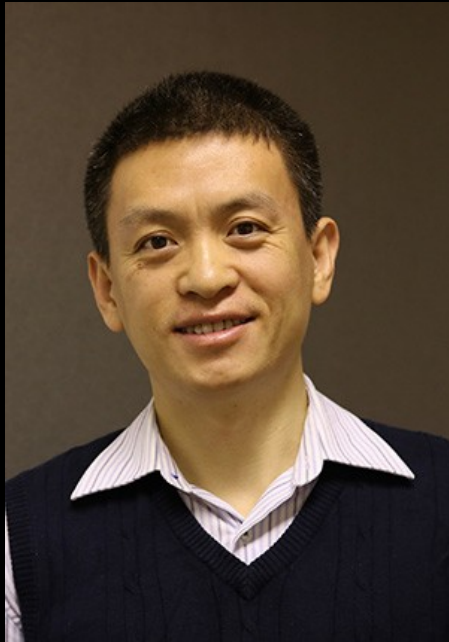It seems like a good idea, but is it scalable?

https://s3.amazonaws.com/lowres.cartoonstock.com/animals-scalable-product-mice-cats-slingshot-mdbn347_low.jpg

COM6012: Scalable Machine Learning - Spring 2021

https://github.com/haipinglu/ScalableML (Since 2019)

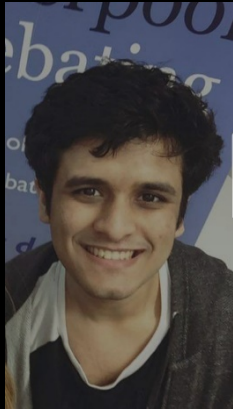The University of Sheffield

Two Lecturers

Haiping Lu
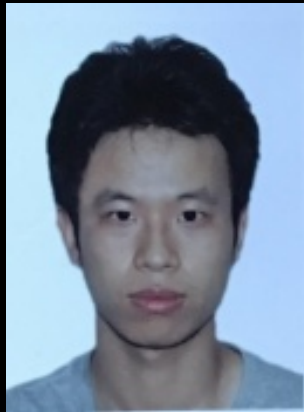Module leader

Mauricio A. Álvarez

Four Demonstrators (TAs)
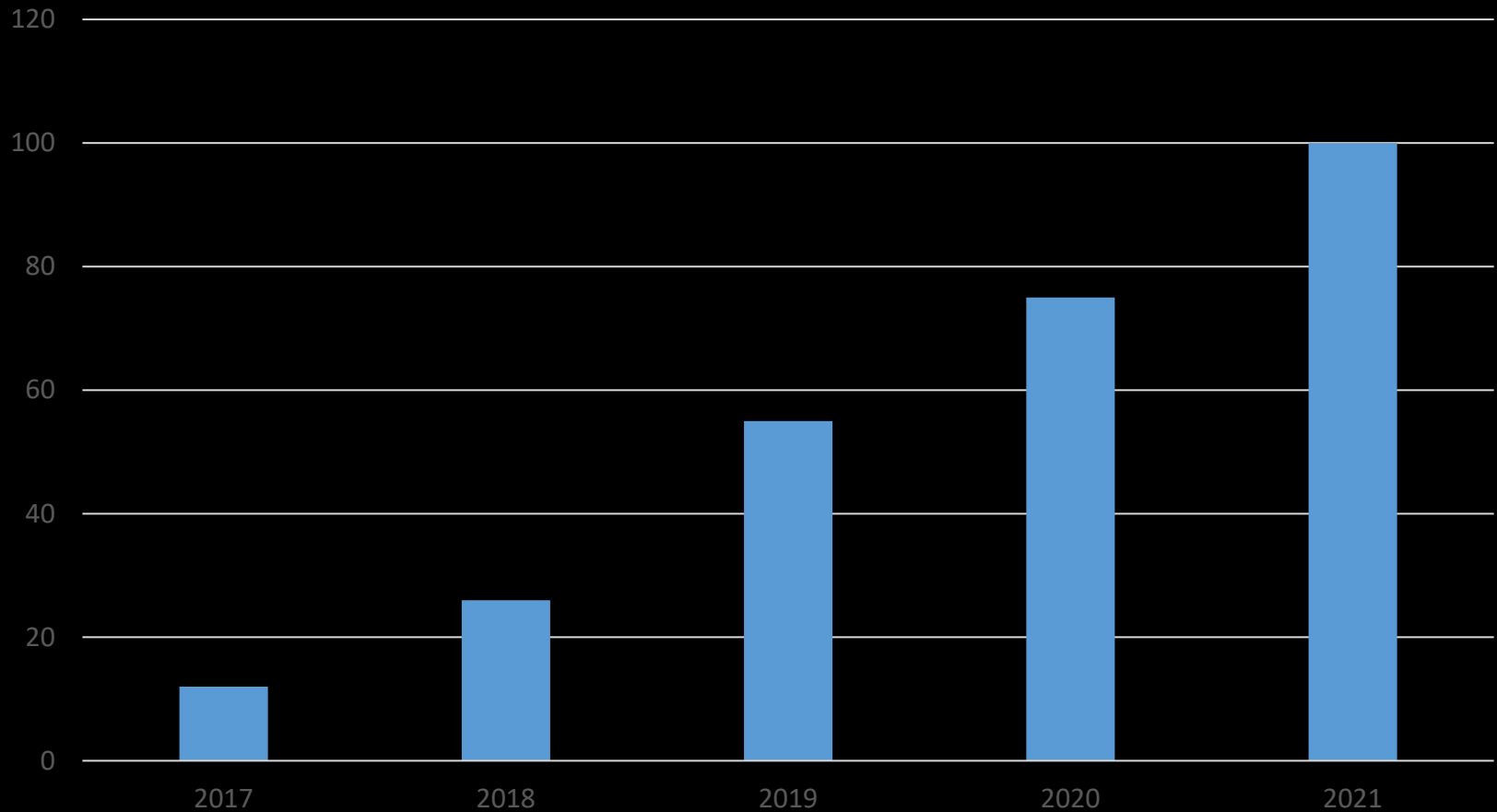

Areeb Sherwani
Head

Mario Alejandro Hevia Fajardo

Mingjie Chen

Tom McDonald

Number of Registered Students

Schedule and Interactions

- Monday 9am: lecturer → each week's video lecture, slides, and lab posted on Blackboard (BB) and GitHub
  - From week 2: lab exercise reference solution posted on BB by Monday
- Tuesday 5pm: students → 1) studied all lecture videos; 2) started the lab with possible questions/problems found
- Wednesday 9-10am: BB Collaborate online with the lecturer
  - Lab demos, question answering, problem solving, material review
  - Get your questions ready before the session starts
- Wednesday 2-4pm: Discord online help-desk with the demonstrators
- Thursday 9-11am: Discord online help-desk with the demonstrators
- Thursday 5pm: students → completed the lab (except exercise)
- Friday 10-11am: BB Collaborate additional session with the lecturer

Other Interactions

- Face-to-face sessions with a demonstrator
  - Cancelled until further notice
- BB discussion board: post your questions to get answered by the lecturer
  - One general forum: general question/feedback
  - Two lecture & lab forums: first half + second half
    - To get help on lecture/lab contents
  - Two assignment forums: one for each assignment
    - To ask for clarification on assignment questions (i.e. the tasks to do)
    - NOT to ask how to solve the problems, the correctness of a specific solution, or share a possible solution. It is an assessment.
- Direct email to the lecturer: personal/private issues

Assessment (2 + 2)

#	Assessment	Release	Due	Mark (total: 100)
1	Blackboard Quiz 1	25 Feb 6pm	26 Feb 6pm	20
2	Assignment 1	26 Feb 11am	12 Mar 11am	30
3	Blackboard Quiz 2	25 Mar 6pm	26 Mar 6pm	20
4	Assignment 2	27 Mar 11am	30 Apr 11am	30

- Marking and feedback (from Student Handbook)
  - Quiz: 1 working week
  - Assignment: 3 working weeks

VPN: Necessary for HPC & Assessment

- See the official guide at https://www.sheffield.ac.uk/it-services/vpn

Remote Access VPN

VPN (Virtual Private Network) allows staff and students secure access to university-restricted services away from campus.

The university has implemented a new VPN service "FortiClient", which builds in support for multi-factor authentication (MFA). It will soon replace the existing (legacy) VPN service, which currently requires a Remote Access (RATS) Password.

- You will connect to the new VPN using your synchronised university password that you already use to connect to services such as MUSE.
- You will now need to perform MFA during the VPN connection process by approving the connection each time on your mobile device or token.
- **You must be set up with MFA before you begin setting up the new VPN.**

Do I need to use VPN?

Find out **when you need to use VPN** before connecting as most university services no longer require the VPN. Please only use a VPN connection if it's essential and disconnect when you're finished.

Setting up and connecting to VPN

Follow these steps to access VPN.

+ Show all

Step 1: Setting up Multi-factor authentication (MFA)

Step 2: Setting up your VPN connection

Step 3: Connecting to VPN with MFA

Contents: More Hands-on

Week	Date	Topic	Lecturer
1	08 Feb	Introduction to Spark and HPC	Haiping
2	15 Feb	RDD, DataFrame, ML pipeline, & parallelization	
3	22 Feb	Scalable matrix fact. for collaborative filtering (RecSys)	
4	01 Mar	Scalable K-means clustering	
5	08 Mar	Scalable PCA for dimensionality reduction	
6	15 Mar	Scalable decision trees	Mauricio
7	22 Mar	Scalable logistic regression	
8	19 Apr	Scalable generalized linear models	
9	26 Apr	Scalable neural networks	
10	03 May	Apache Spark in the Cloud (guest lecturer: Mike Smith)	