2019.01.19 (window 유사분산모드)

2019년 1월 21일 월요일   오후 6:32

1. hadoop-2.7.1 설치
   - C:\opt\hadoop-2.7.1
2. Datanode, namenode, namesecondary, tmp 설치
3. 환경변수

시스템 변수 편집   ✕

변수 이름(N): | HADOOP_HOME |

변수 값(V): | C:\opt\hadoop-2.7.1 |

[디렉터리 찾아보기(D)...] [파일 찾아보기(F)...]   [확인]   [취소]

시스템 변수 편집   ✕

변수 이름(N): | JAVA_HOME |

변수 값(V): | C:\Progra~1\Java\jdk1.8.0_181 |

[디렉터리 찾아보기(D)...] [파일 찾아보기(F)...]   [확인]   [취소]

- Path
  %HADOOP_HOME%\bin
  %HADOOP_HOME%\sbin

4. Hadoop 환경 구축

-

```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

 <configuration>
   <property>
     <name>fs.defaultFS</name>
     <value>hdfs://localhost:9000</value>
   </property>
</configuration>
```

- **opt/hadoop-2.7.1/etc/hadoop/mapred-site.xml**

```xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
```

You may obtain a copy of the License at

<!-- Put site-specific property overrides in this file. -->

<configuration>

 <property>

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

  <value>yarn</value>

 </property>

</configuration>

- opt/hadoop-2.7.1/etc/hadoop/hdfs-site.xml

<?xml version="1.0" encoding="UTF-8"?>
<!--

```xml
  limitations under the License.
-->
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>/opt/hadoop-2.7.1/data/namenode</value>
  </property>
  <property>
    <name>dfs.namenode.checkpoint.dir</name>
    <value>/opt/hadoop-2.7.1/data/namesecondary</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/opt/hadoop-2.7.1/data/datanode</value>
  </property>
  <property>
    <name>dfs.http.address</name>
    <value>hadoop0:50070</value>
  </property>
  <property>
    <name>dfs.secondary.http.address</name>
    <value>hadoop1:50090</value>
  </property>
</configuration>
```

- **opt/hadoop-2.7.1/etc/hadoop/yarn-site.xml**

```xml
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0
```

```
<configuration>

<!-- Site specific YARN configuration properties -->
 <property>
   <name>yarn.nodemanager.aux-services</name>
   <value>mapreduce_shuffle</value>
 </property>
 <property>
   <name>yarn.nodemanager.aux-services.mapreduce_shuffle.class</name>
   <value>org.apache.hadoop.mapred.ShuffleHandler</value>
 </property>
<property>
   <name>yarn.nodemanager.local-dirs</name>
   <value>/opt/hadoop-2.7.1/data/tmp</value>
 </property>
</configuration>
```

- **opt/hadoop-2.7.1/etc/hadoop/hadoop-env.cmd**

```
@echo off
@rem Licensed to the Apache Software Foundation (ASF) under one or more
@rem contributor license agreements.  See the NOTICE file distributed with
@rem this work for additional information regarding copyright ownership.
@rem The ASF licenses this file to You under the Apache License, Version 2.0
@rem (the "License"); you may not use this file except in compliance with
@rem the License.  You may obtain a copy of the License at
@rem
@rem     http://www.apache.org/licenses/LICENSE-2.0
@rem
@rem Unless required by applicable law or agreed to in writing, software
@rem distributed under the License is distributed on an "AS IS" BASIS,
@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
@rem See the License for the specific language governing permissions and
```

```
@rem limitations under the License.

@rem Set Hadoop-specific environment variables here.

@rem The only required environment variable is JAVA_HOME.  All others are
@rem optional.  When running a distributed configuration it is best to
@rem set JAVA_HOME in this file, so that it is correctly defined on
@rem remote nodes.

@rem The java implementation to use.  Required.
set JAVA_HOME=%JAVA_HOME%

@rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
@rem set JSVC_HOME=%JSVC_HOME%

@rem set HADOOP_CONF_DIR=

@rem Extra Java CLASSPATH elements.  Automatically insert capacity-scheduler.
if exist %HADOOP_HOME%\contrib\capacity-scheduler (
  if not defined HADOOP_CLASSPATH (
    set HADOOP_CLASSPATH=%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
  ) else (
    set HADOOP_CLASSPATH=%HADOOP_CLASSPATH%;%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
  )
)

@rem The maximum amount of heap to use, in MB. Default is 1000.
@rem set HADOOP_HEAPSIZE=
@rem set HADOOP_NAMENODE_INIT_HEAPSIZE=""

@rem Extra Java runtime options.  Empty by default.
@rem set HADOOP_OPTS=%HADOOP_OPTS% -Djava.net.preferIPv4Stack=true

@rem Command specific options appended to HADOOP_OPTS when specified
if not defined HADOOP_SECURITY_LOGGER (
  set HADOOP_SECURITY_LOGGER=INFO,RFAS
)
if not defined HDFS_AUDIT_LOGGER (
  set HDFS_AUDIT_LOGGER=INFO,NullAppender
)
```

```
set HADOOP_NAMENODE_OPTS=-Dhadoop.security.logger=%HADOOP_SECURITY_LOGGER% -Dhdfs.audit.logger=%HDFS_AUDIT_LOGGER% %
HADOOP_NAMENODE_OPTS%
set HADOOP_DATANODE_OPTS=-Dhadoop.security.logger=ERROR,RFAS %HADOOP_DATANODE_OPTS%
set HADOOP_SECONDARYNAMENODE_OPTS=-Dhadoop.security.logger=%HADOOP_SECURITY_LOGGER% -Dhdfs.audit.logger=%HDFS_AUDIT_LOGGER% %
HADOOP_SECONDARYNAMENODE_OPTS%

@rem The following applies to multiple commands (fs, dfs, fsck, distcp etc)
set HADOOP_CLIENT_OPTS=-Xmx512m %HADOOP_CLIENT_OPTS%
@rem set HADOOP_JAVA_PLATFORM_OPTS="-XX:-UsePerfData %HADOOP_JAVA_PLATFORM_OPTS%"

@rem On secure datanodes, user to run the datanode as after dropping privileges
set HADOOP_SECURE_DN_USER=%HADOOP_SECURE_DN_USER%

@rem Where log files are stored.  %HADOOP_HOME%/logs by default.
@rem set HADOOP_LOG_DIR=%HADOOP_LOG_DIR%\%USERNAME%

@rem Where log files are stored in the secure data environment.
set HADOOP_SECURE_DN_LOG_DIR=%HADOOP_LOG_DIR%\%HADOOP_HDFS_USER%

@rem The directory where pid files are stored. /tmp by default.
@rem NOTE: this should be set to a directory that can only be written to by
@rem       the user that will run the hadoop daemons.  Otherwise there is the
@rem       potential for a symlink attack.
set HADOOP_PID_DIR=%HADOOP_PID_DIR%
set HADOOP_SECURE_DN_PID_DIR=%HADOOP_PID_DIR%

@rem A string representing this instance of hadoop. %USERNAME% by default.
set HADOOP_IDENT_STRING=%USERNAME%
```

- **opt/hadoop-2.7.1₩sbin₩start-dfs.cmd**

```
@echo off
@rem Licensed to the Apache Software Foundation (ASF) under one or more
@rem contributor license agreements.  See the NOTICE file distributed with
@rem this work for additional information regarding copyright ownership.
@rem The ASF licenses this file to You under the Apache License, Version 2.0
@rem (the "License"); you may not use this file except in compliance with
@rem the License.  You may obtain a copy of the License at
```

```
@rem
@rem     http://www.apache.org/licenses/LICENSE-2.0
@rem
@rem Unless required by applicable law or agreed to in writing, software
@rem distributed under the License is distributed on an "AS IS" BASIS,
@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
@rem See the License for the specific language governing permissions and
@rem limitations under the License.
@rem
setlocal enabledelayedexpansion

if not defined HADOOP_BIN_PATH (
  set HADOOP_BIN_PATH=%~dp0
)

if "%HADOOP_BIN_PATH:~-1%" == "\" (
  set HADOOP_BIN_PATH=%HADOOP_BIN_PATH:~0,-1%
)

set DEFAULT_LIBEXEC_DIR=%HADOOP_BIN_PATH%\..\libexec
if not defined HADOOP_LIBEXEC_DIR (
  set HADOOP_LIBEXEC_DIR=%DEFAULT_LIBEXEC_DIR%
)

call %HADOOP_LIBEXEC_DIR%\hdfs-config.cmd %*
if "%1" == "--config" (
  shift
  shift
)

start "Apache Hadoop Distribution" hadoop namenode
start "Apache Hadoop Distribution" hadoop datanode
start "Apache Hadoop Distribution" hadoop secondarynamenode

endlocal
```

- **opt/hadoop-2.7.1₩sbin₩stop-dfs.cmd**

```
@echo off
```

```
@rem Licensed to the Apache Software Foundation (ASF) under one or more
@rem contributor license agreements.  See the NOTICE file distributed with
@rem this work for additional information regarding copyright ownership.
@rem The ASF licenses this file to You under the Apache License, Version 2.0
@rem (the "License"); you may not use this file except in compliance with
@rem the License.  You may obtain a copy of the License at
@rem
@rem    http://www.apache.org/licenses/LICENSE-2.0
@rem
@rem Unless required by applicable law or agreed to in writing, software
@rem distributed under the License is distributed on an "AS IS" BASIS,
@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
@rem See the License for the specific language governing permissions and
@rem limitations under the License.
@rem
setlocal enabledelayedexpansion

if not defined HADOOP_BIN_PATH (
  set HADOOP_BIN_PATH=%~dp0
)

if "%HADOOP_BIN_PATH:~-1%" == "\" (
  set HADOOP_BIN_PATH=%HADOOP_BIN_PATH:~0,-1%
)

set DEFAULT_LIBEXEC_DIR=%HADOOP_BIN_PATH%\..\libexec
if not defined HADOOP_LIBEXEC_DIR (
  set HADOOP_LIBEXEC_DIR=%DEFAULT_LIBEXEC_DIR%
)

call %HADOOP_LIBEXEC_DIR%\hadoop-config.cmd %*
if "%1" == "--config" (
  shift
  shift
)

Taskkill /FI "WINDOWTITLE eq Apache Hadoop Distribution - hadoop   namenode"
Taskkill /FI "WINDOWTITLE eq Apache Hadoop Distribution - hadoop   datanode"
Taskkill /FI "WINDOWTITLE eq Apache Hadoop Distribution - hadoop   secondarynamenode"
```
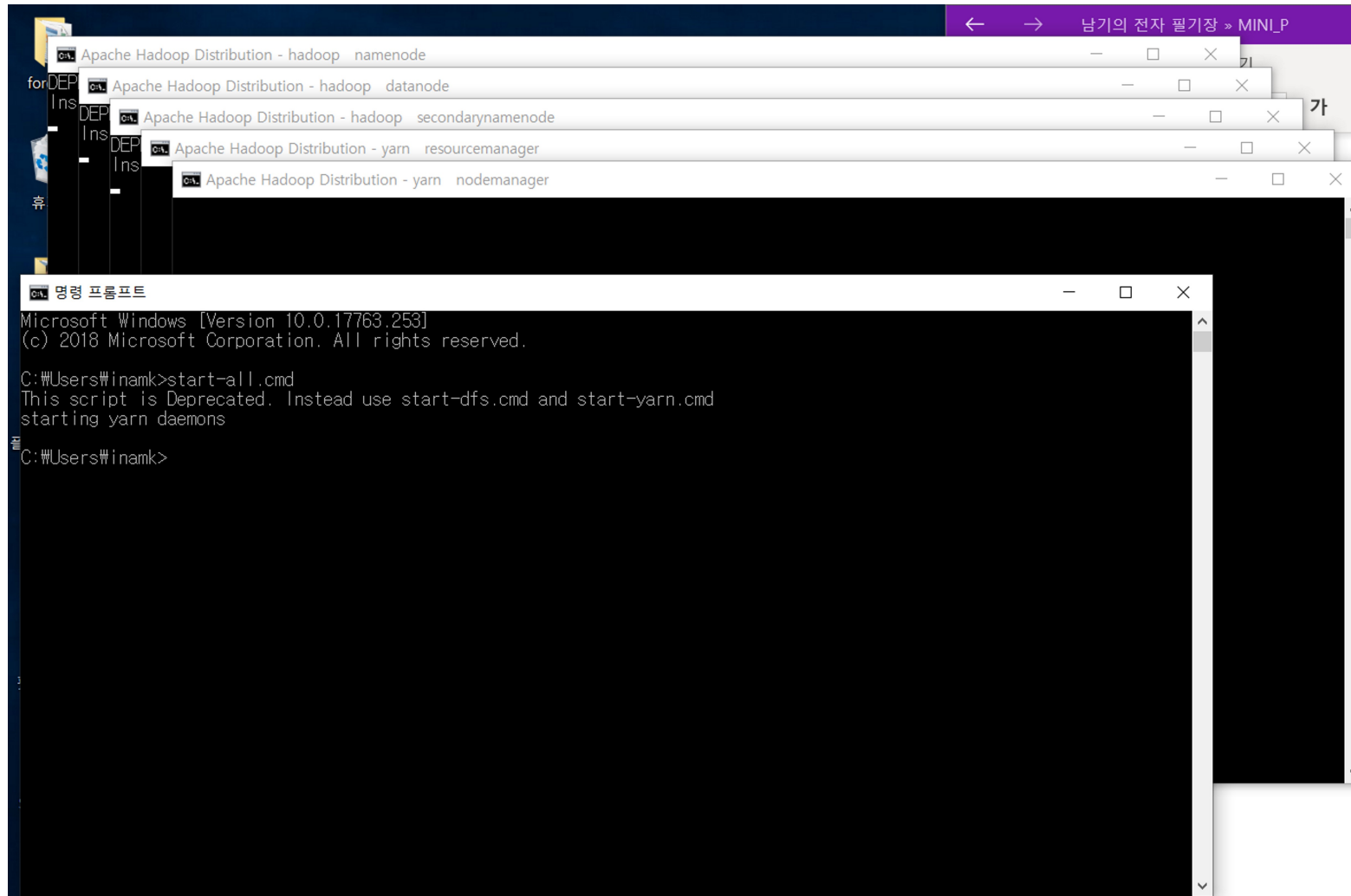
endlocal

5. Namenode 포멧

- C:₩> hdfs namenode -format
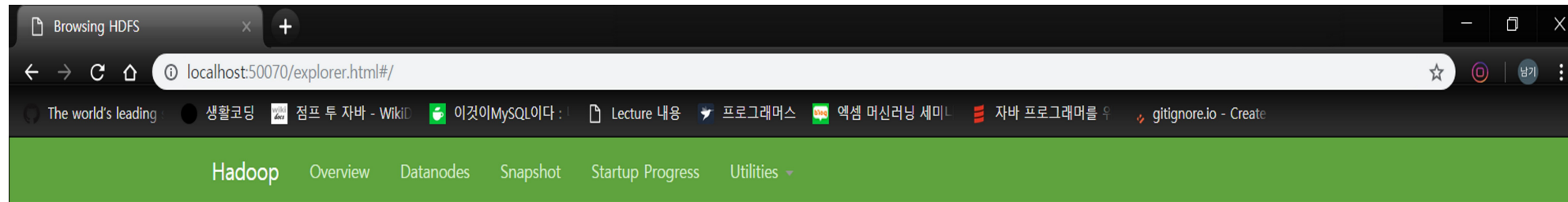
6. Hadoop실행

- start-all.cmd

- C:₩> jps

-
```
C:\Users\inamk>jps
21604 Jps
23140 NameNode
14664 DataNode
2040 ResourceManager
10460 SecondaryNameNode
15804 NodeManager
```

- localhost:50070

-

Browse Directory

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	inamk	supergroup	0 B	2019. 1. 20. 오전 1:58:59	0	0 B	input

Hadoop, 2015.