

install

1. Download anaconda upload to chtc filefold
2. Run the following command
 - `bash Anaconda-latest-Linux-x86_64.sh`
3. Follow the prompts on the installer screens.
4. Say yes at last step

Install package

5. Set environment variable ,point to your anaconda3(or 2) / bin path everytime you login
 - `export PATH=$(pwd)/.../anaconda3/bin:$PATH`
6. Replace the “scipy” into the python package name you want (nltk stopwords cannot use, define manually)
 - `conda install scipy`
7. create your own tarball of the installation:
going to your anaconda3 folder path
 - `tar -czvf anacondas.tar.gz anaconda3/`

.sh file

Your can get this .sh and .sub example at code/peijinli folder

```
#!/bin/bash

# untar your Python installation
tar -xzf anaconda3.tar.gz
# make sure the script will use your Python installation,
# and the working directory as it's home location
export PATH=$(pwd)/anaconda3/bin:$PATH
mkdir home
export HOME=$(pwd)/home
# run your script
python3 data_cleaning_filter1.py
```

untar your anaconda tarball.

your execute .py

If you edit .sh at windows system...you need a little transfet

Run following at your .sh path

- `vi -b <name.sh>`
- `dos2unix <name.sh>`

.sub

```
universe = vanilla
log = hello-htc_$(Cluster).log
error = hello-htc_$(Cluster)_$(Process).err

executable = chtc_datacl.sh
arguments = $(Process)
output = hello-htc_$(Cluster)_$(Process).out

should_transfer_files = YES
when_to_transfer_output = ON_EXIT
transfer_input_files = /home/pli72/628-2/comment.csv, data_cleaning_filter1.py, /home/pli72/628-2/home/ana

request_cpus = 1
request_memory = 3GB
request_disk = 2GB

queue 1
```

your .sh file upload at here

Don't forget add path if the file not in your current folder

your data and .py and all the other file you need

Don't forget add path if the file not in your current folder

your memory should be higher than all the size you file you send. Include anaconda3.tar.gz ask more if necessary

Check and rm job

- `condor_q`
- `condor_rm <name>`