

D4Science Infrastructure - Task #10664

DataMiner - Create a list of the names of Python packages installed on the DataMiner machines

Dec 12, 2017 12:41 PM - Giancarlo Panichi

Status:	Closed	Start date:	Dec 12, 2017
Priority:	High	Due date:	
Assignee:	_InfraScience Systems Engineer	% Done:	100%
Category:	System Application	Estimated time:	0.00 hour
Target version:	DataMiner - Pre Installed Software support		
Infrastructure:	Development, Pre-Production, Production		
Description			
Our users start asking us to install new python packages to be used by their algorithms. It is necessary to create a list of installed packages as we do for R. This list should be reported on svn and available as link for everyone on the wiki:			
https://wiki.gcube-system.org/gcube/Pre_Installed_Packages			
Caution, must also be considered the cases in which installation systems other than Pip are used (for example Conda #10653)			

History

#2 - Dec 20, 2017 03:58 PM - Giancarlo Panichi

- Priority changed from Normal to High

Any news?

#3 - Dec 21, 2017 06:06 PM - Andrea Dell'Amico

- Status changed from New to In Progress

The python environment that's currently installed on all the dataminer servers is based on python 2.7.6 (Ubuntu 14.04).
The python packages installed are (deb packages):

```
dh-python
ganglia-monitor-python
libboost-mpi-python1.55-dev
libboost-mpi-python1.55.0
libboost-python1.55-dev
libboost-python1.55.0
libpython-dev:amd64
libpython-stdlib:amd64
libpython2.7:amd64
libpython2.7-dev:amd64
libpython2.7-minimal:amd64
libpython2.7-stdlib:amd64
libpython3-stdlib:amd64
libpython3.4-minimal:amd64
libpython3.4-stdlib:amd64
python
python-antlr
python-apt
python-apt-common
python-boto
python-bs4
python-bson
python-bson-ext
python-cairo
python-chardet
python-chardet-whl
python-cheetah
python-colorama
python-colorama-whl
python-configobj
python-crypto
```

python-cvxopt
python-dateutil
python-debian
python-decorator
python-dev
python-distlib
python-distlib-whl
python-egenix-mxdatetime
python-egenix-mxtools
python-glade2
python-gobject-2
python-gridfs
python-gtk2
python-html5lib
python-html5lib-whl
python-imaging
python-joblib
python-json-pointer
python-jsonpatch
python-ldb
python-lxml
python-matplotlib
python-matplotlib-data
python-minimal
python-networkx
python-nose
python-numexpr
python-numpy
python-oauth
python-openpyxl
python-pandas
python-pandas-lib
python-patsy
python-pil
python-pip
python-pip-whl
python-pkg-resources
python-prettytable
python-psycpg2
python-pycurl
python-pyinotify
python-pymongo
python-pymongo-ext
python-pyparsing
python-requests
python-requests-whl
python-samba
python-scipy
python-scour
python-serial
python-setuptools
python-setuptools-whl
python-simplejson
python-six
python-six-whl
python-sklearn
python-sklearn-lib
python-software-properties
python-statsmodels
python-statsmodels-lib
python-support
python-tables
python-tables-data
python-tables-lib
python-talloc
python-tdb
python-tk
python-tz
python-urllib3
python-urllib3-whl
python-wheel
python-xapian
python-xlrd
python-xlwt
python-yaml

```
python2.7
python2.7-dev
python2.7-minimal
```

Some python 3.4 packages are also present:

```
python3
python3-apt
python3-commandnotfound
python3-dbus
python3-distupgrade
python3-gdbm:amd64
python3-gi
python3-minimal
python3-pkg-resources
python3-pycurl
python3-software-properties
python3-update-manager
python3.4
python3.4-minimal
```

The pip installed packages are:

```
apt-xapian-index (0.45)
asn1crypto (0.21.1)
backports.shutil-get-terminal-size (1.0.0)
beautifulsoup4 (4.2.1)
boto (2.20.1)
certifi (2017.1.23)
cffi (1.9.1)
chardet (2.0.1)
Cheetah (2.4.4)
cloud-init (0.7.5)
colorama (0.2.5)
configobj (4.7.2)
cryptography (1.8.1)
cvxopt (1.1.4)
decorator (3.4.0)
enum34 (1.1.6)
html5lib (0.999)
idna (2.5)
igraph (0.1.11)
iotop (0.6)
ipaddress (1.0.18)
ipython (5.3.0)
ipython-genutils (0.1.0)
joblib (0.7.1)
jsonpatch (1.3)
jsonpointer (1.0)
lxml (3.3.3)
matplotlib (1.3.1)
ndg-httpsclient (0.4.2)
networkx (1.8.1)
nose (1.3.1)
numexpr (2.2.2)
numpy (1.8.2)
oauth (1.0.1)
openpyxl (1.7.0)
packaging (16.8)
pandas (0.13.1)
pathlib2 (2.2.1)
patsy (0.2.1)
pexpect (4.2.1)
pickleshare (0.7.4)
Pillow (2.3.0)
pip (9.0.1)
prettytable (0.7.2)
prompt-toolkit (1.0.13)
psycpg2 (2.4.5)
ptyprocess (0.5.1)
pyasn1 (0.2.3)
pyparser (2.17)
pycrypto (2.6.1)
pycurl (7.19.3)
Pygments (2.2.0)
```

```
pyinotify (0.9.4)
pymongo (2.6.3)
pyOpenSSL (16.2.0)
pyparsing (2.0.1)
pyserial (2.6)
Pyste (0.9.10)
python-apt (0.9.3.5ubuntu2)
python-dateutil (1.5)
python-debian (0.1.21-nmu2ubuntu2)
pytz (2012rc0)
PyYAML (3.10)
requests (2.2.1)
scandir (1.5)
scikit-learn (0.14.1)
scipy (0.13.3)
setuptools (33.1.1)
simplegeneric (0.8.1)
simplejson (3.3.1)
six (1.11.0)
ssh-import-id (3.21)
statsmodels (0.5.0)
tables (3.1.1)
traitlets (4.3.2)
urllib3 (1.7.1)
wcwidth (0.1.7)
wheel (0.24.0)
xlrd (0.9.2)
xlwt (0.7.5)
```

Most of those packages are installed because some system package requires them or because are dependencies of other packages. The python software explicitly required to be installed on the dataminer is the following:

```
deb:
- python-numpy
- python-matplotlib
- python-matplotlib-data
- python-psycopg2
- python-scipy
- python-sklearn
- python-networkx
- python-pandas

pip:
- igraph
```

#4 - Dec 21, 2017 06:08 PM - Andrea Dell'Amico

I will not track the packages that are not explicitly required for the dataminers btw, because they can be triggered by any software installation or upgrade.

#5 - Dec 21, 2017 08:22 PM - Andrea Dell'Amico

- % Done changed from 0 to 30

Another consideration: the above lists are completely useless to the people that is going to use conda. conda is a completely closed and self contained environment, the python install inside a conda environment cannot see the system installed packages.

It's also completely useless trying to manage the conda installed environment as we do with the R packages: as each installation is self contained, it's executed once.

We can still consider to manage the pip installed packages in the same way we do for the R ones, but we have just one package right now.

#6 - Jan 09, 2018 03:23 PM - Andrea Dell'Amico

- Status changed from In Progress to Closed

- % Done changed from 30 to 100

I'm closing this one. For the aginfra algorithm we needed a custom solution and we do not have any other requests for a more dynamic python packages management right now.