

D4Science Infrastructure - Task #4865

Installing data analysis machine to develop prototypes

Aug 22, 2016 03:44 PM - Gianpaolo Coro

Status:	Closed	Start date:	Aug 22, 2016
Priority:	High	Due date:	
Assignee:	Andrea Dell'Amico	% Done:	100%
Category:	High-Throughput-Computing	Estimated time:	0.00 hour
Target version:	Dataminer and StatMan provisioning		
Infrastructure:	Production		
Description An all-in-one machine is required by the developers of Dataminer algorithms. This machine should be a production Dataminer instance to be placed under the "dataminer-prototype.d4science.org" proxy group. The same machine should be endowed with an RStudio instance. Developers will use the RStudio instance to develop their R script and then the Dataminer instance to run their prototypes. The installation of the R packages will be still at root level and managed in the current way. The installation of the algorithm will pass from our management but will be soon automatised through the Statistical Algorithms Importer.			
Related issues:			
Related to D4Science Infrastructure - Task #4371: same R configuration (R ver...		Closed	Jun 23, 2016 Jul 23, 2016
Related to D4Science Infrastructure - Task #4842: Make the ldap configuration...		Closed	Aug 02, 2016
Blocked by D4Science Infrastructure - Task #4917: Update Dataminer Ansible in...		Closed	Sep 01, 2016

History

#2 - Aug 29, 2016 04:15 PM - Pasquale Pagano

- Tracker changed from Support to Task

#3 - Sep 05, 2016 05:45 PM - Gianpaolo Coro

- Priority changed from Normal to High

Do you have updates on this ticket? Please this is a requirement of the BlueBRIDGE project.

#4 - Sep 05, 2016 05:46 PM - Leonardo Candela

- Parent task set to #4894

#5 - Sep 05, 2016 06:47 PM - Pasquale Pagano

Please @tommaso.piccioli@isti.cnr.it start creating a new VM. Then @andrea.dellamico@isti.cnr.it will run the ansible playbook to configure it.

#6 - Sep 05, 2016 07:12 PM - Tommaso Piccioli

dataminer1-proto.d4science.org created, to be configured as dataminer.

Initial resources are 40 GB disk, 8 cpu and 8 GB ram, they can be modified.

#7 - Sep 05, 2016 07:25 PM - Andrea Dell'Amico

I have some questions:

- rstudio server needs to authenticate in the same way of rstudio.d4science.org?
- If so, against the new ldap server? (rstudio(-d).d4science.org authenticate against the old one, when I tried to switch it didn't work)
- is the new installation procedure for the dataminer needed? (See [#4917](#))

#8 - Sep 08, 2016 11:38 AM - Andrea Dell'Amico

- Blocked by Task #4917: Update Dataminer Ansible installation added

#9 - Sep 08, 2016 11:38 AM - Andrea Dell'Amico

- Related to Task #4371: same R configuration (R version and packages) on R-Studio as the one on machine(s) used by Dataminer. added

#10 - Sep 08, 2016 11:40 AM - Andrea Dell'Amico

- Blocked by Task #4842: Make the ldap configuration on rstudio(-dev).d4science.org point to ldap.d4science.org added

#11 - Sep 08, 2016 11:40 AM - Andrea Dell'Amico

- Status changed from New to In Progress

#12 - Sep 08, 2016 11:40 AM - Andrea Dell'Amico

- Assignee changed from _InfraScience Systems Engineer to Andrea Dell'Amico

#13 - Sep 08, 2016 05:26 PM - Andrea Dell'Amico

@lucio.lelii@isti.cnr.it @roberto.cirillo@isti.cnr.it I need your help with the smartgears contexts configuration. The actual production dataminers have /wps as context:

```
<application mode="online" context="/wps">

  <scope>/d4science.research-infrastructures.eu</scope>

  <scope>/d4science.research-infrastructures.eu/gCubeApps</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/BiodiversityLab</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/BiOnym</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/ScalableDataMining</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/StockAssessment</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/SoBigData.eu</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/PerformanceEvaluationInAquaculture</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/FAO_TunaAtlas</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/ProtectedAreaImpactMaps</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/ENVRIPplus</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/ICCAT_BFT-E</scope>

</application>
```

While rstudio.d4science.org has /r-connector:

```
<application mode="online" context="/r-connector">
  <scope>/d4science.research-infrastructures.eu</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/TabularDataLab</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/FAO_TunaAtlas</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/BOBLME_HilsaAWG</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/StockAssessment</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/BiodiversityLab</scope>
  <scope>/d4science.research-infrastructures.eu/gCubeApps/ICCAT_BFT-E</scope>
</application>
```

Do we need two contexts (it means that I need to refactor the container.xml templating heavily)? or only one is permitted? if so, which one?

#14 - Sep 08, 2016 05:36 PM - Lucio Lelii

you can put both contexts in the container.xml with the same list of scopes.

#15 - Sep 08, 2016 05:43 PM - Andrea Dell'Amico

Thanks.

#16 - Sep 09, 2016 01:27 PM - Andrea Dell'Amico

- % Done changed from 0 to 70

I think only the LDAP configuration for rstudio is missing. In the meantime could you check if the dataminer jars are correctly installed and that the smartgear configuration is correct?

I also see those logs:

```
Connecting...
Updating IS on scope /d4science.research-infrastructures.eu
echo | voms-proxy-init --voms fedcloud.egi.eu --rfc --dont-verify-ac -cert /tmp/x509up_u1000 -out /tmp/proxy4
17816595.pem -pwstdin
Process started !
/bin/sh: 1: voms-proxy-init: not found
Process ended !
Proxy Certificate created
java.io.FileNotFoundException: /tmp/proxy417816595.pem (No such file or directory)
```

```

    at java.io.FileInputStream.open(Native Method)
    at java.io.FileInputStream.<init>(FileInputStream.java:146)
    at org.gcube.vomanagement.occi.utils.ScriptUtil.getScriptFromFile(ScriptUtil.java:26)
    at org.gcube.resources.federation.fhnmanager.impl.ConnectorFactory.createConnector(ConnectorFactory.java:115)
    at org.gcube.resources.federation.fhnmanager.impl.ConnectorFactory.getConnector(ConnectorFactory.java:30)
    at org.gcube.resources.federation.fhnmanager.is.ISProxyImpl.updateIs(ISProxyImpl.java:491)
    at org.gcube.resources.federation.fhnmanager.impl.ISSynchronizer.run(ISSynchronizer.java:47)
    at java.lang.Thread.run(Thread.java:745)

```

voms-proxy-init is an EGI cloud thing, don't know why it's been executed.

#17 - Sep 09, 2016 01:28 PM - Andrea Dell'Amico

The smartgear configuration is

```
<container mode='online'>
```

```

    <hostname>dataminer1-proto.d4science.org</hostname>
    <port>80</port>
    <infrastructure>d4science.research-infrastructure.eu</infrastructure>
    <vo>gCubeApps</vo>

```

```

<site>
  <country>it</country>
  <location>pisa</location>
  <latitude>41.9</latitude>
  <longitude>12.5</longitude>
</site>

```

```

<property name='SmartGearsDistributionBundle' value='UnBundled' />
<property name='SmartGearsDistribution' value='1.2.8-4.0.0-129615' />
<publication-frequency>60</publication-frequency>

```

```

<application mode="online" context="/wps">
  <scope>/d4science.research-infrastructure.eu</scope>
    <scope>/d4science.research-infrastructure.eu/gCubeApps</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/BiodiversityLab</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/BiOnym</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/ScalableDataMining</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/StockAssessment</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/SoBigData.eu</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/PerformanceEvaluationInAquaculture</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/FAO_TunaAtlas</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/ProtectedAreaImpactMaps</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/ENVRIPPlus</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/ICCAT_BFT-E</scope>
</application>

```

```

<application mode="online" context="/r-connector">
  <scope>/d4science.research-infrastructure.eu</scope>
    <scope>/d4science.research-infrastructure.eu/gCubeApps</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/BiodiversityLab</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/BiOnym</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/ScalableDataMining</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/StockAssessment</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/SoBigData.eu</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/PerformanceEvaluationInAquaculture</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/FAO_TunaAtlas</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/ProtectedAreaImpactMaps</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/ENVRIPPlus</scope>
  <scope>/d4science.research-infrastructure.eu/gCubeApps/ICCAT_BFT-E</scope>
</application>
</container>

```

#18 - Sep 09, 2016 02:49 PM - Andrea Dell'Amico

Andrea Dell'Amico wrote:

I think only the LDAP configuration for rstudio is missing. In the meantime could you check if the dataminer jars are correctly installed and that the smartgear configuration is correct?

I also see those logs:

```
Connecting...
```

And those logs go into catalina.out. That should never happen.

#19 - Sep 09, 2016 03:13 PM - Andrea Dell'Amico

The ldap authentication needed by the r-connector is in place. Against the new ldap server, so this has to be tested as well.

#20 - Sep 09, 2016 03:14 PM - Andrea Dell'Amico

- Status changed from In Progress to Feedback

- % Done changed from 70 to 100

#21 - Sep 12, 2016 12:46 PM - Gianpaolo Coro

It seems that the only way to completely test this service is to have a new VRE in the production environment.

#22 - Sep 13, 2016 12:48 PM - Andrea Dell'Amico

Andrea Dell'Amico wrote:

```
Connecting...
Updating IS on scope /d4science.research-infrastructures.eu
echo | voms-proxy-init --voms fedcloud.egi.eu --rfc --dont-verify-ac -cert /tmp/x509up_u1000 -out /tmp/pr
oxy417816595.pem -pwstdin
```

About this one: a service was installed by mistake (wrong defaults into the ansible role). I'm removing it and restarting tomcat.

#23 - Oct 07, 2016 07:41 PM - Pasquale Pagano

- Status changed from Feedback to In Progress

#24 - Oct 07, 2016 07:43 PM - Pasquale Pagano

- Status changed from In Progress to Feedback

#25 - Oct 08, 2016 12:41 PM - Andrea Dell'Amico

- Blocked by deleted (Task #4842: Make the ldap configuration on rstudio(-dev).d4science.org point to ldap.d4science.org)

#26 - Oct 08, 2016 12:42 PM - Andrea Dell'Amico

- Related to Task #4842: Make the ldap configuration on rstudio(-dev).d4science.org point to ldap.d4science.org added

#27 - Oct 08, 2016 12:43 PM - Andrea Dell'Amico

- Status changed from Feedback to Closed

rstudio server is working, and the ldap authentication against the new ldap server too.