

나는 데이터라이즈(DataRize)의 Data Engineer(Junior) 직무에 지원하고자 한다.

1. 지원 회사 및 직무 소개

- 회사명: 데이터라이즈(DataRize)
- 직무: Data Engineer (Junior)
- 채용 링크:

https://team.datarize.ai/job_posting/9f0yBqru?utm_campaign=실적최대화_0613&utm_medium=ppc&utm_source=adwords

2. 직무에서 수행할 구체적인 업무

첫째, 500 개 이상의 고객사 주문 데이터(3 억 건 이상)와 일간 5 천만 건 이상의 행동 로그를 준실시간·일 단위로 수집·저장·정제하는 ETL/ELT 파이프라인을 설계·개발·운영한다.

둘째, 클라우드 기반 데이터 레이크 및 인프라를 구축·운영하며, 내부 구성원과 서비스 요구사항에 맞춰 지속적으로 고도화한다.

셋째, Apache Spark 등 분산 처리 엔진을 활용해 수억 건 규모의 데이터를 효율적으로 처리·서빙하는 애플리케이션과 서빙 시스템을 설계·운영한다.

넷째, 멀티 리전·다중 타임존 환경에서도 안정적으로 동작하는 글로벌 데이터 엔지니어링 방법론을 수립·적용한다.

3. 채용 공고가 요구하는 핵심 역량

1. 데이터 엔지니어링 및 플랫폼 개발·운영 경험(2 년 이하)
2. 분산 처리 이론·기술 이해(Spark 등)
3. Python 등 적어도 한 가지 프로그래밍 언어 숙련
4. 클라우드 환경(AWS·GCP 등) 데이터 엔지니어링 경험
5. 비즈니스 문제 해결을 위한 데이터 활용 역량 및 열의
6. 데이터 분석가·타 팀과의 원활한 커뮤니케이션 능력

우대 내용

- Kafka·Flink 등 대용량 실시간 스트리밍 처리 경험
- 기술 원리에 대한 심층 이해 및 실무 적용 경험
- 스타트업 환경에서 자율적으로 과제를 정의·해결한 경험
- 다국가·다중 타임존 데이터 시스템 운영 경험

4. 요구 역량과 내가 가진 기술·경험의 유사성

- 나는 백엔드 데이터베이스 운영 경험이 있어 로그와 주문 데이터를 수집·적재·관리해 본 ETL/ELT 파이프라인 운영 역량을 갖춘다.

- 나는 백엔드 서버의 동작 원리를 깊이 이해하며, Python 으로 데이터 처리·자동화 스크립트를 작성한다.
- 나는 AWS 환경에서 서버를 배포·관리해 본 경험이 있어 클라우드 기반 인프라 운영이 가능하다.
- 나는 팀 프로젝트에서 데이터 분석가 및 타 팀과 협업하며 문서화·코드 리뷰를 통해 의사소통 역량을 입증했다.

5. 본인의 부족 역량 및 강화해야하는 내용

- 실시간 스트리밍 처리 경험이 부족하므로, Kafka-Spark Streaming 관련 역량 강화
- 분산 클러스터 운영 경험이 미흡하므로, AWS EMR-Dataproc 에서 Spark 클러스터를 구성·운영해 보는 실습 필요
- 글로벌 멀티 리전 운영 지식이 없으므로, 다중 리전 배포 사례를 학습하고 샘플 아키텍처를 설계 필요