Creating a Python Data Science Distribution for ARM with Conda

Jonathan Helmus, Software Engineer

Motivation

Installing software can be difficult

- especially scientific software
- even more so on ARM

Common solutions:

- system package managers
    - limited number of packages
    - often provide older versions
- language specific package managers
    - support for ARM is limited
    - requires compilers, libraries and other tools to be installed

CONDA®

- Conda is a cross platform **package** and **environment** management system
- Written and maintained by Anaconda, Inc
- Open Source, BSD licensed
- Creates reusable, relocatable packages for software written in any language
- Many **Python** and **R** data science, machine learning and AI frameworks
- Typically available by installing the **Anaconda Distribution** or **Miniconda**

package management with conda

- Packages are **binaries**, no compiler or libraries are needed
- Does not require administrator privileges
- Uses a SAT solver for dependency resolution

Typical commands:

- **conda install** : install one or more package(s)
- **conda remove** : remove a package
- **conda update** : update a package
- **conda list** : list the installed packages

conda environments

Conda can create isolated environments that have their own set of packages.

- **conda create** : create a new conda environment
- **conda activate** : activate a conda environment
- **conda deactivate** : deactivate the current conda environment

Great when you need to work with different versions of a library or application.

Environment specification can be exported to a file and recreated.

package tools: conda-build and constructor

conda-build is a tool to create conda packages.

Packages are built from **recipes** which specify package metadata and build steps.

Process

- create an isolated build environment
- execute build steps
- bundle files
- create a test environment
- execute tests

constructor is a tool for creating **installers** from conda packages

conda channels

- Conda packages are provided from different repositories, called **channels**.
- Out of the box, conda installs packages from the “**defaults**” channel.
- Other channels can be enabled to access **additional** collections of packages
- Channels are hosted for free on **Anaconda Cloud** (anaconda.org)

Some key channels are:

- **defaults** : packages from Anaconda, Inc
- **conda-forge** : large community led collection of packages
- **bioconda** : community specializing in bioinformatics packages

CONDA-FORGE

BIOCONDA®

conda-forge

- Numfocus-affiliated community organization made up of volunteers
- One GitHub repository per recipe
- Fine granularity over permissions
- Heavy use of automation for building, deploying, and updating recipes
- Packages built on public CI services (TravisCI, CircleCI, Appveyor, Azure)
- https://conda-forge.org/

CONDA-FORGE

conda on ARM

For conda to support ARM:

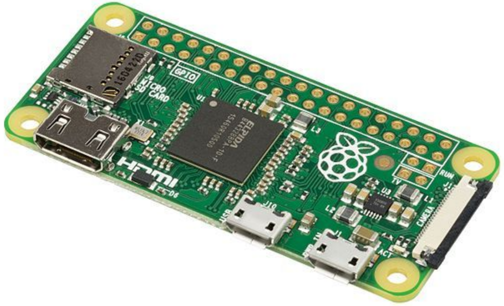
conda itself must recognize the platform

- linux-armv6l supported in conda 1.5.0 (2013)
- linux-armv7l supported in conda 3.15.0 (2015)
- linux-aarch64 supported in conda 4.3.18 (2017)

Installers and **packages** must be available for the platform.

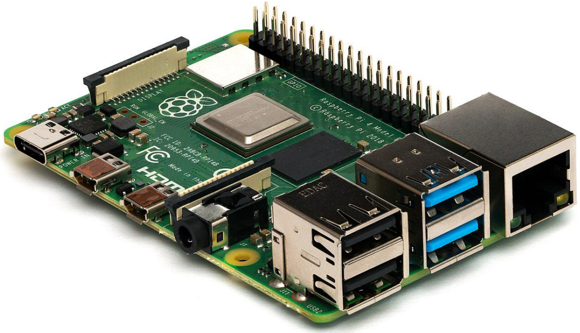
Conda on 32-bit ARM (armv6l and armv7l)

My initial interest in ARM started with a Raspberry Pi



Anaconda (then Continuum Analytics) provided Miniconda and a small collection of packages for these SBCs

I wanted more packages and up-to-date versions



Berryconda was born!

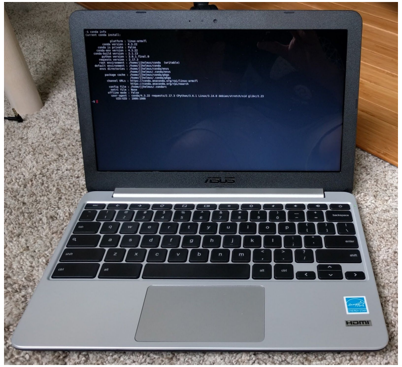
Berryconda

- A conda based Python distribution for the Raspberry Pi
- Installers and packages for:
  - **armv6l** : RPi Zero and 1
  - **armv7l** : RPi 2, 3 and 4
- Designed to work on **Raspbian**
- Over 300 packages in the **rpi channel** on Anaconda Cloud
- Python 2 and 3
- Recipes derived from conda-forge
- Testing ground for noarch packages and NumPy pinning now widely used

https://github.com/jjhelmus/berryconda

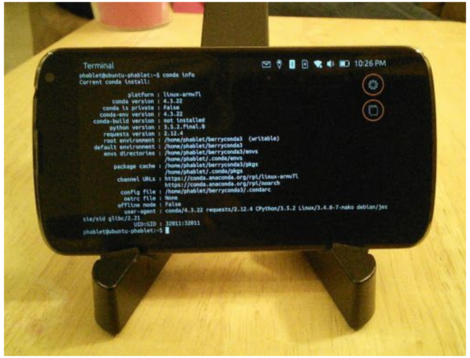
Berryconda

Works on other 32-bit ARM systems



```
~$ conda info
Current conda install:

               platform : linux-armv7l
          conda version : 4.3.22
       conda is private : False
      conda-env version : 4.3.22
    conda-build version : 2.1.13
         python version : 3.6.1.final.0
       requests version : 2.17.3
       root environment : /home/jjhelmus/conda  (writable)
    default environment : /home/jjhelmus/conda
       envs directories : /home/jjhelmus/conda/envs
                          /home/jjhelmus/.conda/envs
          package cache : /home/jjhelmus/conda/pkgs
                          /home/jjhelmus/.conda/pkgs
           channel URLs : https://conda.anaconda.org/rpi/linux-armv7l
                          https://conda.anaconda.org/rpi/noarch
            config file : /home/jjhelmus/.condarc
             netrc file : None
           offline mode : False
             user-agent : conda/4.3.22 requests/2.17.3 CPython/3.6.1 Linux/3.14.0 debian/stretch/sid glibc/2.23
                UID:GID : 1000:1000
~$
```



```
phablet@ubuntu-phablet:~$ conda info
Current conda install:

               platform : linux-armv7l
          conda version : 4.3.22
       conda is private : False
      conda-env version : 4.3.22
    conda-build version : not installed
         python version : 3.5.2.final.0
       requests version : 2.12.4
       root environment : /home/phablet/berryconda3  (writable)
    default environment : /home/phablet/berryconda3
       envs directories : /home/phablet/berryconda3/envs
                          /home/phablet/.conda/envs
          package cache : /home/phablet/berryconda3/pkgs
                          /home/phablet/.conda/pkgs
           channel URLs : https://conda.anaconda.org/rpi/linux-armv7l
                          https://conda.anaconda.org/rpi/noarch
            config file : /home/phablet/berryconda3/.condarc
             netrc file : None
           offline mode : False
             user-agent : conda/4.3.22 requests/2.12.4 CPython/3.5.2 Linux/3.4.0-7-mako debian/jes
sie/sid glibc/2.21
                UID:GID : 32011:32011
phablet@ubuntu-phablet:~$
```

But **not** a general purpose 32-bit ARM distribution

- uses compilers and libraries from Raspbian
- targets Raspberry Pi specific floating point instructions (e.g. neon-vfpv4)

conda for 64-bit ARM (AArch64)

Many Raspberry Pi clones have 64-bit ARMv8 processors

Can we run conda on these systems?

- No packages available
- No installer available

Need to **bootstrap** a new platform, create an **installer** and then build **packages**

Bootstrapping a new platform in conda

Installing Miniconda is not an option for new platforms.

To bootstrap a new platform:

- install or build Python dependencies (bzip2, zlib, etc) from source
- compile Python from source
- pip install conda dependencies (six, tqdm, requests, etc)
- install/compile non-pip conda dependencies from source (cph, libarchive)
- install conda (and conda-build) from source

Docker containers are great for bootstrapping, for example:

https://github.com/jjhelmus/conda_from_scratch

Build an AArch64 installer

Constructor can be used to build an AArch64 installer.

~130 packages needed:

- compilers : gcc 7.3.0, binutils 2.29.1
- build tools : make, autotools, ...
- libraries : zlib, sqlite, readline, …
- Python and Python libraries
- conda, conda-build, constructor, ...

Packages built on a ROCKPro64, available in the c4aarch64 channel

Recipes, build tools, and an installer: https://github.com/jjhelmus/conda4aarch64

Building AArch64 packages on conda-forge

To build packages for the linux-aarch64 platform on conda-forge:

- Create a docker image for the platform
- Teach conda-smithy about linux-aarch64
- Define a migrator for the platform

CONDA-FORGE

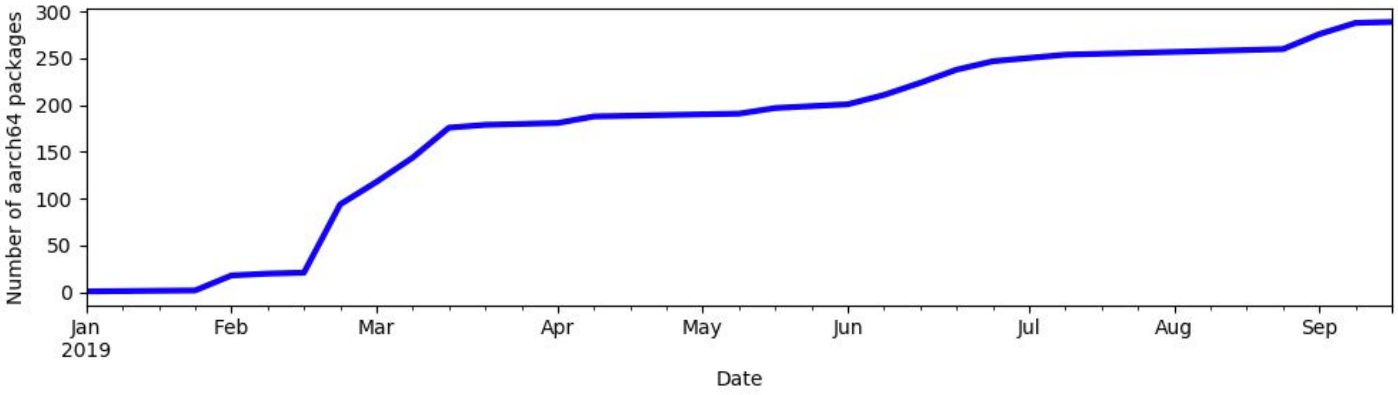
With the migrator defined a conda-forge bots submits pull requests to the feedstocks adding builds for the new platform.

Packages are built and uploaded to the conda-forge channel

Lots of work done by Marius van Niekerk and Mark Harfouche

conda-forge AArch64 status

Conda-forge AArch64 migration was started in January, 2019



Currently 289 AArch64 packages are available along with 4260 noarch packages

Migrator Status: https://conda-forge.org/status/

Contributing

Interesting in conda on ARM?

- Participate in conda-forge discussions: https://gitter.im/conda-forge/conda-forge.github.io
- Help with the aarch64 migration: https://conda-forge.org/status/
- Write documentation: https://conda-forge.org/docs/
- Champion building armv7l packages on conda-forge
  - https://github.com/jjhelmus/berryconda/pull/40
  - https://github.com/jjhelmus/conda4armv7l/