UXL Public Test Infrastructure Update

Aaron Dron – Codeplay Software Ltd

February 2025

Community Runners for Public UXL Repositories

- GitHub self-hosted runners
  - **UXLF GitHub Organisation is now on Enterprise Tier**
  - Free CI time for public repos, up to **500 concurrent (unproven)**
  - Public repos can test on **Linux (x64 + ARM), Windows, and macOS (x64 + ARM)**
- 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 – **likely deprecated**
  - ARM community runner requested
- Codeplay hosted runners
  - NVIDIA GPU and Intel GPU (Battlemage) systems in progress

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

- **Testing new GitHub Enterprise features**
- Looking for any UXL Members interested in providing test hardware
- Discussing potential for more systems from Intel
  - 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?
- **We are planning a proof of concept open-source release of oneAPI**

Reach Out

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