OpenNMT – Industry Bridge

Jean Senellart, SYSTRAN Global CTO

OpenNMT Workshop Paris, March 2018

SYSTRAN Overview

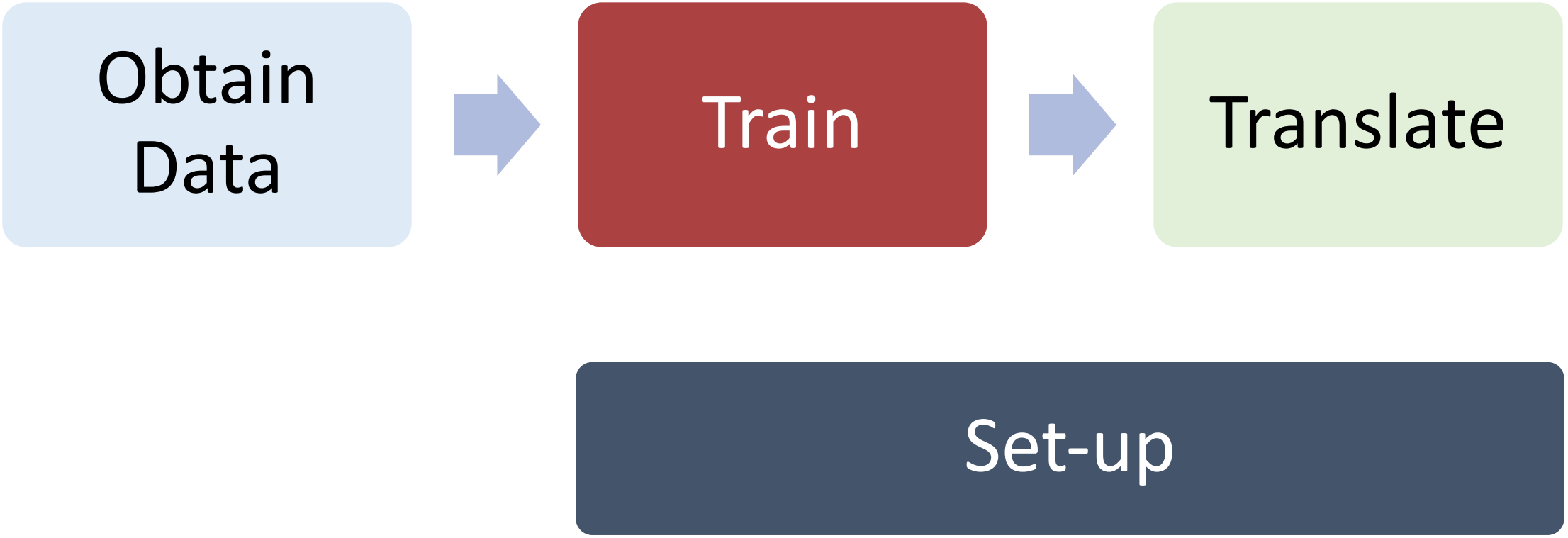
- Pure Machine Translation player founded in 1968
- From the beginning, forerunner in the different MT waves
  - 1968-2005 – rule-based machine translation
  - 2006-2016 – Statistical Machine Translation
  - 2016- Neural Machine Translation
- About 150 employees in 3 sites San Diego, **Paris**, and Seoul
- Key products
  - Enterprise servers for secured on-premise/private cloud installations
  - 150 language pairs, User and Productivity Tools, Customized and/or large volume Translation
  - Target Market: B2B, institutions

SYSTRAN NMT Odyssee

- Started Proof of Concept early 2016
  - Quickly outperformed baseline engines (internal, online)
- Launched Commercial NMT solution in October 2016
  - … one week after Google NMT
- With HarvardNLP, Open-Sourced OpenNMT in December 2016
  - Dedicated team and passionate support of community
- 110 language pairs trained for 8 different domains
  - Hundreds of specific customer Customization
- NMT shifted full company focus

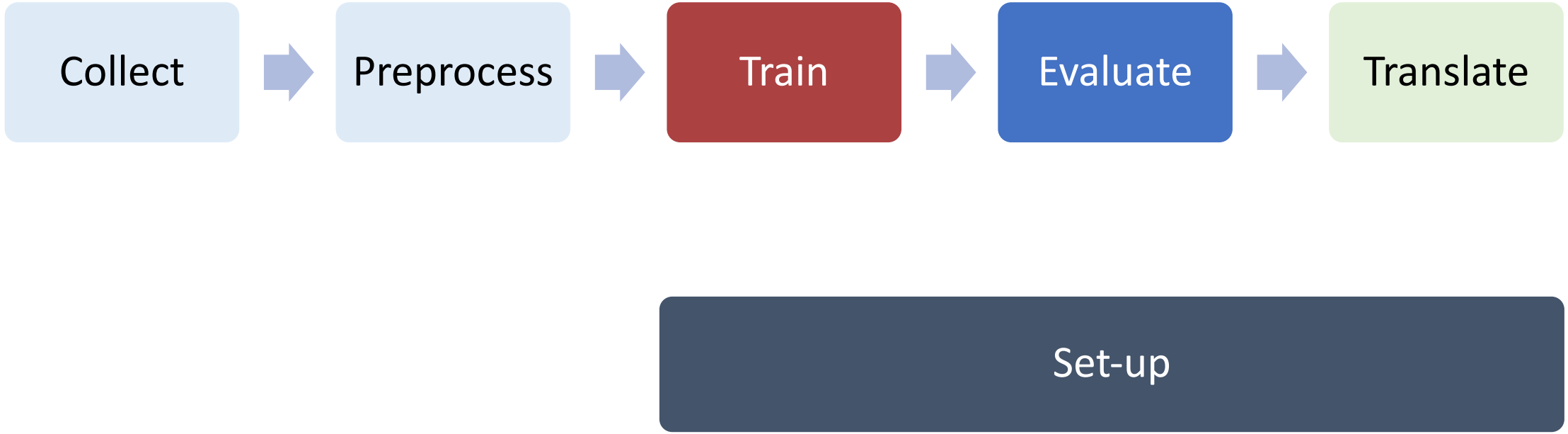
Training a NMT model

- Building a first NMT model is simple and fun!



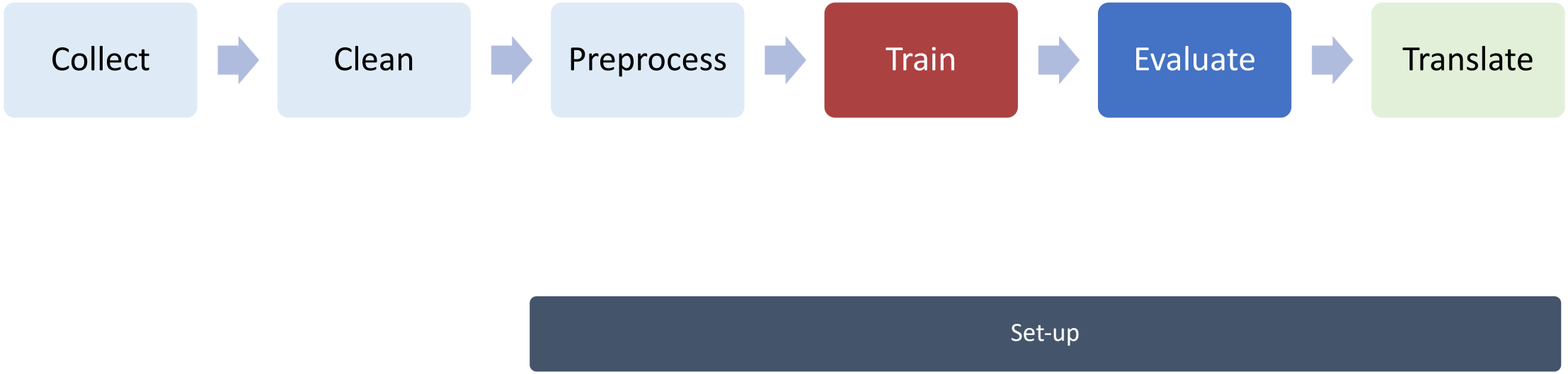
Training a NMT model

- But getting it right requires attention…

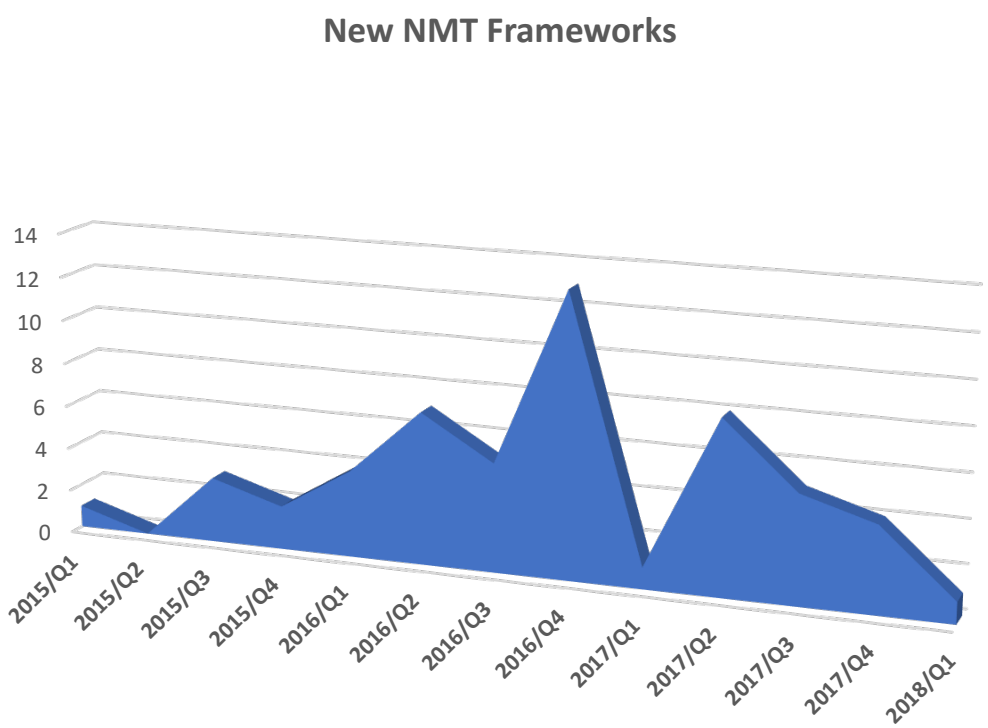


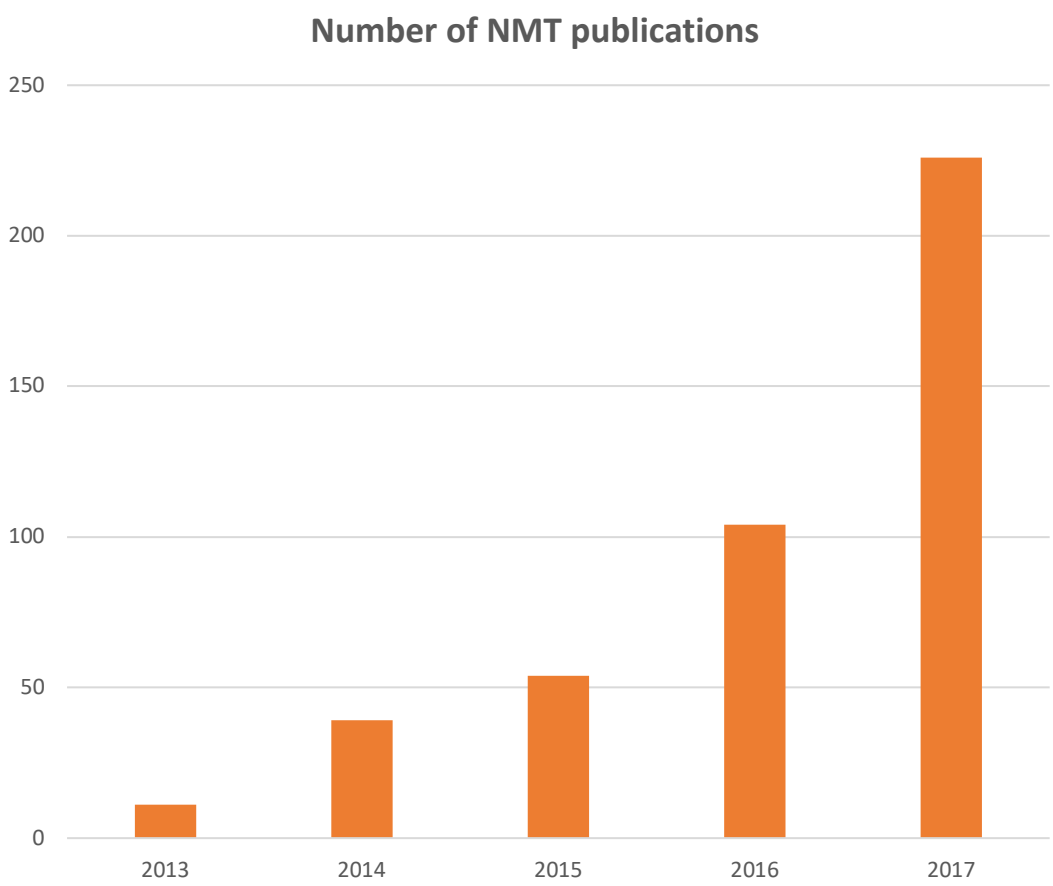
Training a NMT model – reaching state of the art

- And quickly becomes complicated

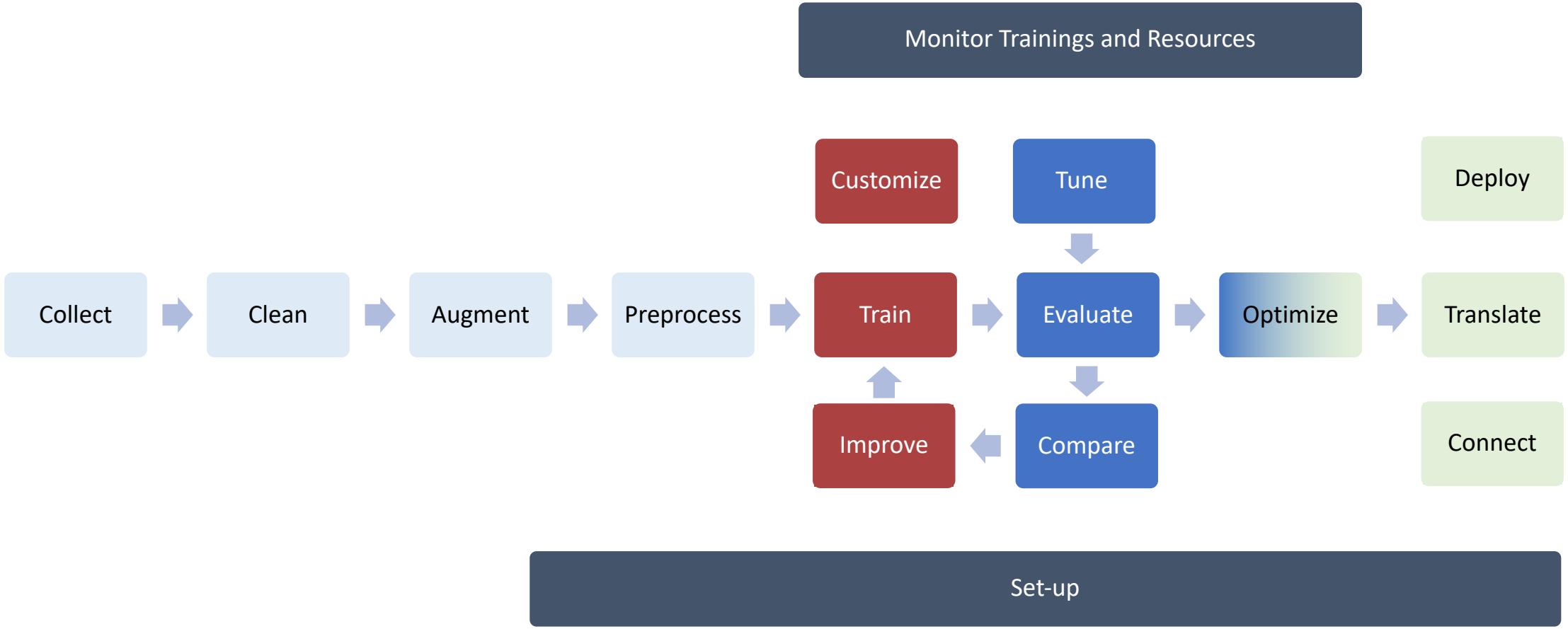


Staying up to date in « competitive » environment





Training a NMT model – Going for production



Open Source motivations and commitment

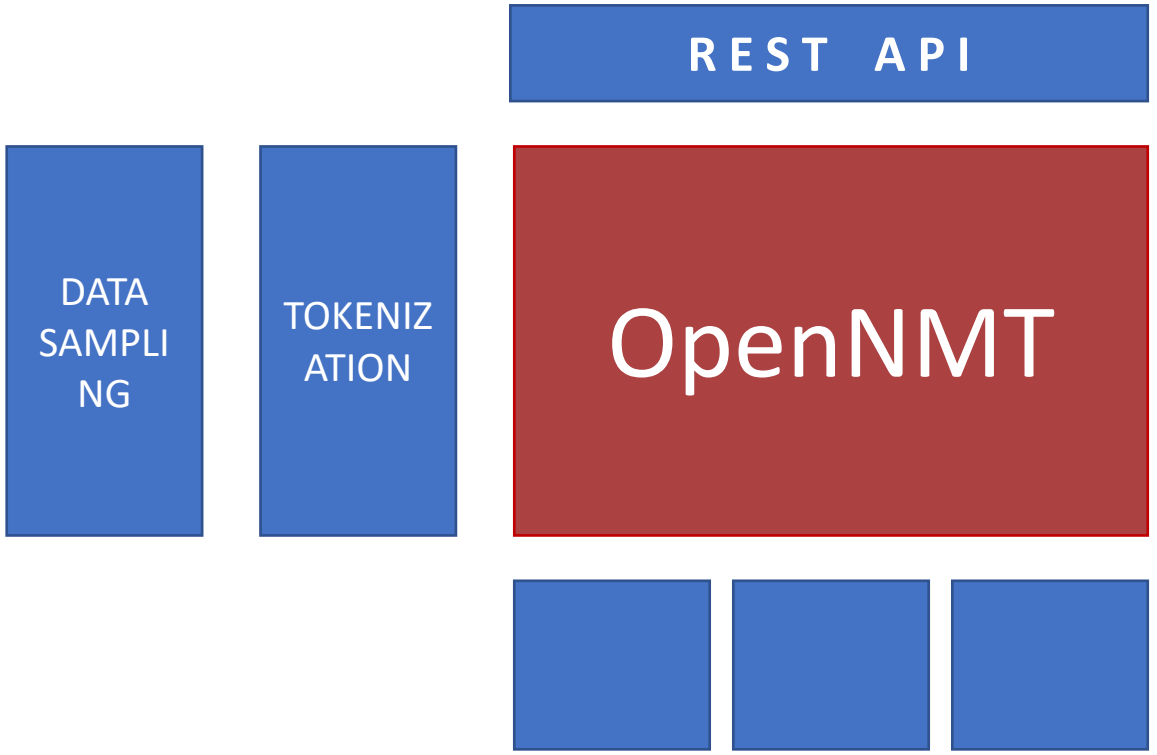
- Should stay state of the art
  - Meaning real focus on evaluation, reading papers, testing ideas
- Sharing should cover:
  - Technology
  - But also know-how and processes
- Must have short term flexibility
  - And provide long term maintenance support
- Has to be **ready to use for business** use case
  - No secret component

Typical questions and limitations

- Is there new cuda drivers?
- I have a problem in compiling Torch
- Which version of OpenNMT should we use?
- But I read that Sockeye is better!
- Do you have this feature, there is a paper about it – it is a must-have!
- Where is the corpus?
- How to tokenize Chinese?
- What are the best parameters?
- How to translate and how to deploy?
- It is taking too long to train!

From a raw Engine to a turn-key solution

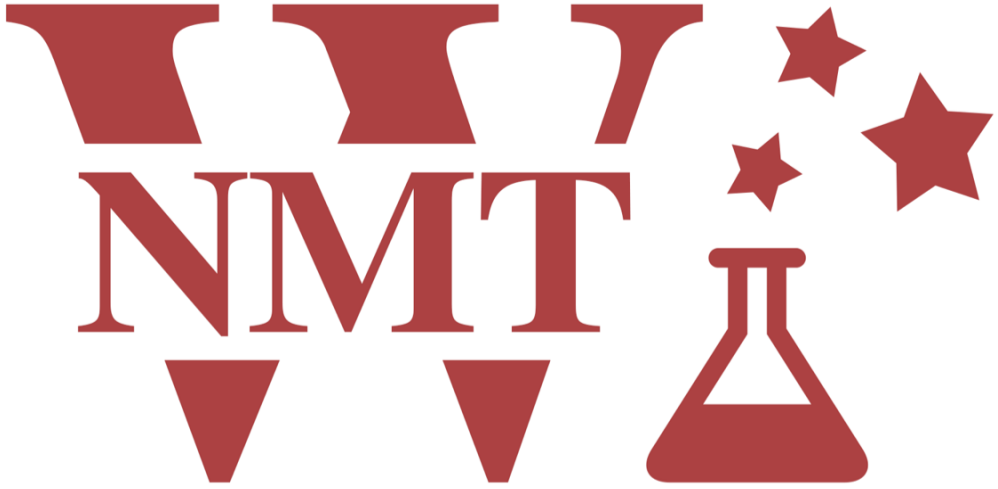
- Integrated tokenization
- Simple API
- Hooks for extended features
- Sampler for handling very large dataset



Going further – announcing the NMT wizard…

- New application part of OpenNMT
- Available today on github

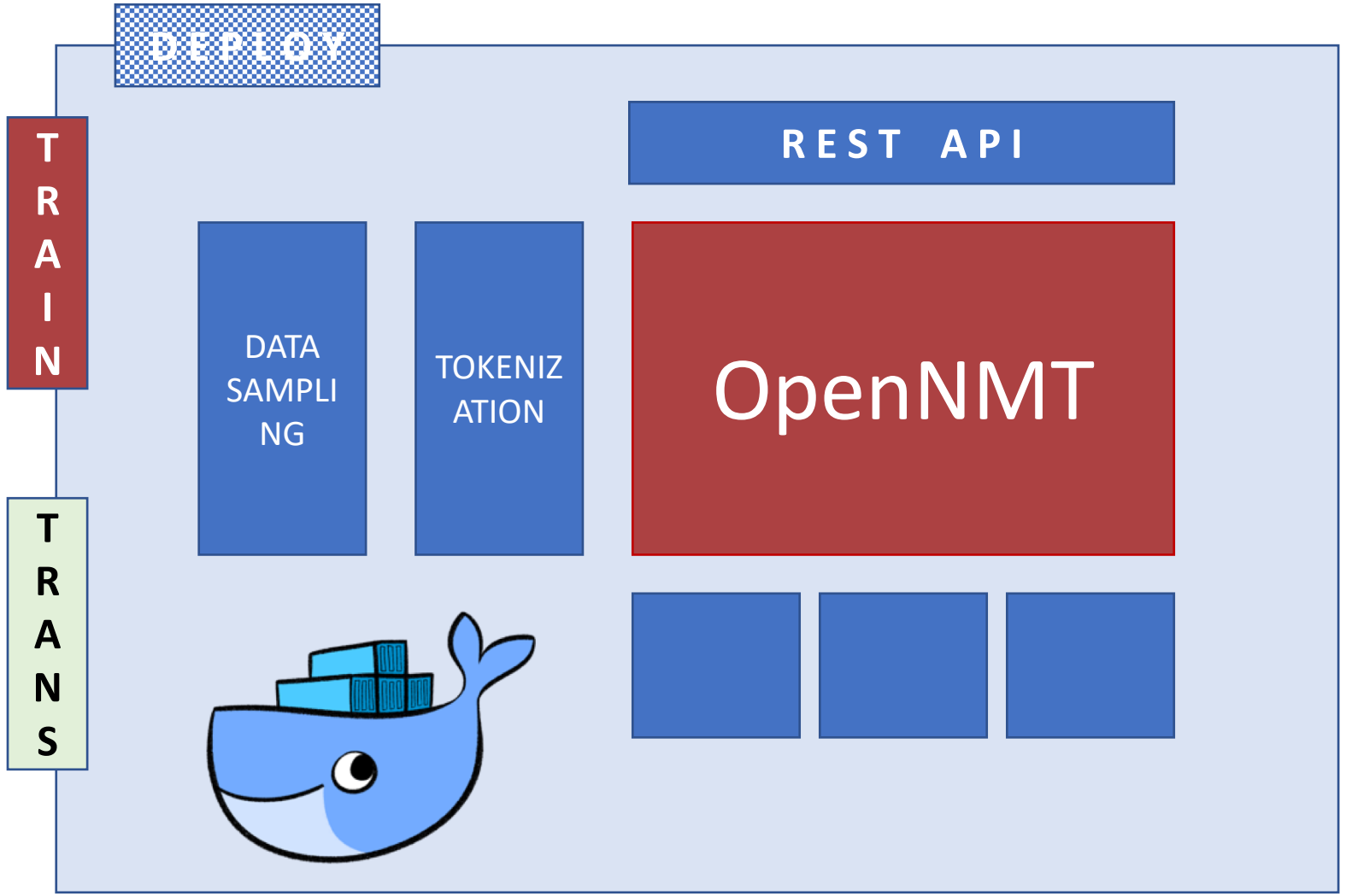
http://github.com/OpenNMT/nmt-wizard



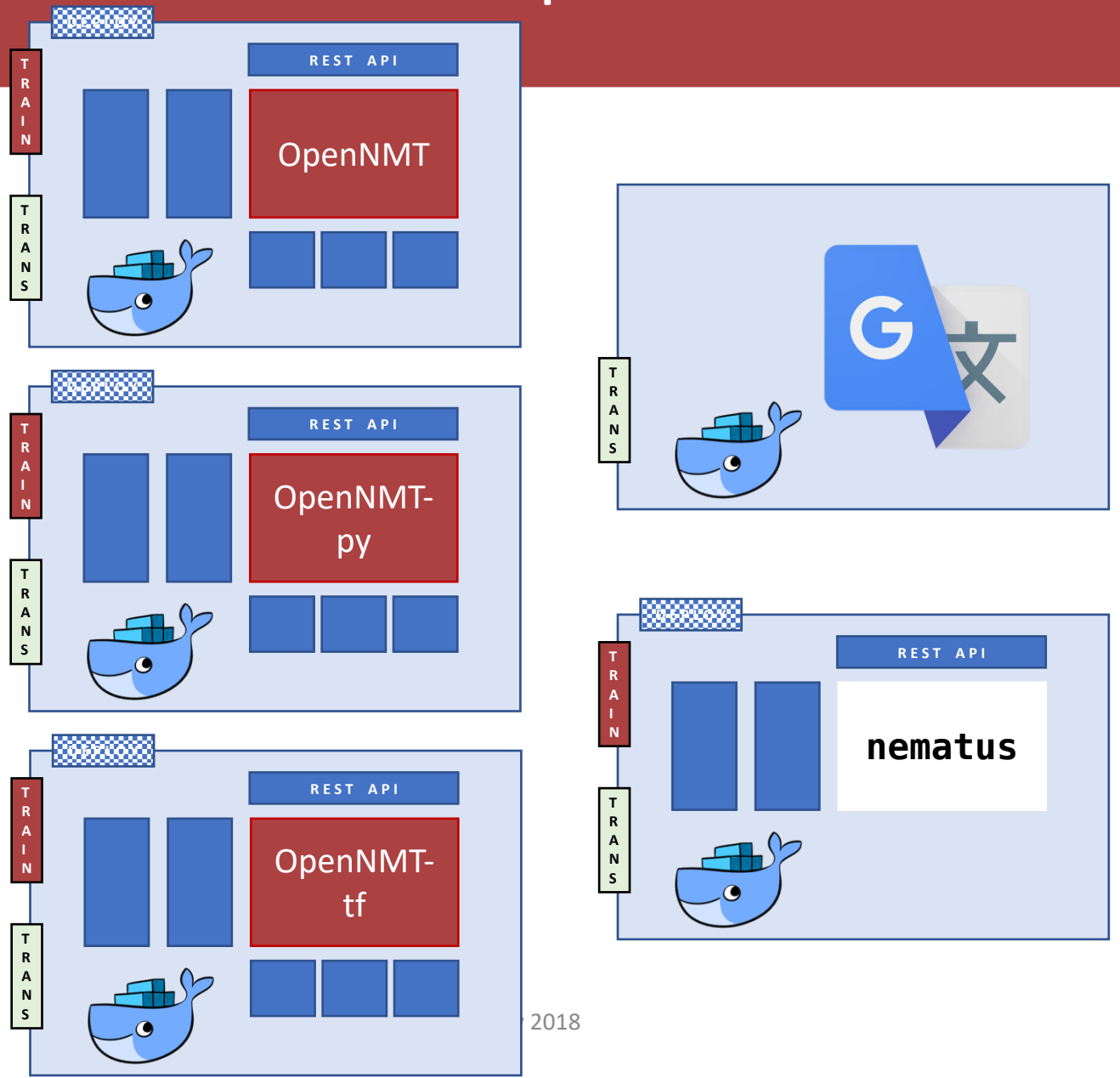
- Generic solution for training, and deploying NMT engines on different types of remote services

The NMT Wizard

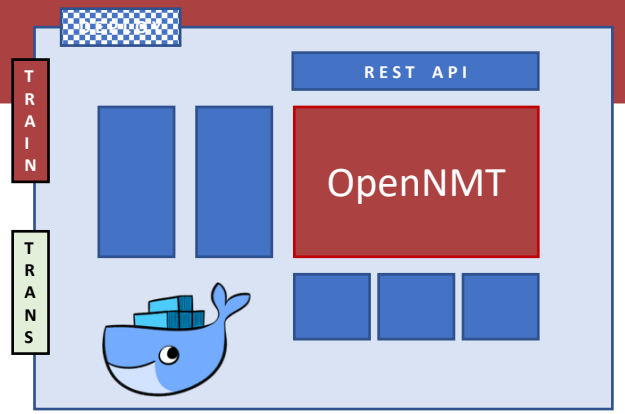
- Generic entrypoint encapsulating the full process and providing:
  - Training
  - Translation
  - Deployment (*coming soon*)

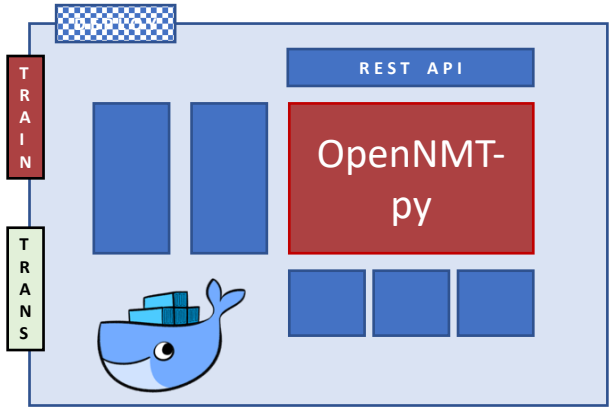


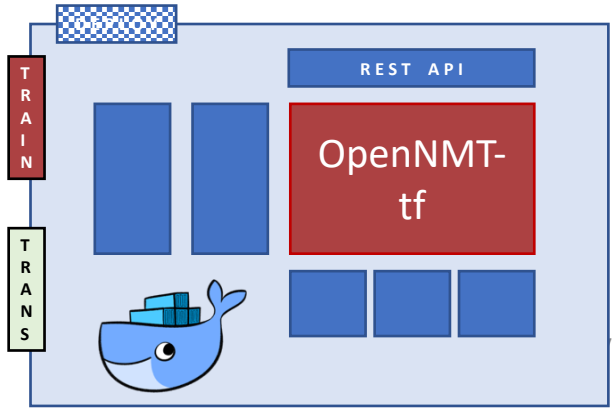
The NMT Wizard – Multiple frameworks

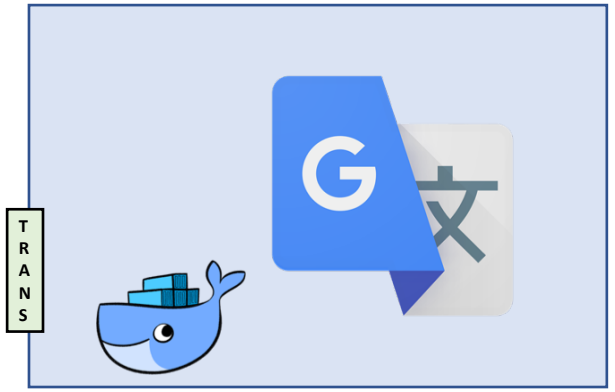


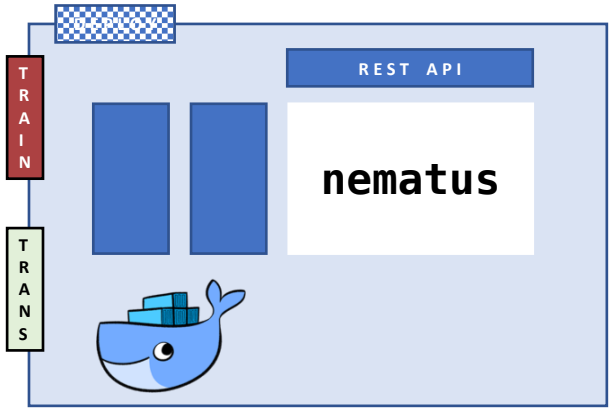
The NMT Wizard – Multiple frameworks



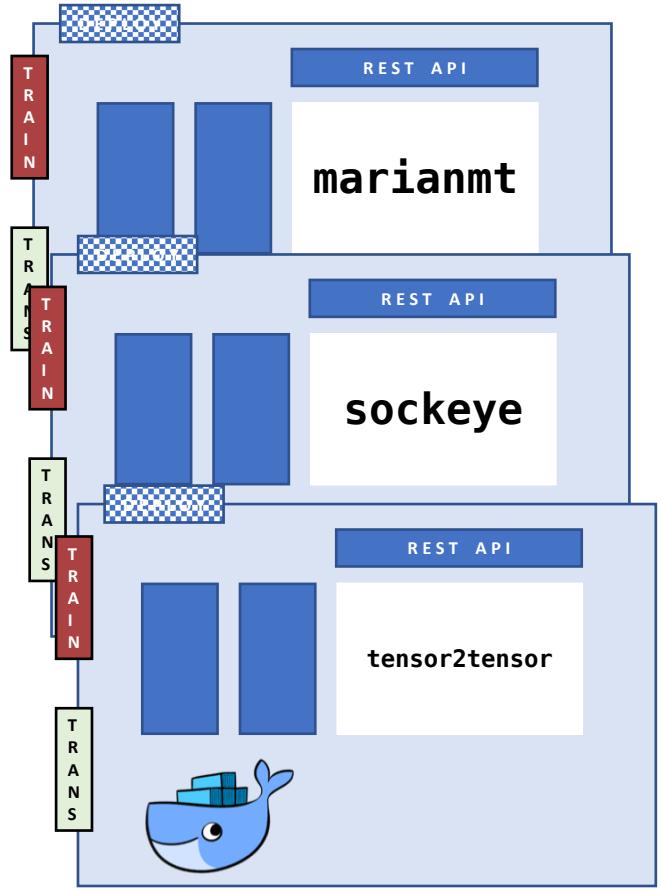








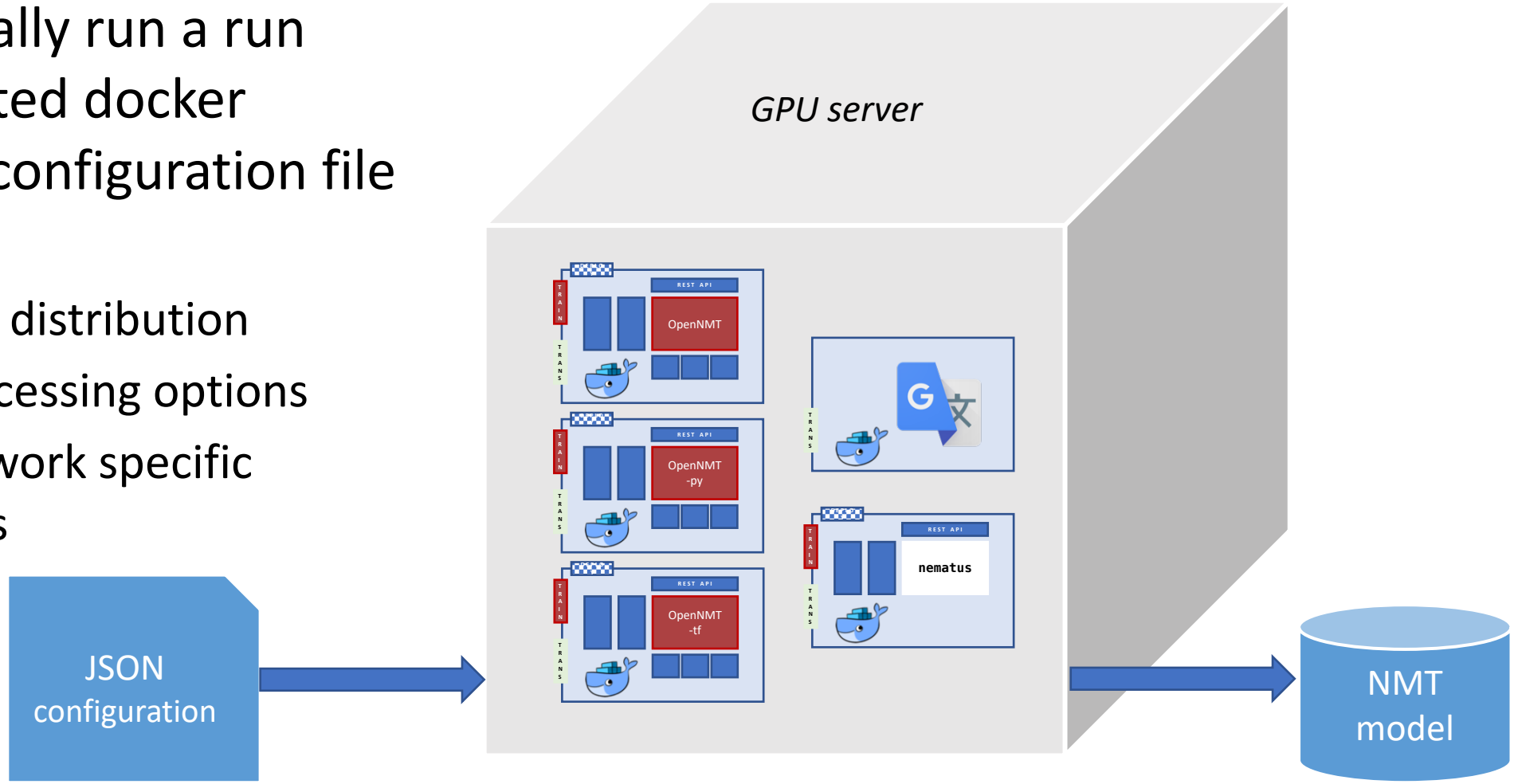
to be further extended:



The NMT Wizard – launching a service

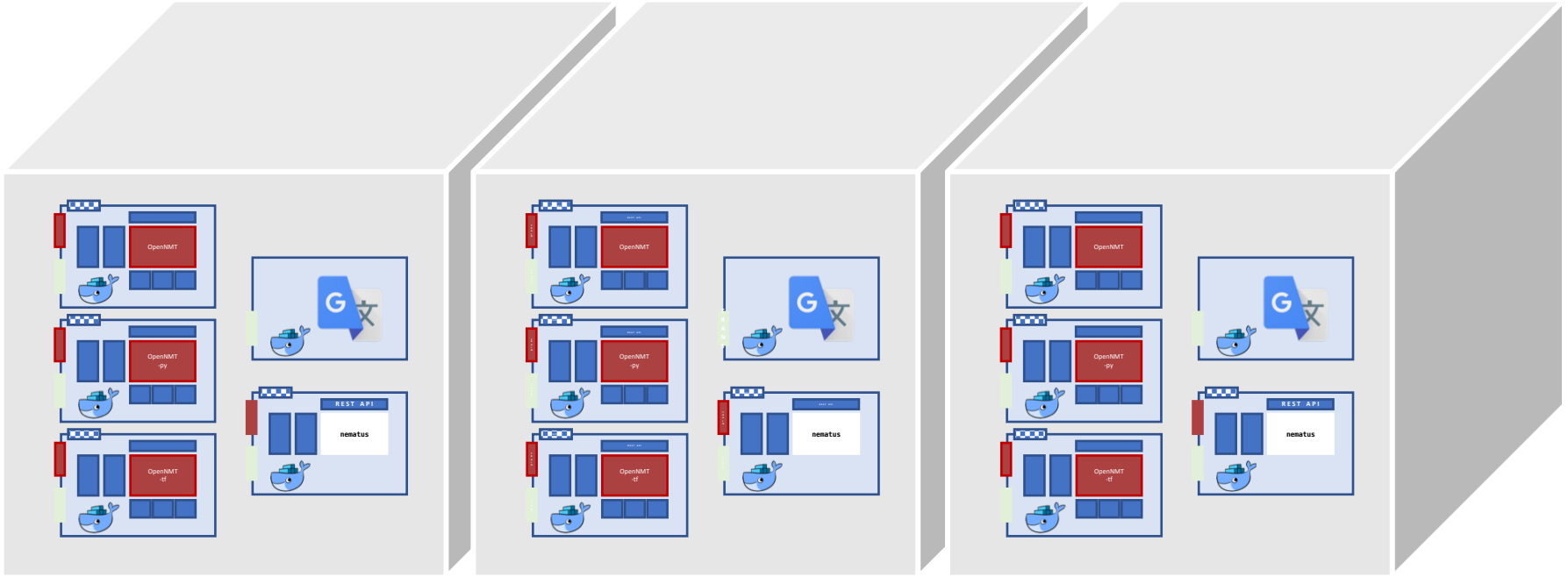
Automatically run a run from selected docker through a configuration file setting:

- Corpus distribution
- Preprocessing options
- Framework specific options

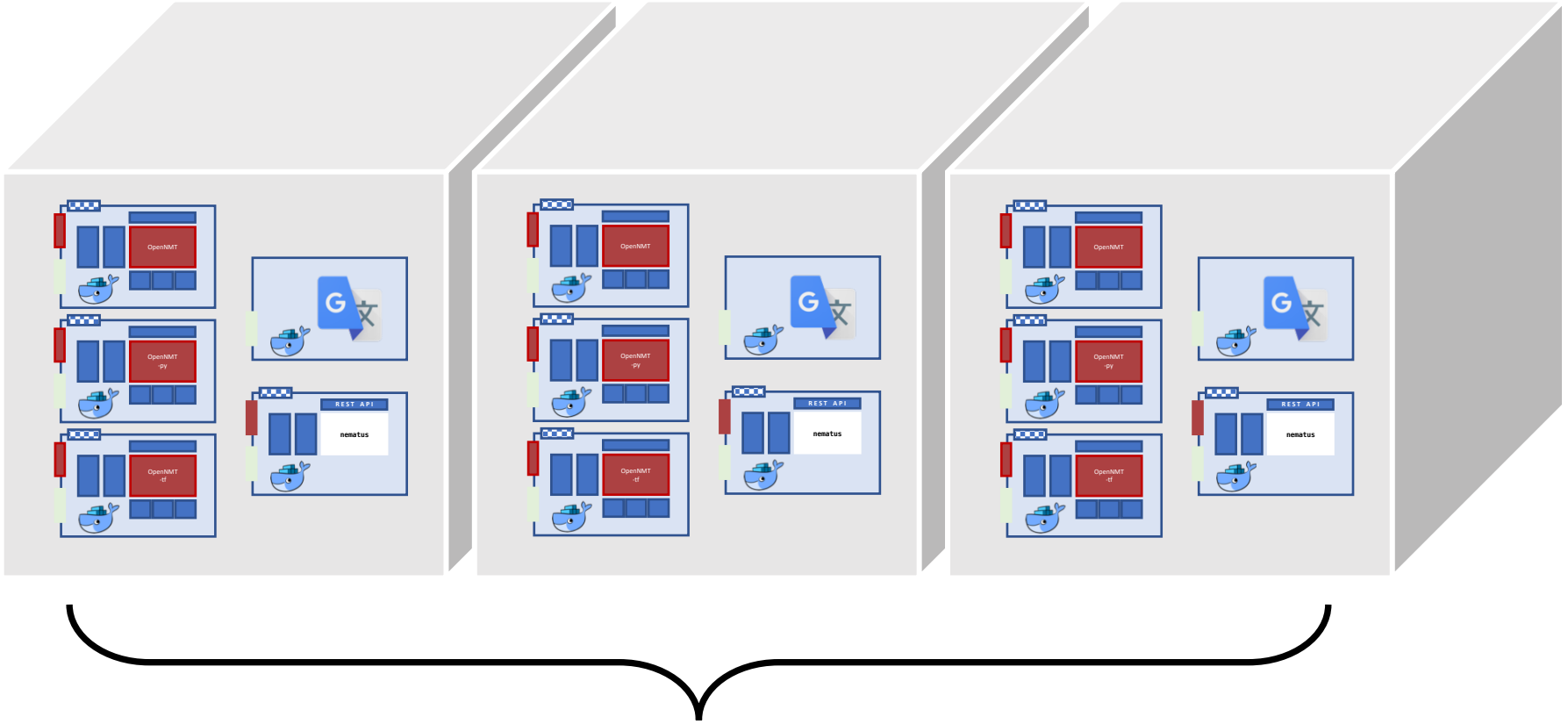


➢ Generates a NMT model

The NMT Wizard – resource management



The NMT Wizard – resource management



Automatically manages multiple resources

Resource Management

The NMT Wizard – service management

Multiple services available for local and cloud training.

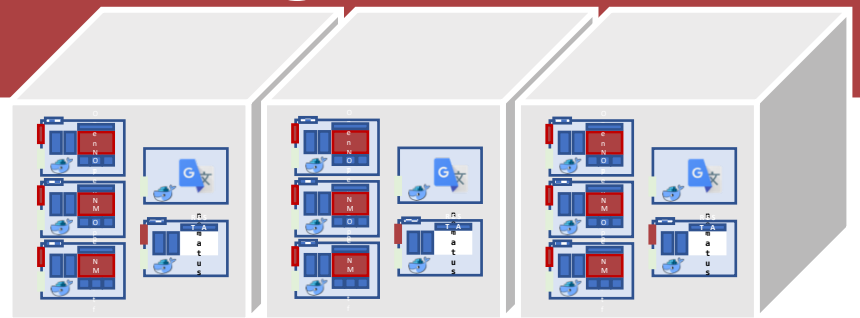
Currently supported:

- ssh service
- EC2 cloud service
- Torque grid system

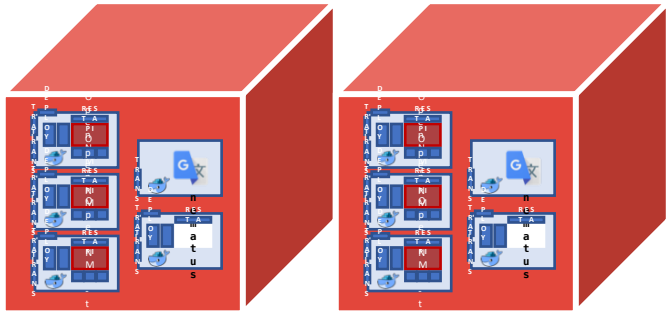
In progress:

- OpenStack services

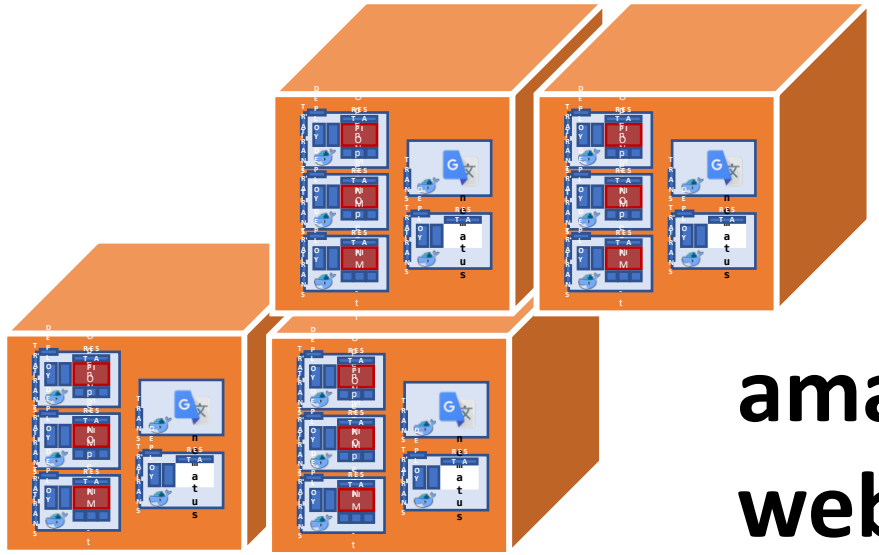
To be extended to other cloud computing services



Dedicated servers



openstack™

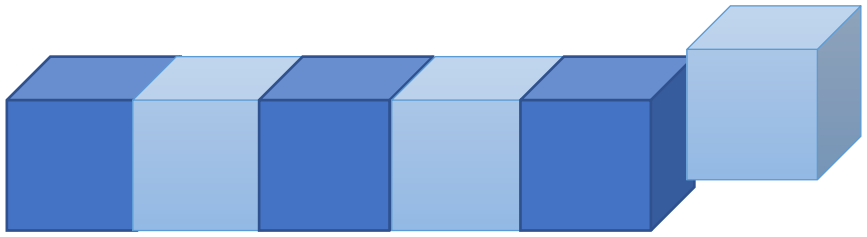


aws

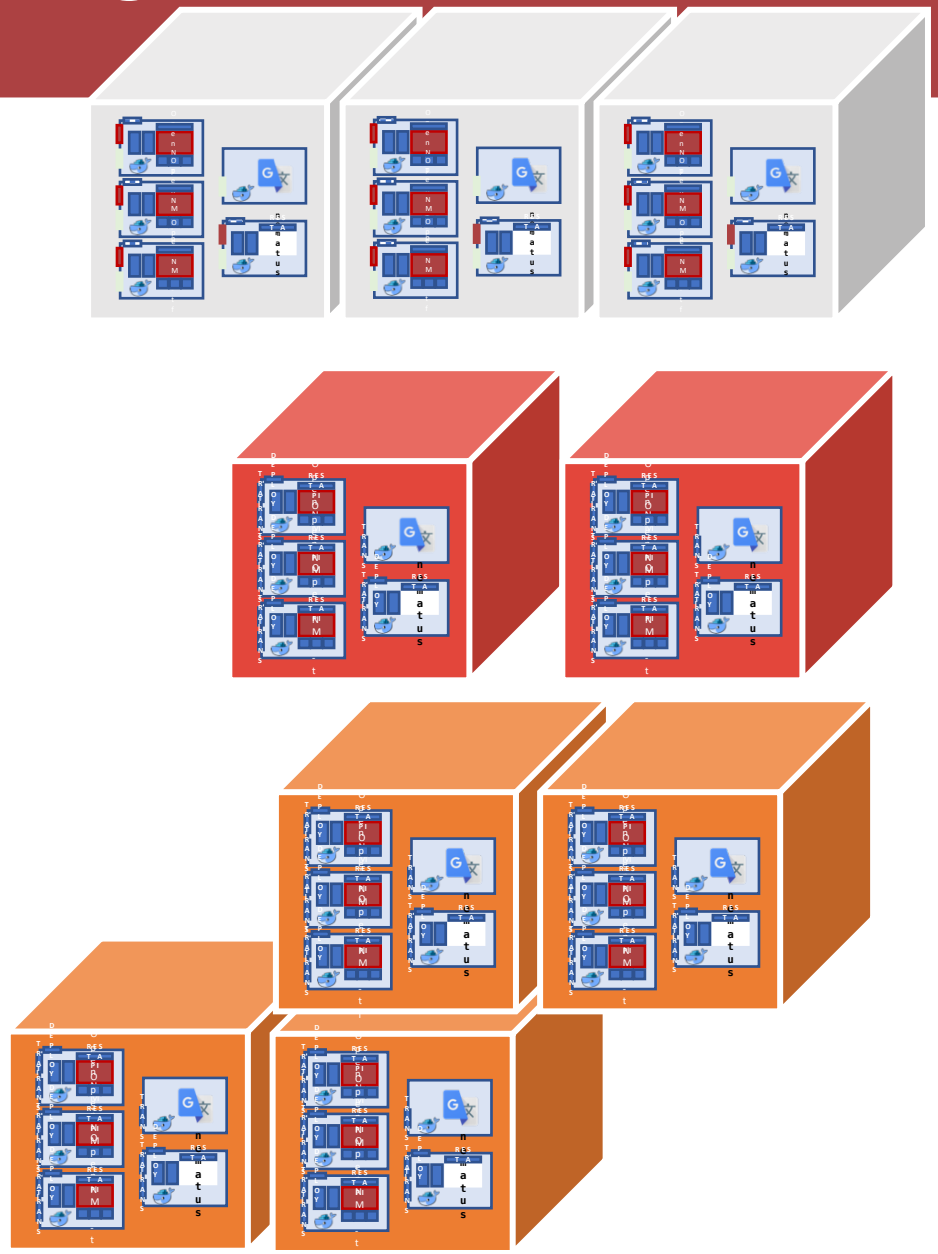
amazon web services

The NMT Wizard – task management

- Built-in queuing and automatically monitoring system for launching batches of runs with limited resources



Task Management



openstack™

aws

The NMT Wizard – in short

- Launch from your laptop with one simple commandline unlimited numbers of training with no server configuration, software installation, local connection and monitoring of tasks:

```
$ python launcher.py launch
            -s ec2
            -o '{"launchTemplateName":"GPU_G3_4xlarge_50Gb"} '
            -i nmtwizard/opennmt-tf
            -- -ms s3_models:
            -c @../example/min-tf.json train
```

- Monitor the progress:

```
$ python launcher.py lt
TASK_ID                              LAUNCH DATE         IMAGE                  STATUS    MESSAGE
JPR_93301441-6c2e-4fb7-8ecc-0cdfa42 2018-03-01 22:11:13 nmtwizard/opennmt-lua stopped   completed
JS_65ecbc81-32b5-46f2-a6e1-cb198119 2018-03-01 22:56:32 nmtwizard/opennmt-tf  allocated
JS_65ecbc81-32b5-46f2-a6e1-cb198119 2018-03-01 22:56:32 nmtwizard/opennmt-py  queued
```

- List services, terminate tasks, pipe training and translation, ….

The NMT Wizard

- Find documentation on github OpenNMT/nmt-wizard
- Learn more during Hackathon day and launch trainings with OpenNMT/OpenNMT-py/OpenNMT-lua/Nematus on EC2 in a single commandline!

- Will facilitate **training and deployment** of OpenNMT models, but also reproducing and benchmarking of other framework results in fully controlled and comparable environment (corpus, tokenization, dictionary, hardware)

Perspectives

- Improving technology bricks need to be done in sync with facilitating training, evaluation and access to neural models
- Need to consider training as a continuous process and not as a one-shot competition
- Focus should be on the data, on the parameters and less on the IT, the process and the actual tool