

Yuelun Yu

中国・上海 | +852 9488 2685 | ar8327k@gmail.com

github.com/ar8327 | LinkedIn/Yuelun Yu

ビザサポートが必要 | CEFR C1, IELTS 7.5 | JLPT N2, 語彙：A, 文法：A

概要

ByteDance にてスケーラブルなバックエンド基盤、APM システム、ITSM プラットフォームを構築した 4 年の経験を持つシニアソフトウェアエンジニア。最近では、マルチモデルフェイルオーバーと統一 API 抽象化を備えた LLM ゲートウェイの開発を主導。Go、Python、分散システムに精通。パフォーマンス最適化とシステムアーキテクチャに情熱を注ぐ。

技術スキル

- **言語:** Go、Python、Shell、SQL、Java
- **フレームワーク:** SpringBoot、MyBatis、Gin、gRPC
- **データベース:** MySQL、Redis、Elasticsearch
- **DevOps/ツール:** Docker、Git、Kafka、RocketMQ
- **AI/LLM:** マルチプロバイダー LLM オーケストレーション、統一 API 抽象化、トラフィックシェーピング、高可用性設計

職務経験

シニアソフトウェアエンジニア - LLM ゲートウェイアーキテクト & 開発エンジニア 2025 年 2 月 – 現在 ByteDance (中国・杭州)

- マルチプロバイダー LLM ゲートウェイを独立設計・開発し、AWS Claude、Azure GPT、GCP Gemini など複数の事業者をサポート、OpenAI 互換インターフェースを提供、数週間で本番稼働を実現。
- 高可用性ロードバランシングシステムを実装し、セッション ID ハッシュにより Prompt Cache ヒット率を 90% に最適化、Redis ZSET + Lua スクリプトによる原子的キューイングと制限により 4-5k QPM をサポート。
- インテリジェント災害復旧メカニズムを構築し、システム可用性を 70-80% から 99.9% に向上；GCP 大規模障害時でも 97% の可用性を維持し、自動トラフィック切り替えを実現。
- 数万行のプロバイダーアダプターコードを開発し、異なるベンダーの API フォーマット、ストリーミングレスポンス、関数呼び出し、エラーコードを統一処理、動的 QPM 調整と優先順位キューイングをサポート。
- 包括的監視システムを構築し、データ分析により GCP の Gemini 課金バグを発見・修復支援、会社到大額返金を獲得し、卓越した問題発見・解決能力を実証。

シニアソフトウェアエンジニア - APM メタデータシステムアーキテクチャ設計・実装 2024 年 4 月 – 現在 ByteDance (中国・杭州)

- 200 万サービスをサポートするメタデータ管理システムを独立設計・実装、QPS 3000+、SLA > 99.9%
- SQL 類似 DSL クエリエンジンを実装し、フィールド依存関係の自動解析により 20+ の外部データソースアクセスを統一化

- 多層キャッシュアーキテクチャ（Redis + MongoDB）とインテリジェント降格戦略を設計し、外部サービス障害のゼロ影響を実現
- HTTP と分散ロックベースのメッセージキュー不要のタスクスケジューリングシステムを実装、毎日 200 万件のデータ更新を処理

シニアソフトウェアエンジニア - CMDB プラットフォーム基盤再構築 2022 年 11 月 – 2024 年 4 月 ByteDance（中国・杭州）

- 主要成果：数百万規模構成項目 CMDB プラットフォーム再構築を主導。500 以上の内部システムに対応し、日平均 5000 QPS クエリリクエストを処理。
- Python/Django サービスをゴースト（Go）マイクロサービスアーキテクチャに再構築。単一インスタンススループットを 500 から 3000 QPS に向上（6 倍改善）。Django ORM N+1 クエリ問題解決と多層キャッシュメカニズム導入により、P99 レイテンシを 10 秒から 200 ミリ秒へ最適化（50 倍改善）。
- MySQL + Elasticsearch デュアルストレージエンジンアーキテクチャを構築。MySQL はトランザクション書き込みとシンプルクエリを処理し、ES は 100 以上フィールドのワイドテーブルで複雑関連クエリを支援。Canal ベースの準リアルタイムデータ同期パイプラインを設計し、RocketMQ メッセージキューによりデータ最終一貫性を秒レベルで確保。
- メタデータ駆動 RESTful API フレームワークを設計し、30 以上の CI 種別の CRUD 操作を動的処理。オブジェクト関連関係を自動解析して SQL に変換。行レベル（データ機密レベル L2/L3/L4 基準）と列レベル細粒度アクセス制御を実装し、データセキュリティを保障。
- SQL 類似構文 DSL クエリ言語を開発し、基盤 MySQL/ES ストレージ差異を抽象化。インテリジェントルーティングによりクエリを最適ストレージエンジンに分散し、Scroll API 対応により深いページングを防止。クエリ深度 1000 件以内制限でシステム安定性を保証。
- ステートレス Worker クラスターによりデータ同期処理。メッセージ分割で同一レコード順序性を保証。全量データ検証フォールバックメカニズムを実装し、毎日自動データ不整合修復。Nginx 重み配置によるグレーリリースを採用し、3 カ月再構築プロジェクト無障害稼働を実現。

ソフトウェアエンジニア - IT 資産管理システムアーキテクチャ 2020 年 11 月 – 2022 年 11 月 ByteDance（中国・北京）

- 100 万台以上のサーバー用スペア部品管理システムを設計・実装。CPU、メモリ、HDD など 10 種類以上の部品タイプをカバーし、毎日数百台のサーバー故障処理を支援。エンドツーエンドの修理時間を 30% 短縮。
- Redis ベースの分散在庫予約メカニズムを構築。アトミック操作と二段階確認により、複数のオペレーターによる同時部品申請時の売れ過ぎを防止し、最終的な一貫性を保証。
- 多層ハードウェア互換性ルールエンジンを設計。粗いスクリーニング + 精密検証の二段階フィルタリングにより、複雑な互換性チェック（CPU-メモリ連携検証、上位互換置換）をサポート。
- ドメイン駆動設計（DDD）を実践。4 つのコアドメインイベント（チケット作成、部品在庫移動、修理完了）を定義し、メッセージキューによる非同期補償と 2 時間のリトライウィンドウで障害耐性を確保。

研究経験

画像データ抽出に関する研究

2019 年 11 月 – 2020 年 3 月

杭州電子科技大学

- 科学チャートからデータを抽出するため、K-Means ベースのセグメンテーションと CRNN OCR パイプラインを開発。

ショート動画内容のフィルタリング

2018 年 11 月 – 2019 年 4 月

杭州電子科技大学

- TensorFlow を使用し、TextCNN ベースの不正短文識別器を構築、98% の再現率を達成。

論文・特許

- 画像中のデータ自動抽出方法. CN201910972334.8、共著。
- デュアルチャネル CNN による違法短文の識別. CN109670041A、共著。

受賞歴

- 教育部主催サービスアウトソーシングイノベーションコンテスト二等賞（2020 年）
- インテル組込みシステムコンテスト二等賞（2018 年）
- 国家奨学金（2018 年、2020 年）教育部
- 学校功労⊠学金（2018 年、2020 年）杭州電子科技大学

学歴

杭州電子科技大学

2017 年 9 月 – 2021 年 6 月

情報システム学士, GPA: 4.18 / 5.0

中国・杭州

リーダーシップ・活動

MOOC クラブ副会長

2019 年 6 月 – 2020 年 6 月

杭州電子科技大学・計算機学院

- 学生向けの技術コンテストを企画し、チーム結成を支援。
- ワークショップを開催し、後輩向けのメンタープログラムを構築。