

D4Science Infrastructure - Task #4782

Provide ansible playbook for dataminer deployment

Jul 22, 2016 09:36 AM - Roberto Cirillo

Status:	Closed	Start date:	Jul 22, 2016
Priority:	Urgent	Due date:	
Assignee:	Roberto Cirillo	% Done:	100%
Category:	Other	Estimated time:	0.00 hour
Target version:	D4Science Infrastructure Upgrade to gCube 4.0.0		
Infrastructure:	Production		

Description

Provide an ansible playbook that deploys the dataminer instances in production environment:

```
dataminer1-p-d4s.d4science.org
dataminer2-p-d4s.d4science.org
dataminer3-p-d4s.d4science.org
dataminer4-p-d4s.d4science.org
dataminer5-p-d4s.d4science.org
dataminer6-p-d4s.d4science.org
```

These instances should run on the following scopes:

```
/d4science.research-infrastructures.eu/gCubeApps/ScalableDataMining
/d4science.research-infrastructures.eu/gCubeApps/BiodiversityLab
/d4science.research-infrastructures.eu/gCubeApps/StockAssessment
/d4science.research-infrastructures.eu/gCubeApps/SoBigData.eu
/d4science.research-infrastructures.eu/gCubeApps/Performance Evaluation In Aquaculture
/d4science.research-infrastructures.eu/gCubeApps/FAO Tuna Atlas
/d4science.research-infrastructures.eu/gCubeApps/Protected Area Impact Maps
/d4science.research-infrastructures.eu/gCubeApps/ENVRIPPlus
```

History

#1 - Jul 22, 2016 09:59 AM - Roberto Cirillo

- Status changed from New to In Progress

#2 - Jul 22, 2016 04:31 PM - Roberto Cirillo

- Status changed from In Progress to Feedback

- % Done changed from 0 to 100

#3 - Jul 22, 2016 04:56 PM - Andrea Dell'Amico

A couple of notes:

```
dataminer1-p-d4s.d4science.org
dataminer2-p-d4s.d4science.org
```

will be upgraded next monday. The other servers are running, and the haproxy configuration changed so that all the requests to dataminer-bigdata.d4science.org will be directed to dataminer[3:6]-p-d4s.d4science.org.

The haproxy stats URL has been opened to our network, so the frontend and backend status can be seen here: <http://dataminer.d4science.org:8880/>

#4 - Jul 23, 2016 02:00 AM - Gianpaolo Coro

The 3:6 services are not running because there are 52North WPS 3.5.0 libraries together with the (required) 3.3.2 libraries. The 3.5.0 libraries should be removed.

#5 - Jul 24, 2016 02:30 PM - Andrea Dell'Amico

I've just added the following files:

```
52n-wps-algorithm-3.5.0.jar
```

```
52n-wps-commons-3.5.0.jar
52n-wps-database-3.5.0.jar
52n-wps-io-3.5.0.jar
52n-wps-io-impl-3.5.0.jar
52n-wps-server-3.5.0.jar
```

to the list of jars that need to be deleted. But this behaviour is unsustainable, we need a more deterministic way to produce a dataminer webapp working instance and we need to be able to reproduce the same installation for each specific version.

@roberto.cirillo@isti.cnr.it I also made the 1.2.0-4.0.0-129209 version the default.

I'm running the playbook on dataminer[3:6] now.

#6 - Jul 25, 2016 08:54 AM - Gianpaolo Coro

Yes Andrea, I will fix the 52-North dependencies as soon as I will be back to work on DataMiner.

However, I have checked the installation. It seems OK as I have run tests to check the capabilities and the correctness of the algorithms descriptions. However, the service cannot run processes completely, because there are issues in the interaction with the Workspace.

I see this error when I run any computation:

```
java.lang.ClassCastException: java.lang.String cannot be cast to org.gcube.common.homelibrary.model.items.ItemD
elegate
    at org.gcube.common.homelibrary.jcr.workspace.servlet.JCRSession.saveItem(JCRSession.java:360)
    at org.gcube.common.homelibrary.jcr.workspace.JCRWorkspaceItem.save(JCRWorkspaceItem.java:234)
    at org.gcube.common.homelibrary.jcr.workspace.JCRAbstractWorkspaceFolder.setSystemFolder(JCRAbstractWo
rkspaceFolder.java:848)
    at org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mapping.dataspace.DataspaceManager.create
FoldersNetwork(DataspaceManager.java:95)
    at org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mapping.dataspace.DataspaceManager.writeP
rovenance(DataspaceManager.java:399)
    at org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mapping.dataspace.DataspaceManager.run(Da
taspaceManager.java:78)
    at java.lang.Thread.run(Thread.java:745)
```

I guess this is due either to a wrong home-library jar or to the fact that the Workspace system is not updated in the production environment. These are the home-library jars on the machine:

```
-rw-r--r-- 1 gcube gcube 302415 Jul 22 14:40 tomcat/webapps/wps/WEB-INF/lib/home-library-jcr-2.4.0-4.0.0-13059
8.jar
-rw-r--r-- 1 gcube gcube 121177 Jul 22 14:40 tomcat/webapps/wps/WEB-INF/lib/home-library-model-1.3.0-4.0.0-129
370.jar
-rw-r--r-- 1 gcube gcube 160385 Jul 22 14:40 tomcat/webapps/wps/WEB-INF/lib/home-library-2.4.0-4.0.0-129376.ja
r
```

@valentina.marioli@isti.cnr.it or @roberto.cirillo@isti.cnr.it could you clarify, please?

#7 - Jul 25, 2016 09:43 AM - Valentina Marioli

The issue regarding home-library is due to the fact that Jackrabbit is not updated, so the folder cannot be set as "system folder".

#8 - Jul 25, 2016 06:14 PM - Gianpaolo Coro

Dataminer 3:6 have been tested on non-geospatial algorithms. I'm waiting for the geospatial infrastructure to be up and running to finish the tests.

#9 - Jul 25, 2016 07:21 PM - Andrea Dell'Amico

dataminer[1:2] has been upgraded (for real, this time).

#10 - Jul 25, 2016 08:10 PM - Gianpaolo Coro

dataminer [1:2] work. Perhaps, tomorrow I will be able to test the geoprocessing part. If you create the groups of dataminers on the proxy, I can test them as well.

#11 - Jul 26, 2016 08:42 AM - Gianpaolo Coro

I have run complete tests, also on the geoprocessing algorithms. All the dataminers works. Now it's time to implement the proxies.

#12 - Jul 26, 2016 08:52 AM - Gianpaolo Coro

I have measured that dataminer4 and dataminer6 are slower than the other machines. Is it possible to move them on a better hardware?

#13 - Jul 26, 2016 09:11 AM - Andrea Dell'Amico

Gianpaolo Coro wrote:

I have run complete tests, also on the geoprocessing algorithms. All the dataminers works. Now it's time to implement the proxies.

If you are referring to ha proxy, it's already configured. The hostname is dataminer-sobigdata.d4science.org

#14 - Jul 26, 2016 10:17 AM - Gianpaolo Coro

Hi Andrea, the "sobigdata" term makes me thrill... :-S

The distribution of the dataminer should be:

```
dataminer.d4science.org: dataminer [1:2]
```

```
dataminer-bigdata.d4science.org: dataminer [3:6]
```

The second proxy is "dataminer-bigdata", NOT "dataminer-sobigdata", because it will serve **real** big data computations.

The "soBigData" Project problems are currently **not real** big data problems and will be managed by the first proxy! Whenever they became real big data problems, I would move them to the other proxy.

Could you please change the dataminer-sobigdata.d4science.org into dataminer-bigdata.d4science.org?

#15 - Jul 26, 2016 11:12 AM - Andrea Dell'Amico

Gianpaolo Coro wrote:

Hi Andrea, the "sobigdata" term makes me thrill... :-S

The distribution of the dataminer should be:

```
dataminer.d4science.org: dataminer [1:2]
```

```
dataminer-bigdata.d4science.org: dataminer [3:6]
```

Could you please change the dataminer-sobigdata.d4science.org into dataminer-bigdata.d4science.org?

It's dataminer-bigdata.d4science.org already. Sorry, I wrote sobigdata without checking first.

#16 - Jul 26, 2016 12:14 PM - Gianpaolo Coro

OK, everything works correctly. I think we can close this ticket.

#17 - Aug 29, 2016 11:13 AM - Roberto Cirillo

- Status changed from Feedback to Closed