UXL Public Test Infrastructure Update

Aaron Dron – Codeplay Software Ltd

January 2025

Community Runners for Public UXL Repositories

- GitHub self-hosted runners
  - Free for public repos, up to 20 concurrent, 4 threads, 16GB RAM
  - Ubuntu, Windows, Mac
  - **New public preview of ARM runners – investigating**
  - **Could GitHub sponsor a higher tier of membership for the UXL Foundation?**
- Intel Tiber AI Cloud hosted runners
  - 2 systems provided, each with 192 threads, 2TB RAM, 8x GPU Max 1550
  - Acting as Kubernetes cluster, using ARC for GitHub
- AWS hosted Graviton runner
  - On loan from Codeplay for the oneAPI Construction Kit
  - ARM community runner requested – **no progress, likely no longer necessary**
- Codeplay hosted Nvidia runners
  - In progress, TBC

Kubernetes Runners

- **Initial deployment is live (but should be considered unstable)**
- Documentation and best practices in progress
- Pods can be provisioned in different specifications (including multiple GPUs)
- Not limited to GPU testing
- Runners are ephemeral and isolated
- Caveats:
  - Linux only
  - Cannot run docker within CI workflows
  - **Uses pre-built runner images which can be customised in advance**
  - No “real world” testing done on Multi-GPU pods
  - Working on process for driver changes

Next Steps

- Looking for any UXL Members interested in providing test hardware
- Discussing potential for more systems from Intel Tiber AI Cloud
  - CPU only test system
  - Intel multi-GPU system
  - **Intel Battlemage GPUs**
- Discussing strategies for driver updates on IDC GPU Max test systems
- **We would like to ensure the minimum viable testing for PRs is available in UXL projects**
  - **What does this look like for each project?**
  - **What resources would be required to achieve it?**

Reach Out

- aaron@codeplay.com
- aaron.dron@intel.com
- UXL Slack