Securing Python Applications with Thoth Recommendations

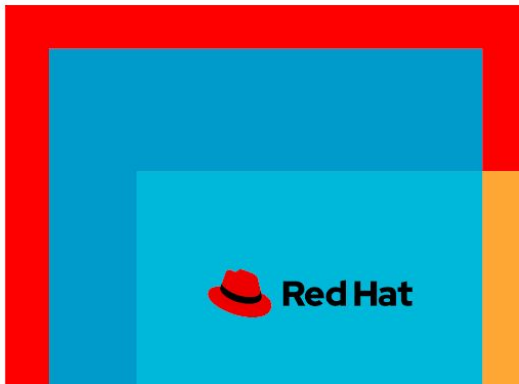
Thoth team - thoth-station.ninja

Fridolín Pokorný
Senior Software Engineer, Red Hat

Maya Costantini
Associate Software Engineer, Red Hat

Agenda

1. $ whoweare
2. Our mission
3. Why should you use Thoth?
4. Thoth's resolution process
5. Heal your Python applications with prescriptions
6. Securing your applications with Thoth



$whoweare

- Thoth - AIDevSecOps
  - Started in 2018 as a research project in Red Hat AICoE team, Office of the CTO
  - thoth-station.ninja

- See our linked YouTube channel for more information
- Follow us on Twitter - @ThothStation

Red Hat

Our mission

- Help Python developers and data scientists create healthy applications

- Project has multiple parts:
  - AICoE-CI - a CI that builds container images
  - Thoth resolver - a recommendation engine for Python applications
    - AIDevSecOps
  - Dependency Monkey - a service that can validate software in a cluster
  - jupyterlab-requirements extension for managing dependencies
  - Bots maintaining GitHub repositories
  - A self hosted Python package index using Pulp available to all Red Hatters
  - Container image analysis and containerized Python applications

Red Hat

Dependencies can be a source of vulnerabilities in a project

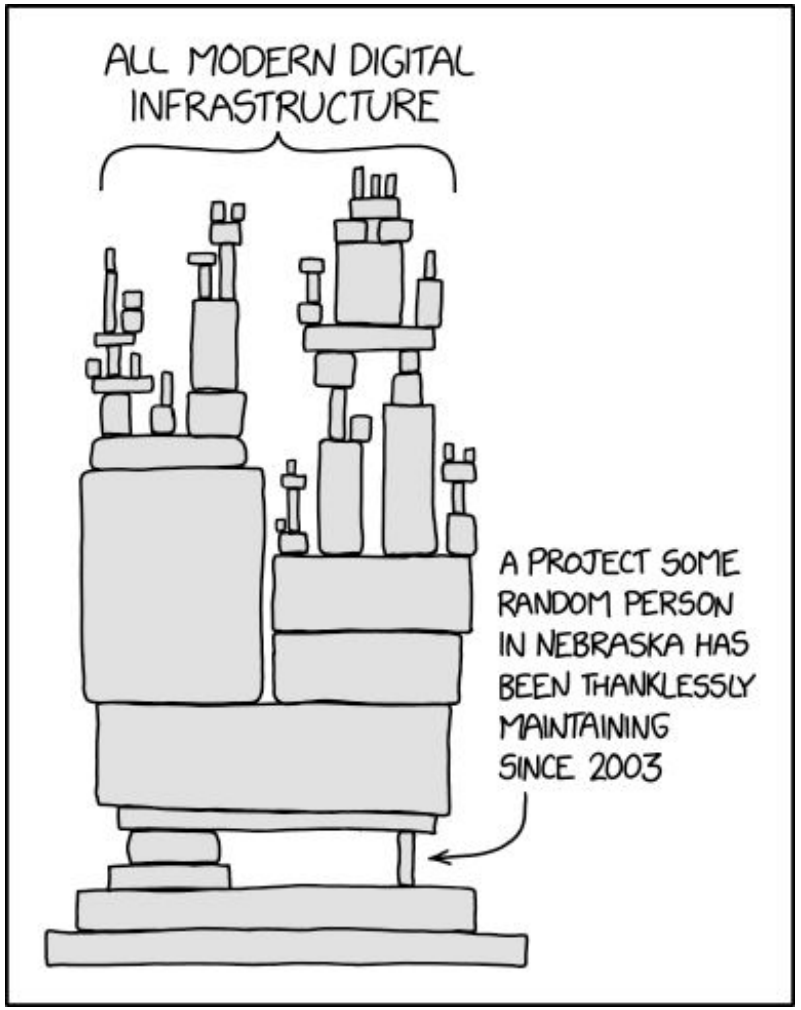
- *"How one developer just broke Node, Babel and thousands of projects in 11 lines of JavaScript"* [1]
- *"Log4j: Google and IBM call for list of critical open source projects"* [2]
- *"Dev corrupts NPM libs 'colors' and 'faker' breaking thousands of apps"* [3]

[1] https://www.theregister.com/2016/03/23/npm_left_pad_chaos/

[2] https://www.zdnet.com/article/log4j-after-white-house-meeting-google-calls-for-list-of-critical-open-source-projects/

[3] https://www.bleepingcomputer.com/news/security/dev-corrupts-npm-libs-colors-and-faker-breaking-thousands-of-apps/

Why should you use Thoth?

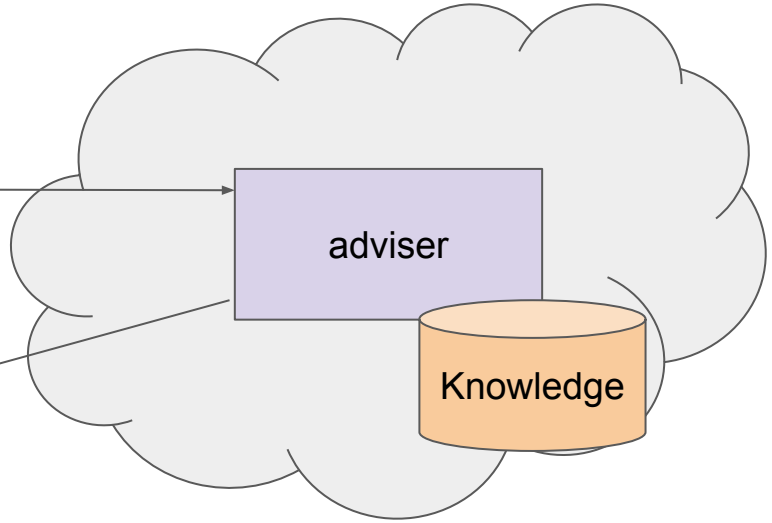


Source: https://xkcd.com/2347/

Red Hat

Python cloud resolver

Requirements
Constraints
Information about software in runtime environment
 - OS, Python version, base image, CUDA, …
Information about hardware in runtime environment
 - CPU, GPU, …
Static source-code analysis
Recommendation type



Lockfile + justification

Heal your Python applications with prescriptions

- Provide a way to declaratively state how the resolution process should look like
- Community driven open database used by the resolver to resolve high quality software - you can contribute!
  - https://github.com/thoth-station/prescriptions/
- A set of YAML files that are automatically consumed by the resolver in a deployment

Prescriptions - Example

💊 Pillow in version 8.3.0 does not work with NumPy

https://github.com/python-pillow/Pillow/issues/5571

```
with PIL.Image.open(filepath) as img:
      numpy.array(img, dtype=numpy.float32)
```

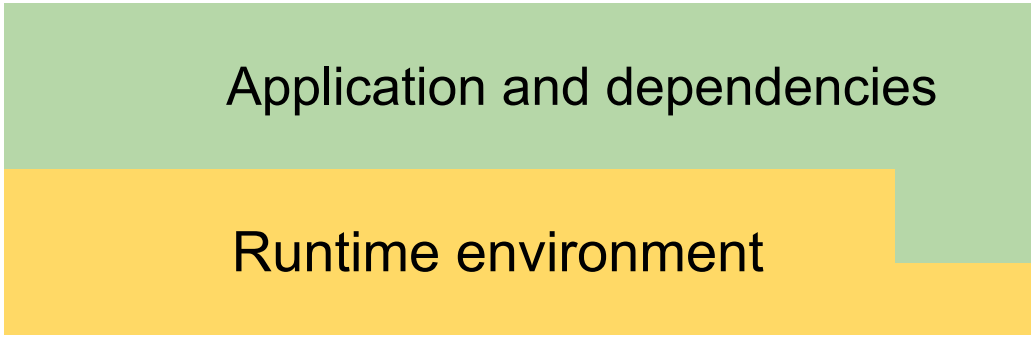
```
>               frame_paletted = np.array(im, np.uint8)
E               TypeError: __array__() takes 1 positional argument but 2 were given

  /lib/python3.9/site-packages/imageio/plugins/pillow.py:745: TypeError
```

```
units:
 steps:
 - name: Pillow830TypeErrorStep
   type: step
   should_include:
     adviser_pipeline: true
   match:
     package_version:
       name: pillow
       version: ==8.3.0
       index_url: https://pypi.org/simple
     state:
       resolved_dependencies:
       - name: numpy
   run:
     not_acceptable: Pillow in version 8.3.0 does not work with NumPy
     stack_info:
     - type: WARNING
       message: Pillow in version 8.3.0 does not work with NumPy
       link: https://github.com/python-pillow/Pillow/issues/5571
```

Fitting the runtime environment

- Checking RPM, Python packages, ABI, CUDA, cuDNN, OpenMKL, ...
- One central place to declare requirements
  - Prescriptions (packages)
  - User's runtime environment (Pipfile, .thoth.yaml file)
- Still optional, the resolution might be "generic", like pip, Pipenv, ...



Security - AIDevSecOps

- Docs: Thoth security advises
- Recommendations based on static source code analysis
  - See recommendations from the Python standard library (example)
- PyPA - advisory-db
  - A database of known vulnerabilities in Python ecosystem
  - https://github.com/pypa/advisory-db
- Security Scorecards by Open Source Security Foundation
  - https://openssf.org/blog/2020/11/06/security-scorecards-for-open-source-projects/
  - Example: see scorecards_prefixed prescriptions for TensorFlow

+ additional information about Python packages not strictly related to security

Demo: Using Thamos to manage vulnerabilities in your application

```
pip install thamos
thamos config
thamos add 'pillow==8.0.0'
thamos advise
```

Look at the tutorial documentation* to reproduce the demo

*https://redhat-scholars.github.io/managing-vulnerabilities-with-thoth

Red Hat Summit

Thank you


linkedin.com/company/red-hat


facebook.com/redhatinc


youtube.com/user/RedHatVideos


twitter.com/RedHat

Red Hat