



# 국내 최대 이미지 데이터를 서비스 하는 네이버 이미지검색 시스템의 10여 년간의 DevOps 변화와 미래

김인겸, 주혜원  
NAVER

# CONTENTS

1. 네이버 검색시스템 소개
2. 10년 전 이미지검색 시스템
3. 변화하는 이미지 검색 DevOps
4. 현재의 이미지검색 DevOps
5. 앞으로의 DevOps 변화
6. QnA

# 1. 네이버 검색시스템 소개

# 1.1 네이버 검색시스템이란?

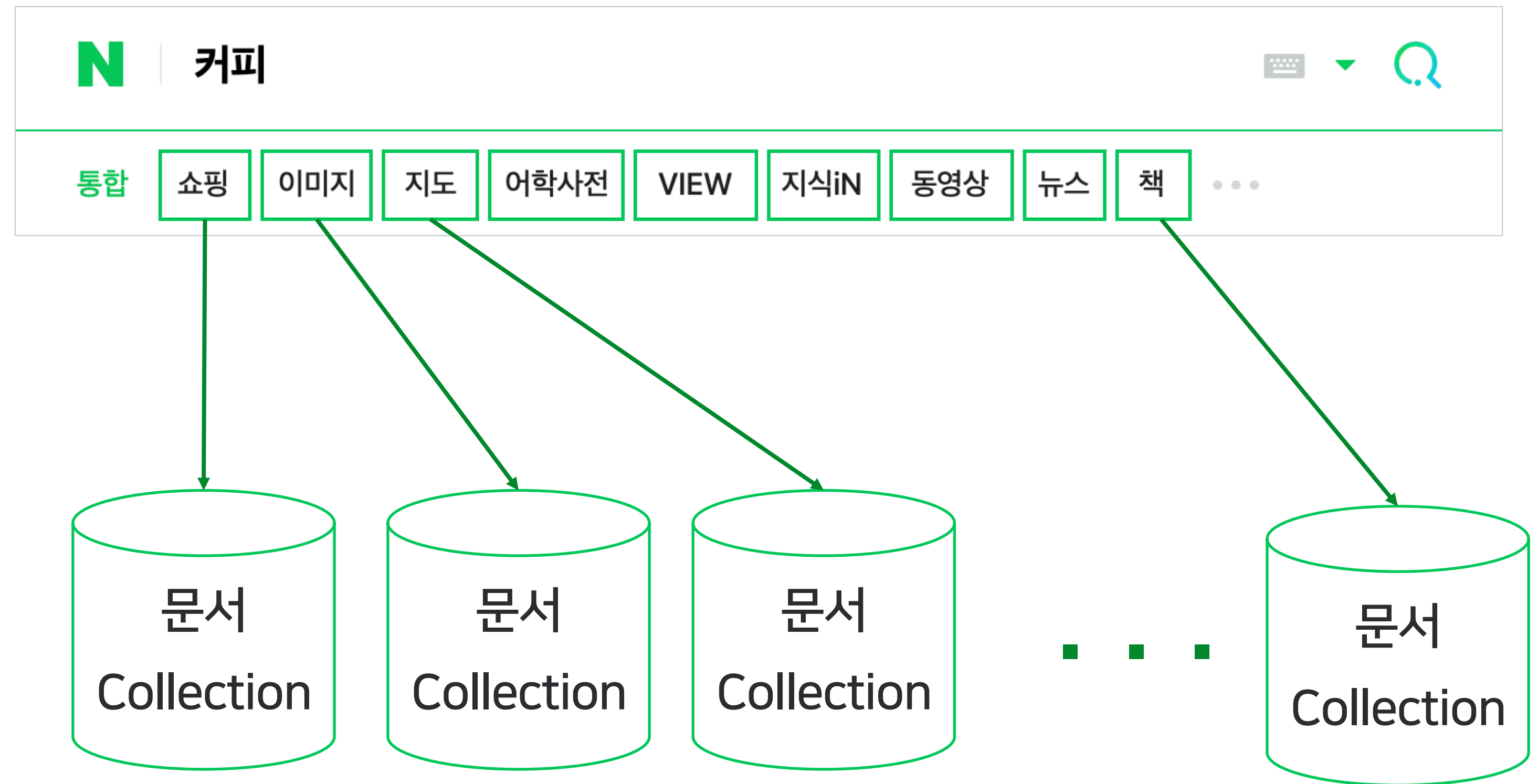
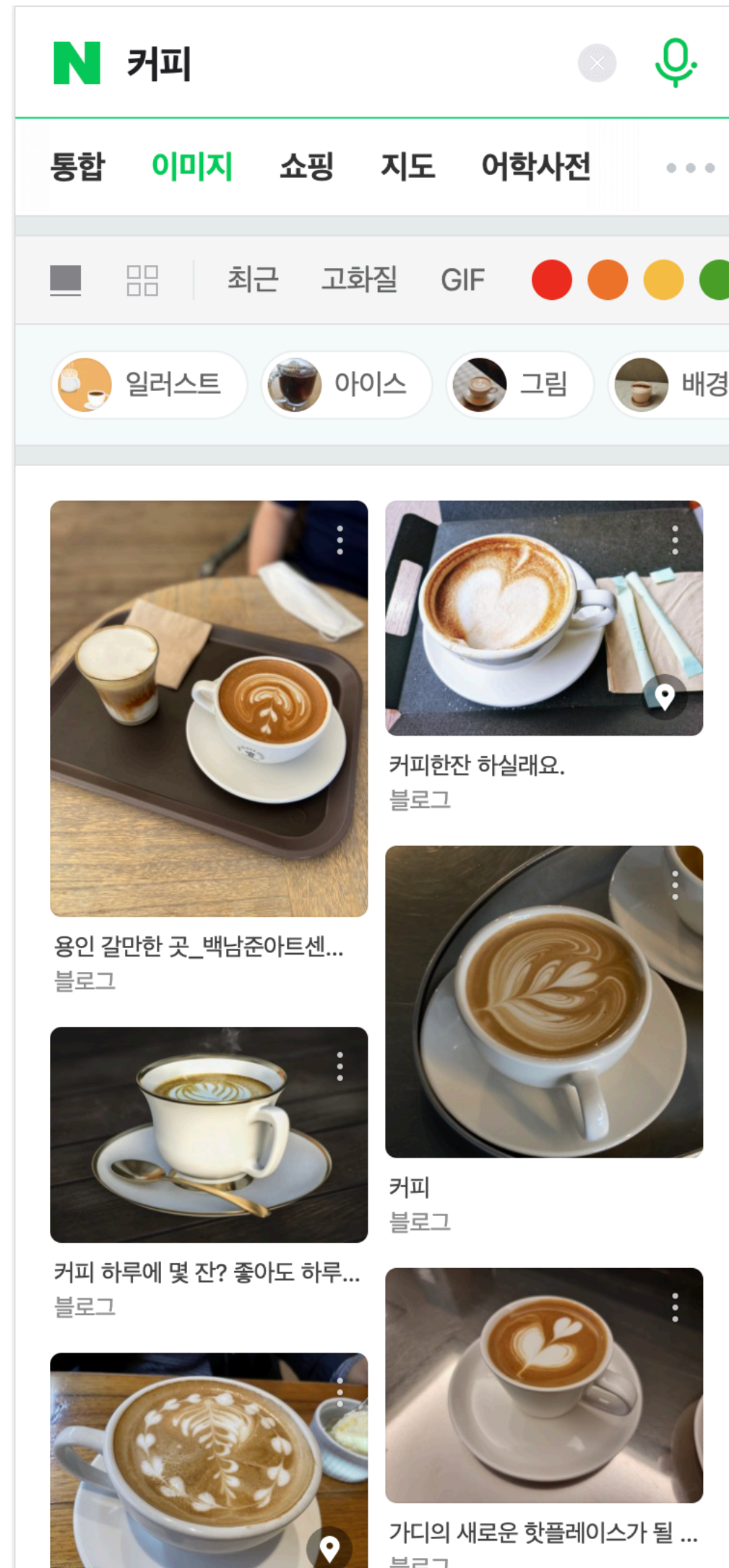
# NAVER

네이버 검색 시스템 DevOps

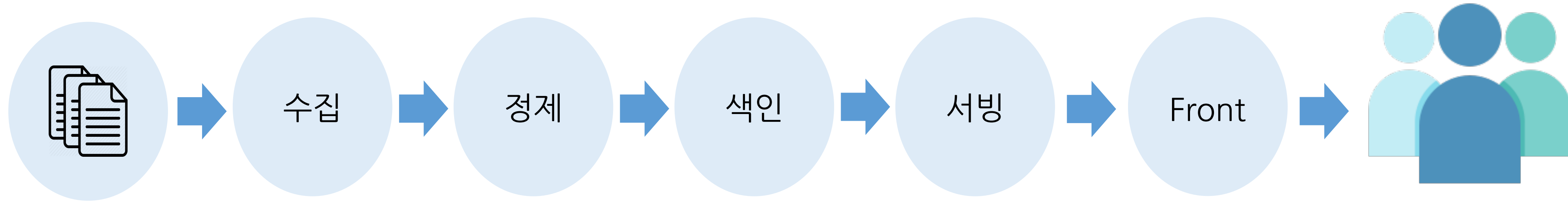




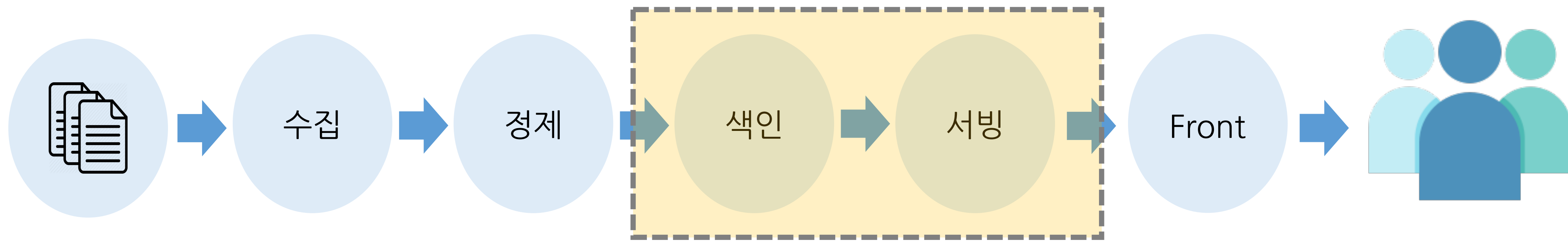
# 1.1 네이버 검색시스템이란?



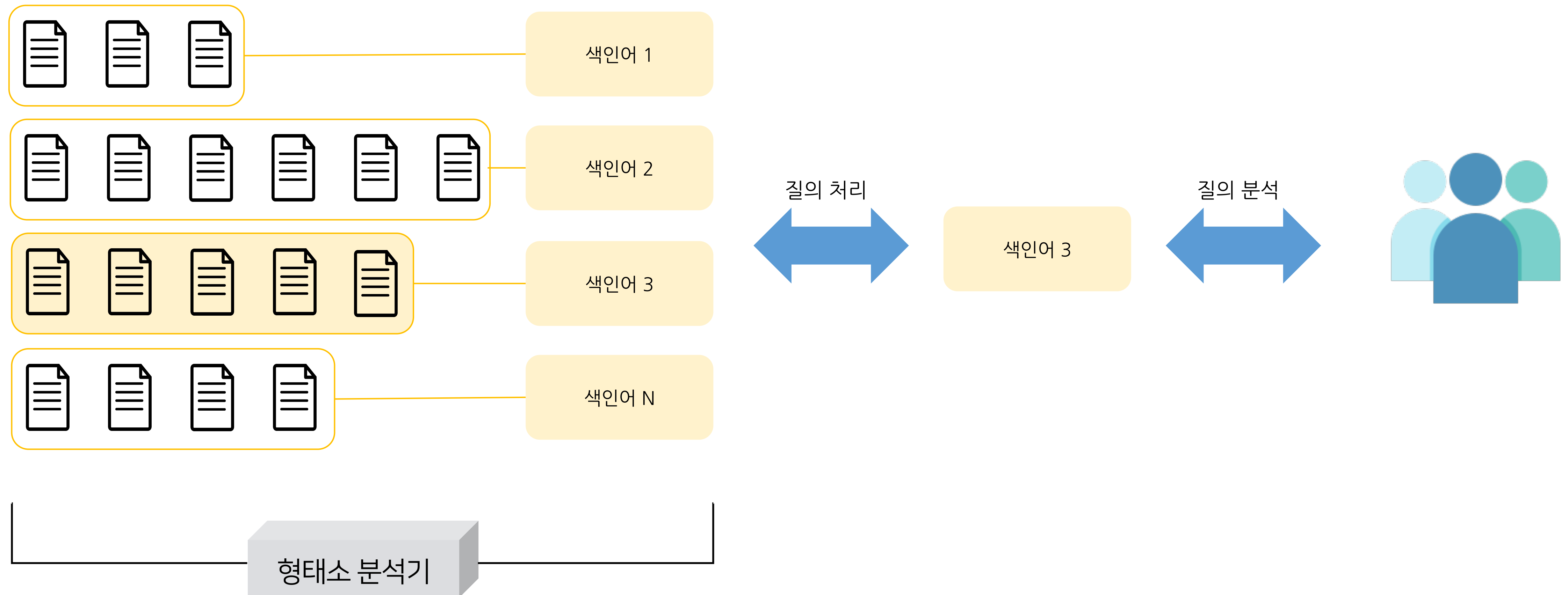
# 1.1 네이버 검색시스템이란?



# 1.1 네이버 검색시스템이란?

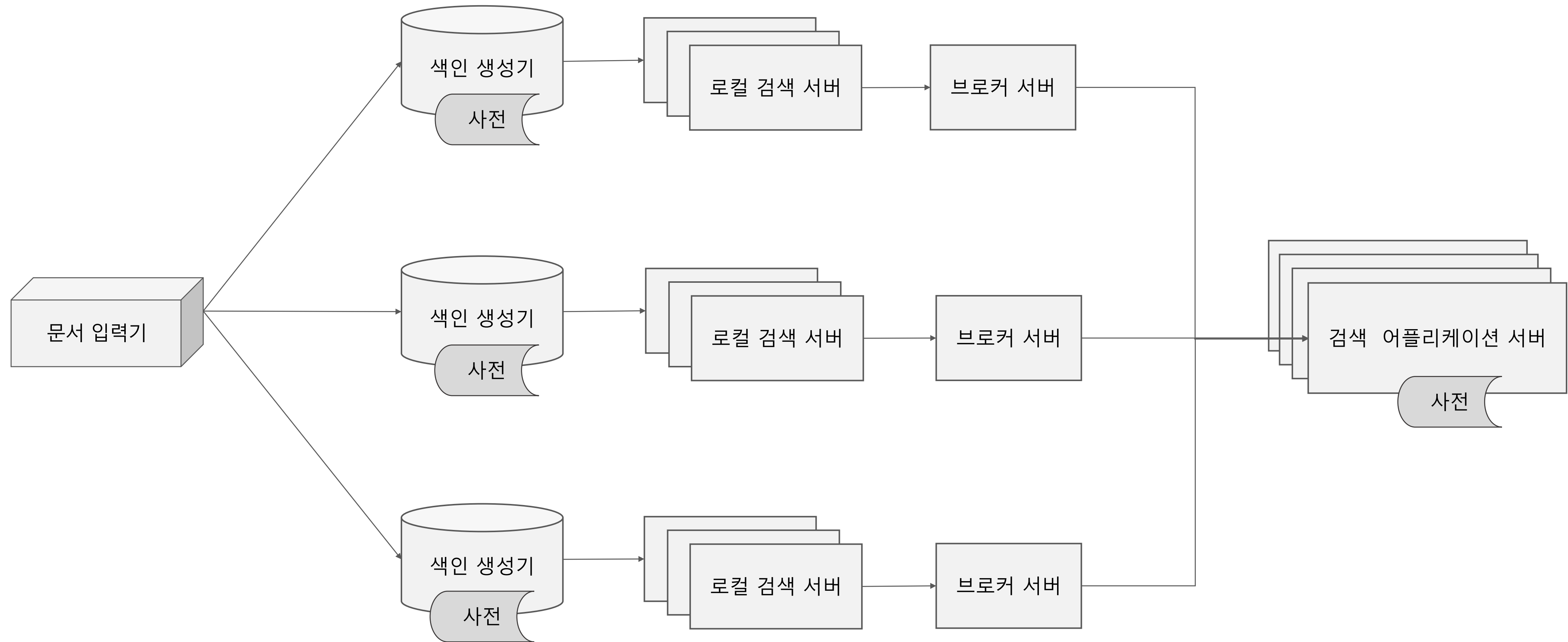


# 1.1 네이버 검색시스템이란?



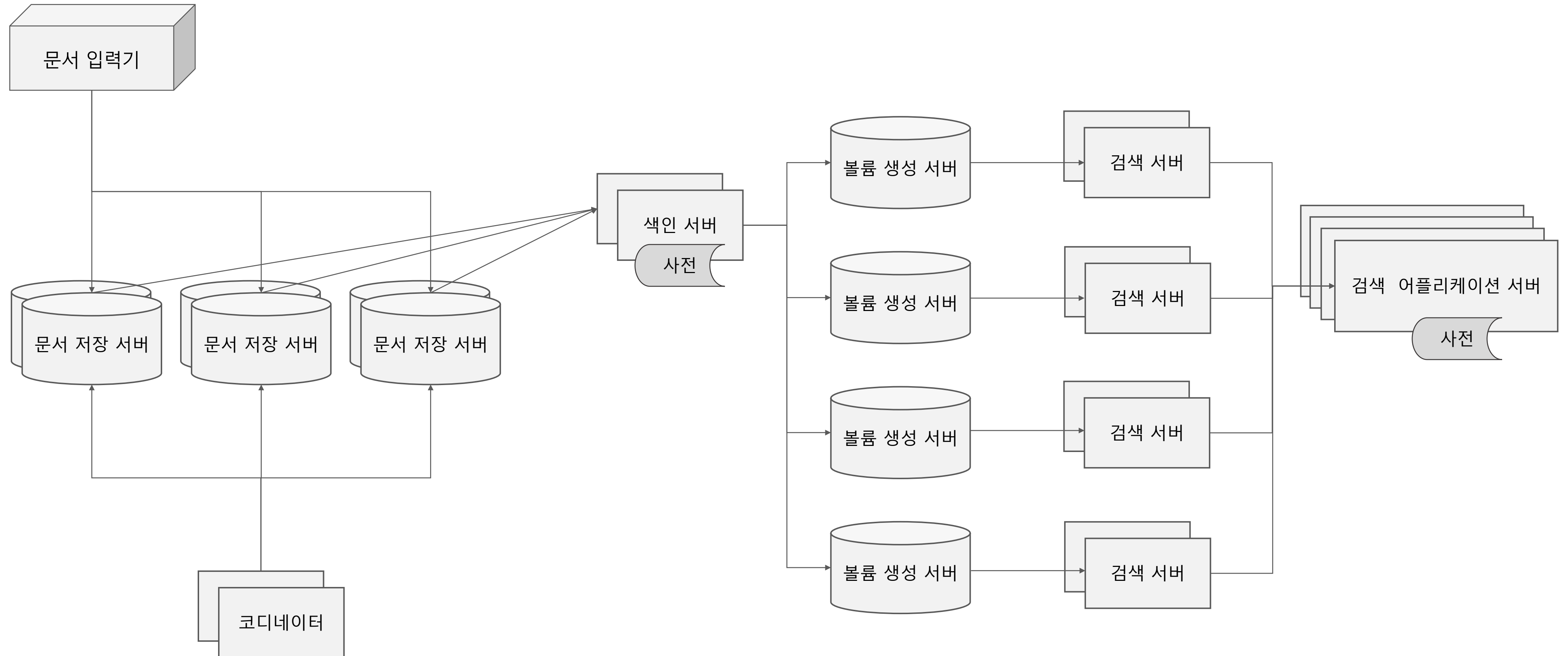
# 1.2 네이버 검색시스템 주요 구성 및 역할

## [배치 검색 엔진]



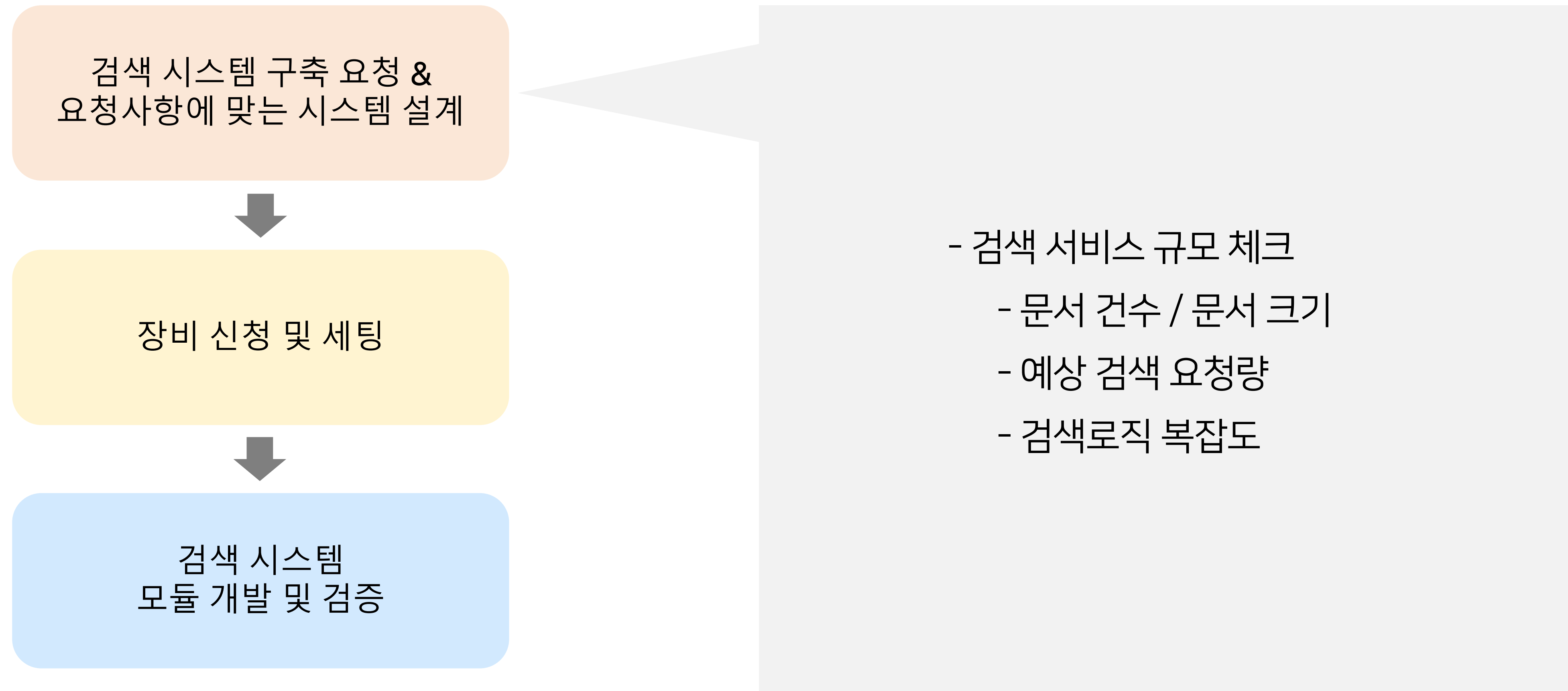
# 1.2 네이버 검색시스템 주요 구성 및 역할

## [증분 검색 엔진]

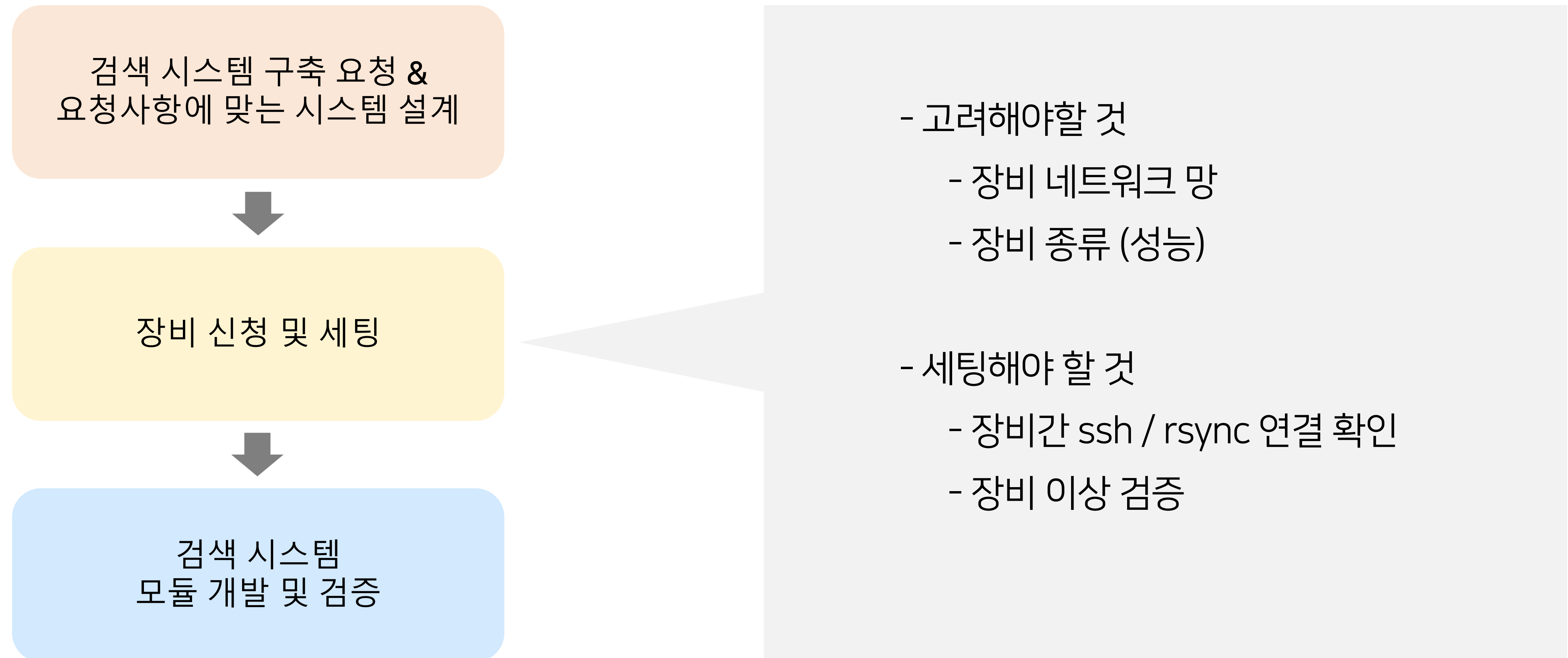




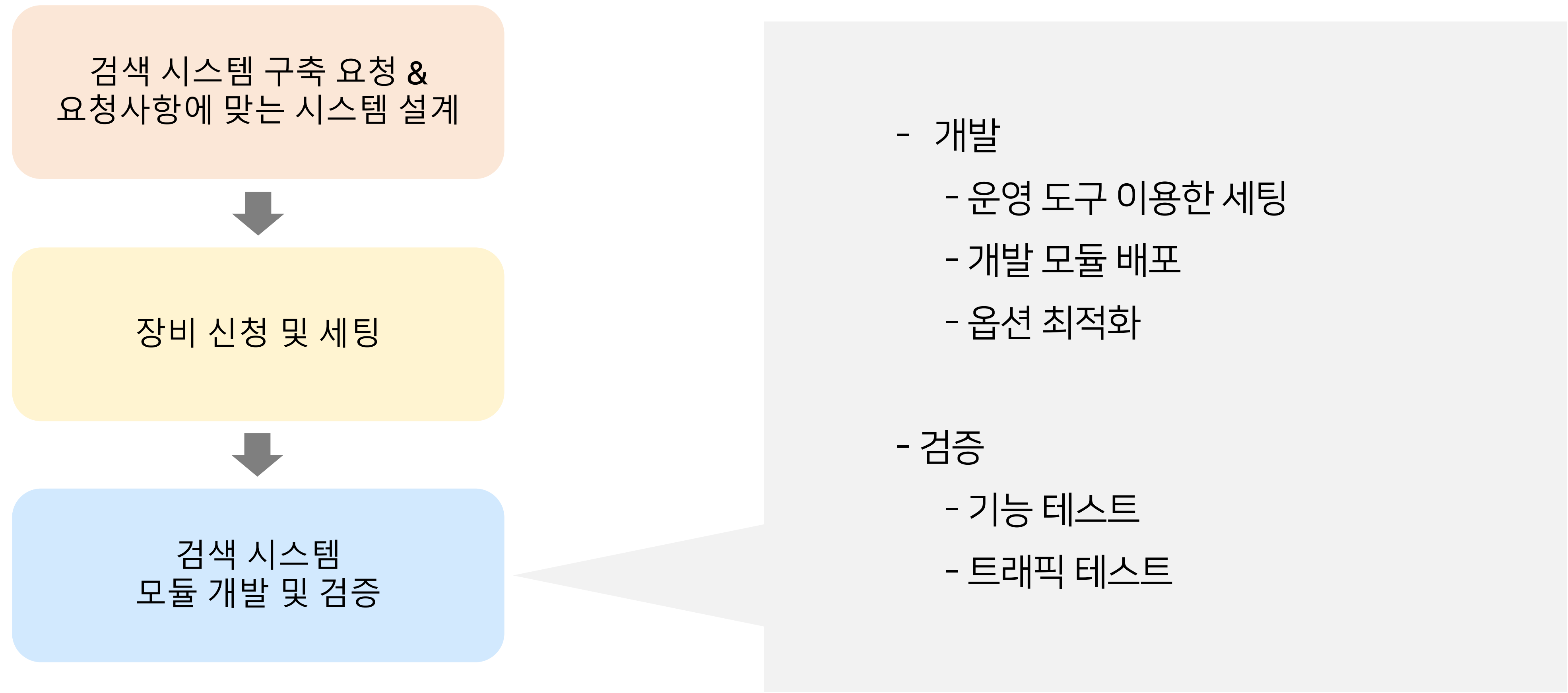
# 1.3 네이버 검색시스템 구축하기



# 1.3 네이버 검색시스템 구축하기



# 1.3 네이버 검색시스템 구축하기



# 1.4 네이버 검색시스템 DevOps 어려운 점

## ✓ 문서 입력 관련 장애

- 신규 문서 반영 지연 : 검색 결과 노출 지연
- 문서 정제 부서와 협업하여 빠르게 해결 필요

## ✓ 장비 장애

- 장비의 역할에 따라 검색 지연 / 결과 미노출 등 다양한 장애 가능성
- 장비 제외 배포 및 장비 복구 요청 필요

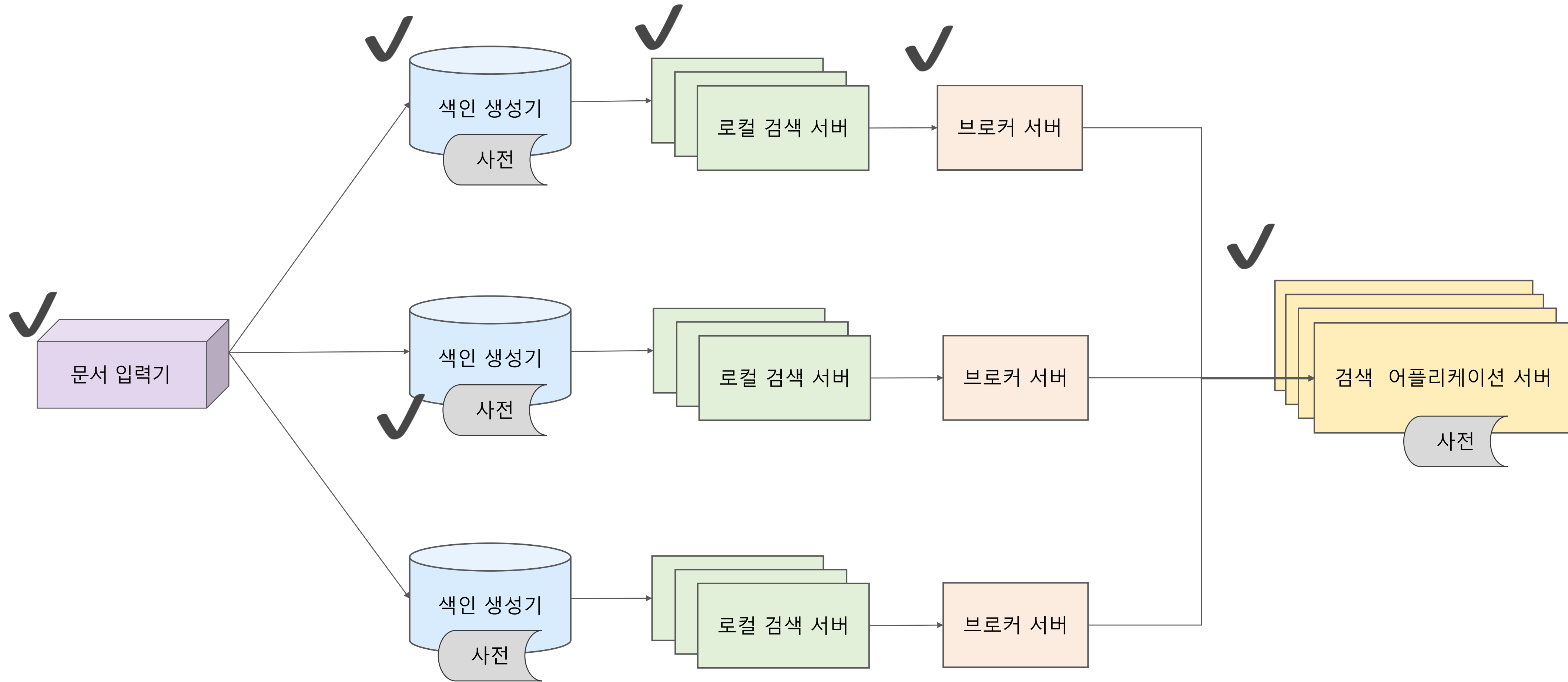
## ✓ 검색 결과 이상

- 원인 분석 및 시급함 따라 대응

## ✓ 안정성 유지

- 24시간 정상 작동
- 기존의 3배 트래픽이 들어와도 안정성 유지 가능해야 함

# 1.4 네이버 검색시스템 DevOps 어려운 점



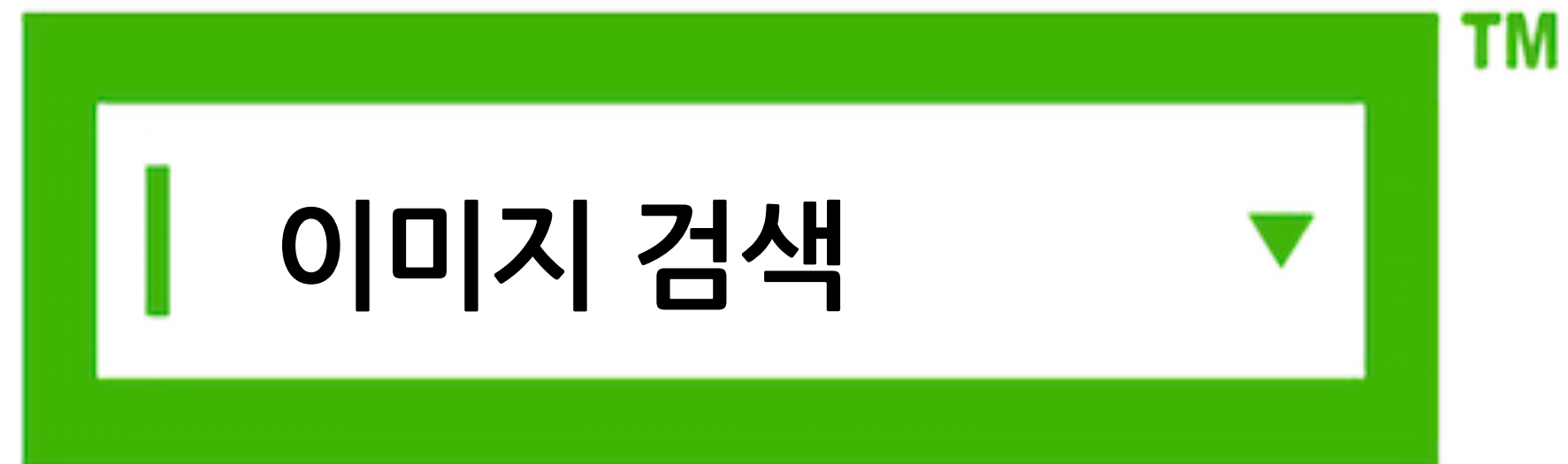
# 1.4 네이버 검색시스템 DevOps 어려운 점





# 1.5 네이버 이미지 검색 소개

**NAVER**



- ✓ 물리 장비 1200대 이상
- ✓ 문서 수 약 60억건
- ✓ 일간 검색량 6억건 이상

지난 10년간

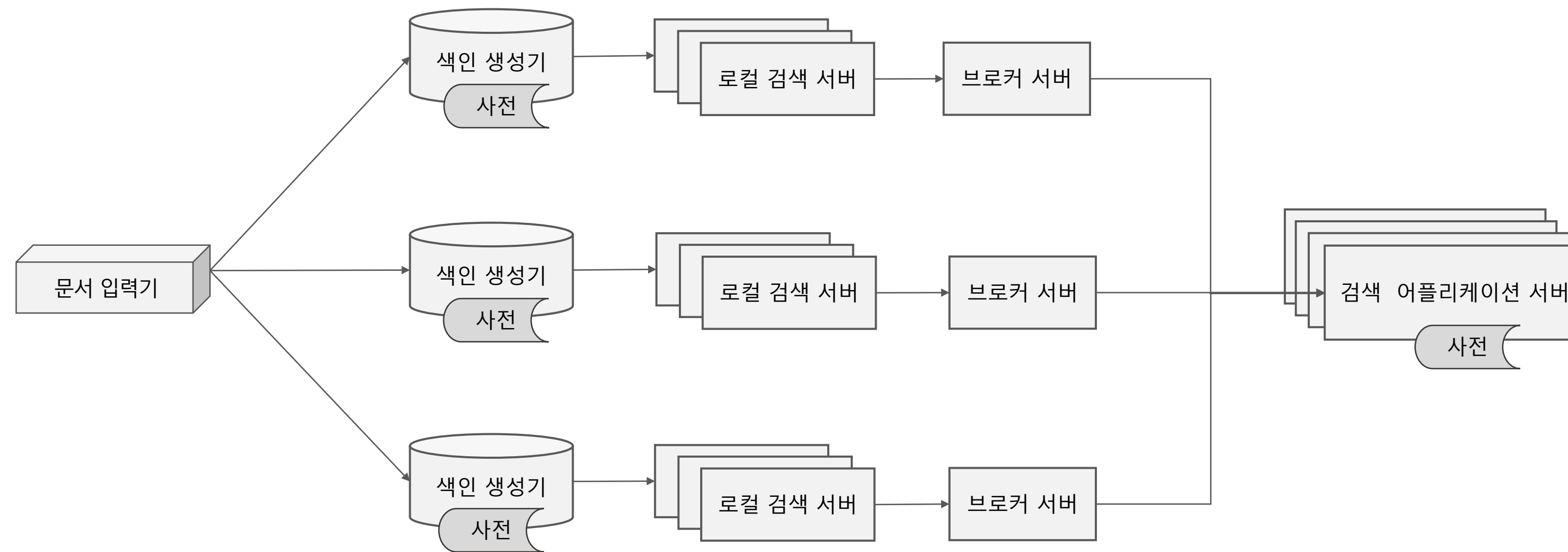
네이버 **이미지 검색**의

**과거와 미래**를 소개합니다.

## 2. 10년 전 이미지검색 시스템

# 2.1 검색시스템 구조

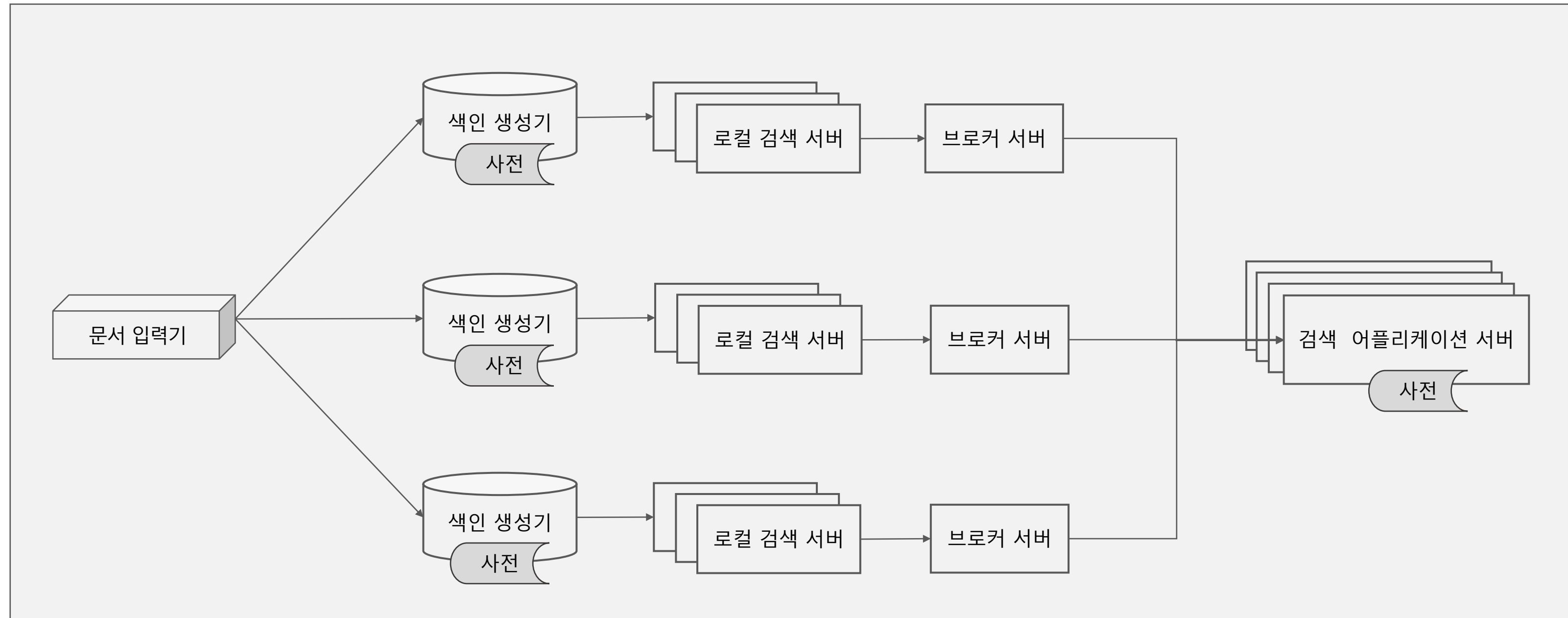
## 기본적인 검색 시스템



# 2.1 검색시스템 구조

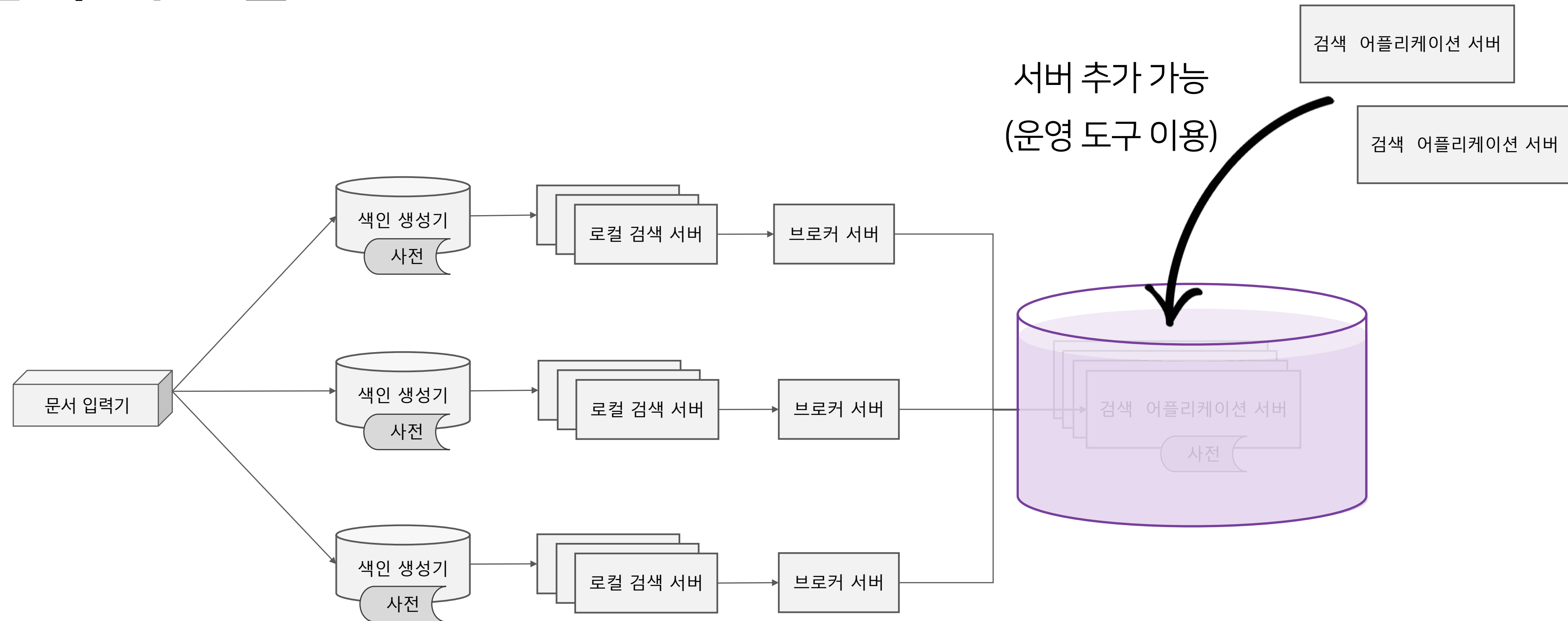
## 기본적인 검색 시스템

Autoserv (검색 시스템 운영 도구)



# 2.1 검색시스템 구조

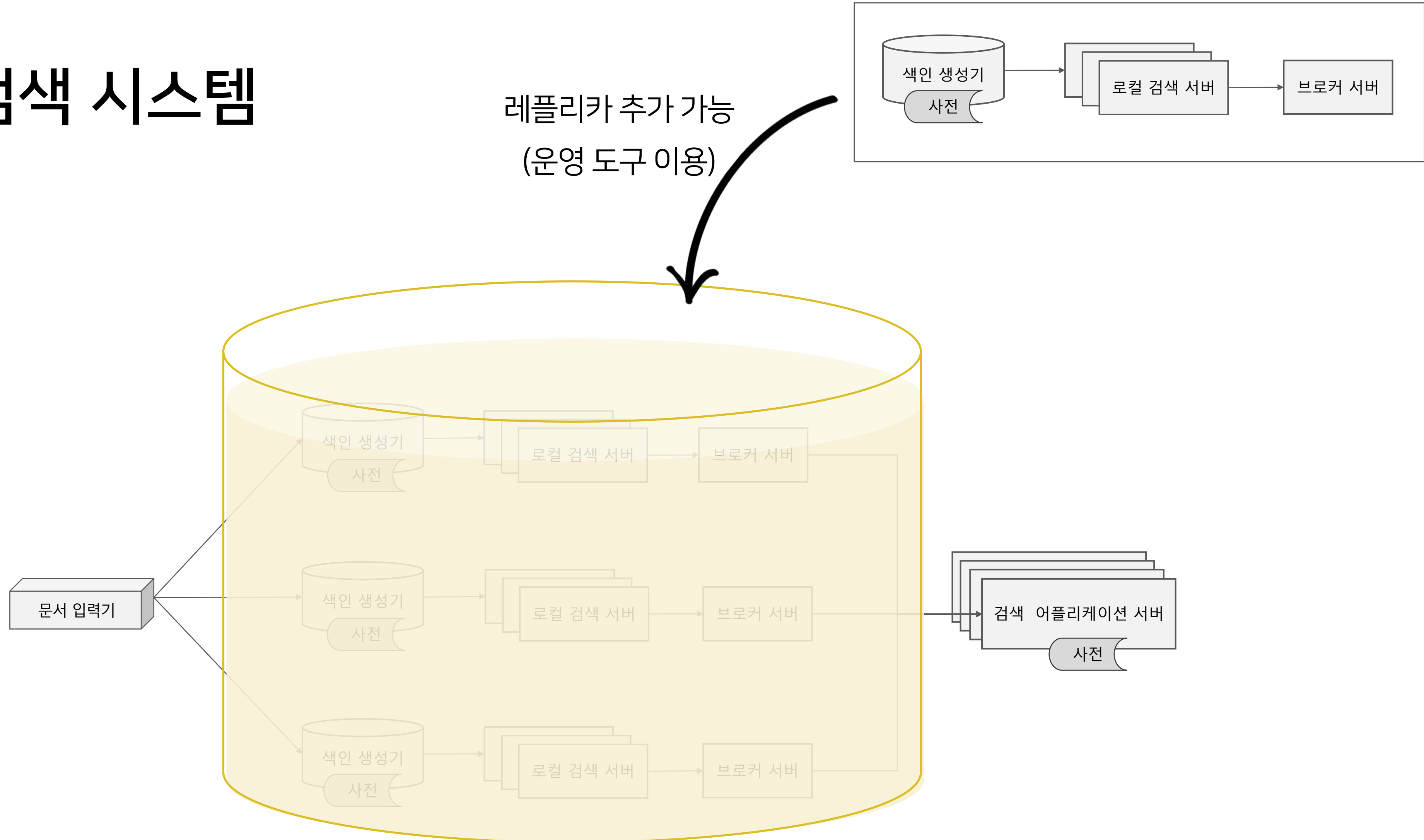
## 기본적인 검색 시스템





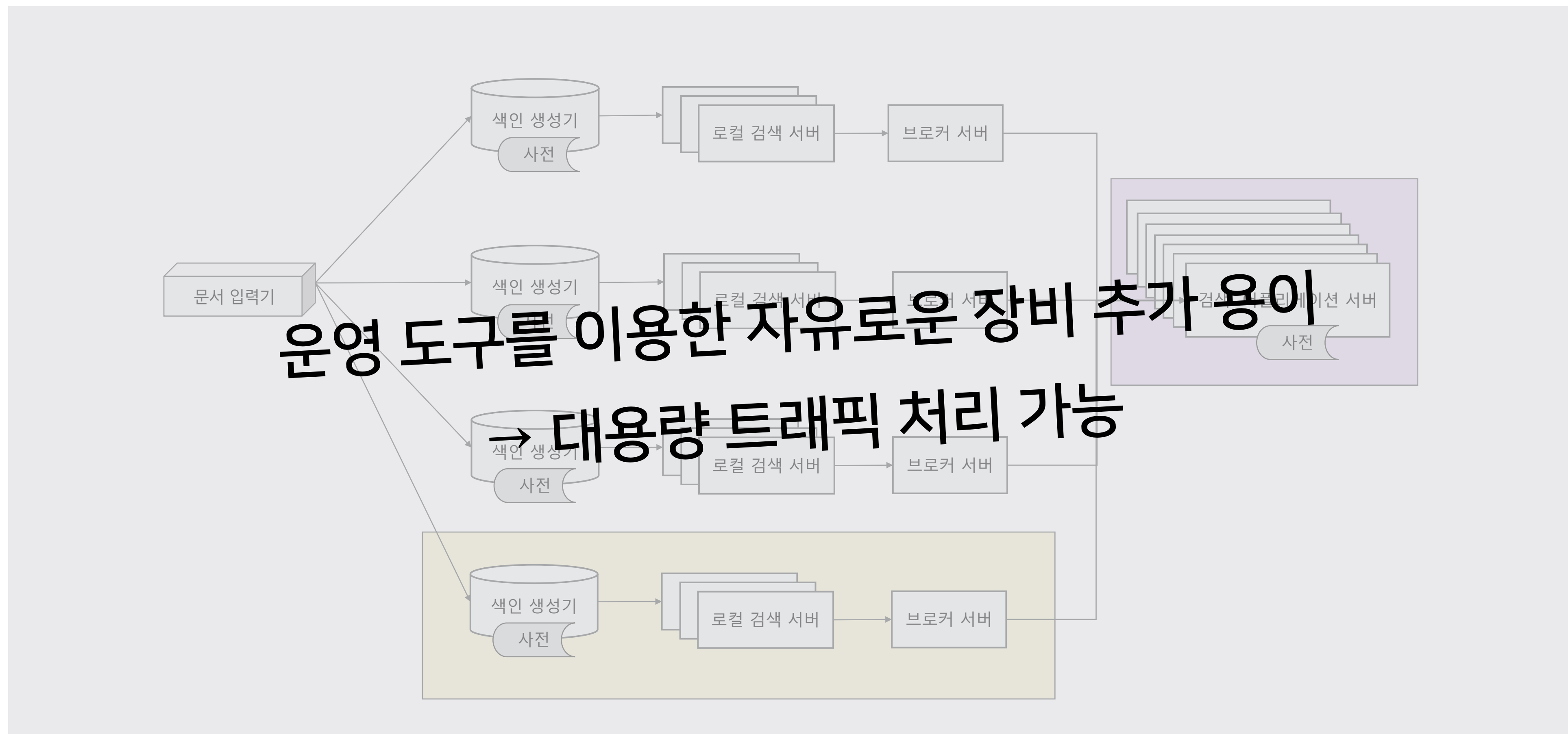
# 2.1 검색시스템 구조

## 기본적인 검색 시스템



# 2.1 검색시스템 구조

## 기본적인 검색 시스템



## 2.2 이미지 검색을 위한 시스템

### 이미지검색을 위해 필요한 시스템들

이미지  
저장소

Thumbnail  
이미지  
생성기

이미지 정보  
추출 &  
중복 이미지  
판독기

노출 제어  
시스템

## 2.2 이미지 검색을 위한 시스템

### 이미지검색을 위해 필요한 시스템들

이미지  
저장소

Thumbnail  
이미지  
생성기

이미지 정보  
추출 &  
중복 이미지  
판독기

노출 제어  
시스템

## 2.2 이미지 검색을 위한 시스템

### 이미지검색을 위해 필요한 시스템들

이미지  
저장소

Thumbnail  
이미지  
생성기

이미지 정보  
추출 &  
중복 이미지  
판독기

노출 제어  
시스템

## 2.2 이미지 검색을 위한 시스템

### 이미지검색을 위해 필요한 시스템들

이미지  
저장소

Thumbnail  
이미지  
생성기

이미지 정보  
추출 &  
중복 이미지  
판독기

노출 제어  
시스템



## 2.2 이미지 검색을 위한 시스템

### 이미지검색을 위해 필요한 시스템들

이미지  
저장소

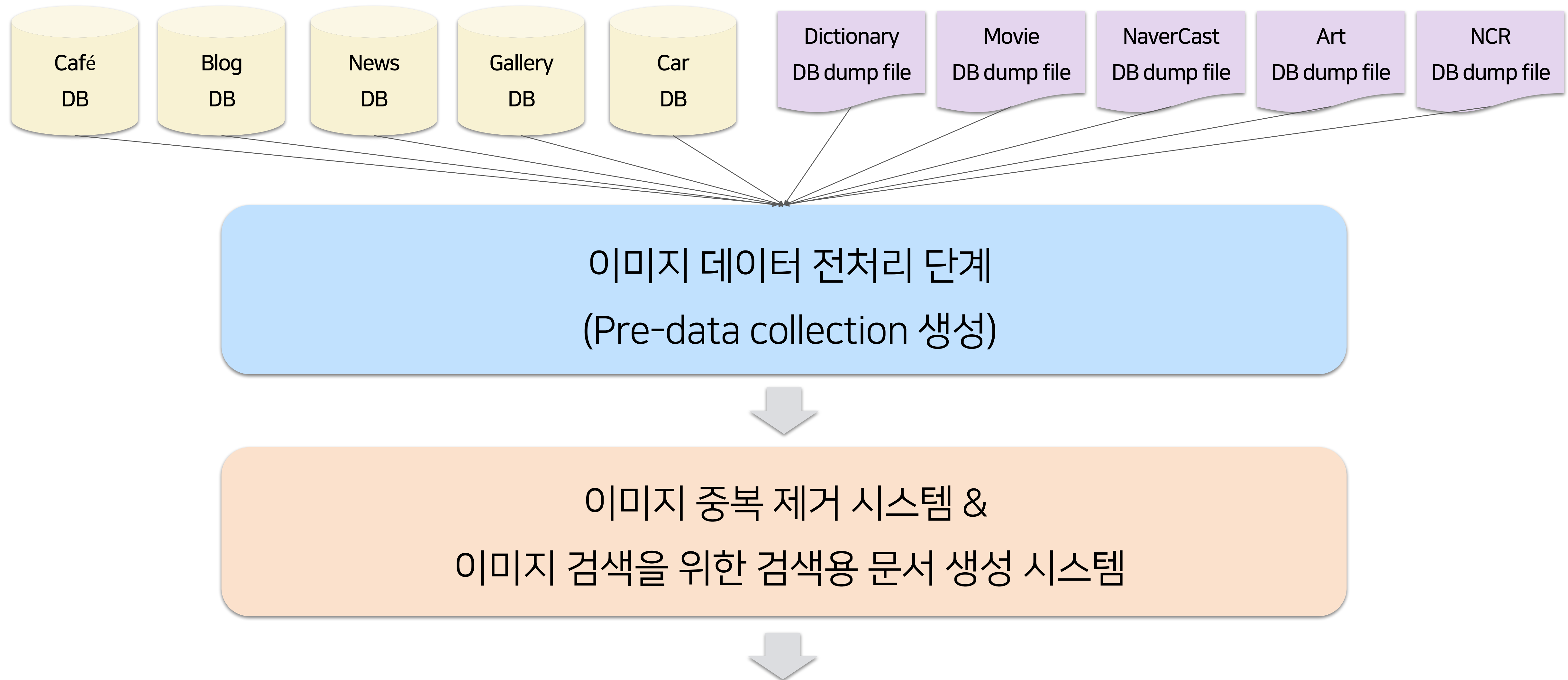
Thumbnail  
이미지  
생성기

개별 시스템으로 독립적 운영

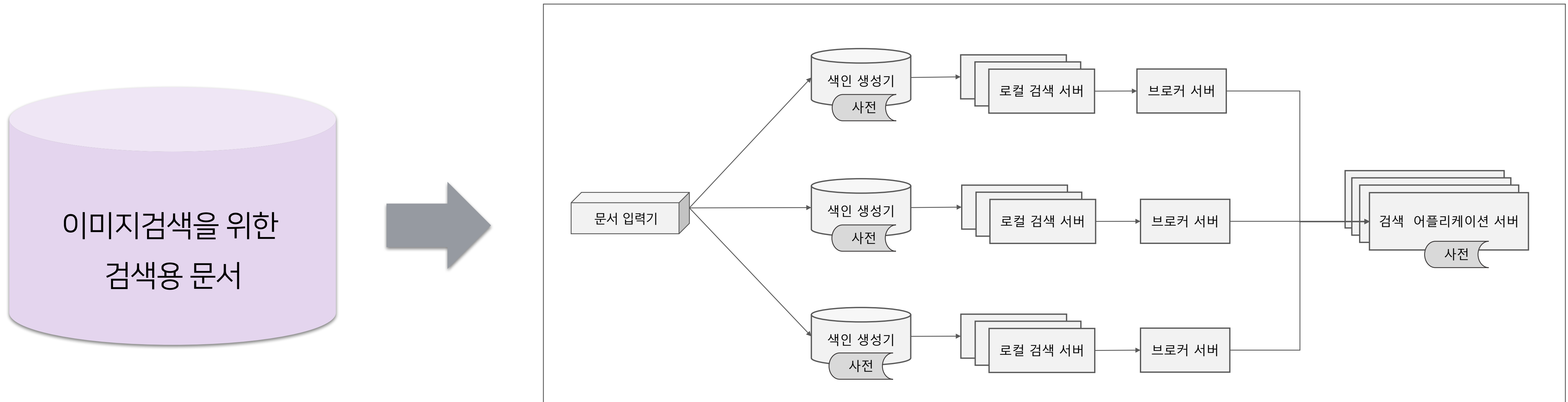
이미지 정보  
추출기  
중복 이미지  
판독기

노출 제어  
시스템

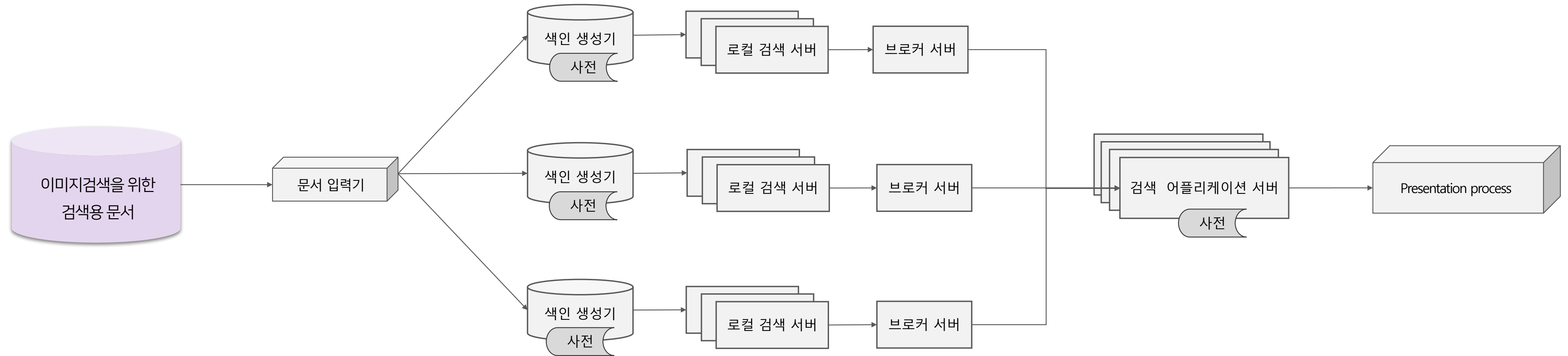
## 2.3 이미지 검색시스템 구조



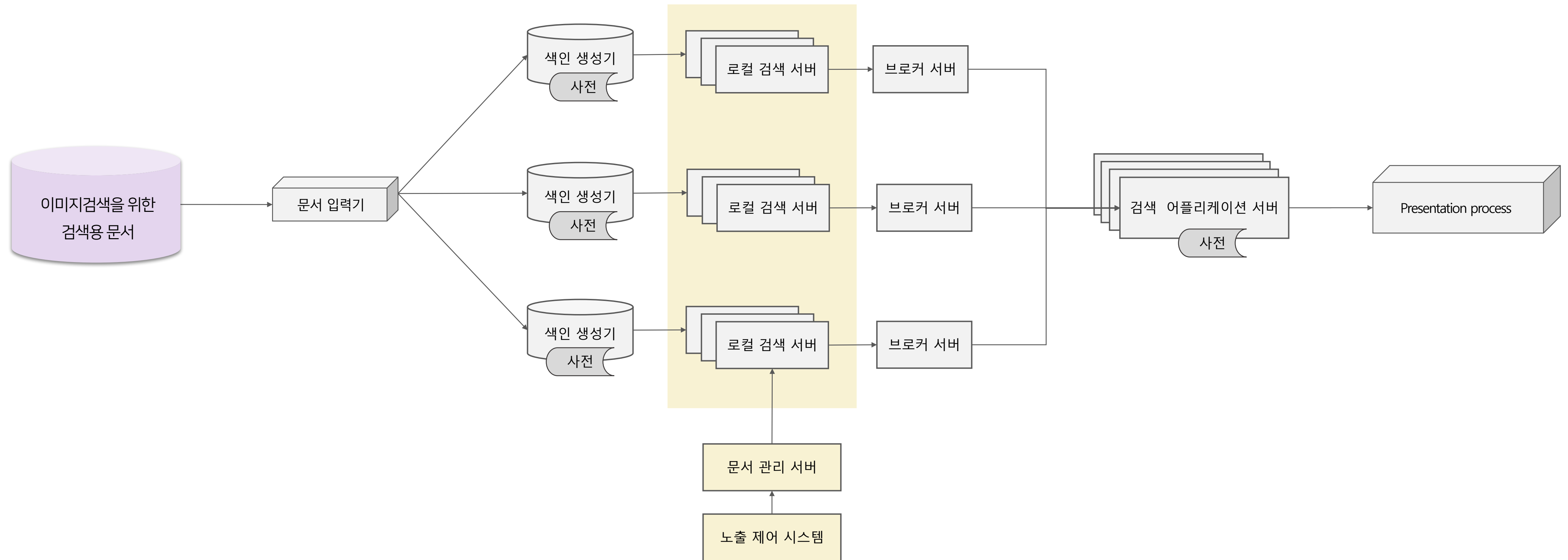
# 2.3 이미지 검색시스템 구조



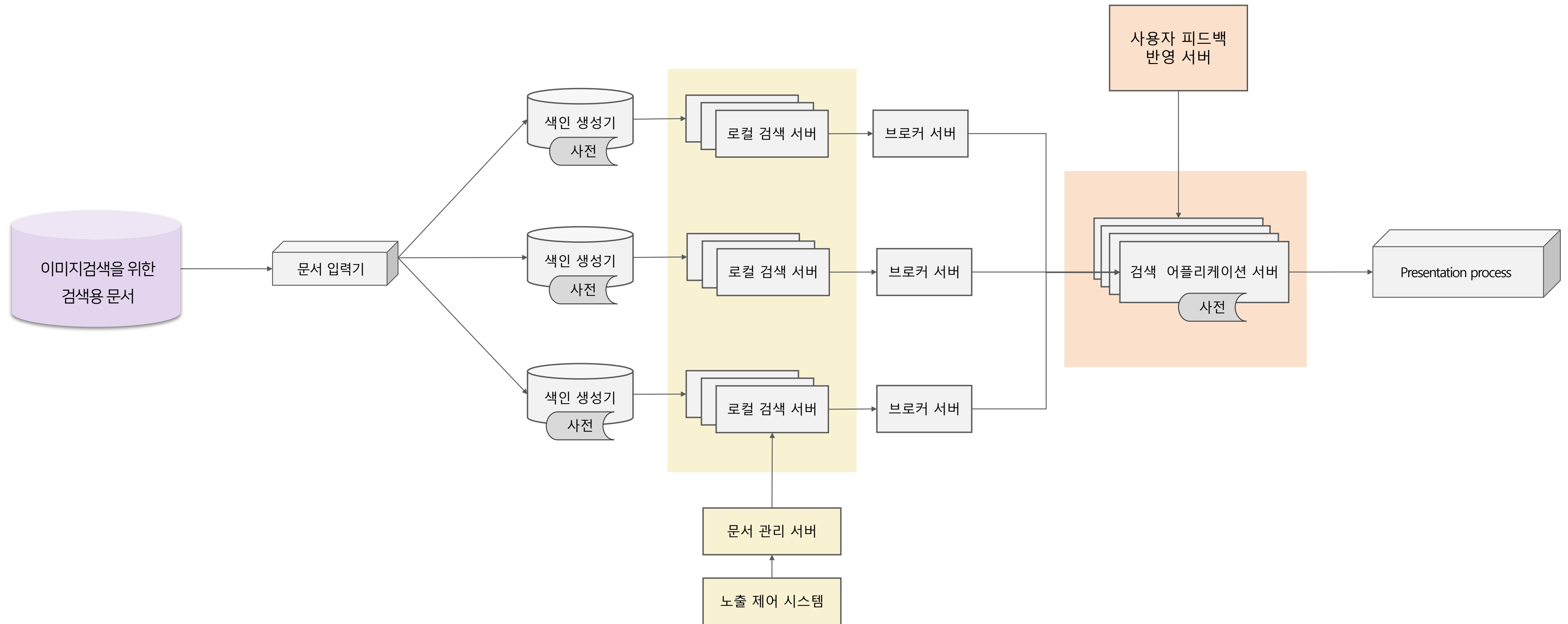
# 2.3 이미지 검색시스템 구조



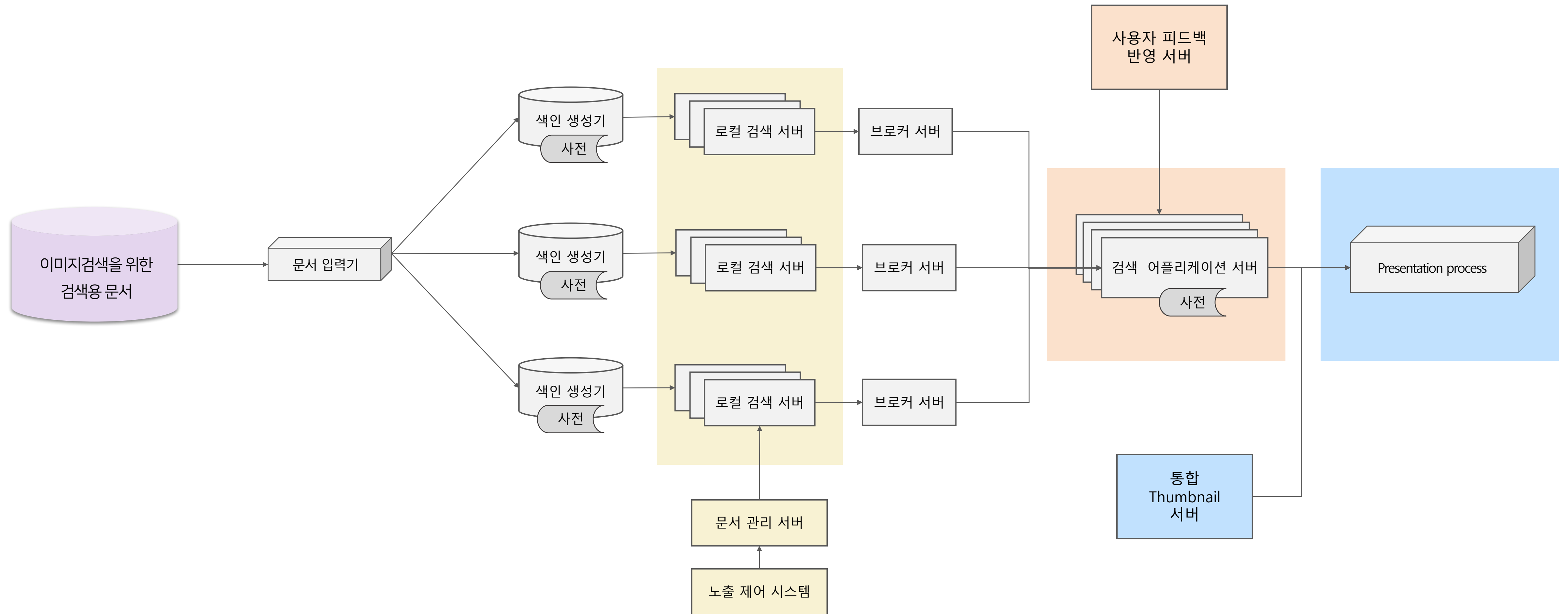
# 2.3 이미지 검색시스템 구조



# 2.3 이미지 검색시스템 구조



# 2.3 이미지 검색시스템 구조



## 2.3 이미지 검색 시스템 구조

### 소스 코드 관리

소스 코드 관리 툴



SVN (SubVersion) 이용  
개발 서버에 1일 단위 백업

코드 리뷰  
✕



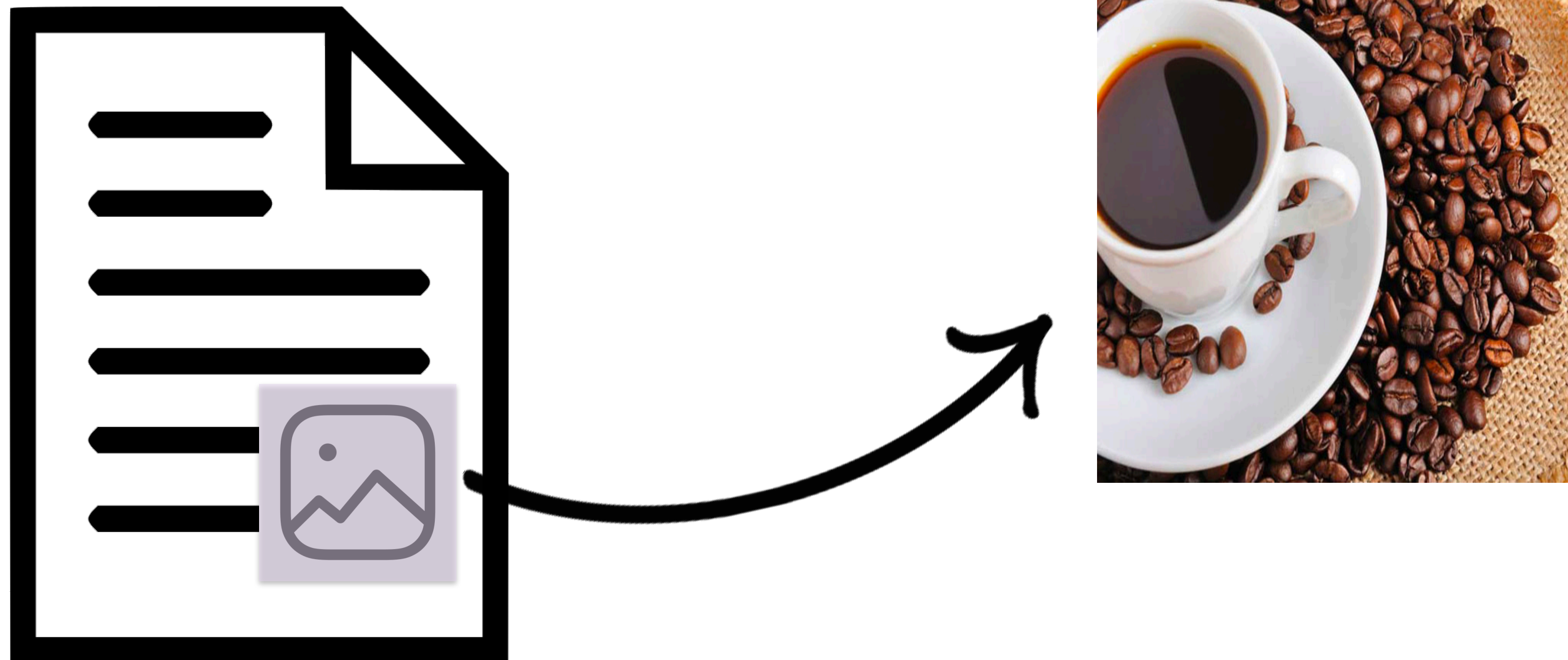
비 정기적 진행 (필요시)  
검색 서비스 담당자 혼자서 개발 및 관리



## 2.4 이미지 검색 랭킹

### 검색 품질에 영향을 미친 주요 요소

1. 이미지가 포함된 문서에서 주요 키워드와 가중치 추출



## 2.4 이미지 검색 랭킹

### 검색 품질에 영향을 미친 주요 요소

1. 이미지가 포함된 문서에서 주요 키워드와 가중치 추출



주요 키워드 : 커피  
가중치 : 90

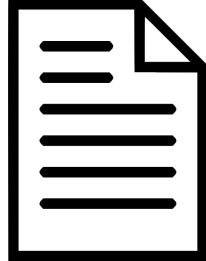

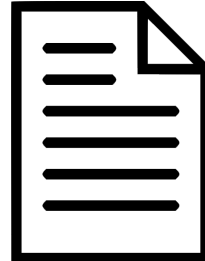







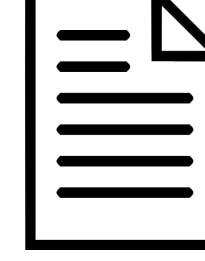


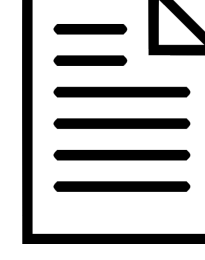
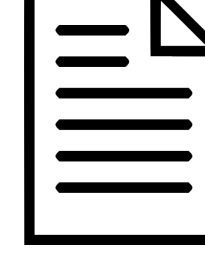
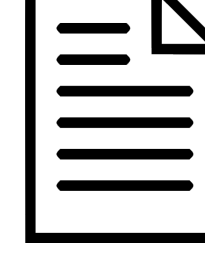
# 2.4 검색 랭킹

## 검색 품질에 영향을 미친 주요 요소

### 2. Word matrix

- 질의와 연관 검색어 간 Word matrix 생성
- 세로축은 질의 / 가로축은 연관 검색어
- 질의에 맞는 연관 검색어의 문서를 부스팅

질의

		연관 검색어			
		가수	노래	아이유	좋은날
질의	가수				
	노래				
	아이유				
	좋은날				

# 2.4 검색 랭킹

## 검색 품질에 영향을 미친 주요 요소

### 2. Word matrix

? [질의] 가수

연관도 높은 문서 Boosting



	가수	노래	아이유	좋은날
가수				
노래				
아이유				
좋은날				

# 2.4 검색 랭킹

## 검색 품질에 영향을 미친 주요 요소

### 2. Word matrix

? [질의] 노래

연관도 높은 문서 Boosting



	가수	노래	아이유	좋은날
가수				
노래				
아이유				
좋은날				



# 2.4 검색 랭킹

## 검색 품질에 영향을 미친 주요 요소

### 2. Word matrix

? [질의] 아이유

연관도 높은 문서 Boosting



	가수	노래	아이유	좋은날
가수				
노래				
아이유				
좋은날				

# 2.4 검색 랭킹

## 검색 품질에 영향을 미친 주요 요소

### 2. Word matrix

? [질의] 좋은날

연관도 높은  
문서 Boosting



	가수	노래	아이유	좋은날
가수				
노래				
아이유				
좋은날				

# 2.4 검색 랭킹

## 검색 품질에 영향을 미친 주요 요소

### 2. Word matrix

	가수	노래	아이유	좋은날
가수				
노래				
아이유				
좋은날				



# 2.4 검색 랭킹

## 검색 품질에 영향을 미친 주요 요소

### 2. Word matrix

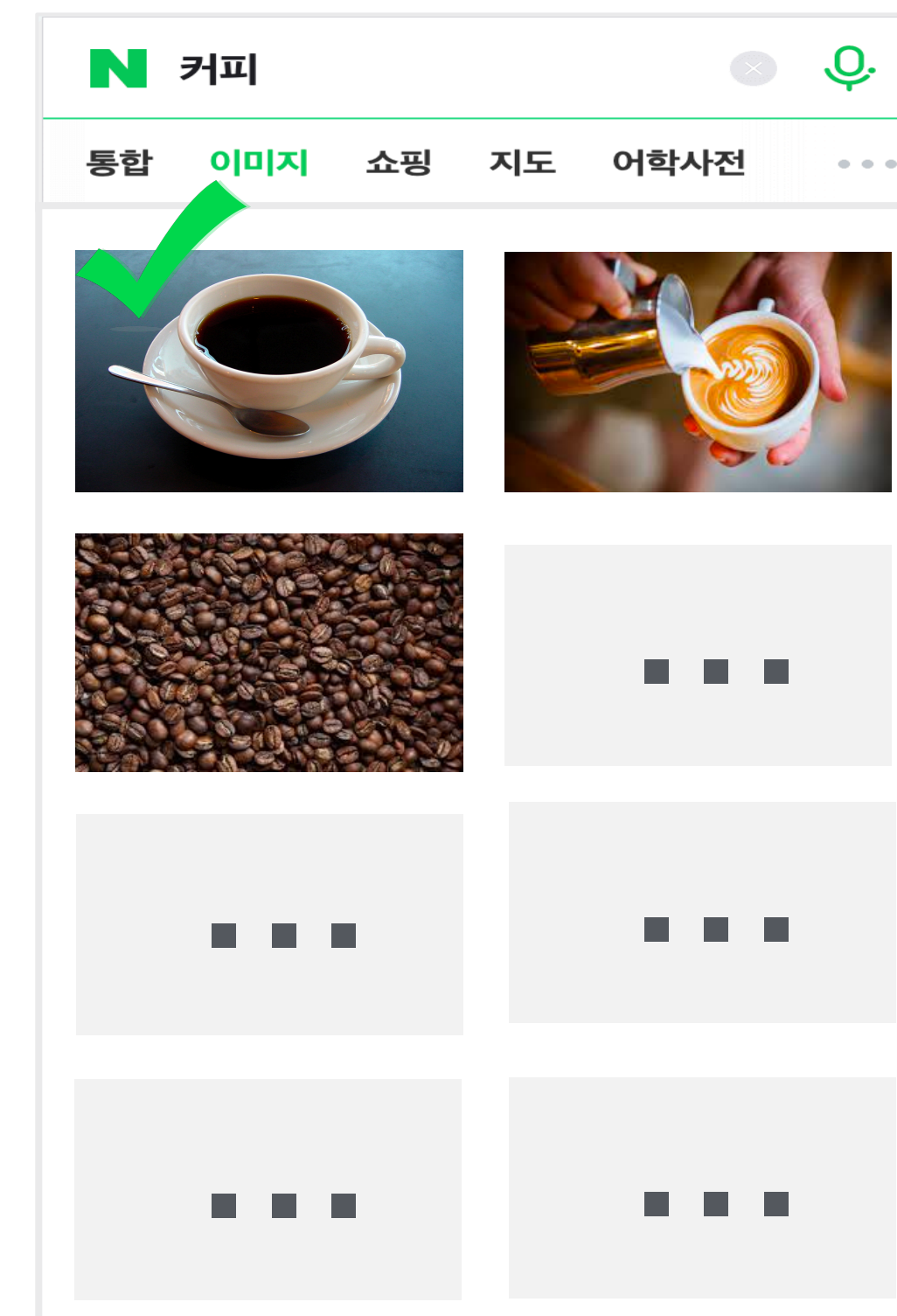
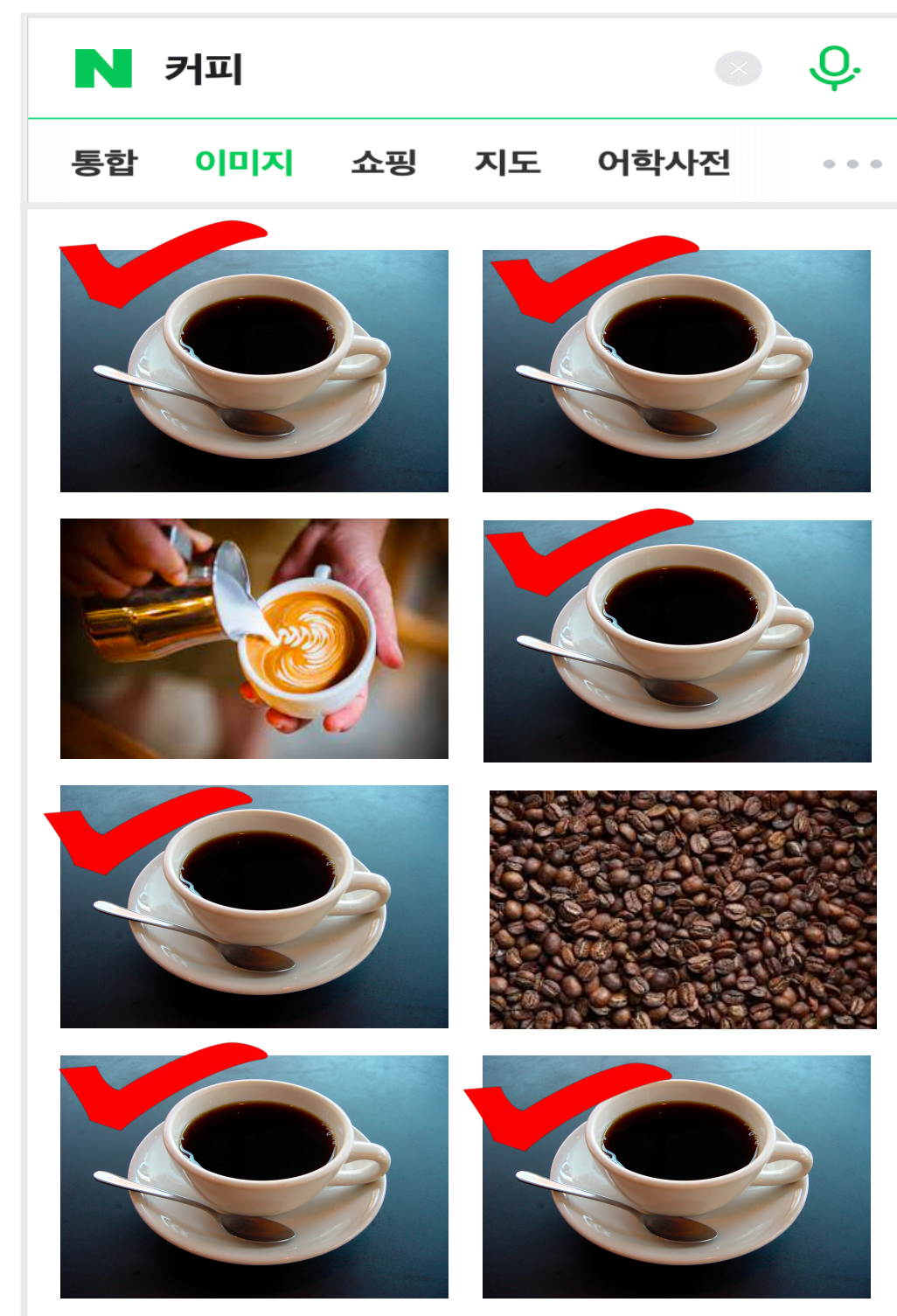
	가수	노래	아이유	좋은날
가수				
노래				
아이유	X			
좋은날		X		

대칭 구조 X

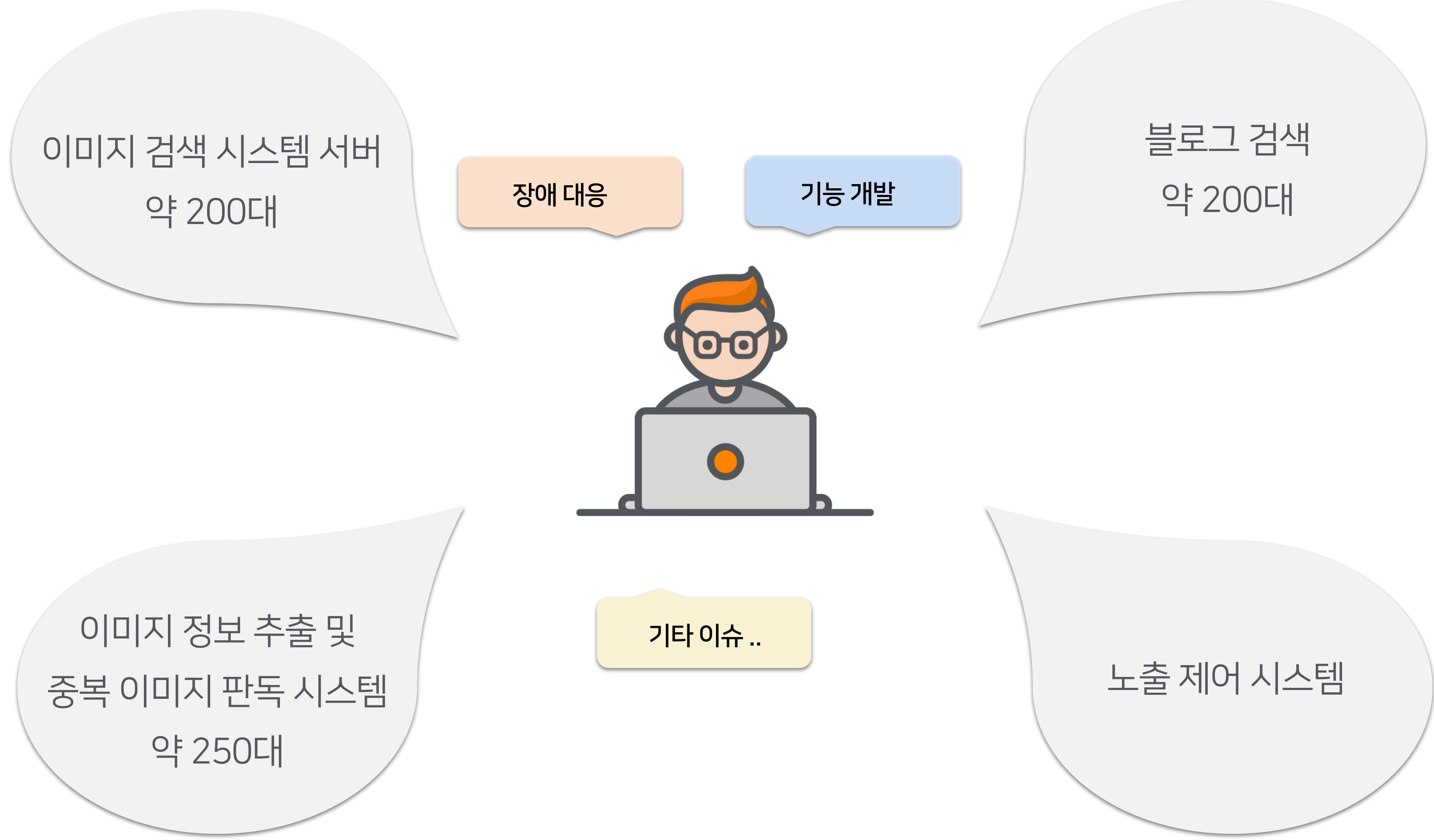
# 2.4 검색 랭킹

## 이미지 중복 추출

- 이미지정보 추출 후 중복 이미지 노출 방지
- 중복 이미지 중 원본 이미지가 우선적으로 노출



# 2.5 2021년 돌아본 그때의 DevOps



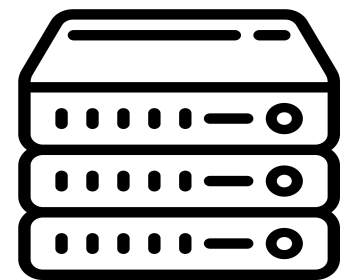
# 2.5 2021년 돌아본 그때의 DevOps

## 검색 시스템 서버 담당자의 장애대응



### 24시간 무중단 서비스를 운영하는 것

- 장비 장애 시 즉시 대응 필요
- 업무 시간 외에도 항시 대기
- 외부 활동 시 장애 발생을 대비해서 노트북 지참 / PC방 활용



### 운영 서버의 성능

- RAID 0 (stripe), SATA / HDD로 구성
- 디스크 장애에 취약
- 시스템 가용량이 넉넉하지 않아 규모가 큰 시스템들은 빠른 장비 교체 필요

# 2.5 2021년 돌아본 그때의 DevOps

## 검색 시스템 서버 담당자의 장애대응 사례



사례 1



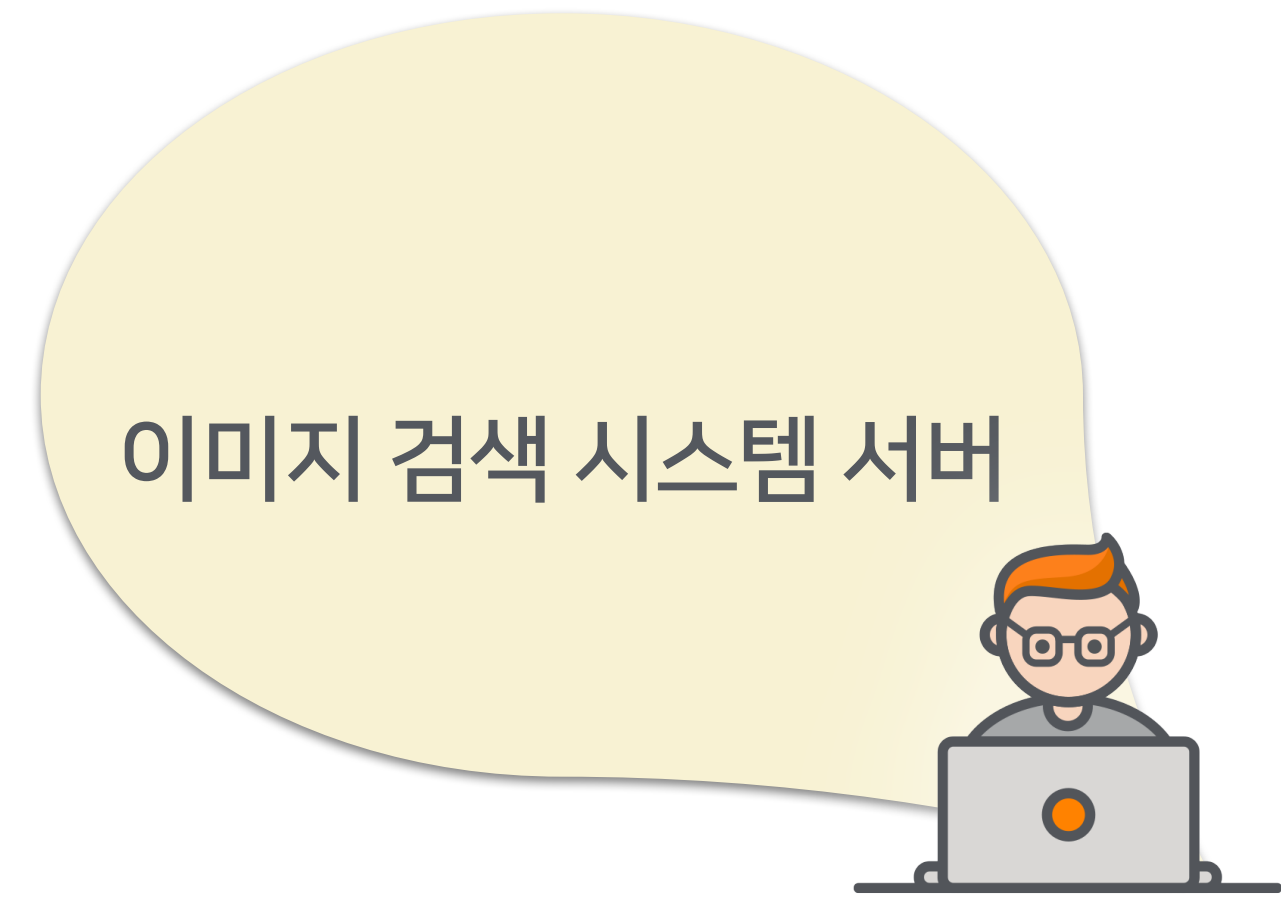
사례 2



사례 3

# 3. 변화하는 이미지 검색 DevOps (2015년, 그때 그랬지)

# 3.1 변화하는 업무구조



담당업무 전문화  
(검색, 정제, 품질 등)





## 3.2 변화하는 검색시스템

### 검색 관련 플랫폼 활용 시작

Step 1. 각 플랫폼 업무 전문화

Step 2. 개별 클러스터 시스템에서 플랫폼을 활용하는 형태로 변화 시작

Step 3. 운영 부담 감소 (관련 시스템 관리 / 장애 대응 시간 감소)



## 3.2 변화하는 검색시스템

### 검색 관련 플랫폼 활용 시작

Step 1. 각 플랫폼 업무 전문화

Step 2. 개별 클러스터 시스템에서 플랫폼을 활용하는 형태로 변화 시작

Step 3. 운영 부담 감소 (관련 시스템 관리 / 장애 대응 시간 감소)

## 3.2 변화하는 검색시스템

### 검색 관련 플랫폼 활용 시작

Step 1. 각 플랫폼 업무 전문화

Step 2. 개별 클러스터 시스템에서 플랫폼을 활용하는 형태로 변화 시작

Step 3. 운영 부담 감소 (관련 시스템 관리 / 장애 대응 시간 감소)

## 3.2 변화하는 검색시스템

### 검색 관련 플랫폼 활용 시작

Step 1. 각 플랫폼 업무 전문화

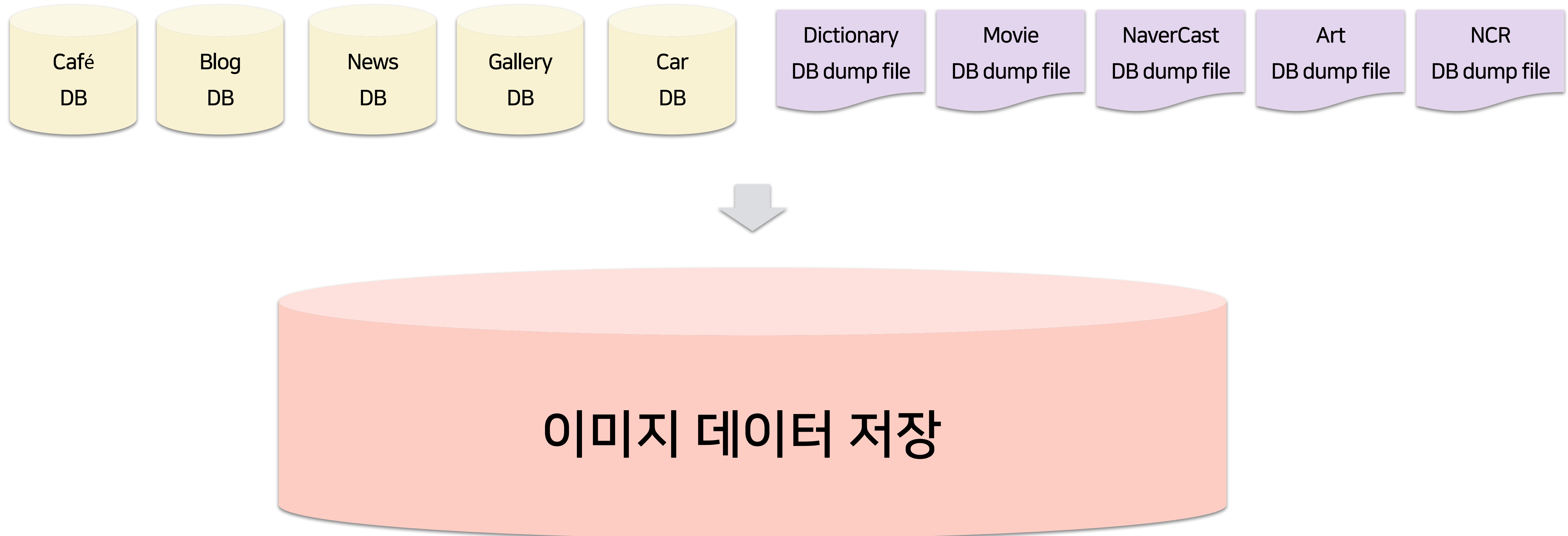
Step 2. 개별 클러스터 시스템에서 플랫폼을 활용하는 형태로 변화 시작

Step 3. 운영 부담 감소 (관련 시스템 관리 / 장애 대응 시간 감소)

**=> 개발 및 안정화 3년 정도 소요**

# 3.3 검색시스템 지원 플랫폼

## CUVE (데이터저장소)



## 3.3 검색시스템 지원 플랫폼

### CUVE (데이터저장소)

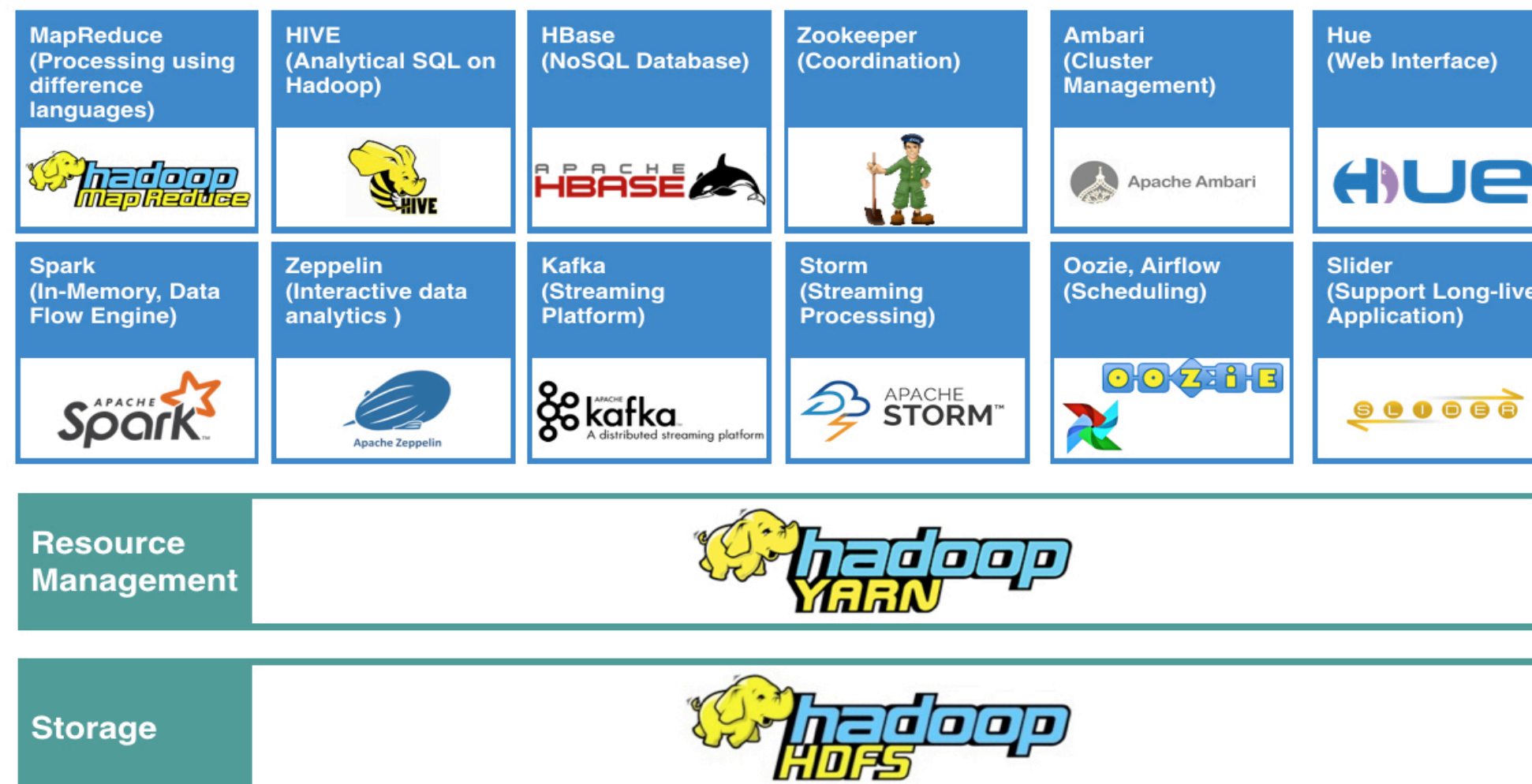
#### 이미지 데이터 저장

- 이미지 검색을 위한 모든 이미지 및 문서 정보가 저장
- 여러 Source 를 통해 얻은 데이터를 일괄적 관리
- 검색을 위한 ETL (Extract, Transform, Load) 처리가 쉽고 데이터 안정성이 높아짐

# 3.3 검색시스템 지원 플랫폼

## C3 (분산 클러스터 프레임워크)

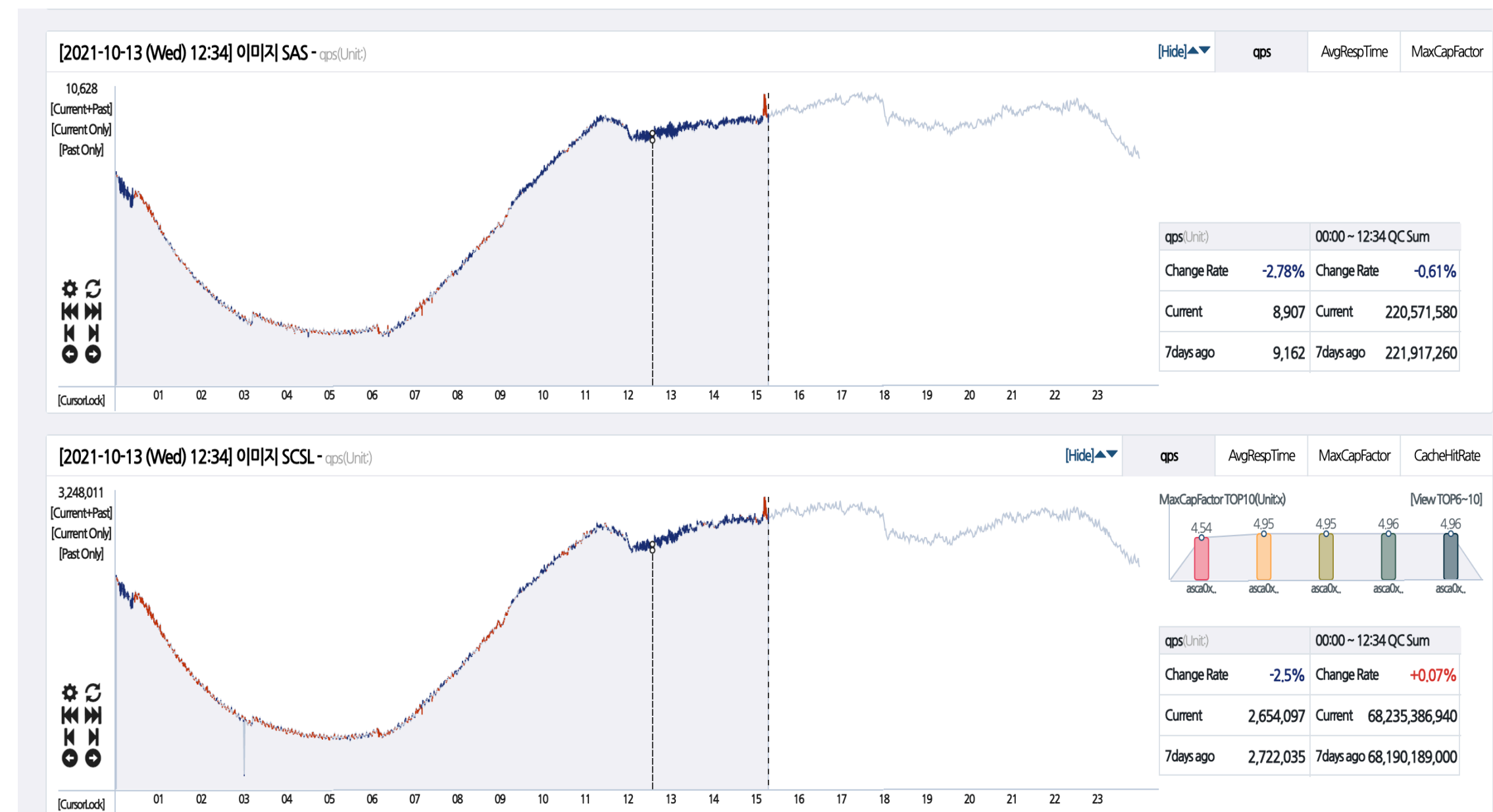
- CUVE에 저장된 데이터를 이용
- 이미지 정보 추출 후 CUVE에 저장
- 검색서비스를 위한 문서 생성 및 HDFS 저장 또는 필요 시스템에 전송



# 3.3 검색시스템 지원 플랫폼

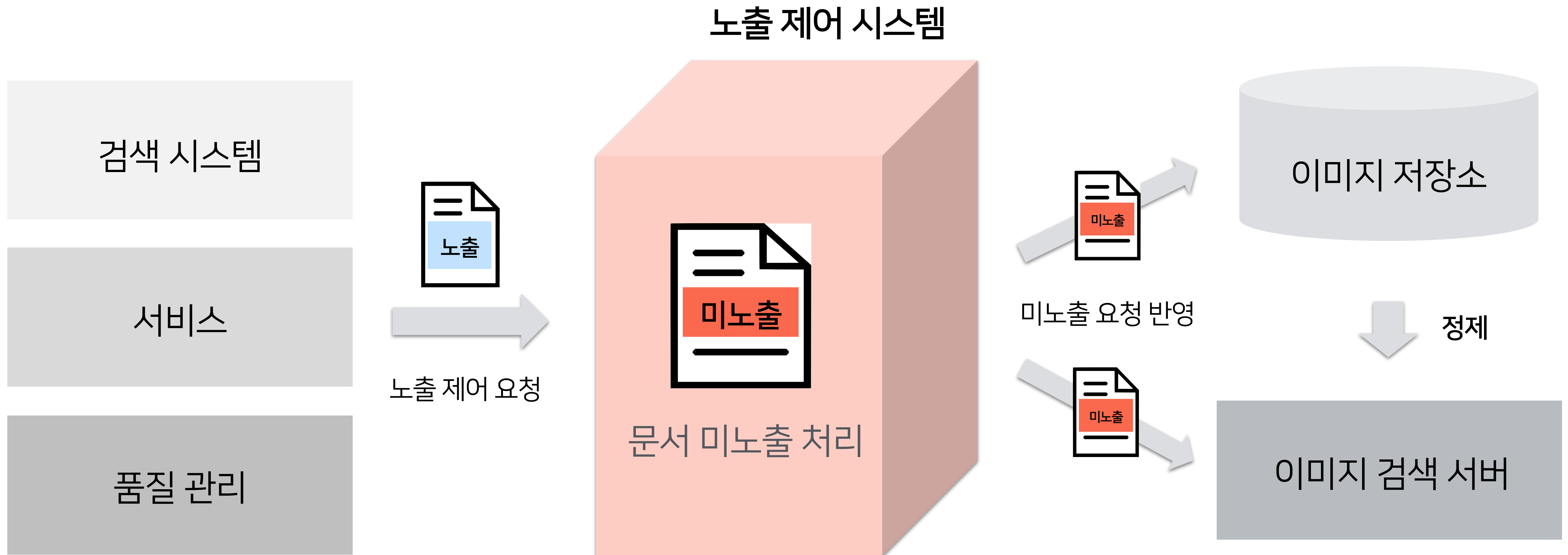
## 모니터링

- 검색 장비 지표 수집 도구
- 검색 트래픽 이상 탐지 도구



# 3.3 검색시스템 지원 플랫폼

## 노출제어





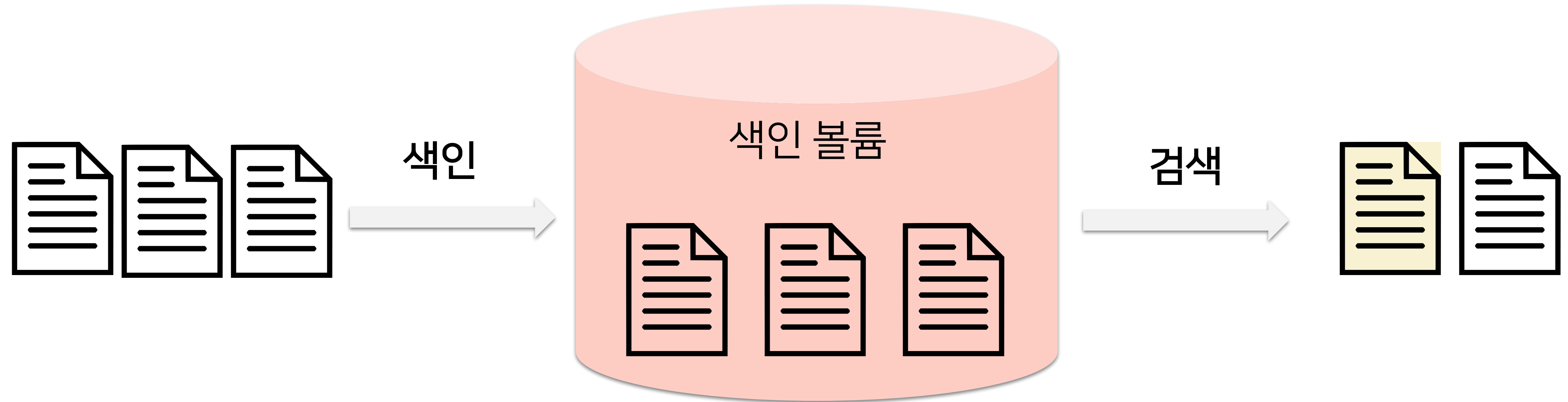
# 3.4 변화하는 검색시스템

## 검색 엔진의 변화 - 배치



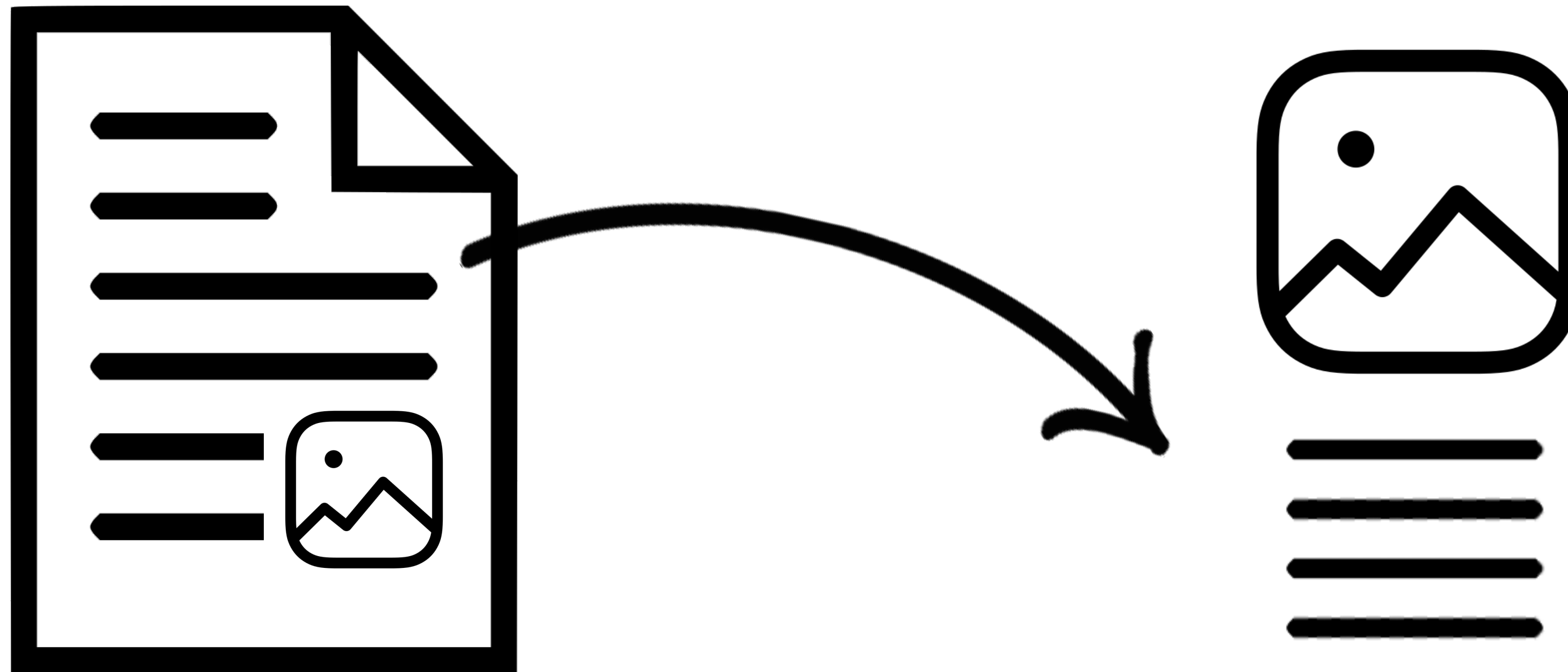
# 3.4 변화하는 검색시스템

## 검색 엔진의 변화 - 증분



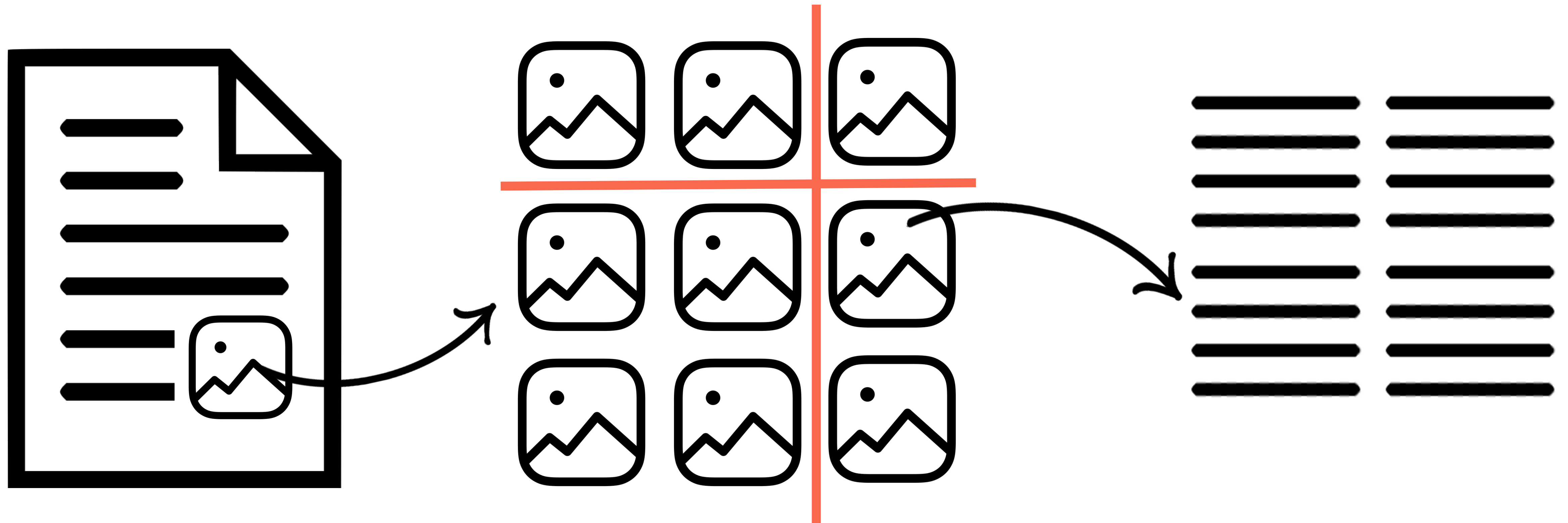
## 3.5 변화하는 검색랭킹

텍스트 중심에서 이미지정보 중심으로



# 3.5 변화하는 검색랭킹

텍스트 중심에서 이미지정보 중심으로



## 3.6 이미지 검색의 과도기



각각 검색 관련 플랫폼들의 독립으로 인한 분업 및 전문화



새로운 검색 관련 플랫폼들을 잘 이해하고 사용할 수 있어야 함



검색 관련 플랫폼 지원으로 인한 사용 가능 리소스 증가 및 이미지 검색 품질 증가

# 4. 현재의 이미지검색 DevOps

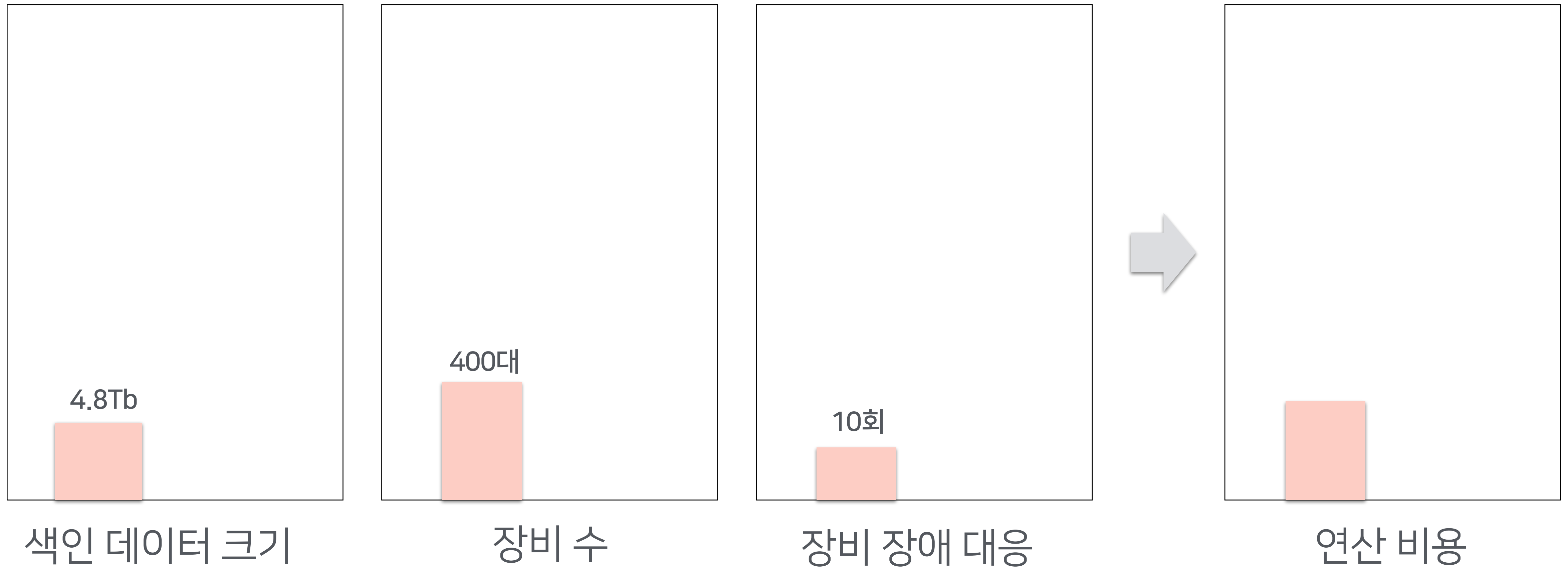
# 4.1 계속 변화하는 검색랭킹

기존  
이미지  
검색

새로운 이미지 검색  
색인 수 증가  
이미지 정보 수 증가  
리소스 증가  
신규 데이터 처리량 증가  
(텍스트 + 이미지)

# 4.1 계속 변화하는 검색랭킹

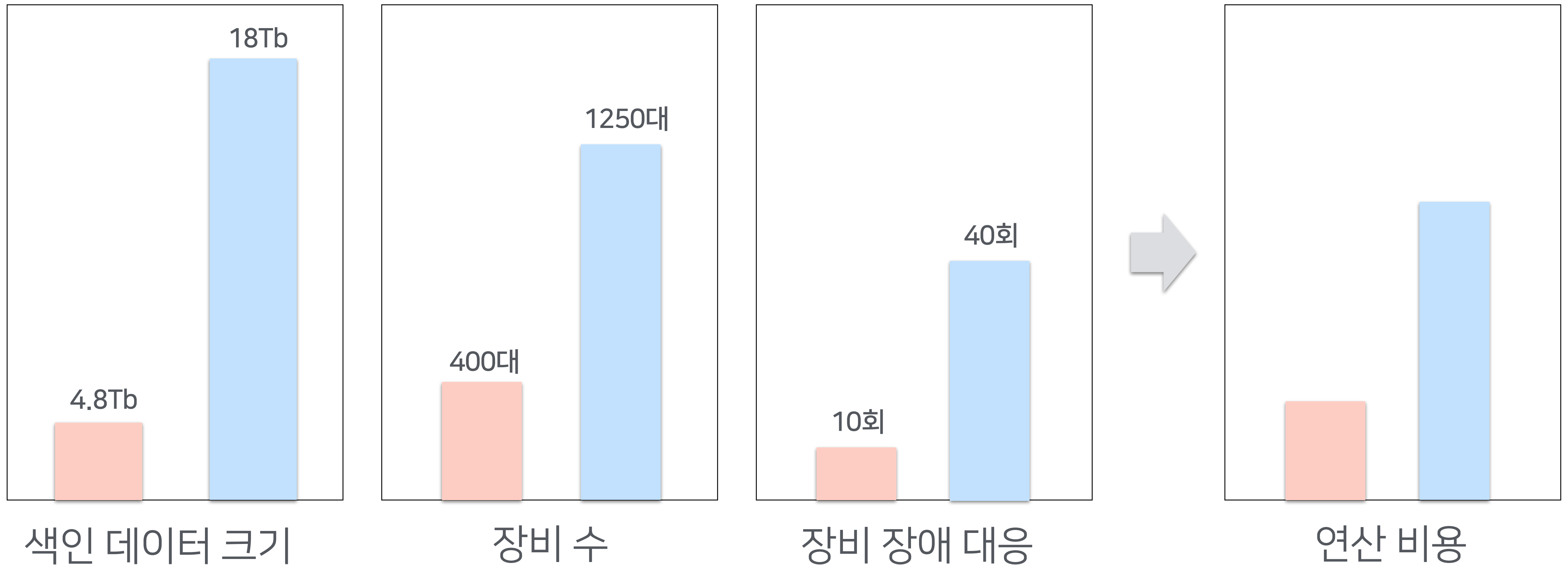
데이터 증가와 랭킹 변화에 따른 검색 시스템의 영향





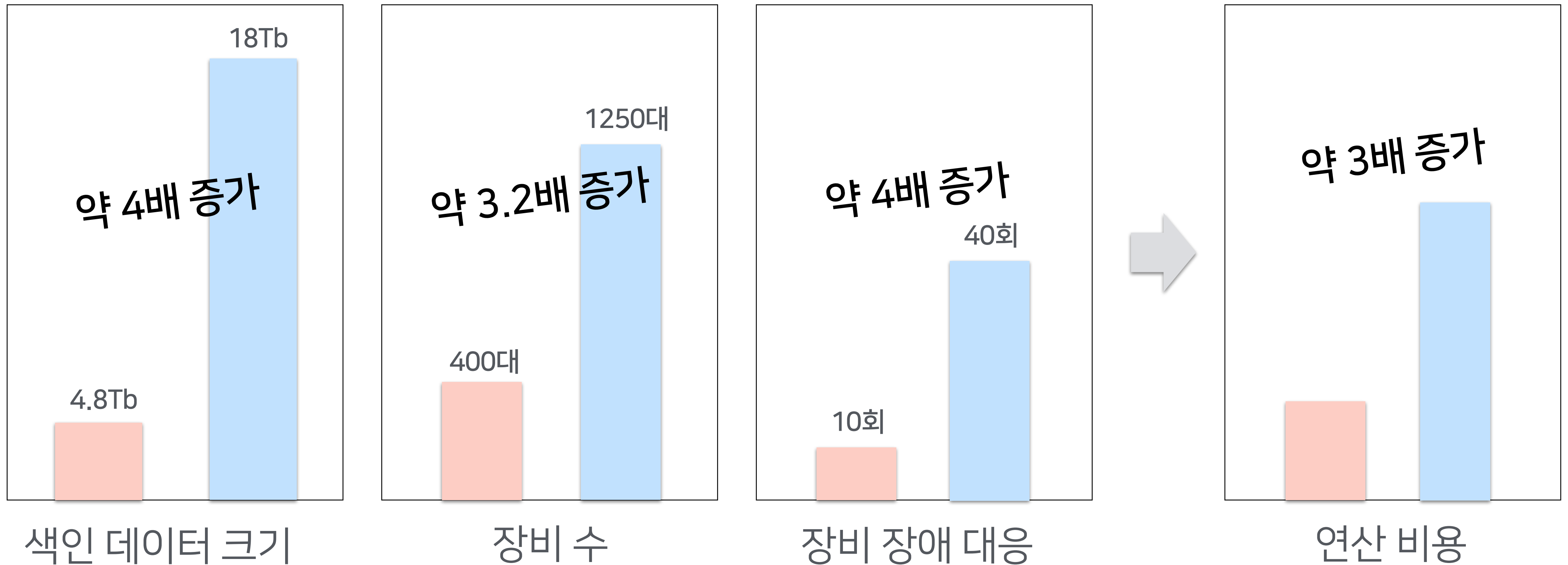
# 4.1 계속 변화하는 검색랭킹

데이터 증가와 랭킹 변화에 따른 검색 시스템의 영향

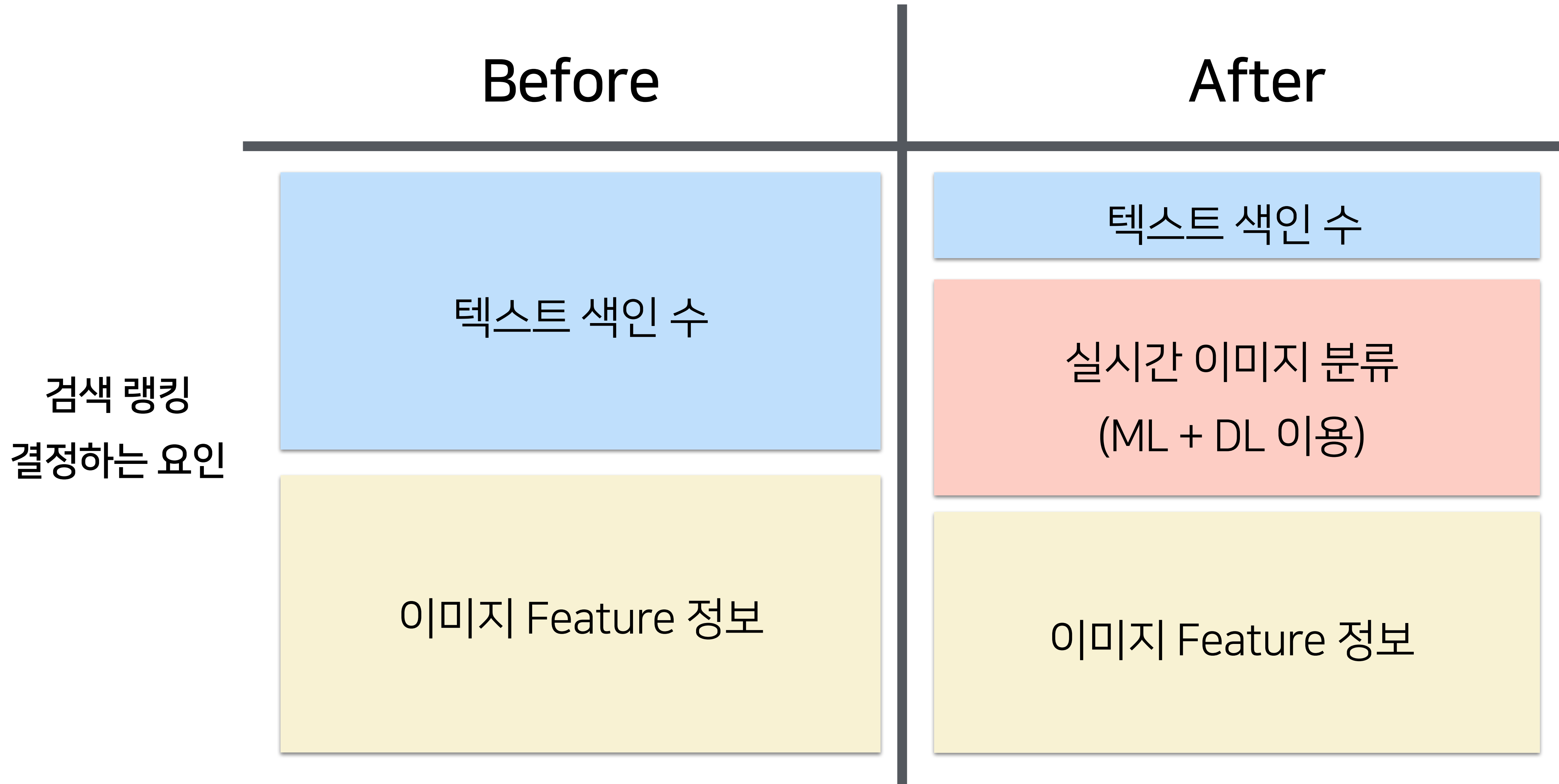


# 4.1 계속 변화하는 검색랭킹

데이터 증가와 랭킹 변화에 따른 검색 시스템의 영향



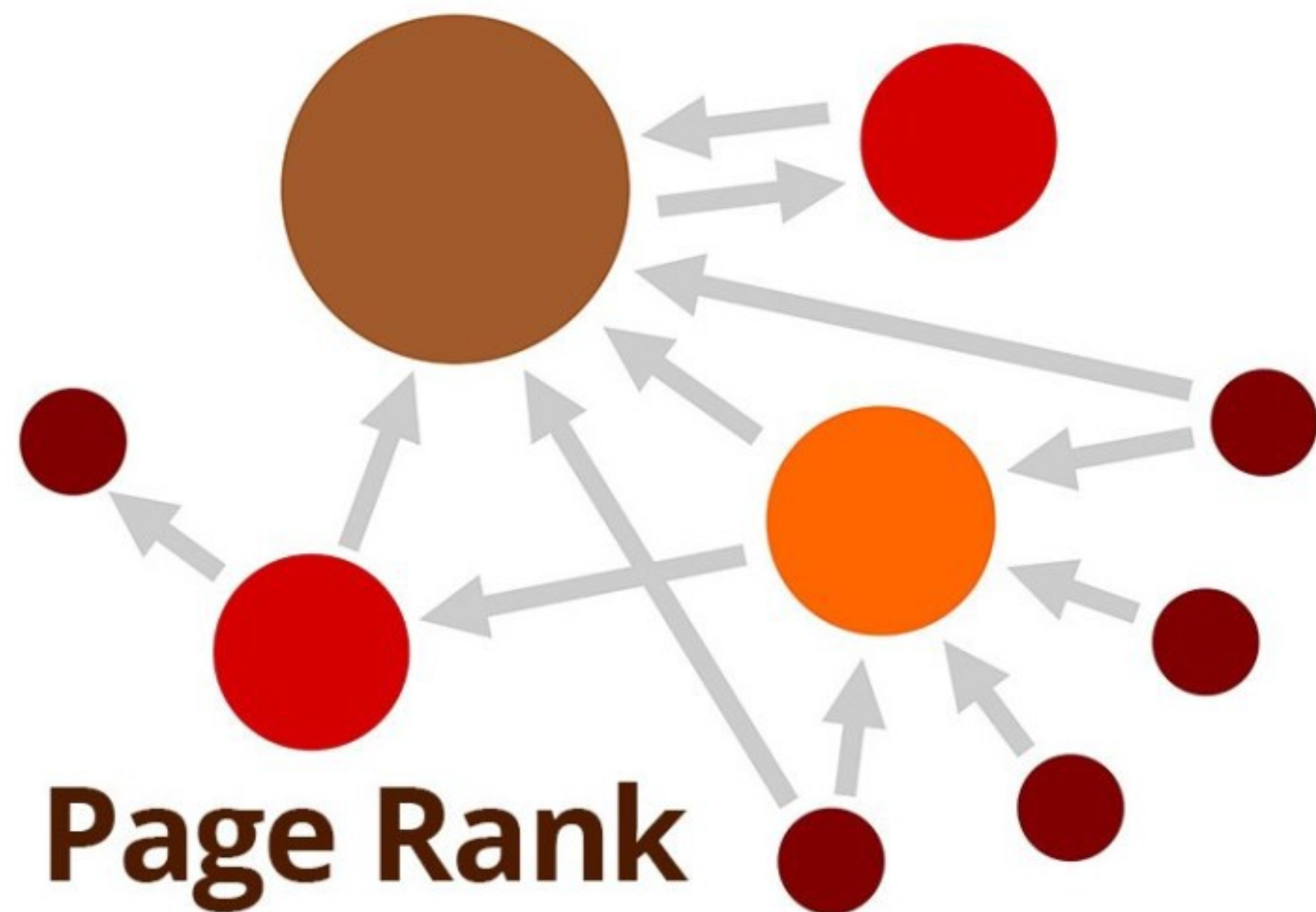
# 4.1 계속 변화하는 검색랭킹



# 4.1 계속 변화하는 검색랭킹

## 실시간 이미지 랭킹 로직

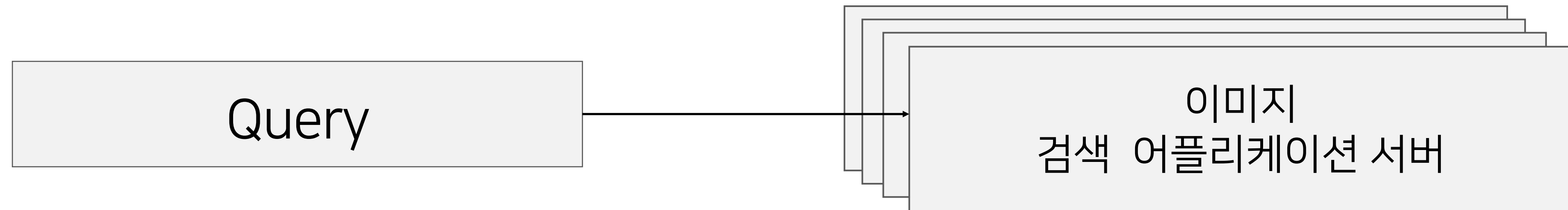
- 구글의 PageRank 개념을 이미지 랭킹 로직에 도입



1. 웹 페이지 간의 관계를 통해 페이지의 중요도를 계산
2. Top-n 이미지들 간의 이미지 유사도를 계산하고 중요도가 높은 이미지에 높은 점수를 줌
3. Top-n에 랭킹되는 이미지들을 상위 결과에 노출

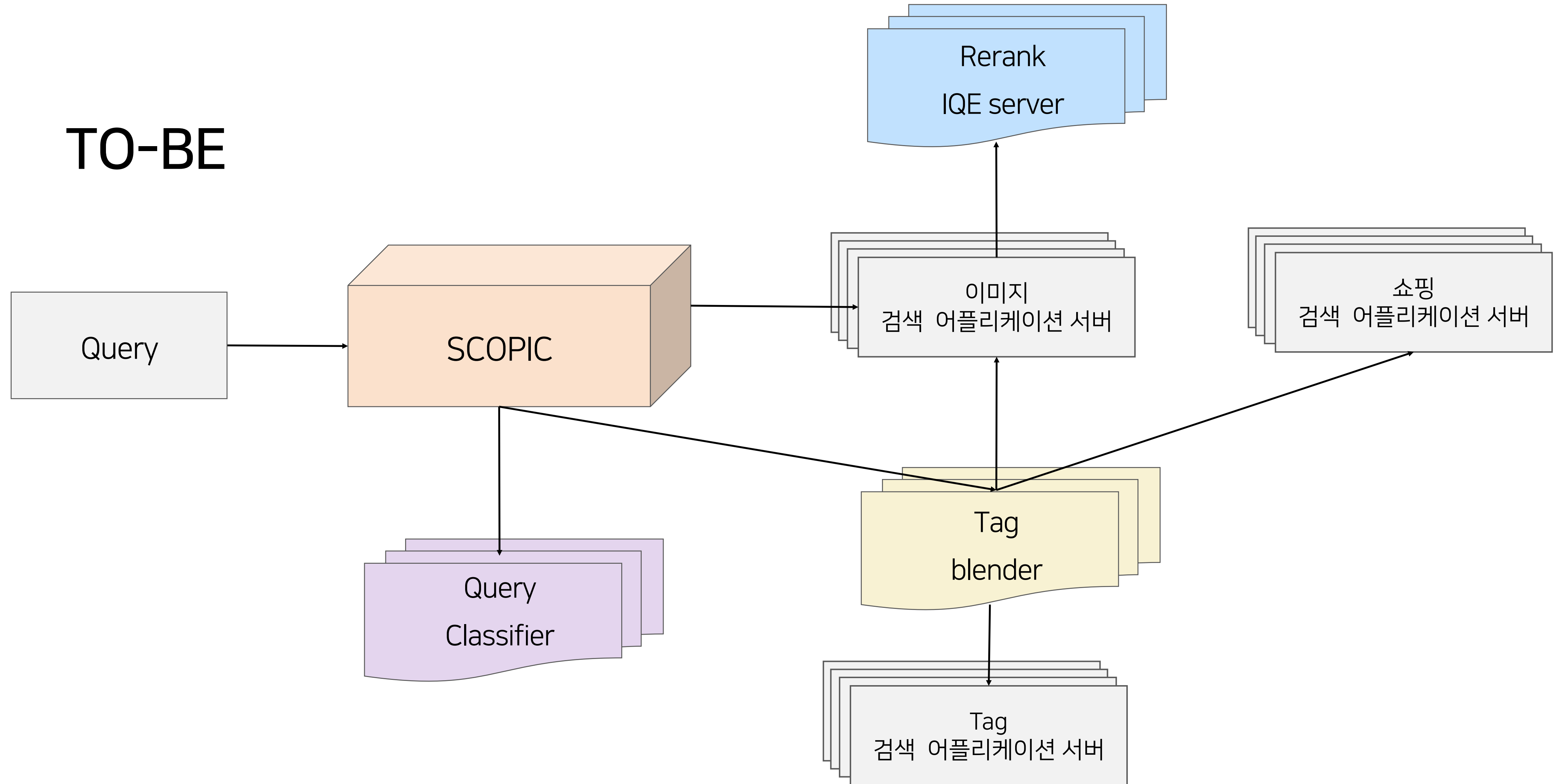
## 4.2 이미지 검색시스템 구조 변화

AS-IS



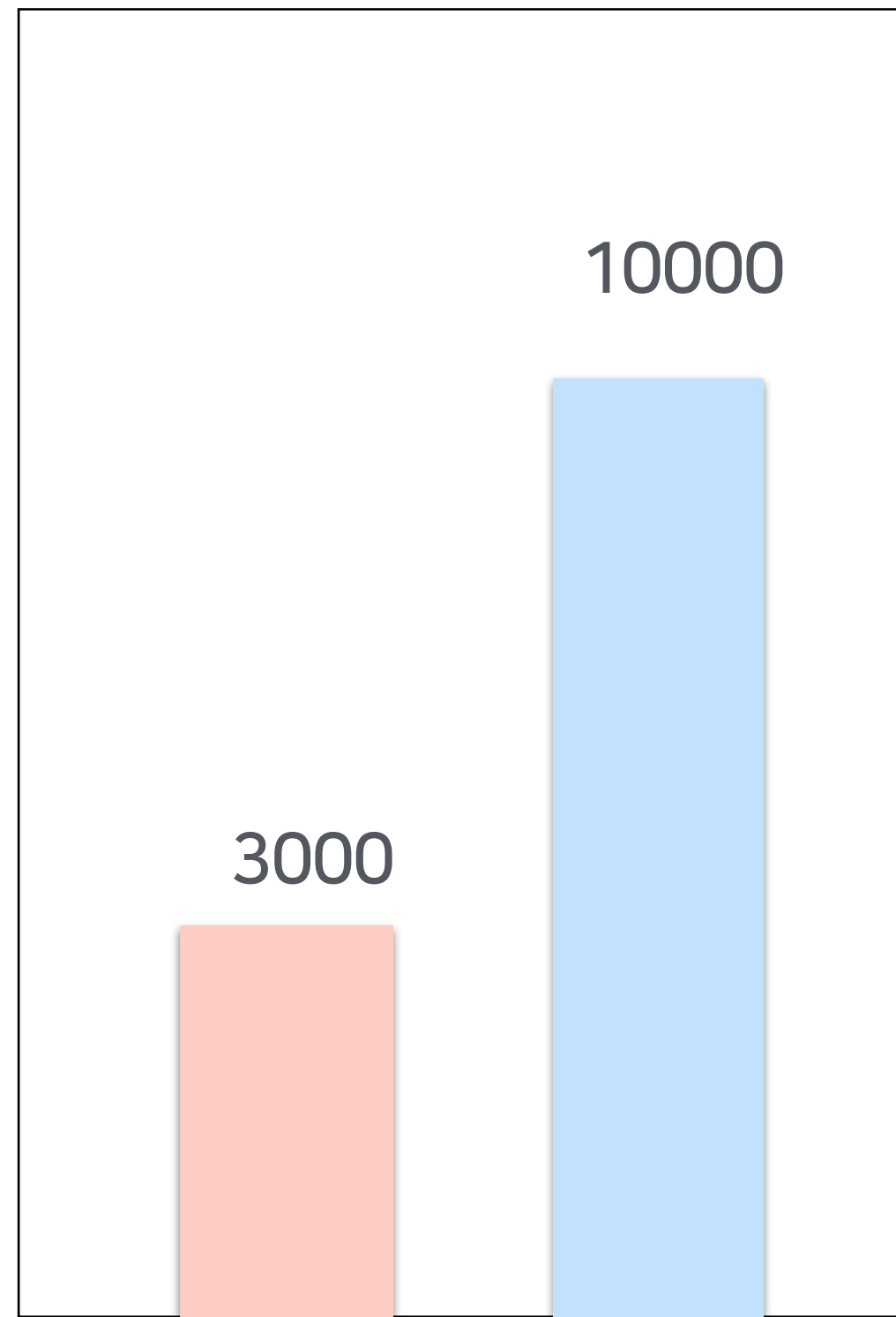
# 4.2 이미지 검색시스템 구조 변화

TO-BE

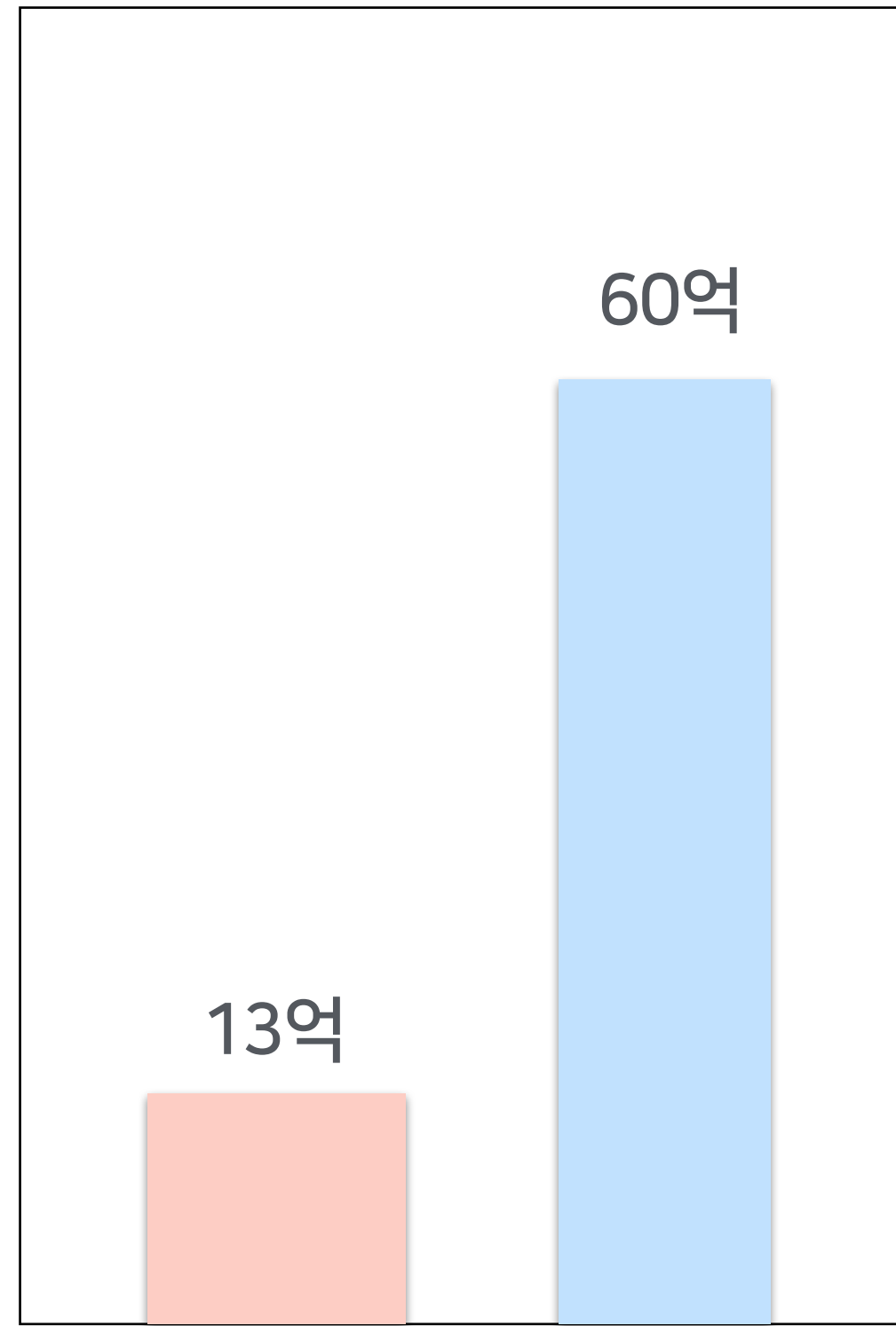


# 4.2 이미지 검색시스템 구조 변화

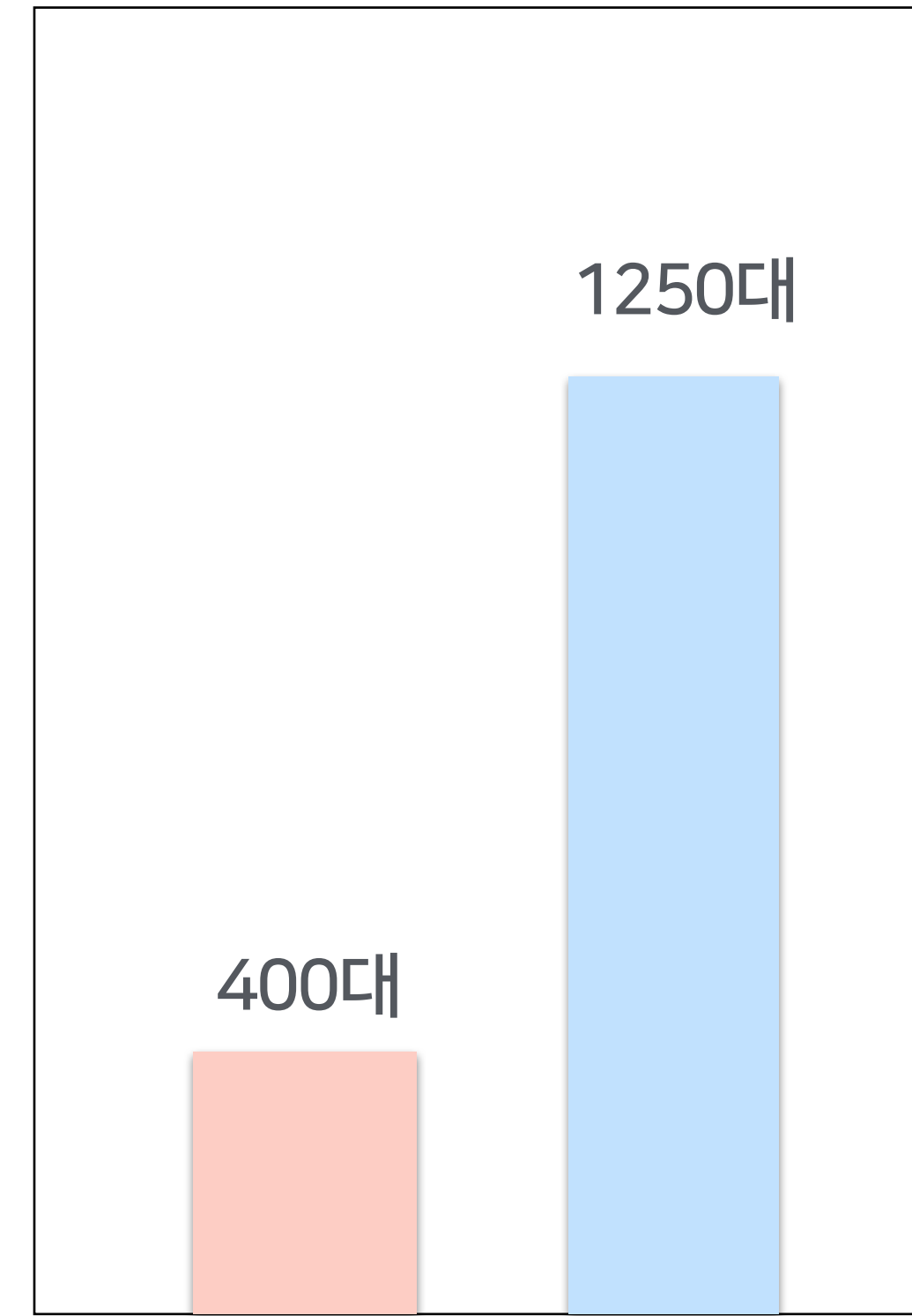
## 이미지 검색 시스템 주요 지표



초당 요청량



이미지 수



서버 수

## 4.2 이미지 검색시스템 구조 변화

### 이미지 검색 변화에 따른 변화

#### - 좋은 점

- 검색 품질 향상

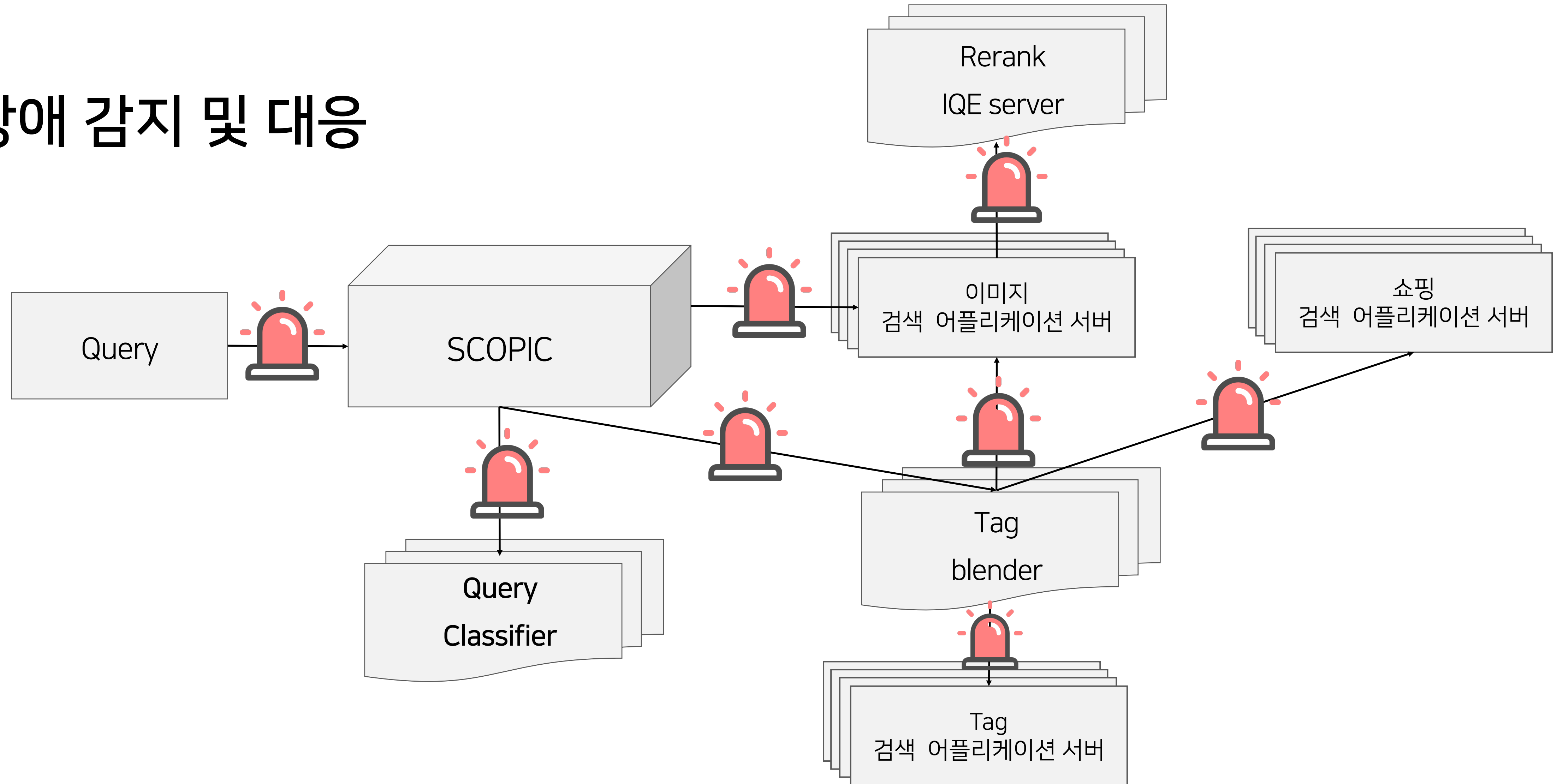
#### - 어려운 점

- 검색엔진에서의 응답시간 외 추가적으로 발생하는 응답시간 고려필요
- 시스템장애 포인트 증가 및 해당 시스템 오류 시 검색결과 이상 발생



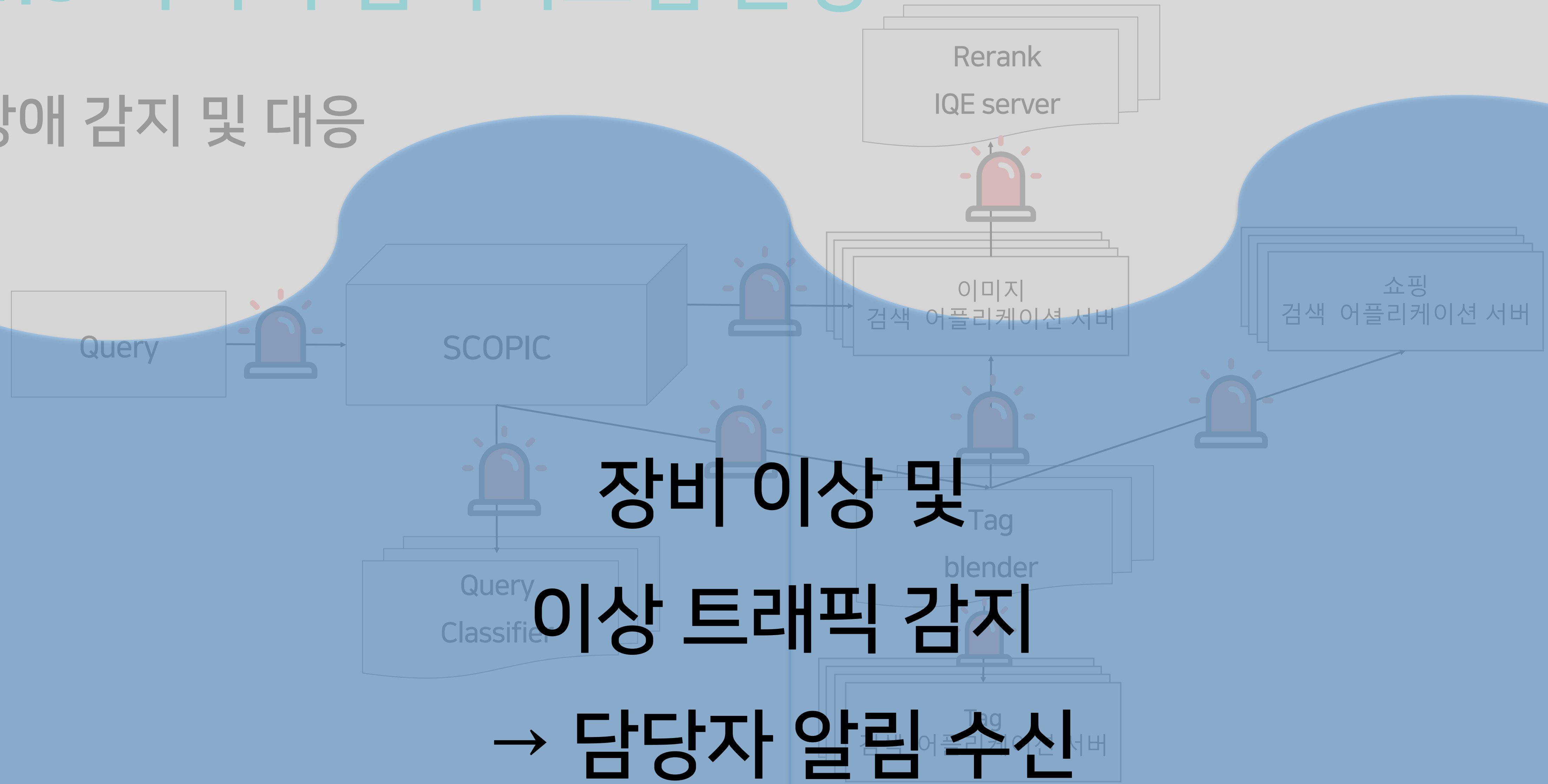
# 4.3 이미지 검색시스템 운영

## 장애 감지 및 대응



# 4.3 이미지 검색시스템 운영

장애 감지 및 대응



**장비 이상 및  
이상 트래픽 감지**

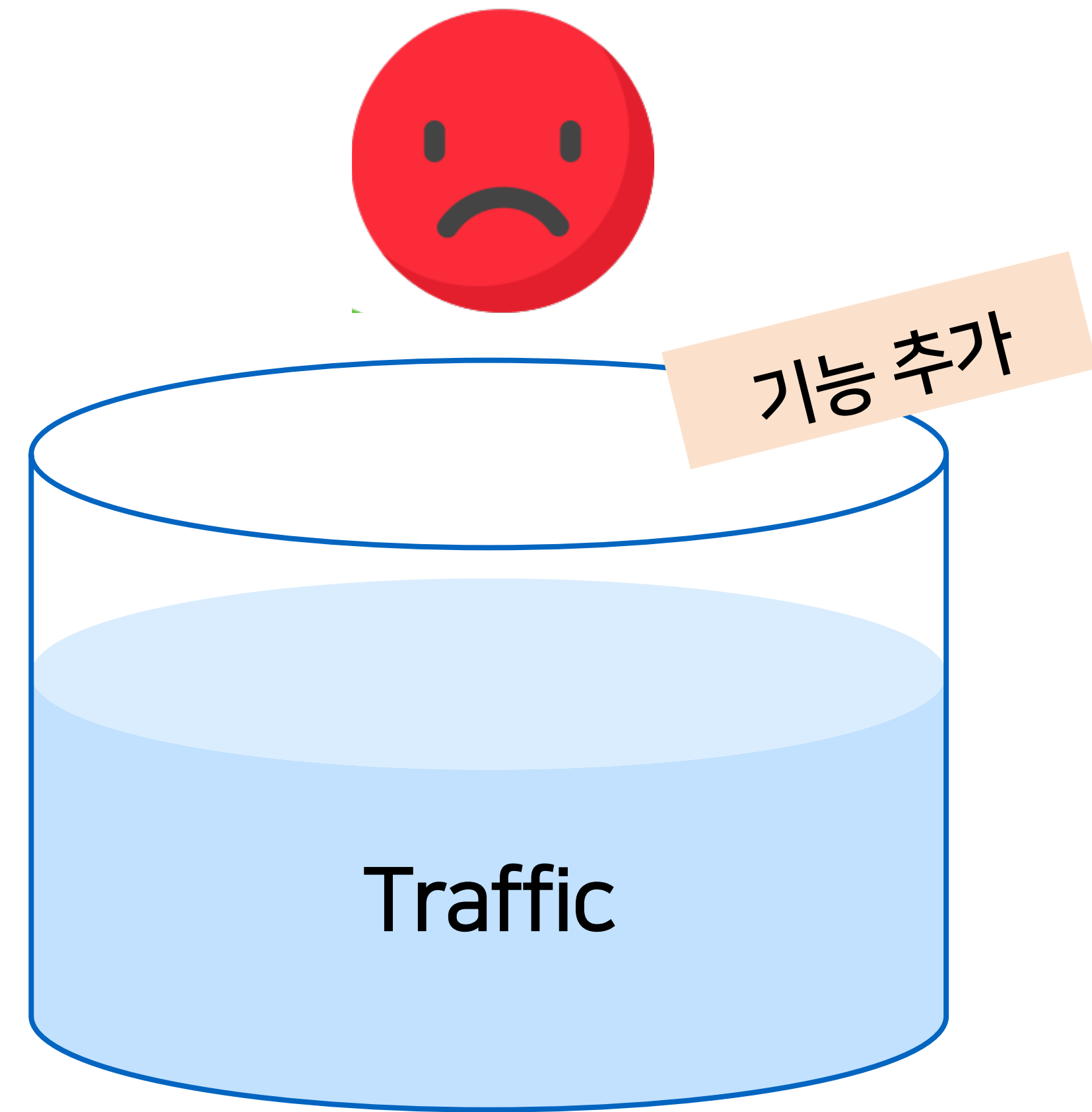
**→ 담당자 알림 수신**

# 4.3 이미지 검색시스템 운영

## 클러스터 관리



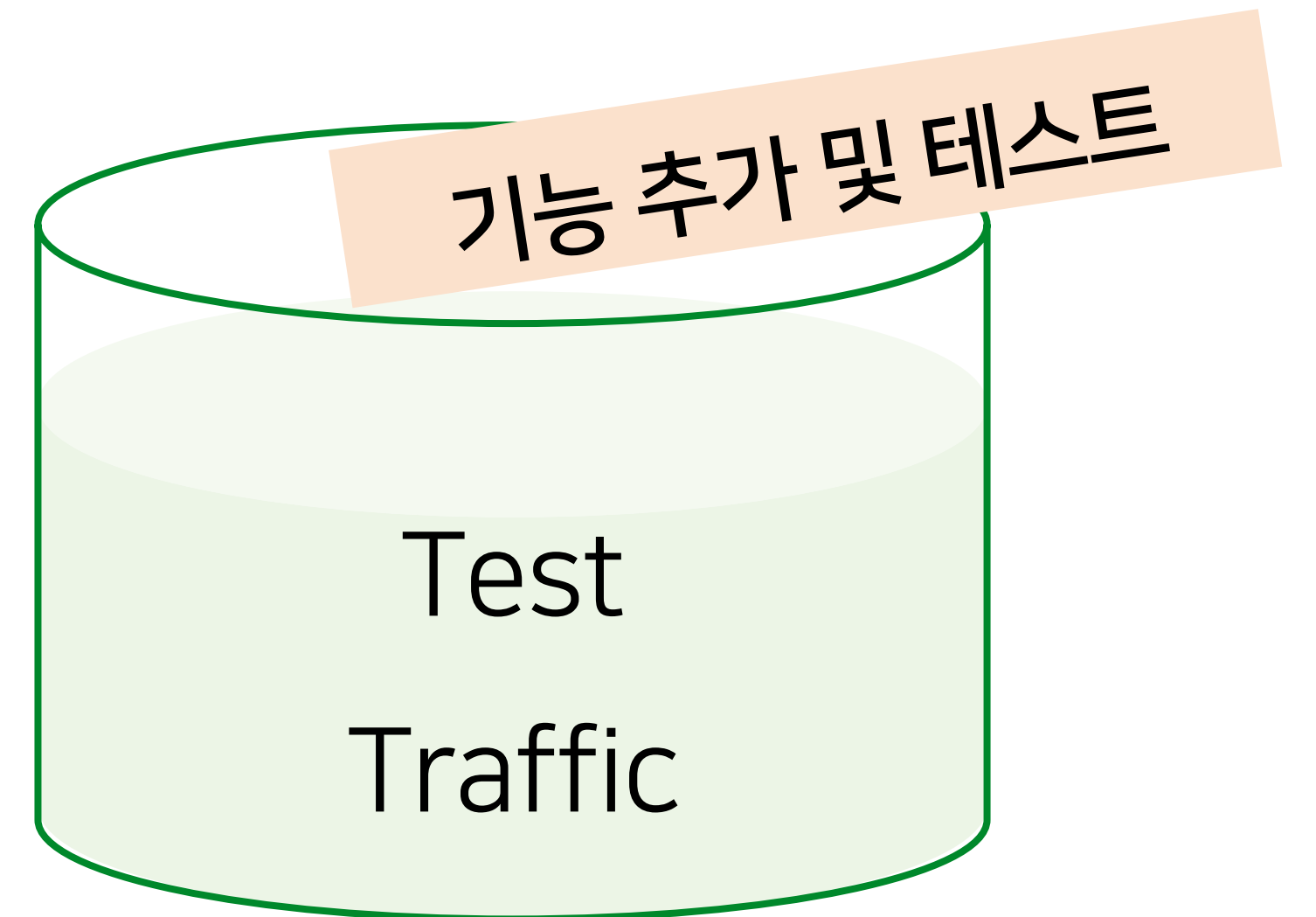
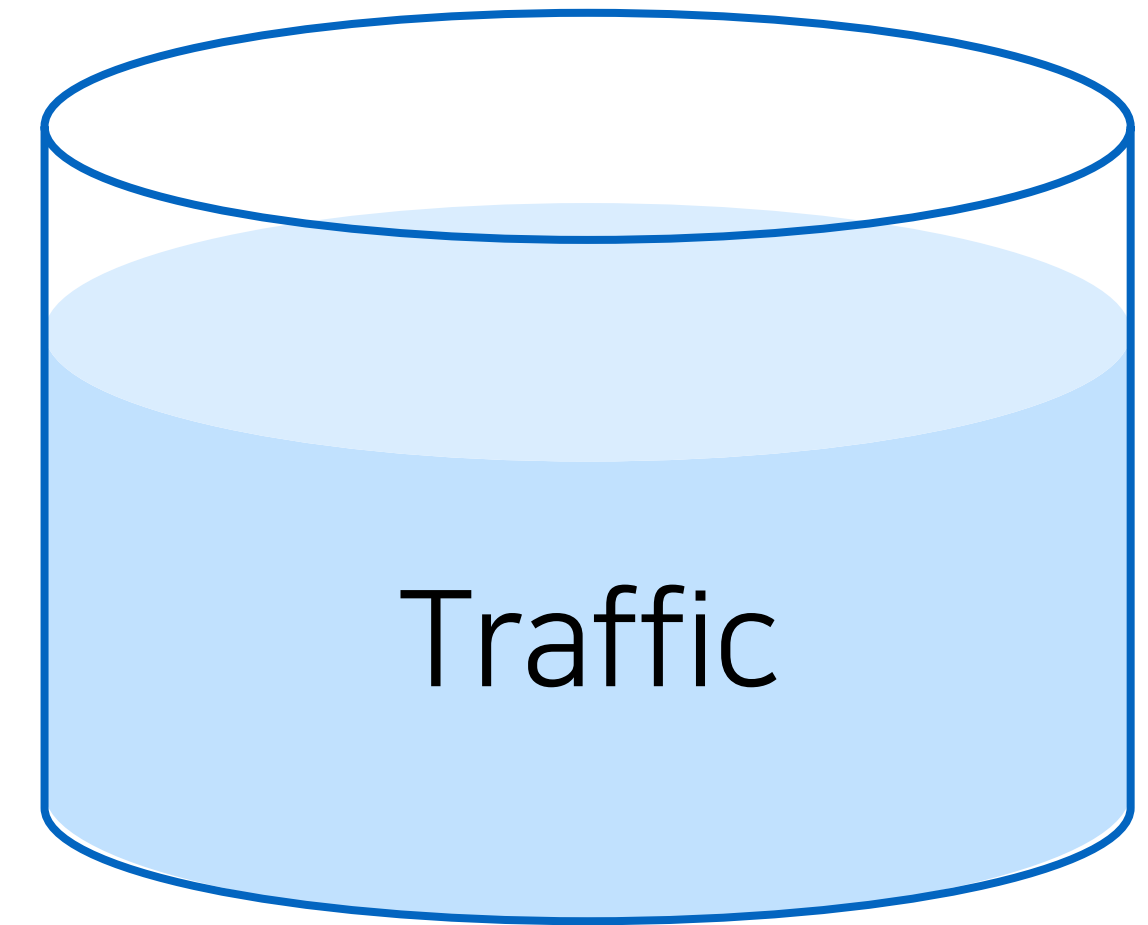
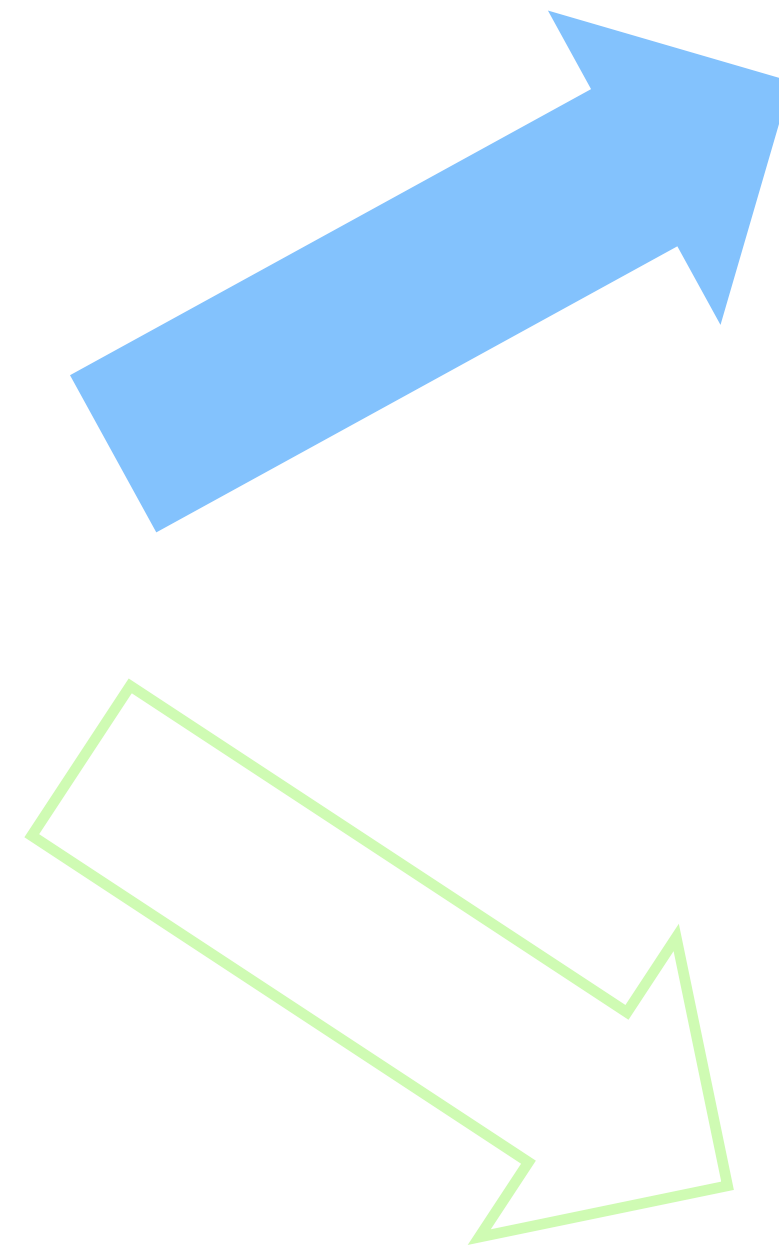
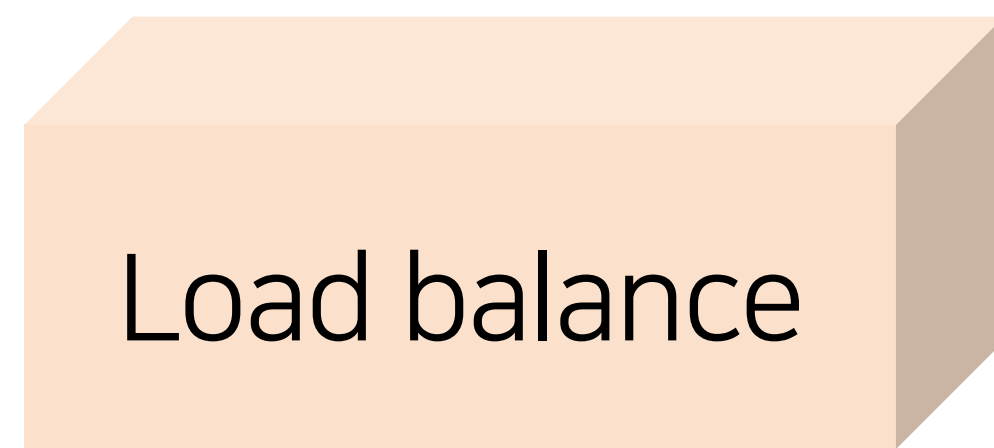
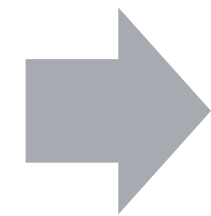
Test



Real

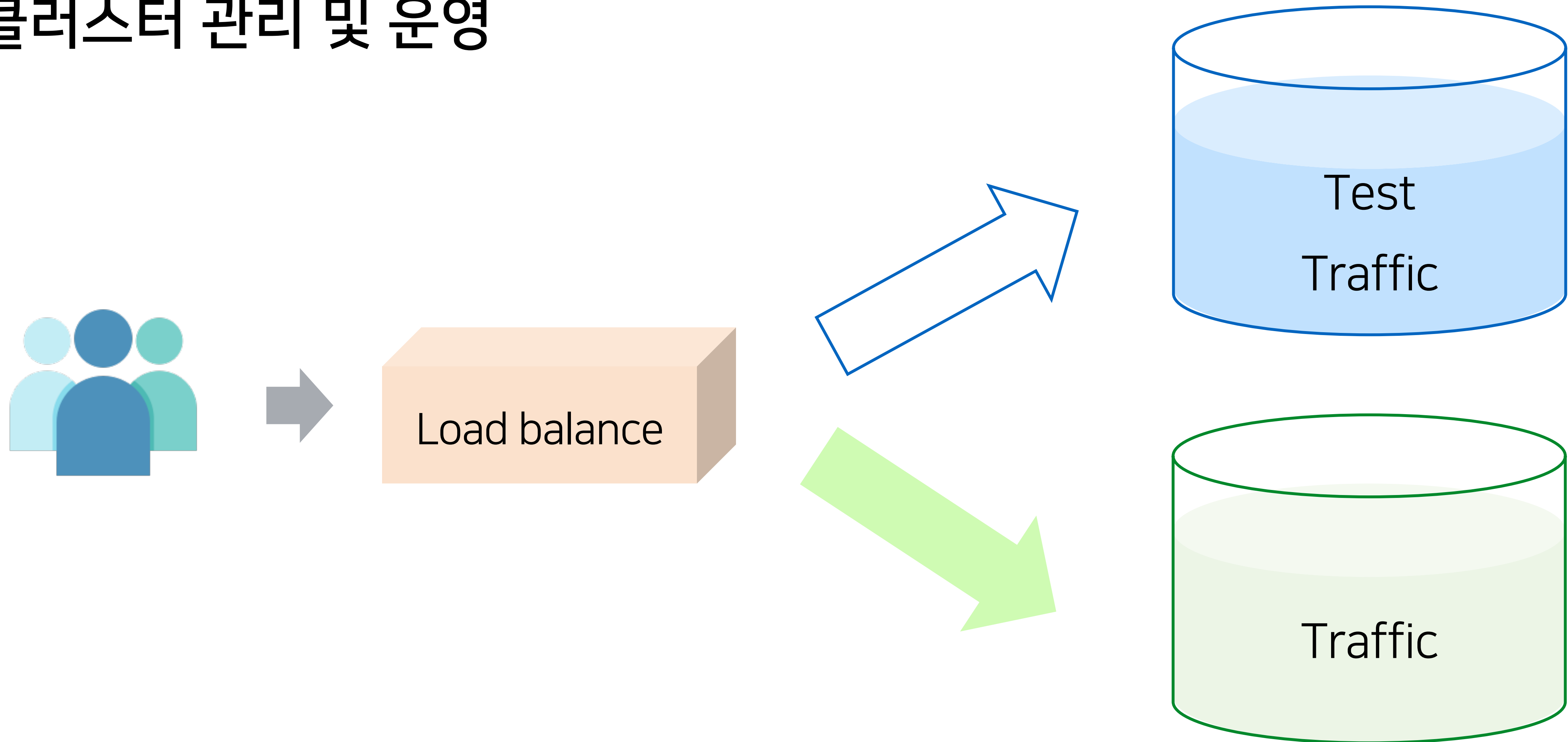
# 4.3 이미지 검색시스템 운영

## 클러스터 관리 및 운영

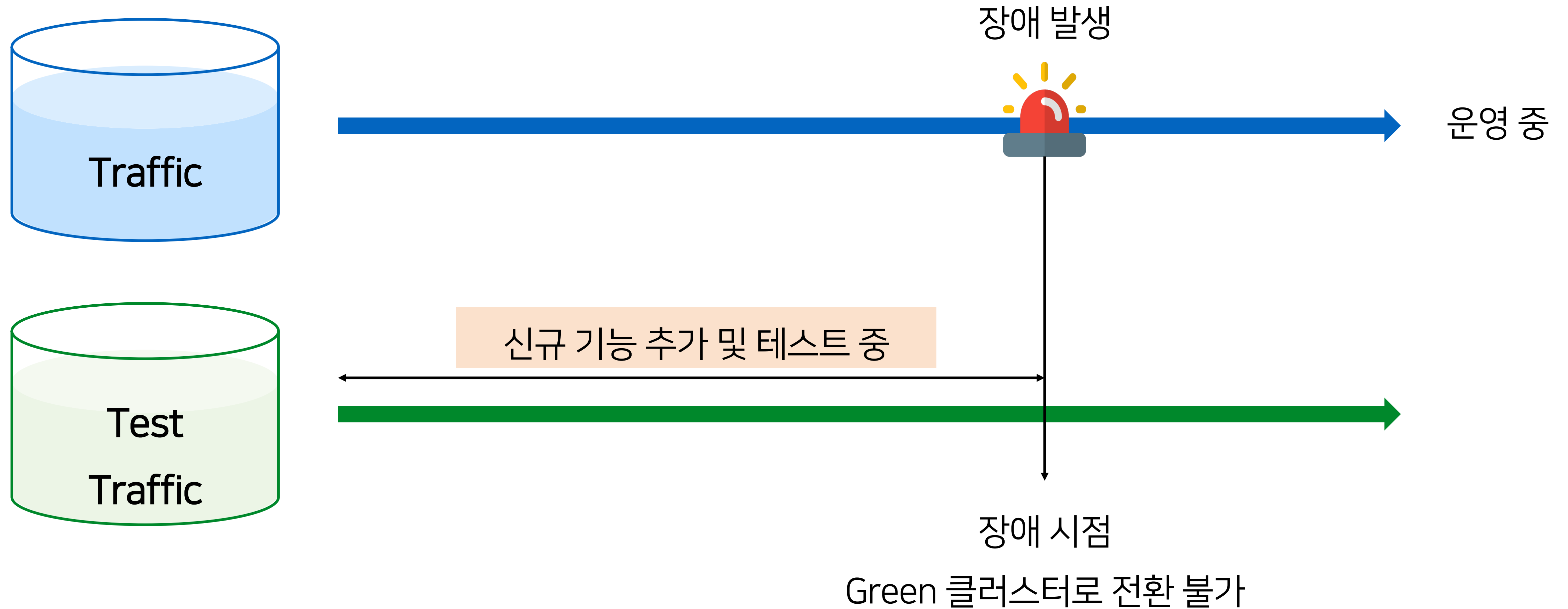


# 4.3 이미지 검색시스템 운영

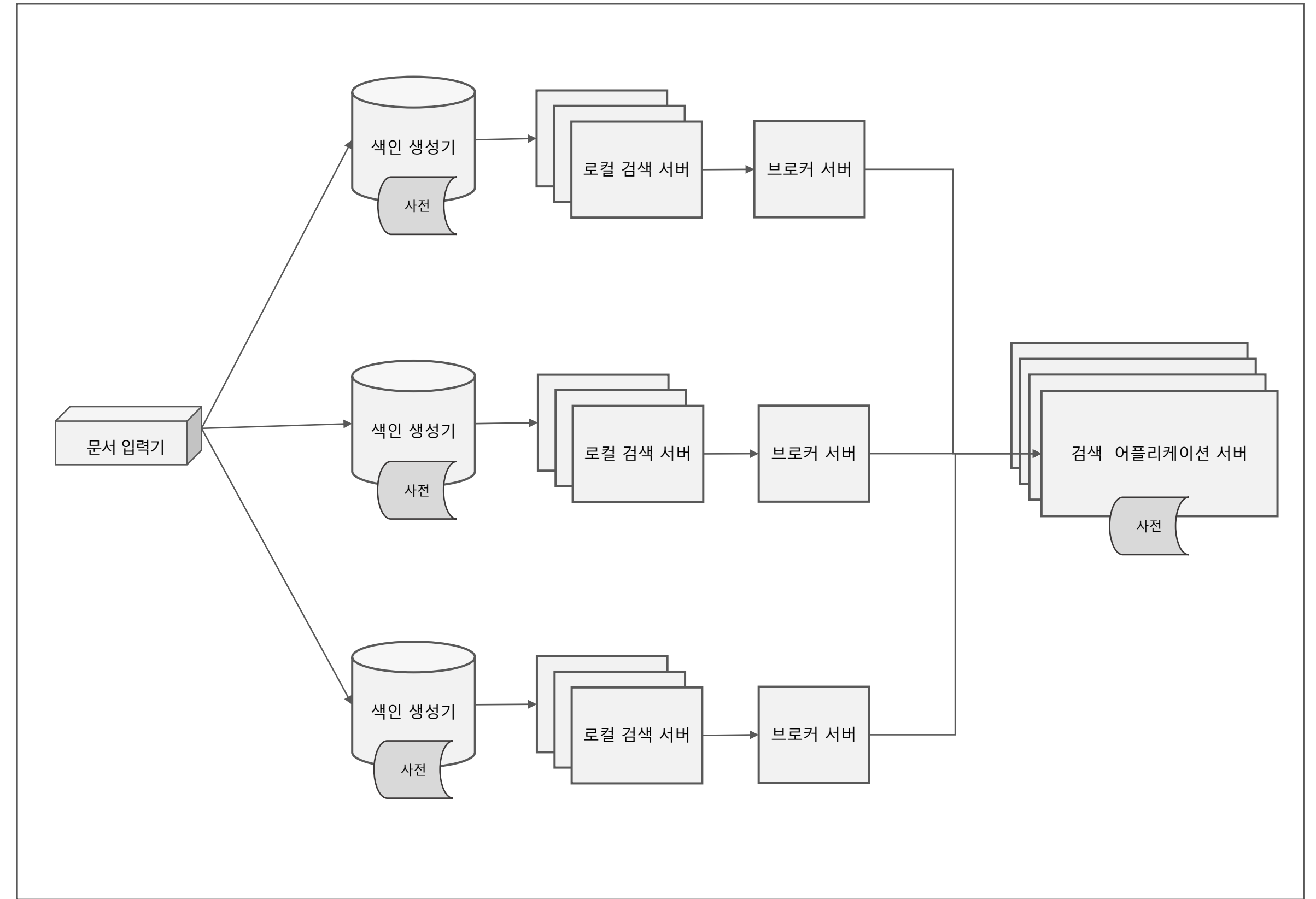
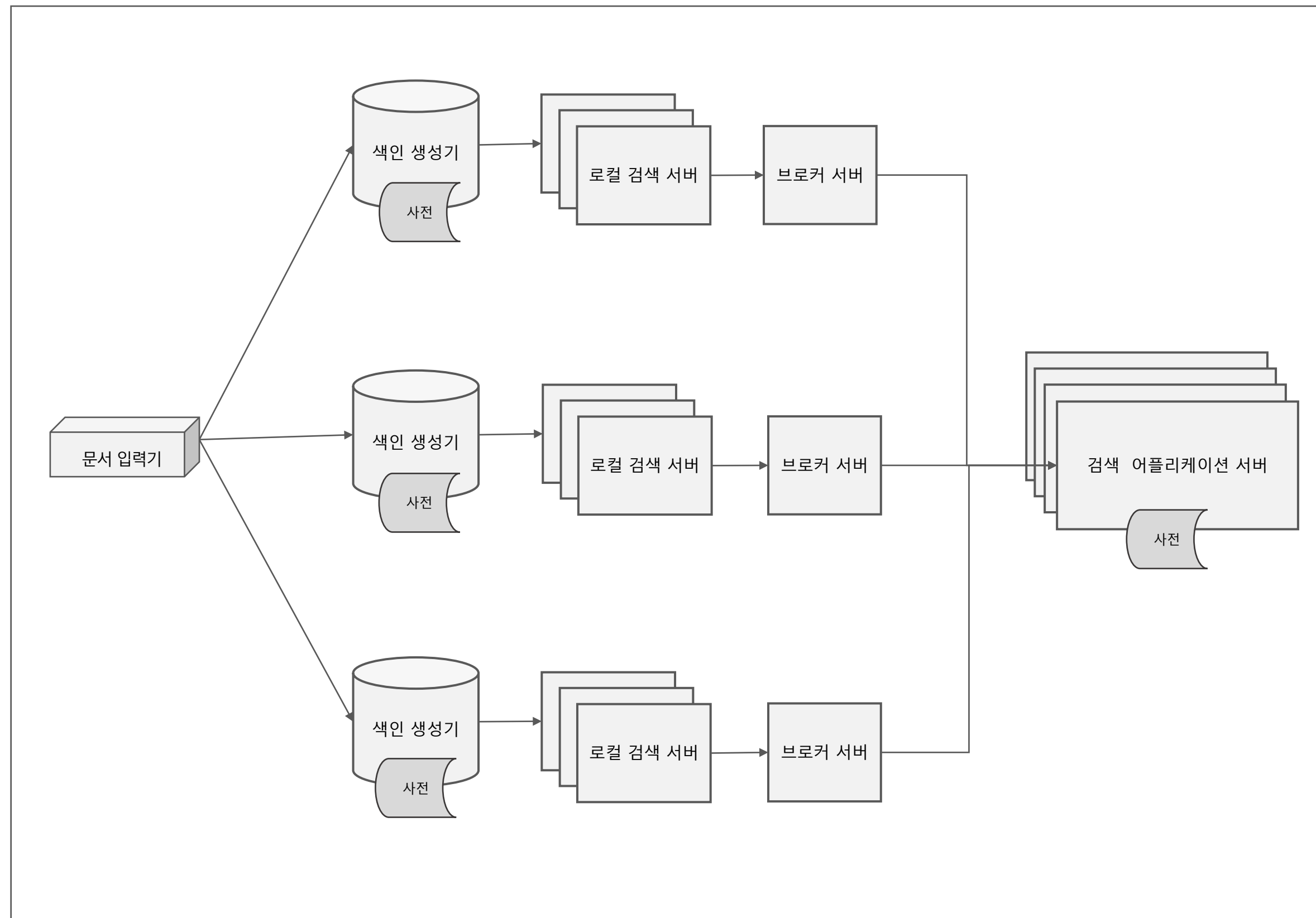
## 클러스터 관리 및 운영



# 4.3 이미지 검색시스템 운영

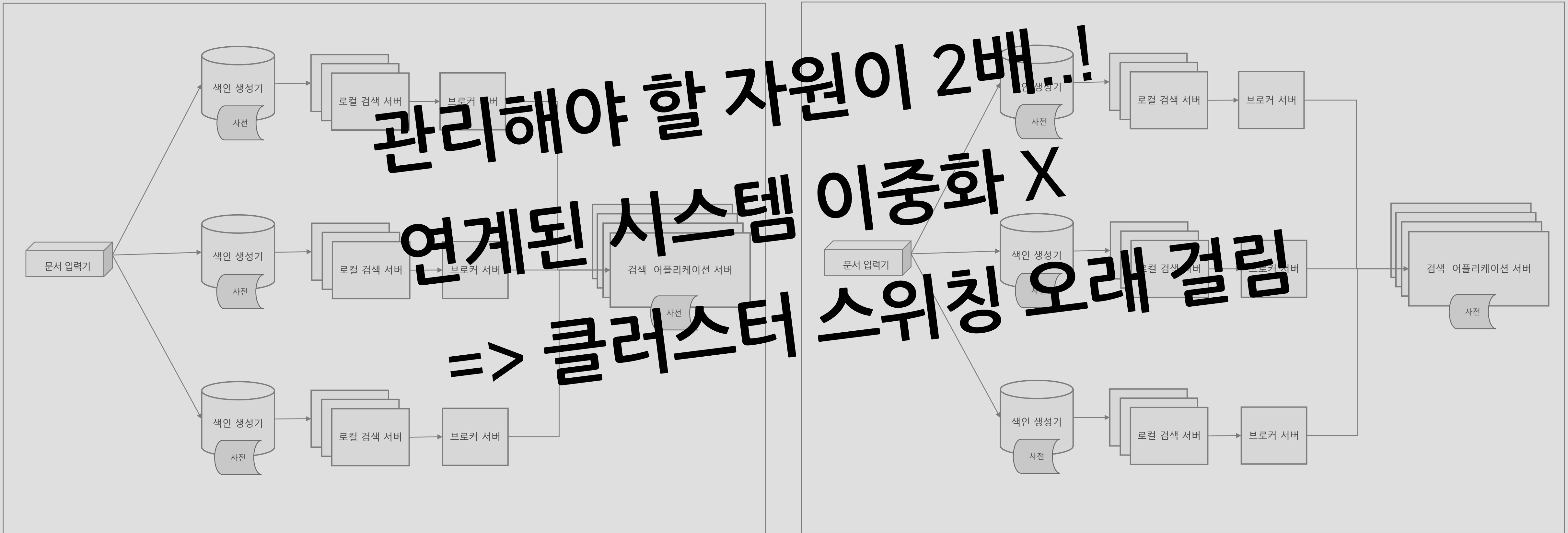


# 4.3 이미지 검색시스템 운영



# 4.3 이미지 검색시스템 운영

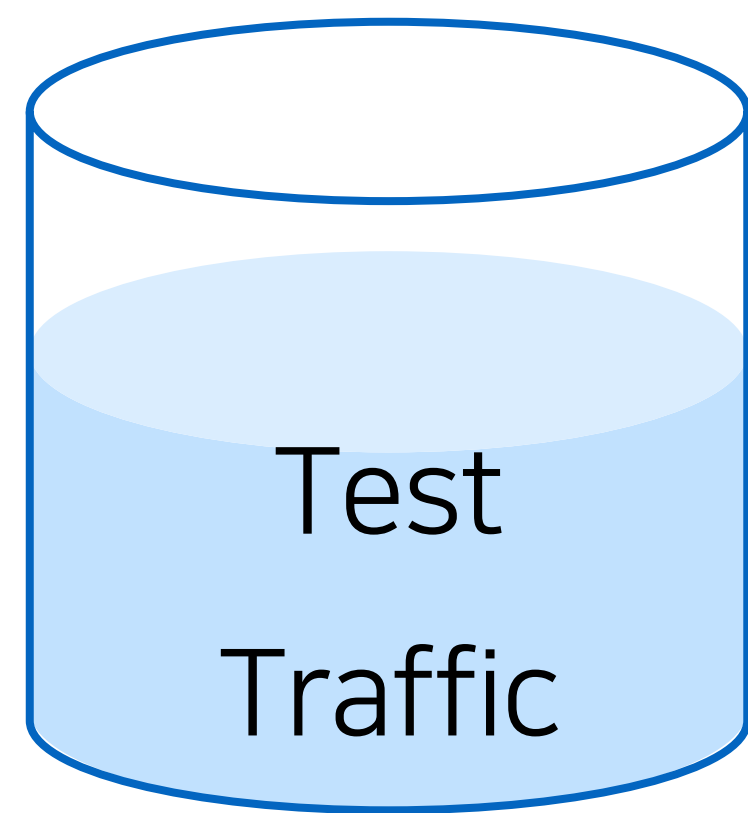
**관리해야 할 자원이 2배..!**  
**연계된 시스템 이중화 X**  
**=> 클러스터 스위칭 오래 걸림**





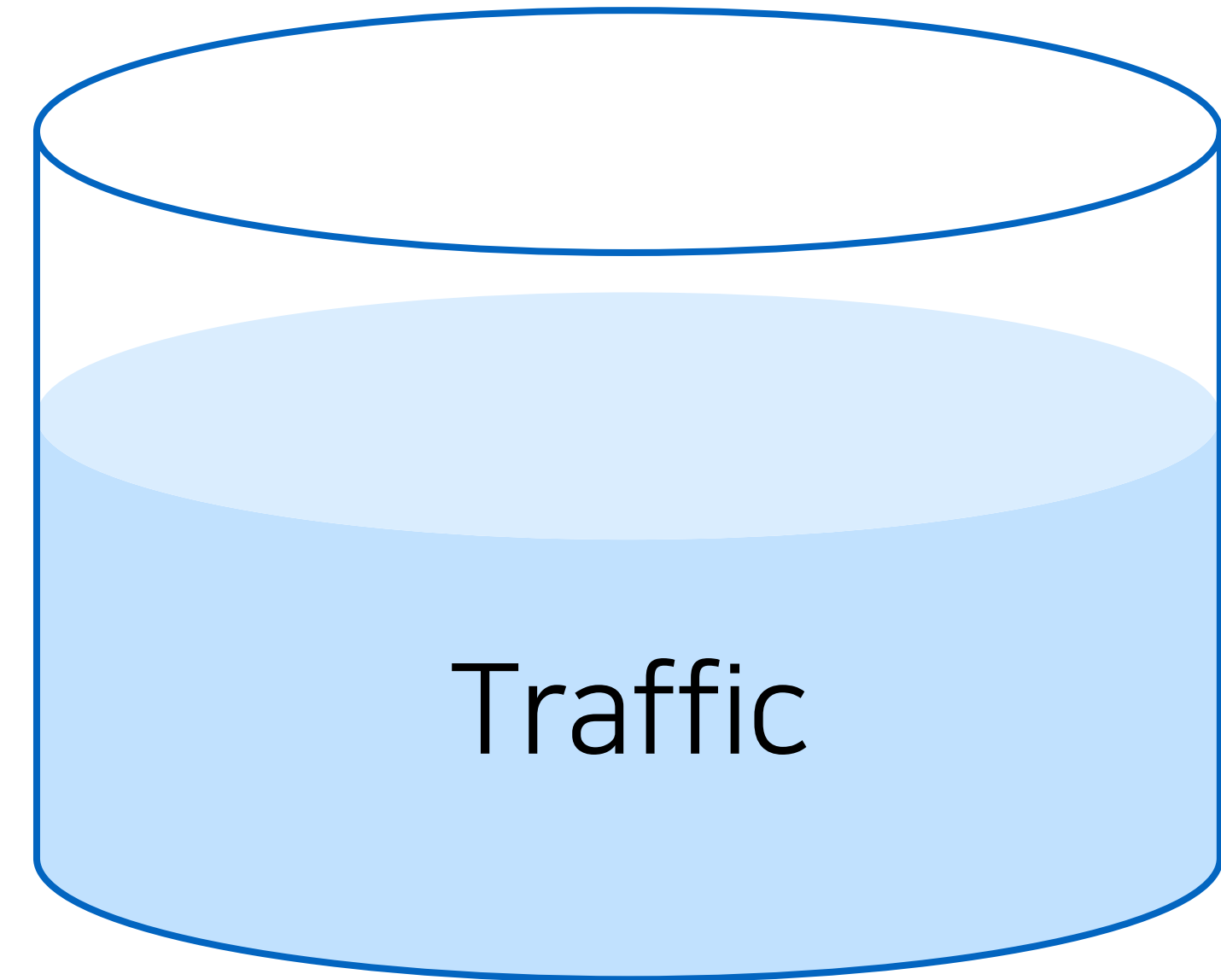
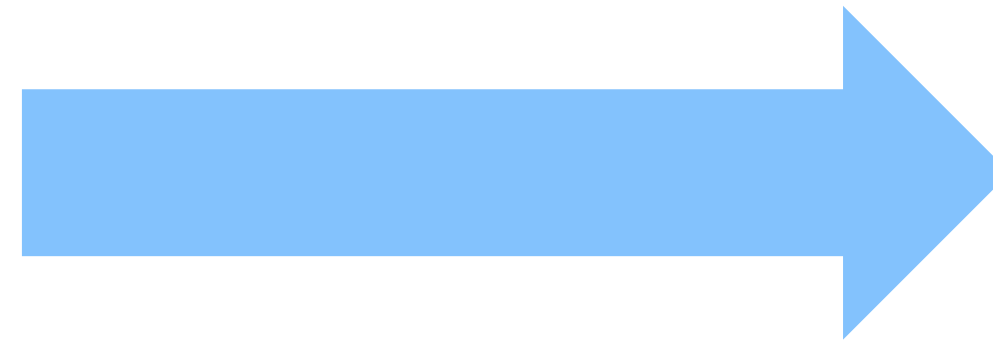
# 4.3 이미지 검색시스템 운영

## Real - Staging 클러스터 운영



Staging

테스트 후 적용



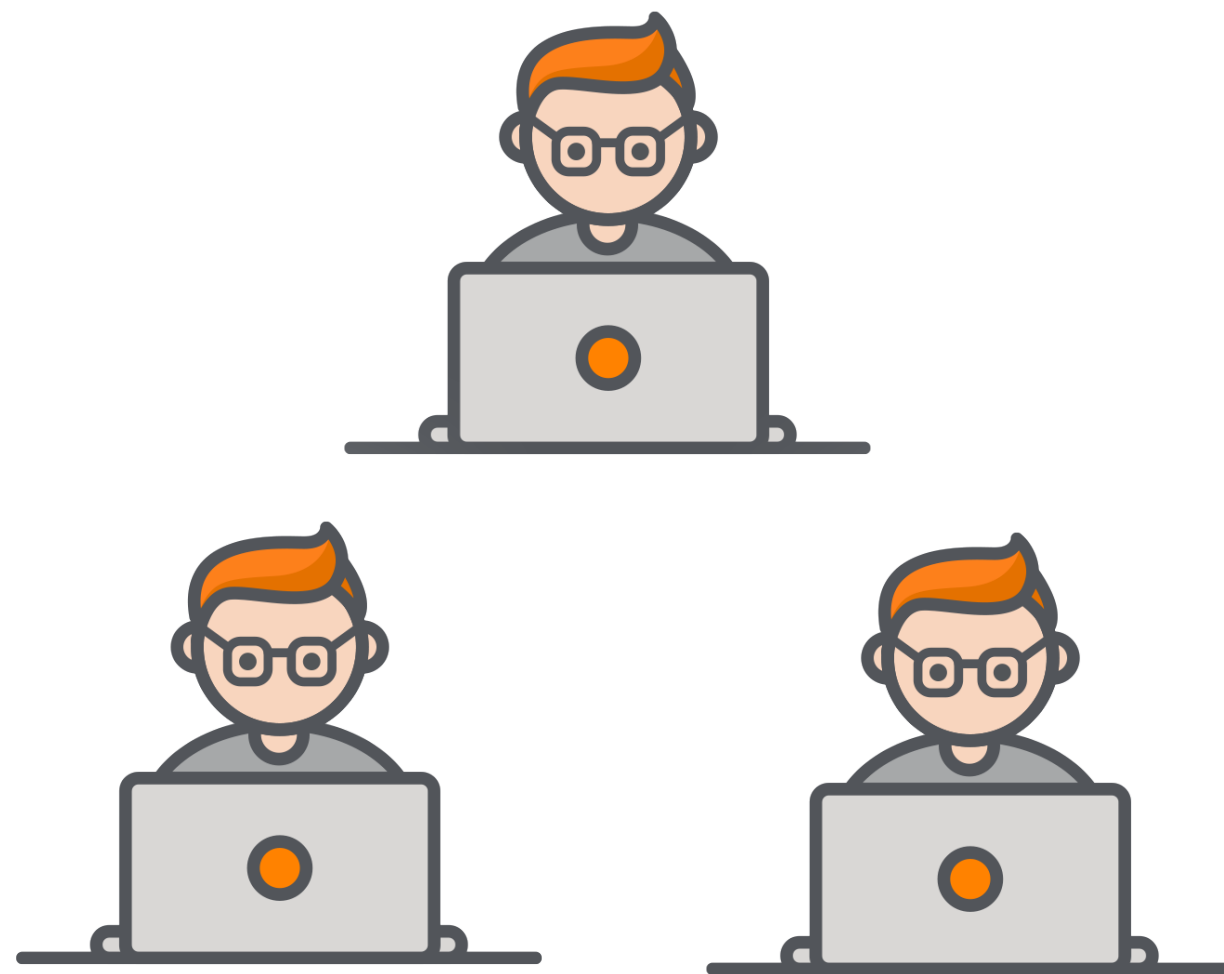
Real

## 4.3 이미지 검색시스템 운영

- ✓ 최근 2년간 이미지 검색의 장애 총 0건
- ✓ Defense-System-Outage(DSO) 작성으로 장애 사례 공유
- ✓ 검색 조직 서비스 안정성에 기여

## 4.3 이미지 검색시스템 운영

### 계속되는 변화와 적응



- 계속해서 변화하는 시스템에 대한 이해와 적응
- 담당자간 최적의 운영 방법 논의
- 연관 부서와의 커뮤니케이션 채널 유지
- 지속적인 검색 시스템 개선 노력

# 5. 앞으로의 DevOps

# 5.1 DevOps에 영향을 미친 주요요소

이미지 검색  
랭킹 로직  
복잡도 증가

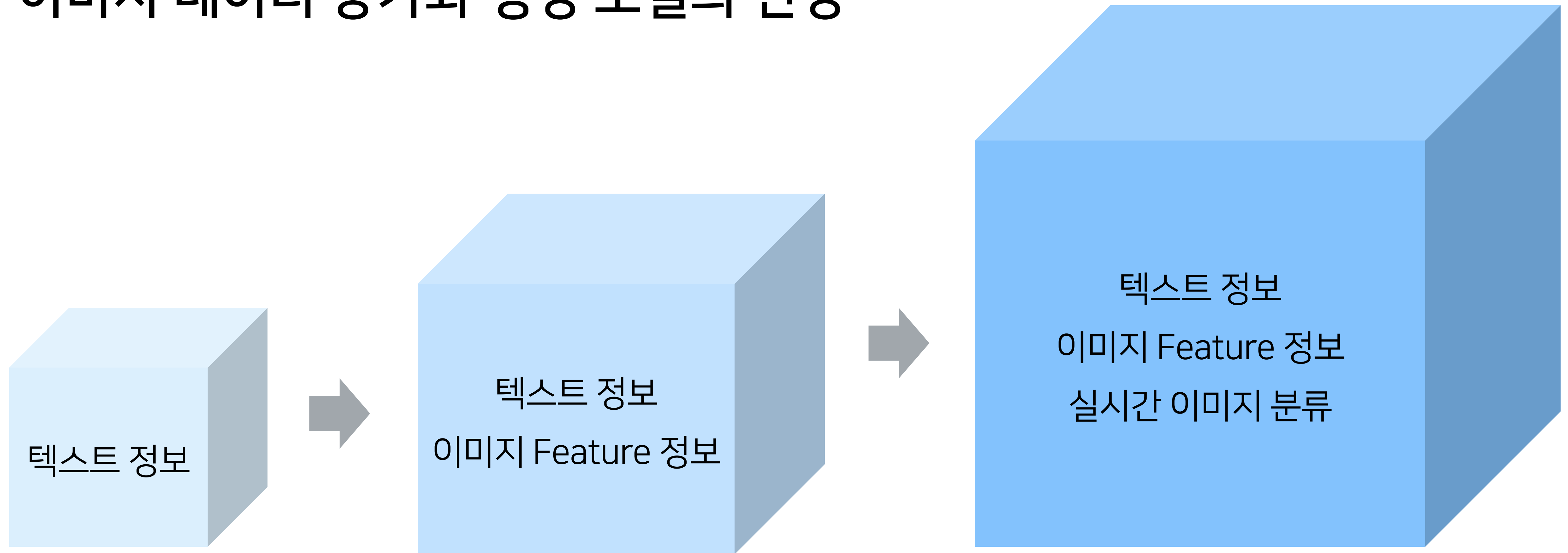
검색 플랫폼  
전문화

검색 교육  
체계화

장애 대응  
고도화

# 5.1 DevOps에 영향을 미친 주요요소

## 이미지 데이터 증가와 랭킹 모델의 변경



13억 건 → 60억 건으로 전체 이미지 데이터 증가

# 5.1 DevOps에 영향을 미친 주요요소

## 검색 관련 플랫폼의 전문화

- 다양한 활용 가능한 플랫폼
- 운영 비용 감소
- 검색 엔진 증분화와 고도화

# 5.1 DevOps에 영향을 미친 주요요소

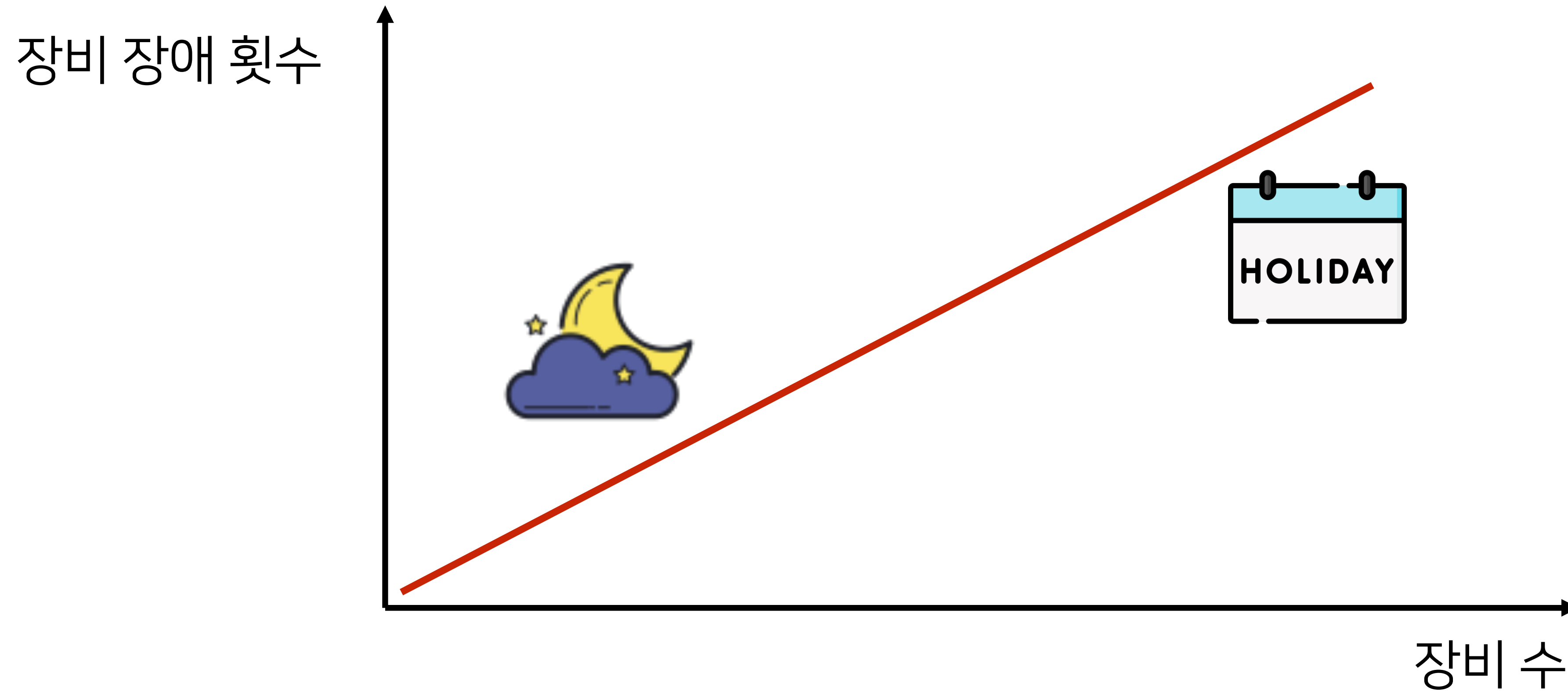
## 검색 교육의 체계화

- 신규사원 교육을 위한 조직 및 팀내 교육 프로그램 준비
  - 10년 전에는 간략한 문서와 구두의 시스템 설명
- 대규모 검색시스템을 운영 할 수 있을 때까지 팀내 교육진행
  - 기초교육 3개월 심화 완료까지 약 12개월 이상 계속 진행



# 5.1 DevOps에 영향을 미친 주요요소

## 장비 수 증가에 따른 장비장애 발생 수 증가



# 5.1 DevOps에 영향을 미친 주요요소

## 검색 시스템 모니터링 구축

- 모니터링을 통한 시스템 지표 확인, 알림 및 대응 체계
- 시스템 구축이 되었다고 DevOps를 할 수 없음
- 검색시스템 구축 시 연관 시스템 포함 관련된 시스템의 이상 지표 확인 가능하도록 세팅
- 오류나 이상 지표가 확인된 경우 관련 담당자에게 알림 발송 되고 상황에 맞게 대응되도록 고도화 되고 있음

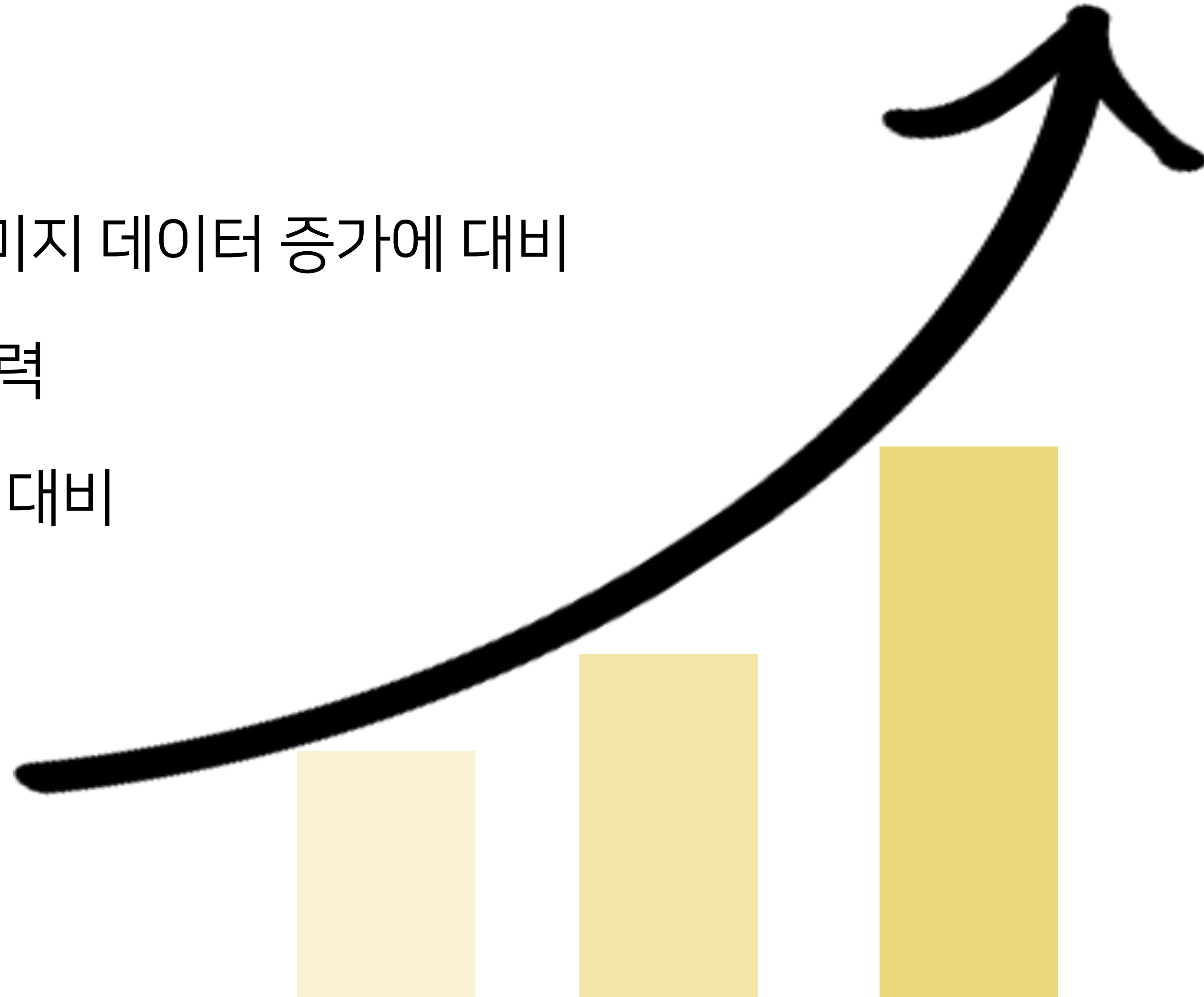
## 5.2 이미지 검색시스템의 미래

### 증분 시스템, 컨테이너 환경으로

- 많은 검색서비스가 증분, 컨테이너 환경에서 서비스되고 있음
- 이미지검색도 증분 시스템으로 전환을 검토 중이나 운영 효율성 측면에서 전환 안함
- 서버 500이상 필요한 대규모 검색시스템이 컨테이너 환경에서 운영되는 아직 없음
- 멀지 않은 미래에 증분 시스템, 컨테이너 환경으로 전환 기대

## 5.2 이미지 검색시스템의 미래

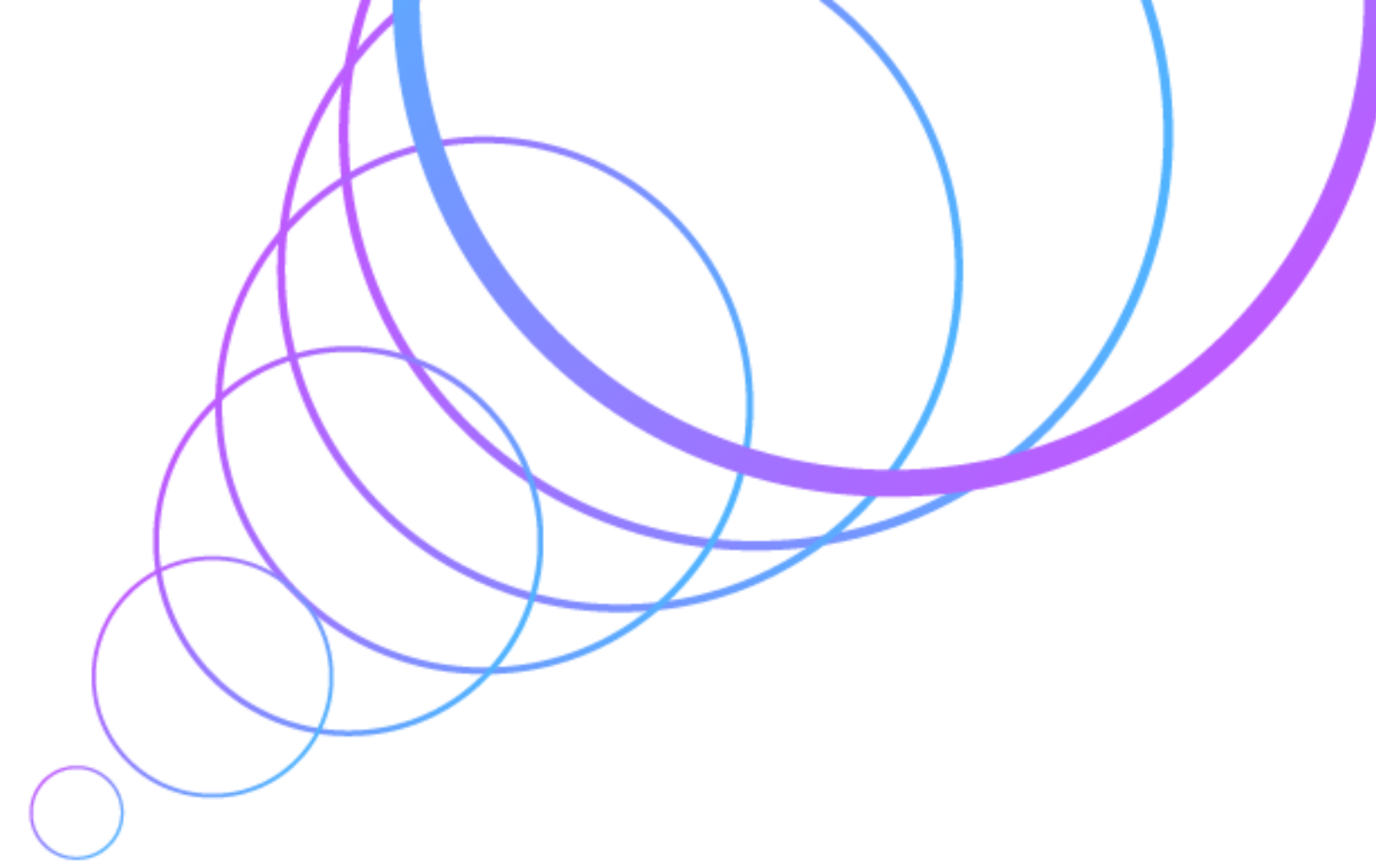
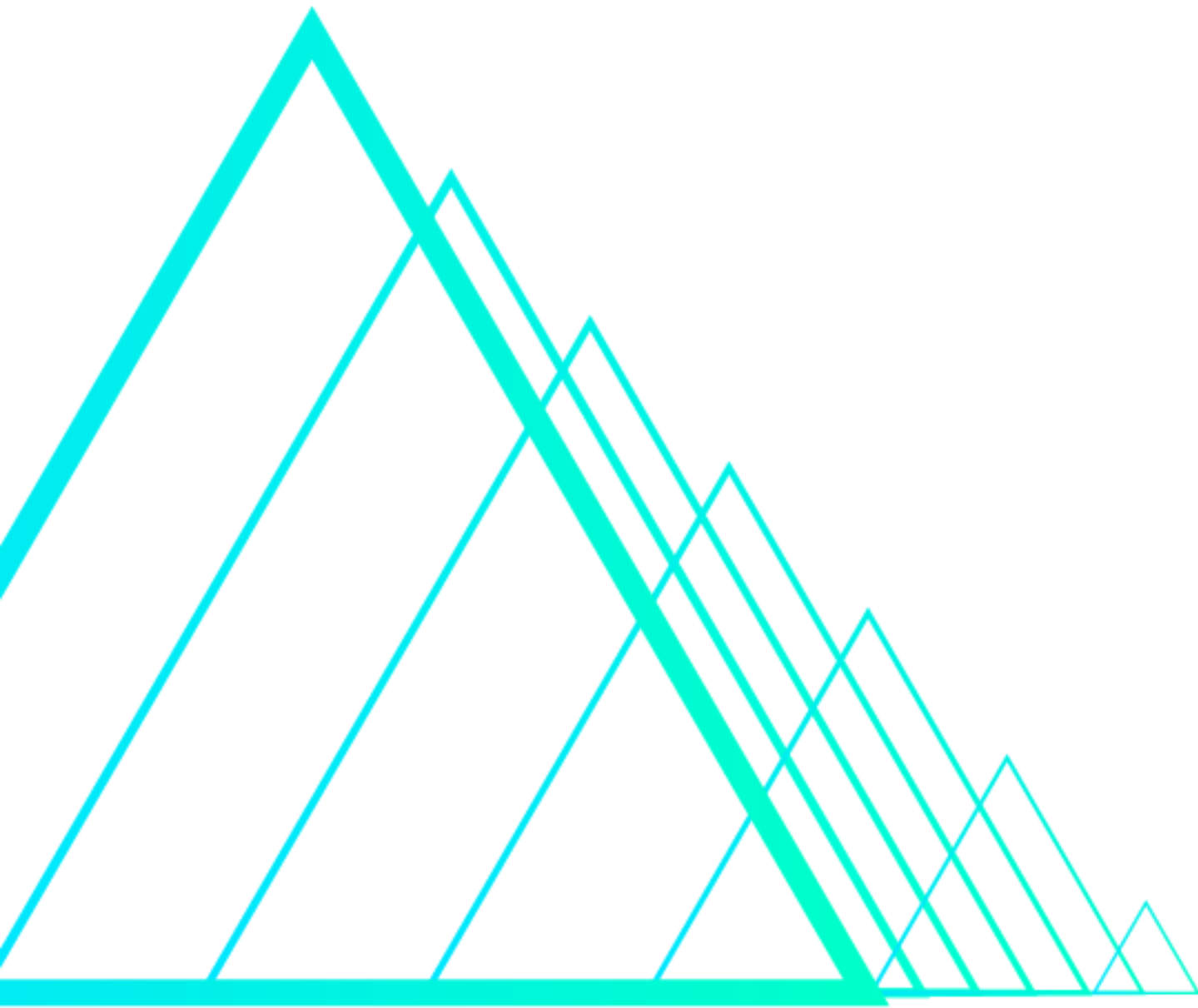
- ✓ 100억 이상의 이미지 데이터 증가에 대비
- ✓ 검색 품질 향상 노력
- ✓ 장비 수 증가 따른 대비



## 5.3 마무리

### 검색 시스템 DevOps

- 다양한 플랫폼을 통한 대한민국 검색 서비스를 서빙 할 수 있는 경험
  - 검색 도메인 습득과 다양한 개발 및 플랫폼 활용
    - 시스템 엔지니어로서 성장 가능
  - 서비스 운영 안정성을 도모하며 문제 해결 능력을 키울 수 있음
- 앞으로도 DevOps의 난이도는 올라가고 고민 point도 많아질 것.



**Thank You**

