Arham Ahmed

(+1) 425-588-8005 | arham.b.ahmed@gmail.com | linkedin.com/in/arhamahmed | arhamahmed.github.io

Education

University of Illinois Urbana-Champaign — Champaign, IL
Master of Computer Science — *May 2021 – May 2023*

University of Waterloo — Waterloo, ON
Bachelor of Applied Science in Honors Computer Engineering — *Sep. 2014 – Apr. 2019*

Technical Skills

Languages: Go, Python, Java, C++, TypeScript, SQL, C#, Rust, BASH, Ruby
Frameworks: Kubernetes, Keda, Istio, Envoy, Jaeger, Prometheus, FluentD, gRPC
Developer Tools: Git, Docker, Helm, Azure, Google Cloud Platform, VS Code, iTerm, UNIX

Experience

Microsoft — Sep. 2019 – Present
Software Engineer 2 — *Redmond, WA*

- Constructed a distributed cluster management tool to streamline ad-hoc scripting/tuning of Kubernetes nodes from a manual O(N) operation to O(1)
- Devised strategy to inject fault tolerance, autoscaling, circuit breakers and other best practices into microservices with helm, improving service uptime by 70%
- Spearheaded the architecture, design, and rollout of the release management system of the new IoT Operations product involving automated gating, branching, and exposing of related components used by 8+ teams
- Lead team of 3 engineers in automation of observability infrastructure with 15x speedup in deployment time
- Leveraged legacy product and crafted extensible/modern APIs to provision, configure secrets, and deploy Kubernetes clusters accelerating partner teams' product launch speed by 9x on average
- Added distributed tracing to 70+ microservices to ease on-call debugging via the Istio service mesh and Envoy
- Designed a system to autoscale K8S pods on resource/external metrics, cutting total operational costs by ∼50%

Amazon — Sep. 2018 – Dec. 2018
Software Engineering Intern — *Palo Alto, CA*

- Crafted a novel genetic algorithm-based SAT solver in Java to solve the NP-hard problem of allocating variable-length ads to videos with multiple constraints in near real-time
- Modeled the ad allocation problem mathematically allowing the use of constraint solvers to find optimal solutions
- Built a stochastic greedy algorithm to allocate video ads resulting in an 8x speedup with near-optimal performance

Microsoft — Jan. 2018 – Apr. 2018
Software Engineering Intern — *Redmond, WA*

- Designed and implemented a sharding strategy for Azure Stream Analytics to optimally partition input events
- Reduced memory usage of sharded output files by a factor of O(N) with a heavy file caching strategy
- Increased write throughput to Blob Storage by 13 % by migrating all IO operations to asynchronous tasks

Google — May 2017 – Aug. 2017
Software Engineering Intern — *Kirkland, WA*

- Designed and implemented a private library hosting service due to high customer demand running on App Engine and Container Engine with a multi-tier architecture saving 25% in operational costs
- Implemented the service with 99.95% uptime and tooling to manage its deployment to Google Cloud Platform
- Reduced service CPU usage by 2.5x with asynchronous file backups triggered by whitelisted requests
- Added OAuth2, GCP-compatible storage/database layer, and server orchestration for load balancing on top of an existing library leveraged for hosting functionality

Projects

Kura | *Go, gRPC, protobuf, GCP, Linux FUSE* — git.io/vpiyJ

- Devised an end-to-end encrypted, synchronized, and distributed file system to host/share personal media
- Formulized a secure file sharing protocol modeled after WhatsApp's proven Signal protocol