# oozie搭建教程

## 简介

Apache Oozie是Hadoop工作流调度框架。它是一个运行相关的作业工作流系统。这里，用户被允许创建向非循环图工作流程，其可以在并列 Hadoop 并顺序地运行。

它由两部分组成：

工作流引擎：一个工作流引擎的职责是存储和运行工作流程，由 Hadoop 作业组成：MapReduce, Pig, Hive. 协调器引擎：它运行基于预定义的时间表和数据的可用性工作流程作业。 Oozie可扩展性和可管理及时执行成千上万的工作流程(每个由几十个作业)的Hadoop集群。

Oozie 也非常灵活。人们可以很容易启动，停止，暂停和重新运行作业。Oozie 可以很容易地重新运行失败的工作流。可以很容易重做因宕机或故障错过或失败的作业。甚至有可能跳过一个特定故障节点。

## 部署

### 编译

Unix box (tested on Mac OS X and Linux)Java JDK 1.7+Maven 3.0.1+Hadoop 0.20.2+Pig 0.7+

1.下载及解压

$ wget <http://apache.fayea.com/oozie/4.3.0/oozie-4.3.0.tar.gz>

$ sudo tar xvf oozie-4.3.0.tar.gz

修改解压的目录名oozie=4.3.0为oozie

sudo mv ./oozie-4.3.0 ./oozie

2.编译

进入解压后的目录 oozie

$ mvn clean **package** assembly:single -DskipTests

在国内的同学可以将中央仓库设成阿里云的地址，这样下载速度能快一点，部分下载不下来的，建议手动在mvnrepository.com中央仓库下载一下

### 配置

编译好的文件在以下路径

$ cd distro/target/oozie-4.3.0-distro/oozie-4.3.0/

**1.环境变量配置**

$ sudo **vim** /etc/**profile**

在profile文件中添加以下内容：

export OOZIE\_HOME=/user/local/oozie/distro/target/oozie-4.3.0-distro/oozie-4.3.0

export PATH=$PATH:$OOZIE\_HOME/bin

执行以下命令生效

$ **source** /etc/profile

**2.hadoop 集群集成**

修改core-site.xml，添加以下信息，该教程全程采用的是root用户，因此配置中填入的是root,如果是其他用户需要改成其他用户，**切记**，不然会报没有权限的错误的。修改该文件后，还需要重启hadoop集群，以便参数生效。

<property>

<name>hadoop.proxyuser.ubuntu.groups</name>

<value>**\***</value></property>

<property>

<name>hadoop.proxyuser.ubuntu.hosts</name>

<value>**\***</value></property>

**修改yarn-site.xml**

**<property>**

**<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>**

**<value>org.apache.hadoop.mapred.ShuffleHandler</value>**

**</property>**

**<property>**

**<name>yarn.resourcemanager.address</name>**

**<value>master:8032</value>**

**</property>**

**<property>**

**<name>yarn.resourcemanager.scheduler.address</name>**

**<value>master:8030</value>**

**</property>**

**<property>**

**<name>yarn.resourcemanager.resource-tracker.address</name>**

**<value>master:8031</value>**

**</property>**

**<property>**

**<name>yarn.resourcemanager.admin.address</name>**

**<value>master:8033</value>**

**</property>**

**<property>**

**<name>yarn.resourcemanager.webapp.address</name>**

**<value>master:8088</value>**

**</property>**

**<property>**

**<name>yarn.resourcemanager.webapp.https.address</name>**

**<value>master:8090</value>**

**</property>**

**<property>**

**<name>yarn.application.classpath</name>**

**<value>**

**/usr/local/hadoop/etc/hadoop,**

**/usr/local/hadoop/share/hadoop/common/lib/\*,**

**/usr/local/hadoop/share/hadoop/common/\*,**

**/usr/local/hadoop/share/hadoop/hdfs,**

**/usr/local/hadoop/share/hadoop/hdfs/lib/\*,**

**/usr/local/hadoop/share/hadoop/hdfs/\*,**

**/usr/local/hadoop/share/hadoop/yarn/lib/\*,**

**/usr/local/hadoop/share/hadoop/yarn/\*,**

**/usr/local/hadoop/share/hadoop/mapreduce/lib/\*,**

**/usr/local/hadoop/share/hadoop/mapreduce/\*,**

**</value>**

**</property>**

**3.oozie配置**

以下配置默认在oozie/distro/target/oozie-4.3.0-distro/oozie-4.3.0/ 目录下操作

* 配置文件修改

修改conf目录下oozie-site.xml文件，默认配置可以运行，但是运行hadoop job会报错的。主要是将hadoop集群的配置信息导入oozie中

<property>

<name>oozie.service.HadoopAccessorService.hadoop.configurations</name>

<value>\*=/usr/local/hadoop/etc/hadoop</value>

<description>

Comma separated AUTHORITY=HADOOP\_CONF\_DIR, where AUTHORITY is the HOST:PORT of

the Hadoop service (JobTracker, HDFS). The wildcard '\*' configuration is

used when there is no exact match **for** an authority. The HADOOP\_CONF\_DIR contains

the relevant Hadoop \*-site.xml files. If the path is relative is looked within

the Oozie configuration directory; though the path can be absolute (i.e. to point

to Hadoop client conf/ directories in the local filesystem.

</description>

</property>

* ext

新建libext目录

$ mkdir libext

$ cd libext

$ sudo wget <http://archive.cloudera.com/gplextras/misc/ext-2.2.zip>

$ cd ..

* 打包

bin/oozie-setup.sh prepare-war

$ cd oozie-server/webapps

解压 oozie.war文件

$ sudo unzip oozie.war –d oozie

$ cd oozie/WEB-INF/**lib**/

删掉所有的2.6.0版本的jar文件

$ sudo rm -rf ./\*-2.6.\*.jar

* 复制依赖jar 到下面路径和libext文件夹下

$ sudo cp -rf /usr/local/hadoop/share/hadoop/\*/**lib**/\*.**jar** ./

$ sudo cp -rf /usr/local/hadoop/share/hadoop/\*/\*.jar ./

删除其中mr1的包，及jsp-api.jar，jasper-compiler-5.5.23.jar

$ cd $OOZIE\_HOME

$ sudo cp -rf ./oozie-server/webapps/oozie/WEB-INF/lib/\* ./libext

* 上传 examples到hadoop hdfs

上传examples及oozie-sharelib-4.3.0

$ tar xvf oozie-examples.tar.gz

$ tar xvf oozie-sharelib-4.3.0.tar.gz

$ hdfs dfs -put examples examples

$ hdfs dfs -put share share

**4.运行**

初始化数据库

$ $OOZIE\_HOME/bin/ooziedb.sh create -sqlfile oozie.sql -run

以下输出即为成功：

Validate DB Connection.

DONE**Check** DB **schema** does **not** exist

DONE**Check** OOZIE\_SYS **table** does **not** exist

DONE**Create** **SQL** **schema**

DONE

DONE**Create** OOZIE\_SYS **table**

DONE

Oozie DB has been created **for** Oozie **version** '4.3.0'

守护进程运行

$ $OOZIE\_HOME/bin/oozied.sh start

验证是否成功

$ $OOZIE\_HOME/bin/oozie admin -oozie http://localhost:11000/oozie -status

输出System mode: NORMAL即表示配置成功，或者在浏览器中打开 <http://localhost:11000/oozie/>

### examples运行

1.执行map-reduce

$ oozie job -oozie http://localhost:11000/oozie -config job.properties -run

2.验证结果

$ oozie job -oozie http://localhost:11000/oozie -info 0000007-161223101553230-oozie-root-W

如下结果则证明成功

Job ID : 0000007-161223101553230-oozie-root-W

------------------------------------------------------------------------------------------------------------------------------------

Workflow Name : map-reduce-wf

App Path : hdfs://192.168.0.105:8020/user/root/examples/apps/map-reduce/workflow.xml

Status : SUCCEEDED

Run : 0

User : root

Group : -

Created : 2016-12-23 08:40 GMT

Started : 2016-12-23 08:40 GMT

Last Modified : 2016-12-23 08:41 GMT

Ended : 2016-12-23 08:41 GMT

CoordAction ID: -

Actions

------------------------------------------------------------------------------------------------------------------------------------ID Status Ext ID Ext Status Err Code

------------------------------------------------------------------------------------------------------------------------------------0000007-161223101553230-oozie-root-W@:start: OK - OK -

------------------------------------------------------------------------------------------------------------------------------------0000007-161223101553230-oozie-root-W@mr-node OK job\_1482454814338\_0025 SUCCEEDED -

------------------------------------------------------------------------------------------------------------------------------------0000007-161223101553230-oozie-root-W@end OK - OK -

------------------------------------------------------------------------------------------------------------------------------------