project market for project#2

you can propose your own thing as well!

goal

- your first project reproduced an existing study
- here, you have a blank sheet to design your own experimental method
  - system?
  - metrics?
  - workload?
  - …
- multiple groups can do the same project

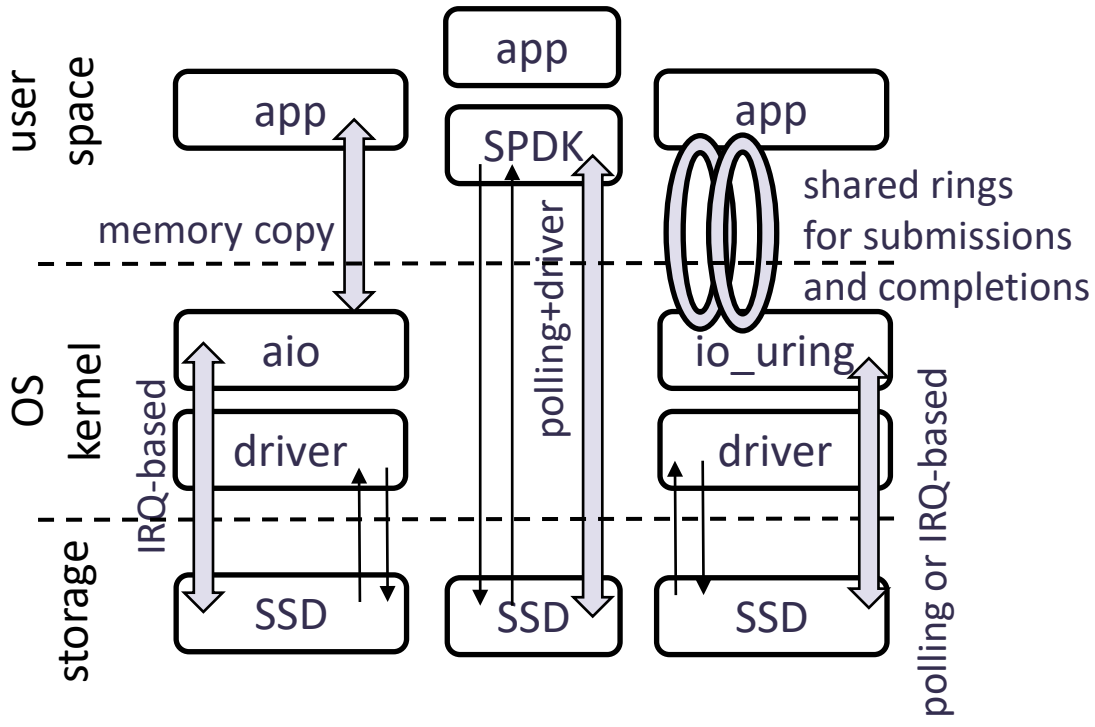
blank template

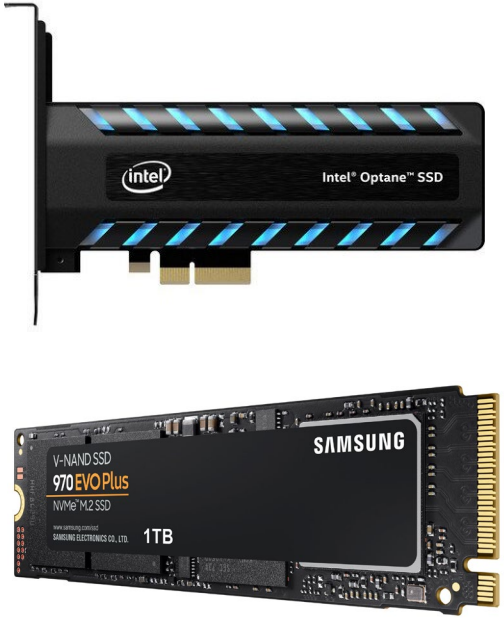
system comparison

- based on your interest
  (e.g., can explore an area for thesis),
  pick
  - a goal
  - a set of systems
  - a set of metrics
  - a set of parameters
  - …

storage

comparing different IO modes or SSDs





e.g., two different SSDs

e.g., io-uring vs aio
via xnvme vs not

using possibly fio or a data management system

processors

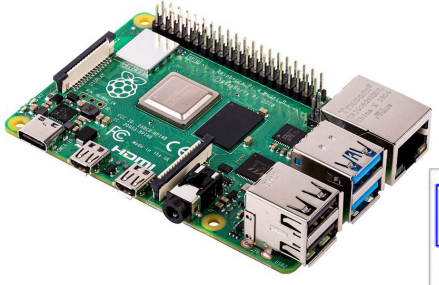
ITU HPC

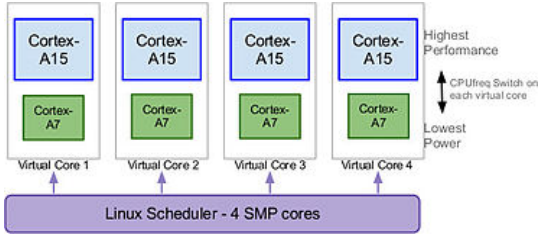
- CPU-GPU cluster built from state-of-the-art commodity server hardware (see hpc.itu.dk)
- running different benchmarks / use-cases of your choice over it to see how it behaves
- you can investigate the impact of different types of GPUs & CPUs or CPUs vs GPUs

low-power device characterization

in our lab, we have bunch of low-power devices

- different generations of raspberry pi
- nvidia jetson
- RISC-V board
- xilinx pynq





- design and conduct experiments to investigate device behavior with benchmark(s) and metric(s) of your choice
  - e.g., picking one device and analyzing it, or comparing two devices, etc.

disclaimer: not all may be available at the time of asking, there are several users in the lab

OS

impact of processor & memory affinity

designing & devising experiments to observe the impact of affinity

- hardware: class server, ITU HPC, one of the low-power devices …
- workload: your choice, but could be
  - operations on an in-memory optimized data intensive system
  - sorting
  - index operations
- metrics: throughput, cpu migrations, cache misses, etc.

applications

other

comparing

- data structures / indexes
- programming languages
- sorting algorithms
- …

similar to the partitioning comparison in project#1.

maybe add additional complexity

- e.g., what happens when things don't fit in memory
  when you need to spill intermediate results to disk

comparison of deep learning libraries

for training or inference

for data loading

workload: something from [MLCommons]

expect using ITU HPC as well

DBMS comparison

- pick a couple of systems
- identify which SQL queries or standardized benchmarks to use to compare them
- a regular performance test or
  a test to see if you can break the systems

big data systems comparison

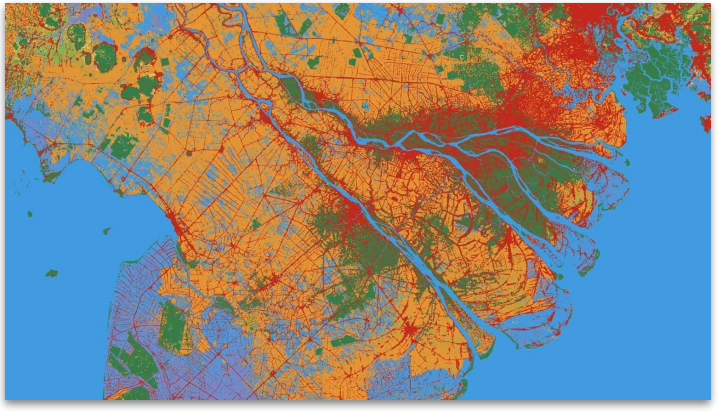
benchmarking Apache Wayang

- compare it to
  - Spark
  - Flink
  - Postgres

- benchmark / workload
  - big data benchmark

offered by our section members

data systems for land use analysis



- land use classifications based on satellite images are updated by Dynamic World every few days
- how can we quantify land use changes (such as deforestation, urbanization) using these maps at the scale of a country or a continent?
- the labelled data is given in raster format
  - Apache Sedona to store and analyze raster data
  - compare with an existent implementation on PostGIS with polygonal data
  - goal is to quantify the differences in efficiency and ease of use

contact: Maria Astefanoaei (msia@itu.dk)

PostGIS Apache Sedona

investigate the issue of an ML training task

throughput
= 15 samples/s on a personal desktop
= 3 samples/s on a big server

why?

investigate performance of an ML training pipeline on a personal laptop vs ITU HPC server

contact: Yucheng Lu (yucl@itu.dk)