

SULTAN MAHMUD

Backend Engineer

Dhaka, Bangladesh . +880 1683152495

kaziirad@gmail.com . [LinkedIn](#) . [Github](#) . [Medium](#)

PROFESSIONAL SUMMARY

Backend Engineer delivering **production-grade distributed systems** with measurable impact: **sub-10ms redirects** at 95% traffic load, **60% infrastructure cost reduction** through intelligent **autoscaling**, and **53% database overhead optimization**. Built **ML-enhanced Kubernetes autoscaling** with 4-layer architecture (time-aware, flash sale detection, Prophet forecasting), **microservices** with independent service scaling, and automated **AWS infrastructure (Pulumi, Ansible)**. Deep expertise in **observability (OpenTelemetry, Tempo)**, **event-driven Lambda systems**, and **fault-tolerant design (circuit breakers, distributed locking, WAL)**. Published technical blogs and educational content impacting 200K+ readers.

EXPERIENCE

BACKEND ENGINEER (Freelance/Contract) | Remote

August 2024 - Present

Providing architecture and development expertise for scalable backend systems and cloud-native applications.

- Architected and delivered 5 production systems across **e-commerce** (2 projects), **FinTech** (1), and **SaaS platforms** (2), managing infrastructure serving 5,000+ combined users with 99.9% uptime using **Python (FastAPI, Django)** and **AWS**.
- Engineered high-concurrency API systems sustaining **1,000+ concurrent** users with **sub-10ms API latency** through Redis caching (**95% hit rate**), **async Python** patterns, and **connection pooling optimization**.
- Implemented **fault-tolerant systems** through strategic implementation of **Redis clustering**, **database replication**, and **automated failover mechanisms**, ensuring **high availability** and **data resilience**.
- Owned full SDLC from architecture design to production deployment: automated infrastructure provisioning of **11+ EC2** instances across **multi-AZ VPC** using **Pulumi (IaC)** and **Ansible** configuration management, reducing deployment time from 4 hours to 15 minutes with zero-downtime releases via **GitHub Actions CI/CD**.

BACKEND DEVELOPER | Cooking Station, Dhaka

June 2024 - August 2024

- Streamlined operations for **200+ users** by designing **role-based admin dashboard** with **real-time meal analytics**, reducing manual subscription **management overhead**
- Eliminated 40% of manual effort** in account management through **automated balance updates** and **error-prevention systems**
- Built production-ready meal scheduling system using **cron jobs** with configurable time boundaries and daily user listing functionality

NOTABLE PROJECTS

ULTRA-FAST URL SHORTENER MICROSERVICE | FastAPI Redis PostgreSQL K3s AWS Ansible OpenTelemetry

Challenge: Design distributed URL shortener with independent service scaling and sub-10ms redirects

Impact: Reduced database load by 53% through connection pooling, achieved sub-10ms redirect latency with cache-first architecture

Technical Achievements:

- Architected** decoupled microservices: **create_service** (write-heavy), **redirect_service** (**95% read traffic**), **worker_service** with independent **scaling**, **Redis-first** caching with **MongoDB** fallback and **Nginx** routing targeting **sub-10ms redirects**
- Implemented** intelligent key **pre-population** using **Celery workers** maintaining a pool of unused **short URL keys** for **instant URL creation** without database latency
- Built repository pattern** with abstract base classes for **PostgreSQL/MongoDB**, **optimized key allocation** using **SELECT FOR UPDATE SKIP LOCKED** for **race-free operations** and parameterized **bulk inserts (100K+ keys)**
- Engineered production resilience:** **PgBouncer connection pooling** achieving **53% overhead reduction** with **exponential backoff retries** and database **timeout protection**
- Implemented circuit breaker pattern** preventing **cascade failures** with configurable thresholds, **automatic degradation**, and **observability** via **OpenTelemetry (B3 propagation)**, **Tempo** tracing, and **Grafana** visualization
- Built testing** with **multi-database mocking**, **async pytest**, and **deployed on K3s/AWS** via **Pulumi IaC**, **Ansible**, and **CI/CD**

Key Innovation: Cache-first architecture with **repository pattern** and **multi-database testing** enabling independent service scaling

Challenge: Build production-grade intelligent autoscaler with ML-based predictive scaling for K3s clusters on AWS

Impact: Event-driven Lambda architecture with 4-layer intelligent scaling achieving 60% cost reduction (\$200→\$111-\$149/month), multi-AZ high availability, and proactive pre-scaling before traffic spikes

Technical Achievements:

- Architected **4-layer intelligent autoscaling**: (1) **Data Collection**, (2) **Time-Aware Scaling** (peak/off-peak: 85%/60% vs 60%/40%), (3) **Flash Sale Detection** (>30% spike in 2min), (4) **Prophet ML forecasting CPU 15min ahead**
- **Built ML training pipeline** with **Kubernetes CronJob** for weekly automated retraining, **feature engineering** (cyclical encoding, lag features), **cross-validation**, and **backtesting**
- **Implemented distributed locking** via **DynamoDB conditional writes**, **Write-Ahead Log (WAL)** for crash recovery, and **EventBridge orchestration** for fault-tolerant scaling
- Engineered **multi-AZ high availability** with round-robin distribution across 3 AZs, **LIFO scale-down** with **kubectl drain** via **SSM**, and single **NAT Gateway** optimization
- Deployed comprehensive **observability**: 17 **CloudWatch alarms** (CRITICAL/WARNING), **Prometheus graceful degradation**, fixed **LogGroups**, **DLQ monitoring**, and **SNS alerting**
- Built **spot instance support** with automatic **On-Demand fallback** handling 3 **AWS capacity exceptions** for **70% cost savings**

Key Innovation: **Hybrid reactive+proactive scaling** combining **time-aware thresholds**, **flash sale detection**, and **ML forecasting** for **SLA-driven pre-scaling**

TECHNICAL SKILLS

- **LANGUAGE & FRAMEWORKS:** Python, Golang, JavaScript, **FastAPI**, Django, DjangoREST
 - **CLOUD & INFRASTRUCTURE:** AWS(EC2, S3, VPC, Lambda, EventBridge, CloudWatch, SNS, SQS), **Docker**, Kubernetes(K3s), Pulumi, Ansible
 - **DEVOPS & CI/CD:** GitHub Actions, Docker Compose, Nginx, Load Balancing
 - **ARCHITECTURE & DESIGN:** Microservices, Distributed Systems, High Availability, Fault Tolerance, Load Balancing, Caching Strategies, Database Optimization
 - **DATABASE & CACHING:** PostgreSQL, Redis, MongoDB, SQLAlchemy, PgBouncer (Connection Pooling), Alembic (Migrations)
 - **API DEVELOPMENT:** Swagger/OpenAPI, RESTful API, Authentication (JWT), Rate Limiting, CORS
 - **MESSAGE QUEUING & ASYNC:** Celery(Worker, Beat, Flower), Redis, RabbitMQ, Async/Await
 - **OBSERVABILITY & MONITORING:** OpenTelemetry (Distributed Tracing), Prometheus, Grafana, Tempo
-

EDUCATION

BACHELOR OF SCIENCE IN COMPUTER SCIENCE & ENGINEERING

Daffodil International University, Dhaka

September 2017 - December 2022

EXTRACURRICULAR ACTIVITIES & COMMUNITY ENGAGEMENT

- **Quora Bengali:** Maintain A Dedicated Blog On **Quora** Offering Insights And Solutions To Inquiries Spanning Technology, Culinary Arts, And Health In Bengali. Accumulated **Readership** Nearing **200,000** With Nearly **200 Followers**.
 - **Medium Technical Blog:** Published 2 in-depth technical articles (1,000+ words each) discussing distributed systems, backend architecture & engineering decisions and trade-offs
 - **YouTube Channel:** Founded And Managed The Youtube Channel "**I.T. Darshonik**" Producing Over **40 Instructional Videos** In Bengali, Aimed At Bridging The Language Gap In Tech Education.
-

AWARDS & ACHIEVEMENTS

- **DIU Take-Off Programming Contest, Spring 2018:** Ranked **6th/300+ participants** (Top 2%)
- **Competitive Programming:** Solved 500+ problems across BeeCrowd, LightOJ, HackerRank, LeetCode.
- **Proficiency:** Dynamic Programming, Graph Theory, Greedy Algorithms, Binary Search