Byte

https://www.byte.eco/job/22955/

Senior DevOps Engineer

Hiring organization
Together AI

Job Location
San Francisco, California, United States

Base Salary
$ 75000 - $ 120000

Date posted
May 13, 2024

Apply Now

Description

We are hiring a talented Senior DevOps Engineer to develop the software and processes for orchestration of AI workloads over large fleets of distributed GPU hardware. In this role, you'll be part of a cloud engineering organization that aims to automate everything and build failure-resistant and horizontally scalable cloud infrastructure for GPU-resident applications. As a Senior DevOps Engineer, you'll build deep understanding of Together AIâ??s services and use that knowledge to optimize and evolve our infrastructure's reliability, availability, serviceability, and profitability. The best applicants for this role are deeply technical, enthusiastic, great collaborators, and intrinsically motivated to deliver high quality infrastructure. You have experience practicing infrastructure-as-code, including the use of tools like Terraform and Ansible. You also have strong software development fundamentals, systems knowledge, troubleshooting abilities, and a deep sense of responsibility. Requirements Minimum of 5 years of prior relevant experience in DevOps, cloud computing, data center operations, SRE, and Linux systems administration Experience in programming in at least one of the following languages: Java, Python, Go, C++ Experience designing and building advanced CI/CD pipeline frameworks Experience with cloud computing toolsets like Terraform, Vault, and Packer Experience with configuration management tools like Ansible, Pulumi, Chef and Puppet Experience with Kubernetes, containerization and VPNs Strong sense of ownership and desire to build great tools for others Self-driven and motivated, with a strong work ethic and a passion for problem solving Experience with AI workloads and blockchain based protocols a plus GPU programming, NCCL, CUDA knowledge a plus Experience with Pytorch or Tensorflow a plus Responsibilities Create a highly automated infrastructure pipeline for deploying and scaling distributed and multi-tenant GPU-resident compute to new cloud and data center environments Create infrastructure to auto-scale AI models, create training clusters, and wrestle with CUDA dependencies Introduce tools to facilitate greater automation and operability of services Design, build, and maintain CI/CD infrastructure Architect, deploy, and scale observability infrastructure Participate in on-call rotation and ensure uptime of services Investigate production issues and help prevent their reoccurrence Create runtime tools/processes that optimize cloud triaging and limit downtime Define best practices to make our systems and services measurable Work closely with internal teams to ensure best practices are appropriately applied Build tools to help engineering and research teams measure and improve their velocity Analyze and decompose complex software systems Collaborate with and influence others to improve the overall design About Together AI Together AI is a research-driven artificial intelligence company. We believe open and transparent AI systems will drive innovation and create the best outcomes for society, and together we are on a mission to significantly lower the cost of modern AI systems by co-designing software, hardware, algorithms, and models. We have contributed to leading open-source research, models, and datasets to advance the frontier of AI, and our team has been behind technological advancement such as FlashAttention, Hyena, FlexGen, and RedPajama. We invite you to join a passionate group of researchers in our journey in building the next generation AI infrastructure. Compensation We offer competitive compensation, startup equity, health insurance and other competitive benefits. The US base salary range for this full-time position is: $160,000 – $230,000 + equity + benefits. Our salary ranges are determined by location, level and role. Individual compensation will be determined by experience, skills, and job-related knowledge. Equal Opportunity Together AI is an Equal Opportunity Employer and is proud to offer equal employment opportunity

to everyone regardless of race, color, ancestry, religion, sex, national origin, sexual orientation, age, citizenship, marital status, disability, gender identity, veteran status, and more. Please see our privacy policy at https://www.together.ai/privacy Please mention the word **GLORIOUS** and tag RMzQuODYuMTYzLjE1Mg== when applying to show you read the job post completely (#RMzQuODYuMTYzLjE1Mg==). This is a beta feature to avoid spam applicants. Companies can search these words to find applicants that read this and see they're human.

Contacts
Job listing via RemoteOK.com