**김영준** Data Engineer 

�� 010-9364-5472

✉️ dr0joon@gmail.com

https://github.com/0BVer

https://www.linkedin.com/in/0joon/

**소개**

● Spring Boot에서 동시에 10000대 이상의 IoT 장비로부터 데이터를 수집하는 ETL 파이프라인 개발 경험이 있습니다. ● Monitoring, Logging, Stress Test를 활용하여 가용성과 성능을 검증하고 이를 바탕으로 개선한 경험이 있습니다. ● End-to-End 프로젝트를 리드한 경험이 있고, Backend, Frontend, DevOps, ML Engineering을 경험하여, 웹 개발에서 데이터를 다루는 방법을 이해하고 있는 데이터 엔지니어입니다.

**경험**

**유니로보틱스 / Server Engineer (2022.03 ~ 2022.07)**

**Langur.io : IoT Data Platform**

IoT 센서 장비에서 데이터를 수집하는 파이프라인을 구축하여, 실시간 장비 제어 및 대시보드를 제공하는 프로젝트입니다. 목표 처리량은 10000대의 장비에서 데이터를 수집하는 것이었고, 부하 테스트와 모니터링을 통해 검증 하였습니다.

● 클라우드 기반 IoT Data Ingestion ETL 파이프라인 구축

○ 파이프라인 구조 : IoT ➞ Server ➞ MQ ➞ Parsing ➞ MQ ➞ Data Warehouse

○ Scale-out, Non-blocking, 신뢰성을 위해 GCP Pub/Sub Message Queue 도입 (500ms ➞ 300ms) ● 실시간 대시보드를 위한 데이터 웨어하우스 BigQuery 시각화 REST API 구현 및 성능 개선 ○ 파티션 정책을 월간에서 주간으로 변경하여 조회할 전체 행의 수를 줄여, 쿼리 성능 향상 (2000ms ➞ 600ms) ○ BigQuery 쿼리문에 Prepare Statement를 사용하여 쿼리 인젝션 공격 방어 및 옵티마이저 최적화 ● Monitoring & Alert System 구축을 통해 서버 부하 및 장애 확인을 자동화

○ 부하 시나리오 테스트를 통해 파이프라인과 API의 성능 및 신뢰성 검증

기술 스택 - Spring Boot, JPA, GCP [BigQuery, Pub/Sub], Nginx, Prometheus, Grafana, K6, Gitlab CI

**기술**

| ● **Language** : Java, Python, Go, TypeScript | ● **Server** : Spring Boot, Django, Gin |
| --- | --- |
| ● **Database** : MySQL, BigQuery | ● **MQ** : Kafka, RabbitMQ |
| **● Data Distribution :** Spark  ● **Logging** : ELK, Filebeat  ● **Client** : React, HTML, CSS | **● Scheduling :** Airflow  ● **Monitoring** : Prometheus, Grafana ● **ETC** : Avro, Nginx, Github Actions |

**블로그**

● 당신은 Kafka를 믿습니까? - Kafka와 Zookeeper 그리고 Replication - 링크

● Kafka를 사용하여 초당 10000건의 로그를 처리하는 파이프라인 구축하기 - 링크 ● Kafka 메시지에 스키마를 정의해 보자 : Apache Avro - 링크

● 쿼리 튜닝을 통해 슬로우 쿼리 개선하기 - 링크

**프로젝트**

**Kafka 기반 ETL 파이프라인 / (2023.05 ~ 진행중)**

분산 이벤트 브로커인 카프카와 분산 클러스터 컴퓨팅 프레임워크인 스파크를 사용하여 실시간 로그 중앙 처리 시스템을 구축한 프로젝트입니다. 실 서비스에 적용 가능함을 검증하기 위해 더미 로그를 사용하였습니다.

● Extract

○ 낮은 결합도로 데이터를 전달하기 위해 Zookeeper Kafka 클러스터를 이용한 이벤트 기반 아키텍처를 구현 ○ 파이프라인의 안정성을 확인하기 위해 Prometheus, Grafana를 통해 실시간 처리량과 consumer lag 모니터링 ● Transform

○ 높은 처리량과 내결함성을 확보하기 위해 Spark Streaming을 통한 분산 병렬처리를 적용

○ 메시지의 스키마를 정의하기 위해 Avro로 데이터를 직렬화하여 어플리케이션간의 데이터 버전과 호환성 관리 ● Load

○ 지속적으로 생성되는 로그 데이터를 시계열 데이터베이스인 InfluxDB에 저장하고 Grafana를 통해 시각화 ○ 비용 최적화를 위해 공간 효율이 좋은 Parquet을 사용하여 로그 데이터 백업

기술 스택 - Kafka, Zookeeper, Spark Streaming, Go, InfluxDB, Prometheus, Grafana, Avro

**Hands Up 영상 기반 Q&A 놀이터 / Backend (2023.04 ~ 진행중)**

짧은 영상을 통해 쉽고 재미있게 질문하고 답변하며 소통할 수 있는 서비스입니다. 10월에 앱 스토어와 플레이 스토어에 동시 출시하는 것이 목표이며 10명의 개발자와 디자이너가 참여하고 있습니다.

● Django를 사용하여 질문, 답변, 댓글, 신고 도메인 기능 개발

○ FCM 푸시 알림을 통해 내 활동과 관련된 게시글 등록시 푸시 알림이 생성되도록 구현

○ 앱스토어 신고 정책에 부합하는 게시글 신고 및 사용자별 블라인드 적용 기능 구현

○ 기존 DB 접근 방식이 DRF의 Serializer와 Repository로 섞여 있어서, Ninja 도입 후 Service에서 접근하도록 통일 ● 멀티 미디어 프로세싱 프레임워크 FFmpeg를 사용하여 영상 및 이미지 처리 기능 개발 ○ 원본 영상의 용량을 감소시키면서, 시각적으로 큰 차이가 없도록 비트레이트에 비례한 영상 압축 구현 (50% 감소) ○ 용량이 큰 영상 로딩 시간에 사용자 경험 개선을 위해 영상 및 이미지 썸네일 추출 기능 개발 ● 영상 프로세싱을 위한 ML 모델 서빙 파이프라인 구축

기술 스택 - Python, Django, DRF, Django Ninja, PostgreSQL, FCM, FFmpeg, Poetry, Black formatter

**Find Spot 이미지 유사도 분석 기반 Photo Spot 인증 서비스 / Full Stack (2022.03 ~ 2022.11)** 이미지 유사도 분석을 통해 관광지의 포토스팟에서 촬영한 사진을 인증하는 서비스입니다.

● Spring Boot를 사용하여 유저, 게시판, 이미지, 팔로우, 그룹 도메인 기능 개발

○ ORM에서 복잡한 쿼리를 표현하는 기능의 한계와 N+1 문제를 해결하기 위해 QueryDSL 도입 ○ 슬로우 쿼리를 튜닝하기 위해 실행 계획을 확인하고 조회에 자주 사용되는 필드에 인덱스를 적용 ○ 개발 생산성 개선을 위한 자동 배포, Blue/Green 배포방식의 CI/CD 도입

● Micro Service를 위해 이미지 유사도 분석 모델 서빙 기능을 Flask를 사용하여 독립적으로 동작하도록 개발 ● Reverse Proxy를 적용하기 위해 React로 구현한 페이지를 Nginx에서 정적 리소스로 제공 ○ ELK + Filebeat 로깅으로 로그를 시각화하여 접속 로그와 장애 분석

기술 스택 - Spring Boot, QueryDSL, JPA, MySQL, Flask, React, Nginx, AWS, Github Actions

**활동**

**Techeer [테커] / 시니어 멘토** 2022.03 ~

● 실리콘밸리 개발자의 **SW 개발자 커리어 그룹**

● 실리콘밸리 한달 살기, 기술 컨퍼런스, 해커톤, 네트워킹 행사 등 다수의 활동 주최 및 경험

**부트캠프 멘토**

● [디캠프 x 테커] **실리콘밸리 SW 부트캠프** 2023.06 ~ 2023.08

● [티타임즈 x 테커] **실리콘밸리 SW 부트캠프** 2023.01 ~ 2023.02

● [성남시] **AI 실리콘밸리 프리 인턴십 부트캠프** 2022.09 ~ 2022.10

**해커톤 주최**

● **2회 Good Night Hackathon** 2023.08.26 ~ 2023.08.27

● [한국공학대학교 x 테커] **2023 실리콘밸리 개발자 아이디어 해커톤** 2023.05.26 ~ 2023.05.27 ● **1회 Good Night Hackathon** 2023.02.24 ~ 2023.02.25

**부트캠프 참가**

● [성남시] **AI 실리콘밸리 프리 인턴십 부트캠프 3기 수료** 2021.12 ~ 2022.02 ● [Posco + 성남시] **AI 실리콘밸리 프리 인턴십 부트캠프 2기 수료** 2021.05 ~ 2021.07

**교육**

소프트웨어전공 / 안양대학교 2019.03 ~ 2023.02