Data Engineer / Senior Backend Engineer

Genggeng, Yuan

📧 kidyuan@foxmail.com | 📱 070-4017-5377

📍 Tokyo, Japan

Visa: Engineer / Specialist in Humanities / International Services (3-year term)

Showcase: Olist Lakehouse Pipeline | perapera.me | PeraTube

PROFESSIONAL SUMMARY

Data Engineer & Senior Backend Engineer with **6+ years** of specialized experience in large-scale distributed systems, high-concurrency database migrations, and cloud-native architecture, backed by a 20-year software engineering foundation. Actively focused on modern **Lakehouse Data Engineering**, proficient in core technologies including **Databricks, PySpark, and Delta Lake (Medallion Architecture, Change Data Feed, Lakeflow Declarative Pipelines, Unity Catalog, CDC & SCD Type 2)**. Experienced in designing and implementing multi-layer incremental data pipelines and **classic Kimball Dimensional Modeling** (Transaction, Accumulating Snapshot, and Periodic Snapshot facts). Highly proficient in the Spring ecosystem, NoSQL (Cassandra, Couchbase), and AI-augmented engineering workflows.

TECHNICAL SKILLS

Core Competencies

- **Data Engineering & Lakehouse**: **Databricks**, **PySpark**, **Delta Lake** (ACID, MERGE INTO, Time Travel, **Change Data Feed / CDF**, Liquid Clustering), **Auto Loader**, **Databricks Asset Bundles (DABs)**, **Lakeflow Declarative Pipelines (DLT / SDP)**, **Unity Catalog**, Medallion Architecture (Bronze/Silver/Gold/Mart), **Kimball Dimensional Modeling** (Transaction Fact, Accumulating Snapshot Fact, Periodic Snapshot Fact), **SCD Type 2**, **CDC**, Window Deduplication, Data Contracts & Quarantine Gates, SQL
- **Distributed Backend & Data Systems**: Java, Kotlin, Python (FastAPI), **Spring Boot (2.x/3.x)**, **Kafka (CDC)**, **Cassandra**, Couchbase, Snowflake, Oracle, Redis, RabbitMQ, Microservices, DDD
- **Search & Real-Time Sync**: **Meilisearch**, Convex (Reactive WebSocket Push), Elasticsearch/Vector Search, RESTful APIs
- **Cloud & DevOps**: **Azure (Static Web Apps, AKS)**, **Railway**, AWS, Docker, Minikube, Podman, CI/CD (GitHub Actions, GitLab CI/CD)
- **AI-Augmented Engineering**: Proficient in **Claude Code, Codex, and Antigravity** for rapid development, code generation, and pipeline refactoring

Languages & Frameworks

- **Languages**: Python, Java, Kotlin, SQL, TypeScript, C#, Go, Rust, C++
- **Web & Frontend**: Next.js, React.js, Vue.js, Convex, Tailwind CSS
- **Languages (Human)**: **Japanese (Business)**, **English (Business)**, Chinese (Native)

SELECTED PROJECTS

Olist Lakehouse Pipeline — Production-Grade Incremental Lakehouse on Databricks

Data Engineering Showcase | Architect & Data Engineer

Repository: https://github.com/kidwon/olist-lakehouse-pipeline

(v0.1.0 Foundation → v0.5.0 Production-Reconciled)

- **End-to-End Medallion Lakehouse**: Built a production-grade incremental lakehouse on Databricks processing 100k+ real e-commerce orders, leveraging **Auto Loader** with schema evolution (`_rescued_data`) and **forward-only CDC MERGE** to resolve out-of-order state updates and duplicate deliveries.

- **Complete Kimball Dimensional Modeling**: Designed a comprehensive Star Schema encompassing **all three classic fact table types**—Transaction (`fact_order_item`), Accumulating Snapshot (`fact_order_fulfillment` for multi-stage lifecycle & lateness attribution), and Periodic Snapshot (`fact_daily_order_backlog` for semi-additive inventory backlog trends).
- **Incremental Gold via Delta Lake CDF**: Implemented an incremental propagation engine powered by **Delta Lake Change Data Feed (CDF)** and high-watermark tracking, eliminating full historical recalculations and reducing daily processing to only **~1% of impacted orders** with 100% consistency parity.
- **Dual-Paradigm Architecture (PySpark vs. Lakeflow DLT)**: Re-implemented the entire pipeline using Databricks' next-gen **Lakeflow Declarative Pipelines (DLT)** alongside the imperative PySpark engine, reconciling 100% data parity across 82k+ orders while evaluating trade-offs in operational control and schema governance.
- **Quality Gates & Enterprise Engineering**: Enforced per-rule automated quarantine and a **5% threshold execution circuit breaker**; established **64 unit/E2E regression tests**, 12 Architectural Decision Records (ADRs), and CI-validated interactive notebooks.

perapera.me — High-Performance Japanese Learning Ecosystem

Personal Project | Lead Full-Stack Architect

- **Architecture**: Built a hybrid multi-service platform using **Next.js/Convex** and **FastAPI**.
- **Search Engine**: Integrated **Meilisearch** for high-performance search across millions of records.
- **Resilient AI**: Implemented a multi-tier fallback architecture (Cache/Key Pools/DeepSeek) to handle strict LLM rate limits.
- **Cloud & Payments**: Orchestrated CI/CD pipelines via GitHub Actions to **Azure** and **Railway**, and integrated **Stripe** for subscriptions.

PeraTube — Japanese Video Example-Sentence Search API (Developer SaaS)

Personal Project | Solo Full-Stack Architect

- **Product**: Indexes massive YouTube subtitle corpora into searchable segments — developers query how Japanese is actually used in real videos via API, receiving video ID, timestamp, transcript, and surrounding context with millisecond latency.
- **Architecture**: **Next.js** dashboard/playground + **Python FastAPI** orchestration layer + **Convex** as the single source of truth (users, quotas, API keys) + **Meilisearch** full-text search.
- **Commercialization**: **Stripe** subscriptions and credit packs, API-key authentication (sha256-hashed, LRU-cached validation), atomic quota deduction in Convex, admin dashboard, and content moderation.
- **Data Pipeline & Quality**: YouTube → VTT subtitles → segment parsing → chunked `.jsonl` → batch import; 15+ pytest test files across the API and pipeline; deployed on **Railway** with a Cloudflare custom domain.

Jessie Signal — Real-Time Market Sentiment Dashboard

Personal Project | Full-Stack Architect

- Built a **React / Convex** serverless web app that aggregates live market data (**VIX, CNN Fear & Greed, major ETFs**) and classifies the market into actionable sentiment scenarios.
- Implemented **user authentication (Clerk)** so users can log in and **subscribe to email alerts** on market-scenario changes, delivered via **Resend**.
- Added a **server-side caching** layer for instant page loads and a gated **admin dashboard**; deployed on **Railway**.

PROFESSIONAL EXPERIENCE

Contract Backend Engineer | NTT DATA

April 2025 - August 2026 | Tokyo, Japan

Insurance System Modernization Project (Struts to Spring Migration)

Responsibilities:

- Contributed to the full-stack modernization of a core insurance application across two major phases, transitioning from legacy frameworks to a modern web application stack:
  - **Legacy Source Stack**: TERASOLUNA 2.2.0 (Struts 1.x) + Oracle 11g + intra-mart 7.2 + JDK 1.7
  - **Target Modernized Stack**: TERASOLUNA 5.x (Spring MVC) + Oracle 18c XE + intra-mart 8.0.36 + JDK 17 *(Architecture Migration, Framework Upgrade, Spring Ecosystem)*
- **Phase 1 (UI/UX Optimization)**: Refactored legacy **JSP** templates and adjusted intra-mart custom tags to improve front-end usability, while maintaining existing Struts backend filters and Java logic. *(JSP Refactoring, intra-mart Ecosystem, Struts 1.x)*
- **Phase 2 (Architectural Upgrade)**: Refactored backend business logic layers into clean, decoupled, and testable Spring-managed beans using **Dependency Injection** and modern Spring paradigms. *(Spring Core, Java 17, Backend Engineering)*

Technologies & Environment: Java 17, Spring MVC (TERASOLUNA 5.x), Struts 1.x (TERASOLUNA 2.2.0), intra-mart 7.2 / 8.0.36, JSP, Oracle 11g / 18c XE, Eclipse.

Contract Backend Engineer | Rakuten Group, Inc.

September 2021 - March 2025 | Tokyo, Japan

Membership System & Database Migration (Kotlin/Spring Boot Microservices)

Responsibilities:

- Conducted a critical Proof of Concept (PoC) to test if Couchbase could handle high-concurrency and high-traffic conditions, generating 300 million mock membership records for benchmarking. *(PoC, Database Benchmarking, High-Concurrency Testing)*
- Participated in executing the database migration from Couchbase to **Cassandra** after deep stress-testing revealed Couchbase's performance degradation and high error rates during node rebalancing under concurrent load. *(Database Migration, NoSQL, Fault-Tolerance)*
- Developed a high-performance benchmarking API server using the asynchronous **Vert.x** framework to generate and process the 300 million mock records for load testing. *(Vert.x, Data Generation, Backend Development)*
- Executed distributed stress testing in the CaaS environment using **Gatling**, integrating automated test scenarios into **GitLab CI/CD** pipelines. *(Gatling, Load Testing, GitLab CI/CD)*
- Developed and packaged a custom **Argon2** hashing library to standardize secure data operations across multiple microservices. *(Kotlin, Argon2, Security)*
- Contributed to code refactoring across membership services, moving duplicate data-access logic into shared core modules to reduce technical debt. *(Code Refactoring, Clean Code)*
- Designed and automated integration and regression test suites for the migrated RESTful endpoints using **Postman and Newman** across local **MiniKube** and staging environments. *(Postman, Newman, API Testing, MiniKube)*
- Updated **OpenAPI (Swagger)** documentation to ensure clear API specifications for cross-team collaboration during the database migration. *(OpenAPI, Swagger, API Documentation)*

Technologies & Environment: Kotlin, Java, Spring Boot, Vert.x, Cassandra, Couchbase, MariaDB, Gatling, Postman, Newman, GitLab CI/CD, Podman, MiniKube, IntelliJ IDEA.

E-commerce Platform (Rakuten Fashion - Agency Order Management System)

Responsibilities:

- Joined the project in its latter phase to contribute to the development and optimization of the Rakuten Fashion Agency Order Management System. *(E-commerce Operations, Order Pipelines)*
- Developed lightweight **Spring Boot scheduled batch tasks** using `@Scheduled` and **RabbitMQ** to automate error handling, periodically routing stalled or unprocessed orders back into the message queue for asynchronous re-processing. *(Spring Boot, @Scheduled, RabbitMQ, Asynchronous Processing)*

- Audited and streamlined legacy **Gradle** build configurations and internal API version management across multiple microservice repositories utilizing **JFrog Artifactory**, resolving version dependency conflicts. *(Gradle, JFrog Artifactory, Dependency Management)*
- Standardized API specifications by integrating the native **Spring Boot OpenAPI (Swagger)** extension across all backend services, establishing a clear contract for cross-team collaboration. *(Spring Boot, OpenAPI, Swagger, API Standardization)*
- Conducted initial business domain analysis by updating enterprise technical specifications in Confluence and maintaining relational database schemas and **DDL** migration documentation. *(Technical Documentation, Confluence, DDL, Database Schema Management)*

Technologies & Environment: Java, Spring Boot, RabbitMQ, Gradle, JFrog Artifactory, OpenAPI, Swagger, MariaDB, Confluence.

Earlier Experience

- **Full-time Senior Software Engineer | Techoes Inc., Tokyo** (Jan 2020 - Aug 2021)
  - Participated in backend development for a multiplayer game, leading the online chat system and AI pathfinding modules using **Go, gRPC, and Protobuf** for high-concurrency, low-latency microservices.
  - Contributed to a live streaming platform, transitioning from interactive front-end development (**Vue.js, Pixi.js**) to backend API development and performance optimization (**PHP/Laravel**).
- **Contract Software Engineer | NEC, Tokyo** (Aug 2018 - Oct 2019)
  - Developed VMS (Video Management System) integrations for the National Police Agency, engineering C# module extensions on top of the **Milestone XProtect VMS** enterprise platform (Windows/VS2017).
  - Engineered backend APIs, web interfaces, and scheduled batch tasks for the **Mobile Suica system**, taking full ownership from core implementation to designing unit test specifications and executing test suites.
- **Lead Client Engineer | Shanghai Thinky Game, China** (Jul 2016 – Feb 2018)
  - Architected and led client-side development for 3D mobile tank battle game *坦克雄心* (**Unity 5.x, C#**).
  - Engineered core combat systems including dynamic shell trajectory physics, armor damage penetration, hit-detection, visual FX managers, dynamic weather systems, and custom editor tools.
  - Conducted deep low-level profiling and optimization across Memory, CPU, and GPU pipelines to achieve high frame rates on low-spec devices.
- **Client Engineer | Shanghai Haymaker Game, China** (Dec 2015 – Jun 2016)
  - Developed core battle logic and character skill execution systems for mobile action RPG *小李飞刀* (**Unity 4.x, C#, SVN**), focusing on high-frequency action inputs and state management.
- **UI & Client Engineer | 2K Games China, China** (Mar 2015 – Nov 2015)
  - Engineered event-driven UI frameworks and HUD systems for AAA PC title *Borderlands Online* (**Unity 4.x, ActionScript 3, C#**), integrating Scaleform/Flash interfaces with C++ low-level game engines using **Perforce**.
- **Client Engineer | Shanghai Hanniu Game, China** (Mar 2014 – Feb 2015)
  - Built real-time combat modules and skill UI systems for mobile MOBA title *暴走DOTA* (**Unity 4.x, C#**), optimizing client-side network packet serialization and rendering performance.
- **UI & Client Engineer | Limbo Studio, China** (Apr 2011 – Mar 2014)
  - Developed PC poker/card gaming platform *新安游* UI and client logic using **C++, ActionScript 2, and Flash**.
  - Engineered client UI modules and HUD mechanics for PC TPS title *枪神* using **Unreal Engine 2.5 (C++, ActionScript 3)**, interfacing UI components with native UE C++ core modules.
- **UI Programmer | The9 Limited, China** (Dec 2007 – Mar 2011)
  - Developed interactive client UI modules and business logic for PC casual online title *小小部落*, leveraging **Gamebryo C++ engine**, ActionScript 2, and Flash pipelines.

- **Game QA & Operations Specialist | Wemade Shanghai, China** (Oct 2003 – Dec 2007)
  - Managed Game Master (GM) operations and technical QA for PC MMORPG *Legend of Mir 3* (*传奇3*), conducting server load verification, live test server environment troubleshooting, and operational feedback loops.
- **Computer Science Teacher | Tiantai Chengxi High School, China** (Sep 2002 – Sep 2003)
  - Taught computer science fundamentals and basic software usage, developing strong technical communication and presentation skills.

EDUCATION

- **Zhejiang University of Technology**
  - Computer Application and Maintenance