Linux Software Installation

Qi Sun
Bioinformatics Facility

Part 1. Components of software

e.g. /programs/bwa-0.7.8/bwa

e.g. /lib64/libz.so.1

e.g. maker requires blast

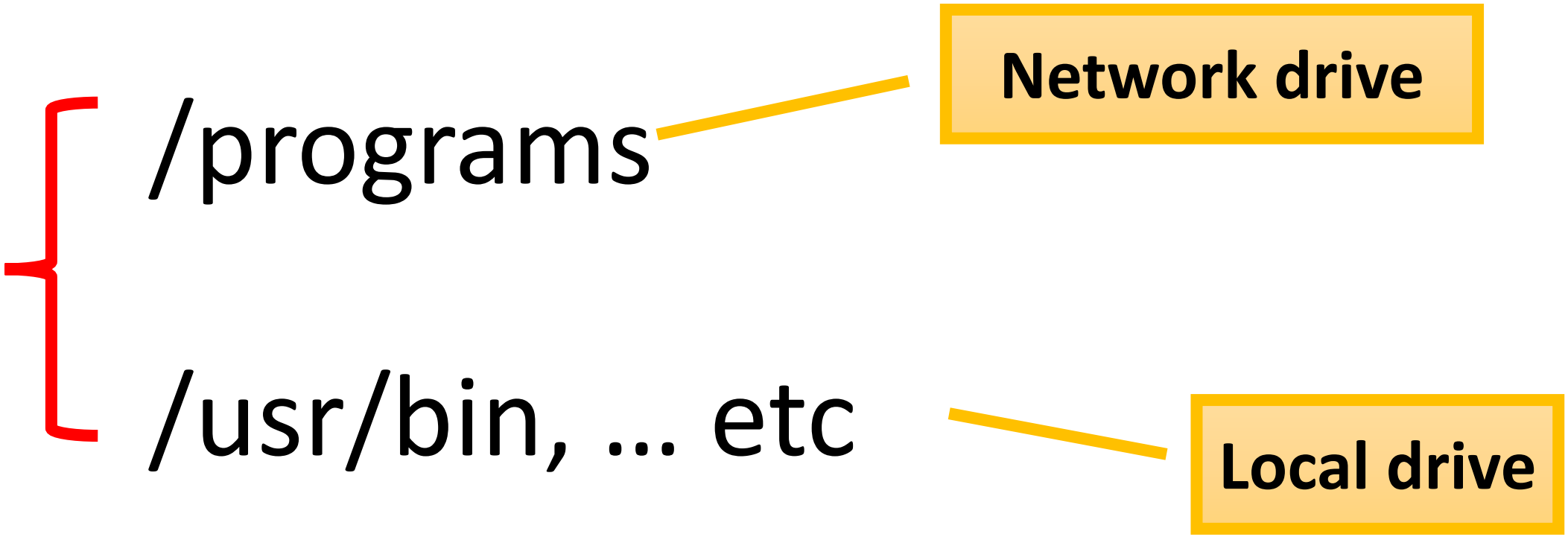
Where are the software files located?

Main code: /usr/bin; /usr/local/bin

Libraries: /lib; lib64; /usr/lib; /usr/lib64;
/usr/local/lib; /usr/local/lib64;

External executables:
/usr/bin; /usr/local/bin

Where are the software files on BioHPC computers?



How to install software by yourself?

All new files need to be in your home directory.

e.g. /home/xxxxx/bin
/home/xxxxx/lib

To run your software:

Specify the PATH of the main code

export PATH=/home/xxxxx/programs:$PATH

Specify the PATH of the libraries

C export LD_LIBRARY_PATH=/home/xxxxx/lib

PERL export PERL5LIB=/home/xxxxx/perl5/5.22.0

PYTHON export PYTHONPATH=/home/xxxxx/python/lib/python2.7/site_packages

Let’s examine a software: “Entropy”

Main code

`/programs/entropy/bin/entropy`

Let’s examine a software: “Entropy”

Libraries

```
/lib64/libgsl.so.0
/lib64/libgslcblas.so.0
/lib64/libz.so.1
/lib64/libdl.so.2
/lib64/libm.so.6
/lib64/libstdc++.so.6
/lib64/libgcc_s.so.1
/lib64/libc.so.6
/lib64/libpthread.so.0
/usr/lib64/atlas/libsatlas.so.3
/lib64/ld-linux-x86-64.so.2
/lib64/libgfortran.so.3
/lib64/libquadmath.so.0
```

Let’s examine a software: “Entropy”

External executables

```
/programs/hdf5-1.10.1/bin
```

To run the software, you need to specify the path of the main code and the libraries:

```
export PATH=/programs/entropy/bin:$PATH

export LD_LIBRARY_PATH=/programs/gcc-
5.3.0/lib:/programs/gcc-5.3.0/lib64
```

- You need to do this every time you start a new “ssh” session;
- Alternatively, put the lines in /home/xxxxx/.bashrc file, which is run automatically when a session starts.

A tool to examine which software file is used:

which bwa

/programs/bin/bwa/bwa

echo $PATH
/programs/docker/bin:/usr/local/bin:/usr/bin:/usr/local/sbin:/usr/sbin:/programs/bin/mummer:/programs/bin/util:/programs/bin/bowtie:/programs/bin/bwa:/programs/bin/cufflinks:/programs/bin/samtools:/programs/bin/tophat:/programs/bin/fastx:/programs/bin/blast:/programs/bin/igv:/programs/bin/velvet:/programs/bin/iAssembler:/programs/bin/GATK:/programs/bin/454:/programs/bin/blat:/programs/bin/perlscripts……

Identify which library file is used:

C

```
ldd /programs/entropy/bin/entropy
```

```
linux-vdso.so.1 =>  (0x00007ffefb1d5000)
libgsl.so.0 => /lib64/libgsl.so.0 (0x00007efe3544a000)
libgslcblas.so.0 => /lib64/libgslcblas.so.0 (0x00007efe3520c000)
libz.so.1 => /lib64/libz.so.1 (0x00007efe34ff6000)
libdl.so.2 => /lib64/libdl.so.2 (0x00007efe34df2000)
libm.so.6 => /lib64/libm.so.6 (0x00007efe34aef000)
libstdc++.so.6 => /lib64/libstdc++.so.6 (0x00007efe347e7000)
libgcc_s.so.1 => /lib64/libgcc_s.so.1 (0x00007efe345d1000)
libc.so.6 => /lib64/libc.so.6 (0x00007efe3420d000)
libpthread.so.0 => /lib64/libpthread.so.0 (0x00007efe33ff1000)
libsatlas.so.3 => /usr/lib64/atlas/libsatlas.so.3 (0x00007efe333a4000)
/lib64/ld-linux-x86-64.so.2 (0x0000556c6875e000)
libgfortran.so.3 => /lib64/libgfortran.so.3 (0x00007efe33081000)
libquadmath.so.0 => /lib64/libquadmath.so.0 (0x00007efe32e45000)
```

Library search path for c programs

- path defined in the code;
- LD_LIBRARY_PATH
- /etc/ld.so.conf.d/*.conf

Examine which library file is used:

python

```
> import numpy
> numpy.__file__
'/usr/local/lib/python2.7/dist-packages/numpy/__init__.pyc'
```

Paths to search for PYTHON libraries

1. **$PYTHONPATH**
2. **sys.path**

['', '/usr/lib64/python2.7/site-packages/Cython-0.23.2-py2.7-linux-x86_64.egg', '/usr/lib64/python2.7/site-packages/HTSeq-0.6.1-py2.7-linux-x86_64.egg', '/usr/lib64/python2.7/site-packages/dadi-1.7.0-py2.7-linux-x86_64.egg', '/usr/lib64/python27.zip', '/usr/lib64/python2.7', '/usr/lib64/python2.7/plat-linux2', '/usr/lib64/python2.7/lib-tk', '/usr/lib64/python2.7/lib-old', '/usr/lib64/python2.7/lib-dynload', '/usr/lib64/python2.7/site-packages', '/usr/lib64/python2.7/site-packages/gtk-2.0', '/usr/lib/python2.7/site-packages', '/usr/lib/python2.7/site-packages/pexpect-3.3-py2.7.egg', '/usr/lib/python2.7/site-packages/numpy-1.9.2-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/scipy-0.16.0-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/simplejson-3.8.0-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/jsonpickle-0.9.2-py2.7.egg', '/usr/lib/python2.7/site-packages/pyamg-2.2.1-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/nose-1.3.7-py2.7.egg', '/usr/lib/python2.7/site-packages/pbcore-1.2.2-py2.7.egg', '/usr/lib/python2.7/site-packages/h5py-2.5.0-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/birdsuite-1.0-py2.5.egg', '/usr/lib/python2.7/site-packages/mpgutils-0.7-py2.5.egg', '/usr/lib/python2.7/site-packages/mock-1.3.0-py2.7.egg', '/usr/lib/python2.7/site-packages/funcsigs-0.4-py2.7.egg', '/usr/lib/python2.7/site-packages/pbr-1.8.0-py2.7.egg', '/usr/lib/python2.7/site-packages/biopython-1.65-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/pandas-0.16.2-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/cutadapt-1.8.3-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/rpy2-2.7.0-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/singledispatch-3.4.0.3-py2.7.egg', '/programs/MACS/lib/python2.7/site-packages/misopy-0.4.6-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/DendroPy-4.0.3-py2.7.egg', '/usr/lib/python2.7/site-packages/openpyxl-2.3.0b2-py2.7.egg', '/usr/lib/python2.7/site-packages/et_xmlfile-1.0.0-py2.7.egg', '/usr/lib/python2.7/site-packages/jdcal-1.0-py2.7.egg', '/usr/lib/python2.7/site-packages/reportlab-3.2.0-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/Pillow-2.9.0-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/deap-1.0.2-py2.7.egg', '/usr/lib/python2.7/site-packages/jcvi-0.5.7-py2.7.egg', '/usr/lib/python2.7/site-packages/screed-0.7-py2.7.egg', '/usr/lib/python2.7/site-packages/khmer-0.4-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/numexpr-2.4.4-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/tables-3.2.1.1-py2.7-linux-x86_64.egg', '/usr/lib/python2.7/site-packages/ordereddict-1.1-py2.7.egg', '/usr/lib/python2.7/site-packages/pysql-0.16-py2.7.egg', '/usr/lib/python2.7/site-packages/setuptools-34.1.1-py2.7.egg']

Part 2. Types of Software

Programming Languages

C / C++	**JAVA**
PYTHON	**R**
PERL	**BASH**

Script vs Binary

Binary: C

Script: PERL, R, BASH

Bytecode: JAVA,PYTHON

A script requires an interpreter software, while a binary software can run directly

Script:

```
python /programs/pybedtools/bin/intron_exon_reads.py
```

Binary

```
bwa
```

* On BioHPC, bwa is in the $PATH. No need to use full path of the file

The **Shebang** line of a script

```
#!/usr/bin/python
```

- **First line in the code**
- **Tell the system what interpreter to use**

python /programs/pybedtools/bin/**intron_exon_reads.py**

or

/programs/pybedtools/bin/**intron_exon_reads.py**

```
ls -l /programs/pybedtools/bin/intron_exon_reads.py

-rwxr-xr-x 1 root root 3362 Mar  5 20:25 /programs/pybedtools/bin/intron_exon_reads.py
```

If you run this way, make sure the intron_exon_reads.py file is readable

```
python /programs/pybedtools/bin/intron_exon_reads.py
```

If you run this way, make sure the intron_exon_reads.py file is readable and executable

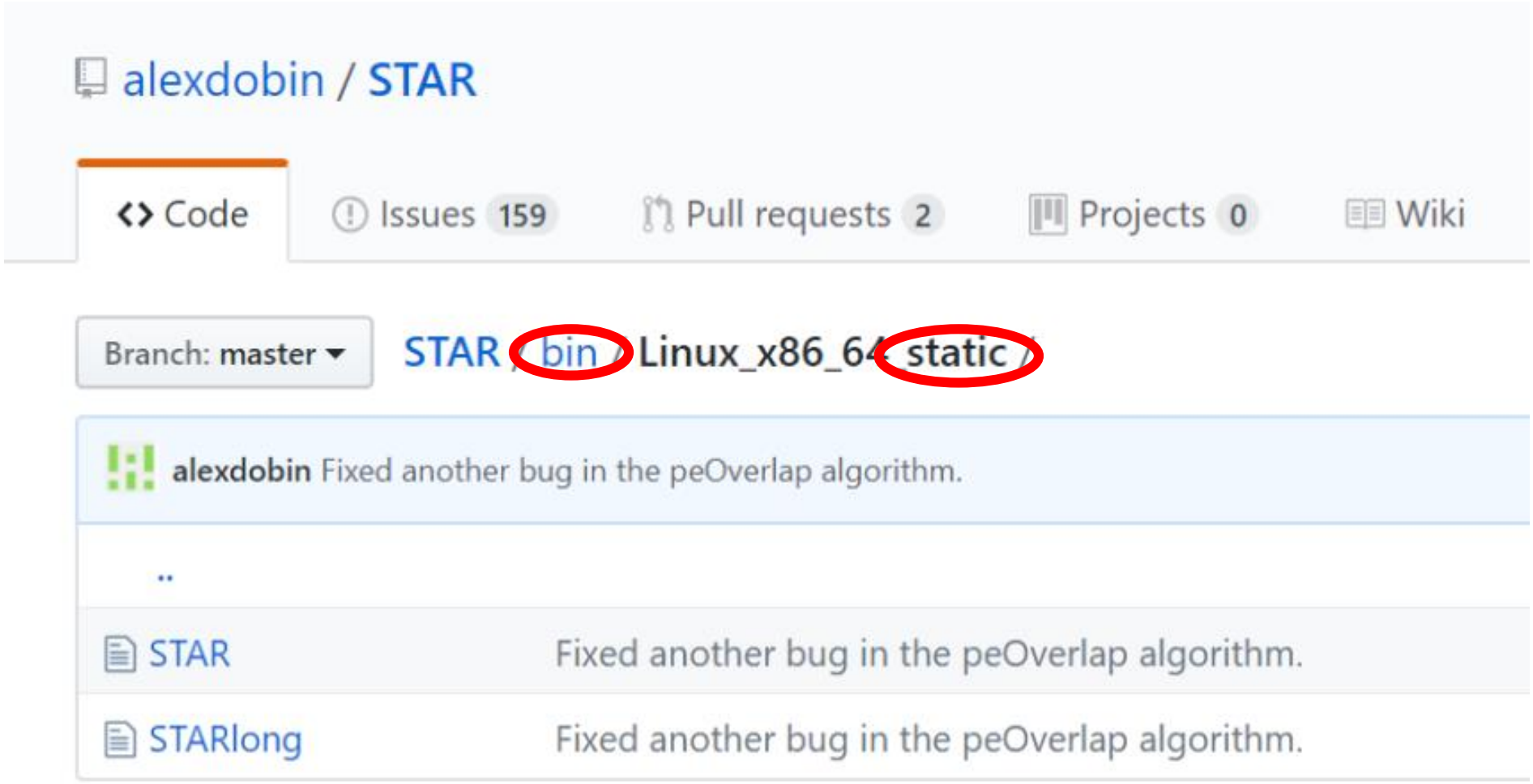
```
/programs/pybedtools/bin/intron_exon_reads.py
```

Part 3. Software installation - an overview

Many programs can be installed by simply downloading the code.

```
wget https://github.com/alexdobin/STAR/raw/master/bin/Linux_x86_64_static/STAR
```

If available, first try to download the “static” “binary” version.



Make the file executable

```
ls -l STAR
-rw-rw-r-- 1 qisun qisun 8805081 May  4 08:52 STAR

chmod uog+x STAR

ls -l STAR
-rwxrwxr-x 1 qisun qisun 8805081 May  4 08:52 STAR
```

Two ways to run the executable

Add the directory to the PATH

```
export PATH=/home/qisun/tools/:$PATH
STAR
```

Use full path

```
/home/qisun/tools/STAR
```

How to install multiple versions of the same software

STAR_2.3.0e.Linux_x86_64
STAR_2.4.0d
STAR_2.4.2a
STAR-2.5
STAR-2.5.2b
STAR-2.5.3a

```
export PATH=/programs/STAR-2.5/bin/Linux_x86_64_static:$PATH

which STAR

echo $PATH
```

Challenges in software installation –Part 1

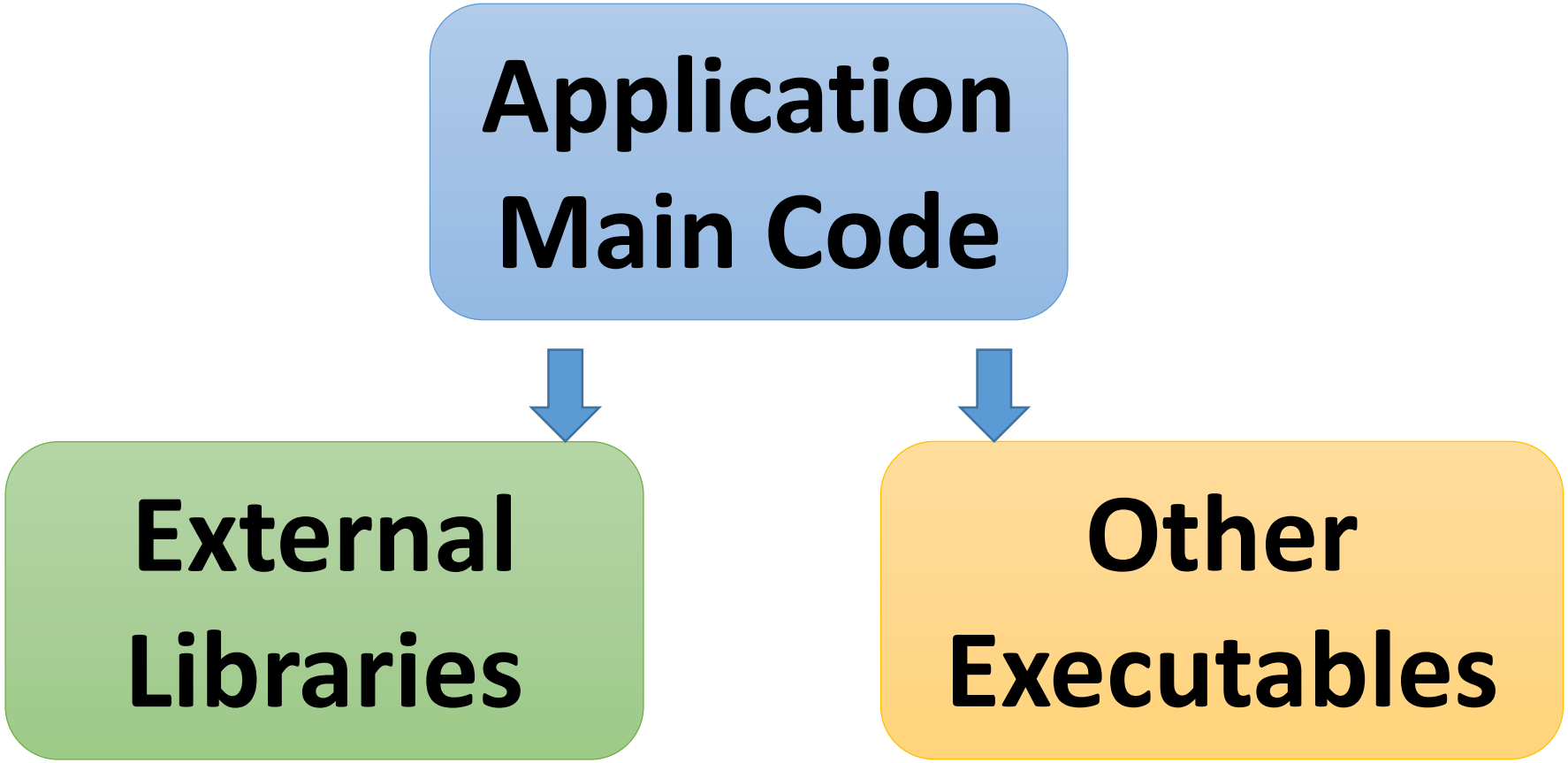
Version compatibility of libraries

Python software: HiCExplorer requires 15 modules

numpy==1.13.*
scipy==1.0.*
matplotlib==2.1.*
pysam==0.11.*
intervaltree==2.1.*
…

Latest numpy version: 1.14.3

Challenges in software installation –Part 2



If you have multiple versions of Numpy on your computer, you need to tell the main executable the path of the right version

Part 4. Conda - One of the solutions

Other solutions:

Language	Repository	Tool
PYTHON	PYPI	pip
PERL	CPAN	cpan
R	CRAN	R
Multiple	Anaconda	conda
Multiple	Docker Hub	docker
Multiple	Singularity Hub	singularity

Step 1. Install Conda in home directory

https://www.anaconda.com/download/#linux

A few more details when installing Conda

Adding Conda path to .bashrc?

You will be prompted to answer this question when install conda:

Do you wish the installer to prepend the Anaconda2 install location to PATH in your /home/qisun/.bashrc ?

no

Step 2. Install software in conda

#start conda

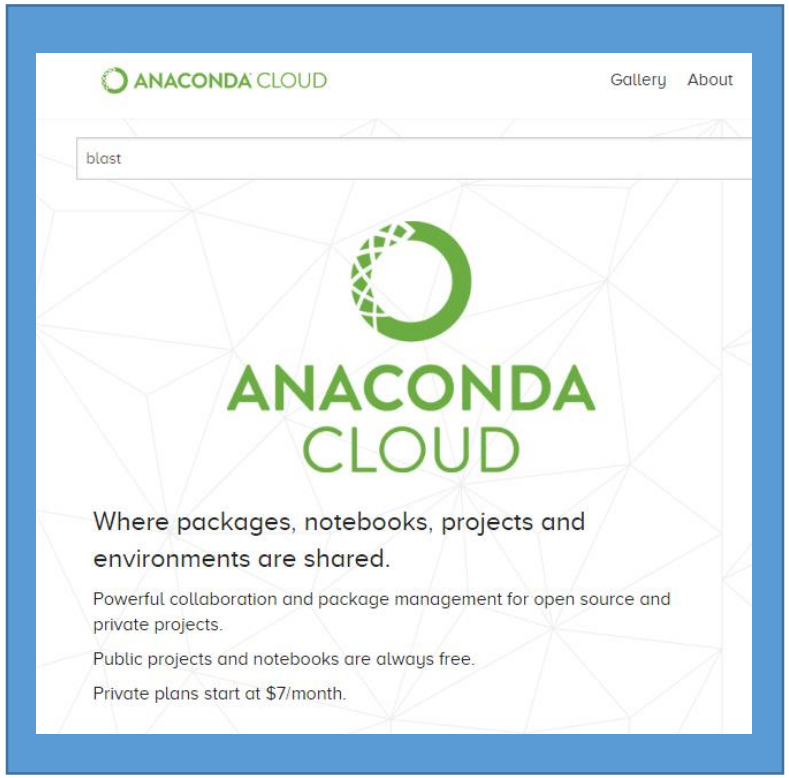
export PATH=/home/qisun/Anaconda2/bin:$PATH

#install software

conda install blast

Check the package availability

https://anaconda.org/



Search for BLAST returns:

Bioconda / blast 2.7.1

BioBuilds / blast 2.6.0

Current version at NCBI: 2.7.1

Check the software version in Conda before you use it

- **Specify channel when installing software.**

```
conda install -c bioconda blast
```

Or

```
conda config -add channels bioconda

conda install blast
```

Step 3. run software

#start conda

export PATH=/home/qisun/Anaconda2/bin:$PATH

#run software

blastn

How does Conda work?

Starting Conda:

export PATH=/home/qisun/Anaconda2/bin:$PATH

What is inside **/home/qisun/Anaconda2/bin:**

python
python3
Pip
…

You are using this copy of python, which manages its own libraries

Conda Virtual Environment

```
conda create -n myPipeLine python=3.4
```

Now, a new directory is created with all required libraries.

Run software in Conda virtual environment

```
# Start Conda virtual environment, add
environment directory to the $PATH:
source activate myPipeLine

# Run software
mySoftware

# Stop Conda virtual environment
source deactivate
```

A few more details when running software

Clear PYTHONPATH and LD_LIBRARY_PATH if they interfere with Conda

```
unset PYTHONPATH
```