Day 4: Essential pieces of reproducibility - Conda, Jupyter, Galaxy, Blogs

Monika Cechova, Anton Nekrutenko | biomonika@psu.edu, anton@nekrut.org

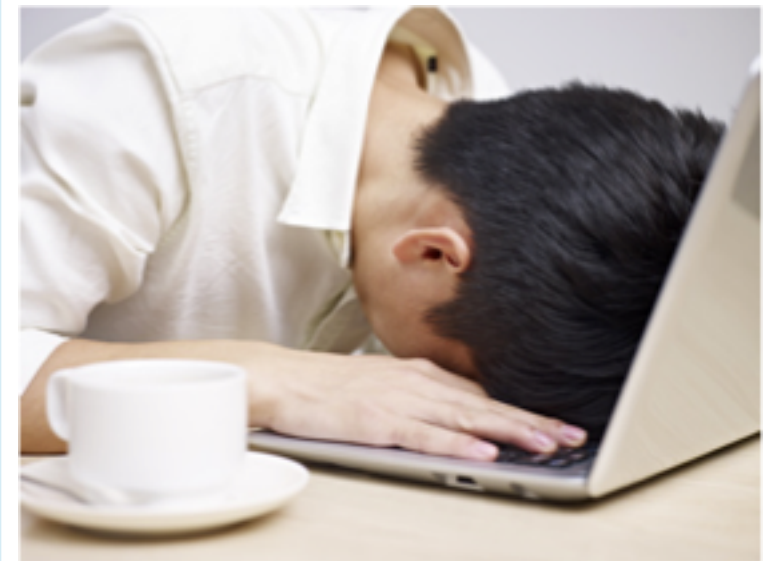Conda and Bioconda

BLACKDUCK | Hub
Hub Detect

python Gradle CONDA

Maven CPAN sbt npm

Linux Package Managers:
APK, DPKG, RPMs

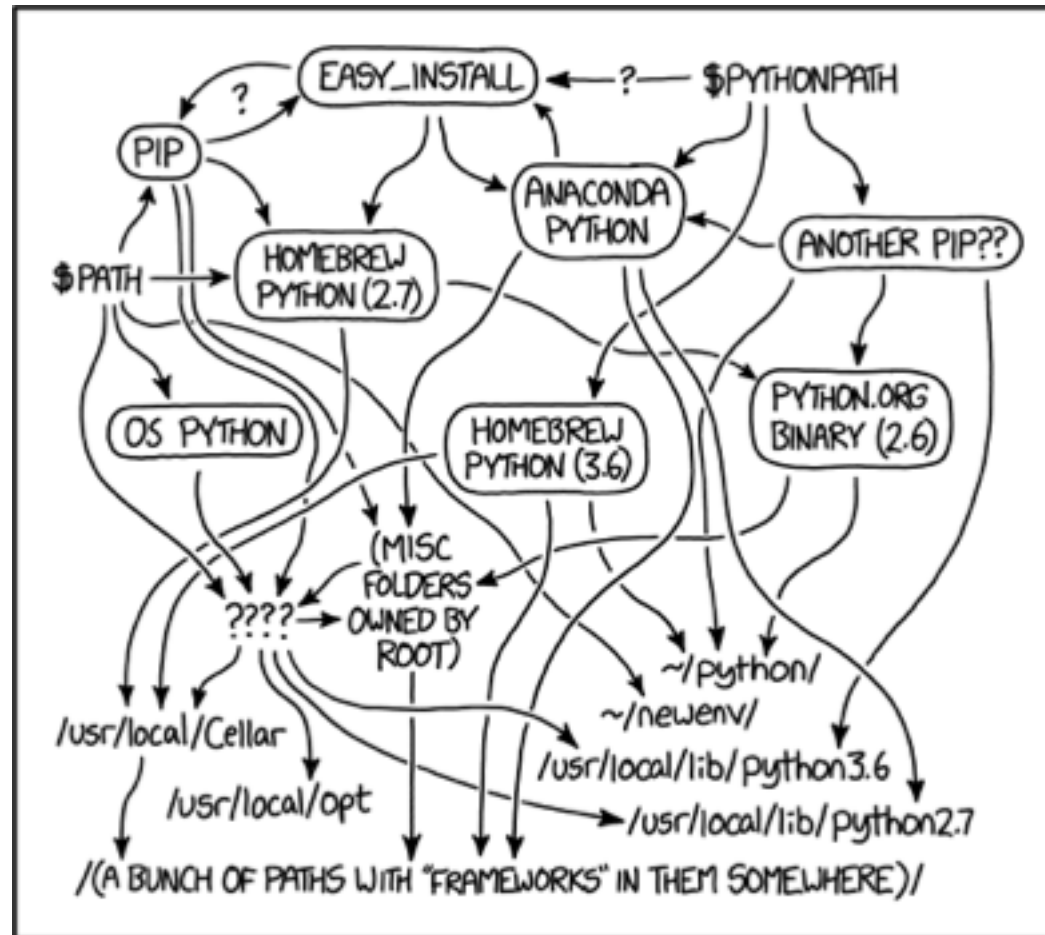Microsoft Visual Studio Team Services TC Bamboo

From: https://www.cad-notes.com/wp-content/uploads/2015/11/frustrated-computer-user.jpg

Conda

Package, dependency and environment management for any language

"Conda is an open source package management system and environment management system that runs on Windows, macOS and Linux. Conda quickly installs, runs and updates packages and their dependencies. Conda easily creates, saves, loads and switches between environments on your local computer. It was created for Python programs, but it can package and distribute software for any language."

from: https://conda.io/docs/

MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

which python

Recommended [blog post](#)

🔒 Secure | https://medium.freecodecamp.org/why-you-need-python-environments-and-how-to-manage-them-with-conda-85f155f4353c ⭐

M | freeCodeCamp(🔥)   [Follow] 🐦 ⓕ   Sign in [Get started]

HOME   DEV   DESIGN   DATA | LEARN TO CODE FOR FREE   🔍
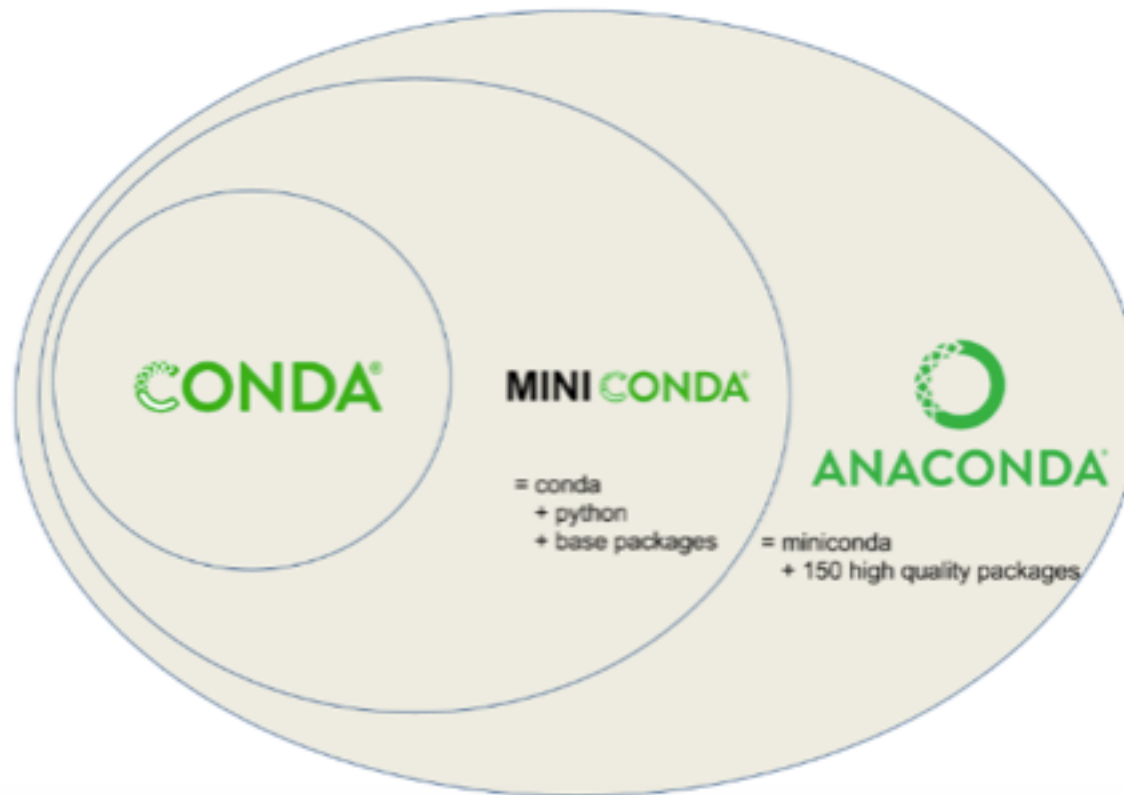
Gergely Szerovay [Follow]
Freelance IT Engineer, developer with over two decades of experience
Feb 21 · 18 min read

Why you need Python environments and how to manage them with Conda

Slide from Galaxy Team and John Chilton

Conda can be used to install:

- Python (with disclaimer)

- Python packages

- R

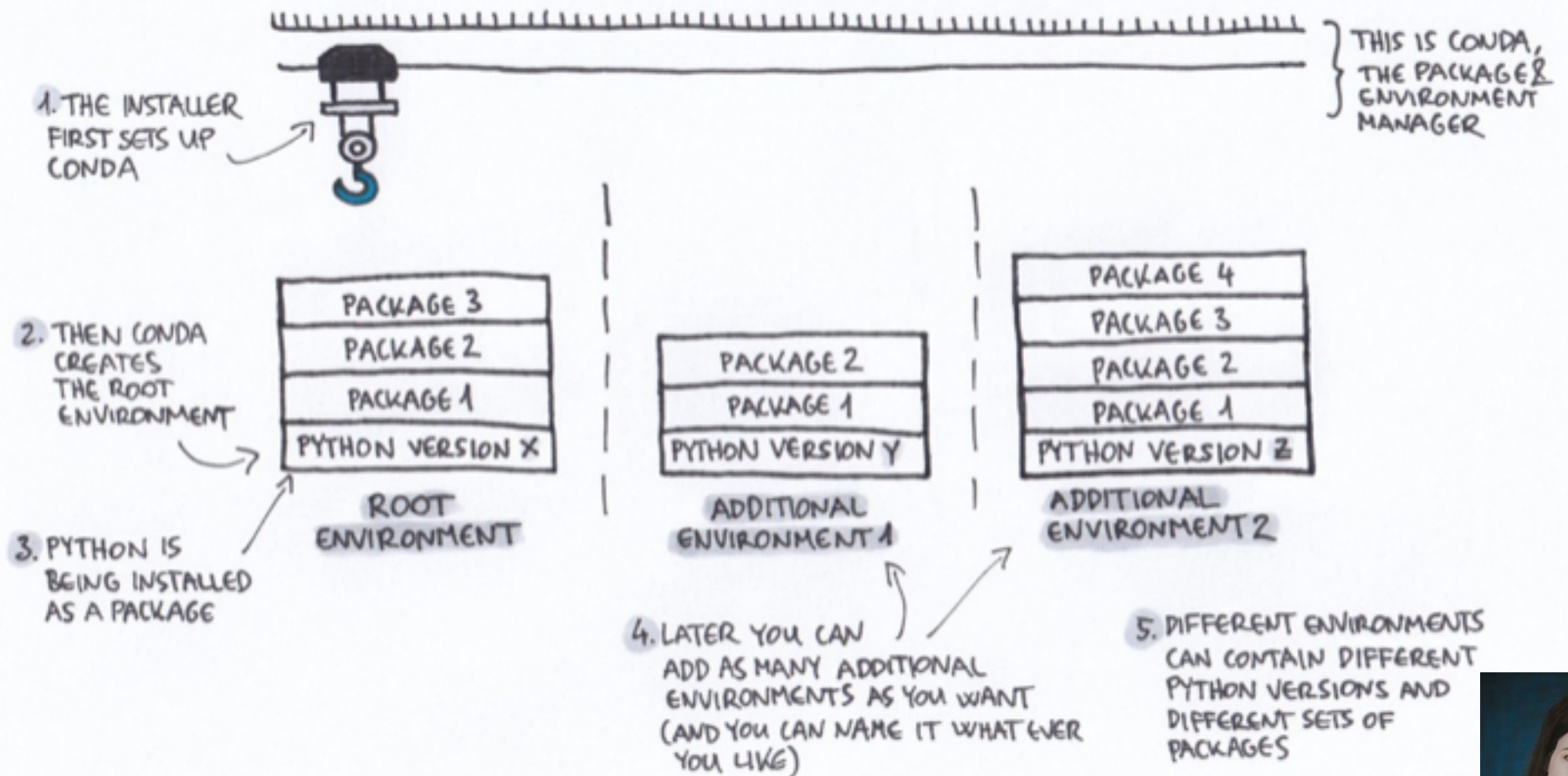- Tools and packages for bioinformatics and data science

✓ Specific versions

1. THE INSTALLER FIRST SETS UP CONDA

THIS IS CONDA, THE PACKAGE & ENVIRONMENT MANAGER

2. THEN CONDA CREATES THE ROOT ENVIRONMENT

3. PYTHON IS BEING INSTALLED AS A PACKAGE

PACKAGE 3
PACKAGE 2
PACKAGE 1
PYTHON VERSION X
ROOT ENVIRONMENT

PACKAGE 2
PACKAGE 1
PYTHON VERSION Y
ADDITIONAL ENVIRONMENT 1

PACKAGE 4
PACKAGE 3
PACKAGE 2
PACKAGE 1
PYTHON VERSION Z
ADDITIONAL ENVIRONMENT 2

4. LATER YOU CAN ADD AS MANY ADDITIONAL ENVIRONMENTS AS YOU WANT (AND YOU CAN NAME IT WHATEVER YOU LIKE)

5. DIFFERENT ENVIRONMENTS CAN CONTAIN DIFFERENT PYTHON VERSIONS AND DIFFERENT SETS OF PACKAGES

Illustration by Krisztina Szerovay

CHANNELS ARE LIKE STORAGES...

CONDA IS

LOOKING FOR
[blue hatched box] THIS PACKAGE

CHANNEL 1

CHANNEL 2

CHANNEL 3

CHANNEL 4

FIRST, IT IS LOOKING FOR THE PACKAGE IN THIS CHANNEL (THIS HAS THE HIGHEST PRIORITY)

IT WAS NOT IN THE FIRST CHANNEL, SO IT MOVES TO THE STORAGE WITH THE 2ND HIGHEST PRIORITY

YAY! CONDA HAS FOUND THE PACKAGE! NOW IT IS ADDED TO YOUR ENVIRONMENT.

THEY ARE CONDA'S DEVELOPERS.
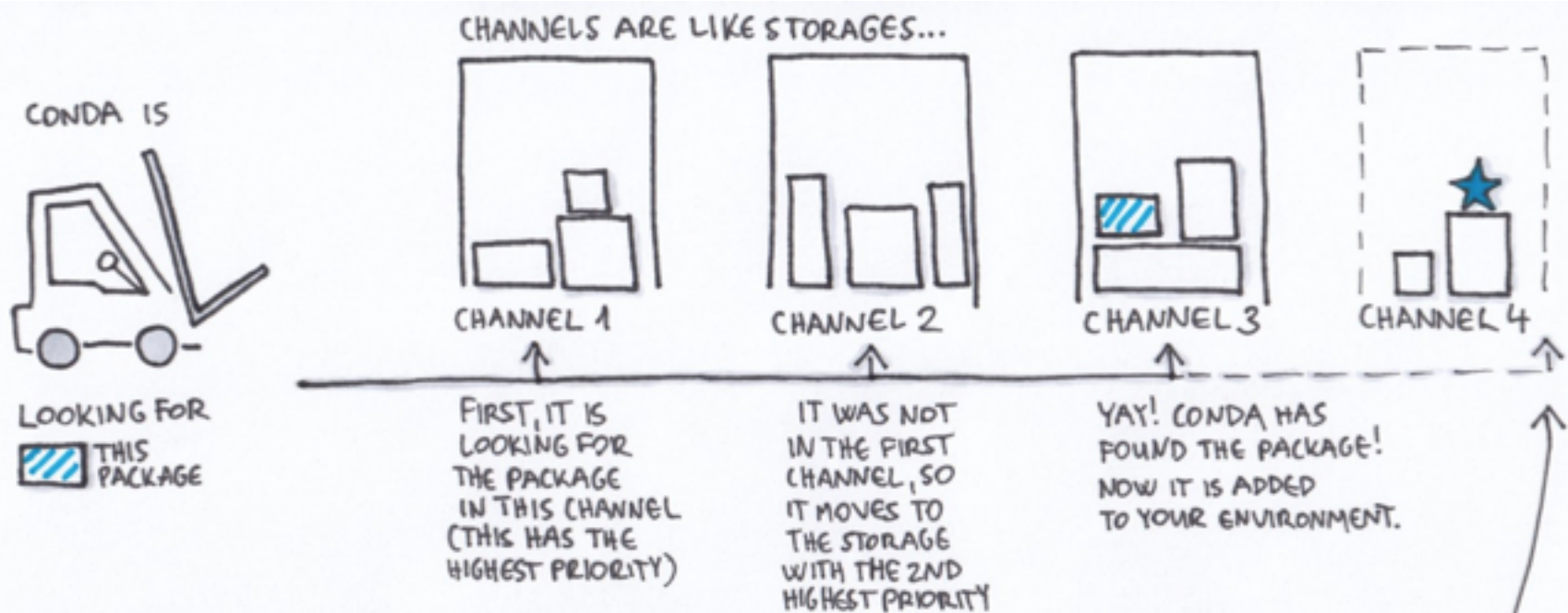
WHAT HAPPENS IF CONDA DID NOT FIND THE PACKAGE?

BY DEFAULT, CONDA LOOKS FOR PACKAGES IN THE OFFICIAL STORAGES OF CONTINUUM.

WHY? BECAUSE BY DEFAULT THESE HAVE THE HIGHEST PRIORITIES.

HOWEVER, YOU HAVE THE POWER TO

[plus icon] ADD NEW CHANNELS (STORAGES) THAT CONTAIN THE PACKAGES YOU NEED!

SO LET'S SAY YOU WANT TO INSTALL THIS PACKAGE : [star]
→ CONDA DOES NOT FIND IT IN THE FIRST 3 CHANNELS
⇒ YOU NEED TO ADD A 4TH ONE!

Illustration by Krisztina Szerovay

Questions?

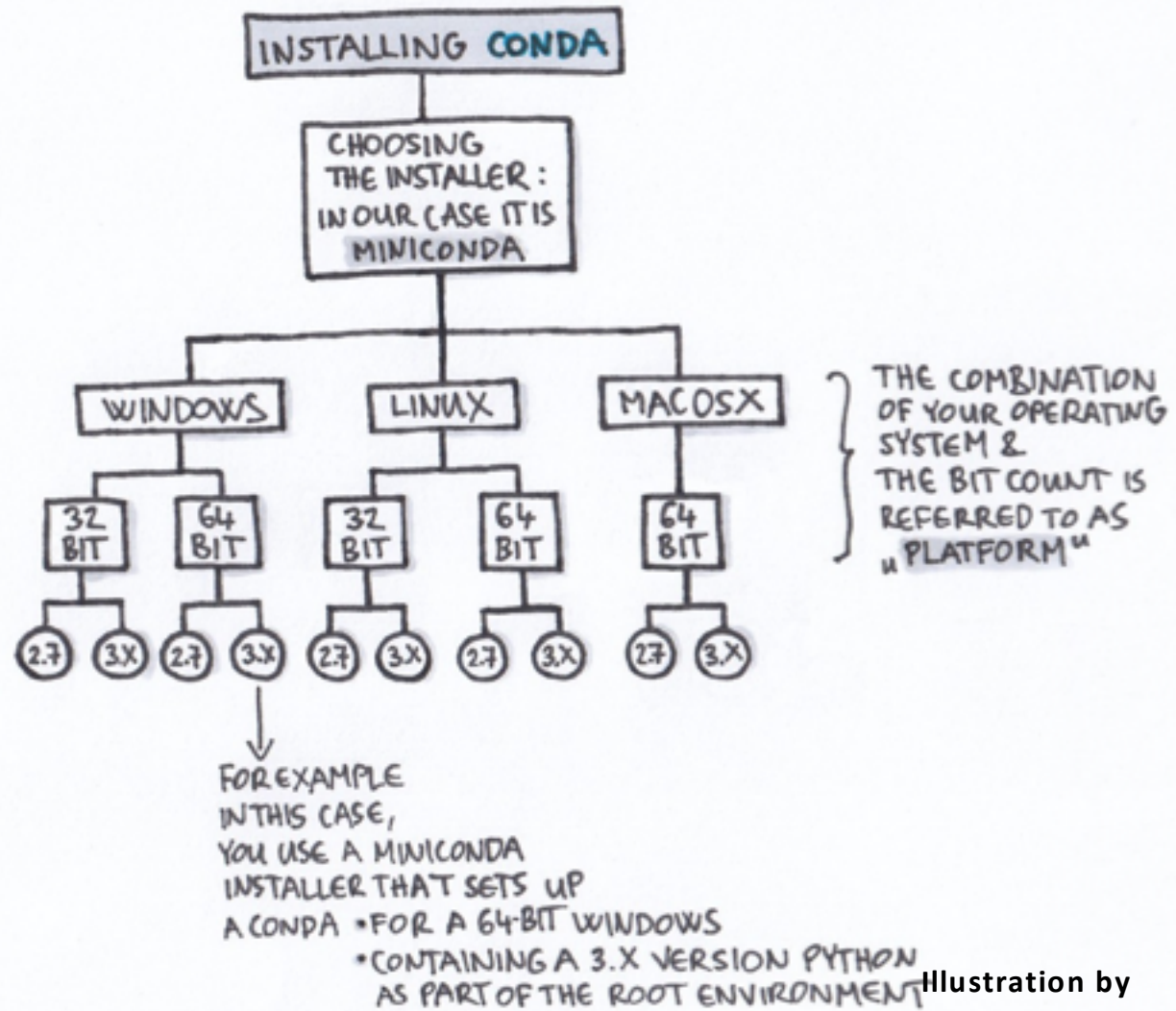https://conda.io/docs/user-guide/install/index.html

General conda commands

- conda --help
- conda list   #all packages available in a given environment
- conda env list   #lists all environments

Create and activate new environment

- conda create --name test
- source activate test

- conda install -c bioconda plink
- conda install -c bioconda minimap
- conda install -c bioconda bioawk
- conda install --channel r r-essentials=1.4
- conda install --channel r r=3.3.1

- source deactivate

Looking for packages to install

Newest version:

Google *conda name_of_the_package*

Which versions of the package are available?
conda search -f samtools

Power of bioconda

- conda search --channel bioconda
- conda search --channel bioconda | wc -l

Installing perl modules via conda

- conda create --name perly_env
- source activate perly_env

- conda install -c anaconda perl # Installing perl
- conda install -c bioconda perl-app-cpanminus # Installing CPAN for managing perl modules
- If -c bioconda does not work, try:

  conda install -c conda-forge perl-app-cpanminus

- cpanm Bio::Perl # Getting any module you like

Questions?