

Big Data: How can I add Apache Oozie to my hadoop Cloudera CDH 5.7.1 instance (based on GPFS)?

How can I add Apache Oozie to my Hadoop instance?

Apache Oozie is a workflow scheduler system to manage Apache Hadoop jobs. Here we describe how to add Oozie to a pre-existing Hadoop instance "gpfs1", based on Cloudera CDH 5.7.1. We then show how to use it to run Mapreduce jobs

1. Add oozie group/user to head node and Hadoop nodes

Execute the following commands on the active head node and in the chroot environment for the software image(s) used by compute nodes.

```
# /usr/bin/getent group oozie || /usr/sbin/groupadd -r oozie
# /usr/bin/getent passwd oozie || /usr/sbin/useradd --comment "Oozie" --shell /bin/bash -m -r -g
oozie --home /var/run/oozie oozie
```

2. Add stanzas (if needed) in core-site.xml (all Hadoop nodes)

The following two stanzas should be present in core-site.xml

```
<property>
  <name>hadoop.proxyuser.oozie.hosts</name>
  <value>*</value>
</property>

<property>
  <name>hadoop.proxyuser.oozie.groups</name>
  <value>*</value>
</property>
```

If `core-site.xml` does not include the stanzas, they can be added using the following commands, which assume that Hadoop nodes are in 'default' category:

Big Data: How can I add Apache Oozie to my hadoop Cloudera CDH 5.7.1 instance (based on GPFS)?

```
# sed -i.bak 's/<Vconfiguration>/
<property>\n<name>hadoop.proxyuser.oozie.hosts</name>\n
<value>*</value>\n</property>\n\n <property>\n
<name>hadoop.proxyuser.oozie.groups</name>\n
<value>*</value>\n</property>\n\n</configuration>/' /etc/hadoop/gpfs1/core-site.xml
```

```
# pdsh -g category=default "sed -i.bak 's/<Vconfiguration>/<property>\n
<name>hadoop.proxyuser.oozie.hosts</name>\n <value>*</value>\n </property>\n\n
<property>\n <name>hadoop.proxyuser.oozie.groups</name>\n
<value>*</value>\n</property>\n\n</configuration>/' /etc/hadoop/gpfs1/core-site.xml"
```

3. Restart all Hadoop services to apply modifications

4. Download Oozie and unpack it

Execute the following commands as root on the active head node. The Ext-2.2 library is needed by the Oozie web console.

```
# cd /tmp/
# curl -O http://archive.cloudera.com/cdh5/cdh/5/oozie-4.1.0-cdh5.7.1.tar.gz
# curl -O http://archive-primary.cloudera.com/gplextras/misc/ext-2.2.zip
# cd /cm/shared/apps/hadoop/Cloudera
# tar xvzf /tmp/oozie-4.1.0-cdh5.7.1.tar.gz
# tar xvzf oozie-4.1.0-cdh5.7.1/oozie-hadoopl原因-4.1.0-cdh5.7.1.tar.gz
# cd oozie-4.1.0-cdh5.7.1/
# tar xvzf oozie-examples.tar.gz
# mkdir libext
# cp hadoop原因/hadoop原因-2.6.0-cdh5.7.1.oozie-4.1.0-cdh5.7.1/* libext/# cp /tmp/ext-2.2.zip
libext/
```

5. Change ownership permissions for some directories

Big Data: How can I add Apache Oozie to my hadoop Cloudera CDH 5.7.1 instance (based on GPFS)?

```
# cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/  
# mkdir logs  
# chown oozie:oozie logs  
# mkdir data  
# chown oozie:oozie data  
# chown -R oozie:oozie oozie-server
```

6. Create Oozie database

```
# su - oozie  
$ cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/bin/  
$ ./ooziedb.sh create -run
```

7. Prepare WAR file

```
# su - oozie  
$ cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/bin/  
$ ./oozie-setup.sh prepare-war
```

8. Create directory for oozie in HDFS

```
# module load hadoop  
# hdfs dfs -mkdir /user/oozie  
# hdfs dfs -chown oozie:oozie /user/oozie
```

Big Data: How can I add Apache Oozie to my hadoop Cloudera CDH 5.7.1 instance (based on GPFS)?

9. Upload sharelib to HDFS

Substitute node001 with the NameNode hostname.

```
# su - oozie$ cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/bin/  
$ ./oozie-setup.sh sharelib create -fs hdfs://node001:8020 -locallib /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/oozie-sharelib-4.1.0-cdh5.7.1-yarn.tar.gz
```

10. Edit Oozie configuration

```
# cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/conf/  
# nano oozie-site.xml
```

Modify <value> to be consistent with the Hadoop configuration directory path:

```
<name>oozie.service.HadoopAccessorService.hadoop.configurations</name>  
<value>*/etc/hadoop/CDH5.2.1</value>
```

11. Start Oozie

Oozie should be started by running it as the oozie user. Use 'run' to run it in the foreground, 'start' to run it in the background. Log files can be found in
/cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/logs

```
# su - oozie  
$ cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/bin/  
  
$ ./oozied.sh run
```

Big Data: How can I add Apache Oozie to my hadoop Cloudera CDH 5.7.1 instance (based on GPFS)?

or

```
$ ./oozied.sh start
```

12. Check web console

The Oozie web console is available on the head node at <http://localhost:11000>

13. Edit Oozie job configuration

```
# cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/examples/apps/map-reduce
# nano job.properties
```

Using nano or another text editor, the following properties should be changed:

```
nameNode=hdfs://node001:8020
```

```
jobTracker=node003:8032
```

Here node001 is the NameNode and node003 is the ResourceManager (YARN server), with default port 8032

14. Upload examples to HDFS

```
# su - oozie
```

```
$ cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/
```

```
$ module load hadoop
```

```
$ hdfs dfs -put examples examples
```

Big Data: How can I add Apache Oozie to my hadoop Cloudera CDH 5.7.1 instance (based on GPFS)?

15. Run job

```
# su - oozie
$ cd /cm/shared/apps/hadoop/Cloudera/oozie-4.1.0-cdh5.7.1/bin
$ ./oozie job -oozie http://localhost:11000/oozie -config
examples/apps/map-reduce/job.properties -run
```

16. Check web consoles

Oozie web console (<http://localhost:11000>) should show the submitted job
YARN web console (<http://node003:8088>) should show the corresponding application, with:
type = MAPREDUCE
name = oozie:launcher:T=map-reduce:W=map-reduce-wf:A=mr-node:ID=0000000-141218162900779-oozie-oozi-W

17. Check job results

```
# su - oozie
$ module load hadoop
$ hdfs dfs -cat /user/oozie/examples/output-data/map-reduce/*
0 To be or not to be, that is the question;
42 Whether 'tis nobler in the mind to suffer
84 The slings and arrows of outrageous fortune,
129 Or to take arms against a sea of troubles,
172 And by opposing, end them. To die, to sleep;
217 No more; and by a sleep to say we end
255 The heart-ache and the thousand natural shocks
302 That flesh is heir to ? 'tis a consummation
346 Devoutly to be wish'd. To die, to sleep;
387 To sleep, perchance to dream. Ay, there's the rub,
438 For in that sleep of death what dreams may come,
487 When we have shuffled off this mortal coil,
531 Must give us pause. There's the respect
571 That makes calamity of so long life,
608 For who would bear the whips and scorns of time,
```

Page 6 / 7

Big Data: How can I add Apache Oozie to my hadoop Cloudera CDH 5.7.1 instance (based on GPFS)?

657 Th'oppressor's wrong, the proud man's contumely,
706 The pangs of despised love, the law's delay,
751 The insolence of office, and the spurns
791 That patient merit of th'unworthy takes,
832 When he himself might his quietus make
871 With a bare bodkin? who would fardels bear,
915 To grunt and sweat under a weary life,
954 But that the dread of something after death,
999 The undiscovered country from whose bourn
1041 No traveller returns, puzzles the will,
1081 And makes us rather bear those ills we have
1125 Than fly to others that we know not of?
1165 Thus conscience does make cowards of us all,
1210 And thus the native hue of resolution
1248 Is sicklied o'er with the pale cast of thought,
1296 And enterprises of great pitch and moment
1338 With this regard their currents turn awry,
1381 And lose the name of action.

Unique solution ID: #1334

Author: Michele Lamarca

Last update: 2016-11-30 18:13