

5+ Years · Data Engineer

나준혁

Project Portfolio

현업의 맥락을 이해하고, 운영 가능한 데이터 흐름을 설계합니다.

CAREER SNAPSHOT

웹 개발로 커리어를 시작해 공공·민간 데이터 플랫폼에서 수집, 적재, 모델링, 오케스트레이션, BI까지 데이터의 전체 라이프사이클을 경험했습니다.

화려한 기술 자체보다 비용효율의 원인을 찾고, 운영 가능한 표준과 자동화로 바꾸는 문제 해결 과정에 강점이 있습니다.

CORE SIGNALS

01

누리꿈소프트(주)

데이터 엔지니어 · 프로

2021.07 - 2026.05 · 4년 11개월

02

루나랩스주식회사

웹 개발자

2021.01 - 2021.07 · 7개월

01

배치 최적화

1,000+ dbt 모델의 워크로드
격리·배치 표준화

02

플랫폼 전환

Oracle·SAS 기반을
Airflow·dbt·Redshift로 전환

03

데이터 신뢰성

품질 검증·모니터링·장애 복구
체계 자동화

기술 역량

언어·쿼리

Python · SQL · Java

오케스트레이션·변환

Apache Airflow · dbt

DW·데이터베이스

Amazon Redshift · Oracle · PostgreSQL/pgvector · MySQL ·
MSSQL · SAP HANA

클라우드·빅데이터

AWS S3 · AWS Glue · Amazon Athena · Hadoop · Hive

개발 운영·CI/CD

Git · Bitbucket · Bamboo · Docker · GitHub Actions

문서화·협업

Confluence · Jira (진척도 확인)

AI·에이전트

HyperCLOVA X · LangGraph · RAG

보유 자격

AWS Certified Solutions Architect - Associate (2026.09.01) · AWS Certified Data Engineer - Associate (2026.08) · 빅데이터분석기사 (2025.07) · SQLD (2023.12) · 정보처리기사 (2024.09) · 리눅스마스터 2급 (2024.1) · 컴퓨터활용능력 2급 (2024.1)

01 대표 프로젝트

GS리테일 HBU 데이터 플랫폼 전환

Oracle-SAS 기반 호스팅 데이터를 Airflow, dbt, Redshift 기반 플랫폼으로 전환하고 배치 시간, 비용, 품질 거버넌스를 개선했습니다.

기간	역할	기여도
2025.06 - 2026.03	배치 최적화 매니저	100% 배치 최적화 담당 영역 전반 주도

01 워크로드 격리와 비용 최적화 문제 1,000개 이상의 dbt 모델이 특정 시간대에 Provisioned 클러스터의 쿼리 큐잉과 리소스 경합을 유발했습니다. 실행 manifest.json 의존성을 파싱하고 헤비 모델을 Redshift Serverless로 분리 실행하는 태그 기반 파이프라인을 설계했습니다. 성과 상시 클러스터 종설 없이 피크 타임 병목을 제거하고 불필요한 비용을 방어했습니다.	02 Stateful Backfill 체계 문제 Hourly 배치의 소급 범위가 4시간으로 하드코딩되어 장시간 장애 시 복구에 제약이 있었습니다. 실행 Airflow Variable로 마지막 성공 data_interval_end를 추적하는 상태 관리 로직을 도입했습니다. 성과 안전 소급 범위를 24시간 이상으로 확장해 데이터 유실 위험을 낮췄습니다.
03 대규모 오케스트레이션 표준화 문제 단일 DAG에서 대규모 모델을 파싱할 때 렌더링 지연과 워커 슬롯 낭비가 발생했습니다. 실행 도메인별 Sub DAG와 Control DAG 구조로 분리하고 deferrable 실행을 템플릿화했습니다. 성과 파싱 부하와 워커 점유를 줄이고 다수 수행사가 함께 사용할 수 있는 운영 표준을 만들었습니다.	04 데이터 품질 가드레일 문제 Redshift가 PK 유일성을 강제하지 않아 멀티 수행사 환경에서 중복 데이터 인입 위험이 있었습니다. 실행 핵심 모델의 dbt test 의무화와 배치 후행 전수 중복 검증 DAG를 구축했습니다. 성과 사람의 주의력에 의존하지 않는 이중 데이터 품질 검증 체계를 확립했습니다.
05 메타 DB 부하 개선 문제 Grafana의 비효율적인 메타 DB 쿼리가 Airflow 스케줄러 지연을 유발했습니다. 실행 인덱스를 고려한 경량 쿼리로 대시보드를 재설계했습니다. 성과 스케줄러 응답을 정상화하고 태스크·Pool Slot·지연율 모니터링을 안정화했습니다.	

STACK Apache Airflow 2.10 · dbt · Amazon Redshift · AWS ECS Fargate · Grafana · Bitbucket · Bamboo · Confluence

공공·민간 데이터 플랫폼 구축 경험

서로 다른 업무 맥락에서도 공통 표준, 자동화와 운영 가능한 구조를 중심으로 문제를 해결했습니다.

PUBLIC SECTOR

차세대 KICS 데이터 플랫폼 구축

2023.09 - 2024.12 공통 지원 선임

기여도 70% · 공통 표준·자동 수집 핵심 담당

다수 사법기관의 이기종·비정형 데이터를 안정적으로 수집하기 위한 공통 표준과 자동 수집 프로세스를 구축했습니다.

이기종 데이터 표준화

적재 실패를 줄이고 파이프라인 안정성을 높였습니다.

법령 수집 무인화

반복 작업을 자동화하고 법령 데이터의 최신성을 보장했습니다.

TeraStream · Hive · Hadoop · Python · Shell

PUBLIC SECTOR

차세대 지방재정관리시스템 데이터 플랫폼

2022.05 - 2023.08 OLAP 개발 선임 · Admin

기여도 70% · OLAP-Admin-외부 연계 핵심 담당

행정안전부와 전국 지자체가 사용하는 정책지원서비스의 통계·BI 화면과 외부 데이터 연계 파이프라인을 구축했습니다.

One-Table 모델 도입

200여 개 화면을 기한 내 완수하고 신규 화면 개발 생산성을 높였습니다.

행정 비즈니스 로직 전산화

유형분류·위기관리 심화 화면 30여 본을 성공적으로 오픈했습니다.

외부 API 수집 기반 마련

내외부 데이터 통합을 위한 실행 가능한 기반을 마련했습니다.

BI-Matrix · InnoQuartz · SAP HANA · Oracle · Python

PRIVATE SECTOR

LG화학 전사 데이터 레이크 구축 1차

2021.07 - 2022.05 데이터 분석 및 ETL 선임

기여도 70% · 원천 분석·적재 전략 핵심 담당

전사 정형 데이터를 AWS 데이터 레이크로 이관하기 위한 원천 분석과 적재 전략 수립을 담당했습니다.

테이블별 적재 전략 최적화

초기 데이터 레이크 파이프라인의 경제성과 처리 효율을 확보했습니다.

AWS S3 · AWS Glue · Athena · Apache Airflow · Python

2021.03 - 2021.05

한화솔루션 ESH Portal

임직원 건강관리와 환경 공정시설을 통합 관리하는 포털을 개발했습니다.

기여도 70% · Backend-Frontend 주요 구현

2021.02 - 2021.03

UBASE 관리자 인사평가 시스템

평가 히스토리 관리를 도입하고 인사평가 메뉴의 전체 로직을 재설계했습니다.

기여도 70% · 인사평가 메뉴 전면 재설계

NEXT CHAPTER

현업에서 쌓은 데이터 신뢰성 경험을 AI Agent 설계로 확장

수집·적재·모델링·배치 운영과 품질 검증 경험을 생성형 AI 서비스에 적용하기 위해 NIPA-NAVER Cloud Sovereign AI PBL에 참여했습니다.

이후 Pop-Talk만 설계 자료가 자세한 이유

현업 프로젝트는 고객사 보안상 공개 가능한 문제·실행·성과를 중심으로 요약했습니다. 교육 팀 프로젝트인 Pop-Talk는 직접 만든 발표자료와 비식별 구현 화면을 활용할 수 있어 기획·Agent Flow-클라우드 아키텍처까지 상세히 보여드립니다.

프로젝트 성격

직무 교육 기반 4인 팀 프로젝트

검증 범위

RAG · LangGraph · NCP 서비스 구현

03 직무 교육 팀 프로젝트 · Pop Talk

추천을 생성하는 것보다 신뢰하게 만드는 문제

사용자의 취향뿐 아니라 누구와 보는지, 현재 기분과 시청 가능 시간 같은 상황까지 이해하고 추천 이유와 검증 근거를 함께 제공하는 Personal Movie AI Agent입니다.

현업 역량의 확장	데이터 파이프라인의 신뢰성·검증·재처리 관점을 영화 추천 Agent의 검색 근거 관리와 품질 검수 흐름으로 연결했습니다.
-----------	---

프로그램	협업	최종 평가
NIPA-NAVER Cloud Sovereign AI PBL 기반 프로젝트	4인 팀 · 기여도 40% 데이터·AI Agent 구현 참여	우수상

해결하려 한 문제	
01	영화 선택에 시간이 오래 걸리지만 기존 추천은 현재 상황을 충분히 반영하지 못했습니다.
02	생성형 AI 추천은 근거가 부족하고 환각 가능성이 있어 결과를 신뢰하기 어려웠습니다.
03	영화 정보, 리뷰, 평점과 관람 기록이 여러 서비스에 분산되어 있었습니다.

01	02	03	04
Personalization	Explainability	Verification	Trust
사용자 취향과 현재 요청 조건을 함께 반영	추천한 이유와 사용한 근거를 사용자에게 설명	영화 정보, 요청 조건과 리뷰 근거를 재검증	검수 결과와 안전한 폴백으로 응답 신뢰성 확보

DATA PIPELINE	RAG CONTEXT	AGENT QUALITY
공공 영화 데이터를 검색 가능한 근거로	메타데이터와 리뷰를 답변 근거로	실패를 원인별로 되돌리는 검수 루프
KOBIS·KMDB 데이터를 스케줄 수집하고 정제·병합해 PostgreSQL에 적재하는 흐름을 구현했습니다.	영화와 리뷰 임베딩을 pgvector에 저장하고, 검색 결과와 근거 ID가 응답 생성까지 이어지도록 구성했습니다.	의도 분류·QueryPlan·검색·생성·검수를 분리하고 실패 원인에 따라 필요한 단계부터 재실행하도록 설계했습니다.

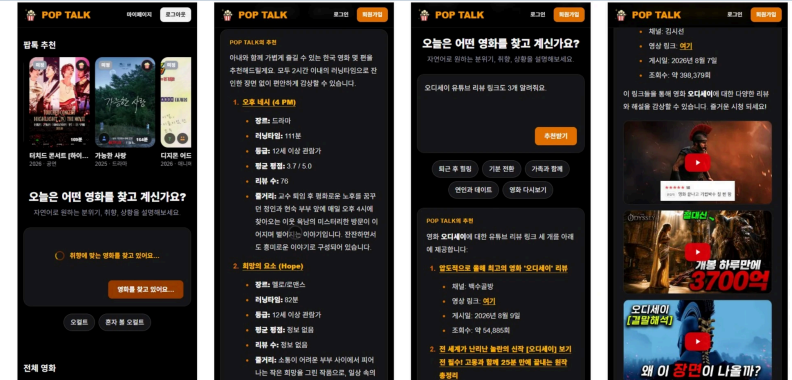
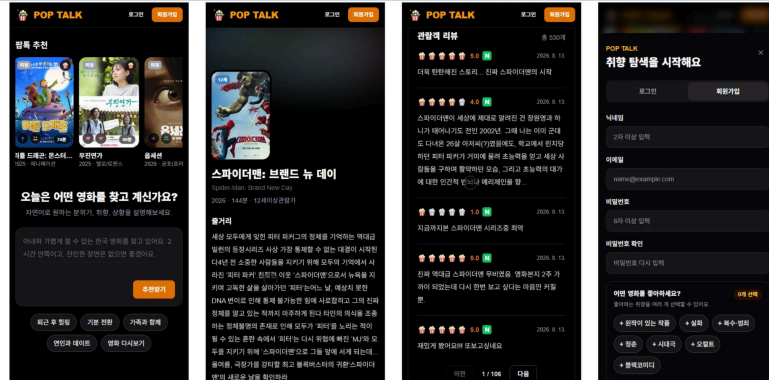
DEMO SNAPSHOT · 2026.08	5,315편	5,309편	59,863건
	영화 메타데이터	인증 완료 영화	누적 감상평

04

Pop Talk · 구현 화면

사용자 경험부터 운영·검수까지

발표자료에 포함된 실제 구현 화면으로 추천 서비스와 데이터 운영 범위를 함께 보여줍니다.



USER EXPERIENCE

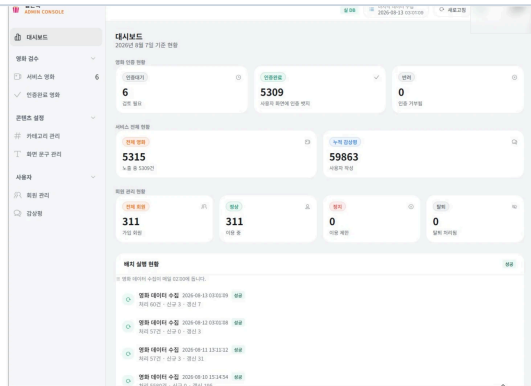
탐색·상세·리뷰·취향 온도당

영화 탐색과 상세 정보, 관객객 리뷰, 가입 시 선호 키워드 선택까지 개인화의 입력과 결과를 하나의 흐름으로 연결했습니다.

USER EXPERIENCE

자연어 추천과 외부 리뷰 근거

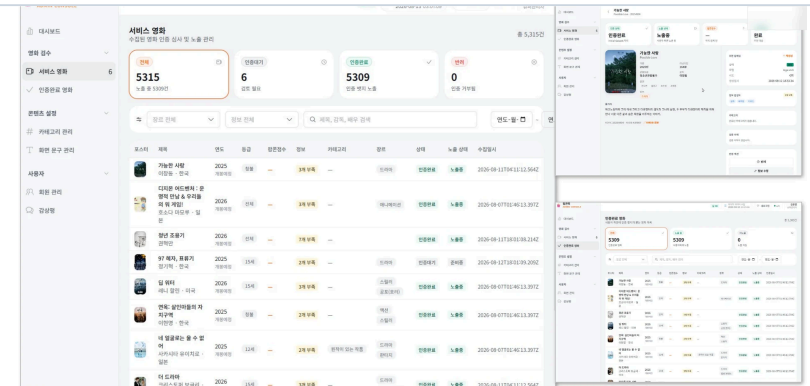
상황을 담은 자연어 요청에 영화별 추천 이유를 제공하고, 관련 리뷰 콘텐츠까지 이어지는 응답 화면을 구현했습니다.



ADMIN OPERATIONS

데이터와 배치를 한눈에 보는 운영 대시보드

영화 인증 상태, 전체 영화·감상평 현황과 매일 실행되는 데이터 수집 배치의 처리 결과를 한 화면에서 확인합니다.



ADMIN OPERATIONS

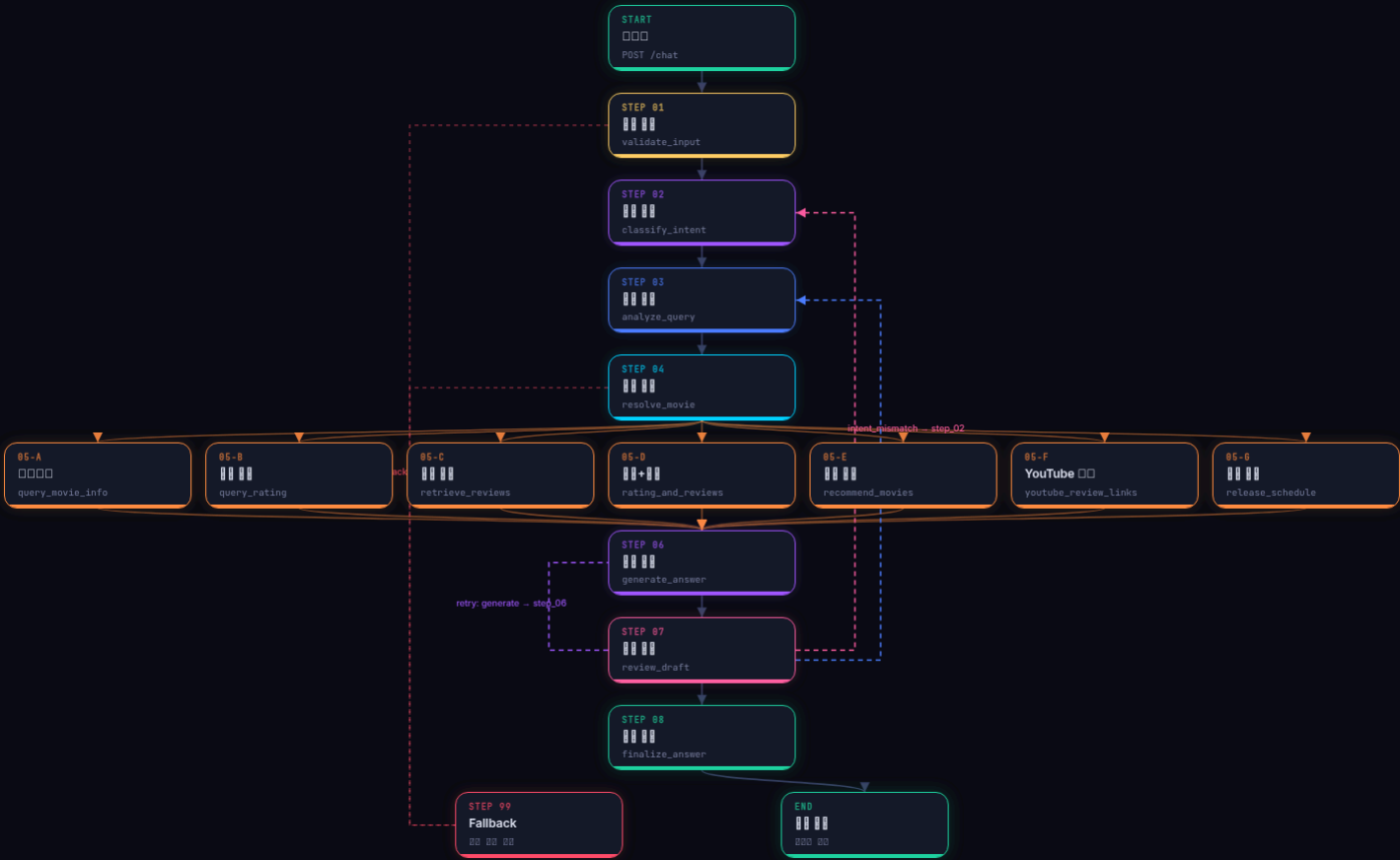
영화 검수와 노출 상태 관리

수집된 영화의 정보 누락, 인증 상태와 사용자 노출 여부를 검색·필터링하고 상세 화면에서 검수할 수 있습니다.

05 Pop Talk · Agent 설계

실패 원인에 따라 필요한 단계부터 다시 실행

입력 검증부터 의도 분류, 검색, 생성, 품질 검수와 자가교정까지 실제 소스의 노드 및 점선 재시도 흐름을 그대로 반영했습니다.



실선은 기본 처리 경로, 점선은 검수 결과에 따른 재분류·재검색·재생성 및 Fallback 경로입니다.

01	02	03	04	05
KOBIS 영화 목록·상세 수집	KMDB 포스터·등급·줄거리·키워드 결합	데이터 정제·병합 후 PostgreSQL UPSERT	영화·리뷰 임베딩을 pgvector에 저장	RAG 검색 결과를 HyperCLOVA X 응답 근거로 사용

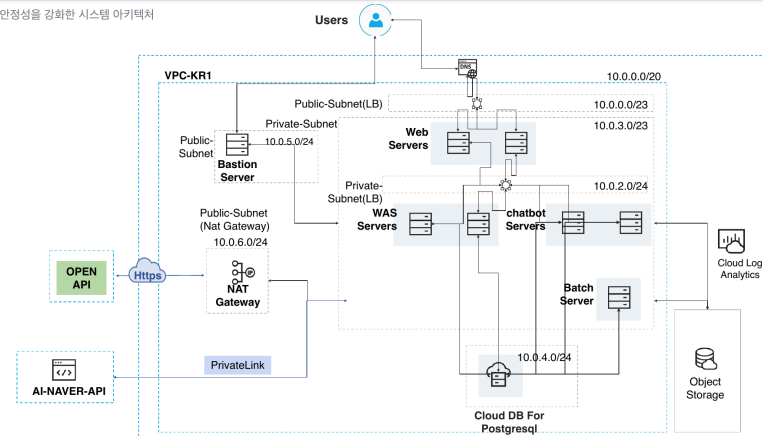
노드별 상세 처리 로직과 설계 의도는 웹 포트폴리오의 인터랙티브 그래프에서 확인할 수 있습니다.

서비스부터 PrivateLink까지 분리한 NCP 구조

사용자 트래픽과 외부 API·AI API의 진입 경로를 분리하고 Web·WAS·ChatBot·DB를 Private Subnet 중심으로 구성했습니다.

NCP 시스템 아키텍처

보안성과 안정성을 강화한 시스템 아키텍처



사용자 트래픽: Public LB → Web → Private LB → WAS/CatBot | 외부 API: NAT Gateway | AI: PrivateLink

HOPE TEAM | Pop Talk

14

Service

Next.js 사용자·관리자 화면, FastAPI 기반 REST API

AI & Agent

HyperCLOVA X, CLOVA Studio Router, LangGraph, RAG

Data

KOBIS·KMDB 수집 배치, PostgreSQL 16, pgvector

NCP Infrastructure

Public-Private Subnet, 이중 Load Balancer, Bastion, NAT Gateway, PrivateLink

다음 고도화 과제

01

추천 검증 고도화

벡터 저장과 RAG 조회 과정에서 아쉬웠던 검색 정확도와 추천 근거의 적합도를 개선합니다.

02

멀티플렉스·OTT 연계

추천 영화가 현재 어느 멀티플렉스와 OTT에서 제공되는지 함께 안내합니다.

03

장기 취향 Profile

최초 선호 장르 선택에 그치지 않고, 시청 영화의 평점과 리뷰를 누적해 사용자 취향 선호도를 지속적으로 고도화합니다.

기술 스택

Front-End

Next.js · React · TypeScript · Tailwind CSS

Back-End

Python · FastAPI · SQLAlchemy

AI & Agent

HyperCLOVA X · LangGraph · RAG · CLOVA Studio

Data

PostgreSQL · pgvector · KOBIS API · KMDB API

Infra

NAVER Cloud Platform · VPC · Load Balancer · Object Storage

운영 가능한 데이터 흐름을 설계합니다.

najunhyeok94@gmail.com · <https://junhyeok94-ls.github.io/>