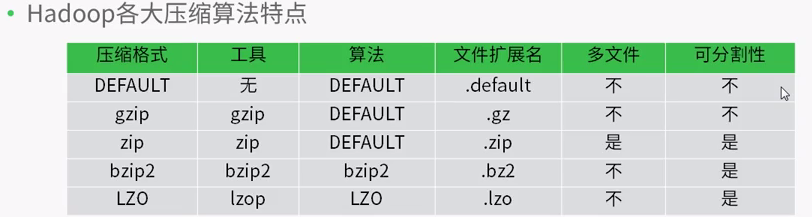
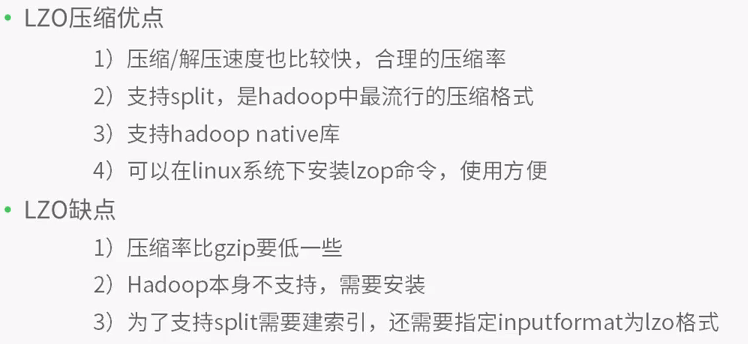
**Hadoop压缩Lzo**







安装配置lzo:5

下载Lzo安装包：

#wget http://www.oberhumer.com/opensource/lzo/download/lzo-2.06.tar.gz

解压：

# tar -zxvf lzo-2.06.tar.gz

# cd lzo-2.06/

编译：

export CFLAGS=-m64

./configure --enable-shared -prefix=/usr/local/lzo

make

make install

安装hadoop-lzo

安装一下系统的依赖关系：

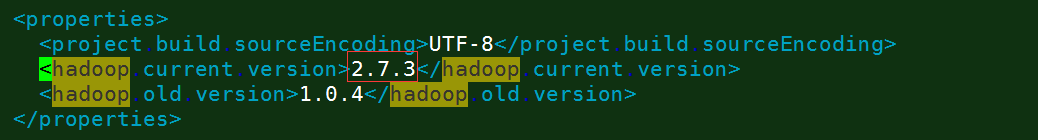
yum -y install ant

wget <https://github.com/twitter/hadoop-lzo/archive/master.zip>

unzip master.zip

# cd hadoop-lzo-master/

修改pom.xml文件，修改对应的hadoop的版本号：



修改完成后进行编译安装：

# export CFLAGS=-m64

# export CXXFLAGS=-m64

# export C\_INCLUDE\_PATH=/usr/local/lzo/include/

# export LIBRARY\_PATH=/usr/local/lzo/lib/

使用mvn进行编译，后面的参数值跳过测试的意思

# mvn clean package -Dmaven.test.skip=true

编译完成后依次拷贝一下依赖：

# cd target/native/Linux-amd64-64

# tar -cBf - -C lib . | tar -xBvf - -C ~

# cp libgplcompression.la /opt/hadoop-2.7.3/lib/native/

# cd /root/source/hadoop-lzo-master/target/

# cp hadoop-lzo-0.4.21-SNAPSHOT.jar /opt/hadoop-2.7.3/share/hadoop/common/

下一步就是整合hadoop-lzo到我们的hadoop软件中：

需要修改付下配置文件：

1、在Hadoop中的$HADOOP\_HOME/etc/hadoop/hadoop-env.sh加上下面配置：

export LD\_LIBRARY\_PATH=/usr/local/lzo/lib/

2、在$HADOOP\_HOME/etc/hadoop/core-site.xml加上如下配置：

<!-- 指定hadoop可用的压缩算法 -->

<property>

<name>io.compression.codecs</name>

<value>

org.apache.hadoop.io.compress.GzipCodec,

org.apache.hadoop.io.compress.DefaultCodec,

com.hadoop.compression.lzo.LzoCodec,

com.hadoop.compression.lzo.LzopCodec,

org.apache.hadoop.io.compress.BZip2Codec

</value>

</property>

<!-- 指定Lzo压缩使用的类 -->

<property>

<name>io.compression.codec.lzo.class</name>

<value>com.hadoop.compression.lzo.LzoCodec</value>

</property>

3、在$HADOOP\_HOME/etc/hadoop/mapred-site.xml加上如下配置

<!-- 开启map输出压缩 -->

<property>

<name>mapreduce.map.output.compress</name>

<value>true</value>

</property>

<!-- 指定map输出的压缩格式 -->

<property>

<name>mapreduce.map.output.compress.codec</name>

<value>com.hadoop.compression.lzo.LzoCodec</value>

</property>

<!-- 开启reduce输出 -->

<property>

<name>mapreduce.output.fileoutputformat.compress</name>

<value>true</value>

</property>

<!-- 指定reduce输出的压缩格式 -->

<property>

<name>mapreduce.output.fileoutputformat.compress.codec</name>

<value>com.hadoop.compression.lzo.LzoCodec</value>

</property>

<!-- 添加lzo本地的本地库 -->

<property>

<name>mapred.child.env</name>

<value>LD\_LIBRARY\_PATH= /usr/local/lzo/lib/</value>

</property>

启动hadoop服务：

# sbin/start-dfs.sh

# sbin/start-yarn.sh

通过登录web界面核实一下服务启动的具体状态

运行命令

# bin/yarn jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar wordcount /words /test01

核实一下输出，发现我们输出是带有压缩的证明我们的配置是成功的

