Meeting Minutes 2023-0523

Time (mins)	Agenda Items	Presented By	Note /Links/
30		MIGU	

MAR 230th, 2023 Attendance list: CMCC hanyu ding Hanyu DingArm TinaMIGU, Huawei @jianpeng heBUPT

MIGU introduce the serveral problems about GPU scheduling:

1GPUGPU

2GPUPODGPUGPU

3cuda/cudnn

4GPU

5k8swindowswindowslinux

1. GPU exclusive mode, low resource utilization. Whether there is a scheme to dynamically adjust the number of instances to release GPU and provide GPU reuse ability through intelligent scheduling;

2. After GPU virtualization, multiple PODs support sharing one GPU card. There is resource competition for services on the same physical card, because the peak and valley of app may overlap, resulting in conflict in the use of GPU computing force when sharing;

3. The versions of cuda/cudnn used for development of different application services are inconsistent, so there are adaptation problems;

4. The service cannot be dynamically adjusted by using GPU;

5. Does k8s support the solution of windows container arrangement? At present, most rendering engines used in rendering business are developed based on windows but not on linux.