

Active Learning in the VO-Cloud Science Platform



P. Škoda^{1,2}, T.Mazel¹, O. Podsztavek¹

¹Faculty of Information Technology, Czech Technical University in Prague

²Astronomical Institute of the Czech Academy of Sciences, Ondřejov



Astronomical Institute
of the Czech Academy of Sciences



**RESEARCH
CENTER FOR
INFORMATICS**

rci.cvut.cz



EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education



MINISTRY OF EDUCATION,
YOUTH AND SPORTS

CZECH TECHNICAL UNIVERSITY IN PRAGUE
FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF SOFTWARE ENGINEERING



Bachelor's thesis

**Design and implementation of a
distributed platform for data mining of big
astronomical spectra archives**

Jakub Koza

Supervisor: RNDr. Petr Škoda, CSc.

12th May 2015

CZECH TECHNICAL UNIVERSITY IN PRAGUE
FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF SOFTWARE ENGINEERING



Master's thesis

**Interactive Cloud-Based Platform for
Parallelized Machine Learning of
Astronomical Big Data**

Bc. Jakub Koza

Supervisor: RNDr. Petr Škoda, CSc.

9th May 2017

Sources of Spectra

Getting spectra + store

(restricted access – big files)

Files

UPLOAD from given local directory (recursive)

DOWNLOAD by http + index, FTP (recursive)

VOTable

UPLOAD VOTable (e.g. prepared in TOPCAT - meta)

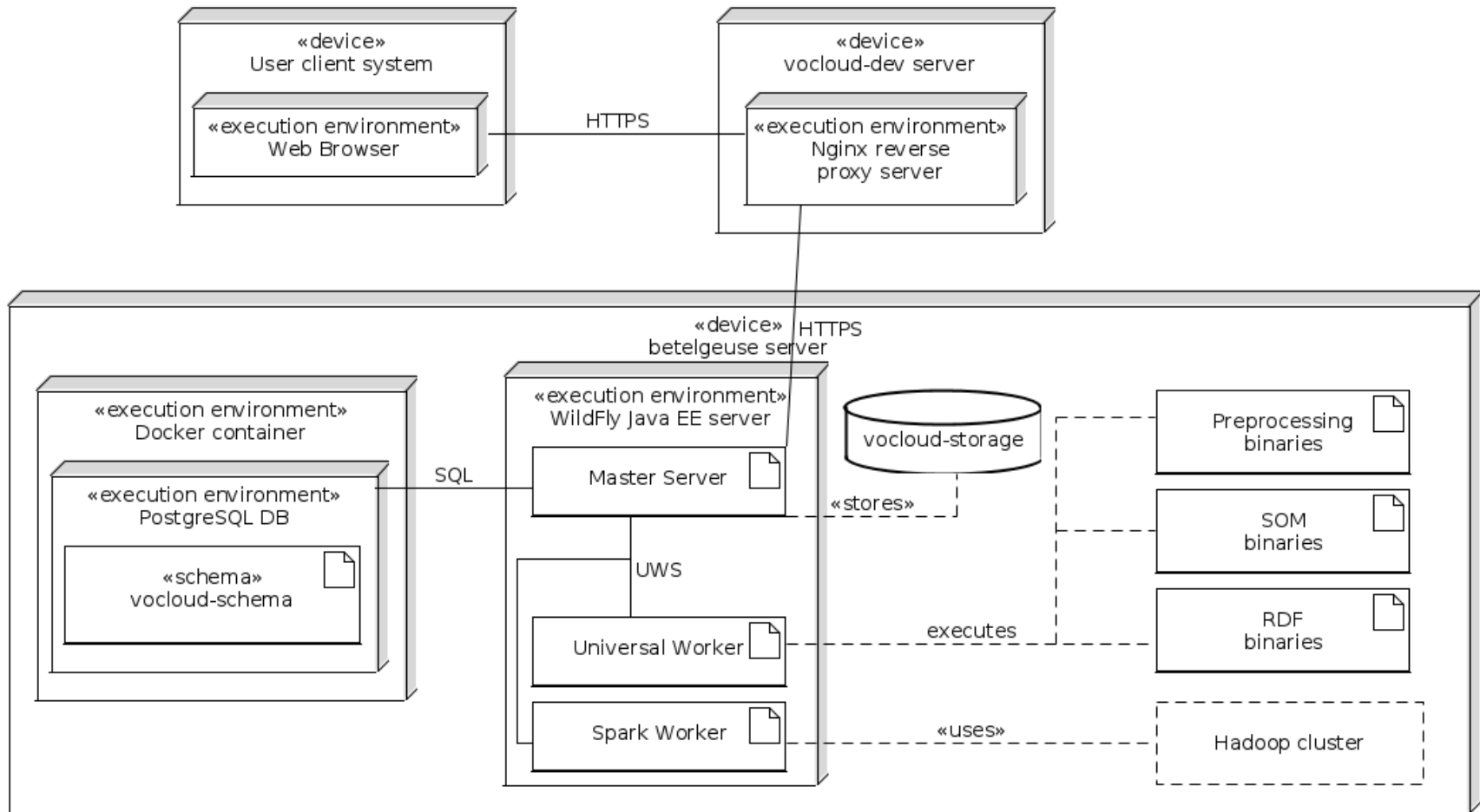
REMOTE VOTable

SSAP query + Accref

+ DataLink + SODA

~~SAMP control - send to SPLAT???~~

VO-CLOUD - UWS driven



VO-CLOUD spectra visualisation

VO-CLOUD MANAGE FILESYSTEM

Home Manage filesystem Jobs Download history Create job Jupyter Settings Admin Help Logout (admin)

New Folder Append new files Delete items Download data

DATA > allond700

| Name | Size | Created | Operation |
|---|---------|-------------------------|------------------------------|
| <input type="checkbox"/> vb040037.fits | | | Download Rename View content |
| <input type="checkbox"/> ue210040.fits | | | Download Rename View content |
| <input checked="" type="checkbox"/> sh180024.fits | | | Download Rename View content |
| <input type="checkbox"/> rd260041.fits | | | Download Rename View content |
| <input type="checkbox"/> vd040029.fits | | | Download Rename View content |
| <input type="checkbox"/> a201503070034.fits | | | Download Rename View content |
| <input type="checkbox"/> a201503240017.fits | | | Download Rename View content |
| <input type="checkbox"/> rg080029.fits | | | Download Rename View content |
| <input checked="" type="checkbox"/> th010022.fits | | | Download Rename View content |
| <input type="checkbox"/> a201503040022.fits | | | Download Rename View content |
| <input type="checkbox"/> a201502150025.fits | | | Download Rename View content |
| <input type="checkbox"/> a201503080034.fits | | | Download Rename View content |
| <input checked="" type="checkbox"/> ti060011.fits | | | Download Rename View content |
| <input type="checkbox"/> va270017.fits | | | Download Rename View content |
| <input type="checkbox"/> a201502200048.fits | 43.2 kB | Apr 26, 2017 1:05:41 PM | Download Rename View content |
| <input checked="" type="checkbox"/> sh150027.fits | 43.2 kB | Apr 26, 2017 1:05:40 PM | Download Rename View content |
| <input type="checkbox"/> a201504060004.fits | 43.2 kB | Apr 26, 2017 1:05:41 PM | Download Rename View content |
| <input type="checkbox"/> ue250024.fits | 43.2 kB | Apr 26, 2017 1:05:40 PM | Download Rename View content |

Spectra plotter

Figure 140706134745720

sh180024.fits: Altair
th010022.fits: HD190603
sh150027.fits: HD190603
ti060011.fits: HD164353

png zoom rect

43.2 kB

Apr 26, 2017 1:05:41 PM

43.2 kB

Apr 26, 2017 1:05:40 PM

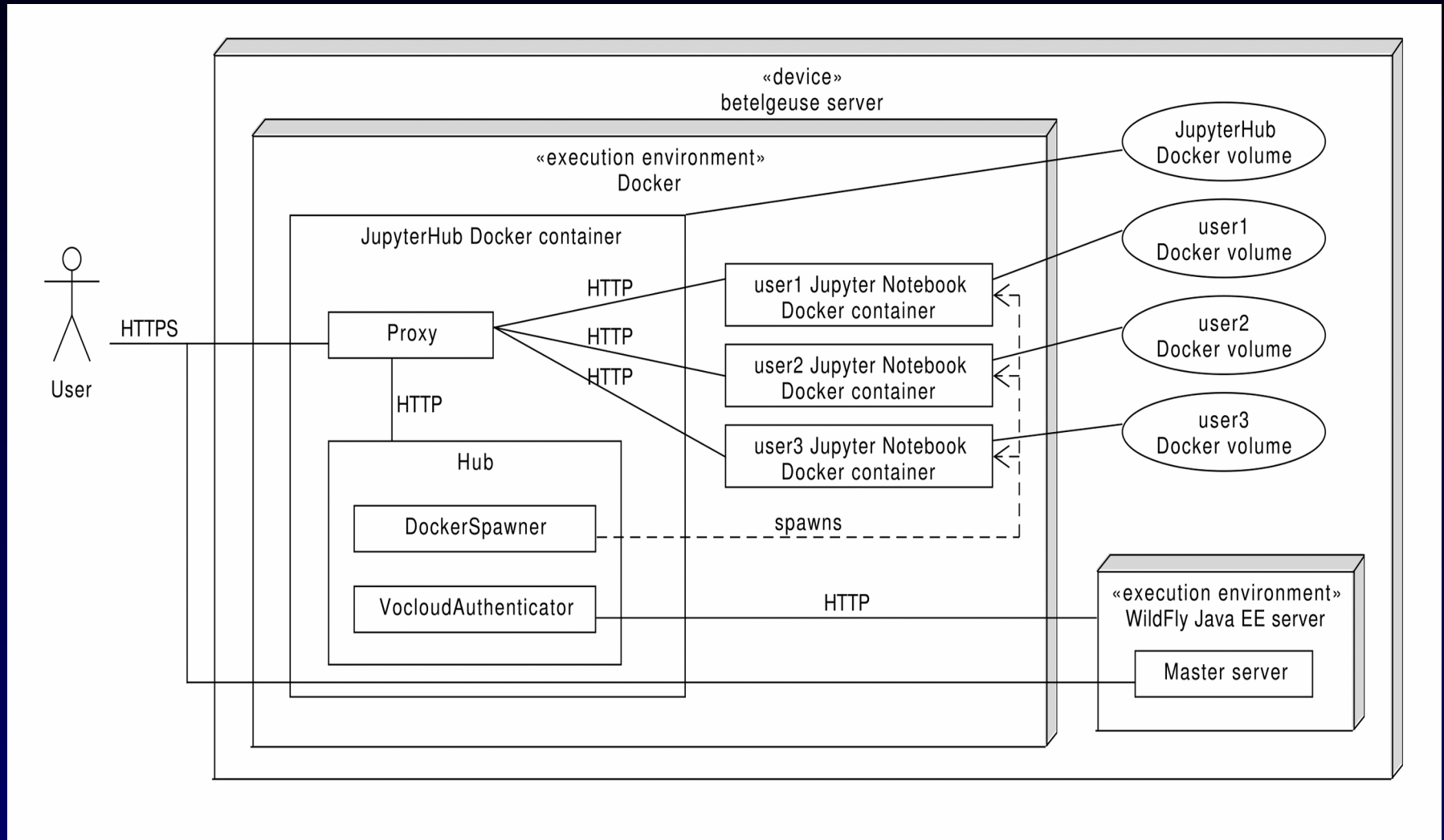
43.2 kB

Apr 26, 2017 1:05:41 PM

43.2 kB

Apr 26, 2017 1:05:40 PM

JupyterHub deployment



JupyterHub example

jupyter example_plotter Last Checkpoint: 05/06/2017 (unsaved changes) Control Panel Logout Python 3

File Edit View Insert Cell Kernel Widgets Help

Code CellToolbar

```
In [7]: path='filesystem/DATA/allond700/'
spectra=['sh180024.fits','th010022.fits','ti060011.fits','sh150027.fits']
files=[path + i for i in spectra]
files
```

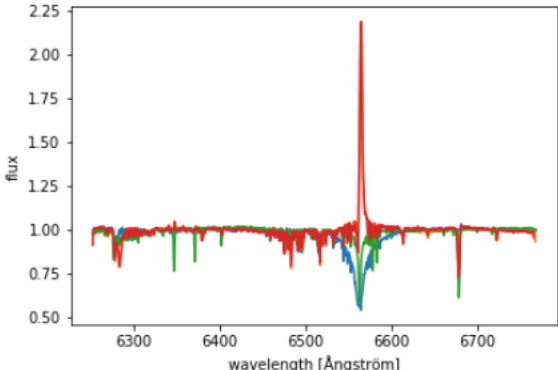
```
Out[7]: ['filesystem/DATA/allond700/sh180024.fits',
'filesystem/DATA/allond700/th010022.fits',
'filesystem/DATA/allond700/ti060011.fits',
'filesystem/DATA/allond700/sh150027.fits']
```

```
In [13]: parsed = [parse_spectrum_file(i) for i in files]
parsed[0]
```

```
Out[13]: {'flux': array([ 0.97623893,  0.97816423,  0.98200884, ...,  0.99071508,
 0.99049042,  0.98766227]),
'name': 'Altair',
'wave': array([ 6252.48405443,  6252.74072204,  6252.99738965, ...,  6764.27926559,
 6764.53593319,  6764.7926008 ])}
```

```
In [12]: for i in parsed:
plt.plot(i['wave'], i['flux'])
plt.xlabel('wavelength [Ångström]')
plt.ylabel('flux')
```

```
Out[12]: <matplotlib.text.Text at 0x7f7f108b8550>
```



Active Learning



FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Bachelor's thesis

Cloud-Based Platform for Active Learning of Astronomical Spectra

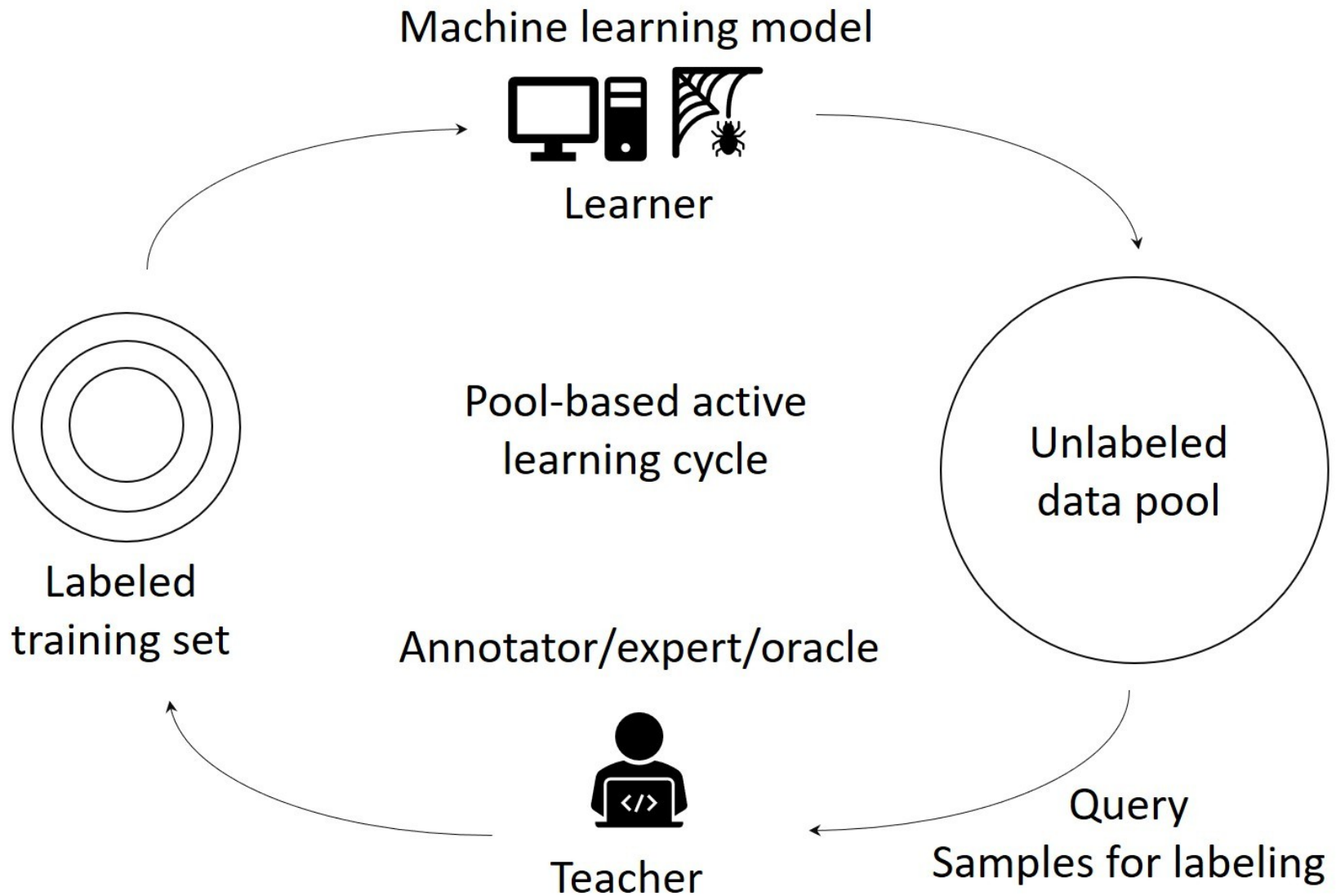
MUDr. Tomáš Mazel, Ph.D.

Department of Software Engineering
Supervisor: RNDr. Petr Škoda, CSc.

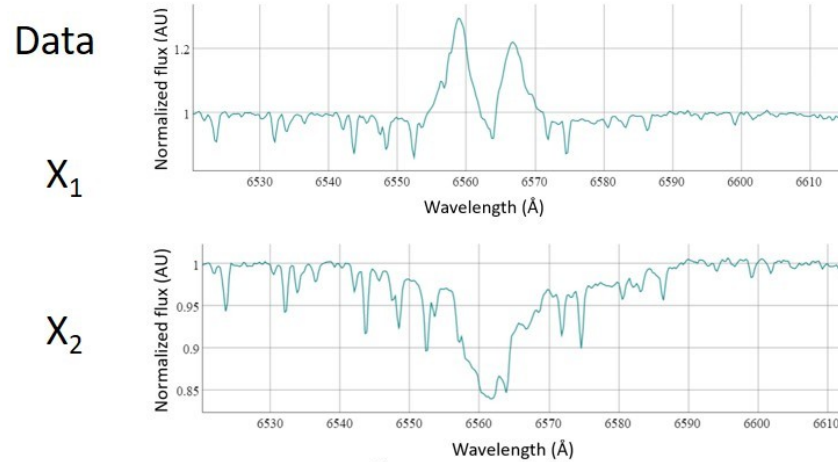
February 12, 2020

- ElasticSearch
- CSV (HDF5)
- Python scripts/nb
- CUDA/TF/KERAS

Active Learning



Active Learning



Neural network



Performance exceeded
a predefined level

oracle data

Help me, I am not sure: $(X_1, ?)$

I think: $(X_1, \text{double peak})$ – add to training set

Am I correct?: $(X_1, \text{double peak})$

Yes, you are excellent!: $(X_1, \text{double peak})$

performance-estimate data

Oracle

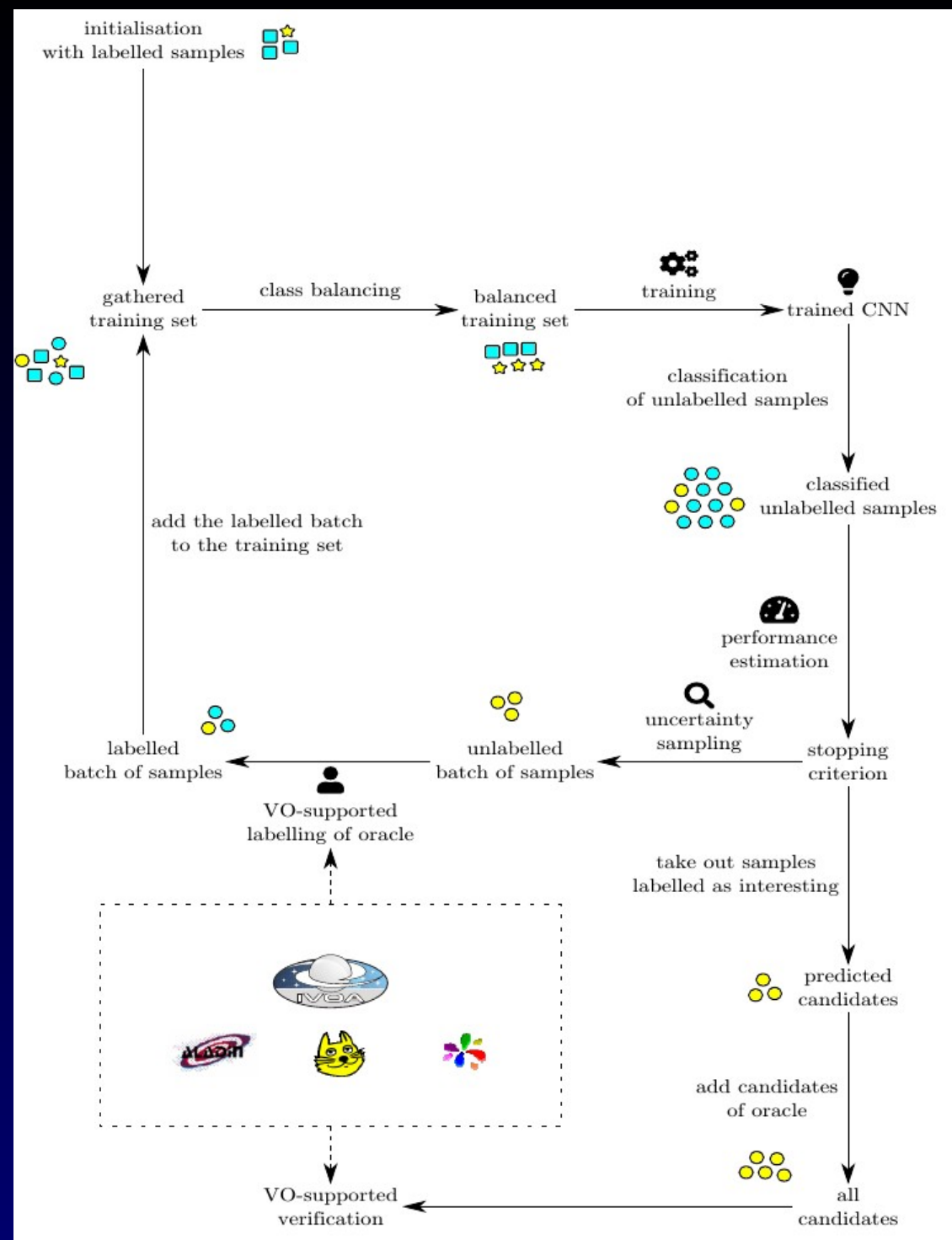


$(X_1, \text{double peak}; X_2, \text{other})$

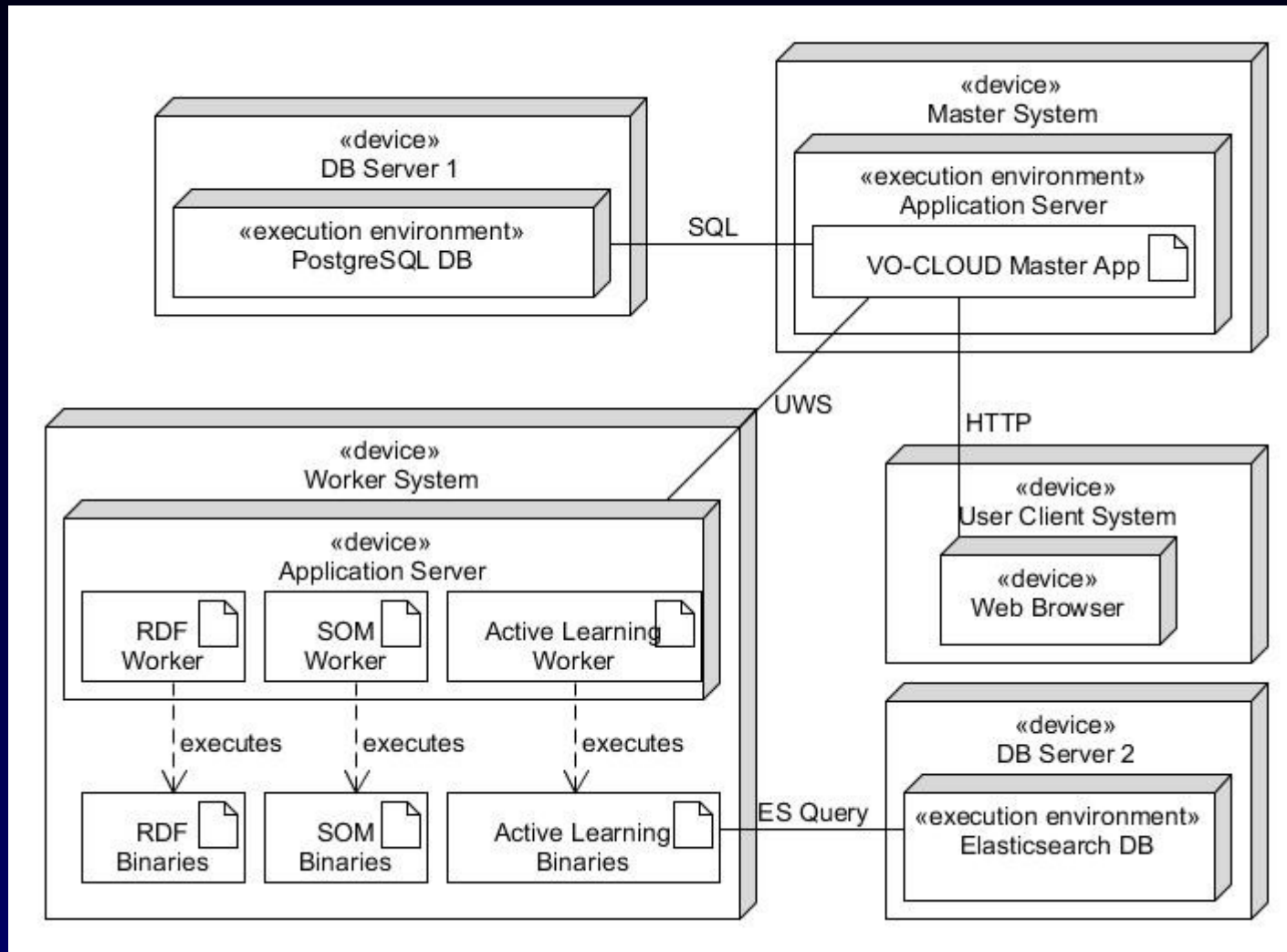
list of interesting objects (candidates)

Active Deep Learning: CNN Chooses Data for Its Training

- Oracle classification (domain expert knowledge)
- Uncertainty sampling (entropy)
- From predicted target classes selected batch of 100
- Batch added to training set
- True positive rate estimation



VO-CLOUD Integration



Future KDD Science platforms

Include human in loop

Expert needs VO to check the nature of suggested candidates

to decide - RE-LABEL OR CONFIRM

Get information OUT of DOMAIN

(e.g. part of spectrum, image not preprocessed)

Thank You

LINKS

<https://zenodo.org/record/573727>

<https://zenodo.org/record/44641>

Source on GitHub

<https://github.com/vodev> (10 repositories)