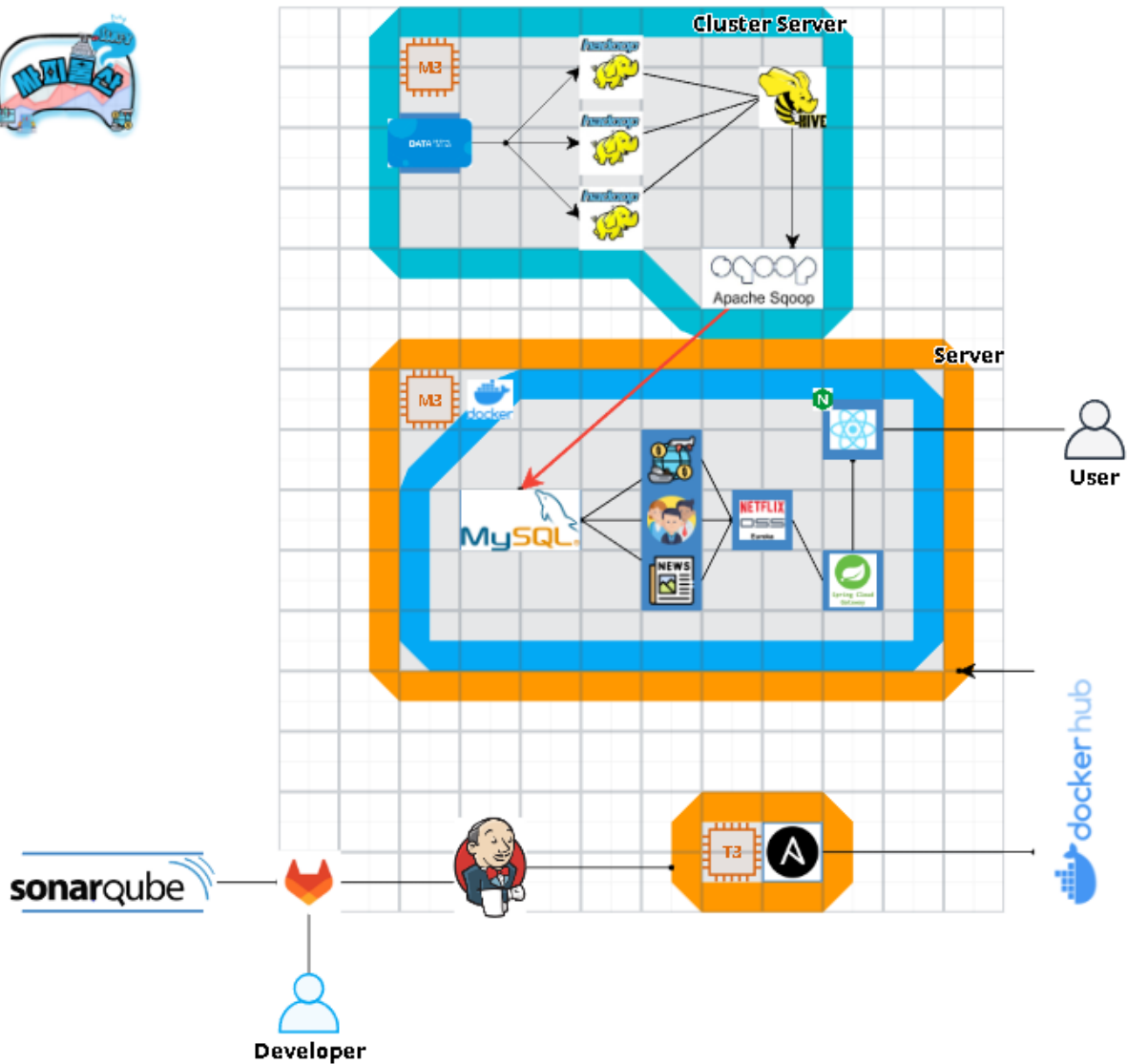
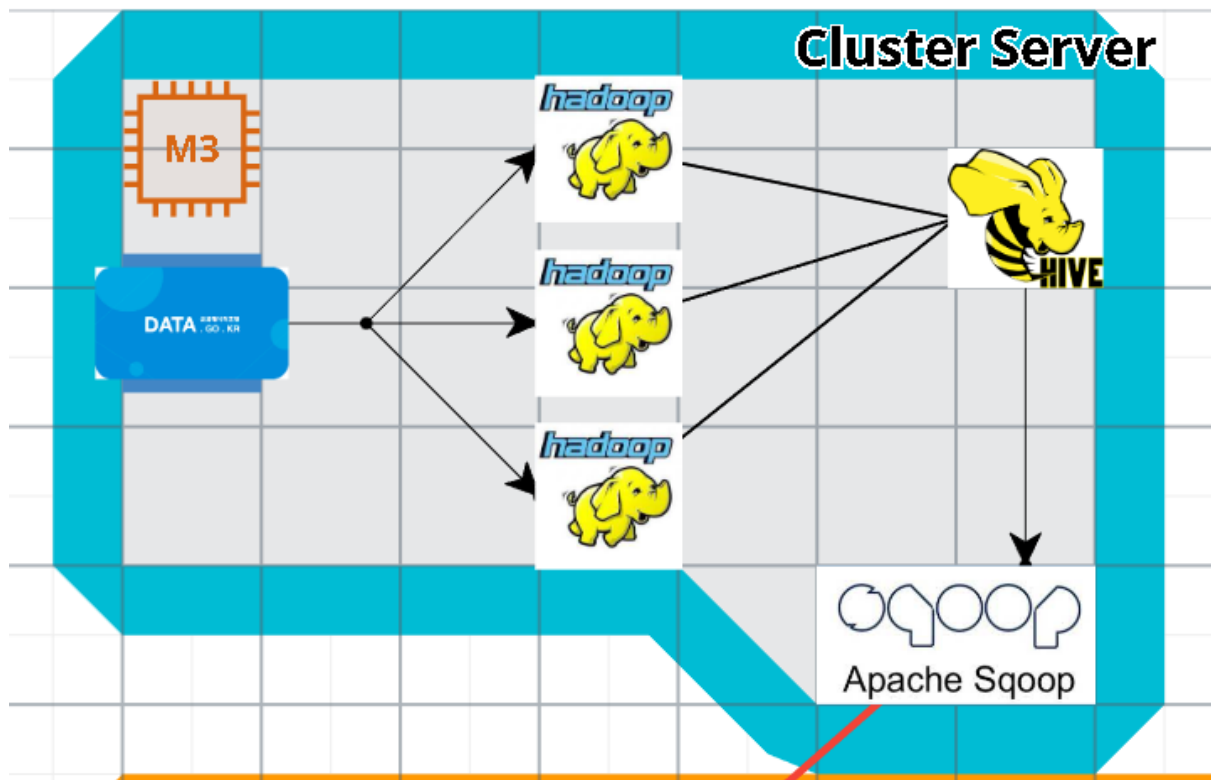




시스템 아키텍처



빅데이터 분산 처리



공공데이터 관세청 자료 API

- 하둡의 hdfs 원천 데이터

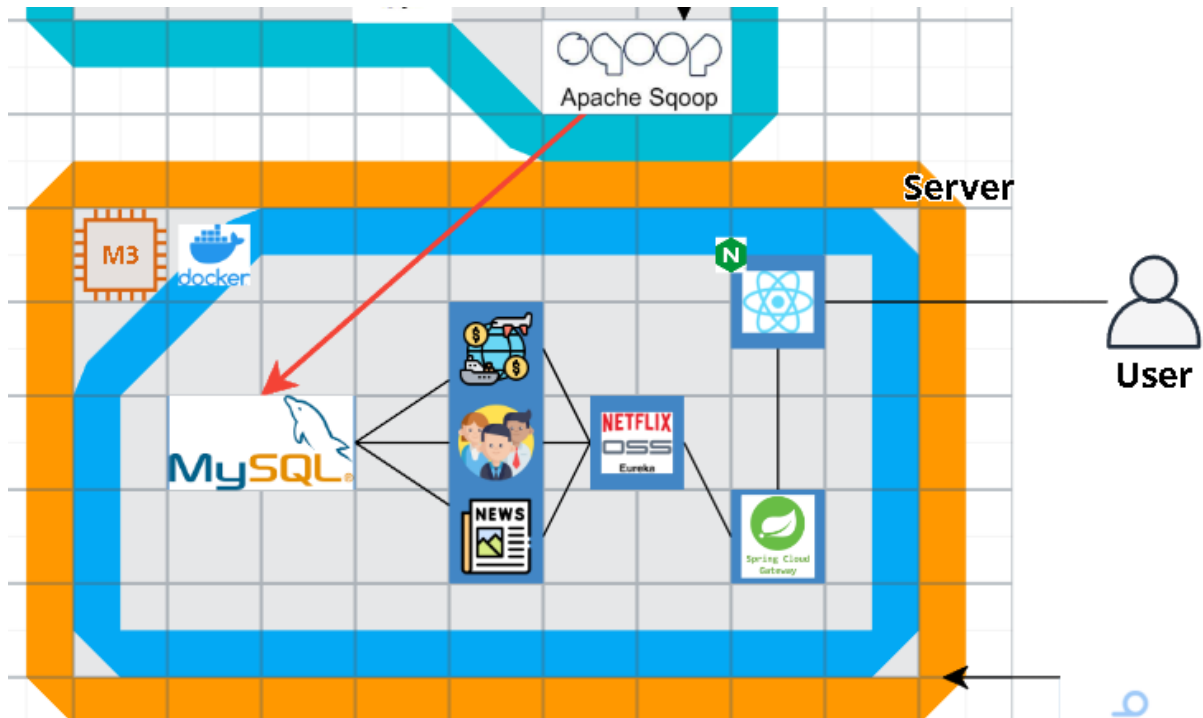
하둡 분산처리

- Hadoop cluster
 - 하둡 분산형 파일 시스템 전체에 걸쳐 다양한 오케스트레이션 작업을 실행
 - MapReduce를 사용해 데이터를 병렬식 연산으로 실행
- HIVE
 - hdfs의 분산 처리된 결과를 결정된 스키마에 따라서 정형화된 데이터로 제공하는 SQL 툴

데이터 적재

- 아파치 스쿱
 - 구조화된 관계형 데이터 베이스 (MySQL)과 아파치 하둡 간의 대용량 데이터를 효율적으로 변환하여 적재하는 인터페이스 애플리케이션

클라우드 서버



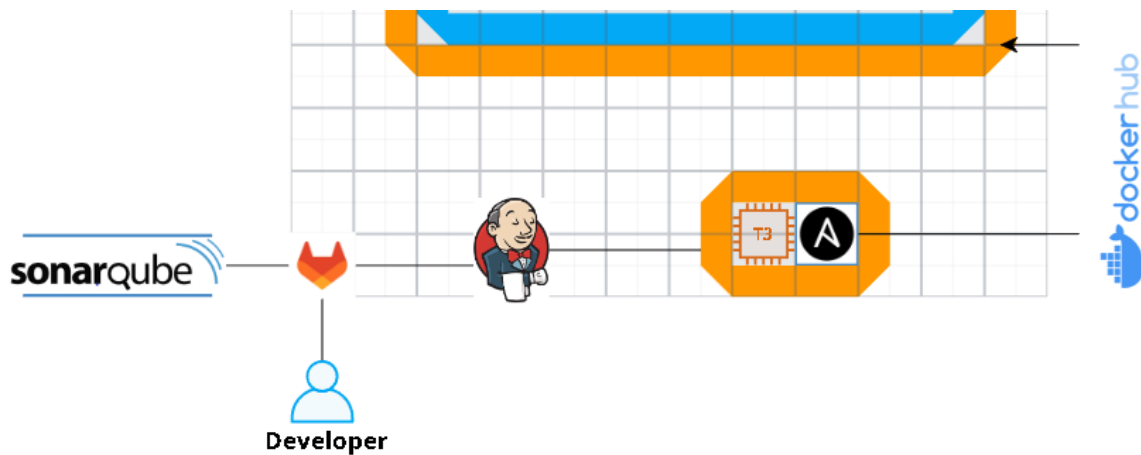
웹 어플리케이션 서버

- Config, Discovery, Api gateway 서버로 각 Micro service가 설정과 연결성을 가짐
 - Config server : 각 마이크로 서버의 환경 설정값을 가짐
 - Discovery server : MSA에서 각각의 서비스에 대한 정보를 저장하여 외부에서 서비스 호출 시 그에 맞는 서버로 요청을 전달해주는 미들웨어
 - ApiGateway server : 마이크로 서비스 로드밸런서, 라우팅, 리버스 프록시 등 여러 공통 모듈 및 관리 역할을 담당하는 서버
- Trade, News, User 서비스
 - Trade server : 하둡 분산처리로 적재된 데이터를 활용해서 API 구현
 - News server : 뉴스 크롤링 데이터를 텍스트 마이닝 형태로 서비스 제공
 - User server : 데이터 차트에 대한 보고서를 사용자 별 관리

웹 클라이언트

- 가상 웹 서버, 리버스 프록시 역할을 위해서 nginx 환경
 - OpenSSL 설정으로 HTTPS 프로토콜 사용
- React 기반으로 SPA 서비스 제공

서버 배포 파이프라인



기본 환경

- Gitlab 코드의 변경사항에 맞춰 빌드 및 배포
- Sonarqube
 - 코드 내용을 통해서 주기적으로 버그 리포팅 → Jira 모니터링 가능

배포 환경

- Jenkins
 - 소스 코드의 지속적인 통합과 지속적인 배포를 자동화해주는 도구
 - CI/CD 파이프라인을 설계하고 과정에 대한 모니터링이 가능
- Ansible
 - 프로비저닝, 구성 관리 등을 자동화해주는 도구
 - CI/CD 과정에서 변경사항을 확인해 배포 단계를 최적화할 수 있음.

- *멱등성 : 연산을 여러 번 적용해도 결과가 달라지지 않는 성질
- Dockerhub
 - public registry로 다양한 유저와 공유하는 이미지 저장소
 - 프로젝트의 이미지를 저장할 수 있음