Robin Thoni

📱 +33 6 34 39 52 26
📱 +1 (424) 242-5068
✉ robin@rthoni.com
⌥ git.rthoni.com/robin.thoni
⌥ robin-thoni

Infrastructure and Automation lover.

Cursus

2016 / 2017
Computer Science Master, *UQAC*, 🇨🇦, Canadian University

2012 / 2017
Master of Engineering, *EPITA*, 🇫🇷, French CS Engineering School

Experiences

2024
Senior Infrastructure Engineer, *FlexAI*, 🇫🇷, PaaS/Infrastructure team
- Designed and implemented from scratch a new platform to train and host AI applications, based on Kubernetes, FluxCD, Helm, GitHub Actions, etc
- Designed and implemented multi-cloud IaC with Pulumi and Ansible (AWS, Azure, Intel, Cisco, Oracle)
- SOC2/GDPR compliance: implemented security measures, automated security audits, and created documentation (Hashicorp Vault, etc)
- Setup and maintained low-level infrastructure: bare-metal, VMs, networking, storage, NVIDIA/Mellanox Infiniband stack, etc
- Worked on custom Kubernetes operators to run AI training and inference workloads

2022 / 2023
Infrastructure Engineer Consultant, *iKe*, 🇱🇺, BGL / BNP Paribas Group
- Migrated manually deployed server applications to CI/CD with Jenkins, Docker, Kubernetes, Kustomize, CA Release Automation, Ansible, AWX, Rancher
- Created a new CI/CD pipeline to deploy Helm charts in IBM Kubernetes Services (IKS)
- Packaging and automation of the deployment of a full Core Banking software in Kubernetes/Helm to replace the existing mainframe

2020 / 2021
Infrastructure Engineer, *NVIDIA*, 🇬🇧 / 🇩🇪, CUDA devtools team
- Maintained devtools' build infrastructure: upgraded >10 years old toolchains, improved existing codebase to fix new compiler warnings/errors, created container-based build environment, added new and improved existing automations with Jenkins, etc

2018 / 2020
Software Engineer, *NVIDIA*, 🇬🇧, CUDA devtools team
- Led the design and implementation of a GPU code coverage tool: generate code coverage reports for kernels written in assembly for internal automotive usage
- Led the design and implementation of a GPU thread-array diversity measurement tool: generate reports on how thread-arrays are dispatched on physical GPU cores to detect hardware errors for internal automotive usage

2016
Infrastructure Engineer, *Betaclean*, 🇫🇷, Family Business
- Setup physical network (ISP / routers / ethernet / Wi-Fi)
- Development of internal intranet web app
- Deployment/management of security systems (Bitwarden / Keycloak / CA / firewall)
- Automation of infrastructure deployment (Saltstack / Kubernetes / Helm)
- Deployment/management of E-mail server (Postfix / Dovecot)
- Deployment/management of DNS authoritative server (internal / external) and DHCP server
- Deployment/management of VPN server (OpenVPN / Wireguard)
- Deployment/management of physical security systems (CCTV / access control / alarm)

2017
Final Internship, *NVIDIA*, 🇺🇸, CUDA devtools team
- Added 'initcheck' support for cuda-memcheck: devtool to check for read-before-write on the GPU
- Optimized Nsight Compute kernel replay feature: restore memory on the GPU only if it has been modified by the profiled kernel

Skills

Languages C; C++; Python; C#; Go; Java; Shell/Bash; PostgreSQL

Frameworks	Android SDK; Qt; ASP.NET Core
Infrastructure	Kubernetes; Docker/docker-compose; E-mail servers; DNS servers, Keycloak; Bitwarden; Iptables; Zabbix; Step-ca; Hashicorp Vault; Linux Debian/RHEL
CI/CD	Jenkins; SaltStack; Ansible; Helm; Kustomize; Pulumi; Terraform

Miscellaneous

Languages

French	Native.
English	Fluent / TOEIC 910.