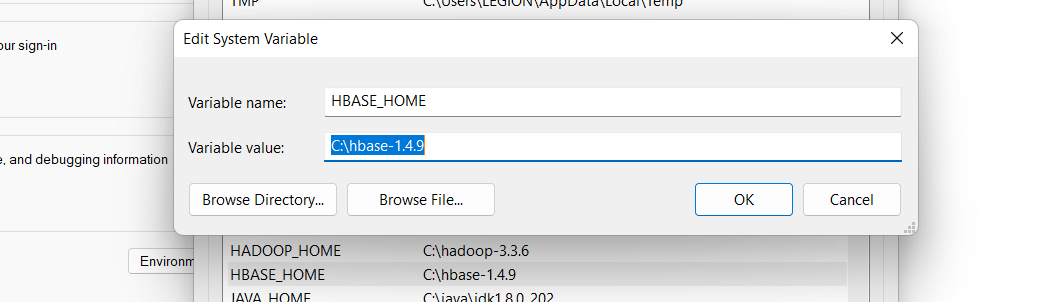
**Tải Hbase:** [Index of /hbase/stable](https://downloads.apache.org/hbase/stable/)



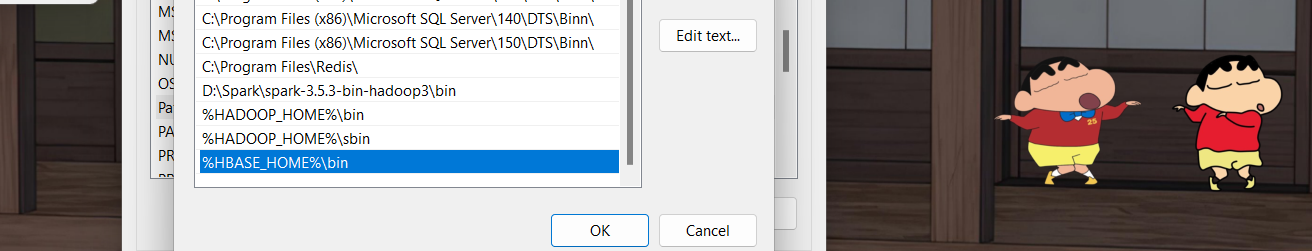
* Tải và giải nén đưa vào ổ C



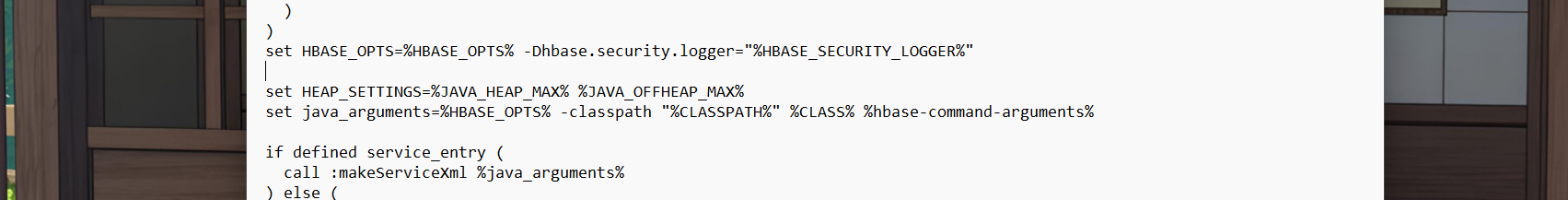
1. **Cập nhật biến môi trường cho Hbase**

****

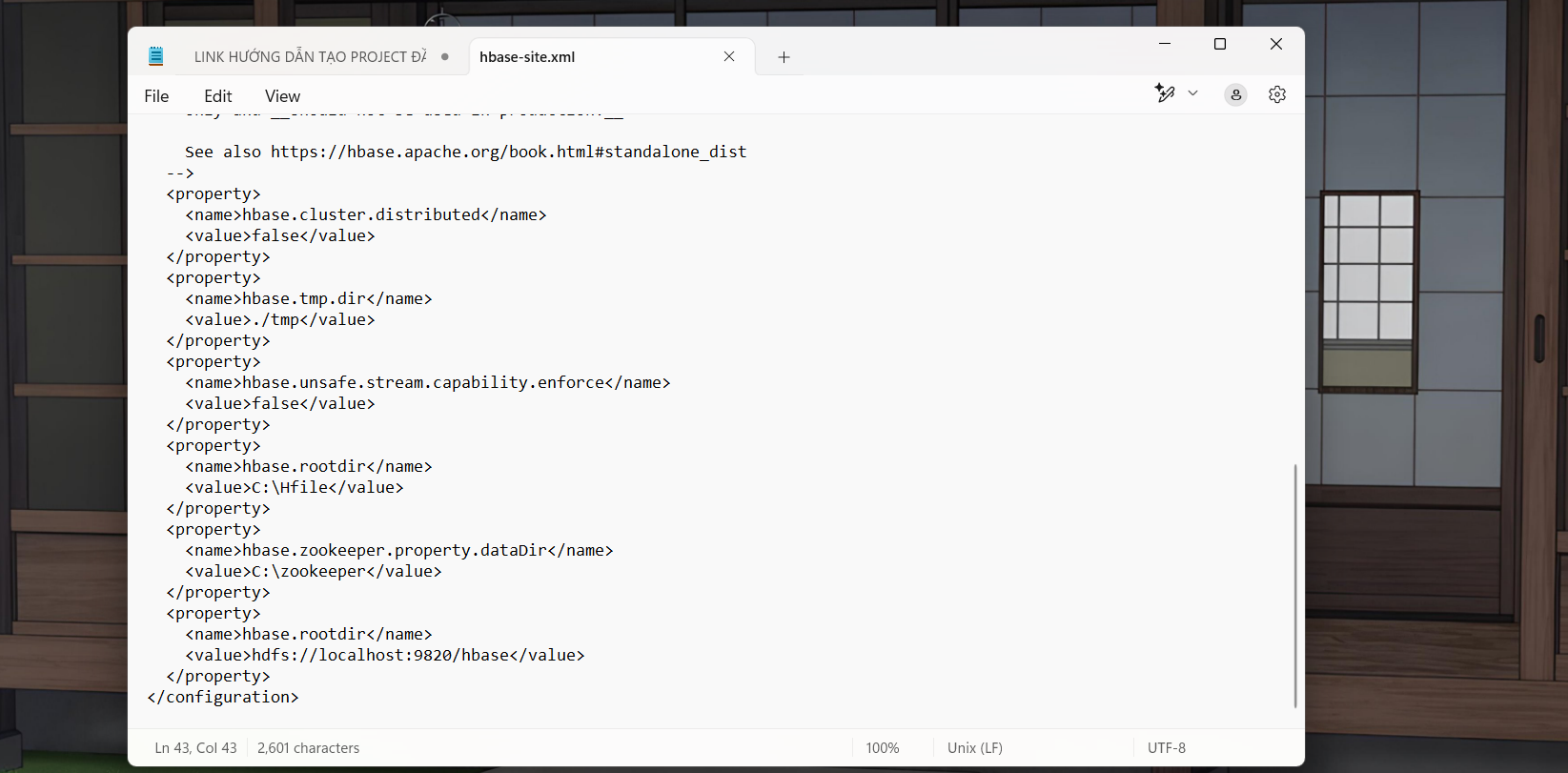
* **Thêm và Path**

****

Chỉnh sửa tệp **hbase-config.cmd**, nằm trong thư mục bin của thư mục đã giải nén, và thêm dòng sau để thiết lập JAVA\_HOME (mở comment ra)



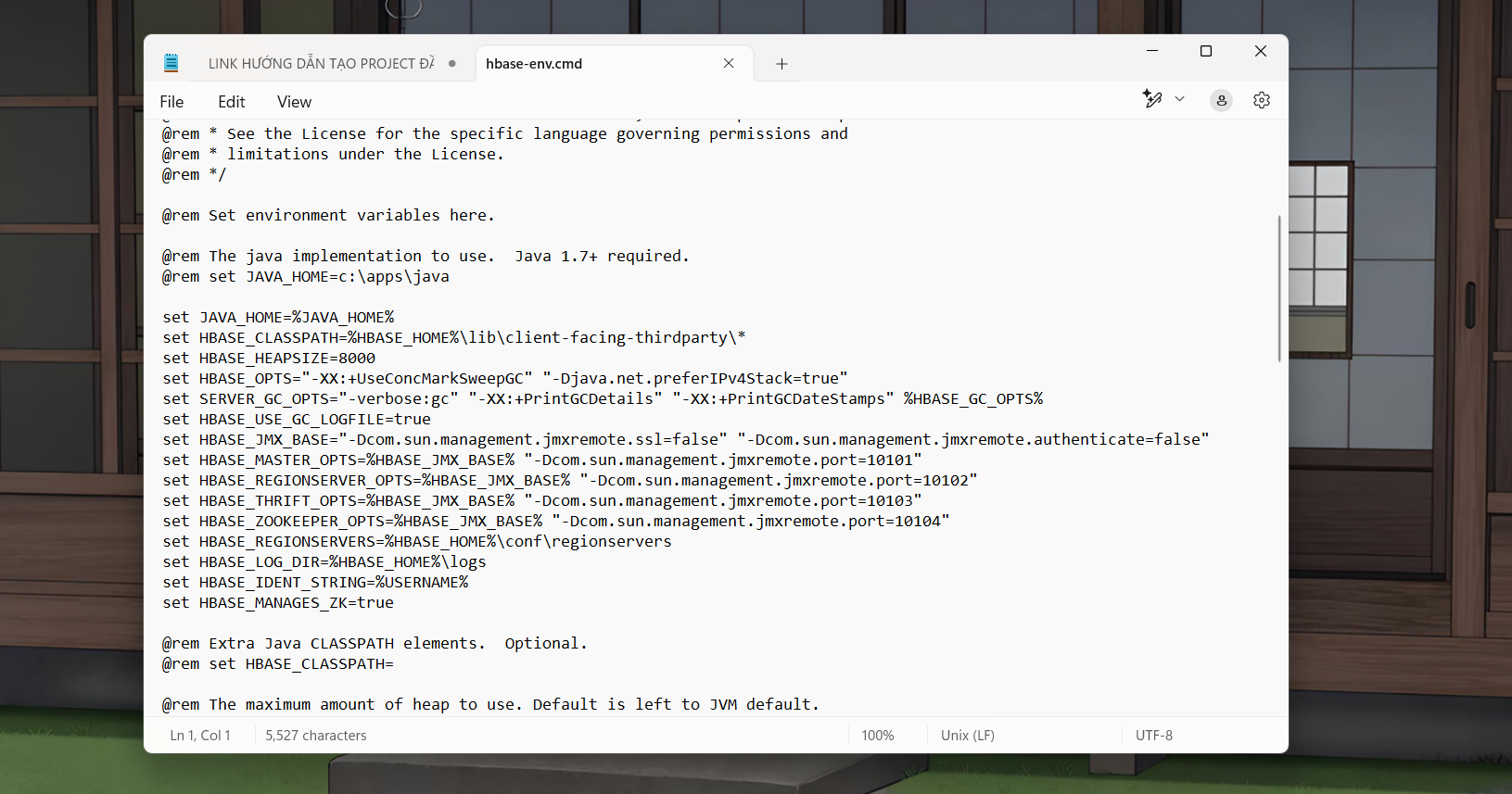
1. **Cấu hình file hbase-site.xml (nằm trong thư mục conf)**

****



* Nhớ tạo thêm trong ổ C:\ hai thư mục là Hfile và zookeeper

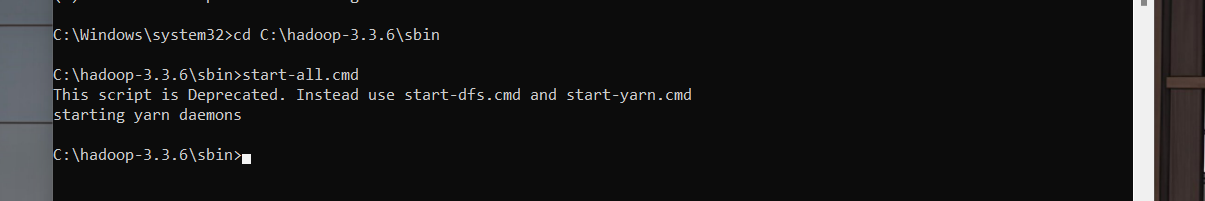
1. **Cấu hình file hbase-env.cmd**

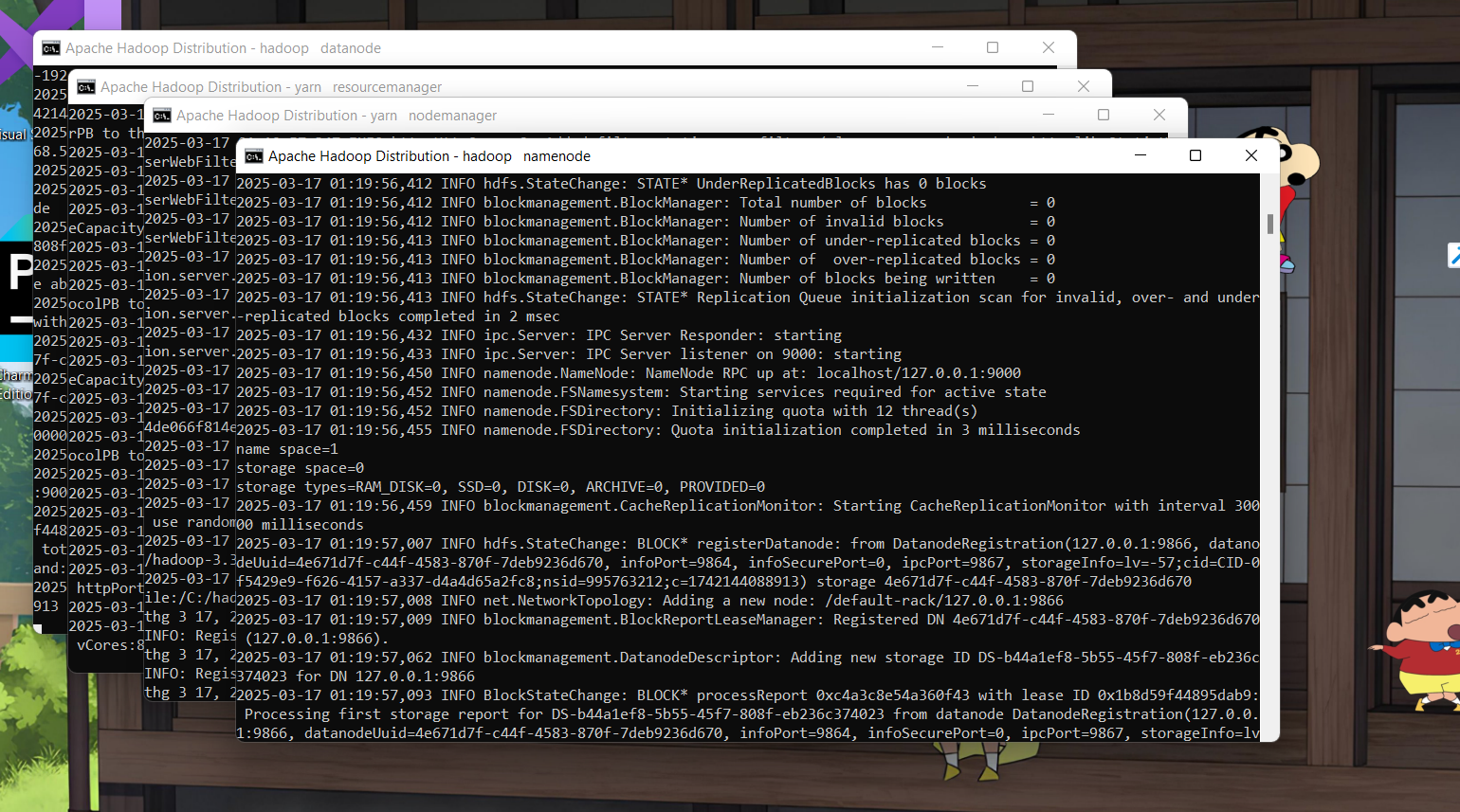
****



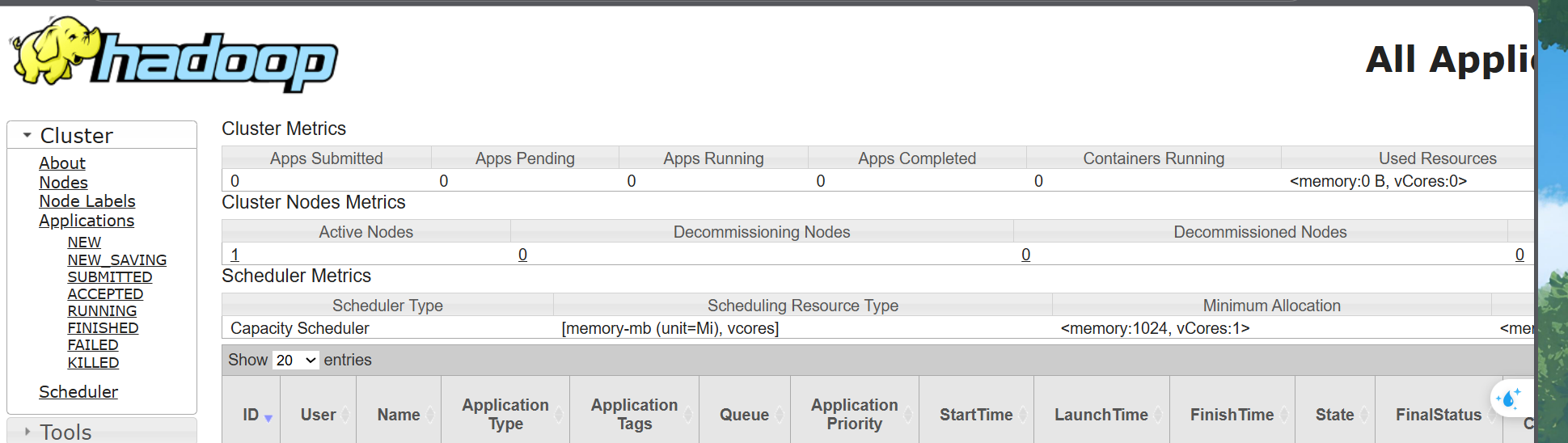
1. **Cấu hình đã xong và tiến hành khởi chạy Hbase**

* Mở cmd và chạy
* Trước khi chạy Hbase, chúng ta cần chạy Hadoop trước

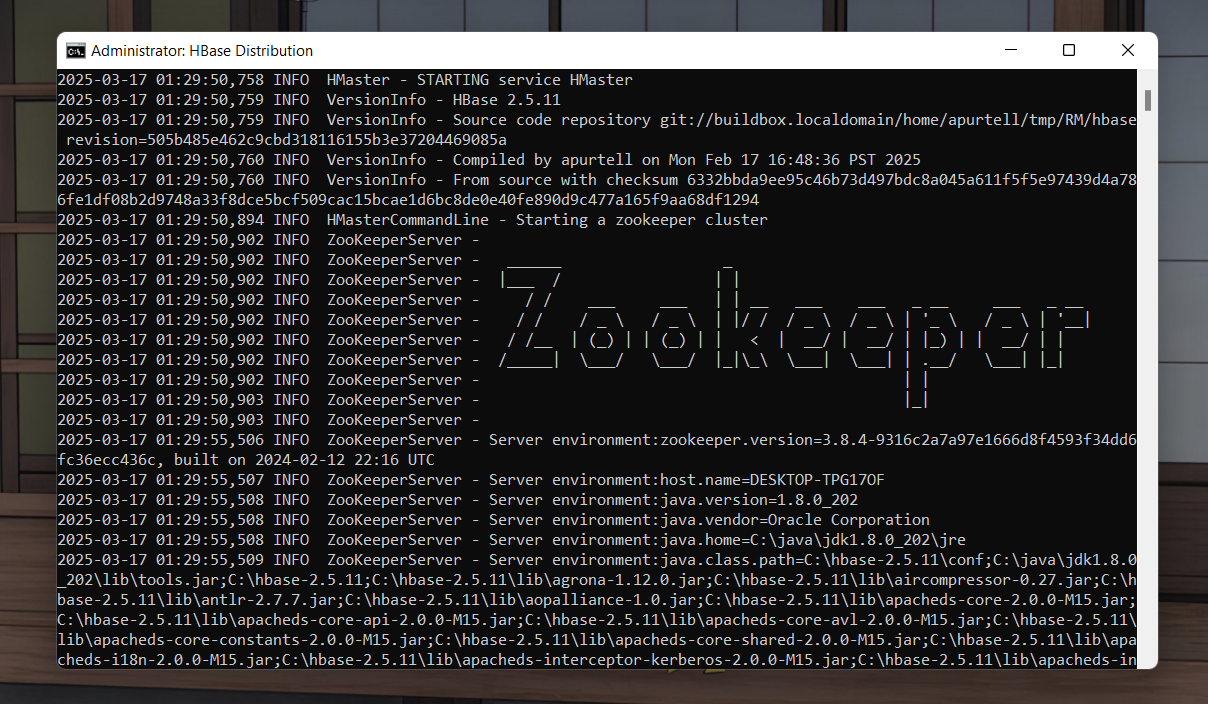


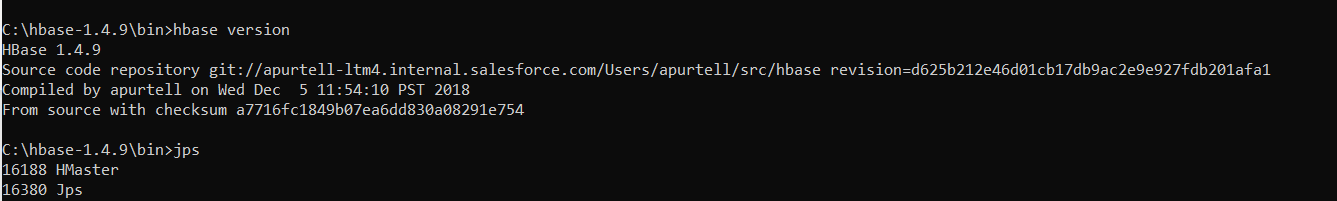


****

****

****

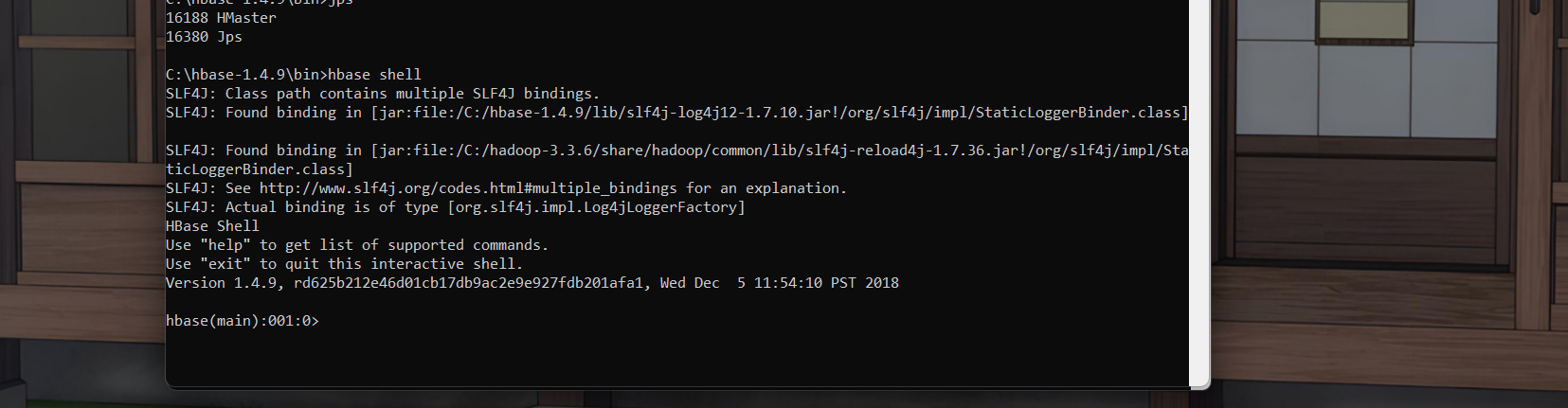
****

****



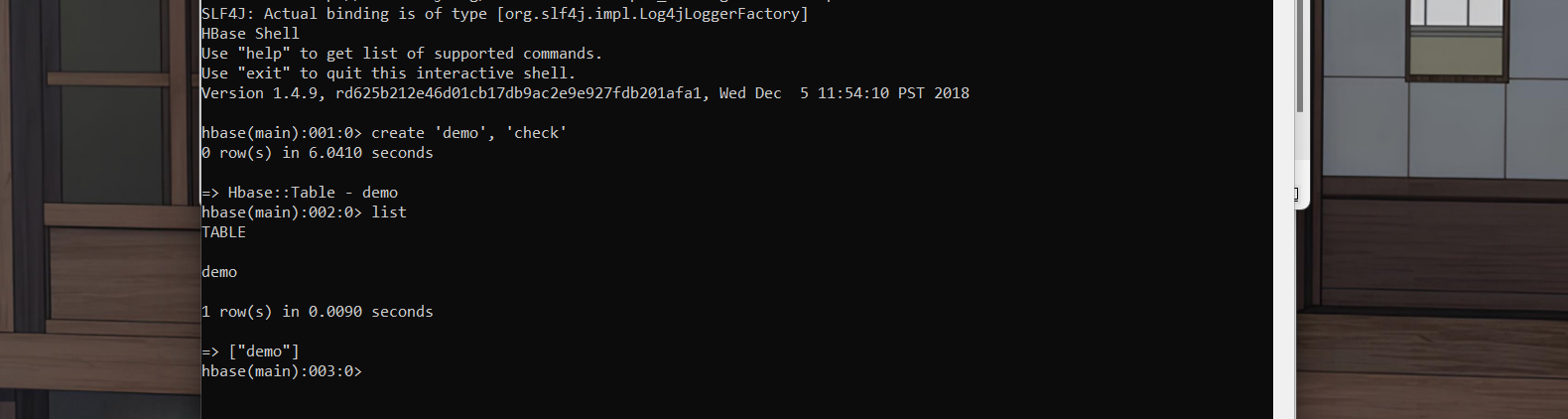
* Kiểm tra version Hbase thành công

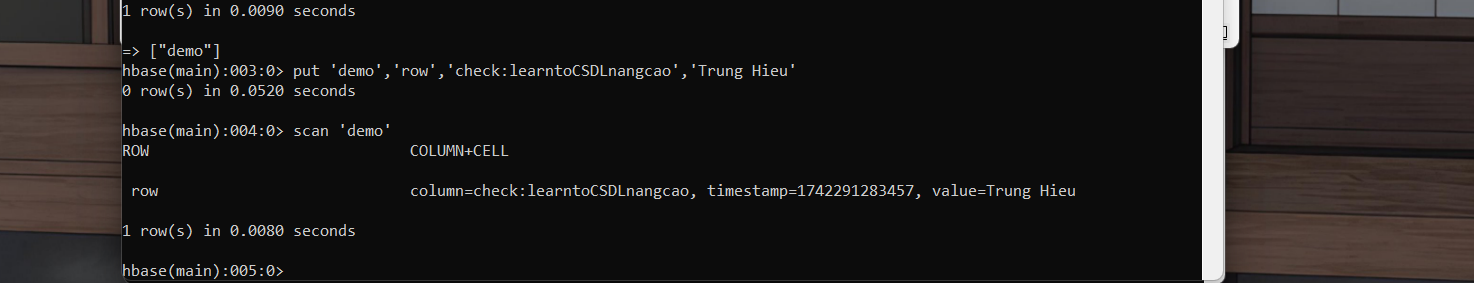
Truy cập Hbase shell và chạy demo



* Truy cập thành công Hbase

**Demo**

****

****

