Virtual machine usage documentation

1. **Open the virtual machine**

图形用户界面

描述已自动生成Start three virtual machines guojiban1, guojiban2 and guojiban3 on VMware Workstation Pro, as shown in Figure 1:

**Figure 1 Virtual machine startup interface**

电脑萤幕的截图

描述已自动生成

**Figure 2 Start the virtual machine guojiban1**

电脑萤幕的截图

描述已自动生成

**Figure 3 Start the virtual machine guojiban2**

电脑萤幕的截图

描述已自动生成

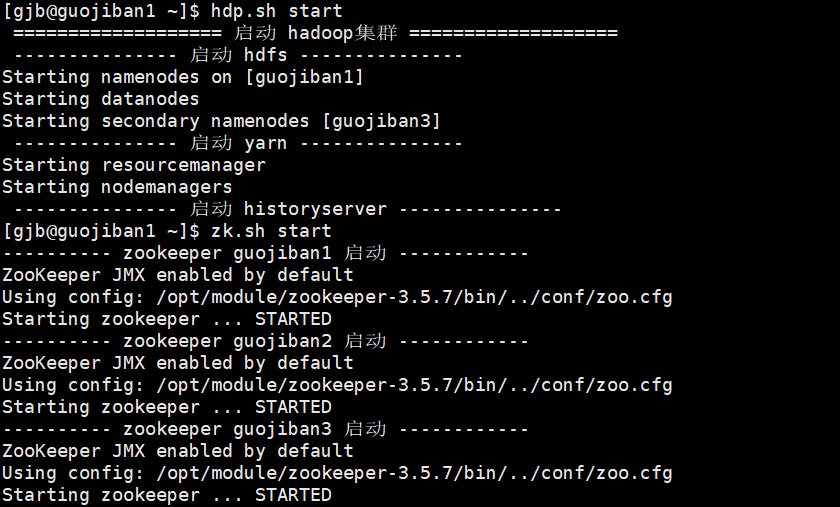
**Figure 4 Start the virtual machine guojiban3**

**2. Log in to the virtual machine**

(1) Enter the virtual machine username gjb and password 000000 to operate the virtual machine (the three virtual machine account passwords are exactly the same), as shown in Figure 5 below:

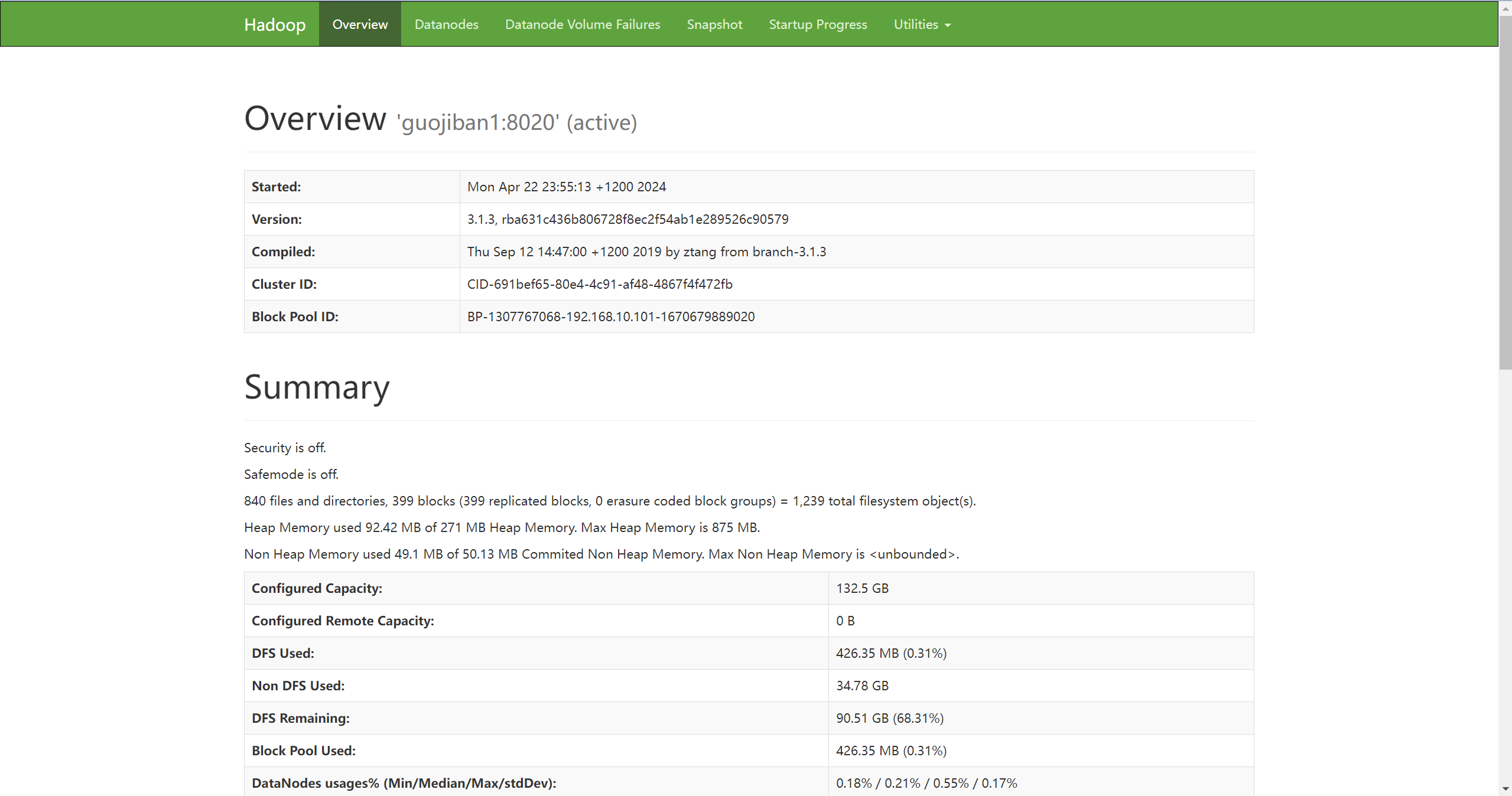


**Figure 5 Enter the virtual machine account (this project uses xshell to simplify virtual machine operations)**

(2) Start the Hadoop cluster: Enter the hdp.sh start and zk.sh start scripts on guojiban1 to start the Hadoop cluster and Zookeeper respectively, as shown in Figure 6 below:

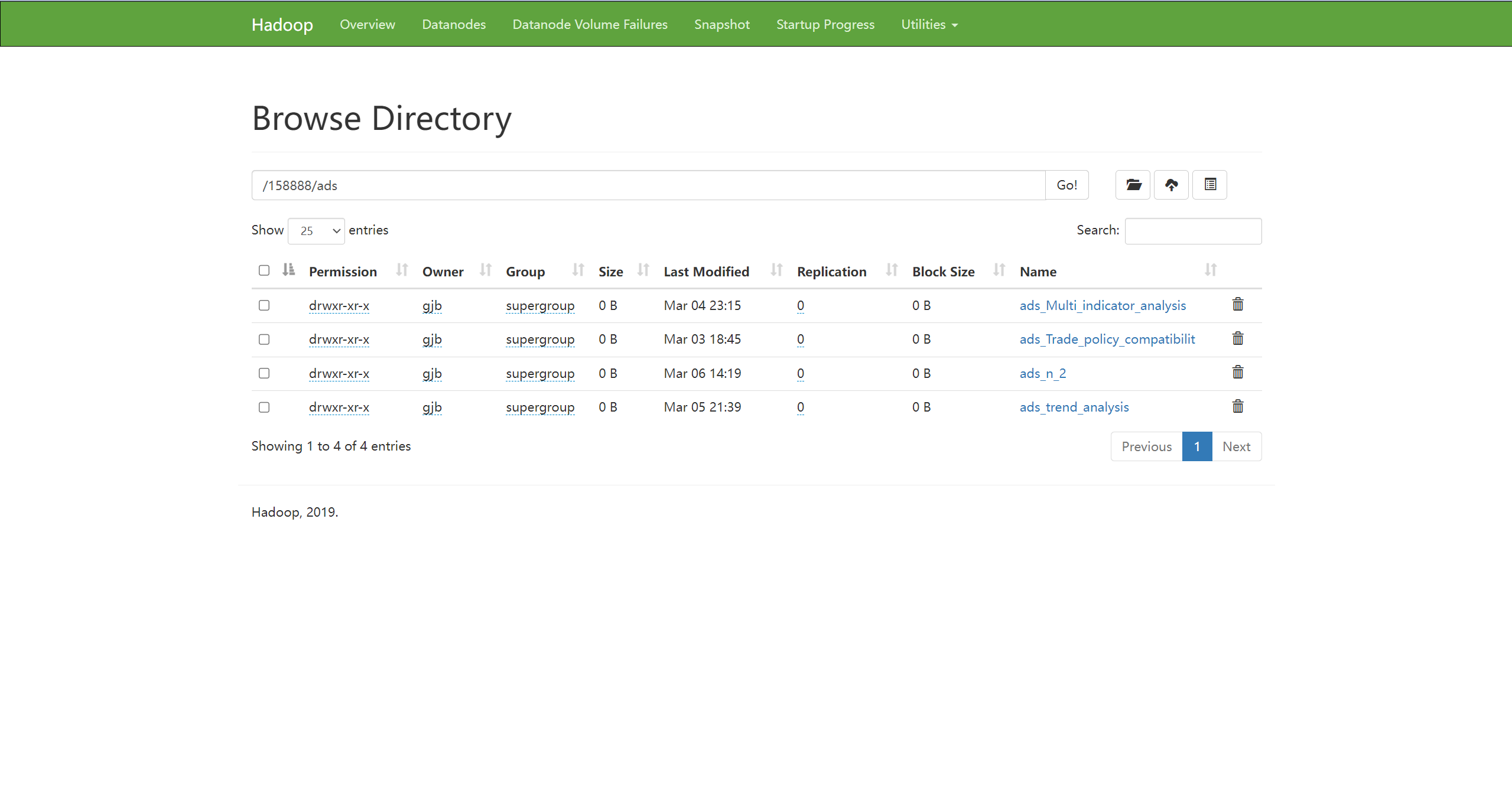
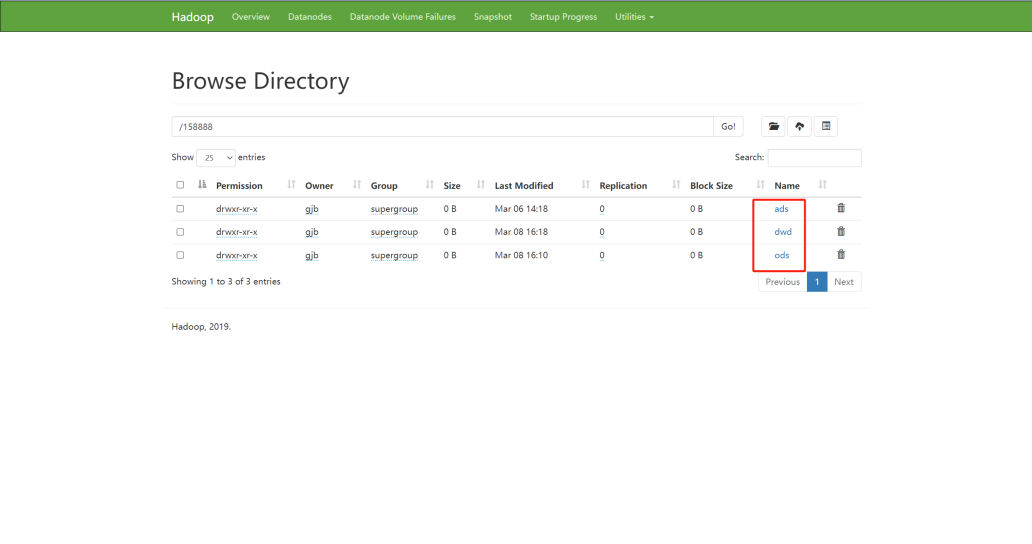
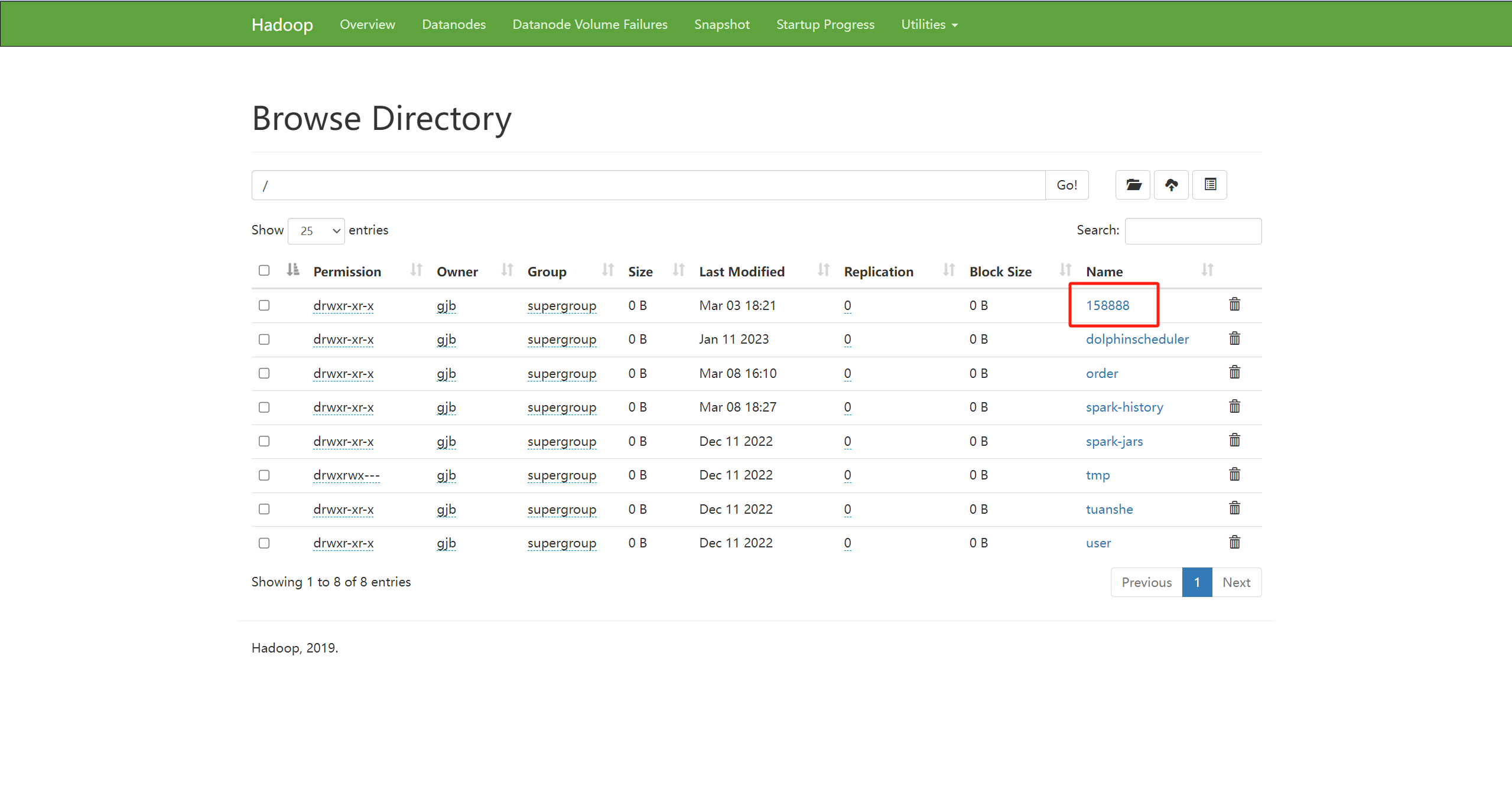
**Figure 6 Starting the Hadoop cluster and Zookeeper virtual machine interface**

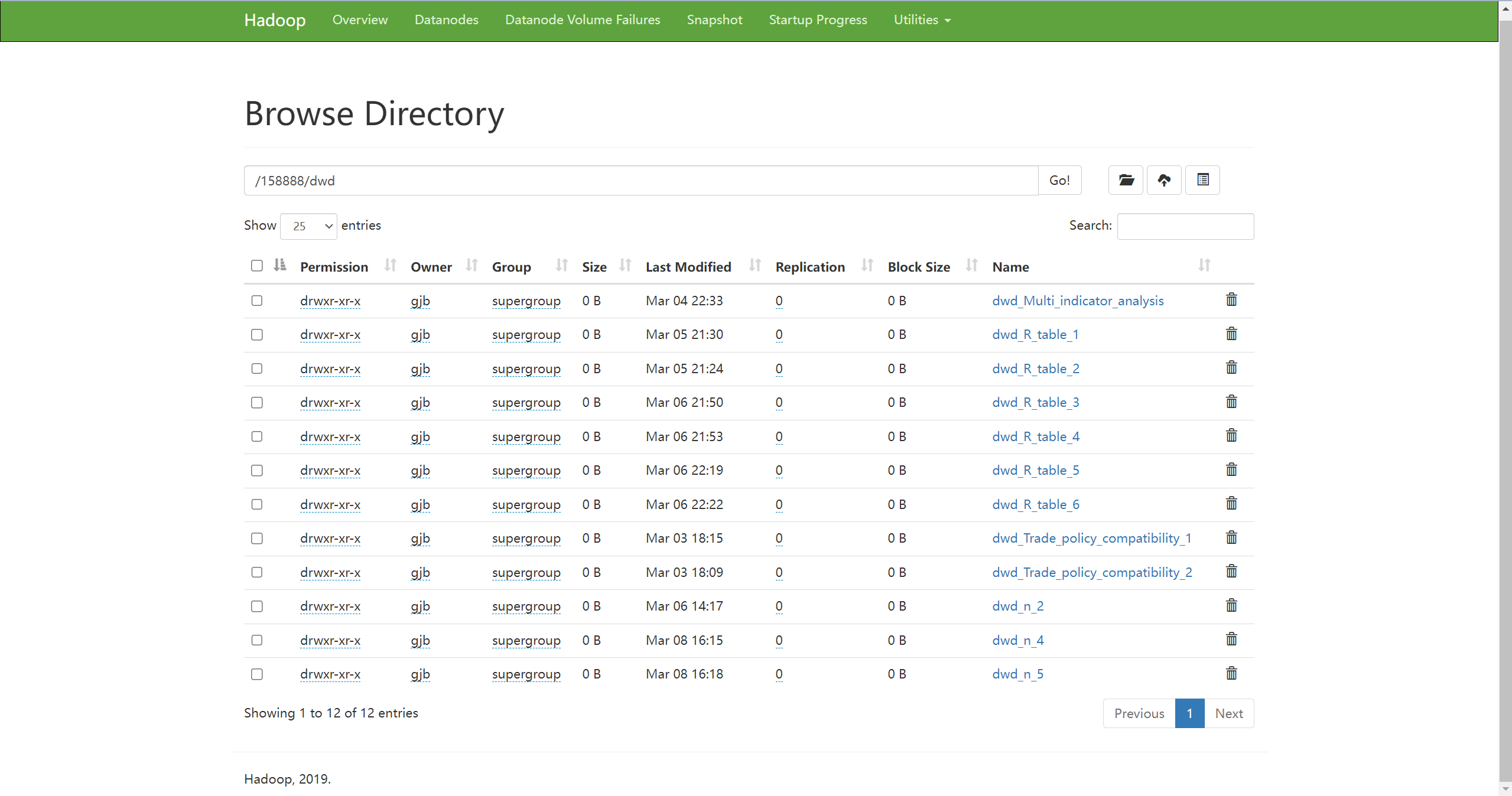
1. Enter http://guojiban1:9870/ to view the cluster startup status and layering status, as shown in Figure 7 below:



**Figure 7 Open the web page to view the cluster startup status**

1. Check the layering situation

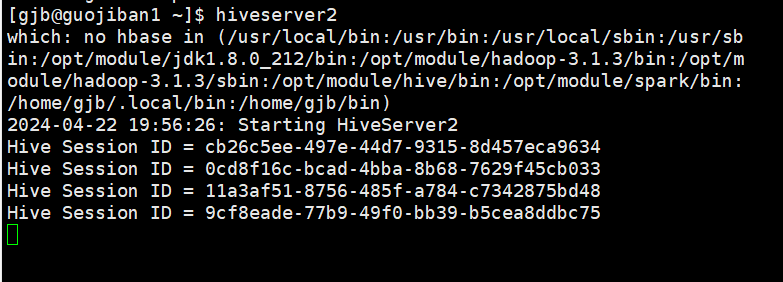




**Figure 8 Check the layering situation**

1. Start and use Hive

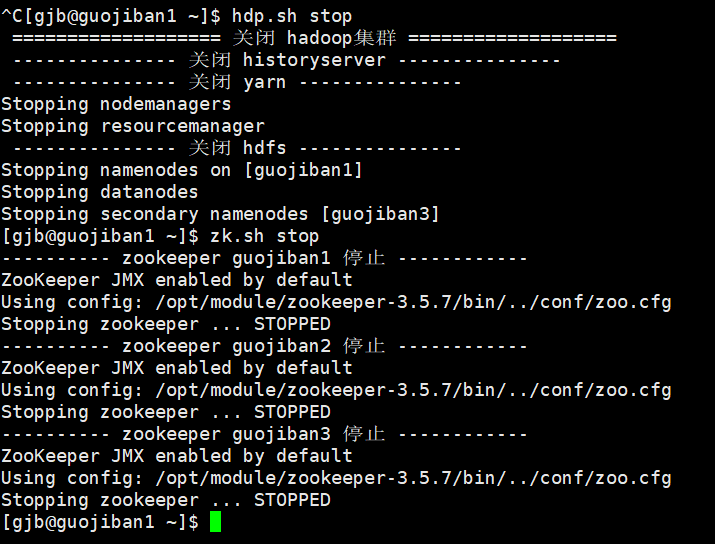
Enter hiveserver2 in the terminal of guojiban1 to open it (it cannot be closed during operation)



**Figure 9 Start and use Hive**

1. Shut down the cluster

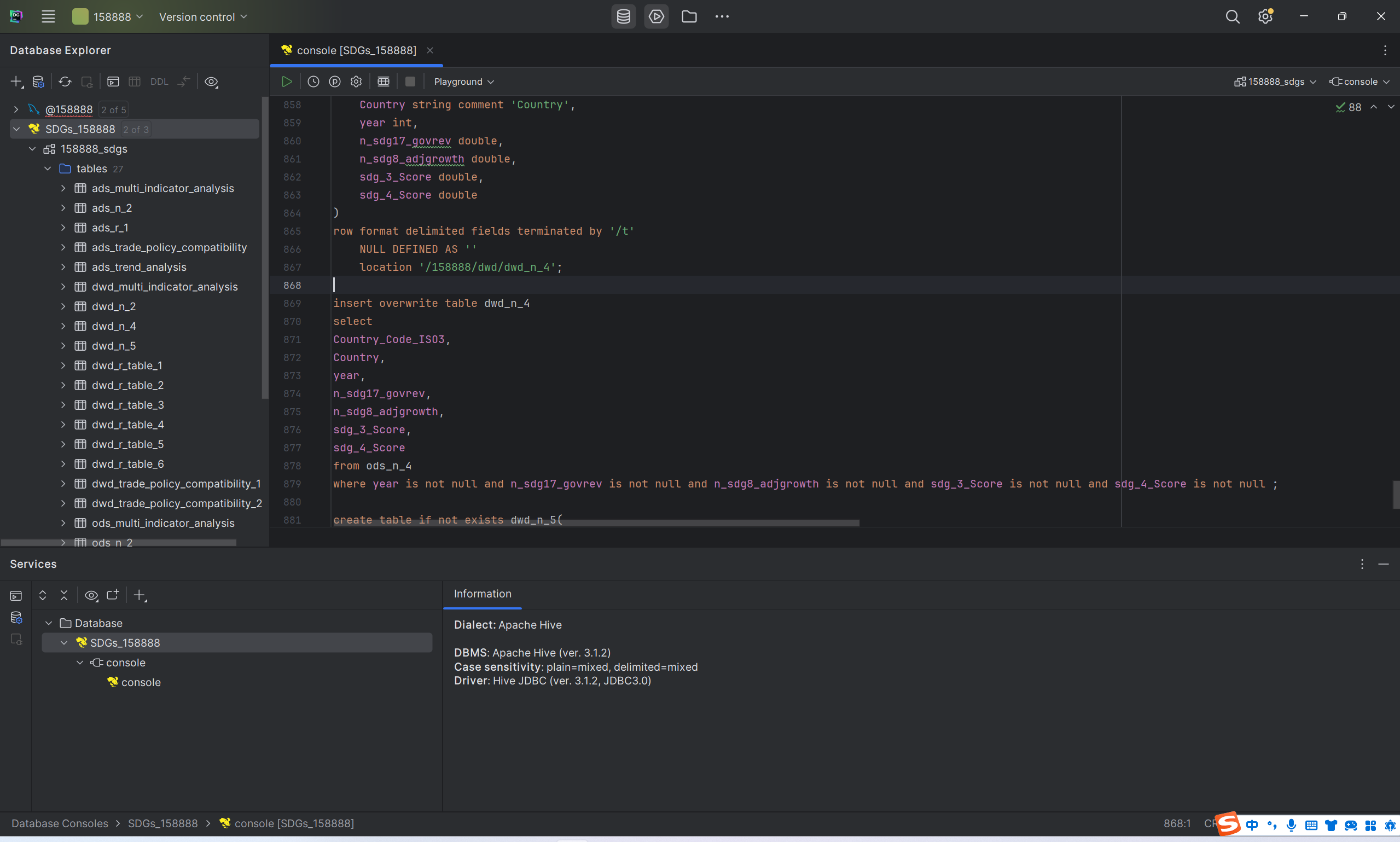
Enter the codes "hdp.sh stop" and "zk.sh stop"



**Figure 10 Shut down the cluster**

**3. Code part**

This project uses Datagrip to connect to the Hive library and write SQL code. As shown in Figure 10: (For code details, please refer to the SQL\_code.txt file uploaded to github)



**Figure 11 Datagrip code usage part**

**4. Cluster scheduling and monitoring (This section is not necessary for the project,)**

(1) On guojiban1, enter the dolphinscheduler directory under the module of opt and enter "./bin/start-all.sh" to start DolphinScheduler. Enter "xcall.sh jps" to query. If there are no failed nodes, enter "sudo systemctl start zabbix-server zabbix-agent httpd rh-php72-php-fpm" and "sudo systemctl enable zabbix-" on the guojiban1 virtual machine. server zabbix-agent httpd rh-php72-php-fpm" these two lines of instructions.

电脑萤幕的截图

描述已自动生成

**Figure 11 Start Zabbix input command interface in the guojiban1 virtual machine**

1. Enter "sudo systemctl start zabbix-agent and sudo systemctl enable zabbix-agent" on the guojiban2 virtual machine, as shown in Figure 12 below:

电脑萤幕的截图

描述已自动生成

**Figure 12 Start Zabbix input command interface in the guojiban2 virtual machine**

(4) Enter "sudo systemctl start zabbix-agent" and "sudo systemctl enable zabbix-agent" on the guojiban3 virtual machine, as shown in Figure 13 below:

电脑萤幕的截图

描述已自动生成

**Figure 13 Start Zabbix input command interface in the guojiban3 virtual machine**

1. Under normal circumstances, the node status needs to be checked. If the node is lost, Zabbix will send emails to our team's mailbox in the following two monitoring. Enter http://guojiban1/zabbix/ to enter Zabbix monitoring. Username Admin, password zabbix. You can query the health of the node. If the node is lost, information will be sent to the email, as shown in the following figure:

图形用户界面, 应用程序

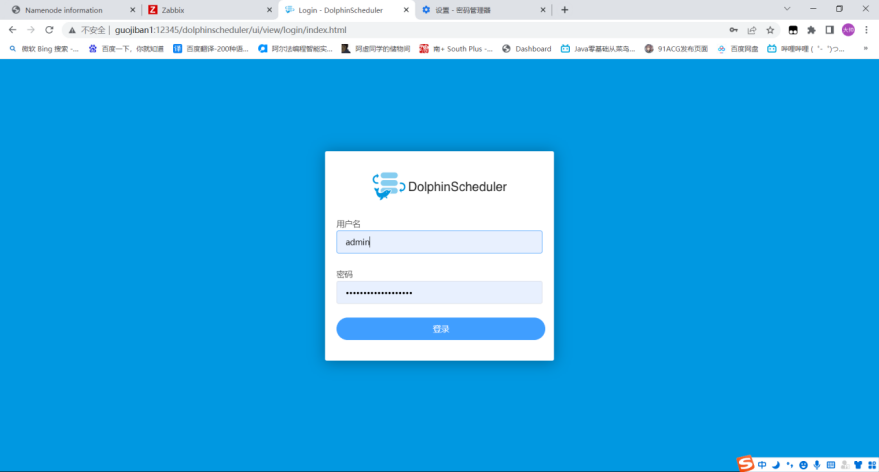
描述已自动生成

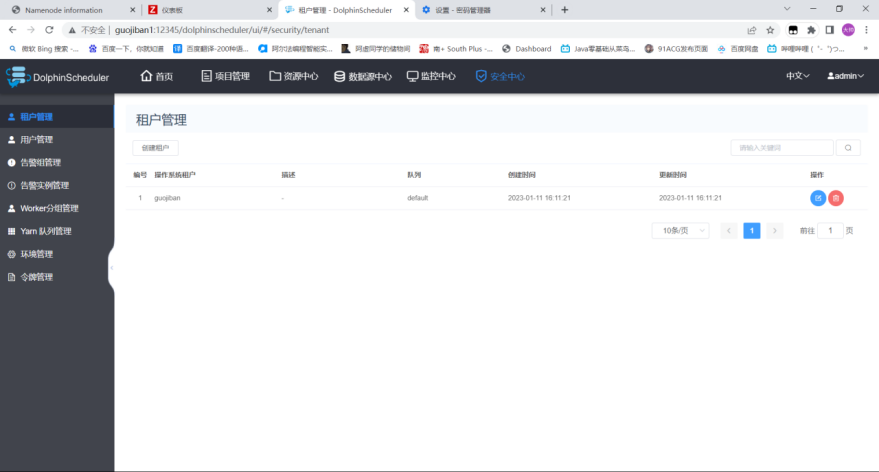
电脑软件的截图

描述已自动生成

**Figure 14 Zabbix monitoring**

1. Enter http://guojiban1:12345/dolphinscheduler/ui/view/login/index.html. User name is admin, password is dolphinscheduler, enter the administrator interface, enter user name gjb, password guojiban000000, and view specific tasks. Enter DolphinScheduler to view task scheduling, as shown in the figure below:

**Figure 15 dolphinscheduler login interface**

**Figure 16 Administrator interface**

图形用户界面, 图表

描述已自动生成**Figure 17 User interface**