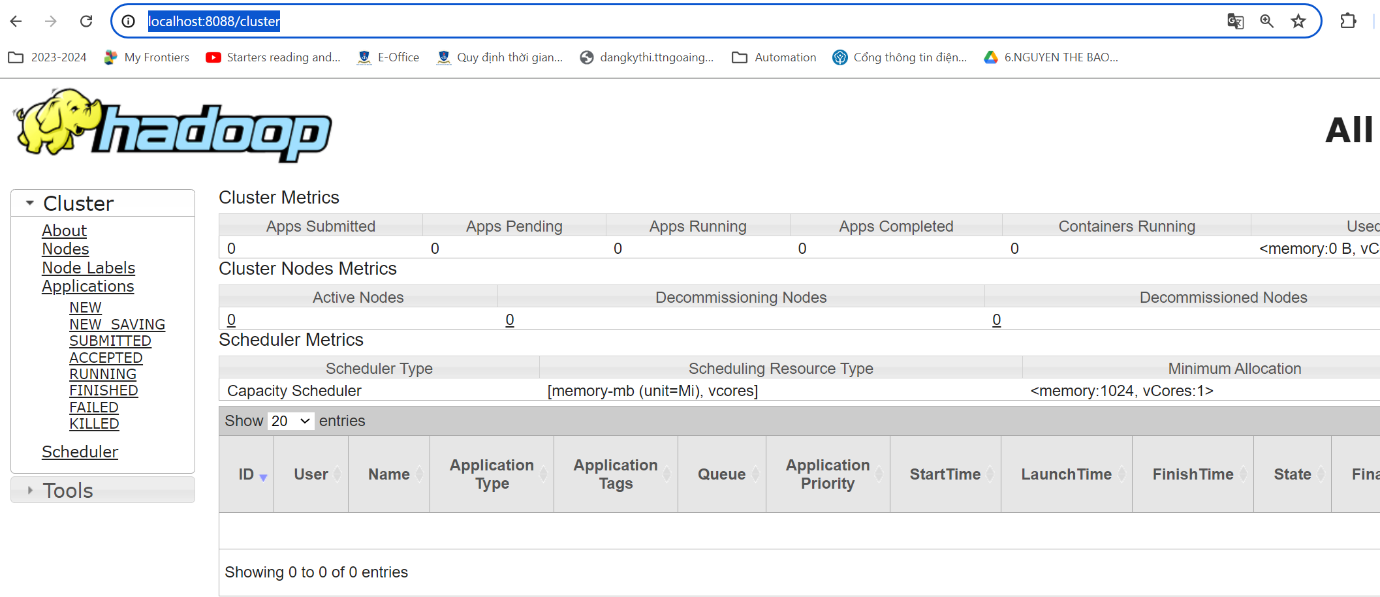
**LAB 2**

***Launching application on Hadoop – Spark Windows***

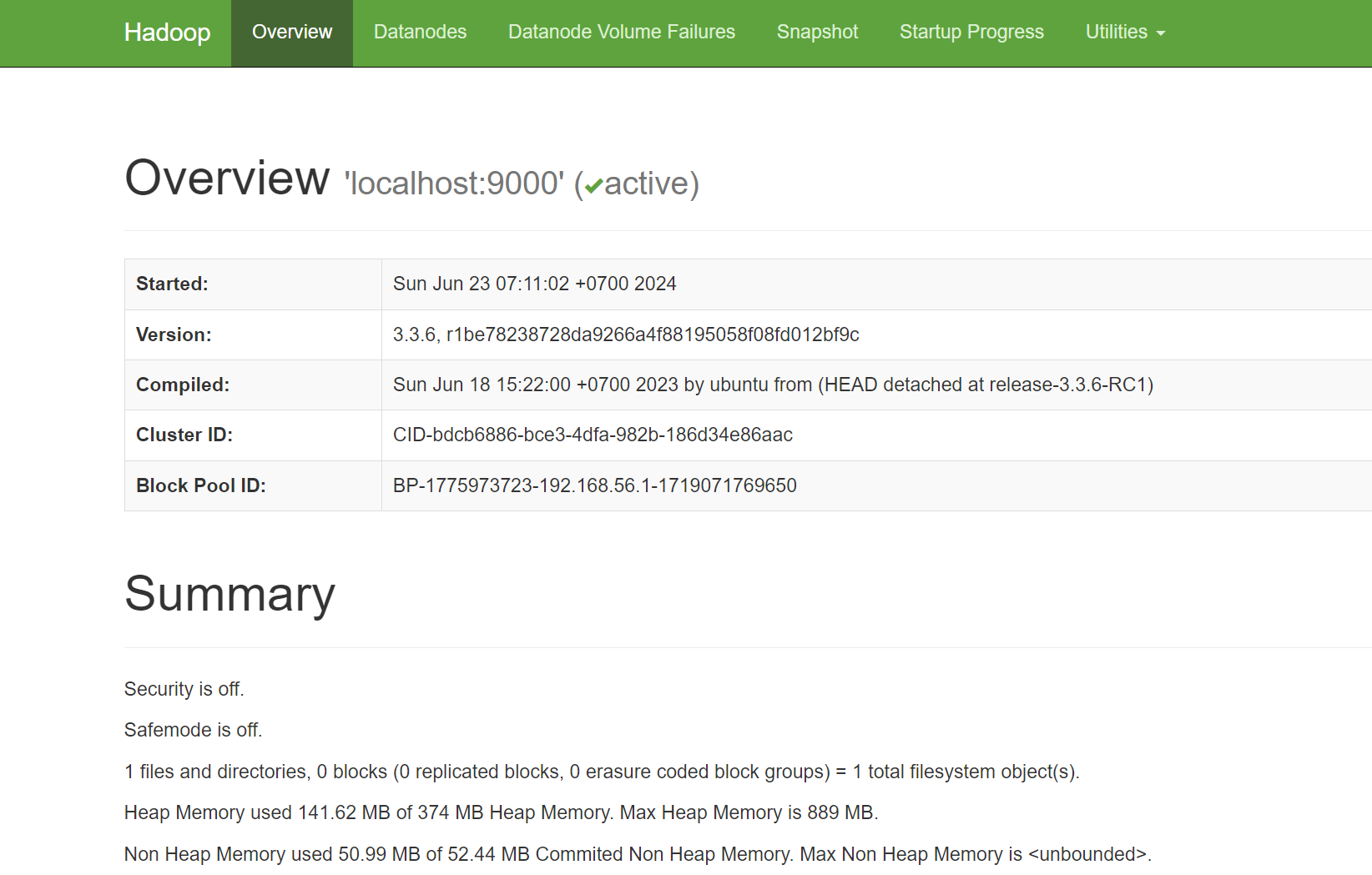
1. **Khởi động hệ thống: Hadoop, Spark.**

* Mở cửa new terminal=> **start-all.cmd** để khởi động hdfs và yarn.

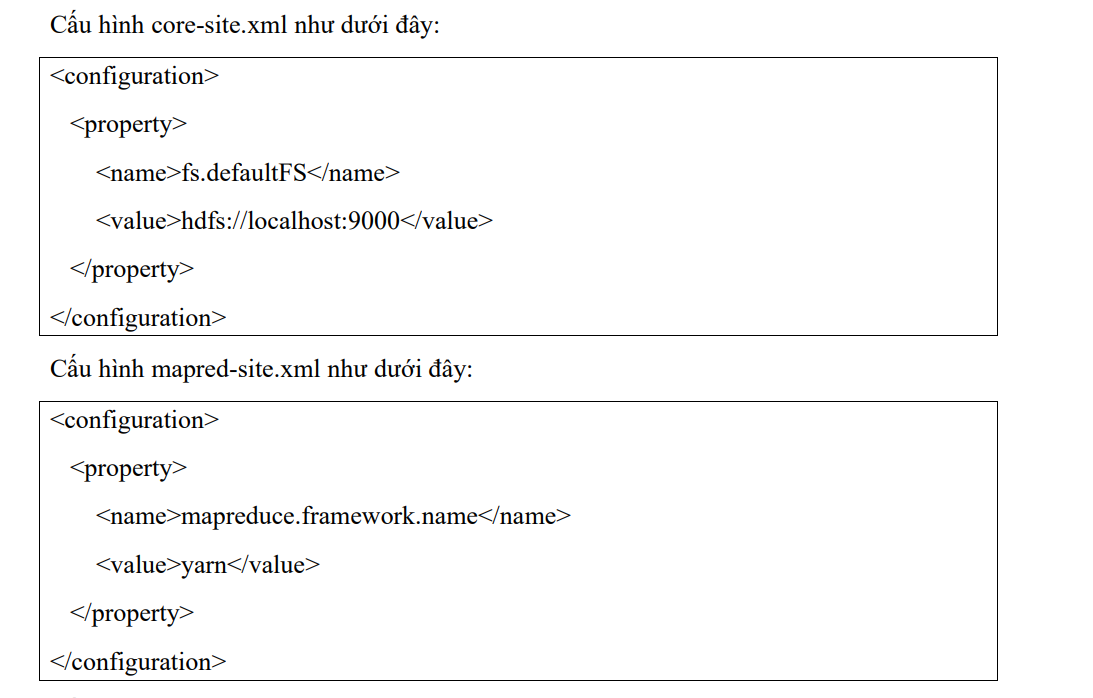
Xem giao diện web UI của Hadoop tại http://localhost:8088



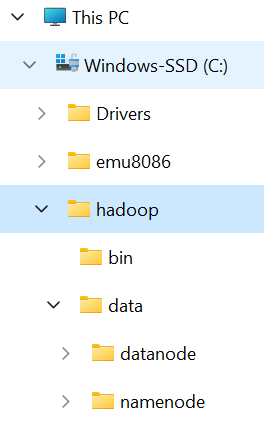
và <http://localhost:9870>



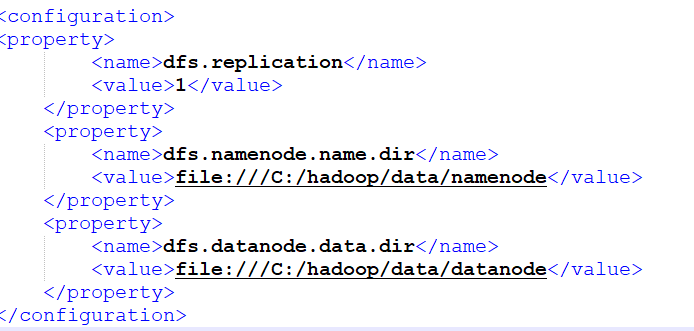
Lưu ý: trong trường hợp không thể start-all on hadoop, kiểm tra lại các file cấu hình trong thư mục cài đặt hadoop **C:\hadoop\etc\hadoop**



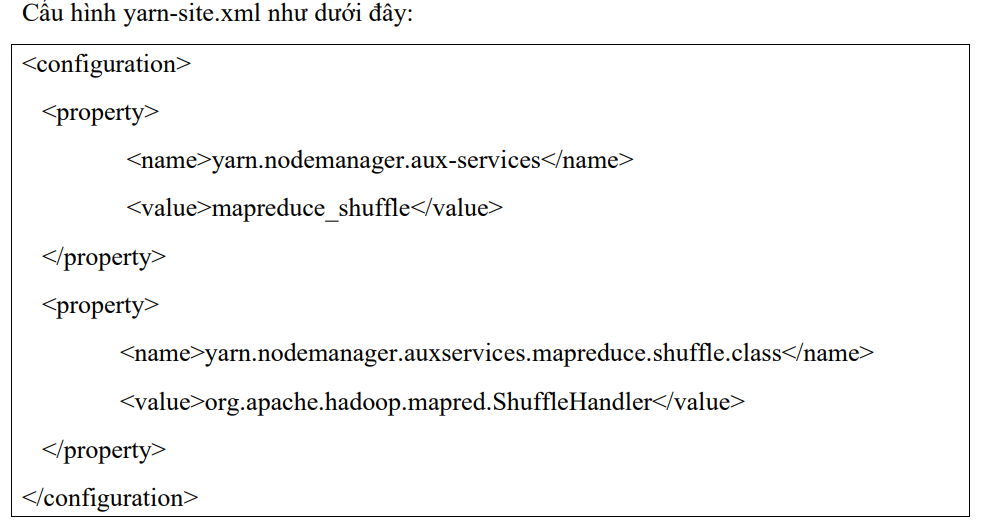
Kiểm tra thư mục cài đặt hadoop đã tạo các thư mục như hình hay chưa.

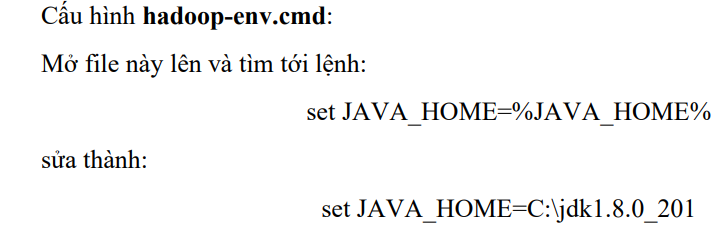


Cấu hình file hdfs-site.xml



Đường dẫn tới hai thư mục namenode và datanode





thay đổi thanh đường dẫn tới JAVA\_HOME của mình

* Tải winutils về giải nén ra thấy thư mục **bin** ở bên trong Chép đè thư mục **bin** này trong thư mục **bin** của hadoop đã cài đặt **C:\hadoop-3.3.6\bin**.
* Sau đó format lại namenode và datanode: mở command line lên, gõ lệnh sau:

**hdfs namenode –format** và **hdfs datanode -format**

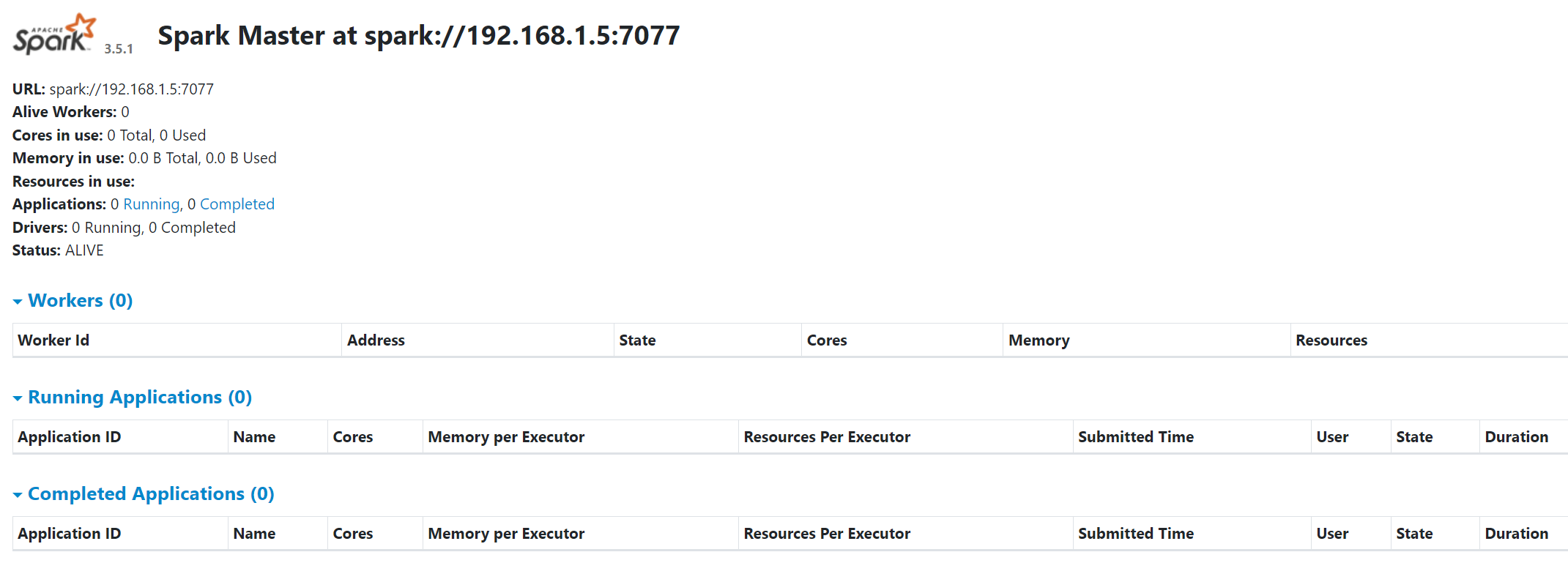
Sao chép file:

“C:\hadoop-3.3.0\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-3.3.0.jar”.

* Mở new terminal => spark-class org.apache.spark.deploy.master.Master để khởi động spark master

spark master sau khi khởi động xong, kiểm tra thông tin tại <http://localhost:8080>



Thông tin IP, Port (SV thay đổi IP, port theo thông tin của mình) của spark master Spark master at spark://**192.168.1.5**:***7077***

KETQUA

A screenshot of a computer

Description automatically generated

* Mở new terminal => spark-class org.apache.spark.deploy.worker.Worker spark://192.168.1.5:7077

spark worker sau khi khởi động xong, kiểm tra thông tin tại <http://localhost:8081>



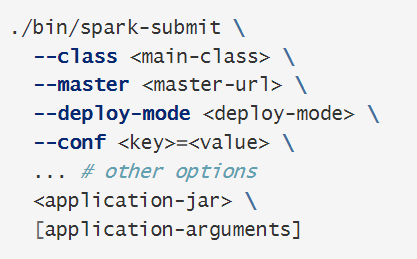
KETQUA

A white background with text

Description automatically generated

1. **Launching application với spark-submit**

* **Cú pháp:**

****

trong đó:

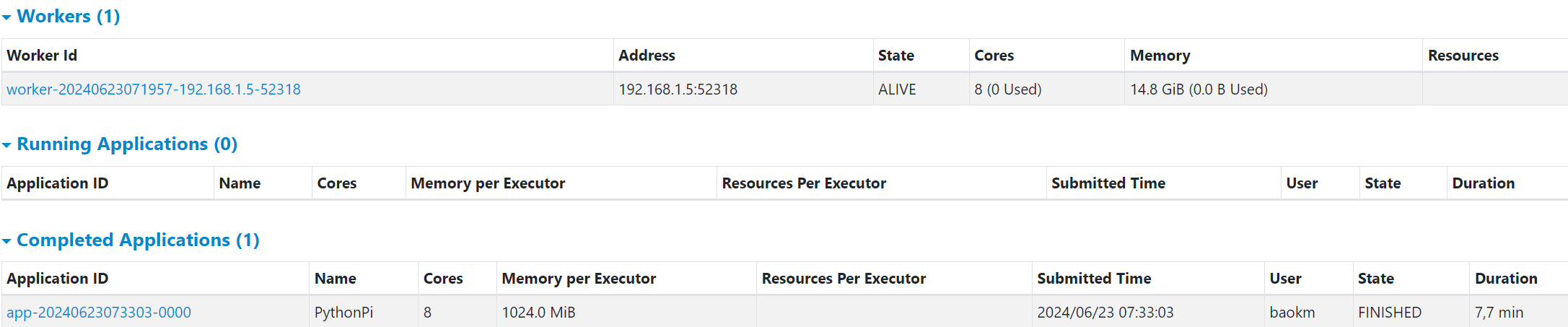
* --class: The entry point for your application (e.g. org.apache.spark.examples.SparkPi)
* --master: The [master URL](https://spark.apache.org/docs/latest/submitting-applications.html#master-urls) for the cluster (e.g. spark://23.195.26.187:7077)
* --deploy-mode: Whether to deploy your driver on the worker nodes (cluster) or locally as an external client (client) (default: client) **†**
* --conf: Arbitrary Spark configuration property in key=value format. For values that contain spaces wrap “key=value” in quotes (as shown). Multiple configurations should be passed as separate arguments. (e.g. --conf <key>=<value> --conf <key2>=<value2>)
* application-jar: Path to a bundled jar including your application and all dependencies. The URL must be globally visible inside of your cluster, for instance, an hdfs:// path or a file:// path that is present on all nodes.
* application-arguments: Arguments passed to the main method of your main class, if any
* Ví dụ 1: Run chương trình py.pi
* Chương trình py.pi được trong thư mục "C:\spark\spark-3.5.1-bin-hadoop3\examples\src\main\python\pi.py".
* Mở new terminal => cd %SPARK\_HOME%



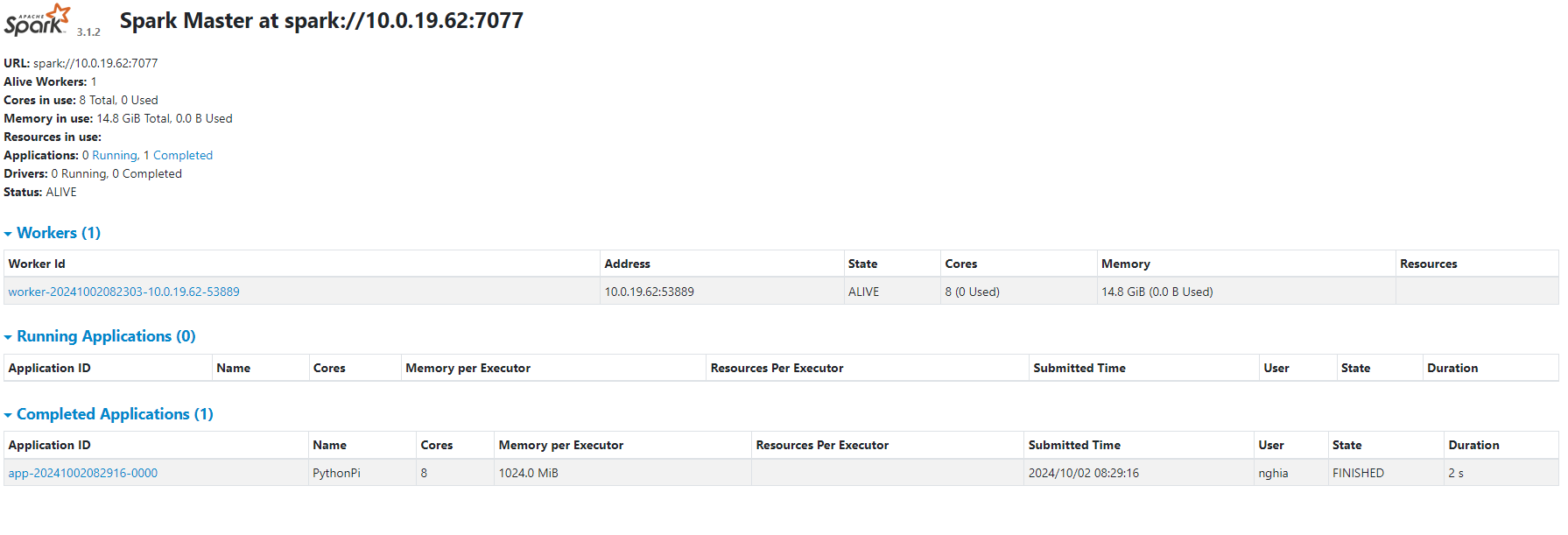
* Nhập lệnh:

***spark-submit --master spark://192.168.1.5:7077 examples/src/main/python/pi.py 1000***

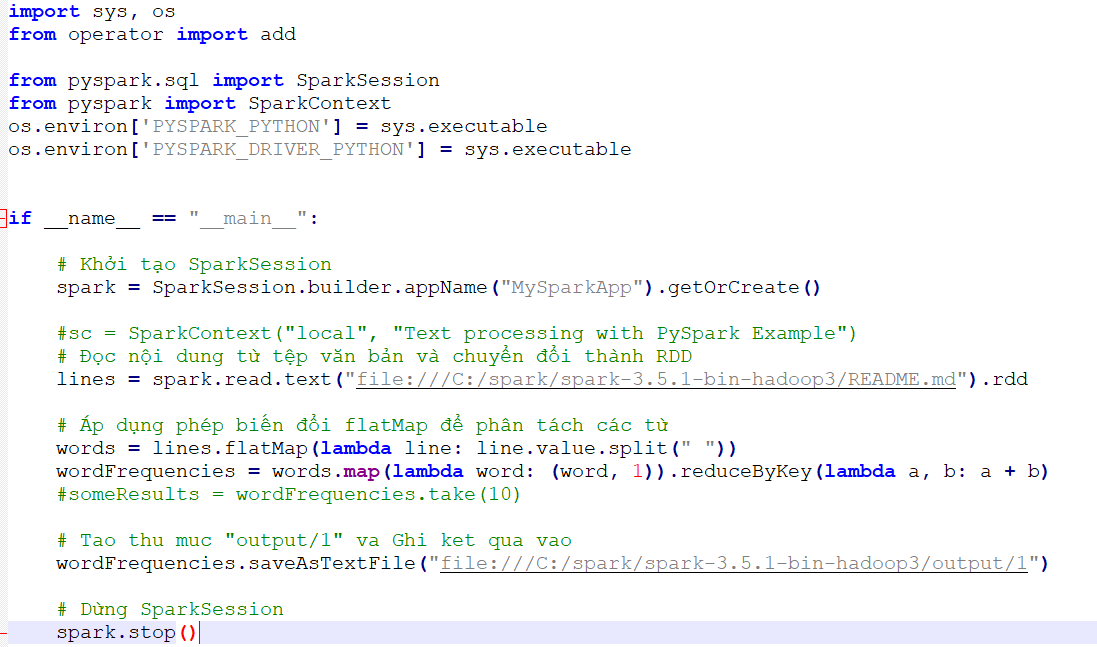
* Xem kết quả tại http://localhost:8080/



KETQUA



Ví dụ 2: viết chương trình wordcount\_demo.py có nội dung như sau

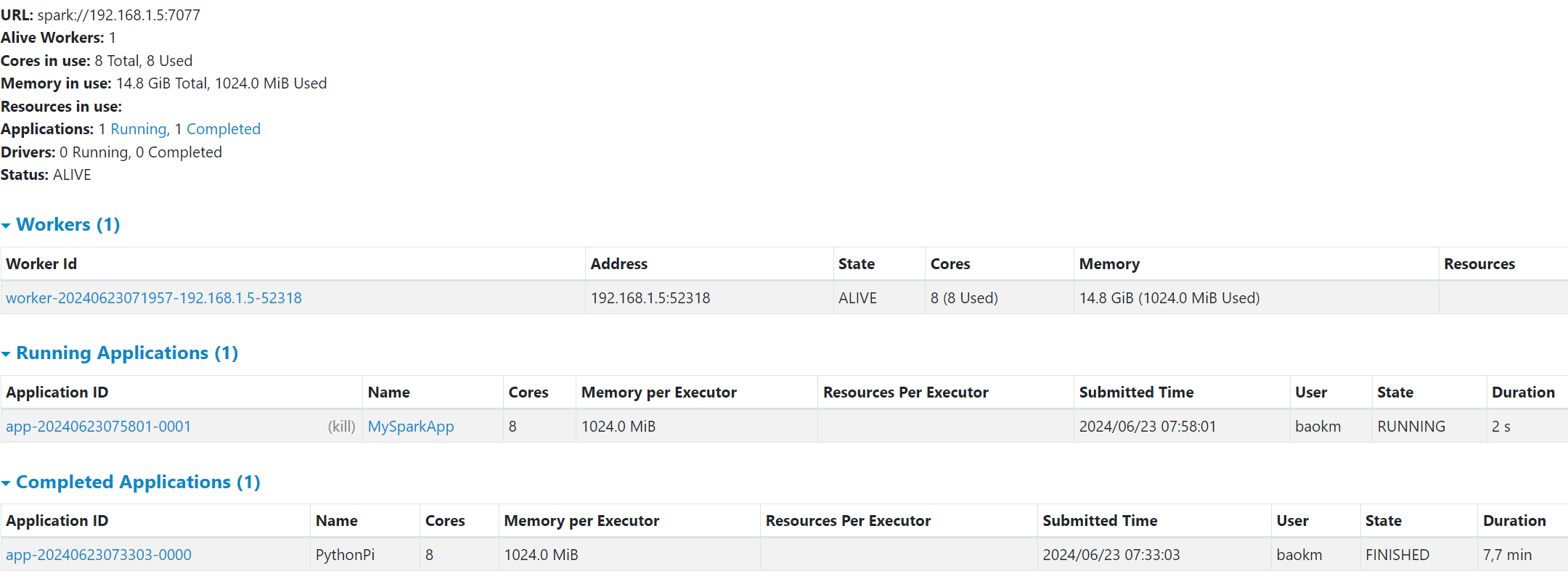


* Mở new terminal => cd %SPARK\_HOME% => Nhập lệnh:

cd %SPARK\_HOME%

***spark-submit --master spark://192.168.1.5:7077 examples/src/main/python/wordcount\_demo.py***

* *Xem kết quả*



ta thấy có một running application. Sau khi application hoàn thành, ở mục Completed application đã có thêm thông tin của application vừa hoàn thành.



KETQUA

A screenshot of a computer

Description automatically generated

1. Exercises

Sinh viên thực thi các chương trình còn lại trong thư mục

**examples\src\main\python**

**A black screen with white text

Description automatically generated**

**Đa số là bị lỗi như hình trên**

**A screenshot of a computer

Description automatically generated**

1. Tham khảo

<https://spark.apache.org/docs/latest/submitting-applications.html>