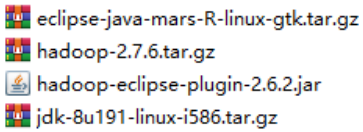
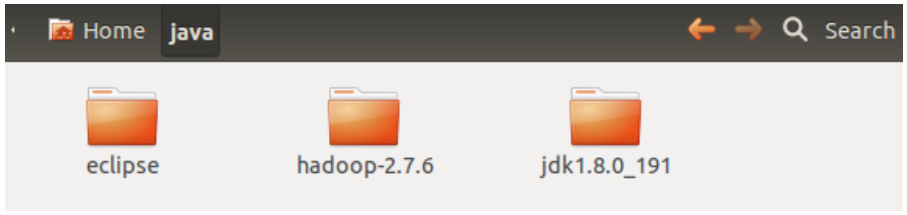
Ubuntu 12.04上安装Hadoop并运行

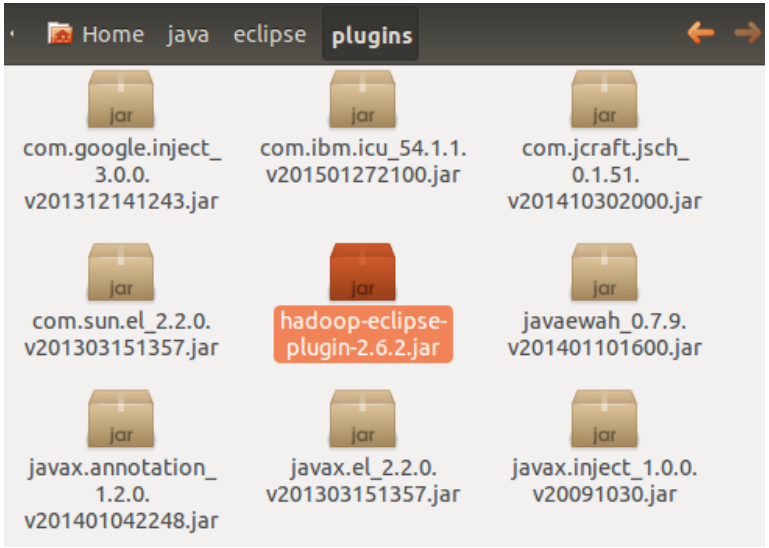
作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

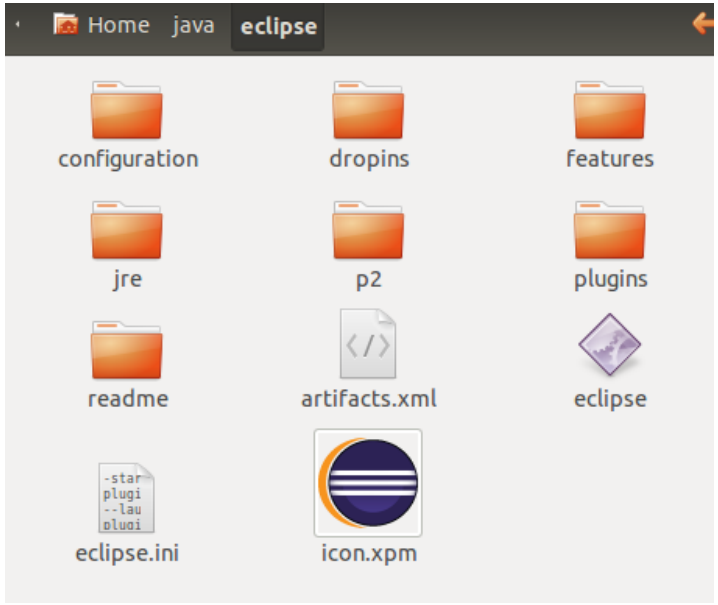
在官网上下载好四个文件



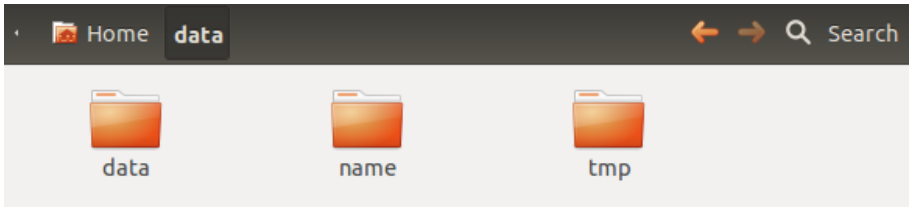
在Ubuntu的/home/wrr/下创建一个文件夹java，将这四个文件拷到Ubuntu的/home/wrr/java/下，将eclipse、hadoop-2.7.6与jdk进行解压，将.iar文件复制到eclipse/plugins，将/jdk1.8.0_191/下的jre文件夹拷到eclipse中，如下图所示







在/home/wrr/下新建文件夹data，里面新建三个文件夹data、name与tmp，创建文件夹的命令：sudo mkdir java



添加环境变量

```
sudo gedit ~/.bashrc
```

打开.bashrc，在后面添加环境变量

```
export JAVA_HOME=/home/wrr/java/jdk1.8.0_191
export JRE_HOME=/home/wrr/java/jdk1.8.0_191/jre
export CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar:$JAVA_HOME/lib
export PATH=$PATH:$JAVA_HOME/bin:$JRE_HOME/bin

export HADOOP_HOME=/home/wrr/java/hadoop-2.7.6
export PATH=$PATH:$HADOOP_HOME/bin
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
```

再键入

```
source ~/.bashrc
```

此时环境变量添加成功，现在输入hadoop version 与 java -version来查看版本。

```
wrr@ubuntu:~$ hadoop version
Hadoop 2.7.6
Subversion https://shv@git-wip-us.apache.org/repos/asf/hadoop.git -r 085099c66cf
28be31604560c376fa282e69282b8
Compiled by kshvachk on 2018-04-18T01:33Z
Compiled with protoc 2.5.0
From source with checksum 71e2695531cb3360ab74598755d036
This command was run using /home/wrr/java/hadoop-2.7.6/share/hadoop/common/hadoo
p-common-2.7.6.jar
wrr@ubuntu:~$ java -version
java version "1.8.0_191"
Java(TM) SE Runtime Environment (build 1.8.0_191-b12)
Java HotSpot(TM) Client VM (build 25.191-b12, mixed mode)
```

配置/home/wrr/java/hadoop-2.7.6/etc/hadoop下的集群参数

hadoop-env.sh

```
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements.  See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership.  The ASF licenses this file
# to you 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.

# Set Hadoop-specific environment variables here.

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

# The java implementation to use.
export JAVA_HOME=/home/wrr/java/jdk1.8.0_191

# The jsvc implementation to use. Jsvc is required to run secure datanodes
# that bind to privileged ports to provide authentication of data transfer
# protocol.  Jsvc is not required if SASL is configured for authentication of
# data transfer protocol using non-privileged ports.
#export JSVC_HOME=${JSVC_HOME}

export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/etc/hadoop"}

# Extra Java CLASSPATH elements.  Automatically insert capacity-scheduler.
for f in $HADOOP_HOME/contrib/capacity-scheduler/*.jar; do
  if [ "$HADOOP_CLASSPATH" ]; then
    export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$f
  else
```

```
    export HADOOP_CLASSPATH=$f
  fi
done

# The maximum amount of heap to use, in MB. Default is 1000.
#export HADOOP_HEAPSIZE=
#export HADOOP_NAMENODE_INIT_HEAPSIZE=""

# Extra Java runtime options.  Empty by default.
export HADOOP_OPTS="$HADOOP_OPTS -Djava.net.preferIPv4Stack=true"

# Command specific options appended to HADOOP_OPTS when specified
export HADOOP_NAMENODE_OPTS="-Dhadoop.security.logger=${HADOOP_SECURITY_LOGGER:-INFO,RFAS} -Dhdfs.audit.logger=${HDFS_AUDIT_LOGGER:-INFO,NullAppender} $HADOOP_NAMENODE_OPTS"
export HADOOP_DATANODE_OPTS="-Dhadoop.security.logger=ERROR,RFAS $HADOOP_DATANODE_OPTS"

export HADOOP_SECONDARYNAMENODE_OPTS="-Dhadoop.security.logger=${HADOOP_SECURITY_LOGGER:-INFO,RFAS} -Dhdfs.audit.logger=${HDFS_AUDIT_LOGGER:-INFO,NullAppender} $HADOOP_SECONDARYNAMENODE_OPTS"

export HADOOP_NFS3_OPTS="$HADOOP_NFS3_OPTS"
export HADOOP_PORTMAP_OPTS="-Xmx512m $HADOOP_PORTMAP_OPTS"

# The following applies to multiple commands (fs, dfs, fsck, distcp etc)
export HADOOP_CLIENT_OPTS="-Xmx512m $HADOOP_CLIENT_OPTS"
#HADOOP_JAVA_PLATFORM_OPTS="-XX:-UsePerfData $HADOOP_JAVA_PLATFORM_OPTS"

# On secure datanodes, user to run the datanode as after dropping privileges.
# This **MUST** be uncommented to enable secure HDFS if using privileged ports
# to provide authentication of data transfer protocol.  This **MUST NOT** be
# defined if SASL is configured for authentication of data transfer protocol
# using non-privileged ports.
export HADOOP_SECURE_DN_USER=${HADOOP_SECURE_DN_USER}

# Where log files are stored.  $HADOOP_HOME/logs by default.
#export HADOOP_LOG_DIR=${HADOOP_LOG_DIR}/$USER

# Where log files are stored in the secure data environment.
export HADOOP_SECURE_DN_LOG_DIR=${HADOOP_LOG_DIR}/${HADOOP_HDFS_USER}

###
# HDFS Mover specific parameters
###
# Specify the JVM options to be used when starting the HDFS Mover.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HADOOP_MOVER_OPTS=""

###
# Advanced Users Only!
###

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

# A string representing this instance of hadoop. $USER by default.
export HADOOP_IDENT_STRING=$USER
```

core-site.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>hadoop.tmp.dir</name>
                <value>/home/wrr/data/tmp</value>
                <description>A base for other temporary directories.</description>
        </property>
        <property>
                <name>io.file.buffer.size</name>
                <value>131072</value>
        </property>
        <property>
                <name>fs.default.name</name>
                <value>hdfs://localhost:9000</value>
        </property>
        <property>
                <name>hadoop.proxyuser.root.hosts</name>
                <value>*</value>
        </property>
        <property>
                <name>hadoop.proxyuser.root.groups</name>
                <value>*</value>
        </property>
</configuration>
```

hdfs-site.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>dfs.replication</name>
                <value>3</value>
        </property>
        <property>
                <name>dfs.namenode.name.dir</name>
                <value>/home/wrr/data/name</value>
                <final>true</final>
        </property>
        <property>
                <name>dfs.datanode.data.dir</name>
                <value>/home/wrr/data/data</value>
                <final>true</final>
        </property>
        <property>
                <name>dfs.namenode.secondary.httpaddress</name>
                <value>localhost:9001</value>
        </property>
        <property>
                <name>dfs.webhdfs.enabled</name>
                <value>true</value>
        </property>
        <property>
                <name>dfs.permissions</name>
                <value>false</value>
        </property>
</configuration>
```

mapred-site.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


    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>mapreduce.framework.name</name>

        <value>yarn</value>

    </property>

</configuration>
```

yarn-site.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

  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.
-->
<configuration>

<!-- Site specific YARN configuration properties -->

        <property>
                <name>yarn.resourcemanager.address</name>
                <value>localhost:18040</value>
        </property>
        <property>
                <name>yarn.resourcemanager.scheduler.address</name>
                <value>localhost:18030</value>
        </property>
        <property>
                <name>yarn.resourcemanager.webapp.address</name>
                <value>localhost:18088</value>
        </property>
        <property>
                <name>yarn.resourcemanager.resourcetracker.address</name>
                <value>localhost:18025</value>
        </property>
        <property>
```

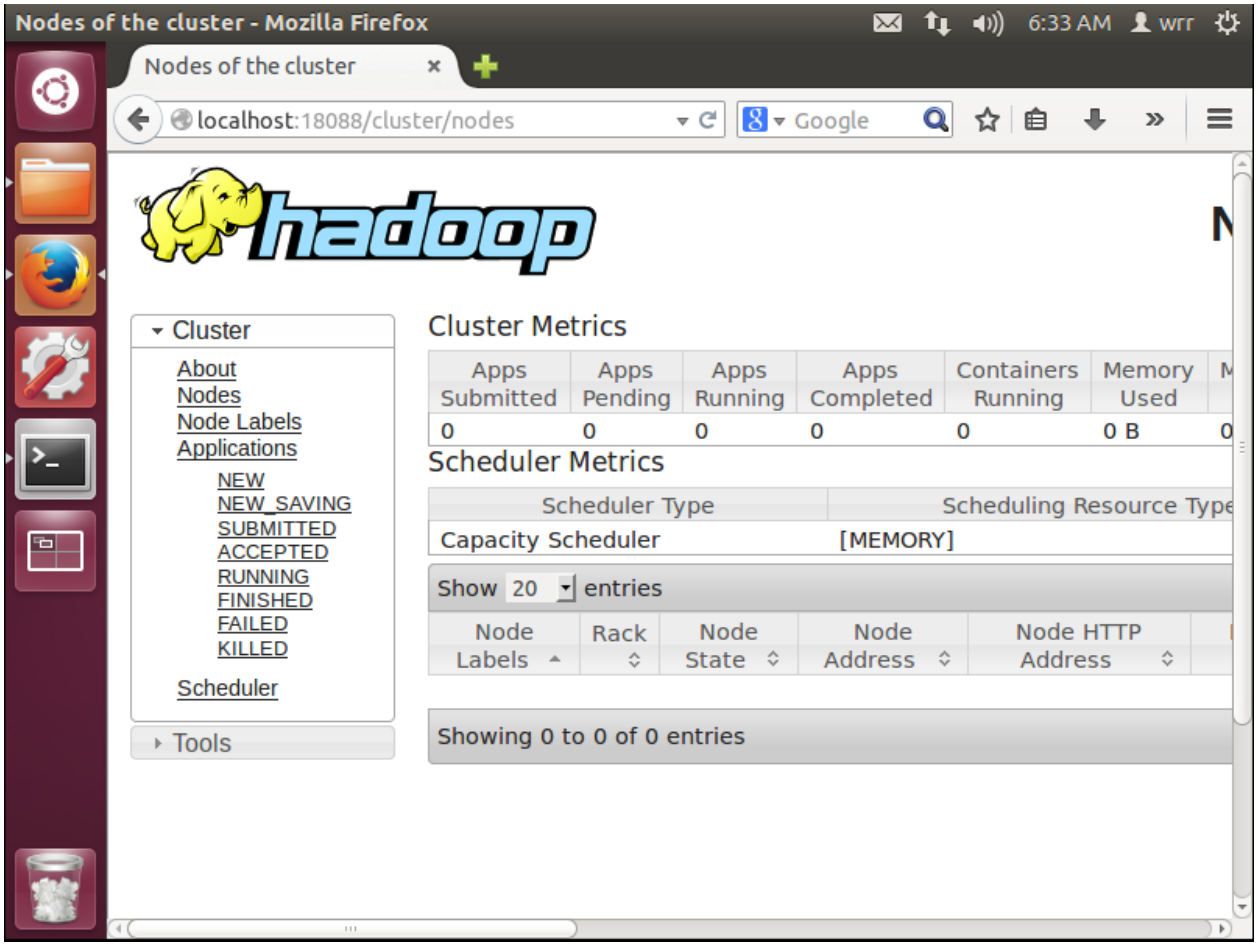
```
                <name>yarn.resourcemanager.admin.address</name>
                <value>localhost:18141</value>
        </property>
        <property>
                <name>yarn.nodemanager.auxservices</name>
                <value>mapreduce.shuffle</value>
        </property>
        <property>
                <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
                <value>org.apache.hadoop.mapred.ShuffleHandler</value>
        </property>

</configuration>
```

启动hadoop。首先导入/home/wrr/java/hadoop-2.7.6/sbin一下目录，再启动namenode，datanode与yarn

```
wrr@ubuntu:~$ cd /home/wrr/java/hadoop-2.7.6/sbin
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ jps
6559 Jps
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ ./hadoop-daemon.sh start datanode
starting datanode, logging to /home/wrr/java/hadoop-2.7.6/logs/hadoop-wrr-datanode-ubuntu.out
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/wrr/java/hadoop-2.7.6/logs/yarn-wrr-resourcemanager-ubuntu.out
localhost: ssh: connect to host localhost port 22: Connection refused
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ ./hadoop-daemon.sh start namenode
starting namenode, logging to /home/wrr/java/hadoop-2.7.6/logs/hadoop-wrr-namenode-ubuntu.out
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ jps
6978 NameNode
6692 ResourceManager
7013 Jps
6587 DataNode
```

都启动之后，在浏览器上输入http://localhost:18088，即可出现如下界面



停止hadoop

```
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ ./hadoop-daemon.sh stop namenode
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ ./hadoop-daemon.sh stop datanode
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ ./stop-yarn.sh
wrr@ubuntu:~/java/hadoop-2.7.6/sbin$ jps
7259 Jps
```

如果想看更详细的解读，请看[Hadoop安装教程_单机/伪分布式配置_Hadoop2.6.0/Ubuntu14.04_厦大数据库实验室博客]