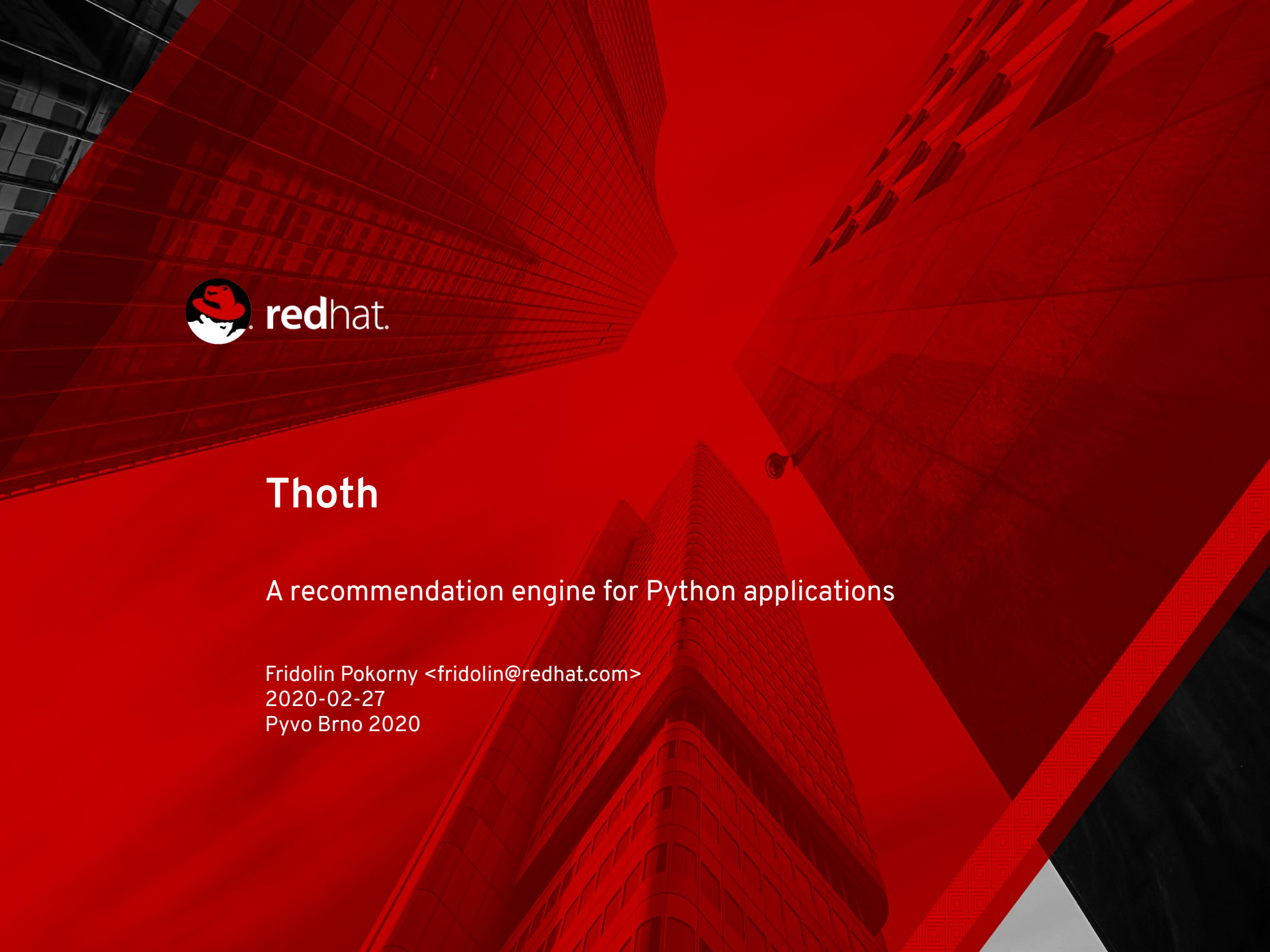


Thoth

A recommendation engine for Python applications

Fridolin Pokorny <fridolin@redhat.com>
2020-02-27
Pyvo Brno 2020

$ whoami

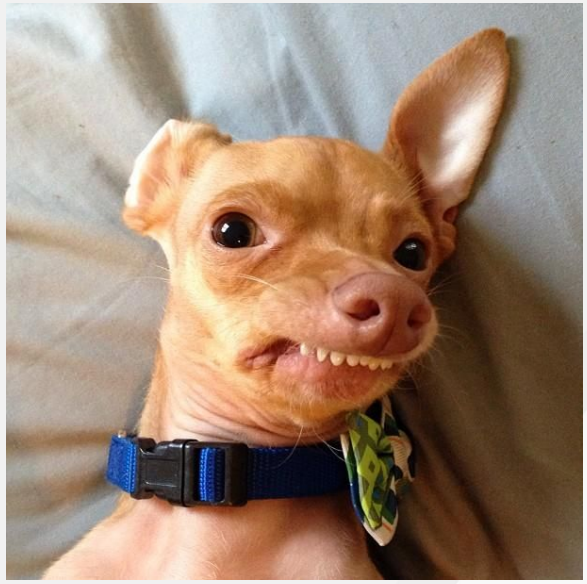
https://fridex.github.io

- Fridolín “fridex” Pokorný
- Senior Software Engineer at Red Hat
- Distributed systems, AI/ML and (of course) Python fan
- Projects:
  - Reverse engineer RetDec (AVG)
  - Linux kernel TLS/DTLS module `AF_KTLS`
  - Selinon - distributed task flows scheduler on top of Celery
  - Project Thoth

What is Thoth?

Why Thoth?

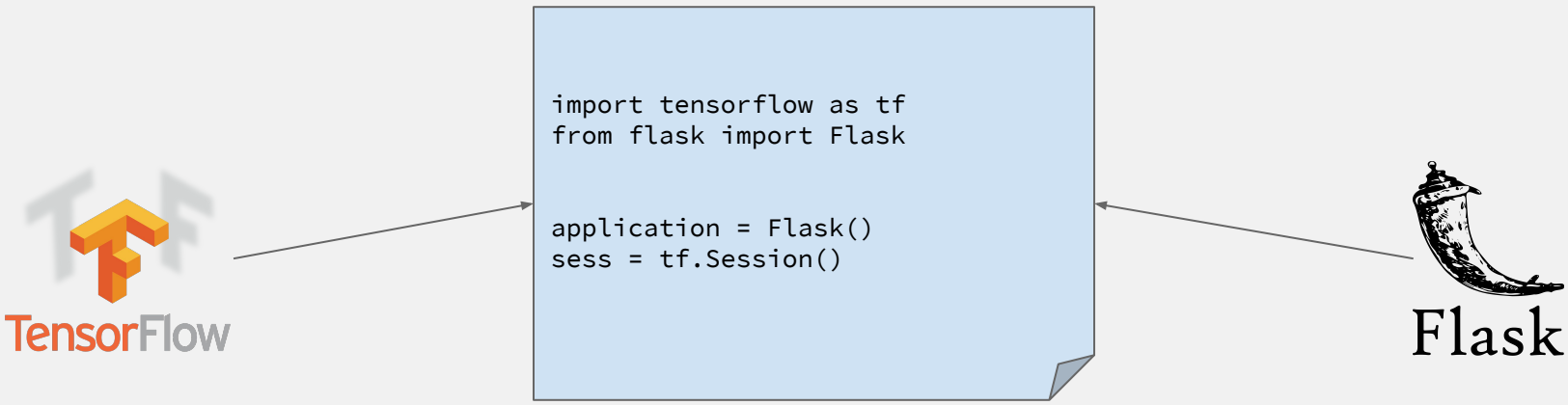
Thoth [tʰɔːtʰ]!



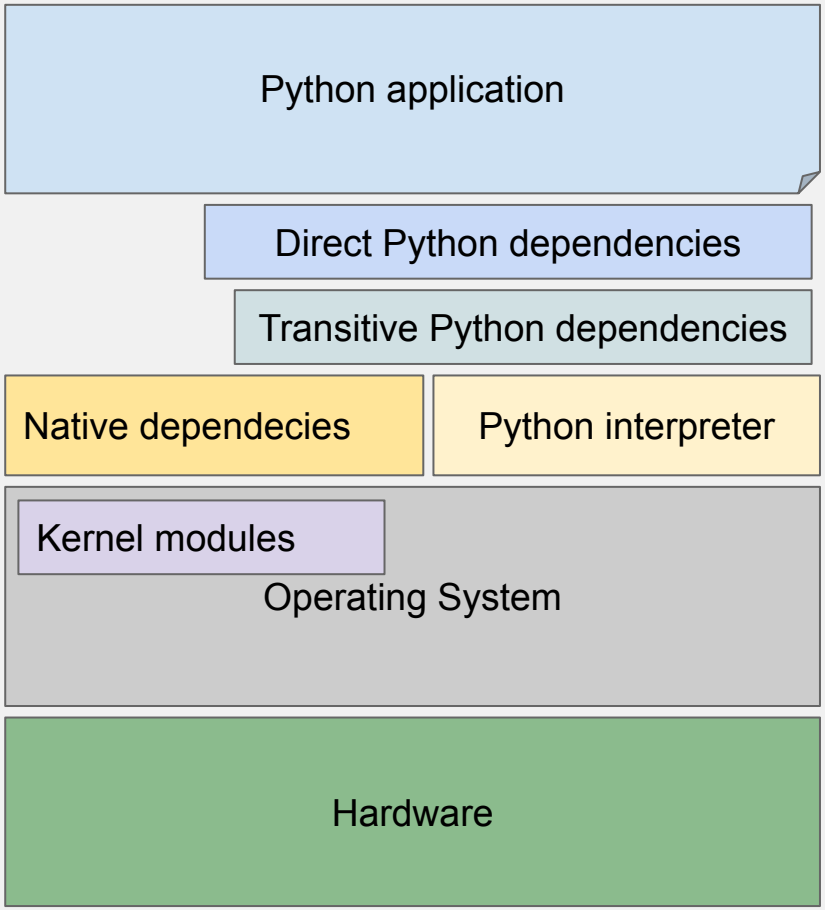
Why Thoth?

- PyPI - Python Package Index
  - https://pypi.org/
  - 215,218 projects
  - 1,645,362 releases (approx. 7 releases per project)

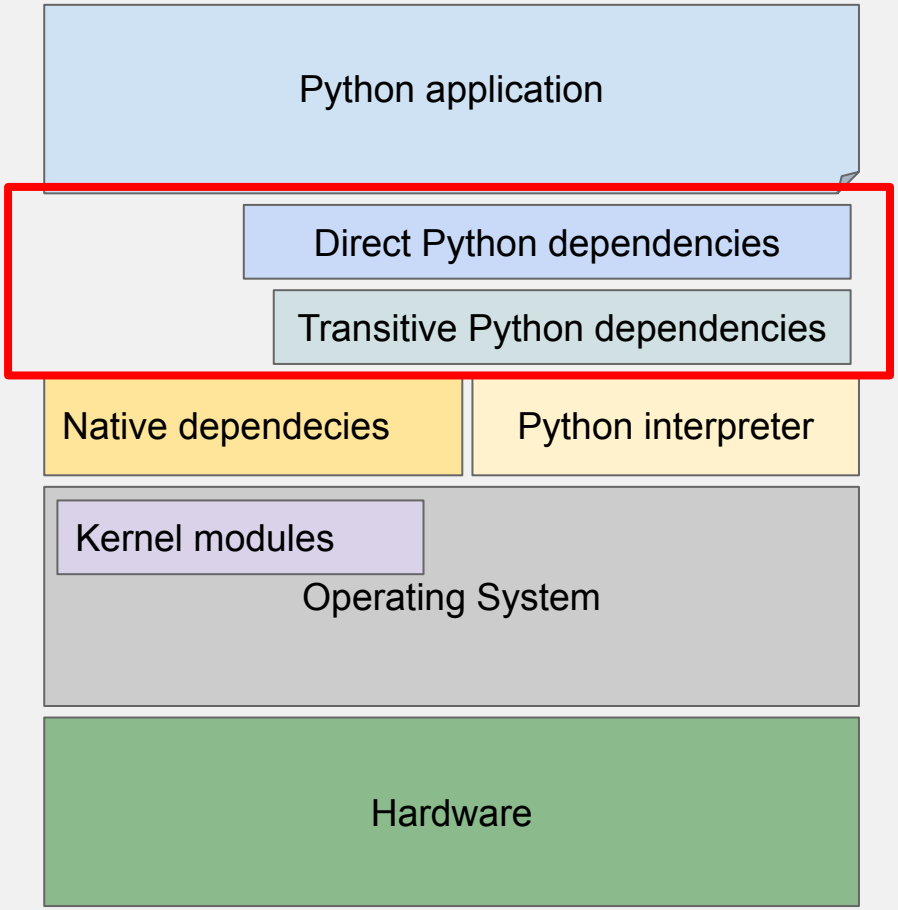
Why Thoth?



Why Thoth?



Why Thoth?



Transitive dependencies

- Flask (33)
  - click, itsdangerous, jinja2, markupsafe, werkzeug

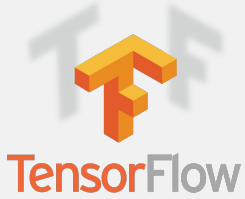
Estimatimated number of combinations: 54,395,000

Transitive dependencies

- TensorFlow (85)
  - absl-py, astor, backports-weakref, bleach, enum34, gast, google-pasta, grpcio, h5py, html5lib, keras, keras-applications, keras-preprocessing, markdown, mock, numpy, pbr, protobuf, pyyaml, scipy, setuptools, six, tensorboard, tensorflow-estimator, tensorflow-tensorboard, termcolor, tf-estimator-nightly, werkzeug, wheel

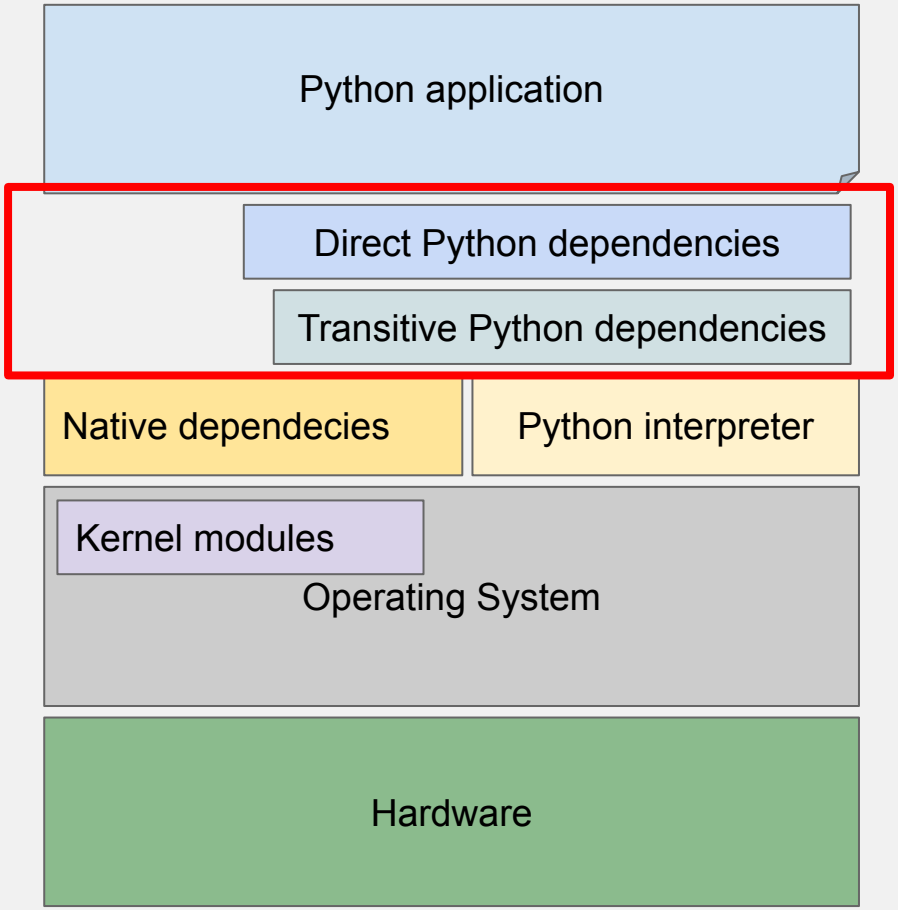
Estimated number of combinations: 139,740,802,927,165,440,000

approx. $1.39*10^{20}$

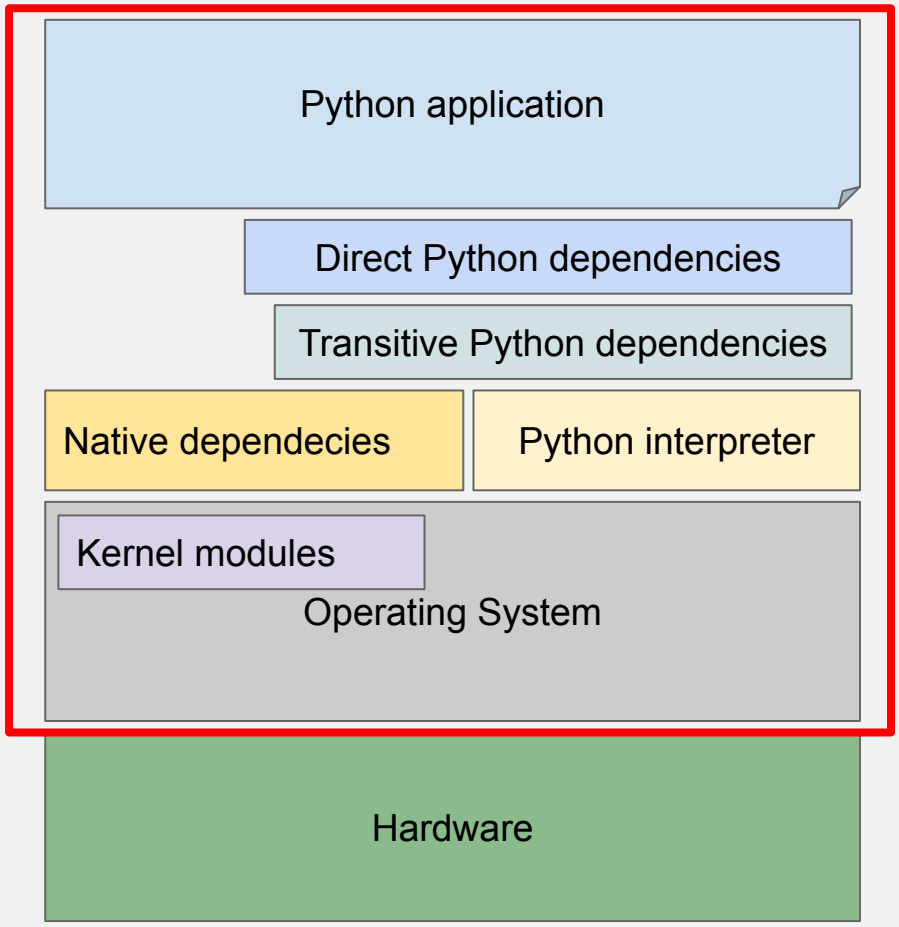


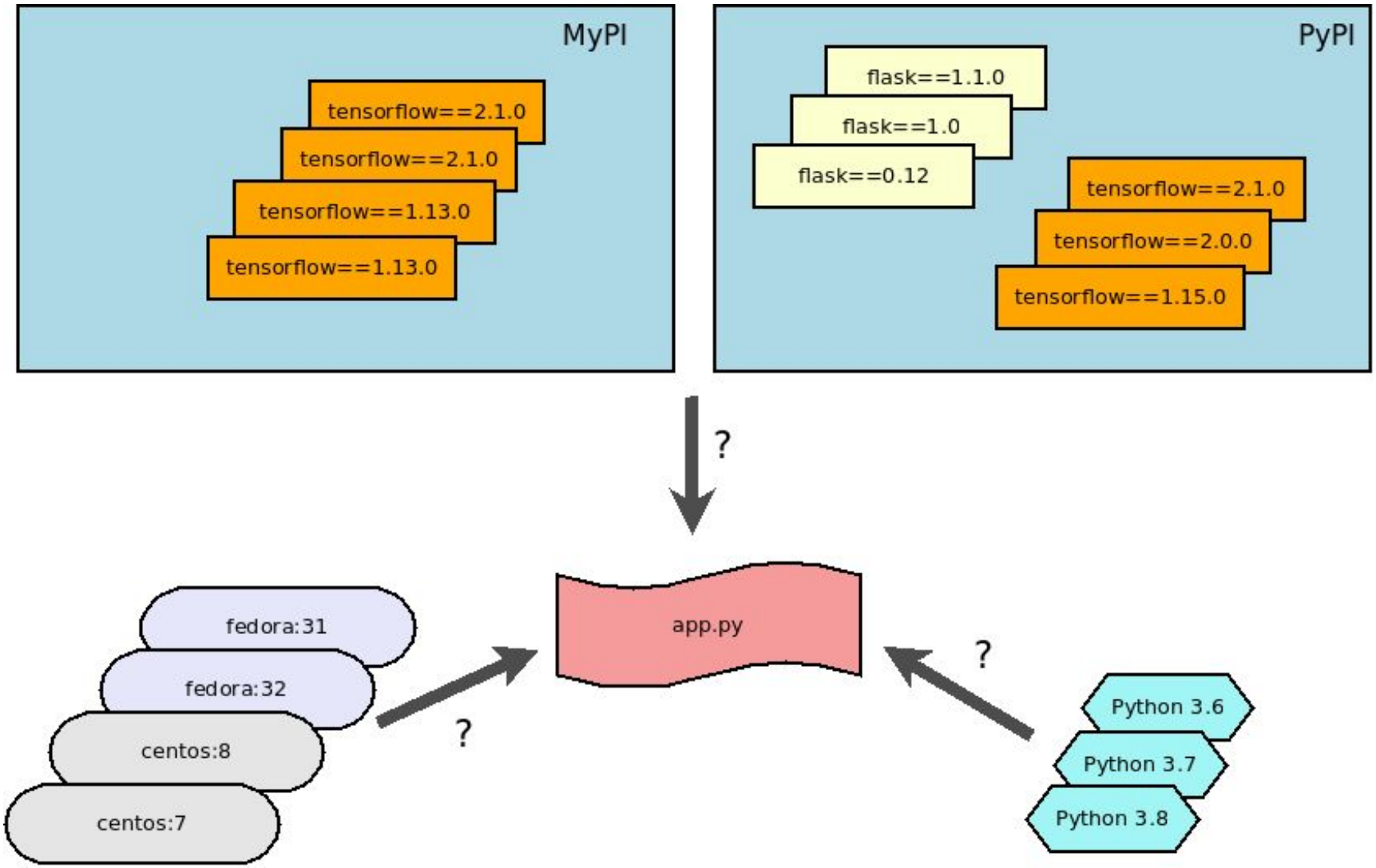
- Go and mathematics - https://en.wikipedia.org/wiki/Go_and_mathematics
  - number of possible game possitions is around:
    - 10^{172}
- Flask, TensorFlow and mathematics
  - Number of possible Python software stacks is around
    - 54,395,000 x 1.39 x 10^{20} = 7.6 x 10^{20} (rough estimate)

Why Thoth?

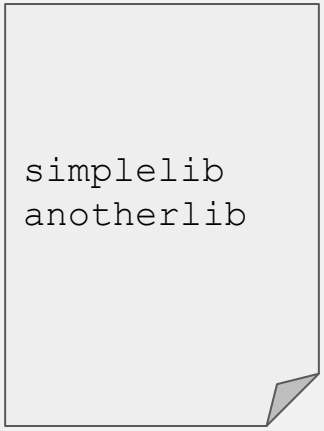


Why Thoth?

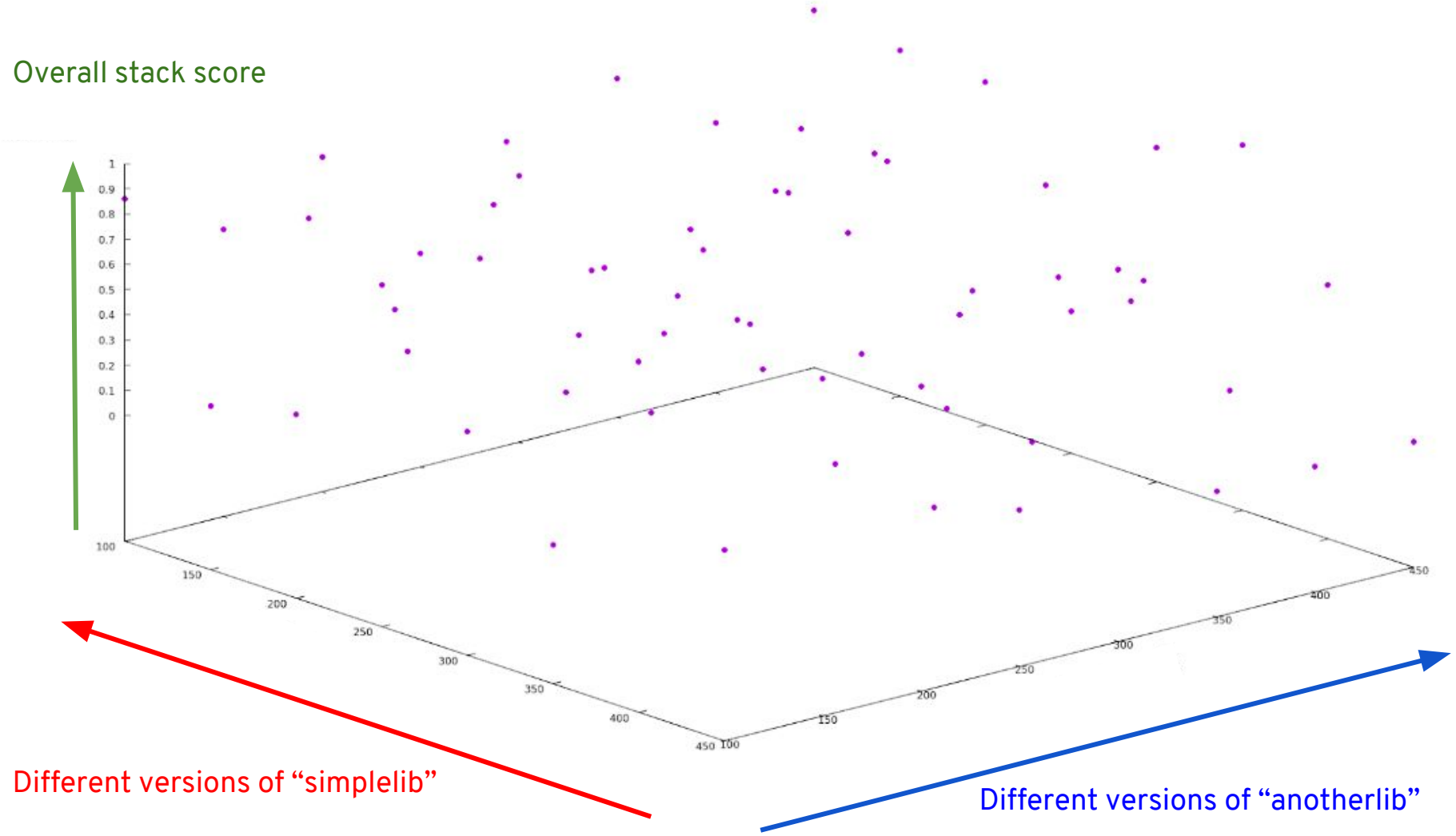



MyPI
tensorflow==2.1.0
tensorflow==2.1.0
tensorflow==1.13.0
tensorflow==1.13.0
PyPI
flask==1.1.0
flask==1.0
flask==0.12
tensorflow==2.1.0
tensorflow==2.0.0
tensorflow==1.15.0
?
app.py
fedora:31
fedora:32
centos:8
centos:7
?
?
Python 3.6
Python 3.7
Python 3.8

How good is my software stack?

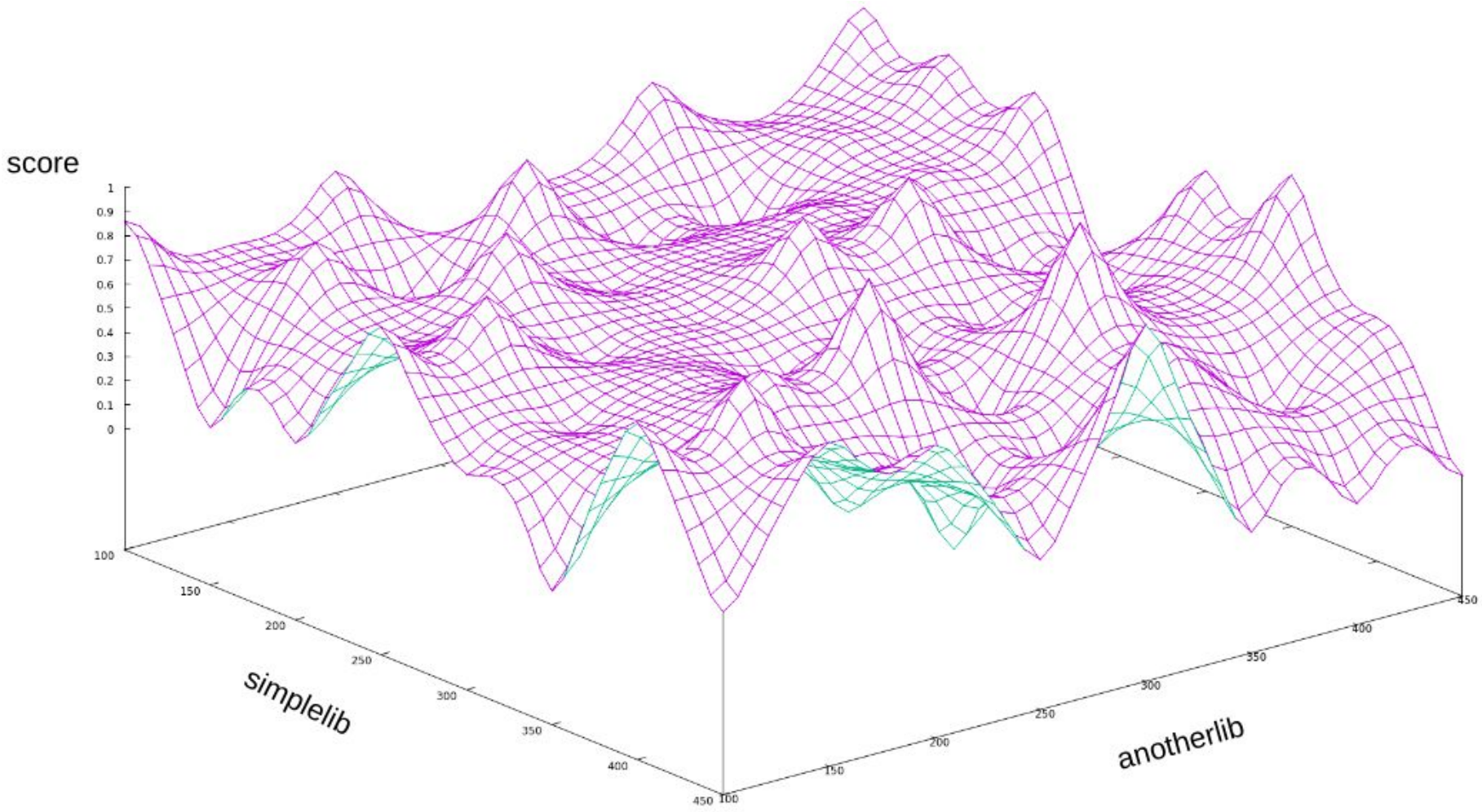


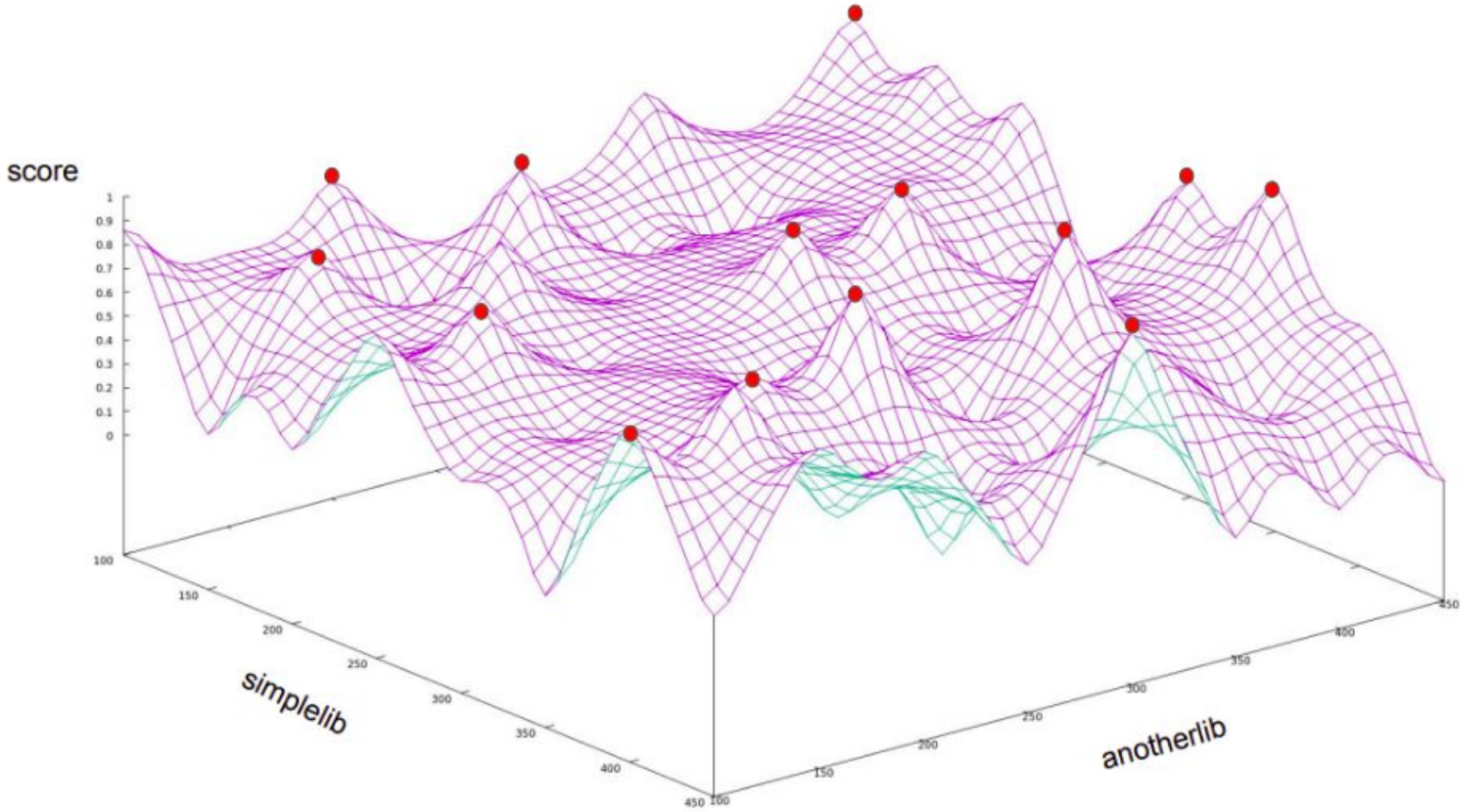
Overall stack score



Different versions of “simplelib”

Different versions of “anotherlib”

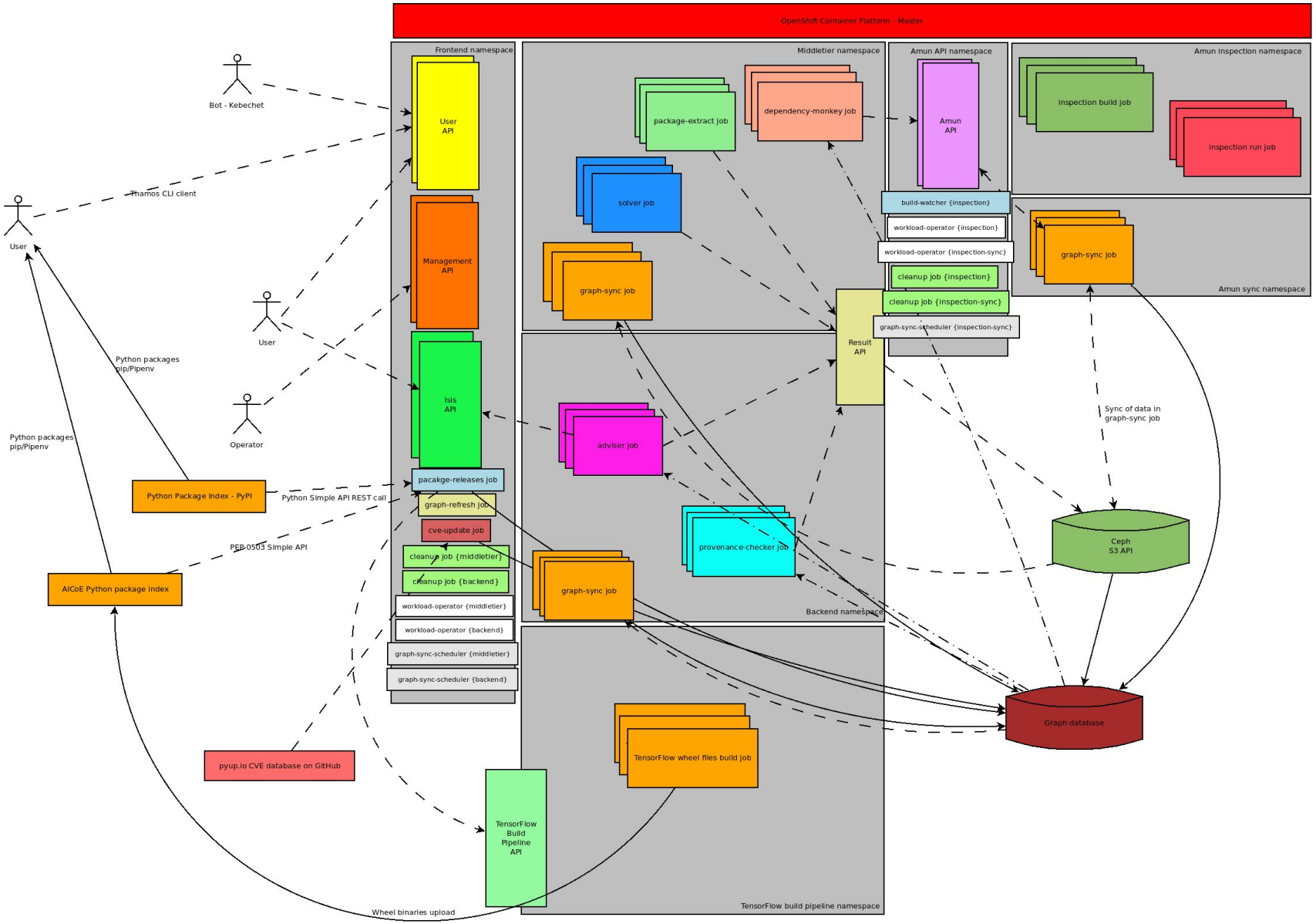

score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100
150
200
250
300
350
400
450
simplelib
100
150
200
250
300
350
400
450
anotherlib


score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100
150
200
250
300
350
400
450
simplelib
100
150
200
250
300
350
400
450
anotherlib

Why Thoth?

- Create knowledge base
  - What packages in which versions should I use?
    - Application builds correctly
    - Application runs correctly
    - Application behaves and performs well
- Create an advanced Python resolver which uses knowledge base to resolve software stacks

Latest versions are not always greatest choices.

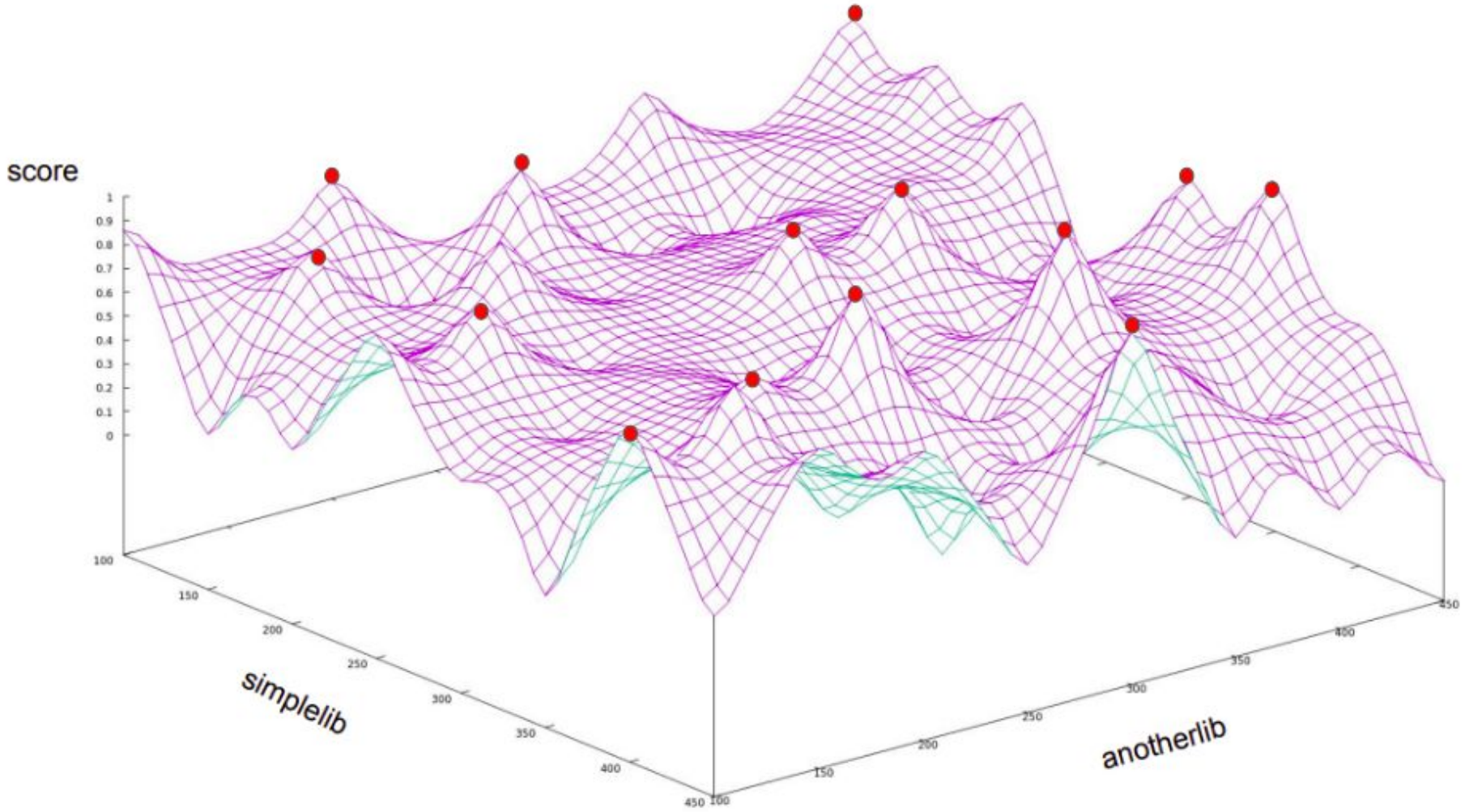

OpenShift Container Platform - Master
Frontend namespace
Middletier namespace
Amun API namespace
Amun inspection namespace
Amun sync namespace
Backend namespace
TensorFlow build pipeline namespace
Bot - Kebechet
User
Thamos CLI client
User
Operator
Python packages pip/Pipenv
Python packages pip/Pipenv
User API
Management API
Isis API
pacakge-releases job
graph-refresh job
cve-update job
cleanup job {middletier}
cleanup job {backend}
workload-operator {middletier}
workload-operator {backend}
graph-sync-scheduler {middletier}
graph-sync-scheduler {backend}
package-extract job
dependency-monkey job
solver job
graph-sync job
adviser job
provenance-checker job
graph-sync job
Result API
Amun API
build-watcher {inspection}
workload-operator {inspection}
workload-operator {inspection-sync}
cleanup job {inspection}
cleanup job {inspection-sync}
graph-sync-scheduler {inspection-sync}
inspection build job
inspection run job
graph-sync job
Sync of data in graph-sync job
Ceph S3 API
Graph database
TensorFlow wheel files build job
TensorFlow Build Pipeline API
Python Package Index - PyPI
Python Simple API REST call
PEP 0503 Simple API
AICoE Python package index
pyup.io CVE database on GitHub
Wheel binaries upload

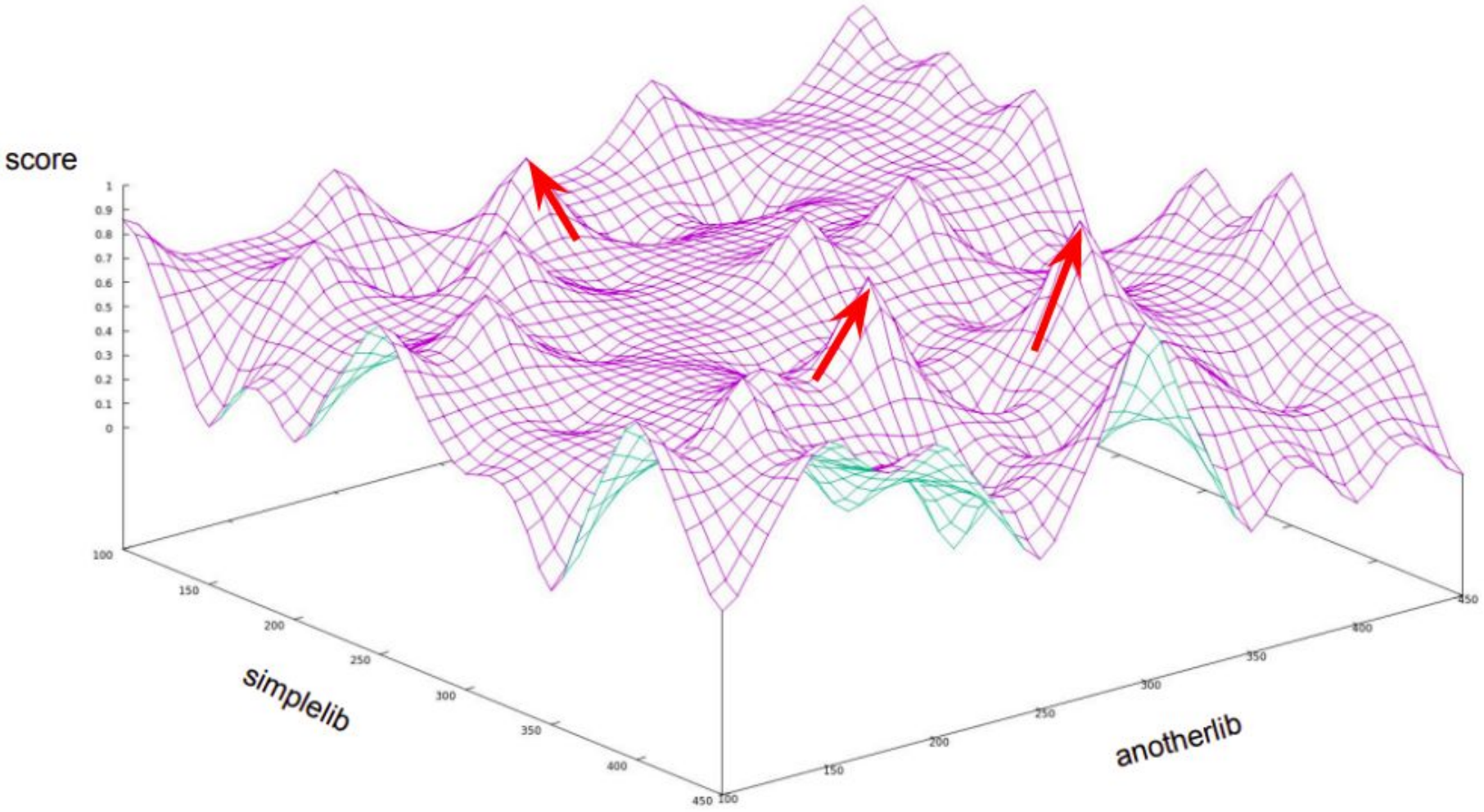
Thoth's adviser

- Server side resolution
- Multiple iterations on implementation
- Pure Python implementation
  - Memory consuption
  - N-ary graph with transactional operations
- Rewritten into C/C++
  - Too many queries to database
  - Cca. 2.5k queries just to obtain TensorFlow dependency graph
  - The main database changed 2 times

Thoth's adviser

- Later stochastic approaches - Operations Research
  - Hill climbing
  - Adaptive Simulated Annealing
- Implementation split into two parts
  - Resolver
    - Resolve software stacks respecting Python ecosystem
  - Predictor
    - Guide resolver in resolution


score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100
150
200
250
300
350
400
450
simplelib
100
150
200
250
300
350
400
450
anotherlib


score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100
150
200
250
300
350
400
450
simplelib
100
150
200
250
300
350
400
450
anotherlib

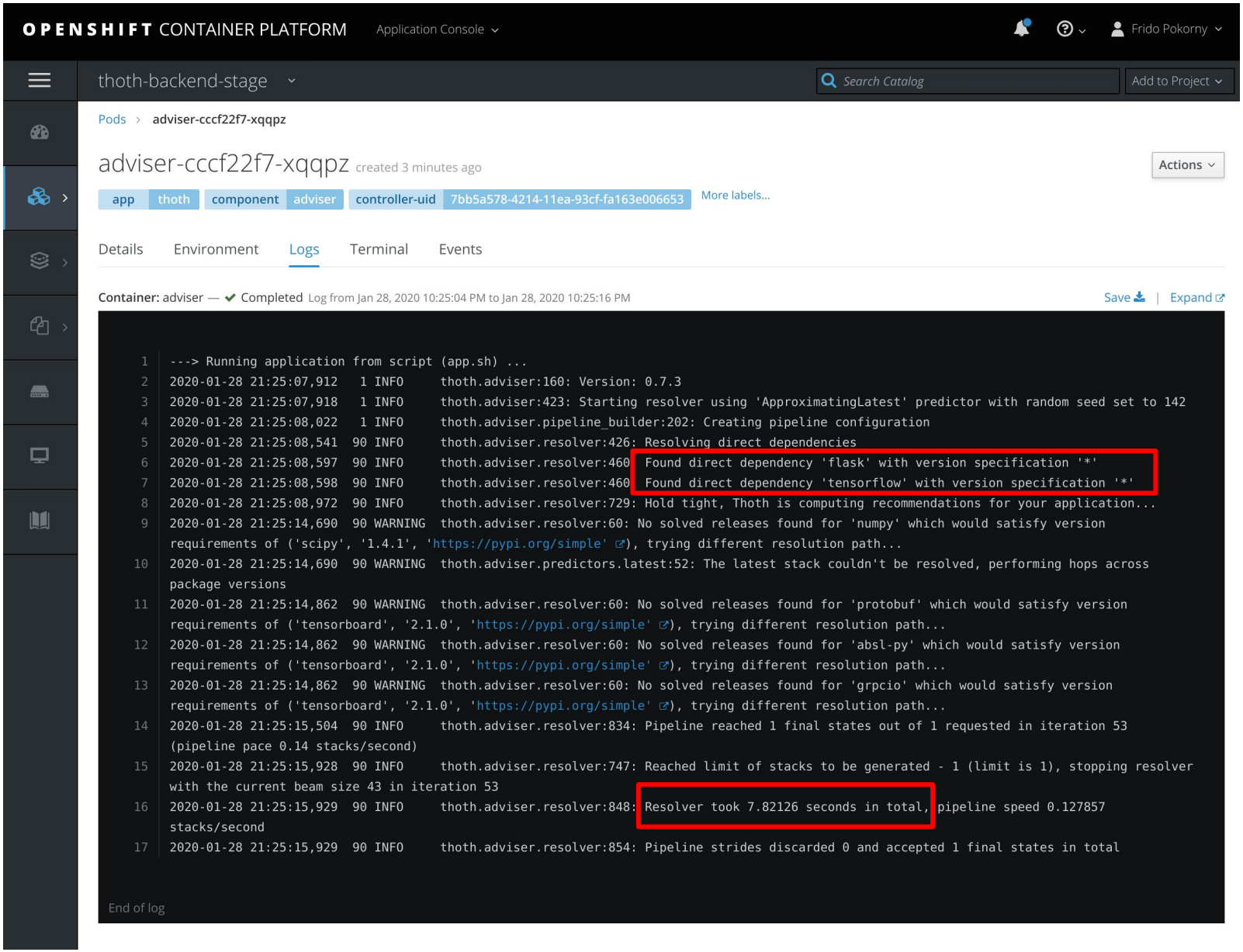
Thoth's adviser

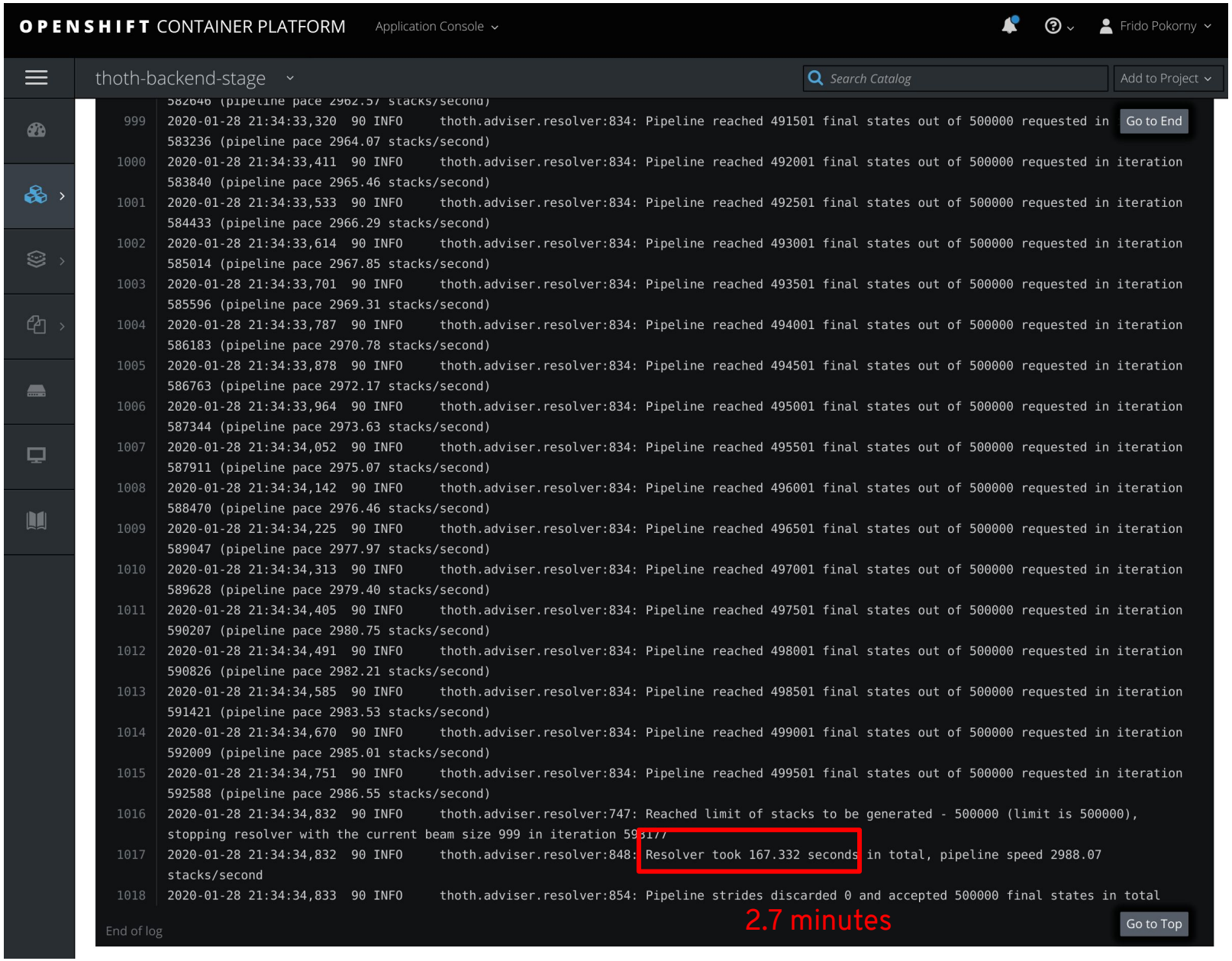
- Reinforcement Learning - Gradient-based methods
  - Not responsive enough
    - Neural Combinatorial Optimization with RL https://arxiv.org/abs/1611.09940

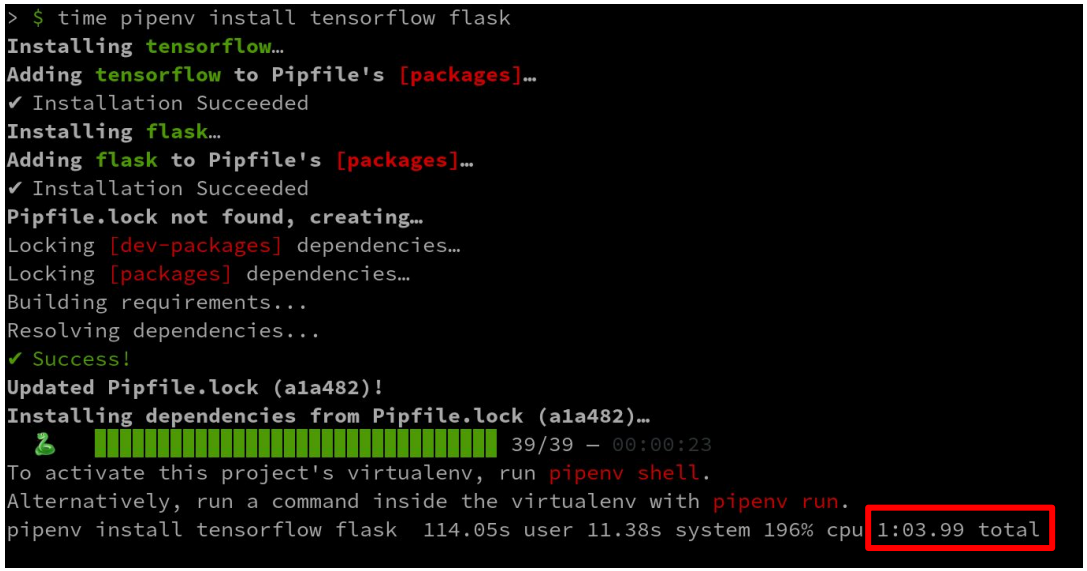
- Reinforcement Learning - Gradient-free methods
  - Temporal Difference, Monte Carlo Tree Search

- Reconfigurable pipeline made out of units
  - Units define scoring function (units of type step)
  - Units define action space (units of type sieve)

- Dependency Monkey
  - Sample state space to gather "observations"

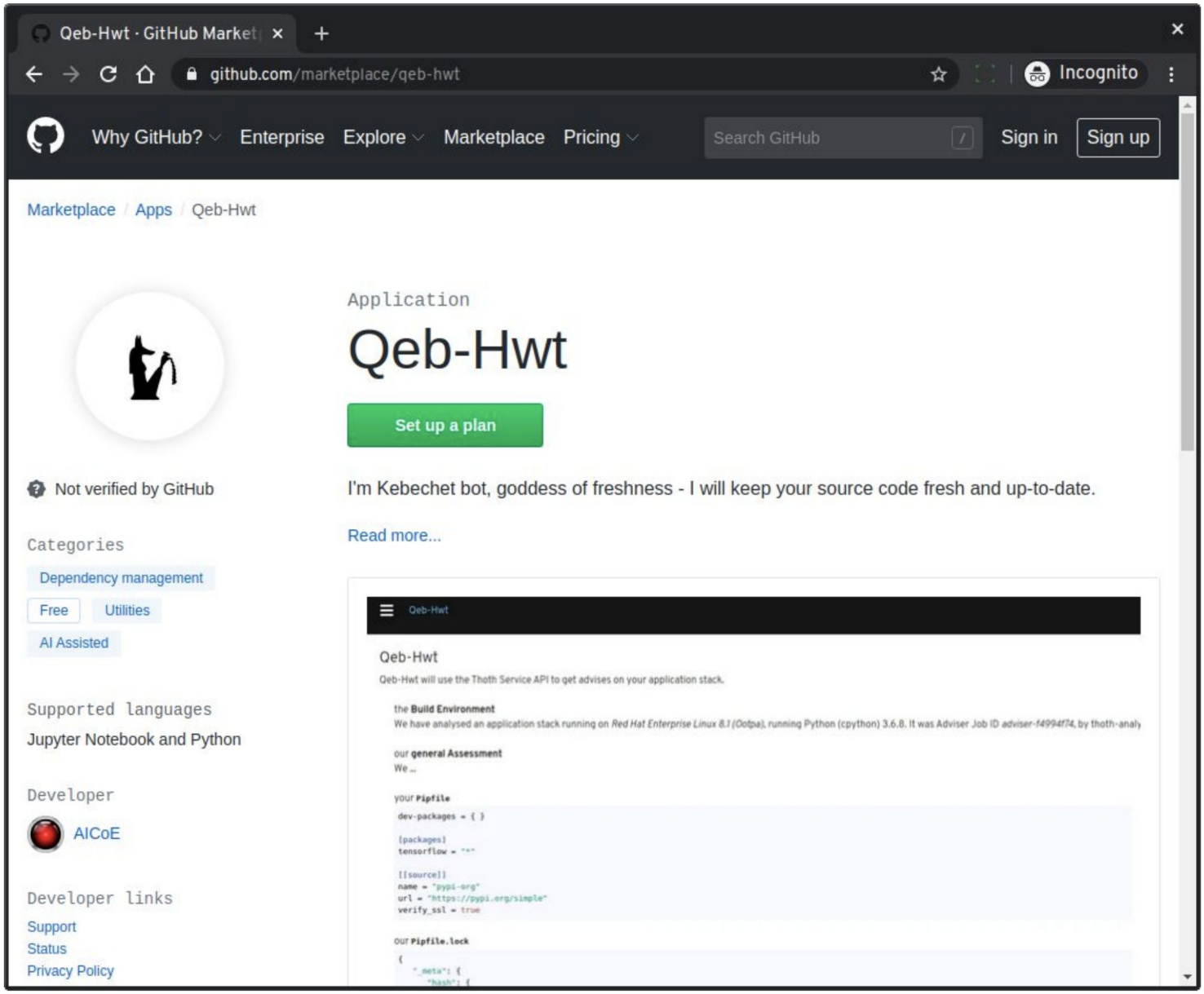

OPENSHIFT CONTAINER PLATFORM
Application Console
Frido Pokorny
thoth-backend-stage
Search Catalog
Add to Project
Pods
adviser-cccf22f7-xqqpz
adviser-cccf22f7-xqqpz created 3 minutes ago
Actions
app thoth component adviser controller-uid 7bb5a578-4214-11ea-93cf-fa163e006653
More labels...
Details
Environment
Logs
Terminal
Events
Container: adviser — Completed Log from Jan 28, 2020 10:25:04 PM to Jan 28, 2020 10:25:16 PM
Save
Expand
1 ---> Running application from script (app.sh) ...
2 2020-01-28 21:25:07,912 1 INFO thoth.adviser:160: Version: 0.7.3
3 2020-01-28 21:25:07,918 1 INFO thoth.adviser:423: Starting resolver using 'ApproximatingLatest' predictor with random seed set to 142
4 2020-01-28 21:25:08,022 1 INFO thoth.adviser.pipeline_builder:202: Creating pipeline configuration
5 2020-01-28 21:25:08,541 90 INFO thoth.adviser.resolver:426: Resolving direct dependencies
6 2020-01-28 21:25:08,597 90 INFO thoth.adviser.resolver:460: Found direct dependency 'flask' with version specification '*'
7 2020-01-28 21:25:08,598 90 INFO thoth.adviser.resolver:460: Found direct dependency 'tensorflow' with version specification '*'
8 2020-01-28 21:25:08,972 90 INFO thoth.adviser.resolver:729: Hold tight, Thoth is computing recommendations for your application...
9 2020-01-28 21:25:14,690 90 WARNING thoth.adviser.resolver:60: No solved releases found for 'numpy' which would satisfy version requirements of ('scipy', '1.4.1', 'https://pypi.org/simple'), trying different resolution path...
10 2020-01-28 21:25:14,690 90 WARNING thoth.adviser.predictors.latest:52: The latest stack couldn't be resolved, performing hops across package versions
11 2020-01-28 21:25:14,862 90 WARNING thoth.adviser.resolver:60: No solved releases found for 'protobuf' which would satisfy version requirements of ('tensorboard', '2.1.0', 'https://pypi.org/simple'), trying different resolution path...
12 2020-01-28 21:25:14,862 90 WARNING thoth.adviser.resolver:60: No solved releases found for 'absl-py' which would satisfy version requirements of ('tensorboard', '2.1.0', 'https://pypi.org/simple'), trying different resolution path...
13 2020-01-28 21:25:14,862 90 WARNING thoth.adviser.resolver:60: No solved releases found for 'grpcio' which would satisfy version requirements of ('tensorboard', '2.1.0', 'https://pypi.org/simple'), trying different resolution path...
14 2020-01-28 21:25:15,504 90 INFO thoth.adviser.resolver:834: Pipeline reached 1 final states out of 1 requested in iteration 53 (pipeline pace 0.14 stacks/second)
15 2020-01-28 21:25:15,928 90 INFO thoth.adviser.resolver:747: Reached limit of stacks to be generated - 1 (limit is 1), stopping resolver with the current beam size 43 in iteration 53
16 2020-01-28 21:25:15,929 90 INFO thoth.adviser.resolver:848: Resolver took 7.82126 seconds in total, pipeline speed 0.127857 stacks/second
17 2020-01-28 21:25:15,929 90 INFO thoth.adviser.resolver:854: Pipeline strides discarded 0 and accepted 1 final states in total
End of log


OPENSHIFT CONTAINER PLATFORM
Application Console
Frido Pokorny
thoth-backend-stage
Search Catalog
Add to Project
582646 (pipeline pace 2962.57 stacks/second)
999 2020-01-28 21:34:33,320 90 INFO thoth.adviser.resolver:834: Pipeline reached 491501 final states out of 500000 requested in
Go to End
583236 (pipeline pace 2964.07 stacks/second)
1000 2020-01-28 21:34:33,411 90 INFO thoth.adviser.resolver:834: Pipeline reached 492001 final states out of 500000 requested in iteration
583840 (pipeline pace 2965.46 stacks/second)
1001 2020-01-28 21:34:33,533 90 INFO thoth.adviser.resolver:834: Pipeline reached 492501 final states out of 500000 requested in iteration
584433 (pipeline pace 2966.29 stacks/second)
1002 2020-01-28 21:34:33,614 90 INFO thoth.adviser.resolver:834: Pipeline reached 493001 final states out of 500000 requested in iteration
585014 (pipeline pace 2967.85 stacks/second)
1003 2020-01-28 21:34:33,701 90 INFO thoth.adviser.resolver:834: Pipeline reached 493501 final states out of 500000 requested in iteration
585596 (pipeline pace 2969.31 stacks/second)
1004 2020-01-28 21:34:33,787 90 INFO thoth.adviser.resolver:834: Pipeline reached 494001 final states out of 500000 requested in iteration
586183 (pipeline pace 2970.78 stacks/second)
1005 2020-01-28 21:34:33,878 90 INFO thoth.adviser.resolver:834: Pipeline reached 494501 final states out of 500000 requested in iteration
586763 (pipeline pace 2972.17 stacks/second)
1006 2020-01-28 21:34:33,964 90 INFO thoth.adviser.resolver:834: Pipeline reached 495001 final states out of 500000 requested in iteration
587344 (pipeline pace 2973.63 stacks/second)
1007 2020-01-28 21:34:34,052 90 INFO thoth.adviser.resolver:834: Pipeline reached 495501 final states out of 500000 requested in iteration
587911 (pipeline pace 2975.07 stacks/second)
1008 2020-01-28 21:34:34,142 90 INFO thoth.adviser.resolver:834: Pipeline reached 496001 final states out of 500000 requested in iteration
588470 (pipeline pace 2976.46 stacks/second)
1009 2020-01-28 21:34:34,225 90 INFO thoth.adviser.resolver:834: Pipeline reached 496501 final states out of 500000 requested in iteration
589047 (pipeline pace 2977.97 stacks/second)
1010 2020-01-28 21:34:34,313 90 INFO thoth.adviser.resolver:834: Pipeline reached 497001 final states out of 500000 requested in iteration
589628 (pipeline pace 2979.40 stacks/second)
1011 2020-01-28 21:34:34,405 90 INFO thoth.adviser.resolver:834: Pipeline reached 497501 final states out of 500000 requested in iteration
590207 (pipeline pace 2980.75 stacks/second)
1012 2020-01-28 21:34:34,491 90 INFO thoth.adviser.resolver:834: Pipeline reached 498001 final states out of 500000 requested in iteration
590826 (pipeline pace 2982.21 stacks/second)
1013 2020-01-28 21:34:34,585 90 INFO thoth.adviser.resolver:834: Pipeline reached 498501 final states out of 500000 requested in iteration
591421 (pipeline pace 2983.53 stacks/second)
1014 2020-01-28 21:34:34,670 90 INFO thoth.adviser.resolver:834: Pipeline reached 499001 final states out of 500000 requested in iteration
592009 (pipeline pace 2985.01 stacks/second)
1015 2020-01-28 21:34:34,751 90 INFO thoth.adviser.resolver:834: Pipeline reached 499501 final states out of 500000 requested in iteration
592588 (pipeline pace 2986.55 stacks/second)
1016 2020-01-28 21:34:34,832 90 INFO thoth.adviser.resolver:747: Reached limit of stacks to be generated - 500000 (limit is 500000),
stopping resolver with the current beam size 999 in iteration 593177
1017 2020-01-28 21:34:34,832 90 INFO thoth.adviser.resolver:848: Resolver took 167.332 seconds in total, pipeline speed 2988.07
stacks/second
1018 2020-01-28 21:34:34,833 90 INFO thoth.adviser.resolver:854: Pipeline strides discarded 0 and accepted 500000 final states in total
End of log
2.7 minutes
Go to Top


> $ time pipenv install tensorflow flask
Installing tensorflow…
Adding tensorflow to Pipfile's [packages]…
✔ Installation Succeeded
Installing flask…
Adding flask to Pipfile's [packages]…
✔ Installation Succeeded
Pipfile.lock not found, creating…
Locking [dev-packages] dependencies…
Locking [packages] dependencies…
Building requirements...
Resolving dependencies...
✔ Success!
Updated Pipfile.lock (a1a482)!
Installing dependencies from Pipfile.lock (a1a482)…
39/39 — 00:00:23
To activate this project's virtualenv, run pipenv shell.
Alternatively, run a command inside the virtualenv with pipenv run.
pipenv install tensorflow flask 114.05s user 11.38s system 196% cpu 1:03.99 total

Thoth parts...

- Bots automating routing tasks
  - Updates of dependencies
  - New releases
- Optimized TensorFlow releases
  - https://tensorflow.pypi.thoth-station.ninja/
- Topics modeling on Python package metadata
- Dependency Monkey + Adviser
- Static source code analysis
- Container image analysis
- Integration with OpenShift, Jupyter Notebooks, CLI
- ...


Qeb-Hwt · GitHub Market
github.com/marketplace/qeb-hwt
Incognito
Why GitHub?
Enterprise
Explore
Marketplace
Pricing
Search GitHub
Sign in
Sign up
Marketplace / Apps / Qeb-Hwt
Application
Qeb-Hwt
Set up a plan
Not verified by GitHub
I'm Kebechet bot, goddess of freshness - I will keep your source code fresh and up-to-date.
Read more...
Categories
Dependency management
Free
Utilities
AI Assisted
Supported languages
Jupyter Notebook and Python
Developer
AICoE
Developer links
Support
Status
Privacy Policy
Qeb-Hwt
Qeb-Hwt will use the Thoth Service API to get advises on your application stack.
the Build Environment
our general Assessment
We ...
your Pipfile
dev-packages = { }
[packages]
tensorflow = "*"
[[source]]
name = "pypi-org"
url = "https://pypi.org/simple"
verify_ssl = true
our Pipfile.lock

Information about Thoth

- Thoth Bot
  - https://bit.ly/a-thoth-bot/
  - Feedback form: https://bit.ly/thoth-feedback/
- Website:
  - https://thoth-station.ninja/
- Twitter
  - https://twitter.com/thothstation
- GitHub
  - https://github.com/thoth-station

redhat.

THANK YOU

plus.google.com/+RedHat

facebook.com/redhatinc

linkedin.com/company/red-hat

twitter.com/RedHat

youtube.com/user/RedHatVideos