Sanghoon Kim

+82-10-2746-1443 / seiker@kaist.ac.kr / seiker.kr / linkedin.com/in/seiker

Summary

대규모 배치 데이터 플랫폼의 설계·운영·고도화를 수행해온 소프트웨어/데이터 플랫폼 엔지니어입니다. 의료 AI 도메인에서 GKE 기반 Research Data Pipeline을 운영하며 누적 3~4PB 규모의 ETL과 수백 개 노드 병렬 처리 워크로드를 다뤘고, 데이터 검증·표준화·메타데이터 관리, Iceberg·BigQuery·Trino 기반 레이크하우스 및 데이터 웨어하우스 고도화를 수행했습니다. Airbyte·Prefect 기반 자동화와 비용/성능 최적화 경험이 있으며, 영어 기반 협업 환경에서 기술 커뮤니케이션을 해왔습니다.

Core Skills

- **Programming Language:** Python, TypeScript
- **Data Platform/Storage:** BigQuery, Iceberg, Trino, PostgreSQL, Airbyte
- **Workflow/Processing:** Prefect, Dask, Ray
- **Infra/Deployment:** GCP, GKE, Kubernetes, Docker, Terraform
- **Language:** Korean (Native), English (Working proficiency; technical documentation and collaboration in multinational team)

Work Experience

Lunit Inc. Seoul, Korea

Aug. 2021 - Present

Senior Data Engineer @ AI Platform DataOps Team (2024.03 - 현재)

대규모 의료 이미지(DICOM/WSI) 데이터의 검증·표준화·메타데이터 적재·AI 인퍼런스를 포함한 배치 파이프라인을 운영·고도화하며, 데이터 품질/거버넌스와 비용·성능 최적화를 주도했습니다. 연간 인사평가에서 3년 연속 최우수 등급을 받았습니다.

- Prefect 기반 Research Data Pipeline을 GKE 환경에서 운영하며, 누적 3~4PB 규모 ETL과 수백 개 노드를 사용하는 병렬 처리 워크로드의 장시간 실행 안정성을 개선
- 데이터 검증·표준화·메타데이터 적재 프로세스를 운영하고, 3계층·4버킷 데이터레이크 구조 및 경로 규칙을 구현해 데이터 관리 일관성과 확장성 확보
- CSV·Parquet + Pandas 기반 데이터 인입 파이프라인을 Iceberg·BigQuery 기반으로 재구축해 57시간 작업을 1분 내/$6.25 이하로 단축, 비용 95%+ 절감
- PostgreSQL 기반 Analytics DB의 Airbyte sync failure/CPU saturation 문제를 분석하고, 사용자 인터뷰와 요구사항 정의를 바탕으로 Iceberg·Trino 기반 Medallion DW 아키텍처를 제안·설계·구축
- GKE 환경에서 Airbyte→Iceberg Bronze, Trino/Prefect 기반 Silver·Gold 빌드, Polaris REST Catalog, DataHub, Superset을 Helm 기반으로 배포·연동해 SQL 조회·데이터 탐색·대시보드 활용 체계를 마련
- 차세대 RDP 스택 선정 과정에서 Spark/Ray/GCP Batch 및 Airflow v3/Prefect v3/Dagster 후보 비교·벤치마크 수행

Senior Frontend Engineer @ AI Platform Frontend Team (2021.08 - 2024.03)

- 사내 데이터 관리 플랫폼(AICP Console) 프론트엔드 개발을 담당하며, 데이터 조회·관리·시각화 기능을 구현하고 데이터 구조와 운영 워크플로우를 반영한 내부 도구 UX를 설계
- Annotation Tool의 50개+ 사용자 행동 이벤트 taxonomy와 로그 수집 체계를 설계해 품질 평가 및 부정행위 탐지 기반을 마련하고 연간 약 $50,000+ 비용 절감에 기여
- 사용자 행동 로깅을 위한 FastAPI 기반 웹 서비스 개발에 기여, 백엔드 엔지니어들과 협업하여 로깅을 실제 운영 환경에 적용

※ 주요 프로젝트 상세는 포트폴리오에 정리되어 있습니다.

Education

한국과학기술원 (KAIST) 전산학 석사 *Sep. 2019 - Aug. 2021*

- 전체 GPA 4.05 / 4.3
- 지도 교수: 고인영 교수 (웹 공학 연구실)
- 학위 논문: "IoT 환경에서 최종 사용자 서비스 매쉬업을 지원하기 위한 대화형 서비스 매쉬업 모델"

울산과학기술원 (UNIST) 전기전자컴퓨터공학부 학사 (우등) *Mar. 2015 - Aug. 2019*

- 전체 GPA 3.78 / 4.3, 전공 GPA 3.99 / 4.3
- 컴퓨터공학 전공 및 경영공학 부전공

Selected Publications

[1] C. Lee, S. Park, H. Song, J. U. Ryu, **S. Kim**, H. Kim, S. Pereira, and D. Yoo. "Interactive Multi-Class Tiny-Object Detection." In **CVPR 2022**.
[2] **S. Kim** and I.-Y. Ko. "A Conversational Approach for Modifying Service Mashups in IoT Environments." In **CHI '22**.
[3] C. Lee, **S. Kim**, D. Han, H. Yang, Y. Park, B. C. Kwon, and S. Ko. "GUIComp: Design and Evaulation of Mobile GUI Design Assistant with Real-Time, Multi-Faceted Feedback." In **CHI '20**.