김승갑 (Seunggab Kim) · Data Engineering Team Lead
seunggabi@gmail.com · LinkedIn · GitHub · Blog · English · PDF

10년 차 개발자이고, 그중 8년 넘게 데이터 엔지니어링을 했습니다. CDC, 스트리밍, 레이크하우스로 이어지는 데이터 플랫폼을 만들고 운영합니다. 문제의 원인이 오픈소스에 있으면 업스트림에 직접 고쳐서 올리고, Apache Spark·Superset·Airflow에 머지됐습니다.

경력

오렌지스퀘어 (WOWPASS) · Data Engineering Team Lead · 2023.10 – 현재 - 운영 MySQL 테이블 197개의 변경분을 Debezium과 MSK Connect로 받아 Iceberg 레이크하우스(S3 Tables)에 3분 커밋 주기로 반영하는 CDC 파이프라인(토픽당 파티션 6개·복제 3, Parquet zstd, 목표 파일 128MB)을 설계·운영해, 분석 쿼리를 운영 DB에서 분리 - Kafka·Flink 스트리밍과 Avro·Schema Registry 기반 앱 이벤트 로그 파이프라인 구축(하루 약 170만 건, 프로듀서 acks=all·idempotence·lz4) - Airflow DAG 1,300여 개(표준 DAG 팩토리 18종, SQL 잡 약 6,000개)와 EMR Spark 배치 운영, MSK·EMR 알람 6종을 Terraform 모듈로 관리 - DataHub 카탈로그 도입, 사내 KPI 대시보드 개발, Athena 위 Superset 운영 참여(대시보드 100여 개, 사용자 약 50명) - 외부 결제·교통카드 연동 서버 API 개발, GitHub Actions 25개로 리뷰·릴리스 승격·의존성 감사 등 개발 프로세스 자동화

티맵모빌리티 · Data Engineer · 2022.10 – 2023.10 - 모노 레포와 공용 core 모듈 도입을 제안하고 적용해 배치 잡 수십 개를 통합, 신규 잡 개발 기간을 3일에서 몇 시간으로 단축. Airflow DAG를 저장소 기반으로 버전 관리 - Kotlin 배치 잡: Hive→DynamoDB 적재, JSON→Hive 적재, 쿼리 결과 S3 추출, S3 버전 복구, SQL 결과 Slack 알림

버킷플레이스 (오늘의집) · Data Engineer · 2022.04 – 2022.10 - 광고 리포트 배치를 Spark 잡과 Spring Batch로 나눠 집계, Kotlin·Spring Boot 서비스의 빌드·배포 구조 - OLAP 데이터 동기화를 API 호출에서 Kafka 비동기 방식으로 전환해 동기화 장애 0건, 모노 레포와 Spring Config Server, 글로벌 로그 스키마 설계 참여

NAVER · 2016.12 – 2022.04 - Data Engineering TL (2021.04 – 2022.04): 글로벌 로그 시스템과 대시보드 과제 리드, NAVER TV 개인화 API - Data Engineer (Back End 포함, 2018.10 – 2021.04): 음악 서비스(VIBE·LINE MUSIC·NOW) 글로벌 로그 시스템(하루 수억 건)을 KR·JP·NOW로 확장 (글), Spark Streaming으로 HDFS 적재 (글), Spark·Hive 지표 배치 수십 개 (글), 사용자 수백만 명의 연말 결산 데이터 (글) - 같은 기간 HBase 개인화 API와 Django 대시보드를 만들고, 노래방·내돈내듣 차트 신설과 LINE MUSIC 비디오 추천 개인화에 참여 - Front End Developer (2016.12 – 2018.09): 블로그·포스트 스마트에디터 컴포넌트(React·MobX)

오픈소스

- **Apache Superset** (머지 3건): 데이터셋 시간 포맷 `epoch_us` (글), MCP API 키 발급 범위 제한 (글), CI 수정. 다운로드 사유 기록 제안 SIP-230 (글)
- **Apache Spark**: `SparkSession.builder.config(Map)` 추가, 3.4.0 포함 (글)
- **Apache Airflow**: `TableauOperator` 템플릿 필드 (글), Helm 차트 release 이름 길이 수정 (글)
- **PyAthena**: Athena 복합 타입 파싱 수정 (글) · **Apache Iceberg**: Kafka Connect 개선 PR 리뷰 중

기술

분야	기술
수집·스트리밍	Kafka, Kafka Connect, Debezium, Amazon MSK, Flink, Spark Streaming, Avro
저장·쿼리	Iceberg, S3 Tables, Hive, Spark SQL, Athena, Presto·Trino, MySQL, HBase, Redis
배치·오케스트레이션	Spark, Airflow, EMR, Spring Batch
서빙·거버넌스	Superset, DataHub, Spring Boot, Django
언어·인프라	Kotlin, Java, Python, SQL · AWS, Terraform, Kubernetes(Helm), GitHub Actions

학력·자격

명지대학교 컴퓨터공학 학사 (학점 4.18/4.5) · 정보처리기사, OCP 11g Administration, 리눅스마스터 2급