Le Xuan Cuong

linkedin.com/in/lxcuong | github.com/lexuancuong

Email : cuongle.software@gmail.com

Mobile : +84 77 888 7894

Summary

Senior Backend Engineer with 6+ years building highly-available, scalable cloud systems processing millions of daily transactions. Proven expertise in system architecture design (10+ Technical Designs approved), microservices, and real-time data pipelines. Passionate about mentoring engineers and collaborating cross-functionally with Product and Design to deliver customer-centric solutions. AWS Certified Solutions Architect.

Education

- **HCM University of Science** — Ho Chi Minh City, Vietnam
  Master of Science in Computer Science — *Sept. 2021 – Sept. 2025*
- **HCM University of Science** — Ho Chi Minh City, Vietnam
  Honor Program - Bachelor of Science in Information Technology; GPA: 8.4/10 — *Sept. 2017 – Sept. 2021*
- **Coder School - Data Science Course** — Ho Chi Minh City, Vietnam
  Machine Learning course sponsored by Facebook (Certificated) — *Aug. 2019 – Dec. 2019*

Experience

- **Parcel Perform** — Ho Chi Minh City, Vietnam
  Senior Backend Engineer — *Aug 2024 - Present*
  - **System Architecture & Design**: Led architecture design reviews for 10+ Technical Designs spanning ML systems, analytics platforms, and real-time pipelines. Authored cross-team TD for ML-based carrier identification across 7 microservices, reviewed by CTO.
  - **High-Availability Cloud Systems**: Built mission-critical Reports & Analytics platform (Jaspersoft/ClickHouse) and Alert Management system using DDD architecture with Dagster orchestration. Integrated Grafana/Sentry for 24/7 monitoring and incident response.
  - **Real-Time Data Pipelines**: Architected ML-based carrier identification pipeline spanning 7 microservices (Flink CoreJobs, Scraper, DataIO, Scheduler), processing millions of daily events. Designed seamless integration with legacy regex flow using feature toggles and Kafka protobuf schemas, improving parcel found rate while preventing recursive scrape loops.
  - **AI/ML Solutions**: Architected L3 AI agents using LangChain/LangGraph with ReAct pattern and AWS Bedrock. Built tool-based architecture for Root Cause Analysis and automated insights generation with LangSmith monitoring.
  - **Engineering Excellence**: Pioneered AI-assisted Technical Design process (first squad at company). Sole BE engineer delivering 105+ tickets with zero missed deadlines. Introduced Django Clickhouse ORM and Django Ninja framework.
  - **Mentorship & Team Growth**: Mentored 2 engineers through architecture discussions, code reviews, and pair programming. Delivered 15+ Open Space sessions on system design, AI/LangChain, and Flink to upskill engineering org.
- **Parcel Perform** — Ho Chi Minh City, Vietnam
  Backend Engineer — *Jul 2022 - Aug 2024*
  - **Scalable API Development**: Architected Public Booking API with OpenAPI spec serving multi-carrier integrations (DHL, MyParcel, Shippo). Implemented S3 pre-signed URLs and Kafka async processing for high-throughput operations.
  - **Extensible Integration Architecture**: Designed base booking integration abstraction supporting 50+ carrier integrations. Gathered requirements through practical usage testing and Product collaboration to build flexible, extensible foundation.
  - **High-Volume Data Pipelines**: Built and maintained data pipeline processing 5M+ daily events across scheduler, scraper, data I/O, and Apache Flink services. Ensured sub-second latency for mission-critical shipment tracking.
  - **Cross-Functional Collaboration**: Authored 4+ Technical Designs collaborating with Product and Design through HLD meetings. Translated customer requirements into scalable, maintainable architectures.

- **Email Extraction Platform**: Designed pattern-based email extraction engine (Regex/XPath/custom formatters) with Kafka CDC pipeline. System running 3+ years in production processing shipment notification emails.
- **Architectural Decision-Making**: Evaluated 4 architectural approaches for email processing pipeline using structured criteria (reliability, scalability, maintainability). Selected app-based CDC over cronjob/Flink alternatives, documenting trade-offs and accepted drawbacks for transparent technical decisions.
- **Sync-to-Async Reliability Pattern**: Designed fault-tolerant Booking API spanning 5+ microservices with sync-to-async fallback pattern. Implemented Kafka-based async processing as fallback when synchronous calls timeout, ensuring booking completion despite component failures.
- **Celery Pipeline Architecture**: Built Collection Point data pipeline using Celery canvas patterns (chain, chord, group) with optimized worker configuration for scalability. Designed simple, maintainable task breakdown (8K+ postal codes chunked into parallel sub-tasks). Evaluated 3 scheduling approaches; selected Celery Beat for auto-retry and fault tolerance.
- **System Reliability**: Resolved critical production incidents: BFCM traffic spikes (10x load), Redis cluster hotspots, and emergency data re-imports (1.6M shipments in 2 days). Established incident response protocols.
- **Open Source & Innovation**: Introduced Django Ninja framework (adopted company-wide), TDD practices, and coverage tooling. Proposed RabbitMQ migration to resolve Redis memory bottlenecks.

- **Cinnamon AI Labs** — Ho Chi Minh City, Vietnam
  Software Engineer — *Aug 2020 - Jul 2022*
  - **Enterprise Client Delivery**: Delivered production systems for Fortune 500 clients (Deloitte, Bridgestone, Daikin, Sumitomo) across document processing and AI automation domains.
  - **Multi-Framework Expertise**: Adapted rapidly across web frameworks (Flask, FastAPI, Django) based on client requirements. Designed RESTful APIs using Repository-Services pattern and SOLID principles.
  - **Multi-Cloud Infrastructure**: Deployed and maintained services across AWS, Azure, and GCP based on enterprise client preferences. Containerized applications with Docker for consistent cross-cloud deployments.
  - **Cross-Team Collaboration**: Partnered with AI researchers and product managers to translate ML models into production-ready solutions meeting client SLAs.

Projects

- **Template/Client Projects (Cinnamon AI Labs)**: Created RESTful API template reused across client projects. Applied Repository-Services Pattern, SOLID principles. Implemented heavy background tasks with Celery and RabbitMQ.
- **Hasiti Application (Final Thesis)**: Deep Learning application to simulate hairstyles. Researched and trained CNN networks for hairstyle transfer. Designed and implemented server architecture for AI model serving.

Certifications

- **AWS Certified Solutions Architect - Associate** — Amazon Web Services
  Valid: Jan 2025 – Jan 2028

Achievements

- **Emotion Recognition 2019**: Hosted by Vintech. Ranked 9th out of 54 teams.
- **Faculty of Information Technology 2018**: Top 10 students with highest ranking. GPA 9.1/10.

Technical Skills

- **Languages**: Python, Scala, Java, JavaScript, SQL
- **Databases & Data Stores**: PostgreSQL, Redis, ClickHouse
- **Web Frameworks**: Django, Django Ninja, FastAPI
- **Big Data & Pipeline**: Apache Flink, Kafka, Celery (canvas: chain/chord/group), Dagster, RabbitMQ
- **Infrastructure**: AWS (EC2, S3, Bedrock), Docker, Grafana, Sentry, CI/CD
- **AI/ML**: LangChain, LangGraph, LangSmith, AI Agents, Prompt Engineering
- **Architecture**: High-Availability Systems, Microservices, DDD, System Design, RESTful APIs, TDD

Languages

- **English**: B2 Upper Intermediate – Aptis ESOL (British Council) 176/200 **Vietnamese**: Native