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

COM6012: **Scalable** Machine Learning - Spring 2022

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

The University of Sheffield

Two Lecturers

Haiping Lu
Module leader

Mauricio A. Álvarez

Five Demonstrators (TAs)

Areeb Sherwani (Head)
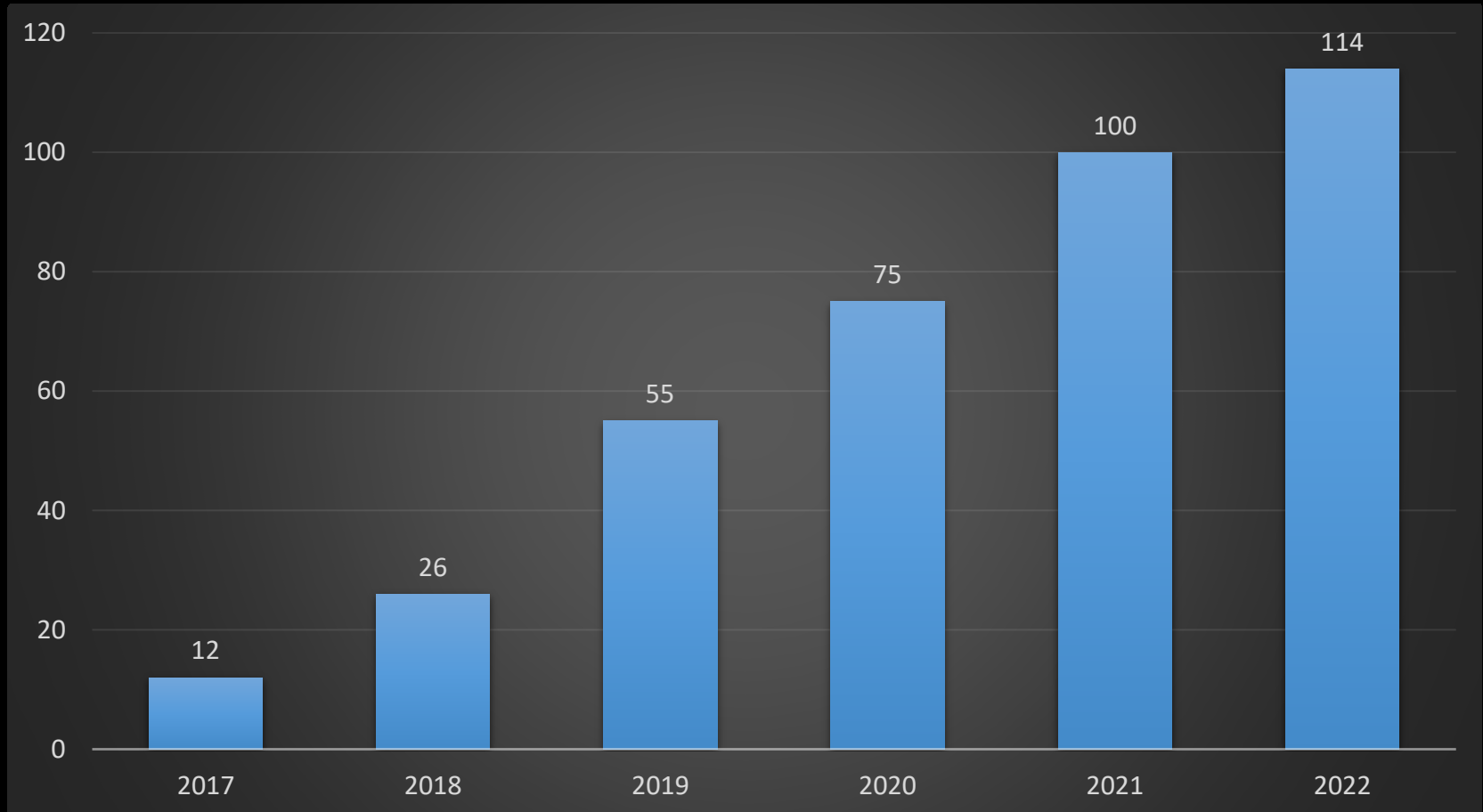
Pawel Pukowski

Nada Yehia

Benediktas Valys

Lawrence Schobs

3

Number of Students

Lectures and Labs

- Lecture: Thursday 10-11am @ Broad Lane Block (Mappin Building) LT 7

- Lab: Thursday 12-2pm @ Diamond Computer Room 3
  - **Finish required part of the labs in this slot!**

Other Interactions

- Online sessions
  - Lecturer blackboard collaborate office hour: Monday 1-2pm
  - Discord help sessions: Friday TBC one-hour slot in 11am-4pm

- BB discussion board: post your questions to get answered by the lecturer
  - One general forum: general question/feedback
  - Two lecture & lab forums: Haiping + Mauricio
    - To get help on lecture/lab contents
  - Two assignment forums: one for each part of the 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

- Lab exercises: 0% (self assessment)
  - Finish lab exercises by Tuesdays
  - Solutions to release on Wednesdays

- Assignment: 50%
  - Release: Part 1 = 17 March, Part 2 = 31 March 2022
  - Deadline: 15:00 on 05 May 2022 (end of lab)
  - Solution release: 12 May 2022
  - Marking and feedback deadline: 26 May 2022

- Final exam: 50%
  - To be scheduled in exam period (23 May – 10 June)
  - Formal exam on Blackboard: 2 hours

Contents: More Hands-on

Week	Date	Topic	Lecturer
1	10 Feb	Introduction to Spark and HPC	Haiping
2	17 Feb	RDD, DataFrame, ML pipeline, & parallelization	Haiping
3	24 Feb	Scalable decision trees	Mauricio
4	03 Mar	Scalable logistic regression	Mauricio
5	10 Mar	Scalable generalized linear models	Mauricio
6	17 Mar	Scalable neural networks	Mauricio
7	24 Mar	Scalable matrix fact. for collaborative filtering (RecSys)	Haiping
8	31 Mar	Scalable K-means clustering	Haiping
9	28 Apr	Scalable PCA for dimensionality reduction	Haiping
10	05 May	Apache Spark in the Cloud	Mike