

Codestates & AFI R&D OT

기업 협업 프로젝트 개요

E. yongs@afidev.com

A. 서울시 성동구 아차산로17길 49 생각공장 데시앙플렉스 1217호



CONTENTS

목차

01. 회사소개

뒤끝 소개

02. 프로젝트

프로젝트 목표 | 인프라 현황
기대하는 결과물 | 데이터 구조

03. 기타

연락 채널, Q&A 등.



회사소개

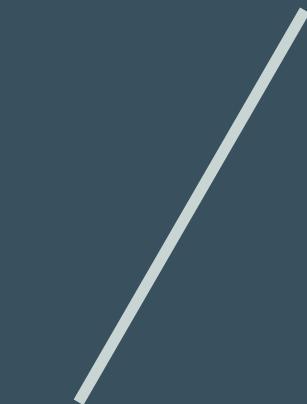
ABOUT COMPANY



뒤끝(THEBACKEND)

프로젝트

ABOUT PROJECT



- 01 개요
- 02 뒤끝 시스템 현황
- 03 데이터 구조



ABOUT PROJECT

데이터 압축 파이프라인 구축

Building a data compression pipeline



데이터 압축 알고리즘

대용량 로그 데이터를 실시간으로 빠르게 압축하는 알고리즘



데이터 저장 및 로드 파이프라인

압축된 데이터를 저장하고, 다시 복원하여 로드하는 파이프라인

PROJECT OVERVIEW

프로젝트 개요

뒤끝에는 방대한 양의 데이터가 오가고 있습니다.
이들을 활용하고 싶지만, 규모가 커서 다루기가 힘듭니다.
이들을 효율적으로 다룰 수 있는 파이프라인이 필요합니다.

방대한 로그 데이터

기능 트랜잭션 로그만
하루 200GB



ML 연구의 필요성

뒤끝의 데이터로
ML R&D가 진행 중



실시간성 파이프라인

저장 -> 후처리가 아닌
실시간 압축 필요



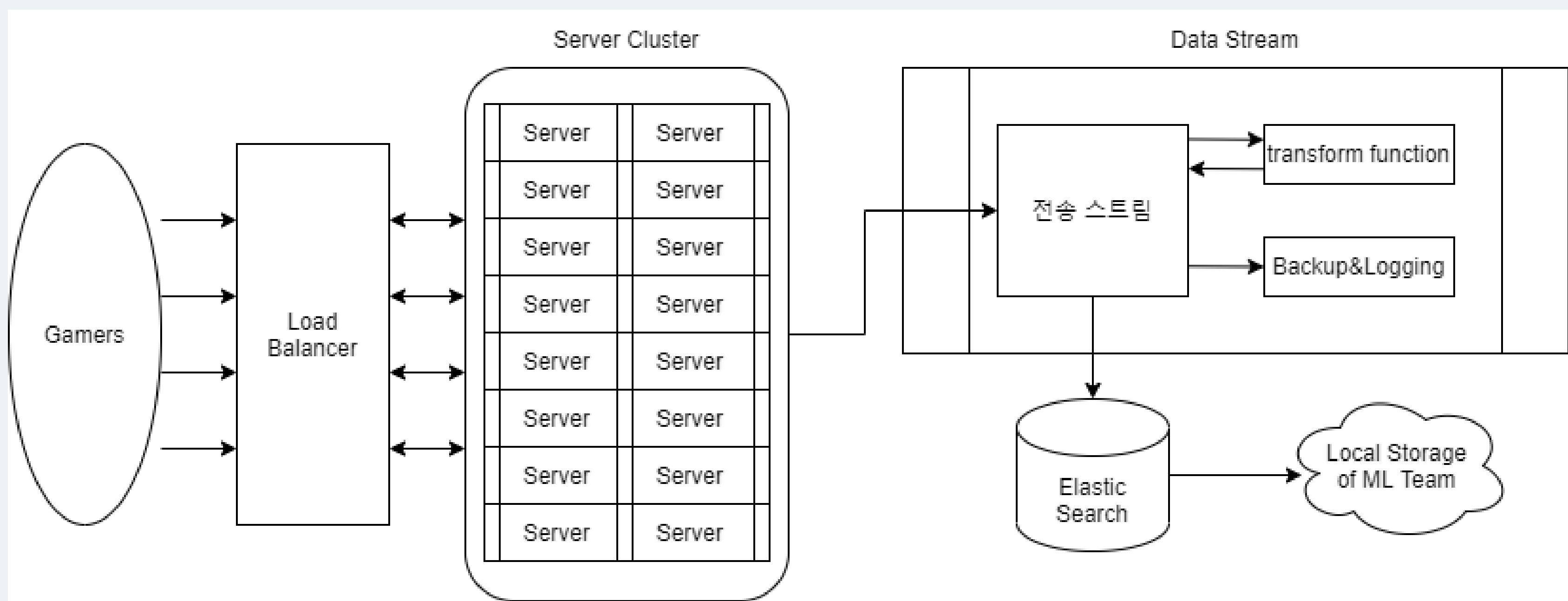
1. 실시간으로 입력되는 로그 데이터를 압축하여 저장하자.
2. 이를 ML Engineer 및 서빙 머신이 쉽게 로드할 수 있게 하자.

INFRASTRUCTURE

뒤끌 시스템

뒤끝 AWS 인프라에 기반한 서비스입니다.

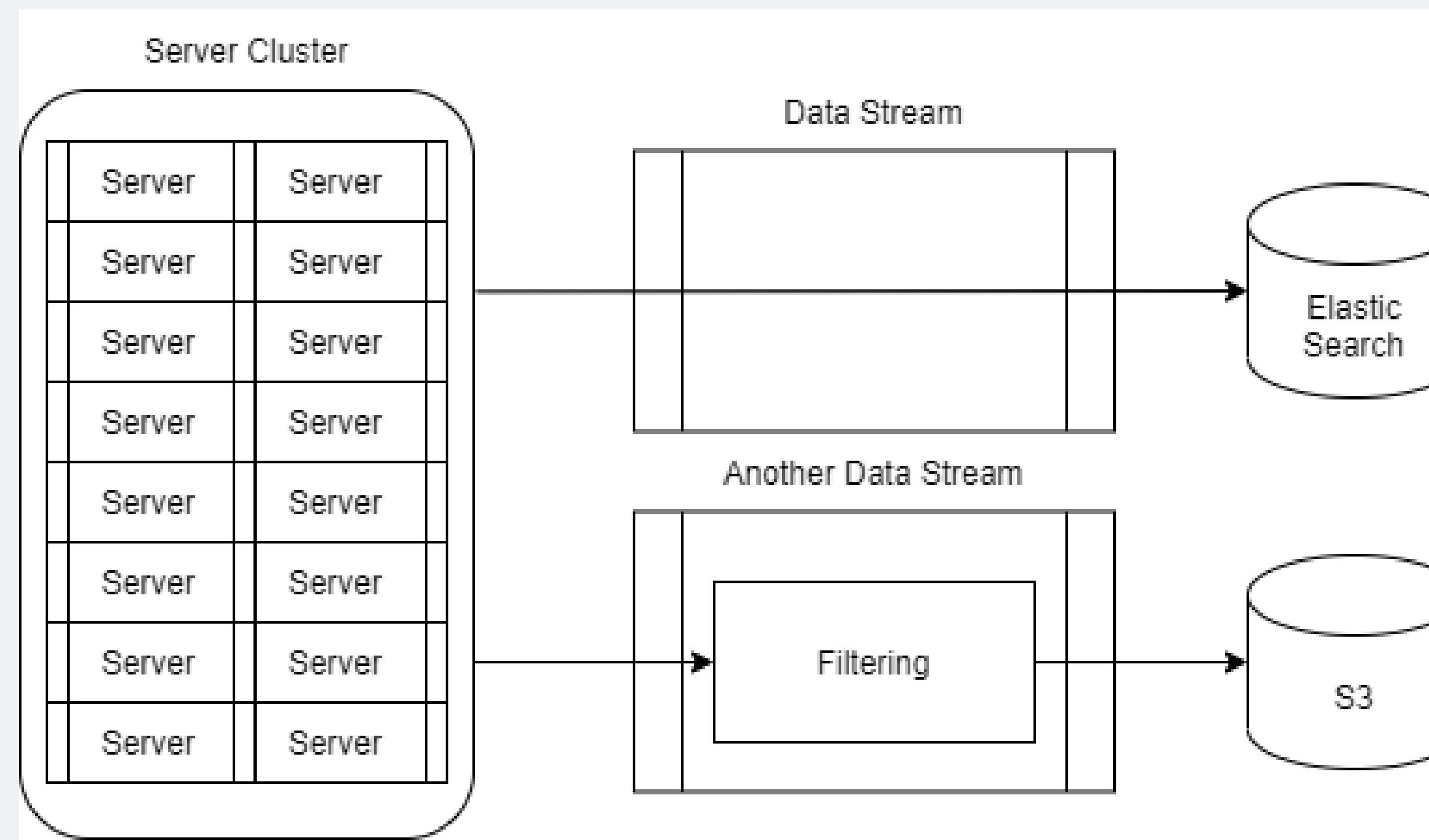
현재 AFI의 ML팀은 비효율적으로 데이터를 ETL하고 있습니다.



INFRASTRUCTURE

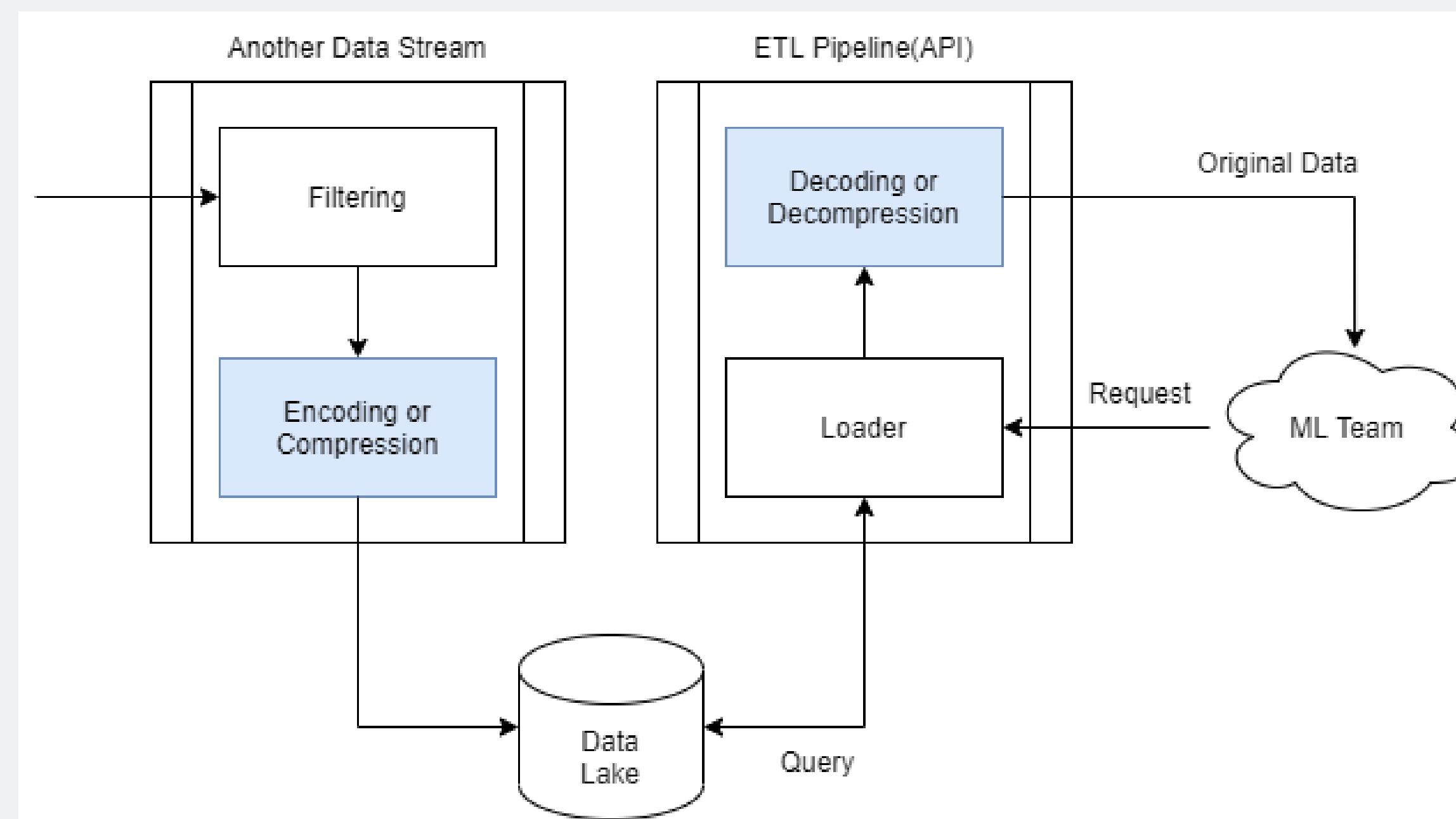
기대 결과 1안

새로운 데이터 스트림을 도입합니다.



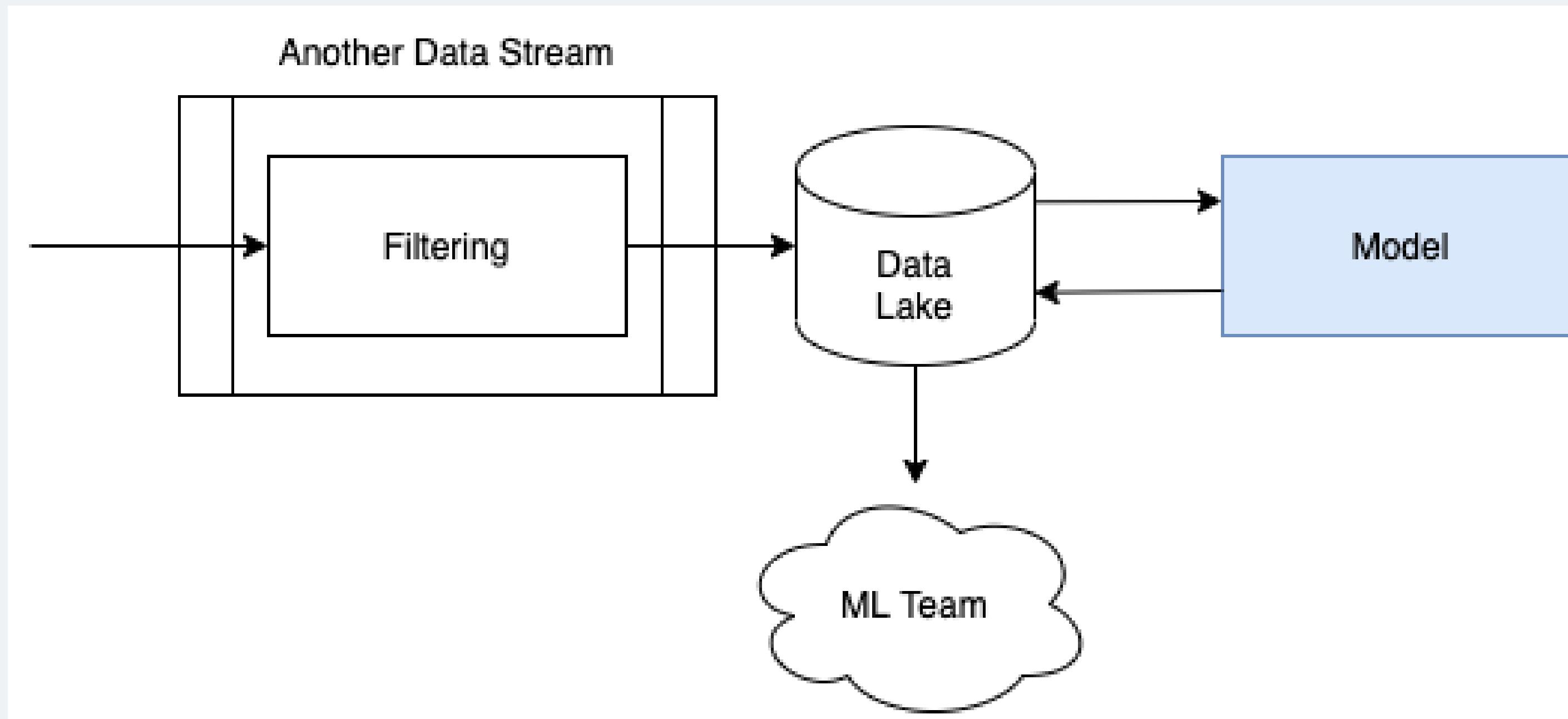
기대 결과 1안

압축 및 복원부를 작성합니다.



INFRASTRUCTURE
기대 결과 2안

실시간 스트림이 아닌 외부 스케줄러에 의한 압축&인코딩을 진행합니다.
그 결과를 다시 데이터 레이크에 저장합니다. 모델을 도입해도 됩니다.



DIRECTION

방향성

다음의 사항들이 이 프로젝트의 메인 포인트입니다.

효용

Effectiveness

효율

Efficiency

복원이 정확하게 되거나

VecRepr의 표현력이 우수하거나

적절한 FeatureEng를 도입하거나

처리 시간이 빠르고

저장 용량이 매우 적고

데이터 로드&복원 비용이 낮고

DATA STRUCTURE

데이터 구조

아래와 같은 OLTP 로그가 발생하고 있습니다.

Expanded document

Table JSON

```
{
  "_index": "sdk-log-2021-12-26",
  "_type": "_doc",
  "_id": "49623735927354776182646864847313121695850829503304564770.0",
  "_version": 1,
  "_score": null,
  "_source": {
    "client_app_id": "e085f380-dd96-11eb-a904-599414fa8c5e2319",
    "func_name": "Backend.GameData.TransactionWrite",
    "nickName": "호랑이들어와요",
    "request_begin_timestamp": "2021-12-26T15:10:59.884+09:00",
    "request_end_timestamp": "2021-12-26T15:10:59.966+09:00",
    "request_url": "https://api.thebackend.io/data/gameinfo/v3.1/union/transactionWrite",
    "response_elapsed_time": 81,
    "response_reason": "Success",
    "response_status_code": 204,
    "sdk_version": "5.4.5",
    "timestamp": "2021-12-26T15:11:04Z",
    "user_inDate": "2021-08-22T07:51:26.636Z"
  },
  "fields": {
    "request_begin_timestamp": [
      "2021-12-26T06:10:59.884Z"
    ],
    "user_inDate": [
      "2021-08-22T07:51:26.636Z"
    ],
    "request_end_timestamp": [
      "2021-12-26T06:10:59.966Z"
    ],
    "timestamp": [
      "2021-12-26T15:11:04.000Z"
    ]
  },
  "sort": [
    1640531464000
  ]
}
```



THANK YOU

Q&A Time

www.afidev.com
www.thebackend.io
©AFI RIGHTS RESERVED.