Extreme Computing
Semester 1 (2023 - 2024)
Luo Mai, Amir Shaikhha
Course Timetable

Friday

13:10 - 14:00 and 14:10 - 15:00

AT_Lecture Theatre 3, Appleton Tower
Course Webpage

- The course webpage is hosted on Learn: http://course.inf.ed.ac.uk/exc

- “OR” simply Google: “Extreme computing Edinburgh 23/24”

- The webpage contains ALL the information
Course Staff

● **Luo Mai**
  ○ Assistant Professor (Lecturer), School of Informatics
  ○ Research interests: computer systems, machine learning, data management
  ○ Homepage: [https://luomai.github.io](https://luomai.github.io)

● **Amir Shaikhha**
  ○ Assistant Professor (Lecturer), School of Informatics
  ○ Research interests: programming languages, compilers, data management
  ○ Homepage: [https://amirsh.github.io](https://amirsh.github.io)
Guest lecturers

- Dr. Dimitrios Vytiniotis, Researcher, Google DeepMind
  - Date: 01/12/2023
  - Topic: TBD

- We may have another guest lecturer from GraphCore, UK
  - Date: TBD
  - Topic: TBD
Communication

● Piazza
  ○ https://piazza.com/class/lmgiwvg3vlj5ec

● Link is available on the course webpage
Course assessment

- Final exam: 100%
Preferred prerequisites

● Programming languages
  ○ Strong programming skills
  ○ Java/Scala/C++
  ○ Python/bash

● Courses
  ○ Operating systems
  ○ Machine learning
Questions?
Course Overview

Luo Mai
Big data

Internet services

Online news, blogs, forums
Content sharing (e.g., Instagram)
Recommendation systems
Social networks

Heterogeneous data

Text
Images
Events (Stream)
Graph

......
Challenges for processing big data

● **Increasing data scale**
  ○ Internet size: 10s millions of Terabytes (Estimated in 2021)
  ○ New Internet data per day: 100,000s Terabytes (Estimated in 2021)
  ○ 100 GBs - 10s PBs data per job (According to Google and Microsoft)

● **Increasing processing complexity**
  ○ Data aggregates (e.g., min, max, average)
  ○ Machine learning (e.g., clustering)
  ○ Deep learning (e.g., GPT-3)
Limitations of conventional data-centric systems

- **Databases**
  - Examples: Oracle DB, MySQL
  - SQL-oriented query workloads
  - Structured data
  - Expensive custom servers

- **High-performance computing systems**
  - Examples: Message-Passing-Interface (MPI)
  - Scientific workloads: physics simulation, weather prediction, …
  - Computation-oriented architecture

**Insufficient flexibility and performance**
Scaling-up and scaling-out

Adding accelerators (e.g., GPUs)  How to design expressive interfaces?

Adding machines (e.g., data centres, clouds)  How to achieve linear scalability?
Extreme computing course

- **Batch processing**
  - MapReduce, Spark

- **Stream processing**
  - Spark Streaming, Flink

- **Graph processing**
  - Pregel

- **Query processing**
  - SparkSQL, Pig, Hive

- **Linear algebra and ML**
  - MLlib

- **Deep learning frameworks**
  - TensorFlow, PyTorch

- **Distributed machine learning**
  - Data parallelism, pipeline parallelism, model parallelism

- **Practical machine learning systems**
  - Collective communication, parameter servers, data centre networks

- **Distributed file systems**
  - Google file systems

- **Cloud resource management**
  - Virtual machines, containers
  - Cluster resource scheduling (Kubernetes)
Approach

Extreme computing

Application programmers
System developers
Researchers
Questions?
Break