developer insights: machine learn and analytics on src/

DevConf.CZ 2020
Friday, 2020-Jan-24 at D0207

Christoph Görn,
Engineering Manager, AI CoE

Francesco Murdaca
Data Scientist, AI CoE

Red Hat

How to choose the dependency that fits the developer's need?

Things get faster, maybe more regulated

Things get critical, and more complex

How to stay up to date?

- Research more framework alternatives
- Test with different versions of your framework
- Analyse more build logs/failures
- What else… ?

How to stay up to date?

- Research more framework alternatives
- Test with different versions of your framework
- Analyse more build logs/failures
- What else… ?

More research => less coding?!

Ask others...

- to do the mundane work
- Identify patterns and drawbacks
- Give some advice on your code

Ask others...

- to do the mundane work
- Identify patterns and drawbacks
- Give some advice on your code

Cyborgs to the rescue?!

Project Thoth

Create a knowledge graph that can be used by cyborgs to give some good advice to developers

https://thoth-station.ninja/
https://github.com/thoth-station/

goals

The AI Center of Excellence's Project Thoth has three strategic objectives

- Deliver optimized AI stacks as container images to demonstrate the concept
- Prototype AI-backed guidance to delivering value add information on application stacks
- Advanced Analytics and AI-backed automation CI/CD and software maintenance

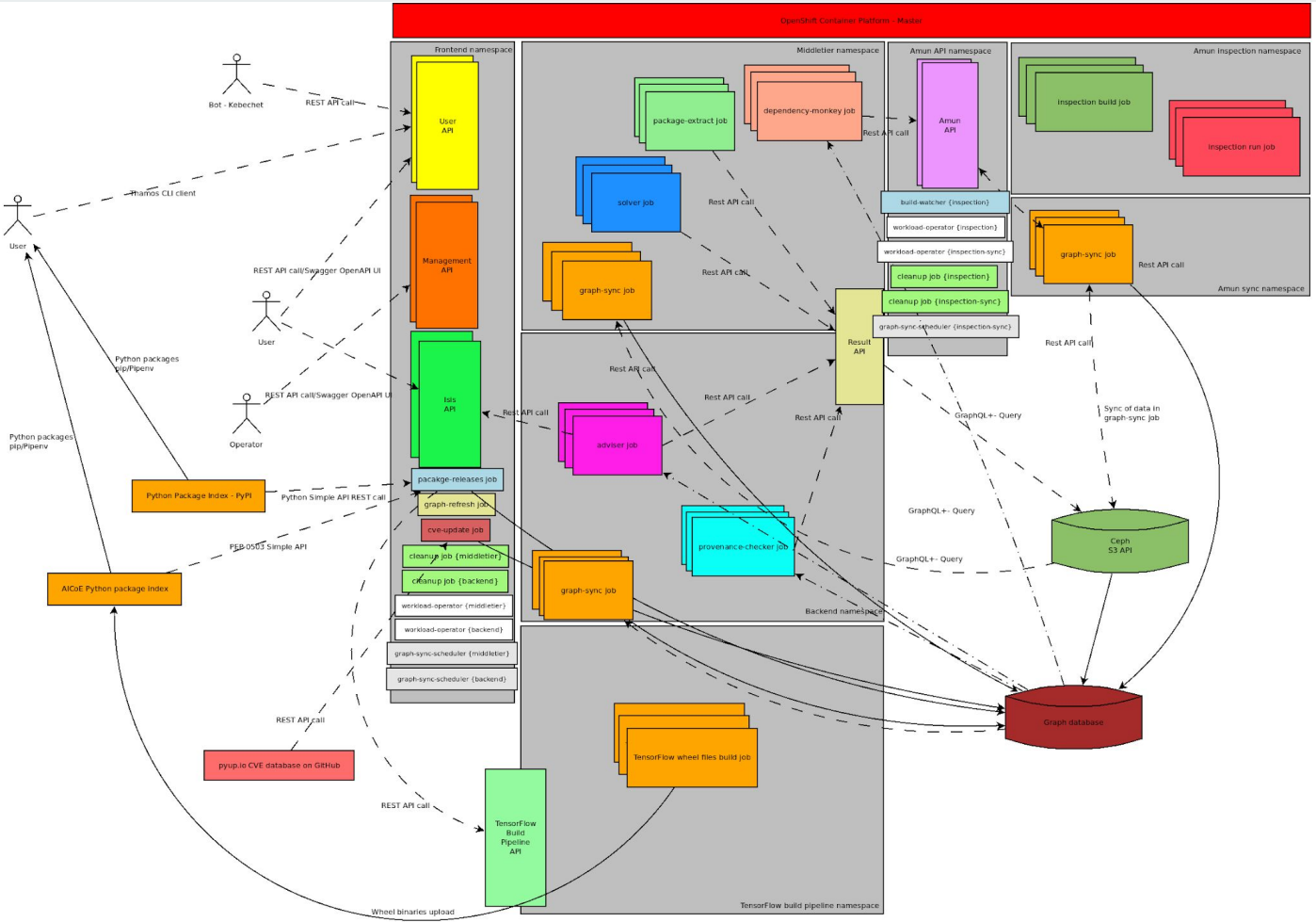
goals

Increase developer velocity

The AI Center of Excellence's Project Thoth has three strategic objectives

- Deliver optimized AI stacks as container images to demonstrate the concept
- Prototype AI-backed guidance to delivering value add information on application stacks
- Advanced Analytics and AI-backed automation CI/CD and software maintenance

Arch overview



what we observe and store in our knowledge graph

Application Stack related:

- Buildtime and runtime environment
- Dependencies
- Performances

Software Package:

- Application Binary Interfaces (ABI)
- CVE

Source Code Meta Information:

- Project features (TTR, TTCI, etc,..) from different software development platform (Github, GitLab, Pagure, etc…)

what we observe and store in our knowledge graph

Application Stack related:

- Buildtime and runtime environment
- Dependencies

Thoth solver

Analyze each Python package version

- Get its dependencies by installing it
- Aggregate information about possible installation errors
  - Analysis of installation logs
  - Package was not installed due to

what we observe and store in our knowledge graph

Application Stack related:

- Performances

Thoth Dependency Monkey and Amun

- Creates combination of software stacks and create inputs for Amun
- Run Performance Indicator to gather performance observations:
  - ML frameworks
  - Python Interpreter

what we observe and store in our knowledge graph

Package-extract

- Extract dependencies from an image and image build logs

Software Package:

- Application Binary Interfaces (ABI)
- CVE

CVE Update job

- checks for new CVE records in the safety-db provided by PyUp.io.
- store them in the database

what we observe and store in our knowledge graph

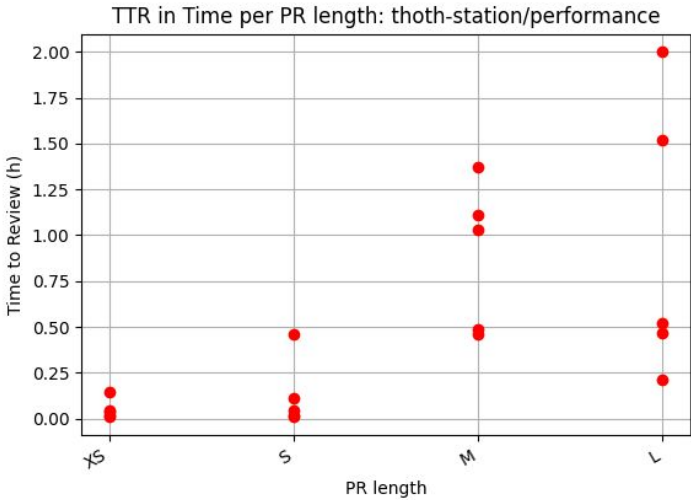
Src Ops Metrics

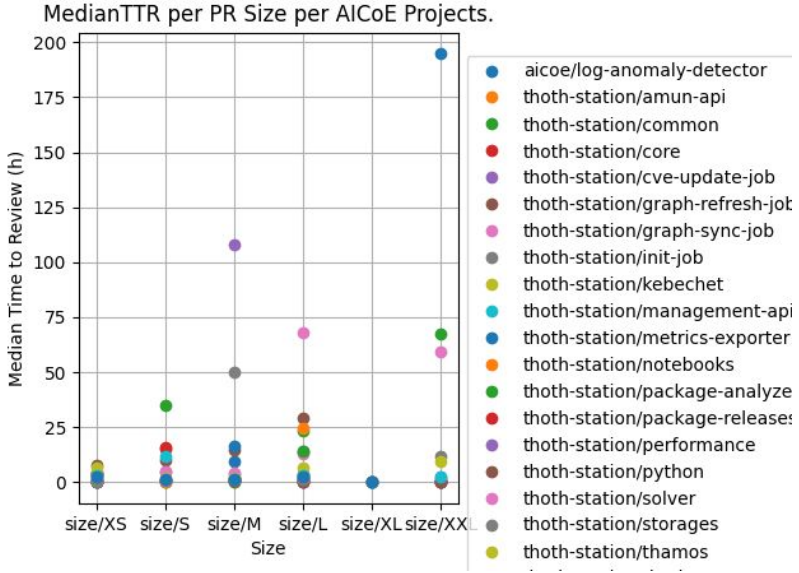
- Gather information from software development platforms:
  - GitHub
  - GitLab
  - Pagure

Source Code Meta Information:

- Project features (TTR, TTCI, etc,..)

what we observe and store in our knowledge graph





Source Code Meta Information:

- Project features (TTR, TTCI, etc,..)

How to choose the dependency that fits the developer's needs?

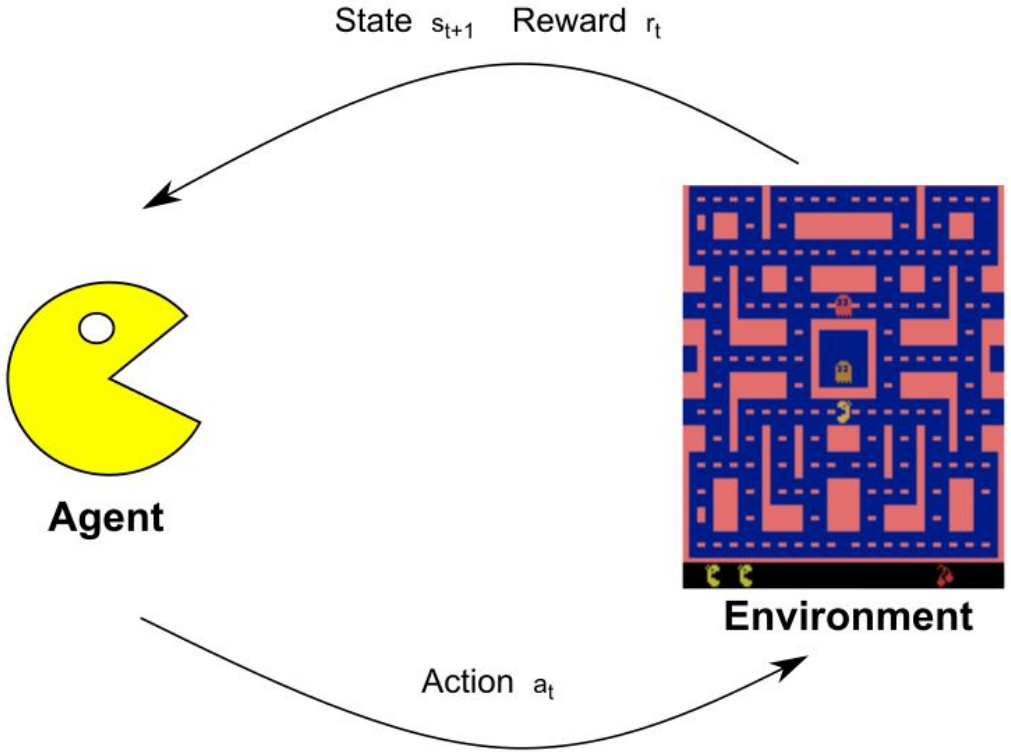
~~How to choose the dependency that fits the developer's needs?~~

Let the machine decide how to satisfy developer's needs!

ML in Thoth

- Recommender system is called **Adviser** in Thoth.
- It uses Reinforcement Learning (RL).

RL in general



ML in Thoth - RL

- **Agent:** resolver
- **Environment:** package dependencies
- **State**: resolver state
- **Actions**: set of actions given by the predictor
- **Policy**: maximize the cumulative reward (long-term)

ML in Thoth - Resolver

- Respect the ecosystem
- Works offline because all dependencies are collected in the database already
- Resolve software stacks using new type of observations;
- Resolution is considered as Markov Decision Process (MDP);
- Resolver uses pipelines Steps and Sieves to compute Rewards and Filter out packages;
- It is tightly coupled with Predictor;

ML in Thoth - Predictor

- Guide resolver in the expansion of state, until a finite state is reached :
    - Expand states that are the most promising ones to be used by users - used for recommending software stacks in adviser
    - Expand states for which Thoth has no observation about - used for filling Thoth's knowledge base using Dependency Monkey and Amun
- It’s trained using observations gathered
- Temporal Difference (TD), Monte Carlo Tree Search

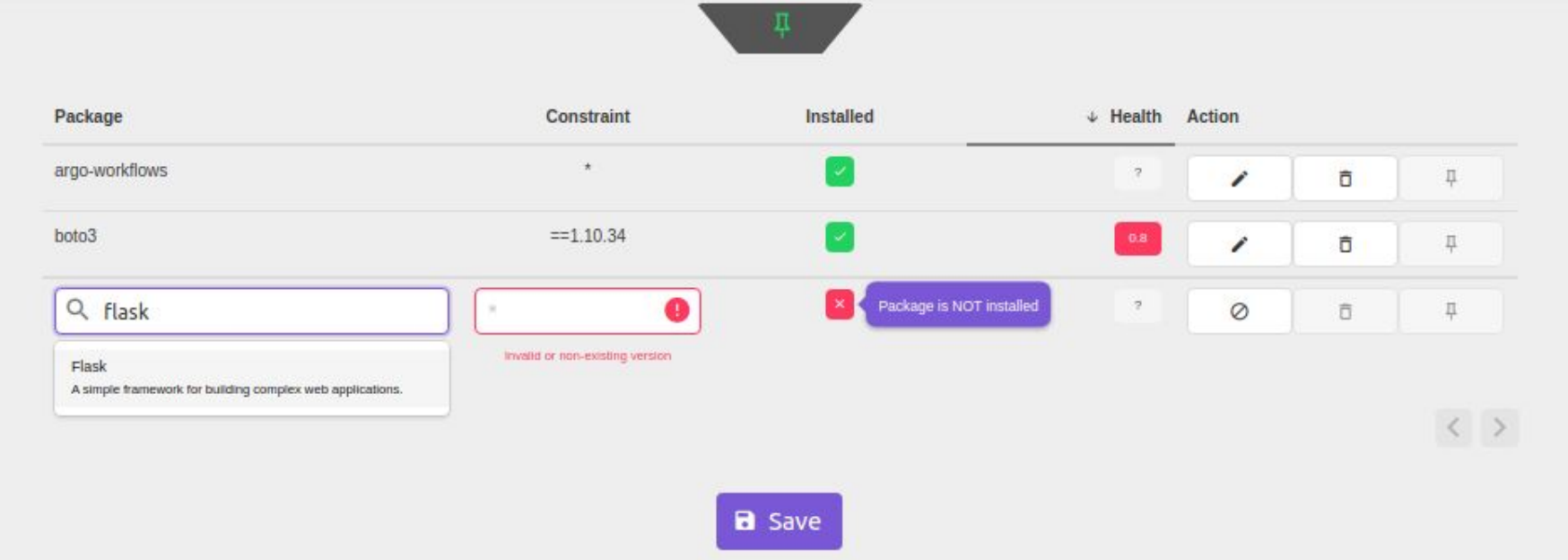
How we help Devs

- Keep dependencies up to date.
- Maintain software stack secure avoiding CVE.
- Recommend the most performant software stacks for AI Apps.
- Integrate source metadata information related to the packages used in the software stack to give advice to users.
- Integrate Thoth in day-to-day developers/data scientists tools.

cmd, OCP, GitHub, Notebook integration

- Command line tool thamos (developer laptop)
- Cyborg Kebechet (pull request creator)
- Source-to-Image (container builder)
- Jupyter Notebooks (data scientist browser)
- GitHub App Qeb-Hwt (CI pipeline check)

Example: nb-requirements



Call to action, participate

- `Pipenv install thamos ; thamos advise`
- Create PR fixing #hacktoberfest
- Let us know what kind of observation you made/want

references

- Website:
  - https://thoth-station.ninja/

- Twitter
  - https://twitter.com/thothstation

- GitHub
  - https://github.com/thoth-station

Thank you

Red Hat is the world's leading provider of enterprise open source software solutions. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500.


linkedin.com/company/red-hat


youtube.com/user/RedHatVideos


facebook.com/redhatinc


twitter.com/RedHat

Red Hat