

- [Close Window](#)
- [Print This Page](#)

Case: 00004924

Case Number	00004924	Contact Phone	none
Case Owner	Kenny Zhang	Contact Email	lmartin@hortonworks.com
Contact Name	Lester Martin		
Deployment Phase	Development		
HDP Version	2.0.6		
Operating System	CentOS-6 (64 Bit)		
Severity	S3 - Minor Feature Inoperative		
Priority	P3 - Med - Next day assist acceptable		
P1 Justification			
Product Component	Hue		
Problem Type	Executing Jobs		
Status	Open		
Sub Status	Resolution Provided - Pending Confirmation		
Apache BugID URL			
Hortonworks BugID URL			
Product Subcomponent	WebUI		
Targeted Fix Version			
Problem Statement/Question	Pig within Hue reports "Please initialize HIVE_HOME"		
Description	<p>Working with my client, Georgia Pacific, who wants to utilize Hue for the customers, I am having trouble getting Pig to run.</p> <p>I have a 5 node (2 masters, 3 workers) 2.0.6.0 cluster we are using for development activities that I installed Hue on based on our product documentation. Hue seems to be working fine for everything except Pig.</p> <p>Any Pig script I run via Hue echoes "Please initialize HIVE_HOME" after it visually reports the job fails.</p> <p>For example, I loaded up Batting.csv as identified in http://hortonworks.com/hadoop-tutorial/how-to-process-data-with-apache-pig/ and then try to run a script with just the load command as follows.</p> <pre>batting = load 'Batting.csv' using PigStorage(',');</pre> <p>I even tried it with the full path name of /user/hue/Batting.csv with the same results.</p> <p>No other messages are present when I click on the "Logs..." link at the bottom of the Hue page. The attached jobLog.txt file is the referenced job that fails. It only shows some WARN messages such as the following.</p> <pre>2014-04-21 10:50:10,737 WARN [Thread-61] org.mortbay.log: Job end notification to http://m2.hdp2:50111/templeton/v1/internal/complete/job_1398085104859_0005?status=SUCCEEDED failed with code: 400 and message "Bad Request"</pre> <p>I only found the following line from /var/log/hue/error.log (didn't see anything else remotely</p>		

related in other log files in this dir).

[21/Apr/2014 07:48:55 +0000] upload ERROR Left-over upload file is not cleaned up: /user/hue/Batting.csv.tmp

This is just a simple example, but I get the exact same error presentation with any Pig script I run. Also, these same Pig scripts run just fine with done via the command line as shown in attached pigCLI.txt file.

Just to show that Pig (by itself) is working fine, I went ahead and added a few more lines to determine the run totals by year and dump them as seen in pigCLIprintYearlyRuns.txt.

The message "Please initialize HIVE_HOME" seems to be coming from /usr/lib/pig/bin/pig which looks to have a mechanism to create a HIVE_HOME variable to be set to /usr/lib/hive if there isn't one present. The following shows that this folder is present on the host that Hue runs on.

```
[root@m1 hive]# pwd
/usr/lib/hive
[root@m1 hive]# ls
bin conf docs lib LICENSE NOTICE RELEASE_NOTES.txt scripts
[root@m1 hive]#
```

So... as the error says, I *just* need to initialize the HIVE_HOME directory, but sadly I'm not sure what to do next. I can't tell if Hue is trying to run the underlying Pig script (which works fine directly) or if it has its own setup/config for Pig (and Hive). I'm hopeful this is a somewhat common problem.

Can anyone help me get this going for my client?

Thanks,
Lester Martin

Current Status/Resolution

ROOT CAUSE: Product defect.

RESOLUTION: Hive was installed and configured on every node in the cluster, this is a work around until the fix is released on HDP.

Date/Time Opened

4/21/2014 12:02 PM

Date/Time Closed

Contact Information

Contact Owner	Eric Walmer
Name	Lester Martin
Account Name	Hortonworks
Contact Status	
Inferred Industry	

Marketo Sales Insight

Marketeto Sales Insight

[Help](#)

Priority:

Actions:

Add to Watch List
▾

Go!

Interesting Moments
Web Activity
Score
Email

Interesting Moment	Date
Milestone: Original source: salesforce.com Go to list (1) »	2/9/2014 7:56 PM

DiscoverOrg

<p style="margin: 0;">DiscoverOrg ID</p> <p style="margin: 0;">DiscoverOrg Company ID</p> <p style="margin: 0;">DiscoverOrg Last Update</p> <p style="margin: 0;">DeletedFromDiscoverOrg false</p> <p style="margin: 0;">Exclude from DiscoverOrg Auto-Updates <input type="checkbox"/></p>	<p style="margin: 0;">IT Org Chart N/A</p> <p style="margin: 0;">NickName</p> <p style="margin: 0;">MiddleName</p> <p style="margin: 0;">DiscoverOrg ReportsTo</p> <p style="margin: 0;">LinkedIn URL</p> <p style="margin: 0;">Twitter URL</p>
---	---

DiscoverOrg Technology Profile

Additional Information

Last Stay-in-Touch Request Date

Last Stay-in-Touch Save Date

System Information

Created By Eric Walmer, 2/9/2014 7:38 PM

Last Modified By Alice Anderson, 2/22/2014 10:05 PM

Custom Links [Google Search](#)
[Yahoo! Weather](#)

[Google Maps](#)

[Send Gmail](#)

Customer Correspondence

4/22/2014 5:54 PM

User Lester Martin

Yep, those two approaches have worked for me in the past when running Pig from the command-line, but never did for using Pig within Hue, so I tried them again.

```
REGISTER hdfs:///user/hue/jars/piggybank.jar;
yielded...
2014-04-22 02:20:25,534 [main] ERROR
org.apache.pig.tools.grunt.Grunt - ERROR
2999: Unexpected internal error. Pathname
/tmp/udfs/hdfs:/user/hue/jars/piggybank.jar
from
/tmp/udfs/hdfs:/user/hue/jars/piggybank.jar is
not a valid DFS filename.
```

and

```
REGISTER /usr/lib/pig/piggybank.jar;
yielded...
2014-04-22 02:14:55,147 [main] ERROR
org.apache.pig.tools.grunt.Grunt - ERROR
101: file
'hdfs://m1.hdp2:8020/usr/lib/pig/piggybank.jar'
does not exist.
```

Comment

I then saw the pattern that Pig within Hue uses paths provided on the REGISTER command to be locations within HDFS w/o the hdfs:/// prefix. I then realized that I already figured this out before as shown in the comments section of <https://martin.atlassian.net/wiki/x/C4BRAQ>.

So... the real problem was fixed when I was asked to deploy clients to all nodes in the cluster.

THANKS! and yes, we can close this one out now. Thanks again, Lester

4/22/2014 1:36 PM

User Lester Martin

I just attached the webhcat.log. There were some errors in there stating the following:

```
Callback failed
http://m1:8000/pig/notify/$jobId/ for job_...
```

I cleared those up as I have two masters m1 (m1.hdp2) on 192.168.56.41 and m2 (m2.hdp2)

4/22/2014 3:04 PM

User Kenny Zhang

Hi Lester,

For the original webhcat issue, you can fix it by adding m2 to the /etc/hosts in all nodes within the cluster. In this way, the pig job can report back to webhcat for the status, then the call back will succeed with both hostname and IP.

For the pig issue, you can fix it by two method:

1. Copy the piggybank.jar to HDFS,

```
# hadoop fs -copyFromLocal
/user/lib/pig/piggybank.jar
/directory/you/want/to
```

Comment

then register the piggybank.jar by:

```
REGISTER hdfs:///packetpig/piggybank.jar
```

2. Leave the piggybank.jar in the local fs on where you run the pig job (in this case, on the hue host), then register by:

```
REGISTER /user/lib/pig/piggybank.jar
```

Thanks,
Kenny

4/21/2014 8:41 PM

User Kenny Zhang

Hi Lester,

From the 400 error, it indicates the job succeeded, but when it report back to the Hue, webhcat complains about the bad request.

Comment

Could you please share the /var/log/webhcat/webhcat.log ?

Thanks,
Kenny

4/21/2014 12:37 PM

User Kenny Zhang

on 192.168.56.42 with Ambari, Hue, and NameNode running on m1 and m2 has Hive, WebHCat, and some other components as shown in the attached clusterDetails.png file. I was accessing hue with the url http://m1:8000 (m1 being resolved by a /etc/hosts entry on my host mac laptop) and when I swapped that out to http://192.168.56.42:8000 those errors went away.

Hi Lester,

This is a bug for Hue and hive which will be fixed in HDP 2.1 where ambari will install the hive client on the nodes which Hue requires.

<https://hortonworks.jira.com/browse/BUG-7466>

Comment

related to <https://issues.apache.org/jira/browse/HIVE-5547>

To workaround this, please install the hive client on all the YARN nodes within the cluster.

Please let me know how this goes.

Thanks,
Kenny

Comment

I then played with everything for a while and it looks like now that my only problem with Pig in Hue is that it can't allow "REGISTER piggybank.jar;" to function correctly.

Using the example at <https://martin.atlassian.net/wiki/x/BQBFAQ> I now get the following problem identified in the "Logs..." section of the Hue/Pig UI:

```
2014-04-21 23:58:41,228 [main] ERROR
org.apache.pig.tools.grunt.Grunt - ERROR
101: file
'hdfs://m1.hdp2:8020/tmp/udfs/piggybank.jar'
does not exist.
```

I tried it with single quotes around it as in "REGISTER 'piggybank.jar';", but this obviously made it much worse.

```
2014-04-22 00:14:00,678 [main] ERROR
org.apache.pig.tools.grunt.Grunt - ERROR
2997: Encountered IOException. File does not
exist:
hdfs://m1.hdp2:8020/tmp/udfs/'piggybank.jar'
```

I verified that /user/lib/pig/piggybank.jar is present (on m1.hdp2) and checked the same job in a 2.0 sandbox installation's Hue instance to make sure it still runs.

4/21/2014 6:31 PM

User Lester Martin

I used Ambari to deploy "Clients" to all three of the worker nodes as both master nodes already had this in place. I was then able to run the script in Hue that is detailed in pigCLIprintYearlyRuns.txt.

Comment

However, when I try to do another simple Pig script (see "Converter Script" in <https://martin.atlassian.net/wiki/x/BQBFAQ> for the example) I fail very similarly to before except this time the message block in Hue itself is blank and jobLog2.txt shows I got the same basic 400/BadRequest warning and nothing else.

I restarted the whole cluster and Hue and got the same results. Any additional suggestions?

4/21/2014 12:03 PM

User **Eric Hsu**

Hi,

Thank you for contacting Hortonworks Support.

This notification is to acknowledge that we have received your case. Our support team will review your case and get back to you.

Comment

Thanks,
-Eric

Hortonworks, Inc.
Technical Support Analyst
My office hours are Mon-Fri: 8am to 5pm PST
Office: 650.319.8315
How am I doing? Please feel free to provide feedback to my manager, Rick Morris, at rick@hortonworks.com
<http://support.hortonworks.com>

Attachments

webhcat.log

Size **7KB**
Ownership **Lester Martin**
View [View file](#)
Last Modified **4/22/2014 1:36 PM**

jobLog2.txt

Size **28KB**
Ownership **Lester Martin**
View [View file](#)
Last Modified **4/21/2014 6:30 PM**

pigCLIprintYearlyRuns.txt

Size **36KB**
Ownership **Lester Martin**
View [View file](#)
Last Modified **4/21/2014 12:03 PM**

pigCLI.txt

Size **24KB**
Ownership **Lester Martin**
View [View file](#)
Last Modified **4/21/2014 12:02 PM**

jobLog.txt

Size **27KB**
Ownership **Lester Martin**
View [View file](#)
Last Modified **4/21/2014 12:02 PM**