Software Engineer II, GPU System, Google Cloud Platforms

Minimum qualifications:

- Bachelor's degree in Computer Science, a related technical field, or equivalent practical experience.
- 1 years of experience in system software development (e.g., kernel, device drivers, firmware, software that manages kernels, machines, or devices).
- 1 year of experience coding in C or C++ or Python.
- Experience with Linux or Unix development environments.

Preferred qualifications:

- Experience with developer operations, release management, and integration testing.
- Experience with designing and developing device drivers for peripherals (e.g., GPUs, PCIe Switches, and connectivity buses like I2C, USB, PCIe).
- Experience coding in Python or with scripting languages (e.g., shell).
- Experience with open source development.

About the job

Google's software engineers develop the next-generation technologies that change how billions of users connect, explore, and interact with information and one another. Our products need to handle information at massive scale, and extend well beyond web search. We're looking for engineers who bring fresh ideas from all areas, including information retrieval, distributed computing, large-scale system design, networking and data storage, security, artificial intelligence, natural language processing, UI design and mobile; the list goes on and is growing every day. As a software engineer, you will work on a specific project critical to Google’s needs with opportunities to switch teams and projects as you and our fast-paced business grow and evolve. We need our engineers to be versatile, display leadership qualities and be enthusiastic to take on new problems across the full-stack as we continue to push technology forward.

The ML, Systems, & Cloud AI (MSCA) organization at Google designs, implements, and manages the hardware, software, machine learning, and systems infrastructure for all Google services (Search, YouTube, etc.) and Google Cloud. Our end users are Googlers, Cloud customers and the billions of people who use Google services around the world.

We prioritize security, efficiency, and reliability across everything we do - from developing our latest TPUs to running a global network, while driving towards shaping the future of hyperscale computing. Our global impact spans software and hardware, including Google Cloud’s Vertex AI, the leading AI platform for bringing Gemini models to enterprise customers.

Responsibilities

- Develop, integrate, debug, and validate Data Center GPUs system software, resolve Data Center Graphics Processing Unit (GPU) machines issues.
- Integrate and validate GPU kernel drivers and firmware, enabling GPU Software bundle on the Data Center machines.
- Collaborate with Google Cloud cross-teams to enable Software (SW) and solve the issues, improve Data Center GPU machines reliability, stability, and repairability.
- Write specifications for software architecture and GPU systems.
- Develop test suites that enable unit, integration, and system level testing of system software.