**HADOOP KURULUM**

Kurulum işlemleri 3 adet 8GB memory 2cpu ve 20gb disk alanına sahip sunucu üzerinde yapılmıştır. 1 Namenode ve 2 DataNode ile cluster oluşturulmuştur.

1. **Tüm sunucularda vi komutu ile /etc/hosts dosyası açılır, üç sunucunun da IP adresleri ve hostname’leri bu dosya içine aşağıdaki formatta yazılır.**

| 34.30.111.172 masternode01 : namenode  104.197.161.170 slavenode01 : datanode  35.188.42.123 slavenode02 : datanode |
| --- |

1. **Sunucular üzerinde Firewall ve selinux servisleri durdurulur. Bu servisleri durdurmak için aşağıdaki komutlar çalıştırılır.**

| [yyukselveysel44@masternode01 ~]$ sudo systemctl stop firewalld  [yyukselveysel44@masternode01 ~]$ sudo systemctl disable firewalld  Removed /etc/systemd/system/multi-user.target.wants/firewalld.service.  Removed /etc/systemd/system/dbus-org.fedoraproject.FirewallD1.service.  [yyukselveysel44@masternode01 ~]$ getenforce  Enforcing  [yyukselveysel44@masternode01 ~]$ sudo setenforce 0  [yyukselveysel44@masternode01 ~]$ getenforce  Permissive |
| --- |

1. **Sunuculara aşağıdaki komut ile Java yüklenir. Yükleme işleminden sonra java –version komutu ile kontrol edilir.**

| [yyukselveysel44@masternode01 ~]$ sudo yum install java-1.8.0-openjdk  [yyukselveysel44@masternode01 ~]$ echo "export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.422.b05-2.el8.x86\_64" >> ~/.bashrc  [yyukselveysel44@masternode01 ~]$ sudo yum install java-<version>-openjdk-devel.x86\_64 |
| --- |

1. **Hadoop kullanıcısı kurulum aşamasında oluşturulmadı ise, aşağıdaki gibi useradd komutu ile 3 sunucuda da Hadoop kullanıcısı oluşturulur ve şifresi passwd komutu ile belirlenir.**

| [yyukselveysel44@masternode01 ~]$ sudo adduser hadoop  [yyukselveysel44@masternode01 ~]$ sudo passwd hadoop  Changing password for user hadoop.  New password:  BAD PASSWORD: The password fails the dictionary check - it is based on a dictionary word  Retype new password:  passwd: all authentication tokens updated successfully. |
| --- |

1. **Hadoop kullanıcısı oluşturulduktan sonra, hadoop kullanıcısına geçilir ve master node’un slave node’lara bağlanması için passwordless ssh konfigürasyonu yapılır. Bu adım sadece masternode01 sunucusunda yapılır. Şu ayarın Yes olması gerekir. Sshd\_conf dosyasında gerekli değişiklik yapılmalıdır.  
   PubkeyAuthentication yes**

**PasswordAuthentication yes # (ilk bağlantı için geçici olarak gerekebilir)**

**PermitRootLogin prohibit-password**

**sudo cat /etc/ssh/sshd\_config | grep -Ei 'PubkeyAuthentication|PasswordAuthentication|PermitRootLogin'**

| [hadoop@masternode01 ~]$ssh-keygen -t rsa -b 2048  [hadoop@masternode01 ~]$ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop@slavenode01  [hadoop@masternode01 ~]$ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop@slavenode02  [hadoop@masternode01 ~]$ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop@masternode01 |
| --- |

1. **Hadoop 3.4.0 kurulum dosyası masternode01’de /opt altına indirilir.**

| [yyukselveysel44@masternode01 ~]$sudo wget <https://archive.apache.org/dist/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz> |
| --- |

1. **3 sunucuda da hadoop binary’lerinin kurulacağı /cluster dizini oluşturulur ve sahipliği hadoop kullanıcısı olarak ayarlanır.**

| [yyukselveysel44@masternode01 /]$ sudo mkdir /cluster  [yyukselveysel44@masternode01 /]$ sudo chown -R hadoop:hadoop /cluster |
| --- |

/home/hadoop dizininde yer alan .bash\_profile dosyası vi komutu ile açılır. .bash\_profile dosyasının içine HADOOP\_HOME ve PATH sistem değişkenleri eklenir.

1. **Masternode01’de /opt altına indirilen hadoop-3.4.0 kurulum dosyası /cluster dizini altına mv edilir. Daha sonra tar komutu ile arşiv açılarak binary dosyalar sisteme yüklenmiş olur. Arşivden hadoop-3.4.0 adı ile çıkan klasör, hadoop olarak yeniden isimlendirilir.**

| [yyukselveysel44@masternode01 /]$ sudo mv /opt/hadoop-3.4.0.tar.gz /cluster/.  [yyukselveysel44@masternode01 /]$ sudo chown -R hadoop:hadoop /cluster  [yyukselveysel44@masternode01 /]$ su - hadoop  [hadoop@masternode01 ~]$ cd /cluster/  [hadoop@masternode01 cluster]$ ls  hadoop-3.4.0.tar.gz  [hadoop@masternode01 cluster]$ tar -zxvf hadoop-3.4.0.tar.gz  hadoop@masternode01 cluster]$ mv hadoop-3.4.0 hadoop  [hadoop@masternode01 cluster]$ ls  hadoop hadoop-3.4.0.tar.gz  [hadoop@masternode01 cluster]$ rm -rf hadoop-3.4.0.tar.gz |
| --- |

1. **Bu adım ile birlikte masternode01 üzerinde hadoop namenode servisi ve yarn için /cluster/hadoop/etc/hadoop dizininde bulunan dosyalarda konfigürasyon ayarları yapılmaya başlanır. *Bu adım sadece masternode01 üzerinde yapılır.***

**9.1 core-site.xml dosyasının içeriği aşağıdaki gibi düzenlenir.**

| [hadoop@masternode01 hadoop]$ pwd  /cluster/hadoop/etc/hadoop  [hadoop@masternode01 hadoop]$ cat core-site.xml  <!-- Put site-specific property overrides in this file. -->  <configuration>  <property>  <name>**fs.defaultFS**</name>  <value>hdfs://masternode01:9000</value>  </property>  </configuration> |
| --- |

**9.2 hdfs-site.xml dosyasının içeriği aş. Gibi düzenlenir.**

| [hadoop@masternode01 hadoop]$ cat hdfs-site.xml  <!-- Put site-specific property overrides in this file. -->  <configuration>  <property>  <name>**dfs.replication**</name>  <value>1</value>  </property>  <property>  <name>**dfs.namenode.name.dir**</name>  <value>/hadoop/hdfs/name</value>  </property>  <property>  <name>**dfs.datanode.data.dir**</name>  <value>/hadoop/hdfs/data</value>  </property>  </configuration> |
| --- |

**9.3 mapred-site.xml dosyasının içeriği aş. Gibi düzenlenir.**

| [hadoop@masternode01 hadoop]$ cat mapred-site.xml  <!-- Put site-specific property overrides in this file. -->  <configuration>  <property>  <name>**mapreduce.framework.name**</name>  <value>yarn</value>  </property>  </configuration> |
| --- |

**9.4 yarn-site.xml dosyasının içeriği aş. Gibi düzenlenir.**

| <configuration>  <property>  <name>**yarn.nodemanager.aux-services**</name>  <value>mapreduce\_shuffle</value>  </property>  <property>  <name>**yarn.resourcemanager.hostname**</name>  <value>masternode01</value>  </property>  <property>  <name>**yarn.nodemanager.local-dir**</name>  <value>/hadoop/yarn/local</value>  </property>  </configuration> |
| --- |

1. **3 sunucuda da HDFS namenode ve datanode’un kullanacağı dizinler aşağıdaki gibi oluşturulur.**

| [yyukselveysel44@masternode01 ~]$ sudo mkdir -p /hadoop/hdfs/name  [yyukselveysel44@masternode01 ~]$ sudo mkdir -p /hadoop/hdfs/data  [yyukselveysel44@masternode01 ~]$ sudo mkdir -p /hadoop/yarn/local  [yyukselveysel44@masternode01 ~]$ sudo chown -R hadoop:hadoop /hadoop |
| --- |

1. **Masternode01 üzerindeki /cluster/hadoop klasörü slavenode01 ve slavenode02 ye /cluster dizini altına kopyalanır.**

| [hadoop@masternode01 ~]$ scp -r /cluster/hadoop hadoop@slavenode01:/cluster/.  [hadoop@masternode01 ~]$ scp -r /cluster/hadoop hadoop@slavenode02:/cluster/. |
| --- |

1. **Slavenode01 ve slavenode02 sunucularında /cluster/hadoop/etc/hadoop dizinindeki hdfs-site.xml ve yarn-site.xml dosyalarının içeriği aşağıdaki gibi değiştirilir. Masternode sunucusunda bu dosyaların içeriği namenode ve resourcemanager servislerini çalıştıracak şekilde ayarlanmıştır. Slave node’larda bu dosyaların değiştirilme sebebi budur.**

Hdfs-site.xml

| <configuration>  <property>  <name>dfs.datanode.data.dir</name>  <value>/hadoop/hdfs/data</value>  </property>  </configuration> |
| --- |

Yarn-site.xml

| <configuration>  <property>  <name>yarn.nodemanager.local-dir</name>  <value>/hadoop/yarn/local</value>  </property>  </configuration> |
| --- |

1. **3 sunucuda da /cluster/hadoop/etc/hadoop dizininde yer alan hadoop-env.sh dosyasına aş. JAVA\_HOME değişkeni eklenir.**

| export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.452.b09-2.el8.x86\_64/jre |
| --- |

1. **3 sunucuda da hadoop kullanıcısına geçildikten sonra /home/hadoop altında yer alan .bashrc dosyasına aşağıdaki değişkenler eklenir.**

| export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.452.b09-2.el8.x86\_64/jre  export HADOOP\_HOME=/cluster/hadoop  export PATH=${PATH}:${HADOOP\_HOME}/bin:${HADOOP\_HOME}/sbin |
| --- |

1. **Masternode01 sunucusunda /cluster/hadoop/etc/hadoop altında yer alan workers dosyasının içeriğine slavenode01 ve slavenode02 eklenir.**

| slavenode01  slavenode02 |
| --- |

1. **Masternode01’de hdfs namenode format komutu çalıştırılarak diskin formatlanması sağlanır.**

| [hadoop@masternode01 ~]$ source .bashrc  [hadoop@masternode01 ~]$ hdfs namenode -format  WARNING: /cluster/hadoop/logs does not exist. Creating.  2024-09-02 20:12:13,701 INFO namenode.NameNode: STARTUP\_MSG:  /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  STARTUP\_MSG: Starting NameNode  STARTUP\_MSG: host = masternode01.us-central1-c.c.praxis-life-428819-h1.internal/10.128.0.8  STARTUP\_MSG: args = [-format]  STARTUP\_MSG: version = 3.4.0 |
| --- |

\*\*\*\*\*

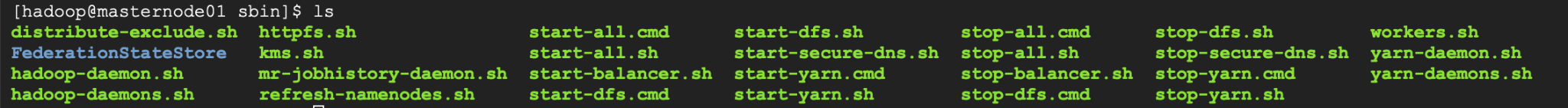
Kurulum işlemleri böylelikle tamamlanmış olur. Artık Hadoop servislerini başlatabiliriz. Start-dfs komutu ile NN ve DN servisleri başlatılır.

| [hadoop@masternode01 ~]$ start-dfs.sh  Starting namenodes on [masternode01]  Starting datanodes  slavenode02: WARNING: /cluster/hadoop/logs does not exist. Creating.  slavenode01: WARNING: /cluster/hadoop/logs does not exist. Creating.  Starting secondary namenodes [masternode01] |
| --- |

Servisler başlatıldıktan sonra jps komutunu kullanarak sunucularda hangi servislerin ayakta olduğunu görebiliriz. Masternode01 üzerinde Namenode ve SecondaryNameNode servisi ayakta iken, slavenode’lar üzerinde ise DataNode servisi ayaktadır.

| [hadoop@masternode01 ~]$ jps  4913 NameNode  5266 Jps  5144 SecondaryNameNode |
| --- |

start-yarn.sh komutu ile de yarn servisleri başlatılır. start-all.sh ile de hem hdfs, hem yarn servisleri başlatılır.



Start-all komutu ile cluster servisleri başlatıldıktan sonra aşağıdaki gibi sunucularda servislerin durumu kontrol edilir.

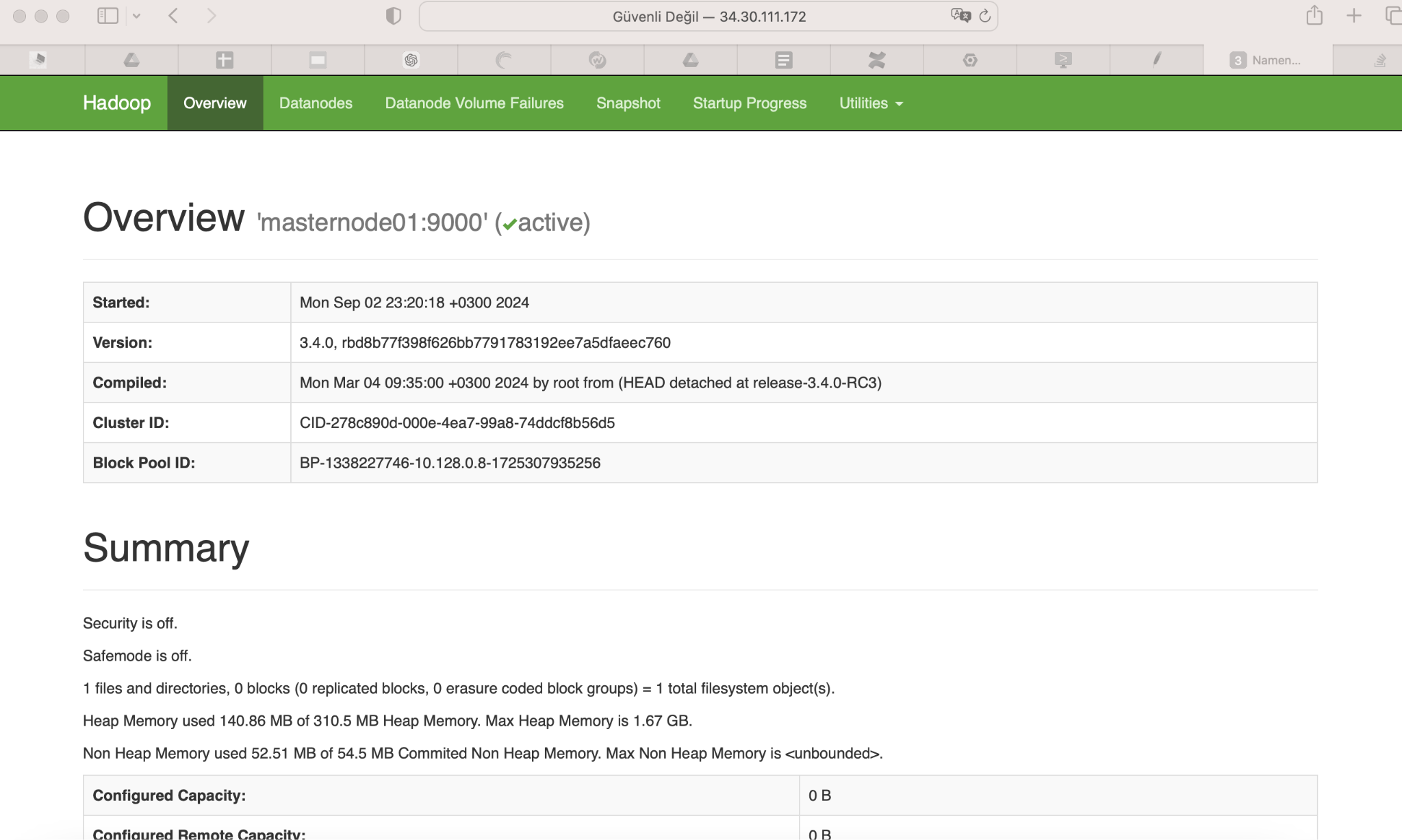
| **[hadoop@masternode01 sbin]$** ./start-all.sh  WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.  WARNING: This is not a recommended production deployment configuration.  WARNING: Use CTRL-C to abort.  Starting namenodes on [masternode01]  Starting datanodes  slavenode01: Warning: Permanently added the ECDSA host key for IP address '34.67.114.150' to the list of known hosts.  slavenode02: Warning: Permanently added the ECDSA host key for IP address '34.44.221.24' to the list of known hosts.  Starting secondary namenodes [masternode01]  Starting resourcemanager  Starting nodemanagers  **[hadoop@masternode01 sbin]$** jps  4032 SecondaryNameNode  4260 ResourceManager  4569 Jps  3791 NameNode  **[hadoop@slavenode01 ~]$** jps  1940 Jps  1690 DataNode  1818 NodeManager |
| --- |

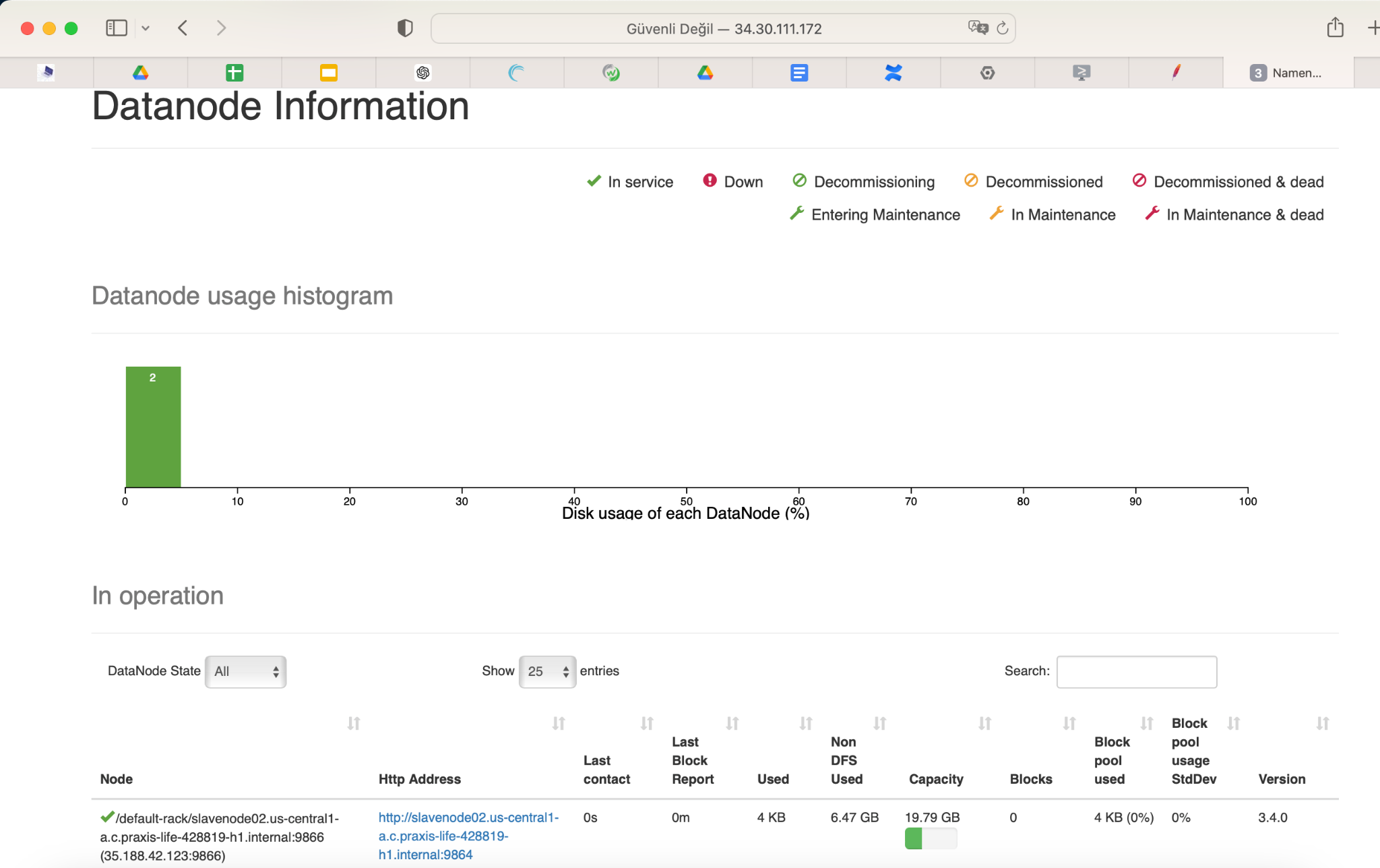
Servisler başladıktan sonra, NameNode ve Resource Manager’ın web arayüzlerine bağlanarak durumları hakkında bilgi alınabilir.

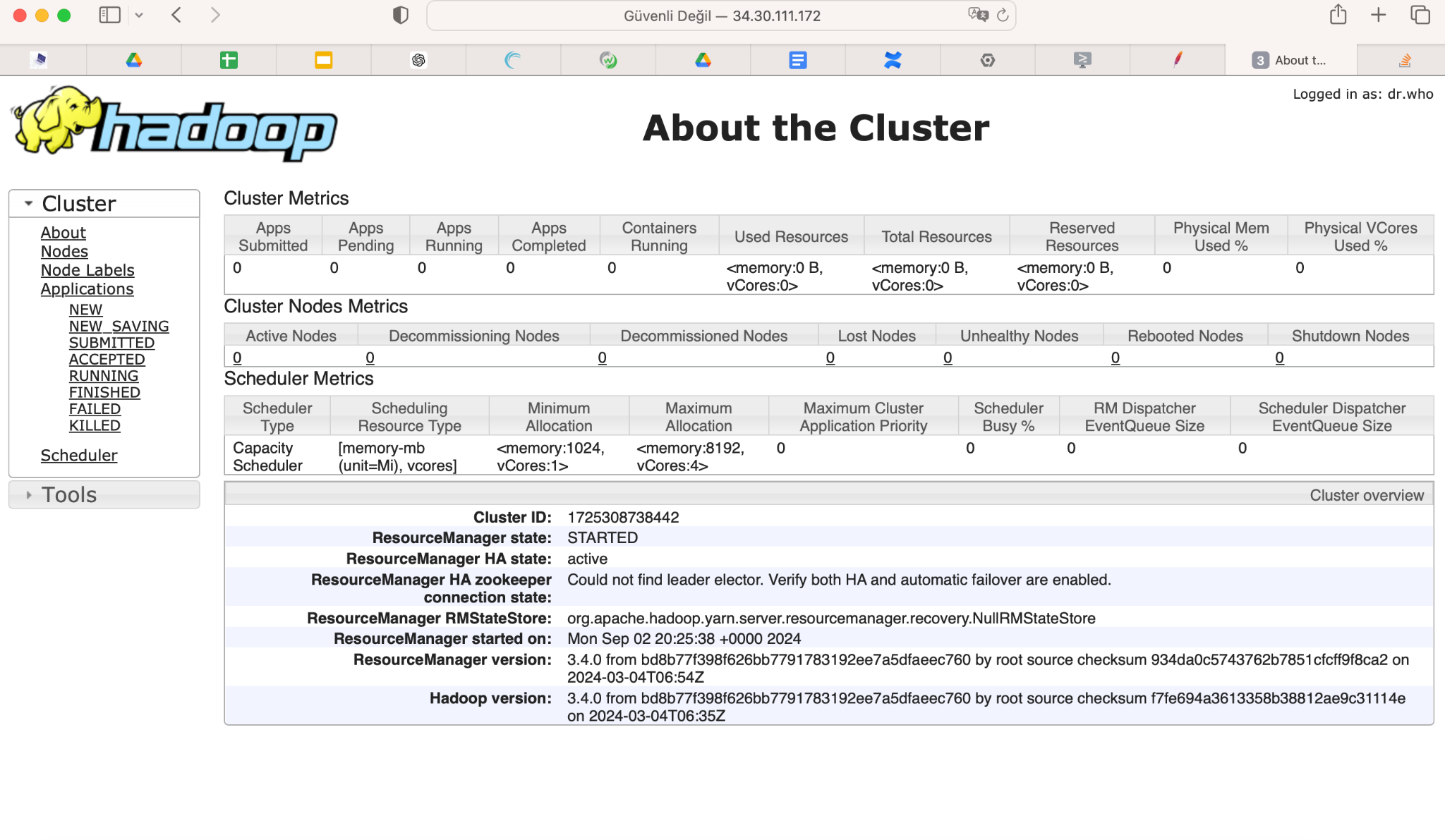
NameNode Web Arayüzü **(Port 9870)** : masternode01:9870

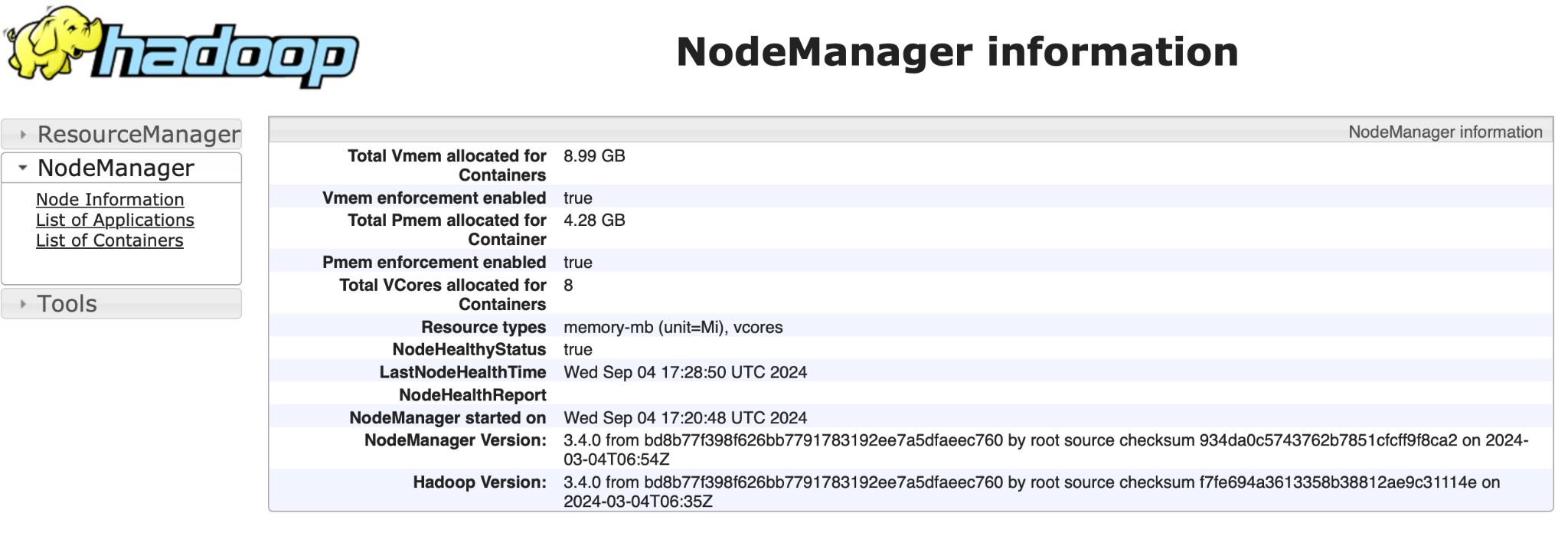
ResourceManager Web Arayüzü **(Port 8088)** : masternode01:8088

Nodemanager Web arayüzü **(Port 8042)**: slavenode01:8042 slavenode02:8042









Clusterdaki tüm servisleri durdurmak için ise stop-all.sh komutu kullanılır.