Research Computing @LSE

Neil Prockter – 17th September 2019

LSEs current HPC/Research Computing services

is called Fabian it offers:

- batch processing
- desktop applications (for shorter interactive use)
- jupyter notebooks
- separate secure environments for sensitive data

Contact fabian@lse.ac.uk in order to request an account on the service

LSEs current HPC/Research Computing services

Rstudio Server Pro for teaching and non-professional research.

https://rstudio.lse.ac.uk

Like Rstudio but in your web browser leave a session running attach to it from anywhere, even an ipad/tablet

Contact fabian@lse.ac.uk in order to request an account on the service

LSEs current HPC/Research Computing services

Fabian consists of

- 17 high spec compute servers (400+ CPU cores, 3TB memory)
- 1 lower spec server for running services / 1 shared storage server / 4 supporting storage servers

its small as clusters go but far bigger than any desktop

Batch processing

Benefits

- submit a job it runs and the system emails you when finished
- leaves your pc free to, watch youtube, take notes, reboot for patches etc
- run many (some use 1000s) copies of you code in parallel over many machines

Downside, it is some extra work to set things up

This year we will move to the cloud

How will this benefit users of the service

- Fabian is a fixed set of servers which are sometimes fully committed and sometimes under used. The cloud will mean we can automatically add and remove servers to meet demand, greatly increasing the potential capacity and reducing the wait when servers are busy (reduces the time to research)
- The cloud offers hardware on a pay by use basis. For example if you want to see if your problem can be solved faster with a GPU or more memory you can use tht for a few minutes or hours and we pay for just that use. There is no way we can do that on site.
- The servers are getting old and need replacing/expanding and that means its a good time to look into other more cost effective options

This year we will move to the cloud

How will it make the service cheaper/better or more flexible?

- We spend a good deal of time looking after the servers, time that could be better spent improving the service and better supporting the users
- We spend time installing applications and software, in the cloud there are machine images which have all this done and optimised ready for use
- We spend money on running a physical Datacenter to host Fabian, money that could be spent on more resources (cloud providers do this cheaper and better)

What you can easily do in the cloud today

Jupyter Notebook (option 1)

MS Azure notebooks https://notebooks.azure.com/

- Its free
- Sign-in using your LSE account
- Python 3 and R support
- Store multiple projects there, well integrated with github
- Lots of examples included

Jupyter Notebook (option 2)

AWS - many options

- https://aws.amazon.com/emr/
- https://aws.amazon.com/sagemaker/ - good git support
- not always free but you can get started using the free tier for one year (per email address) and get $150 credit through AWS Educate program
- Great support for ML (deep learning)
- Lots of examples included

Jupyter Notebook (option 3)

Google Co-Lab

- https://colab.research.google.com/notebooks/welcome.ipynb
- Its free
- Not quite so many examples but enough to get going
- But it looks a bit different to jupyter (it’s jupyter + google docs)

Tips

Other interests

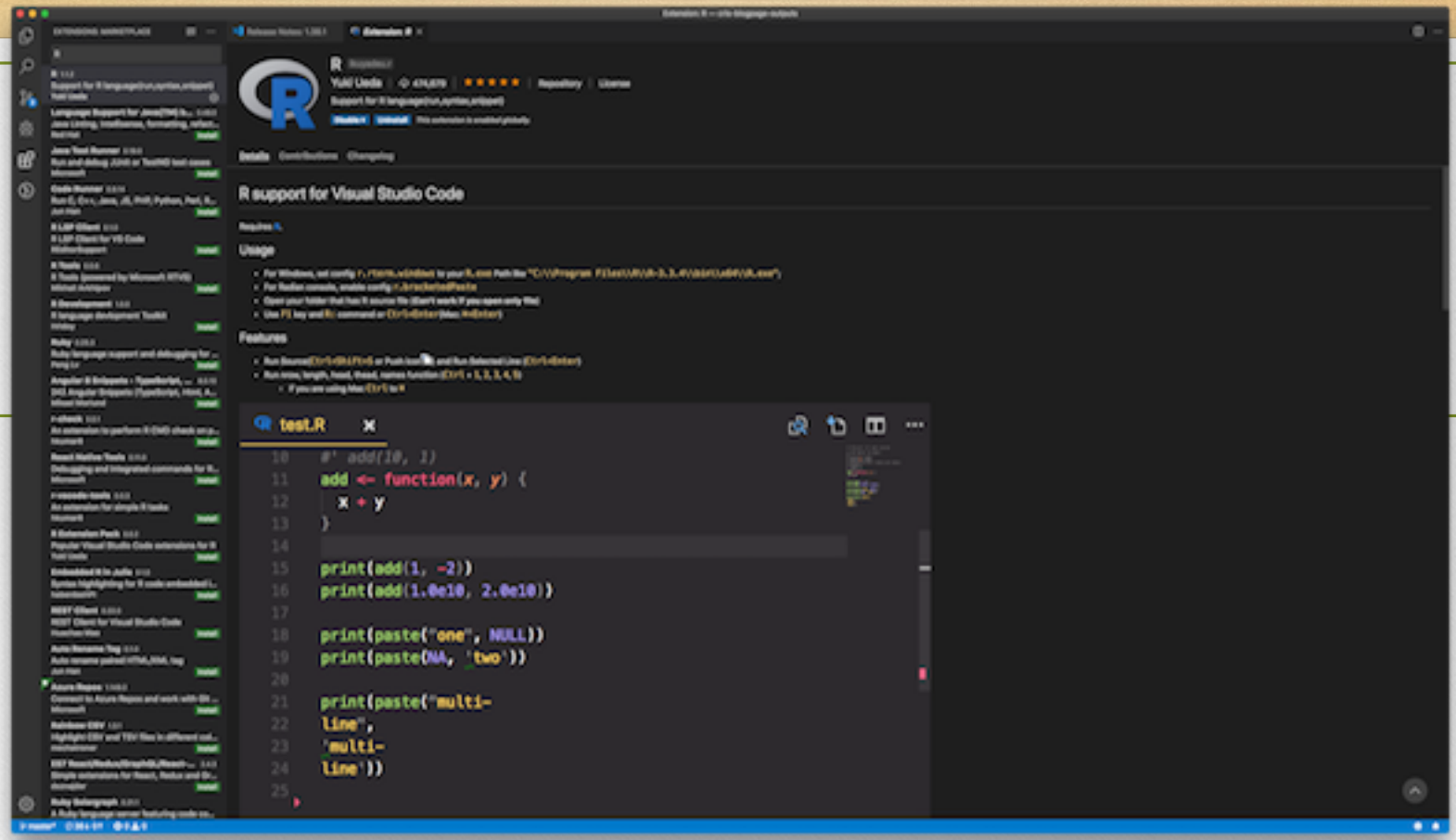
With Azure Machine Learning studio you can visually design workflows to process your data

- https://azure.microsoft.com/en-gb/services/machine-learning-studio/

With AWS Glue you treat csv or json files as a database

- https://github.com/aws-samples/aws-glue-samples

Editors



- Visual Studio Code is a great multi-platform editor https://code.visualstudio.com/ with git support built-in and great support for python and R through extension packages

Questions