The TUD-wide JupyterHub: A computational home for students at TU Delft

We propose to create and maintain a JupyterHub (a cloud computational environment) that is always available to all TUD students and staff.

Current situation

Programming in education at TUD

What do we want to say here?
- Digital skills?
- EWI courses?
- TNW courses?
- Overview of programming courses across campus?
- Core Faculty Module in CiTG?

Overview of existing tools at TUD

- there are many home-grown tools already:
  - Weblab: web-based system for managing submission, auto-running code, managing grading of code
    - No support yet for jupyter notebooks
    - It is not a "cloud-based" service, students still need to install the full development tools on their own laptops
  - CPM: for submission and autograding of code
- None of these are computing platforms:
  - The workflow is that the student sets up a complete programming environment on their own laptop / computer
  - Develops and runs it there
  - The upload it to be graded
  - To perform small tests for programming and math skills
- In addition, we currently have a license for Vocareum
  - This is an online, cloud-based computing platform linked to a specific course
  - Rich and sophisticated assignment submission, management, and grading interfaces
  - Direct linking to Brightspace
- Our proposal is different:
  - We want to streamline the access that students have to programming tools (including in-house software) and computation
  - They log in, they get going *straight away*
  - No installation, no software configuration
  - We do not aim to provide or replace existing course management tools

- Instead we aim to facilitate students and teachers bringing programming and computational tools with *no barrier* into their learning
- In our proposal, the students log in to a pre-configured computational environment on the web via the "Jupyter Hub" and can immediately start running / modifying / writing code and performing computational work
- This could be just instructive for demonstrating phenomena we want to teach using interactive computation
    - Lecturers could provide notebooks easily via the cloud for the students to run in lectures or practicals
- It would also be just a "computational home" for the students throughout their education
    - Data analysis for research practica could be done there
    - Students could use it for their Bachelor research projects as well as minor and honor programs
- It could also be combined with a submission system:
    - It would be easy to establish automated links to existing programming assignment submission systems (weblab, cpm, etc)
- Once it is running, we could also consider tools like NBgrader, which could be deployed on it to provide full course submission / grading management
    - It could form an alternative eventually to Vocareum, hosted here in house, but not right away

About Jupyter and Jupyterhub

Project Jupyter is a broad collaboration that develops open-source tools for interactive and exploratory computing. The tools include: over 100 computer languages (with a focus on Python), the Jupyter Notebook, JupyterHub, and an ecosystem of extensions contributed by a large community. It has also grown as a platform to use in the classroom, to develop teaching materials, to share lessons and tutorials, and to create computational stories.

Jupyter is being adopted for teaching and learning in a growing number of institutions. For example, Jupyter has been adopted as the primary learning platform by UC Berkeley's new undergraduate Data Science program which now serves thousands of students per year. At TU Delft, Jupyter is already being used in TN (TN1405, TN3155), NB (NB2211), and TODO. JupyterHub can really enhance the development of Core Faculty Module, two faculty-wide 3EC courses on modeling and programming at CiTG.

JupyterHub is software designed so that users can log in via a web browser from any operating system or device to access large sets of software tools used in companies and research. It reduces the barrier to entry, because all of the software is available without having to install complex tools, provides some user storage space, can give access to larger computational resources than a laptop can provide, allows instructors and IT to manage what software is available, and provides mechanisms for assessment tools. For education, it is an ideal way for instructors and instructional support to deliver interactive computational content. These tools are also now widely used in companies, which provides a direct connection for the students moving into industry. Given the coronavirus measures, the

JupyterHub offers an effective way to enhance online learning through computational thinking and experimentation.

Examples of existing large scale Jupyterhub deployments at other institutions

- BIDS (gke) (cloud)
- UC Davis self-hosted jupyterhub (bare metal)
- UC San Diego (bare metal deployment)
- Columbia University (Abernathy's earth sciences courses and research consortium)

Stakeholders

- TUD AI institute
- Programming skills program
- Instructors teaching Python, R, Julia, C++, Sage, and [other supported languages](#)
- Instructors who may wish to integrate computational aspects into a non-programming course
  - Simulations / demonstrations of concepts in interactive notebooks
    - Example: QuTiP simulations as part of TN3155 Quantum Sensing and Measurement
    - Math and programming convergence course in AESM1305, CiTG (https://github.com/dvoskov/convergence_modelling)
    - Exploratory computing with Python, CiTG (https://mbakker7.github.io/exploratory_computing_with_python/)
    - **Potentially lots of examples across campus!!!!!**
  - Jupyter notebooks as lab notebooks
    - Deployed already by Margreet Docter for NB2211 in Vocareum
    - Also being actively developed by Freek Pols for TN1405
  - Could pitch as a virtual lab or lecture hub of sorts to help accommodate the coronavirus situation
- ICT innovation
- Extension school
- Library data managers
- Researchers?
  - Could be used for BEP and MEP computational resources
  - Eliminate the challenge of finding fixed "desktop" computers for students to work at in the case that their own laptops are insufficient due to age, computational power, software installation (request arises often in our department)
  - Potentially could be offered to groups / departments / faculties

Goals

- Provide a universally available computational environment to students and staff

- Implement this in a highly scalable and affordable manner
- Enable TUD instructors to use it as a computational platform for their courses
- Plan using this platform to provide extra services in the future

Proposal

- Launch a pilot project for a university-wide Jupyterhub
- Allocate sufficient FTE to set up and maintain this pilot
- Create a Jupyterhub cluster hosted on Azure cloud computing service
- Connect this cluster to a NetID login system

Possible alternatives

Jupyterhub using university hardware

This would also be a serious option, with [some paths](#) towards making this possible, but would require a moderate investment in hardware, combined with a more extensive commitment from the university to provide technically trained staff to maintain it. A cloud backend based on eg. Azure would provide much easier deployment and a much higher potential for flexible scaling.

Vocareum or a similar commercial provider

Gary: I understand that ICT Innovation team was exploring this option. Current Vocareum platform is great for programming courses, where it is important to manage the submissions and grading of programming material inside a specific course. However, it does not fulfil the role of a student "digital home", as the environment is only active for a specific course. Furthermore, while they offer good support for a standard computational environment, customizing this is not easy (it is currently a beta feature that did not work very well when I tried it).

Update: Vocareum also supports deployment of a "plain vanilla" jupyterhub (single container per user, no course management tools) to AWS / Azure / etc. Basically, the would provide the IT support for SAML, persistent storage management, kubernetes, container management and culling, all the stuff that we do not have the expertise to do in-house. Also, many things that 2i2c also cannot do (yet). Erna will look into this for an initial pilot phase.

SURF Jupyterhub service

When we discussed with them, this was offered on only a relatively small scale and did not employ the flexibly scalable design that would be crucial for this service (clusters could be reserved for specific courses for certain periods of time, but could not be dynamically scaled up and down based on demand.)

About the authors

- Anton Akhmerov
  - Ass. Prof. TNW, actively used jupyter both in research and courses for 6 years.
- Lora Armstrong
- Arco van Geest
- Jason Moore
  - Assistant Professor BME/3ME, Co-wrote book "Teaching and Learning with Jupyter", established JupyterHub for UC Davis and LibreTexts, SymPy developer
- Gary Steele
  - Professor TNW, TN Teacher of the Year 2019, Teacher of 1st year programming lectures of TN1405 since 2017/2018, Teacher of TN2513 Computational Science since 2018/2019 revamping course to transition from Matlab to Jupyter Notebooks in 2018/2019 and also then to Cloud platform Vocareum in 2019/2020
- Mark Schenk
- Denis Voskov, Associate Professor at CiTG, actively using Jupyter notebook for teaching several courses in the new Geo-Energy Engineering MSc track and various simulation workshops

References

- JupyterHub: https://jupyter.org/hub
-