Supporting and providing GPU resources and Machine Learning

Format

- 4 questions
- ~7min discussion on each

QUEEN'S UNIVERSITY BELFAST

1. GPU investment

- NVIDIA Tesla v100s are expensive (£45K for 4xV100 node)
- Researchers compare with their workstation GPU which they have immediate access to
- GPUs go out of date quickly – not great for capital spends
- GPU prices seem to be going up

⇒ Assuming a mix of workloads, what sort of split are institutions planning for, e.g. 80% cores, 20% GPU?

⇒ Are institutions buying upfront

QUEEN'S UNIVERSITY BELFAST

2. Adding ML to traditional HPC cluster

- Mixed workloads – batch and interactive jobs, GPU enabled applications and ML/DL.
- Does the Beowulf type architecture fit ML workloads?
- Do we need to be investing in new IO subsystems and memory to keep the GPUs busy?
- Do we understand our user workloads and software?

3. Many researchers are talking about ML but we get feeling they don't know how to start, i.e. how do we on-ramp this?

- NVIDIA training materials?
- Vendor workshops?
- Is this provided as central IS service training?

QUEEN'S UNIVERSITY BELFAST

4. ML – should we be wary of overhype?

- Reproducibility?
- Explanation?
- Obtaining quality datasets for training in some domains
- ML is not going to be practical for every problem!
- IS view is ML is just another tool, useful for some researchers?