Hardware/Software setup

Hardware

- Most of competitions (expect image-based) can be solved on:
  - High-level laptop
  - 16+ gb ram
  - 4+ cores
- Quite good setup:
  - Tower PC
  - 32+ gb ram
  - 6+ cores

Hardware

Really important things:

- **RAM**

  If you can keep data in memory — everything will be much easier

- **Cores**

  More cores you have — more (or faster) experiments you can do

- **Storage**

  SSD is crucial if you work with images or big datasets with a lot of small pieces

Cloud resources

Cloud platforms can provide you with a computational resources.

There are several cloud options:

- Amazon AWS
- Microsoft Azure
- Google Cloud

Cloud resources

Cloud platforms can provide you with a computational resources.

There are several cloud options:

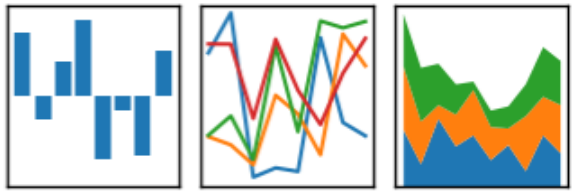
- Amazon AWS spot option!
- Microsoft Azure
- Google Cloud

Software: language

Most of competitors use Python data science software stack.

Basic stack

Most of competitors use Python data science software stack.



IDE

Special packages

dmlc XGBoost

Microsoft / LightGBM

Keras

danielfrg / tsne
forked from osdf/py_bh_tsne

External tools

VOWPAL WABBIT

srendle / libfm

guestwalk / libffm

baidu / fast_rgf

Conclusion

- Anaconda works out-of-box
- Proposed setup is not the only one, but most common
- Don’t overestimate role of hardware\software