

쿠팡 추천시스템 변천사

상품추천에서 실시간 개인화로

오성민

Recommendation

coupang

CONTENTS

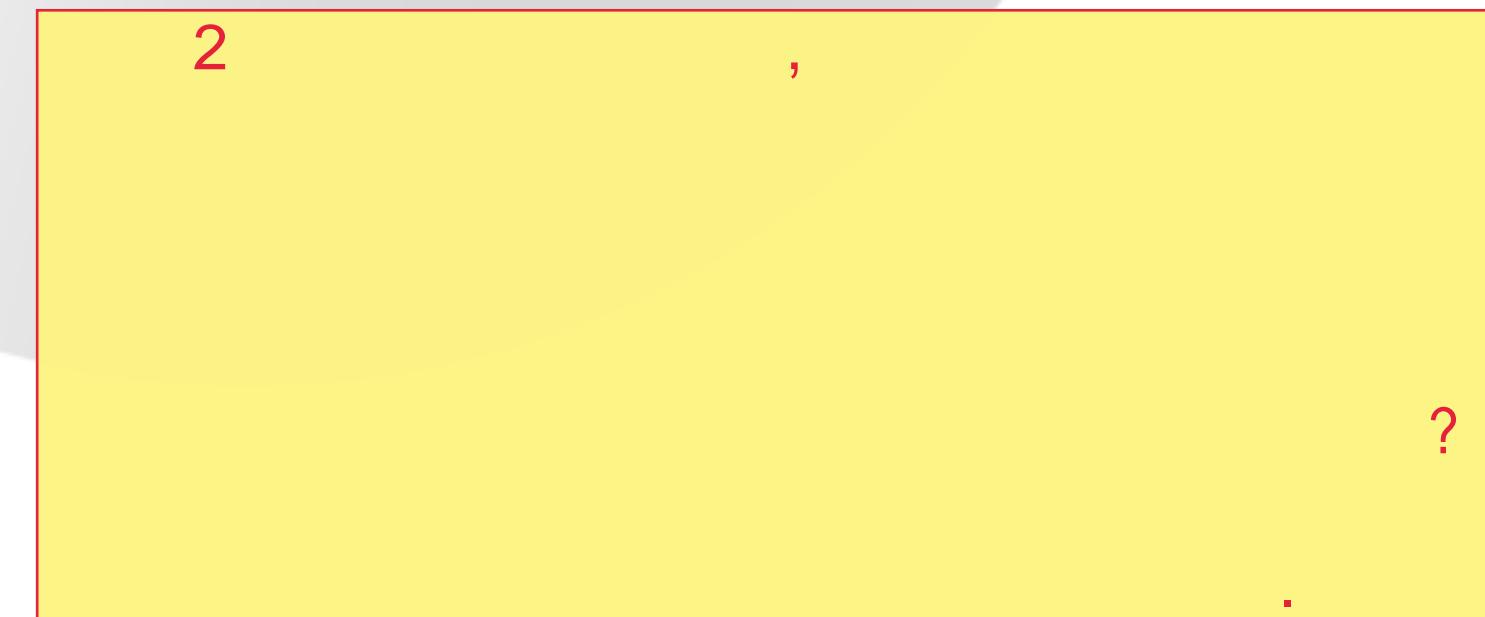
I. 쿠팡 추천 시스템의 변화

1. 쿠팡 추천 팀이 하는 일
2. 과거, 추천 모델 중심의 플랫폼
3. 현재, 서비스와 모델을 분리하는 플랫폼

II. 더 나은 추천을 위해

4. Learning to Rank
5. 실시간 개인화

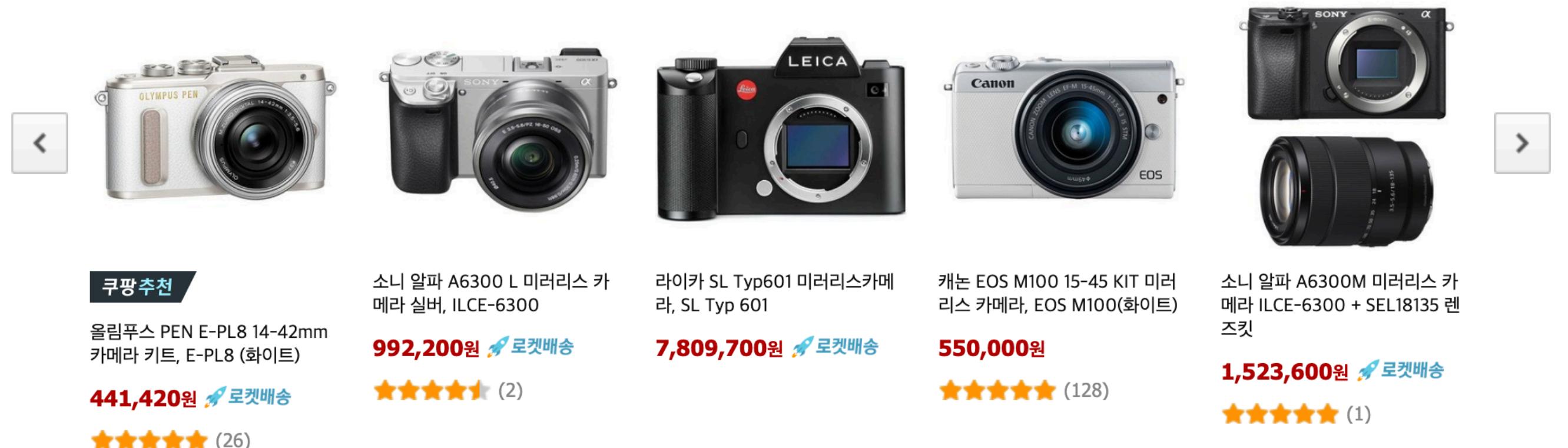
III. 요약 & 팁



1. 쿠팡 추천 팀이 하는 일

1. 과거, 추천 모델 중심의 플랫폼
2. 현재, 서비스와 모델을 분리하는 플랫폼

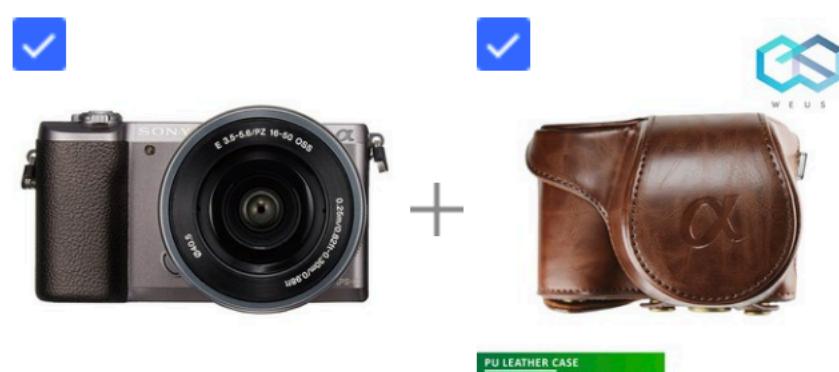
다른 고객이 함께 본 상품



함께 본 대체재 함께 산 보완재



다른 고객이 함께 구매한 상품



현재 상품 : 소니 미러리스카메라, ILCE-5100L 64G 패키지...
499,000원

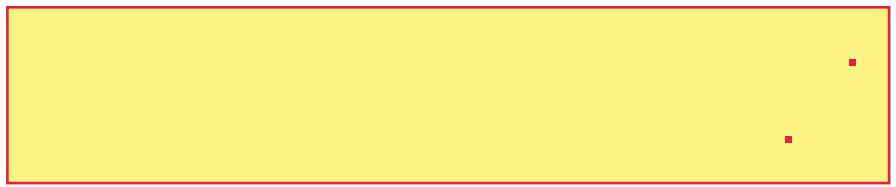
소니 A5100 케이스
11,900원

2개 상품 장바구니 담기

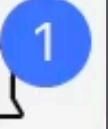
이 외에도 구매한 상품



소니 A5000 A5100 A6000 A6300 케이스, 미러리스 삼각대 17,500원 ★★★★★ (32)	소니 알파 단렌즈 Sonnar T* FE 55mm F1.8 ZA... 926,000원 ★★★★★ (61)	소니 알파 줌렌즈 E PZ 18-105mm F4 G OSS... 695,000원 ★★★★★ (68)	시그마 AF 30mm 소니E마운트용 480,000원 ★★★★★ (59)
--	---	--	--



하고 싶었던 것



(가)

가 ,

context
context
context
context

()

- > *

()

()

- > *

Private Brand (PB)

주말장보기

내가 본 상품의 연관상품

더보기 >



로켓프레시

과일 채소 정육 계란/유제품 수산/건어물

베이커리 디저트/음료

자주 산 상품

전체보기 >

존쿡델리미트 이탈리아 프로슈토, 50g, 1개
6,500원

로켓프레시

장바구니 담기

테이트 이탈밀 치킨머쉬룸 치아바타 샌드위치 (냉동), 1개
8,040원

로켓프레시

장바구니 담기

테이트 누우 치아바타 샌드위치 (냉동), 1개
8,550원

로켓프레시

장바구니 담기

로켓

보관방법 > **간편/대용식 섭취방법 >** **용기 탑입 >**

전시레이인지 맨위로

5,209개의 상품

쿠팡 랭킹순

1

남성패션

실시간 인기상품

바이멜로우 하이스판 슬림일자 라이트옐로우블루 밴딩…
21% ~~18,900원~~
14,900원
★★★★★ (9)

맨즈북 남성용 크롭 슬랙스 8부
51% ~~34,000원~~
16,500원
★★★★★ (134)

직수입 남성 융털내포 모 청바지 sj0063
41% ~~56,000원~~
32,800원
★★★★★ (101)

이런 상품 어때요? 광고 i

티센스 반팔 쿨론 기능성 카라티 쿨버튼
5,900원
★★★★★ (18)

맥태비쉬 슈퍼 롱패딩 커플파딩 빅사이즈패딩 노옵션군…
39,900원
★★★★★ (1,912)

핫코드 남성용 빅사이즈 HC100 맨위로
18,900원
★★★★★ (115)

1

Category
+ Time
+ User
+ Items

item 가 item

Filter
Weight
Relations
Item
Reco
UI

filter ..
context ..
context weight ..
...
theme ..
UI ..

변화

	과거	현재
Input	상품	상품, 유저, 카테고리, 등
Output	상품	상품, 추천 영역, 등
Feature	모델 점수	모델 점수들, 개인화, 등

1. 쿠팡 추천 팀이 하는 일
2. 과거, 추천 **모델** 중심의 플랫폼
3. 현재, 서비스와 모델을 분리하는 플랫폼

Item—**Relation**—Item

Relation → **Service**
= Model

Relation Model

2.1.1 아키텍처 - 데이터



상품 정보



유저 정보



후보 상품들

필터 적용 후

임베딩

결과

Source Product	Target Product	Score
카메라	케이스	0.4
카메라	삼각대	0.3



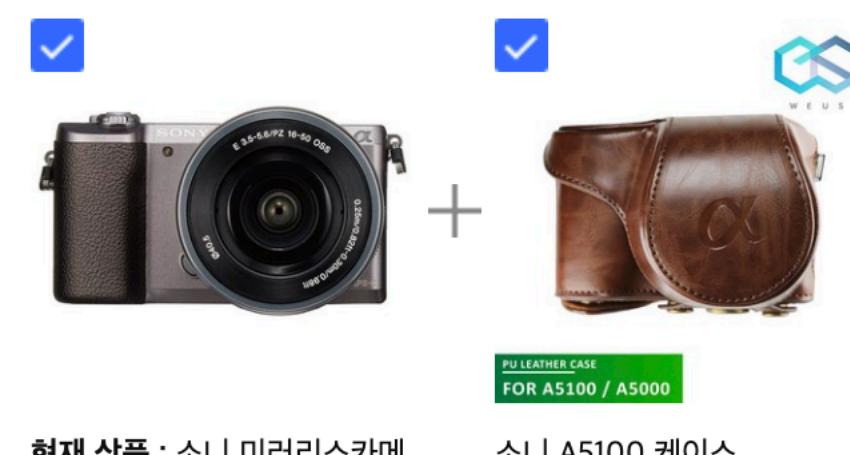
2.1.1 아키텍처 - 데이터

함께 산 상품

Source Product	Target Product	Score
카메라	케이스	0.4
카메라	삼각대	0.3



다른 고객이 함께 구매한 상품



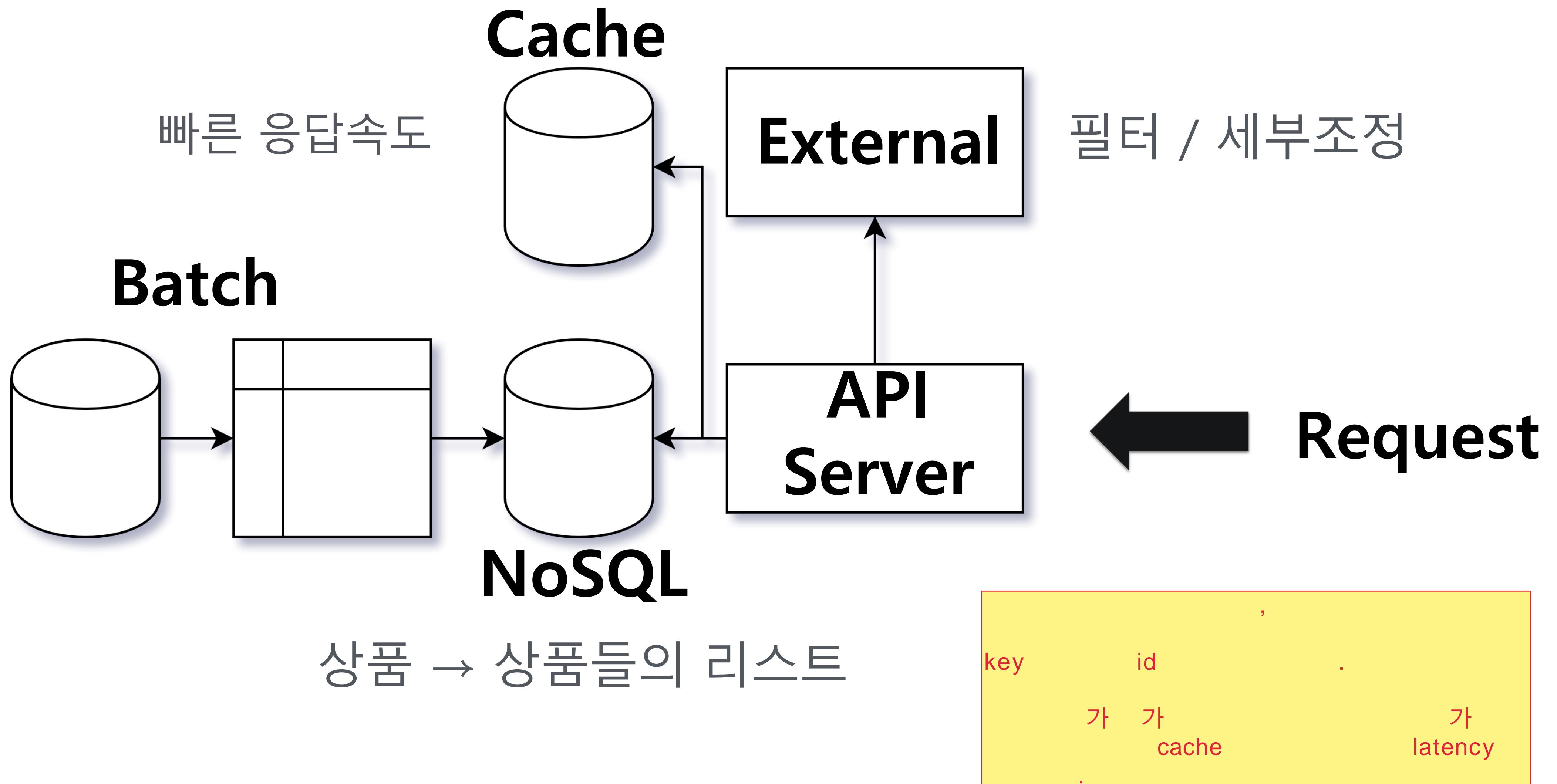
2개 상품 장바구니 담기

이 외에도 구매한 상품



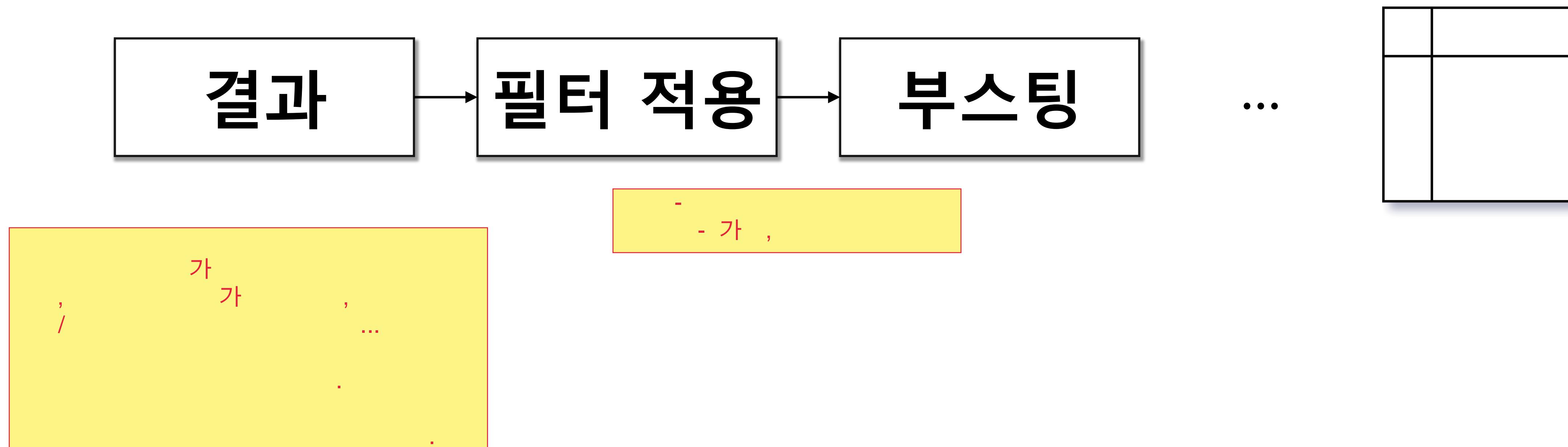
가 .

2.1.2 아키텍처 - 서버



2.2 단점/한계

1. 모델 변경에 따라 길어지는 파이프라인 (필터, 부스팅, ...)
2. 추가 요청사항을 처리하기 어렵다.
3. 완성 전까지 결과를 알 수 없다.
4. 개발에 시간이 오래 걸린다.



2.2 단점/한계

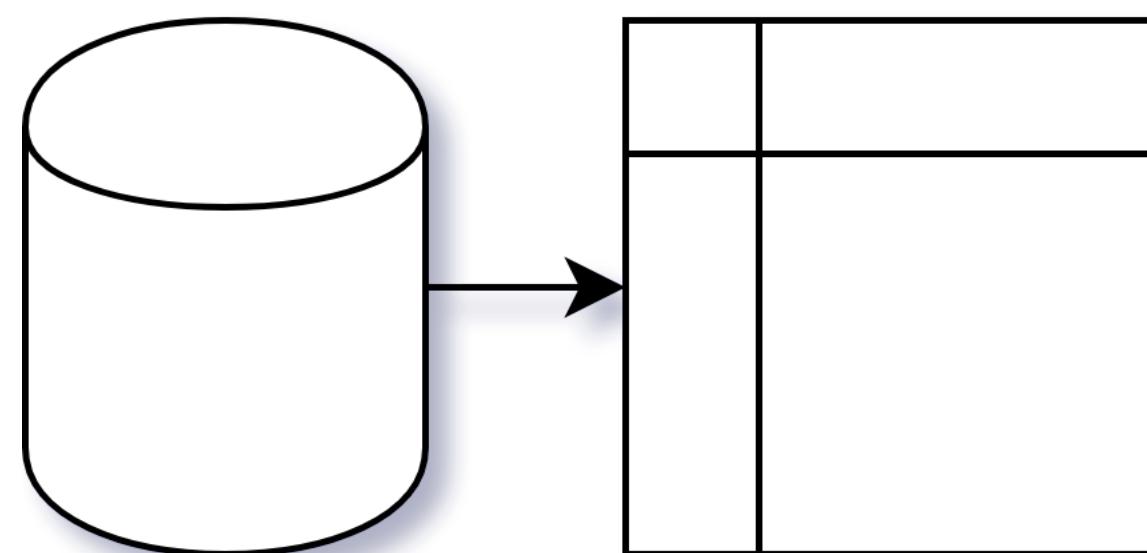


상품 정보 등은
데이터가 생성될 때만 가용

상품 정보

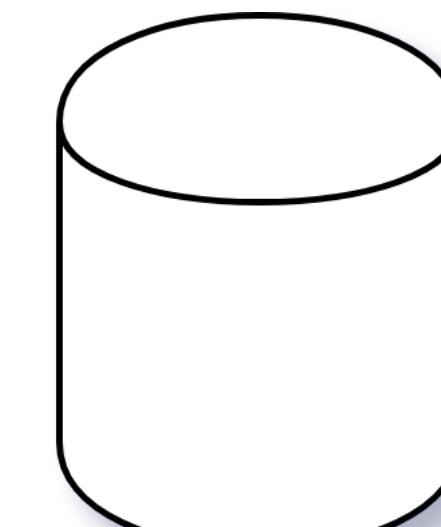


유저 정보



Batch

Cache



External

API Server

NoSQL

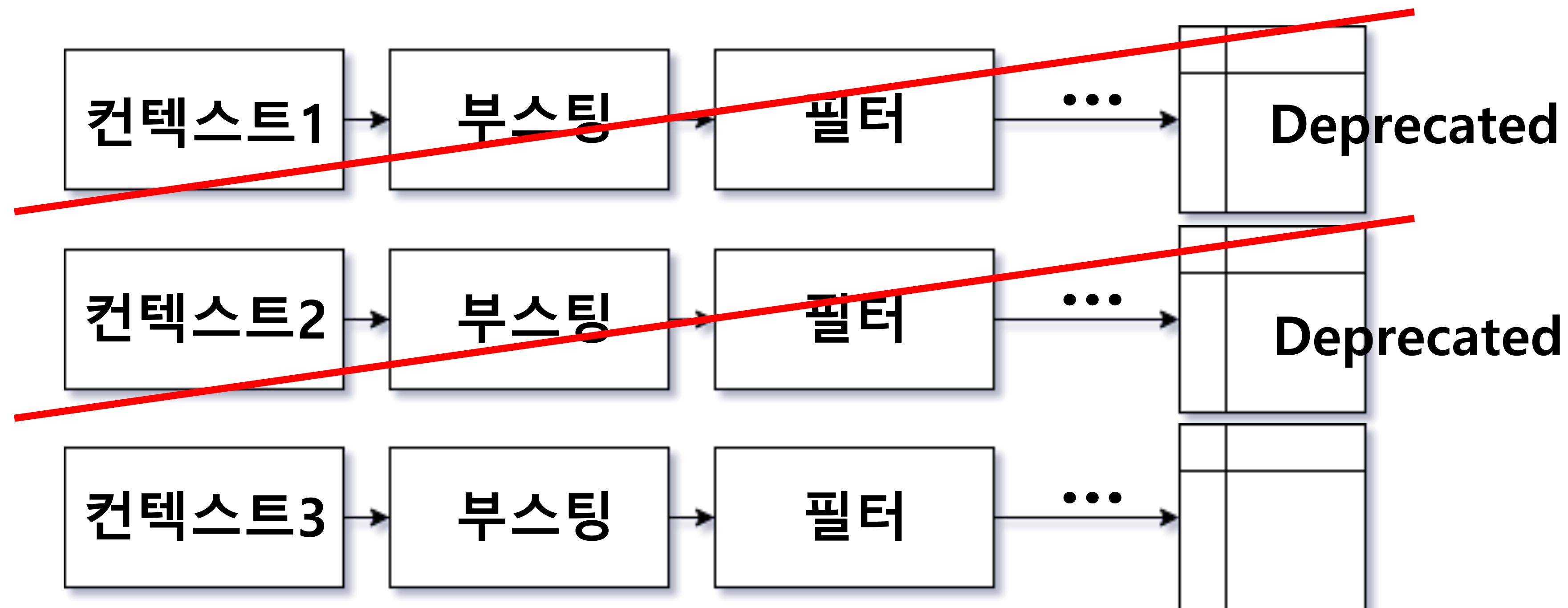


서빙 시에 접근이 어렵다

2.2 단점/한계

- 점진적인 개선이 힘들다.
- 모델의 재활용이 어렵다.

가 .



2.3 목표

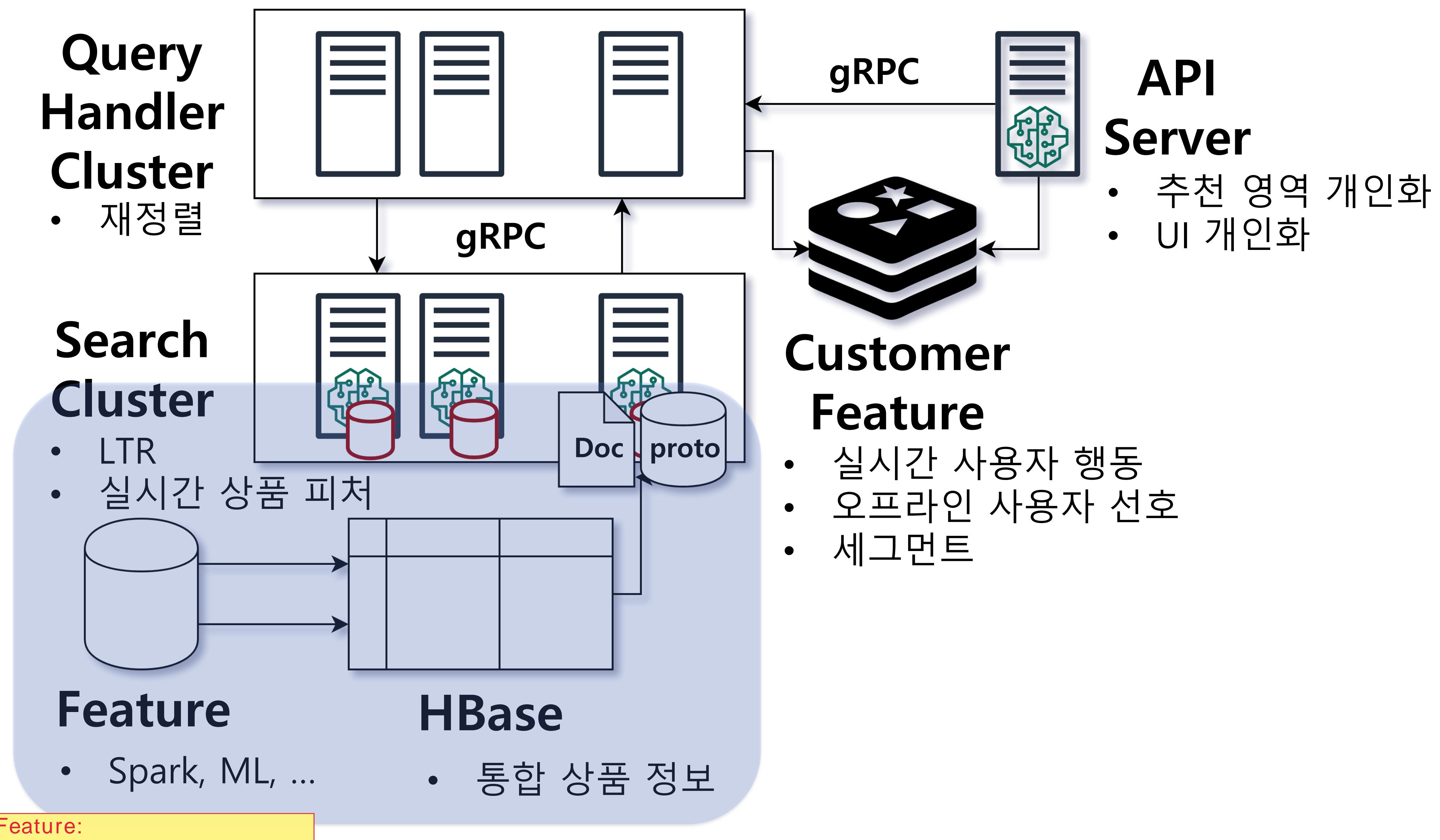
1. 추천 모델과 서비스를 **분리**시킬 것
2. 상품 정보나 유저 정보를 서빙 타임에 접근 가능할 것
3. 필터, 부스팅 등의 변경이 쉽고 빠를 것.

1. 쿠팡 추천 팀이 하는 일
 2. 과거, 추천 모델 중심의 플랫폼
- 3. 현재, 서비스와 모델을 분리하는 플랫폼**

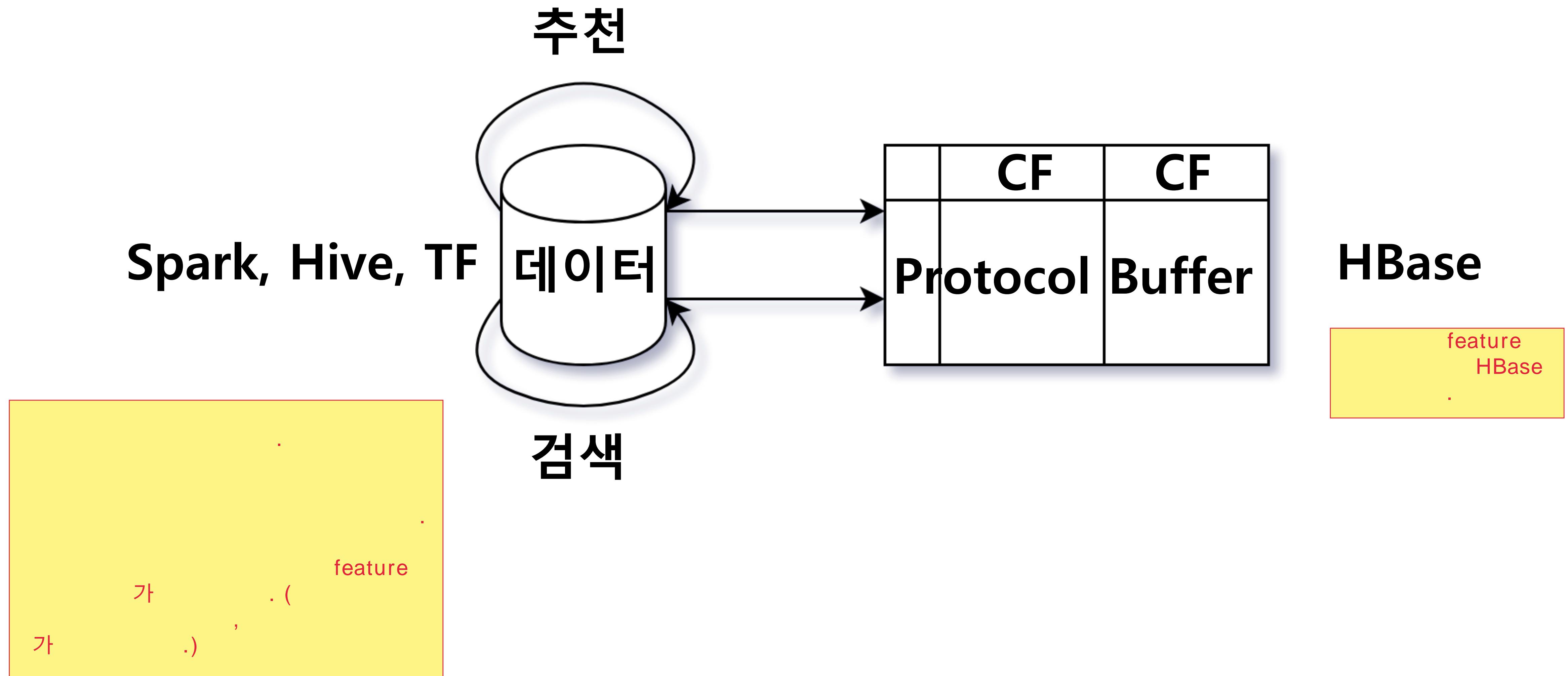
3.1 검색과 추천

검색	자연어	키워드 매칭	관련도
추천	상품/유저	모델 결과	모델 스코어





3.2 중앙화 된 상품 정보



3.2.1 Protocol Buffer

`search.mergeFrom(reco)`

```
message Product {
    int64 product_id = 1;
    Image image = 2;
    Category category = 3;
    int32 rating = 4;
    SearchSignals search_signals = 5;
    RecoSignals recommendation_signals = 6;
}

message RecoSignals {
    repeated RecoSource viewed_together = 1;
    repeated RecoSource bought_together = 2;
}

message RecoSource {
    int64 product_id = 1;
    double score = 2;
}
```

가
protocol buffer mergeFrom

장점

- gRPC
- 빠른 (디)시리얼리제이션
- 압축 효율
- 스키마 변경

주의

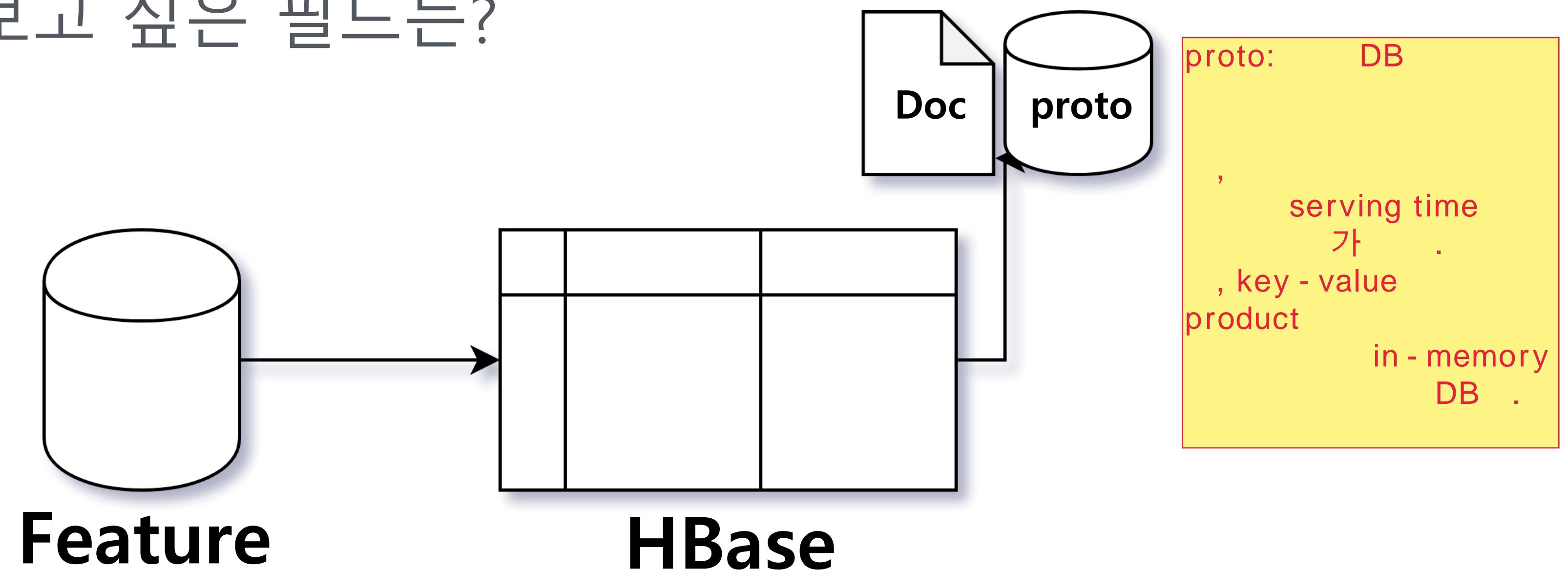
- 태그 넘버
- Null이 없음
- Hive 안됨

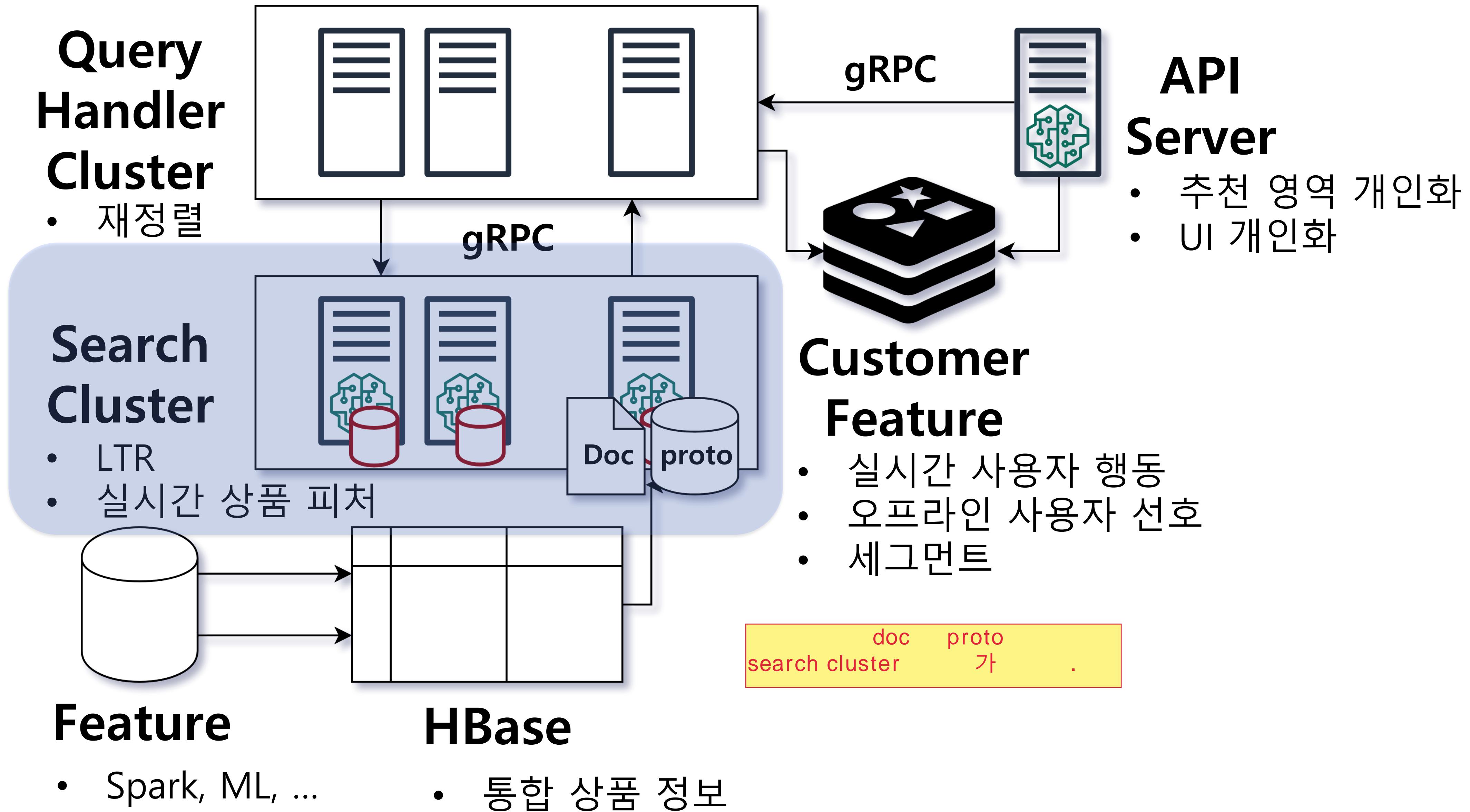
3.3 Feature

연관성 피처	상품 자체 피처
<ul style="list-style-type: none">• 추천 모델 결과• 이미지 유사도• 카테고리 관련성 <p>...</p>	<ul style="list-style-type: none">• 카테고리• 상품평• 가격 <p>...</p>

3.4 인덱싱

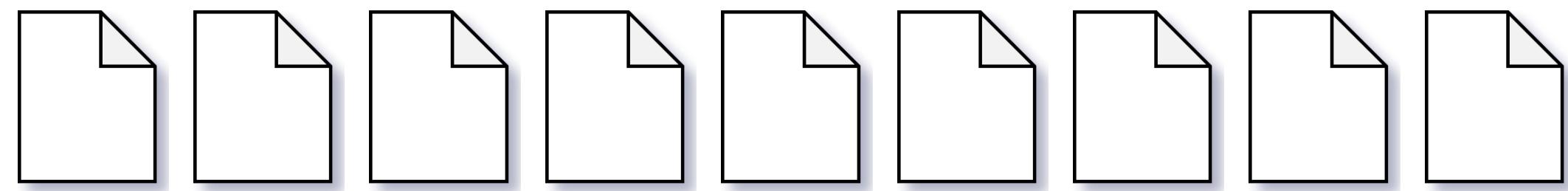
- 어떤 상품에 검색이 되어야 하나? viewed_together 123:0.4, 456:0.3, 789:0.1
- 필터나 부스팅은? category, is_adult, is_rocket
- 오브젝트 그대로 꺼내보고 싶은 필드는?

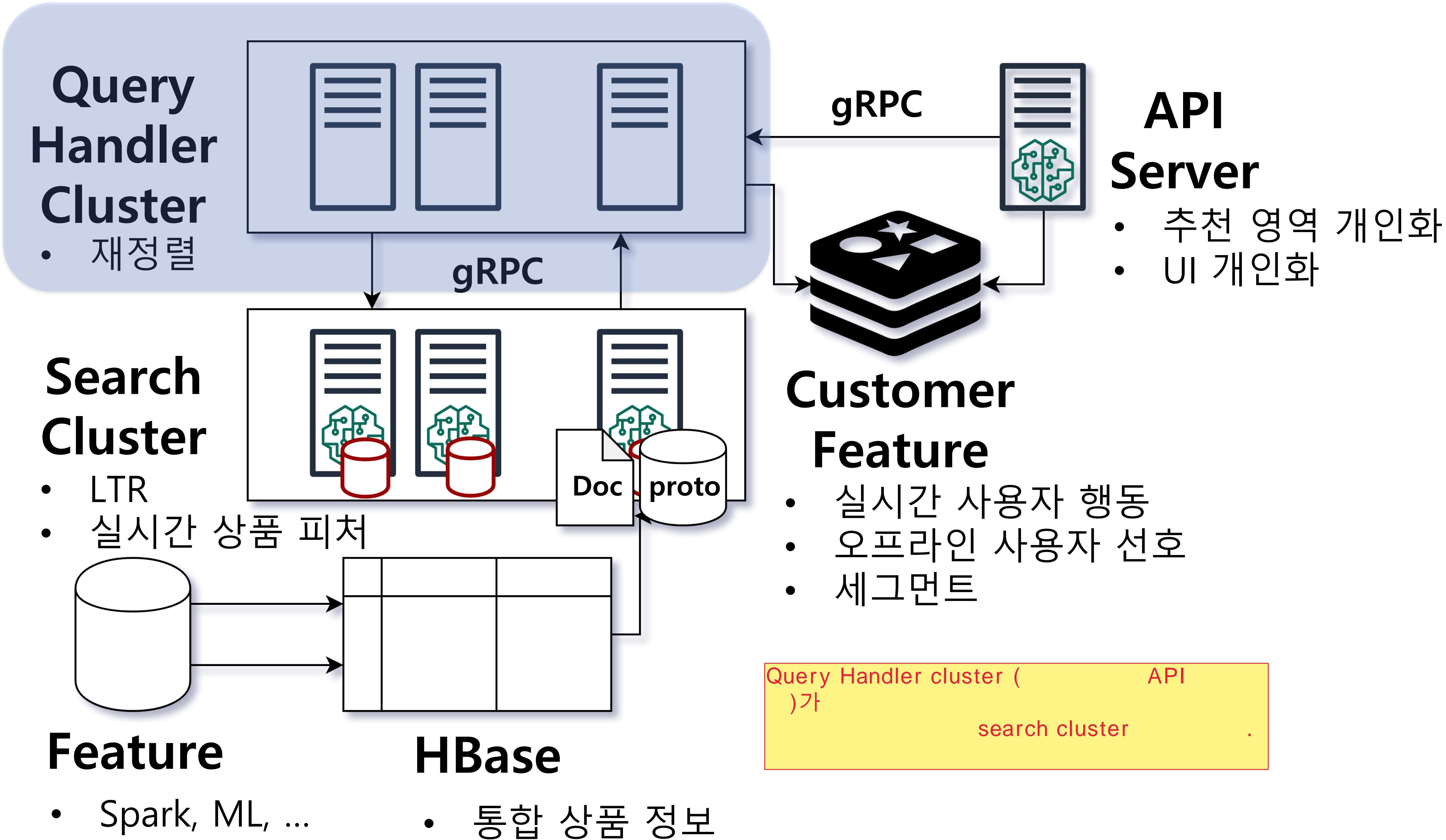




3.5 Search Cluster

- 컨텍스트와 관련된 상품을 찾고
- 조건에 따라 필터 한 다음
- 점수에 따라 정렬





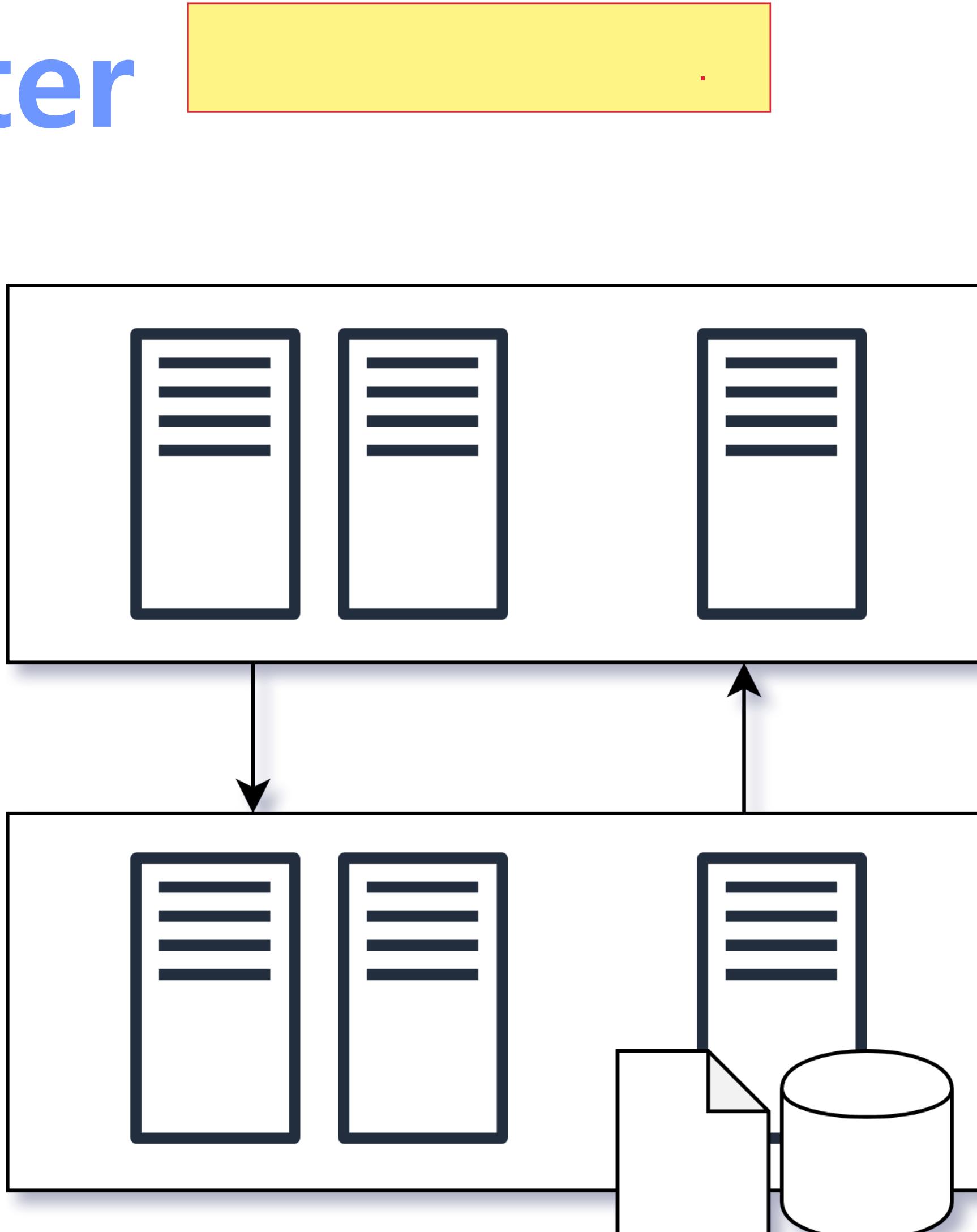
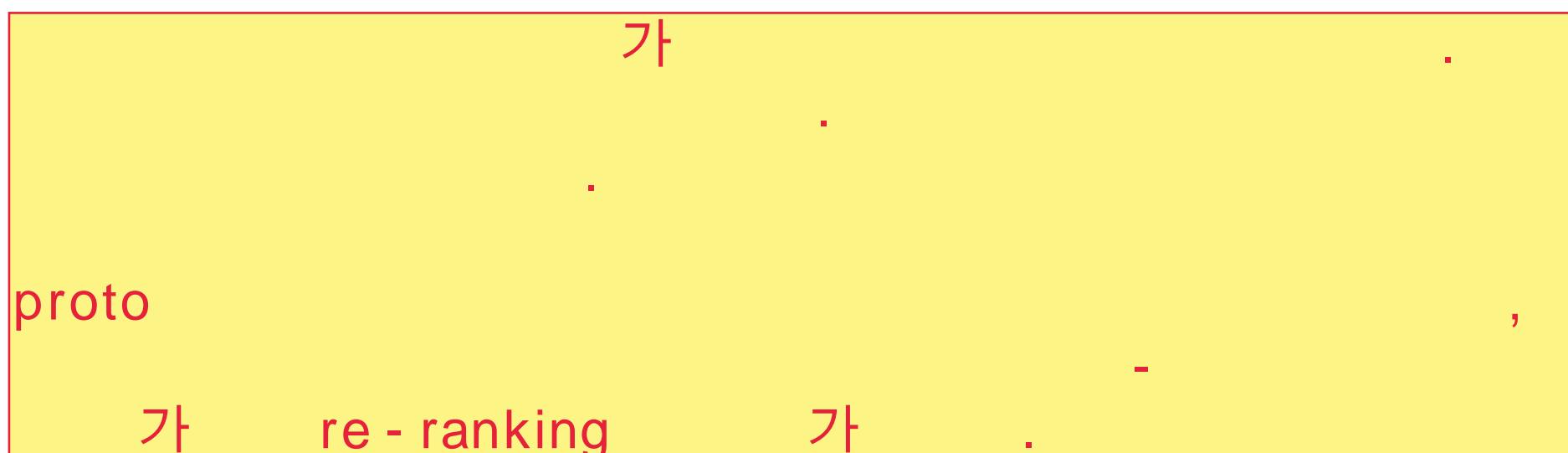
3.6 Query Handler Cluster

서비스 정의: 냄비와 함께 살 할인 식품

- Query: bought_together: 
- Filter: category: 
- Boost: discount_rate 

추가 정렬: 고기, 생선, 야채 번갈아 보여주자

```
sort(    ...  )
```



3.6 Query Handler Cluster

빠른 서비스 개발

1. 피처 조합과 가중치에 따라 새로운 서비스
2. 필터로 새로운 추천 서비스를 만들 수 있다.
(ex. 요즘 인기있는 000, 특가 상품)
3. 검색 결과를 필요에 따라 새롭게 정렬할 수 있다.
(ex. Round robin)



블랑제이 사각 파우치, 와인
레드, 1개

즉시할인가
30% 6,730원
4,700원
 로켓woo
 (1,059)



DM 다이나 화장품 파우치
백, 업그레이드된 버전, 1개

즉시할인가
35% 16,000원
10,270원
 로켓woo
 (48)



쿠첸 전기밥솥,
CJS-FB1011V

[더보기 >](#)



키친웰 벨로스세라믹코팅냄
비 4종 세트, 혼합 색상, ...

72,000원
59,000원
 로켓woo
 (7)



키친아트 프리미엄 세라믹 냄
비세트 3개, 검정, 편수냄...
골라담기 기획전, 01

28,000원
 로켓woo
 (85)

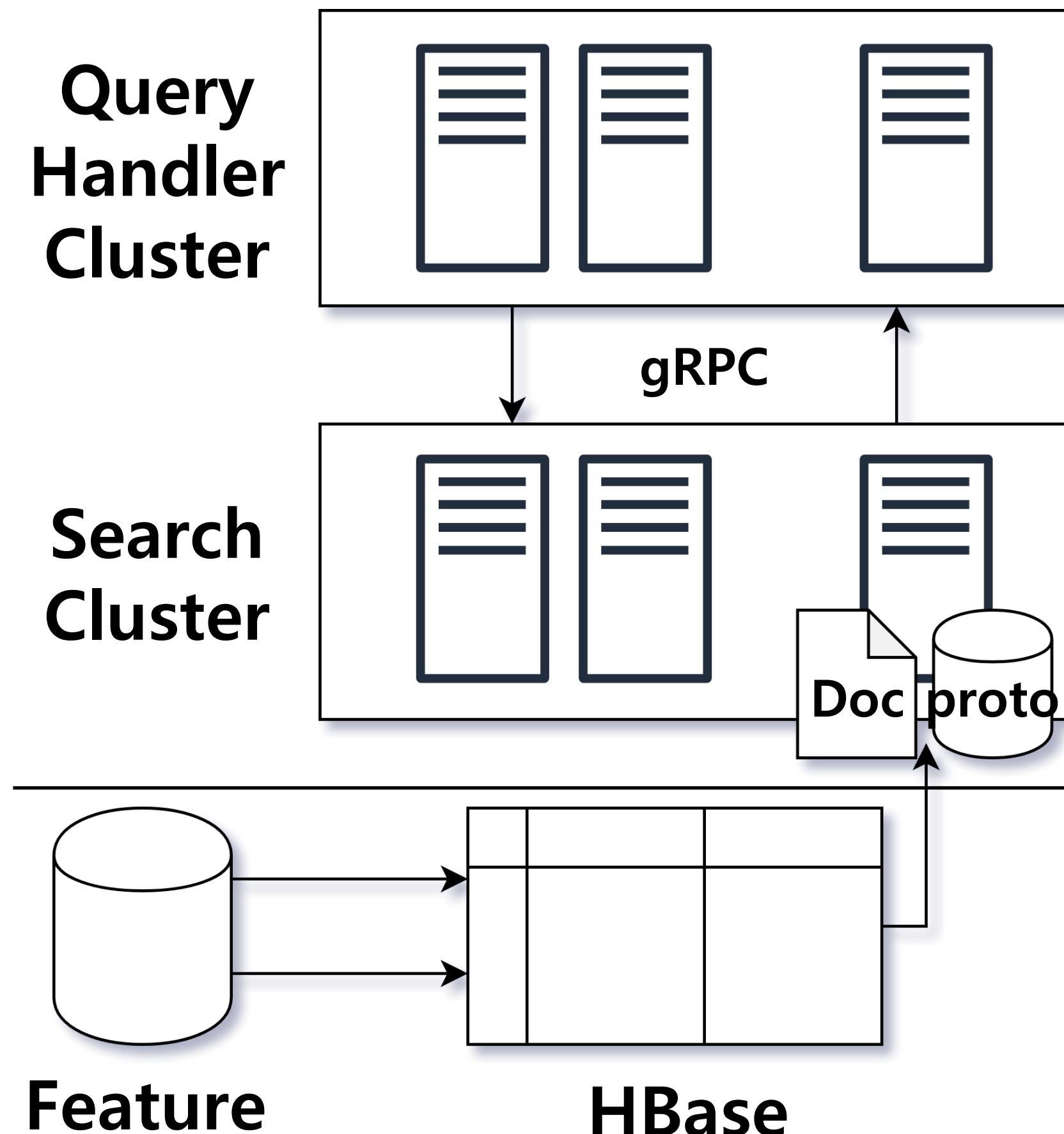


토프 스마트 인덕션
골라담기 기획전, 01

4,900원
 (2)
 맨위로



3.7 모델과 서비스의 분리



서비스 정의

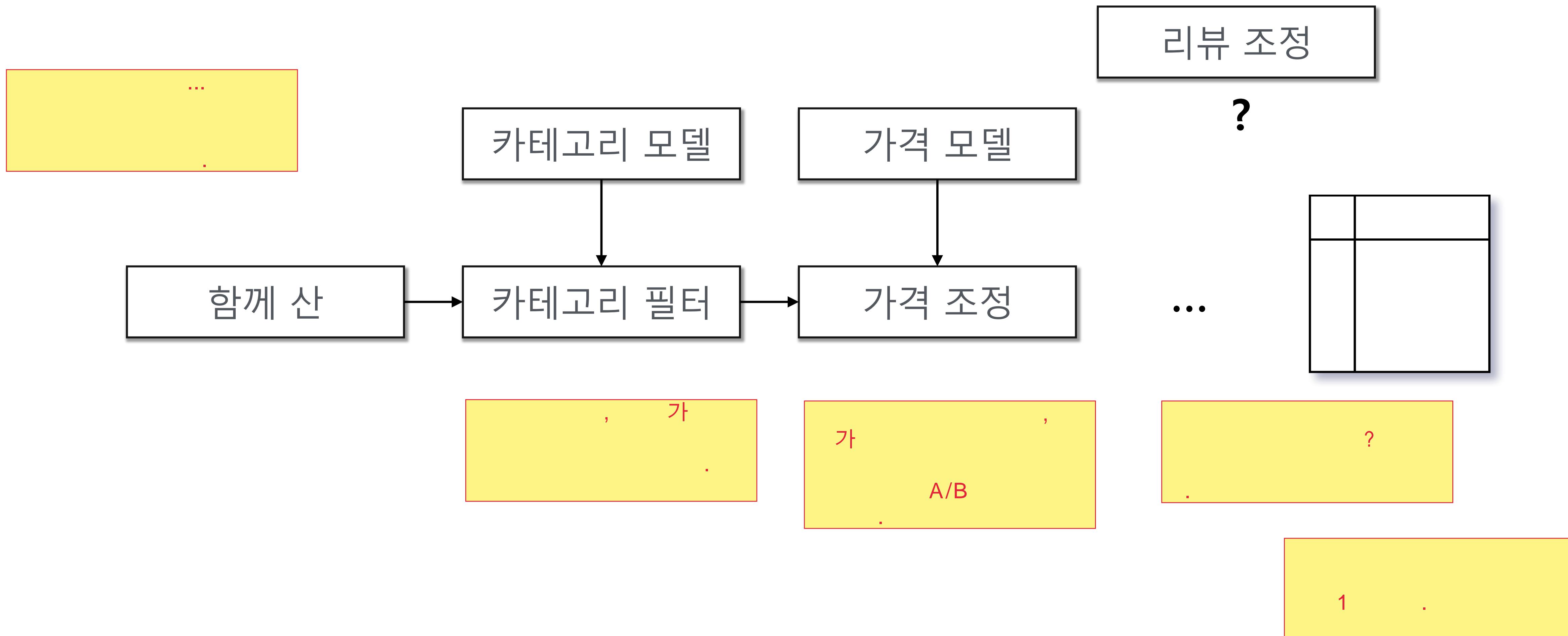
- 서비스는 있는 피처들을 어떻게 쓸지 고민

모델 개발, 피처 생성

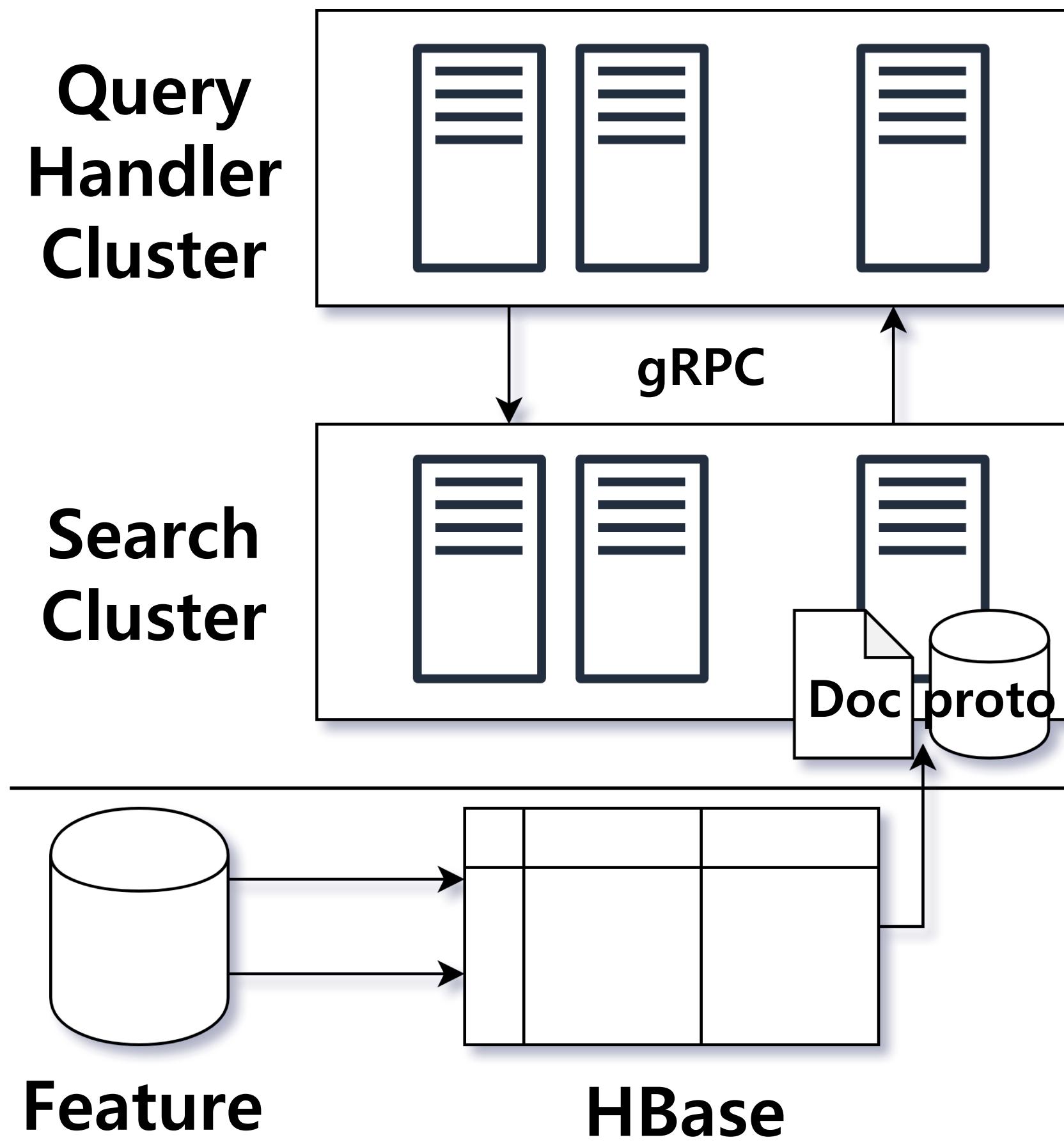
- 모델은 좋은 피처를 만들기 위해 노력

3.7.1 모델과 서비스의 분리 - 기존

함께 살 저렴한 다른 카테고리 추천



3.7.2 모델과 서비스의 분리 - 현재



함께 살 저렴한 다른 카테고리 추천

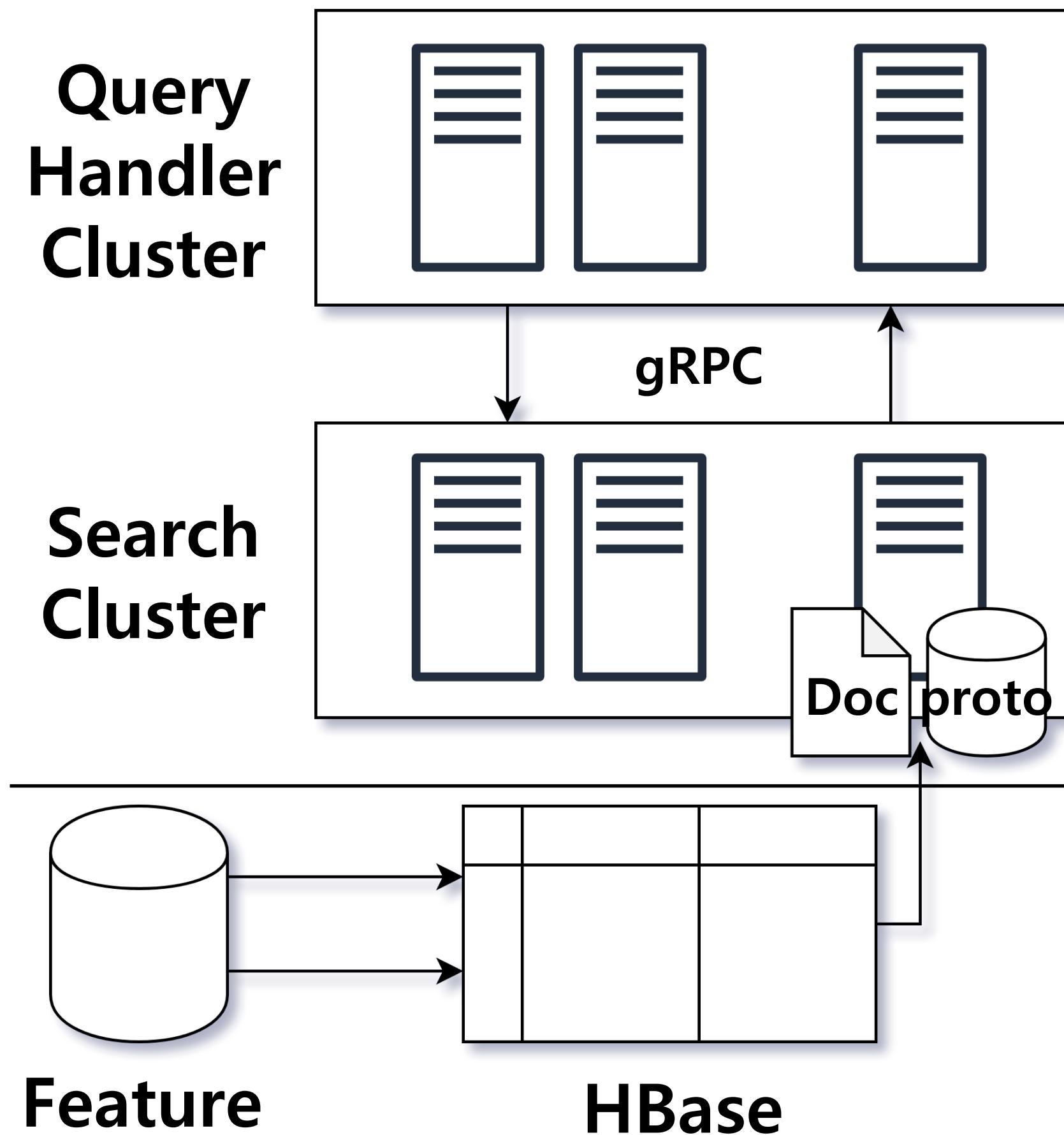
서비스 정의

- 함께 산 상품 데이터 사용
- 매뉴얼 카테고리 맵 (주방용품 → 식자재)
→ AB 테스트 시작

모델 개발, 피처 생성

- 카테고리 모델 개발
- 함께 구매 할 만한 가격대 분포 계산

3.7.2 모델과 서비스의 분리 - 현재



함께 살 저렴한 다른 카테고리 추천

서비스 정의

- 새 피처 사용
→ 새 AB 테스트

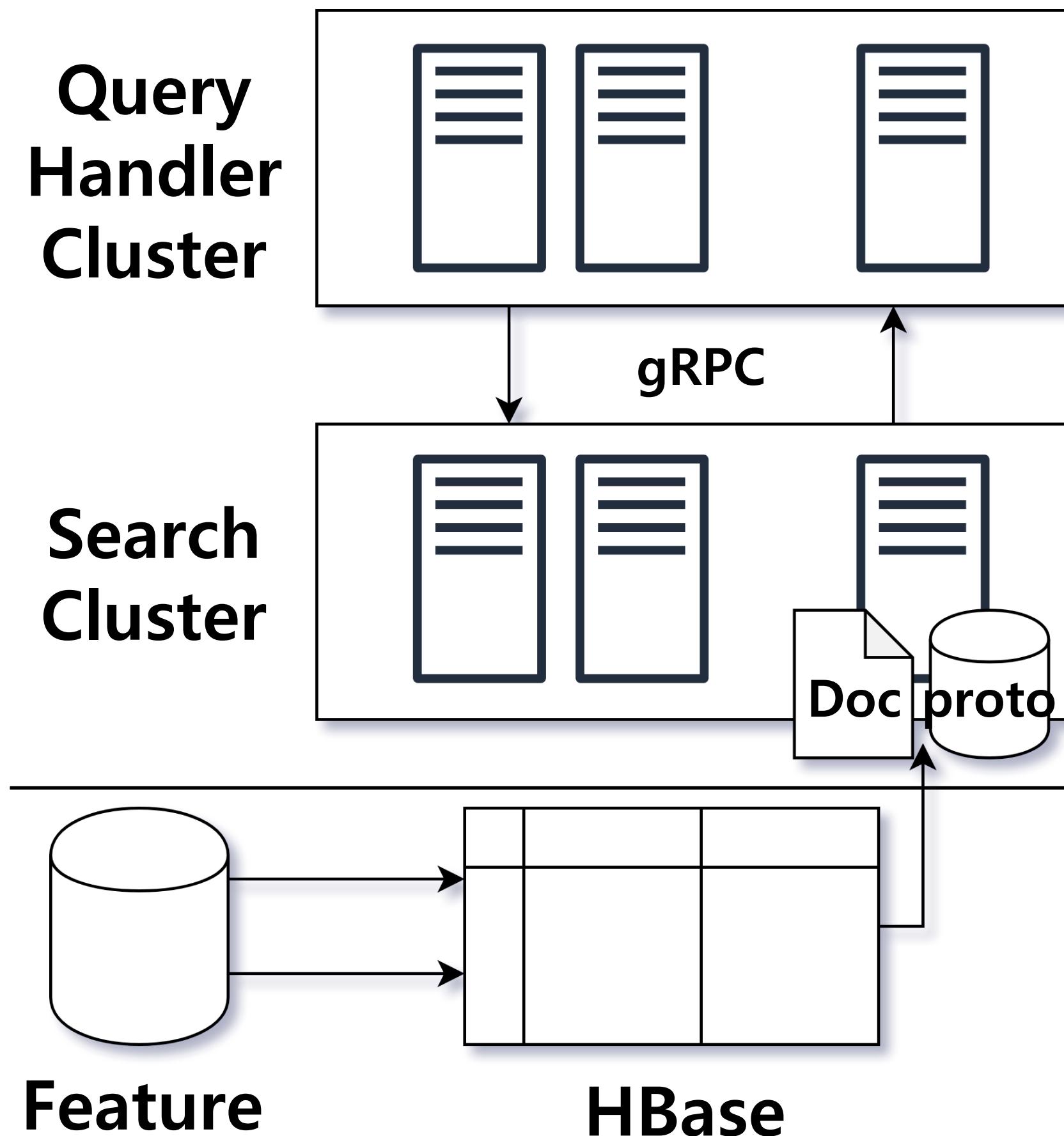
모델 개발, 피처 생성

이전 AB 테스트 반영 →

- 새 피처 개발

A/B
improve

3.7.2 모델과 서비스의 분리 - 현재



함께 살 저렴한 다른 카테고리 추천

서비스 정의

- 지속적 쿼리 튜닝

모델 개발, 피처 생성

- 만들어진 피처는 다른 서비스에도 사용

가
, 가

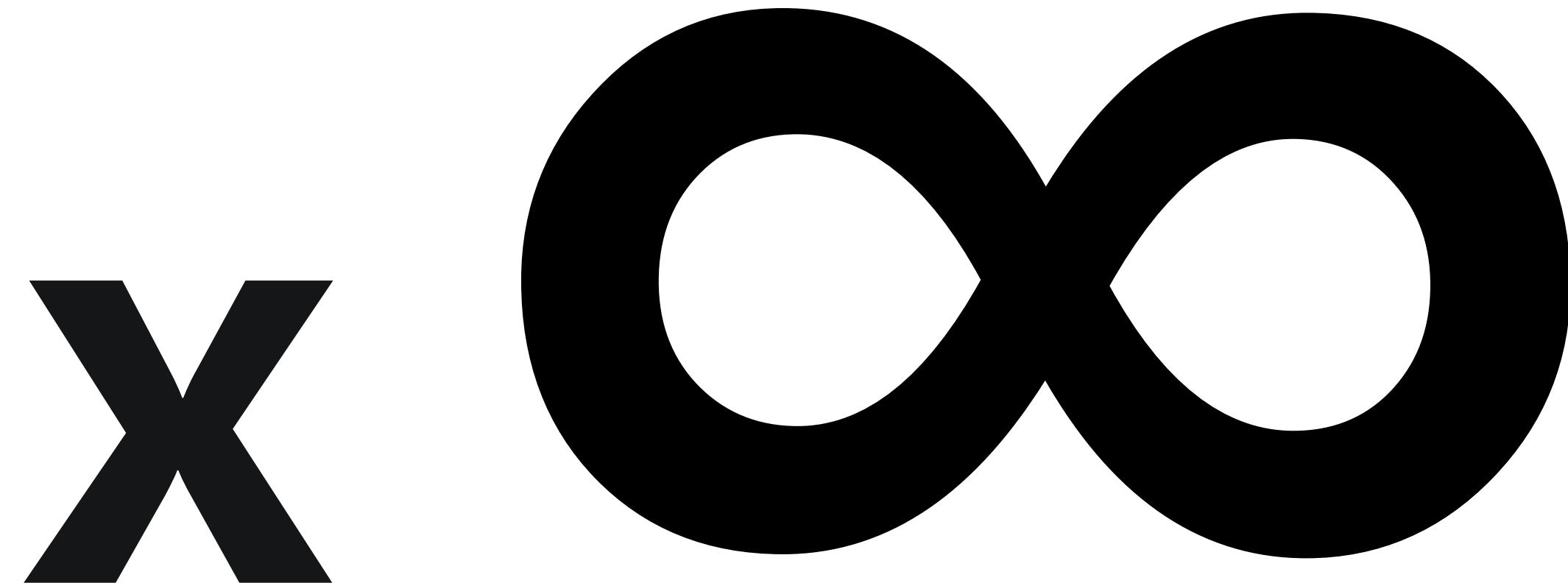
3.8 변화

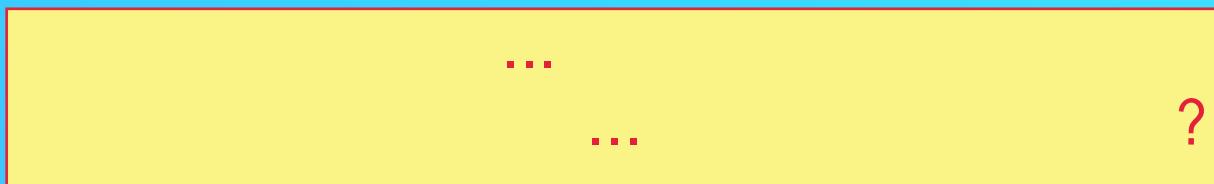
추천 영역의 수

X 7

3.8 변화

추천 노출 / 클릭 / 구매

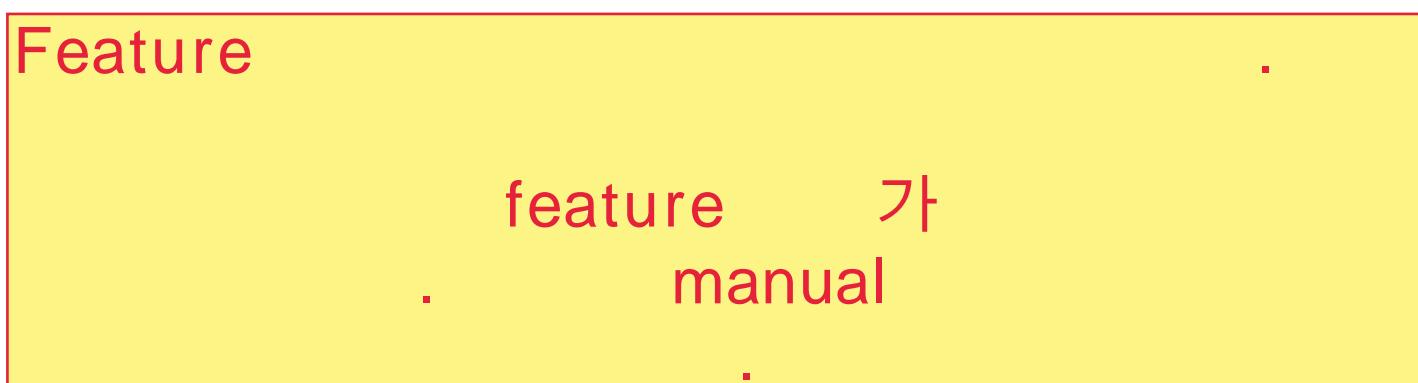




더 나은 추천을 위해

쿼리 튜닝

모델 점수 카테고리 점수 가격 상품평?
 $F(f_1, f_2, \dots, f_n)$ 할인?
피드백?



1. Learning to Rank

2. 실시간 개인화

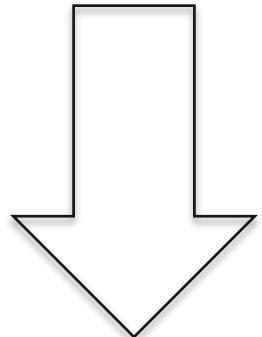
쿼리 투닝을 모델에게 맡기자

4. Learning to Rank

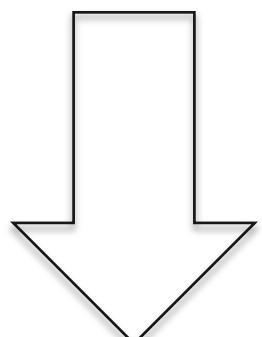
피처 벡터

상품 피처

상품 연관성 피처



모델



점수

4.1 Inference

- Search Cluster 각 노드는 모델을 들고 각자의 추천 후보 상품을 **리랭킹**
- 모델 결과가 뭔가 이상할 수도 있다 → 리뷰 / CTR 등으로 보정

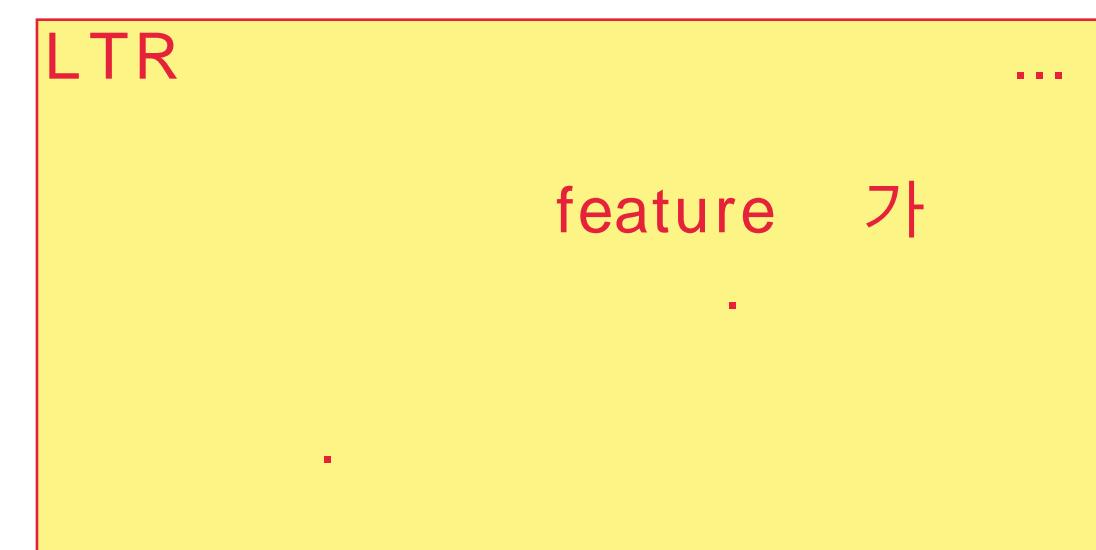


Search Cluster

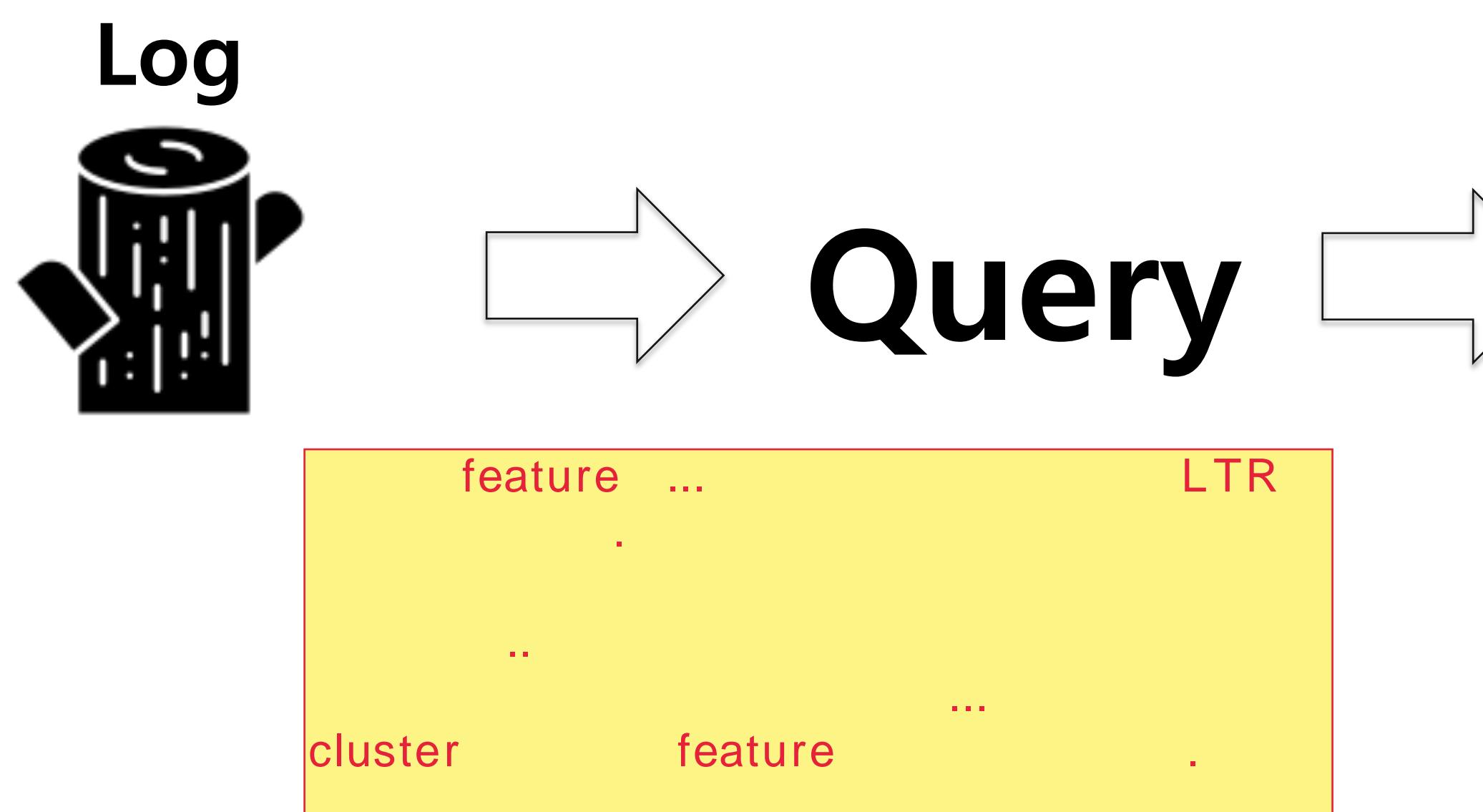
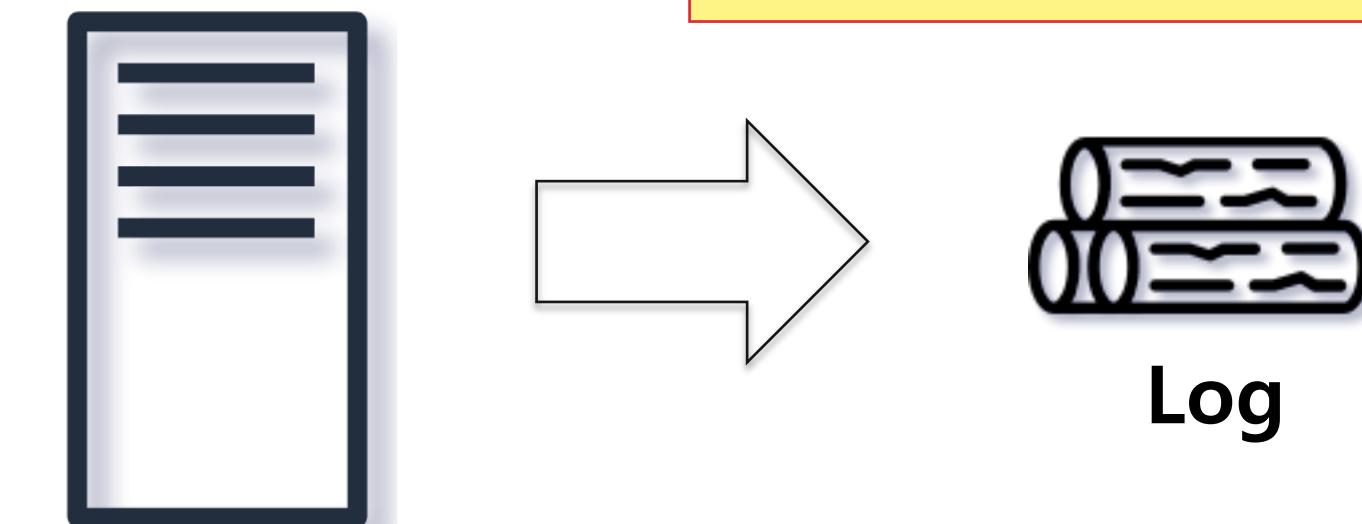


4.2 Training

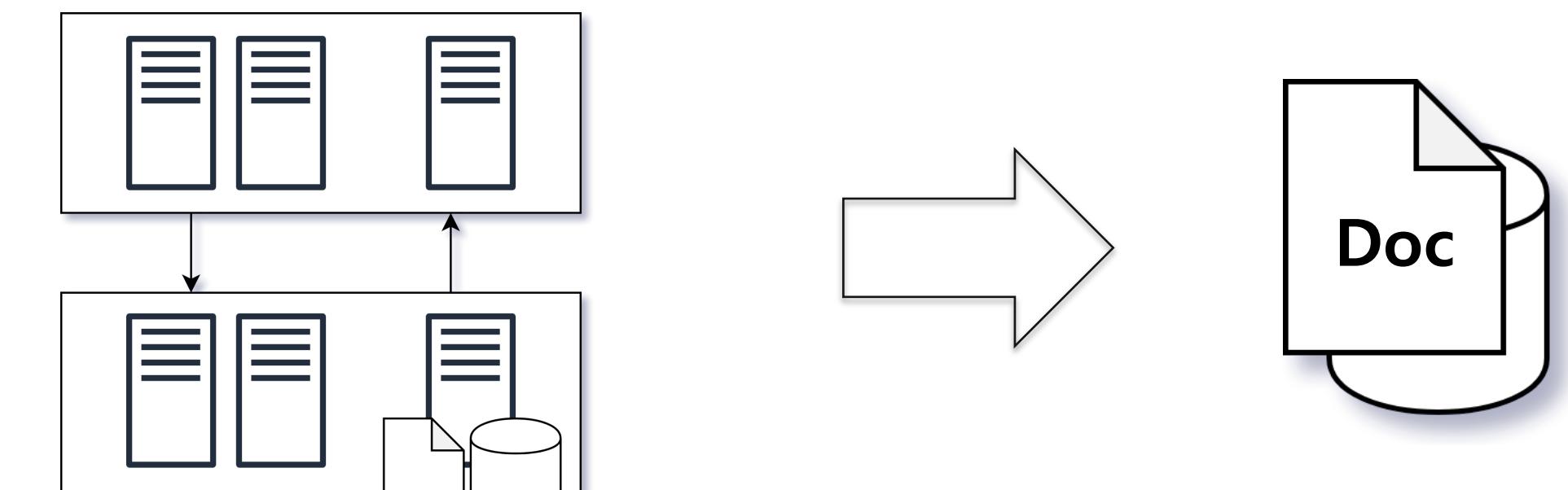
- 실시간 로깅
- 새로운 피쳐는 쿼리를 재현하여 Crawl
- 유저 피드백 Backfill



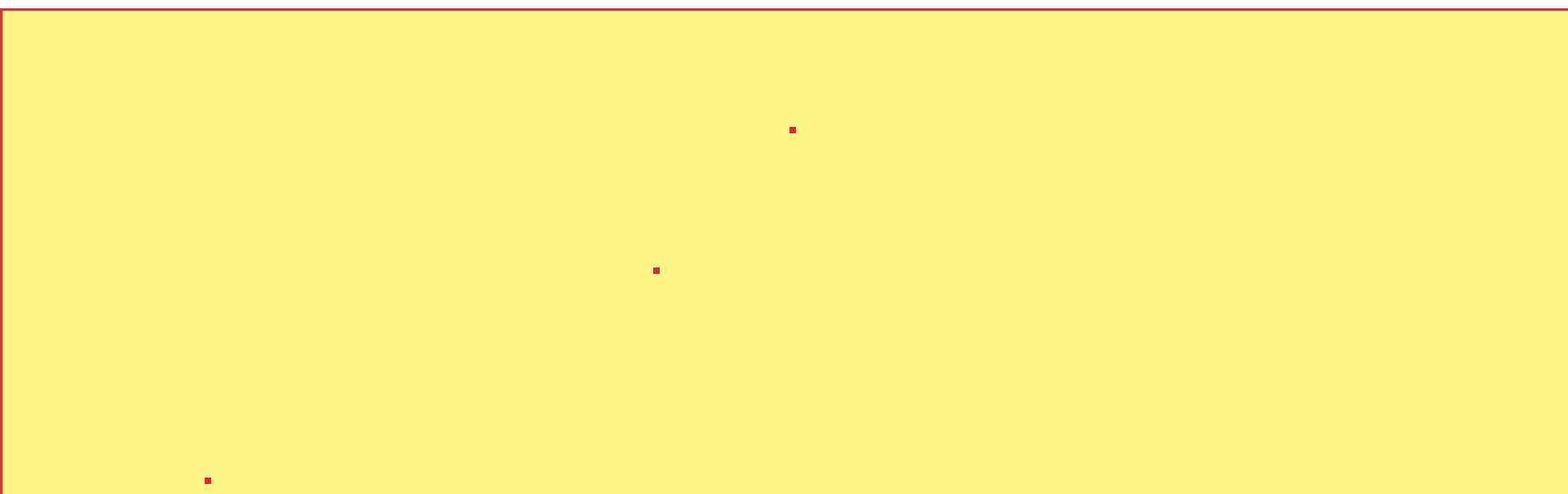
API Server



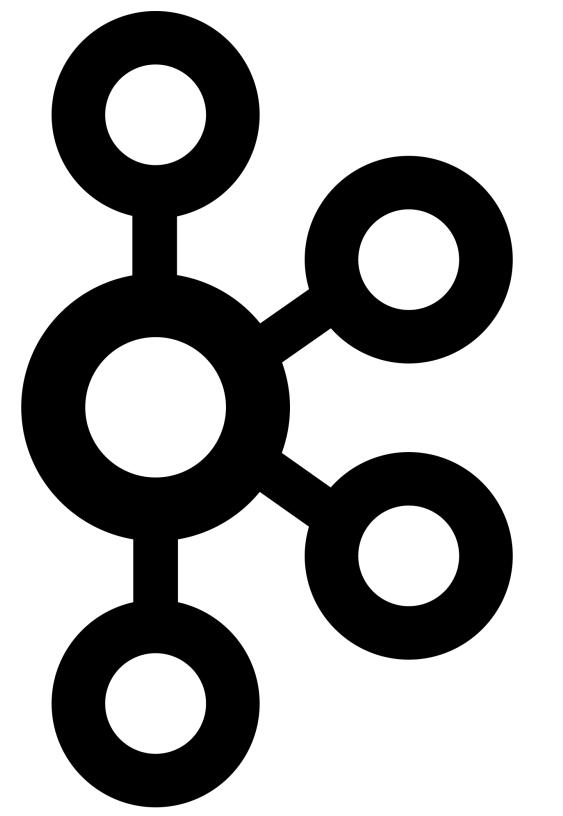
Crawling Cluster



모델을 위한 피처 엔지니어링



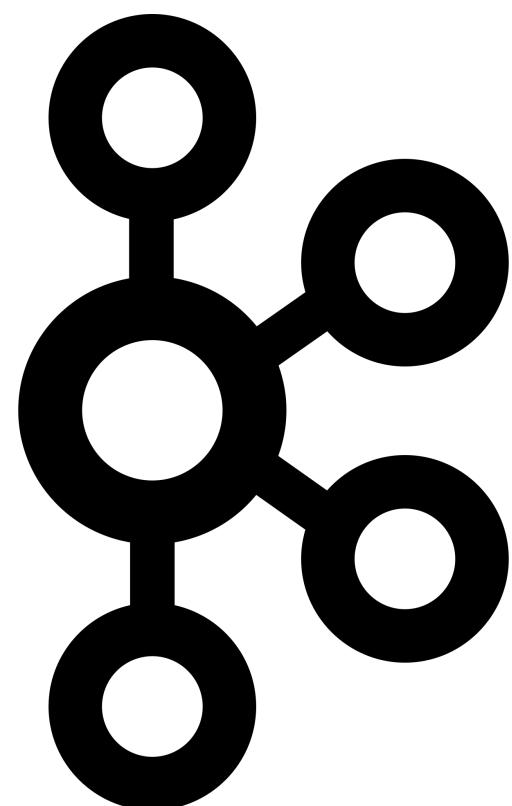
1. Learning to Rank
2. 실시간 개인화



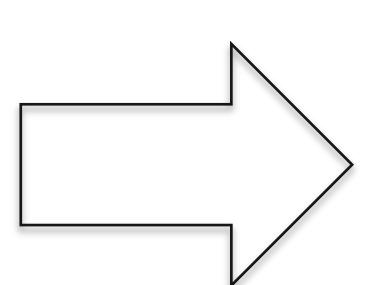
Kafka

5.1.1 실시간 사용자 피처

- 실시간 로그 → Redis



Kafka



암호화된 유저 ID
→ 사용자 정보

- 최근 본 상품
- 최근 산 상품
- 최근 검색한 쿼리

내가 본 상품의 연관상품



리빙온 세라믹 냄비 양수, 혼합 색상, 16cm

10,900원

10,360원

로켓와우

★★★★★ (136)



키친아트 세라믹 냄비 2종 세트, 혼합색상, 18편수 냄비

20,800원

18,610원

로켓와우

★★★★★ (2,295)

더보기 >



키친아트 88오페라 양수냄비, 24cm

18,400원

로켓와우

★★★★★ (753)

5.1.2 오프라인 사용자 피처

사용자 분석 모델

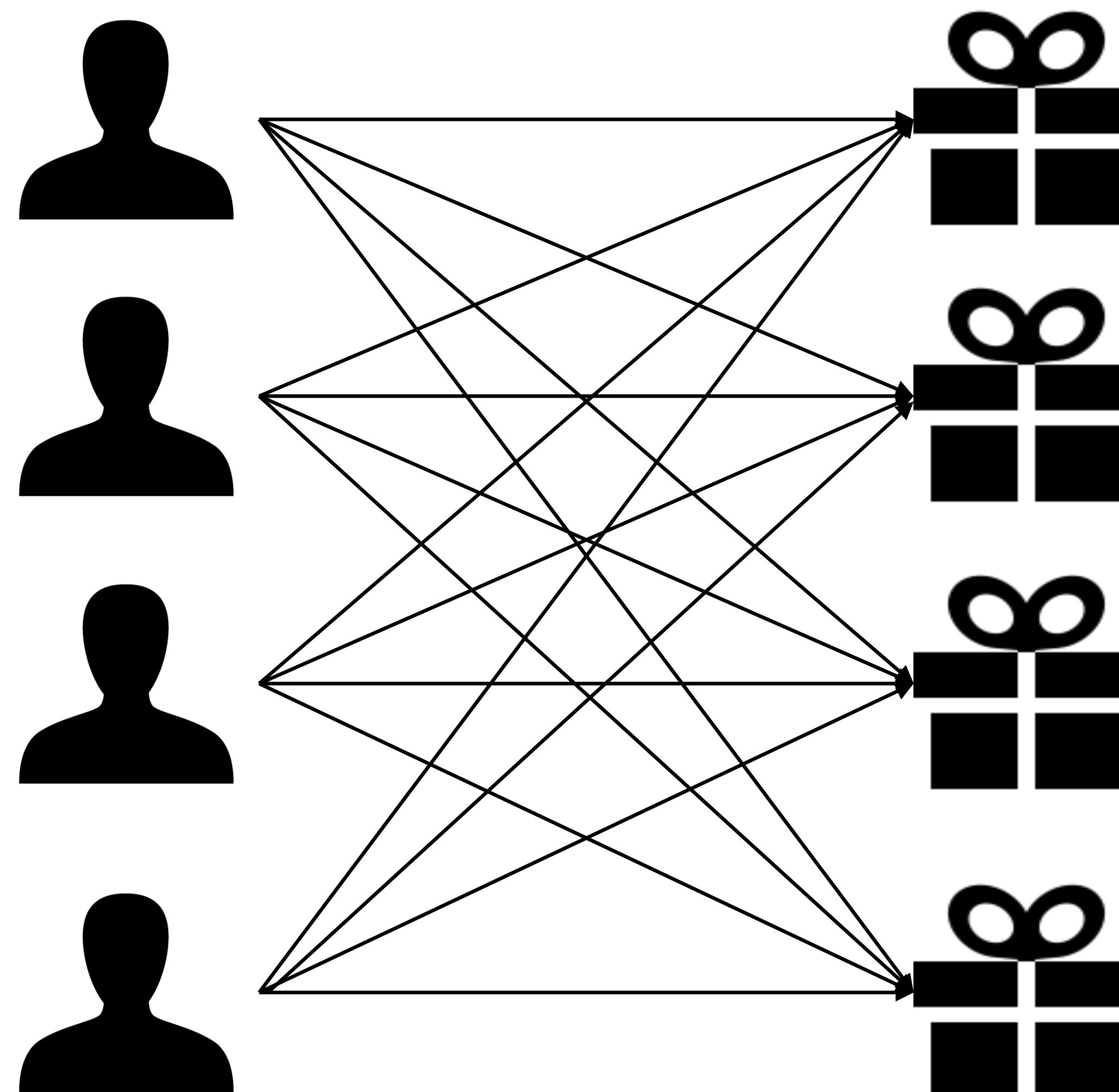
- 선호 카테고리
 - 선호 가격대
 - 선호 브랜드
 - Rocket Wow 멤버십 가입 여부
- ...

feature 가 .

5.2 유저 상품 연관성

유저 상품 관계

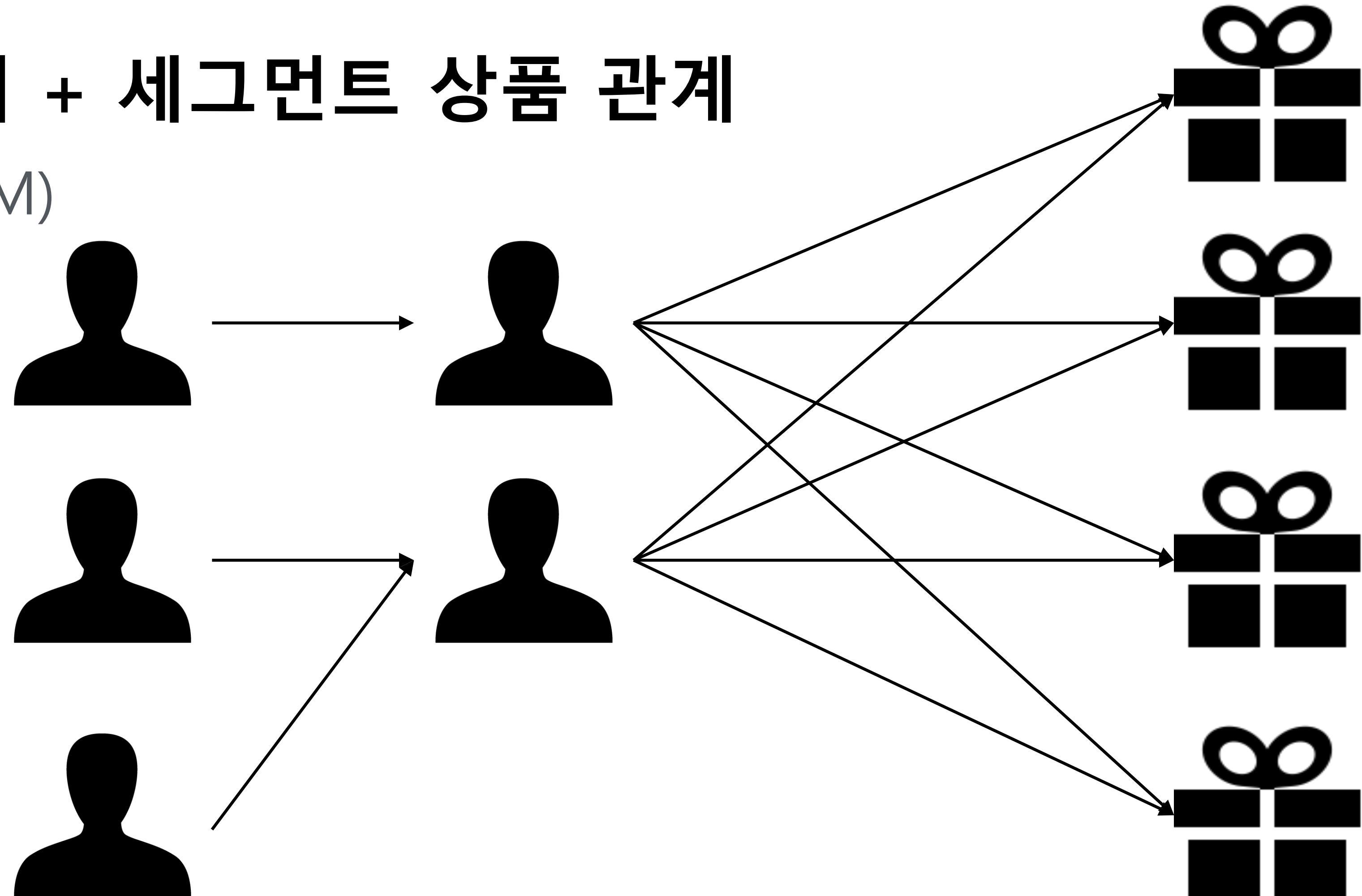
- $M \times N = MN$



5.2.1 세그먼트 상품 연관성

유저 세그먼트 관계 + 세그먼트 상품 관계

- $S \times N = SN$ ($S \ll M$)



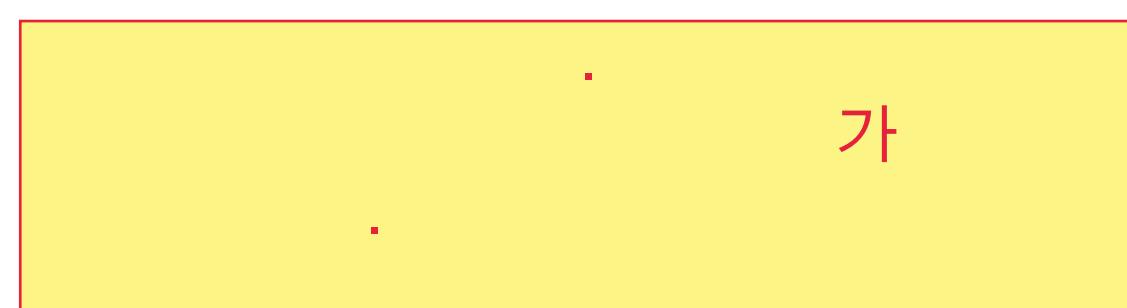
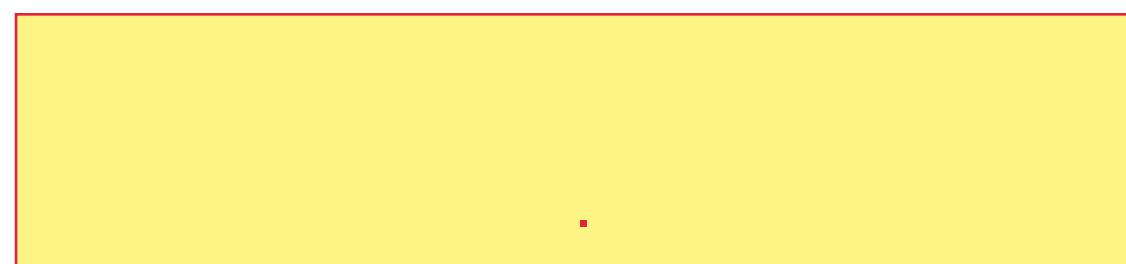
5.2.2 유저 세그먼트

<예>

RFM score ? ? ?	Recency	분위수
	Frequency	분위수
	Monetary	분위수
	Device	아이폰 안드로이드
	Parent	유아 용품 구매 이력
	Car Owner	차량 용품 구매 이력
	Gender	여성/남성 패션 등 → U/M/F

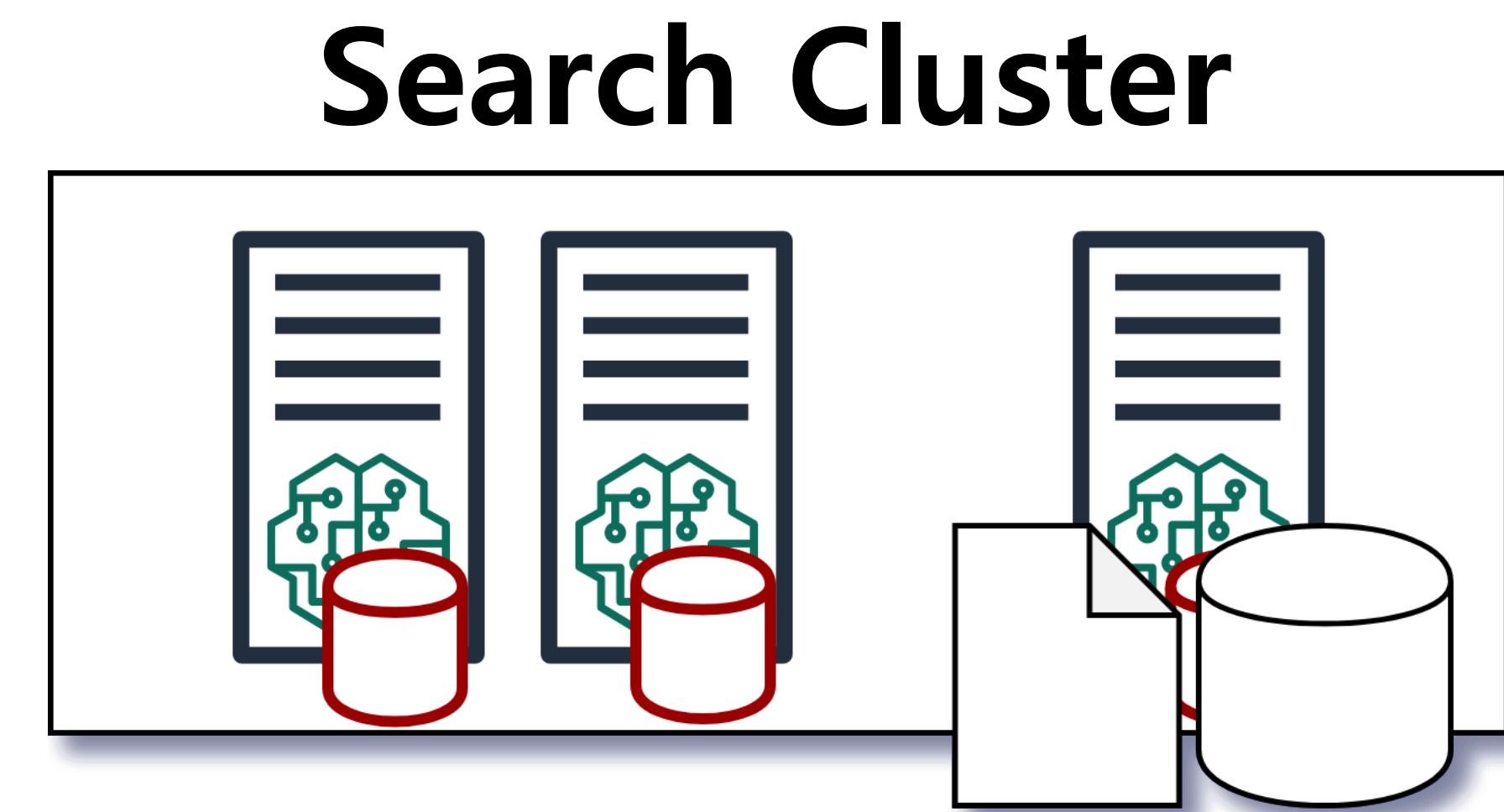
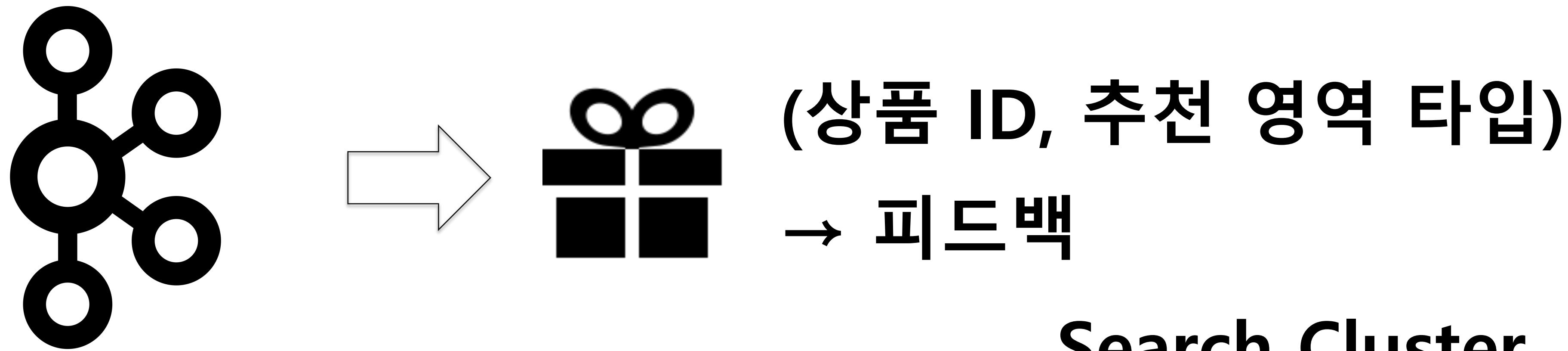
$$5 * 5 * 5 * 2 * 2 * 3 = 1,500$$

[1분위, 5분위,
 2분위, 4분위,
 3분위, 1분위,
 아이폰, 안드로이드,
 O, X,
 O, O,
 여 남]]



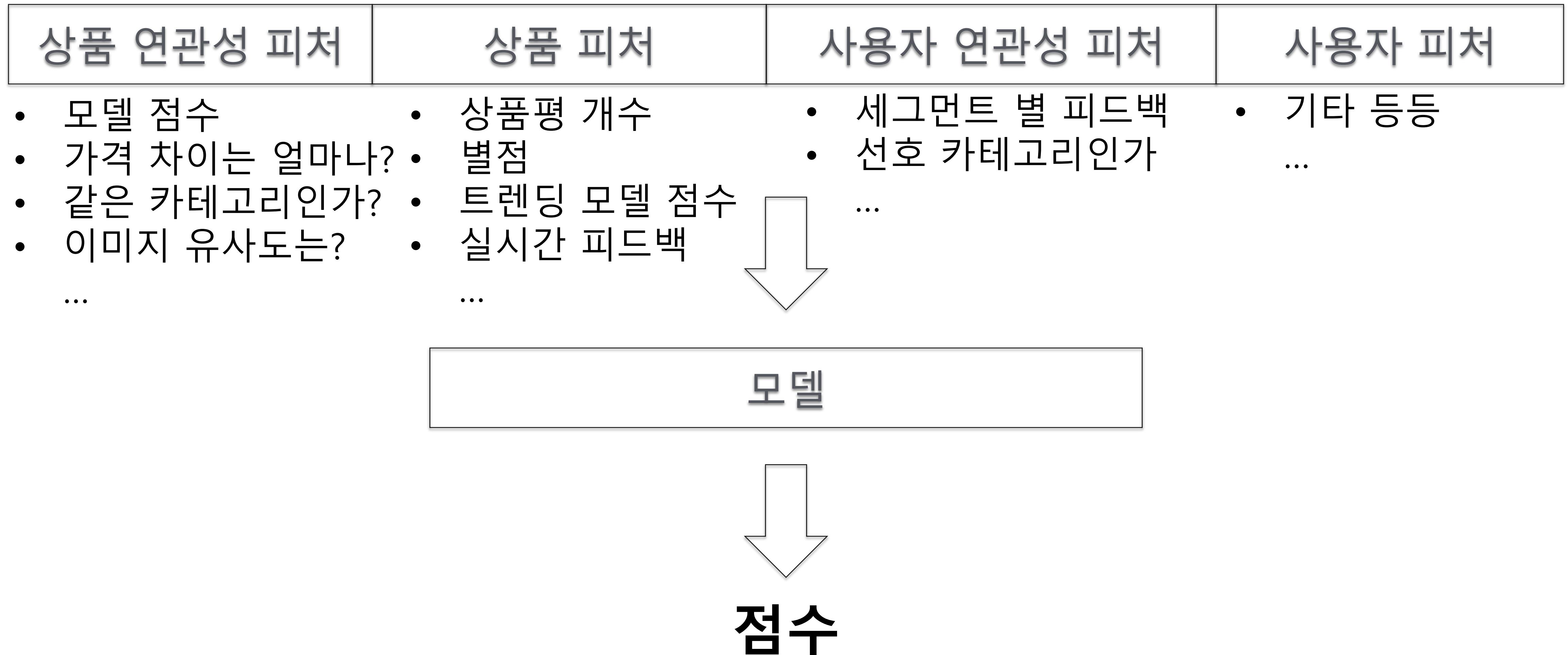
5.3 실시간 상품 피처

- 실시간 로그 → Redis → 각 Search Cluster 노드

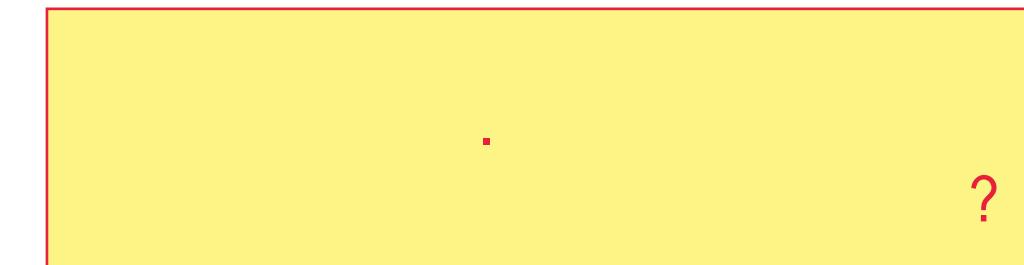


5.4 실시간 개인화

LTR

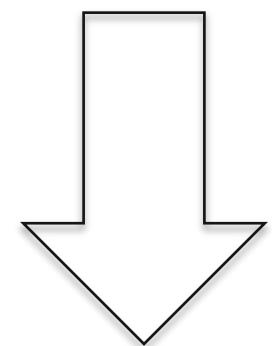
feature
feature
가

5.5 추천 영역 개인화

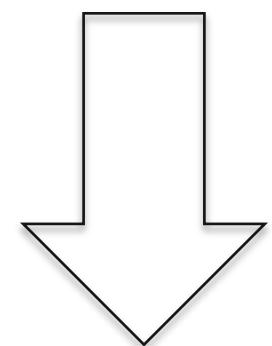


사용자 피처

추천 영역 선호도



모델



점수

5.6 Feature Selection

모든 피처 사용

- 성능이 더 좋을 수 있음

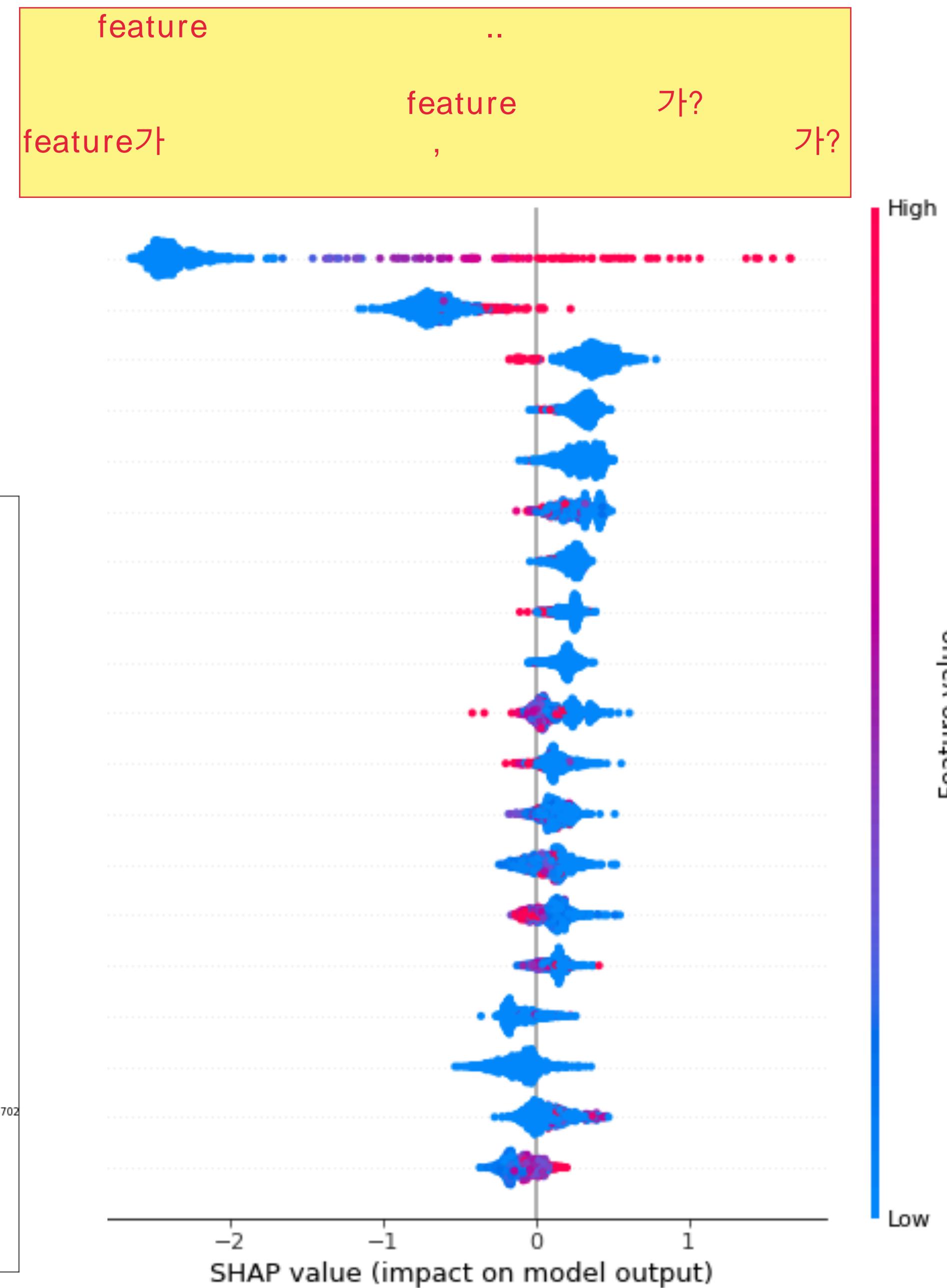
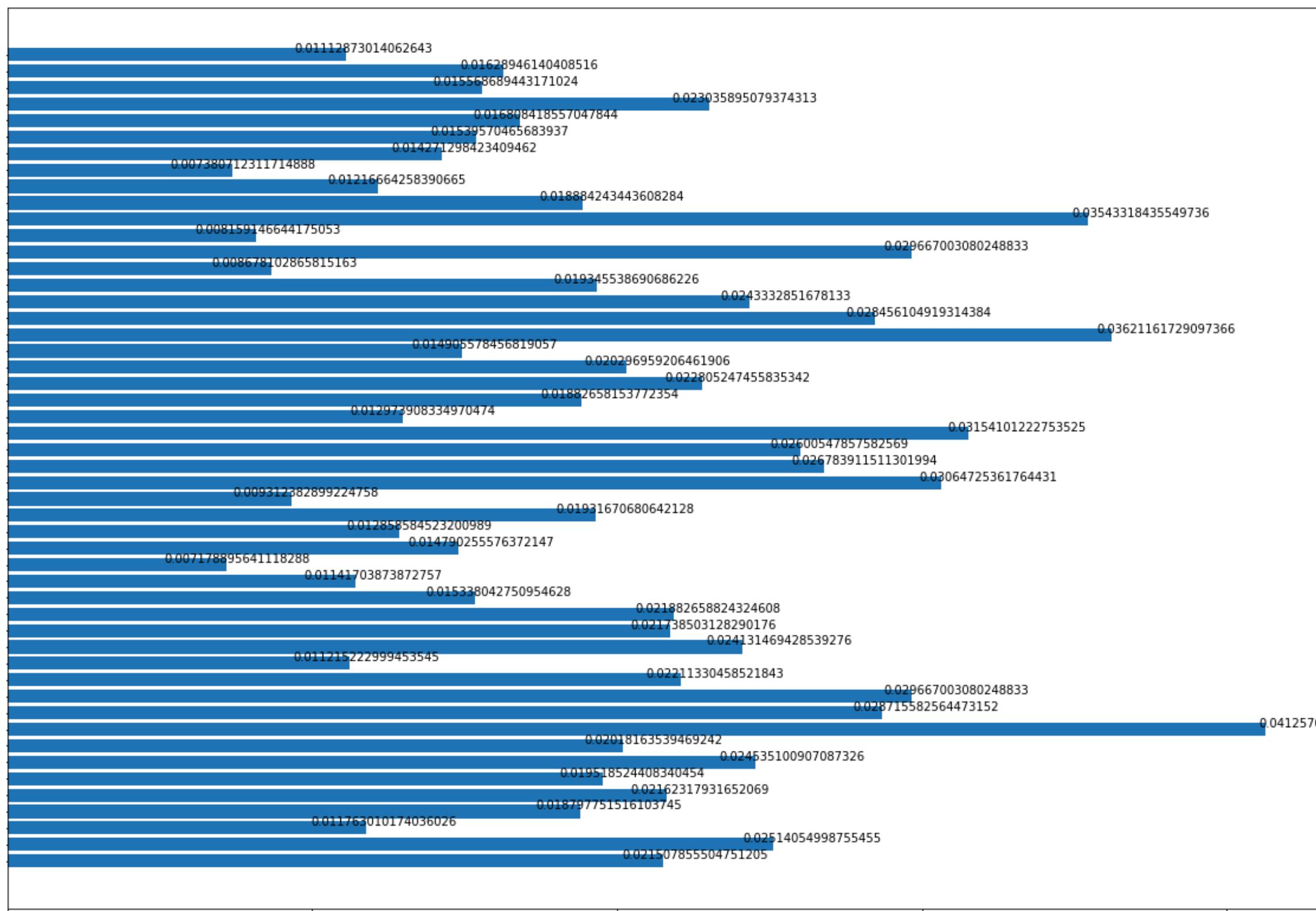
하지만 ...

- 새로운 피처가 어떻게 작동하는가
- 어떤 피처를 우선적으로 개선해야 하는가
- 모델을 어떻게 디버깅하고 개선 해야 하는가



5.6 Feature Selection

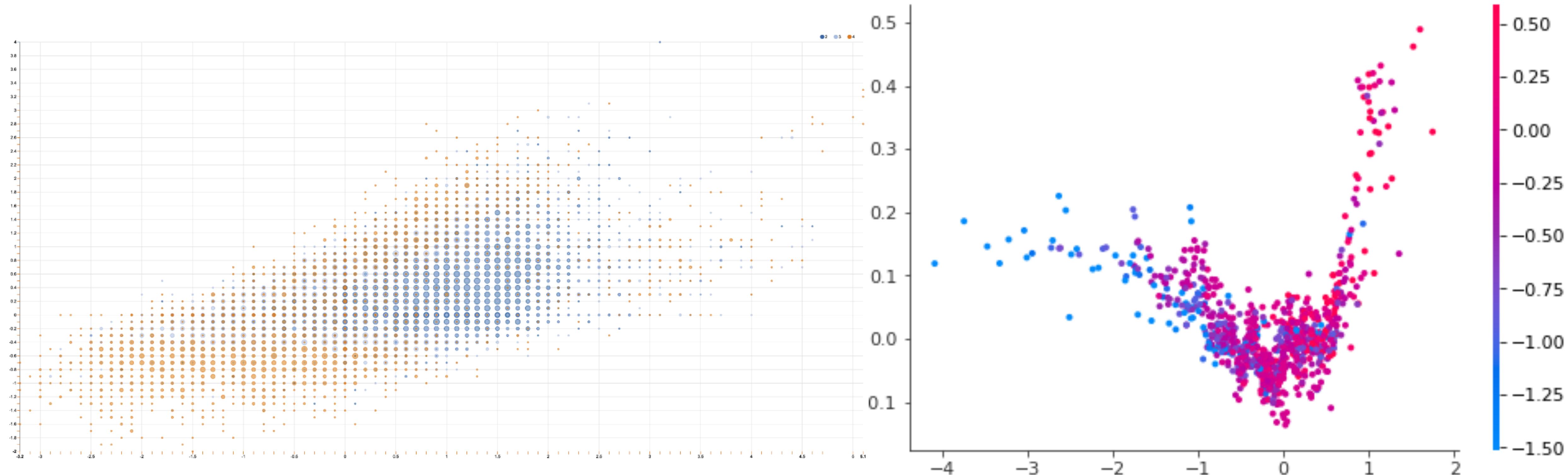
모델이 피처를 어떻게 사용하지?



5.6 Feature Selection

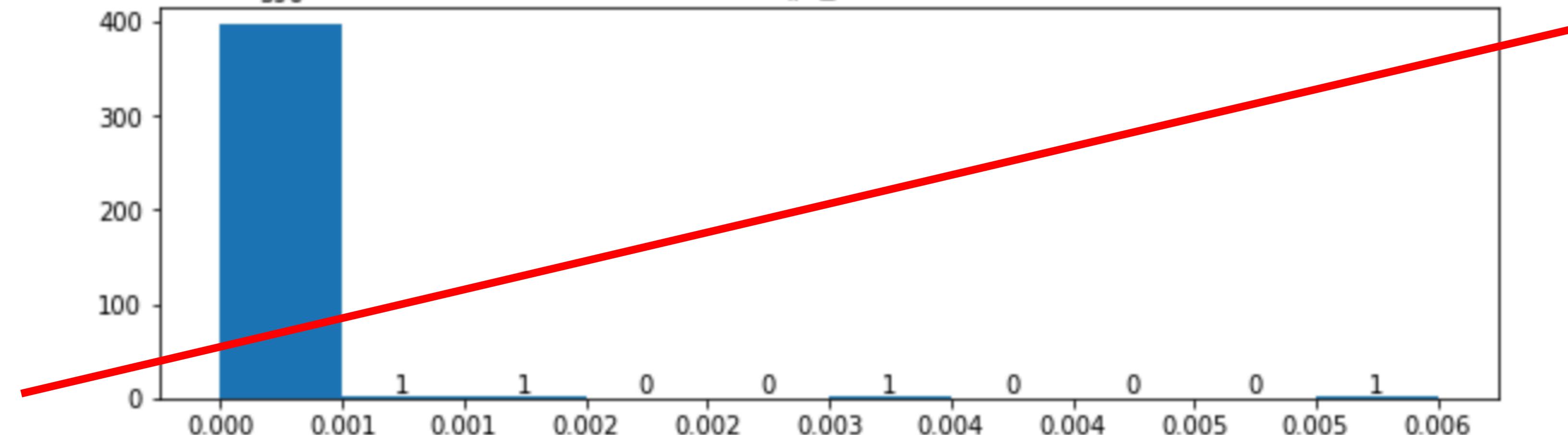
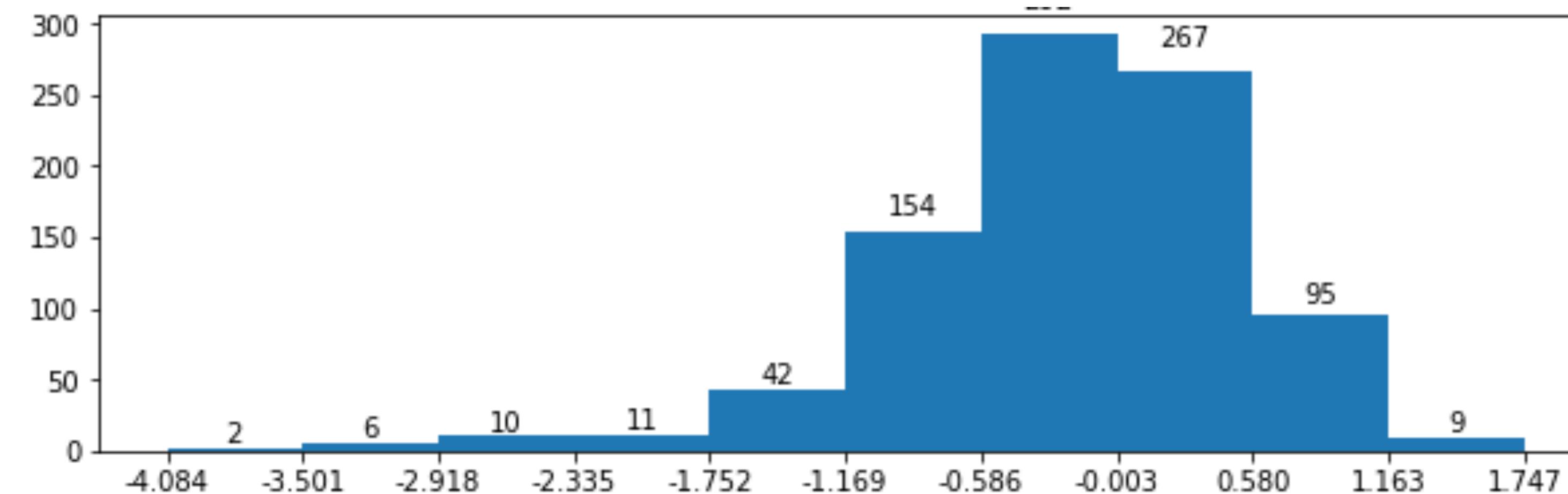
피처끼리는 관련이 없나?

feature	correlation
1	0.5



5.6 Feature Selection

피처는 잘 분포해 있나?



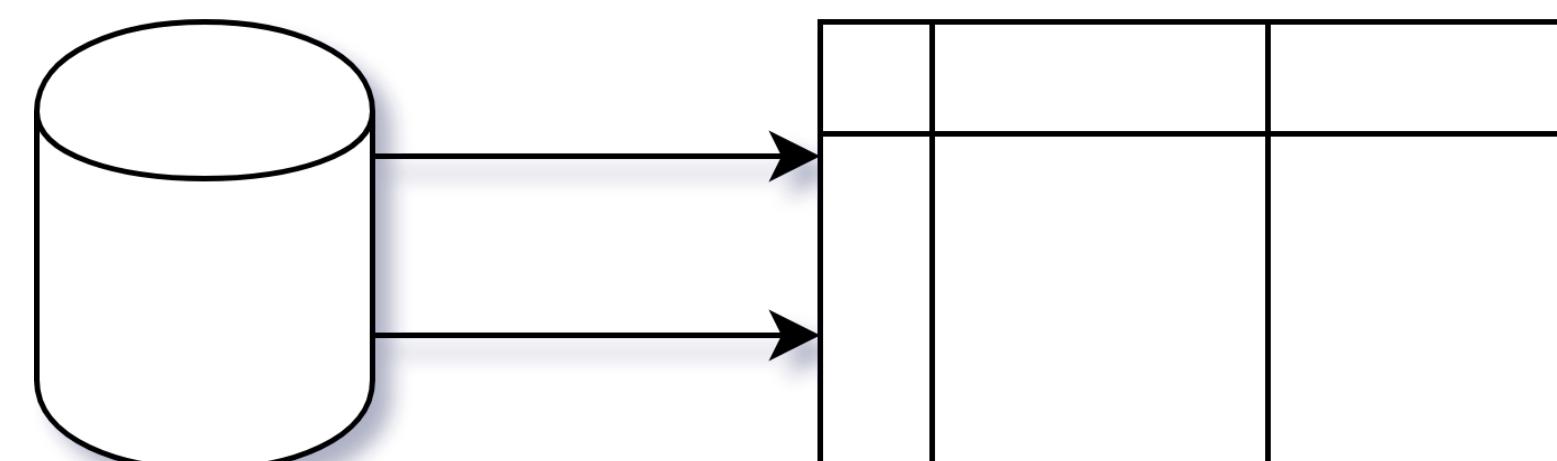
요약 & 팁

Query Handler Cluster

- 재정렬

Search Cluster

- LTR
- 실시간 상품 피쳐

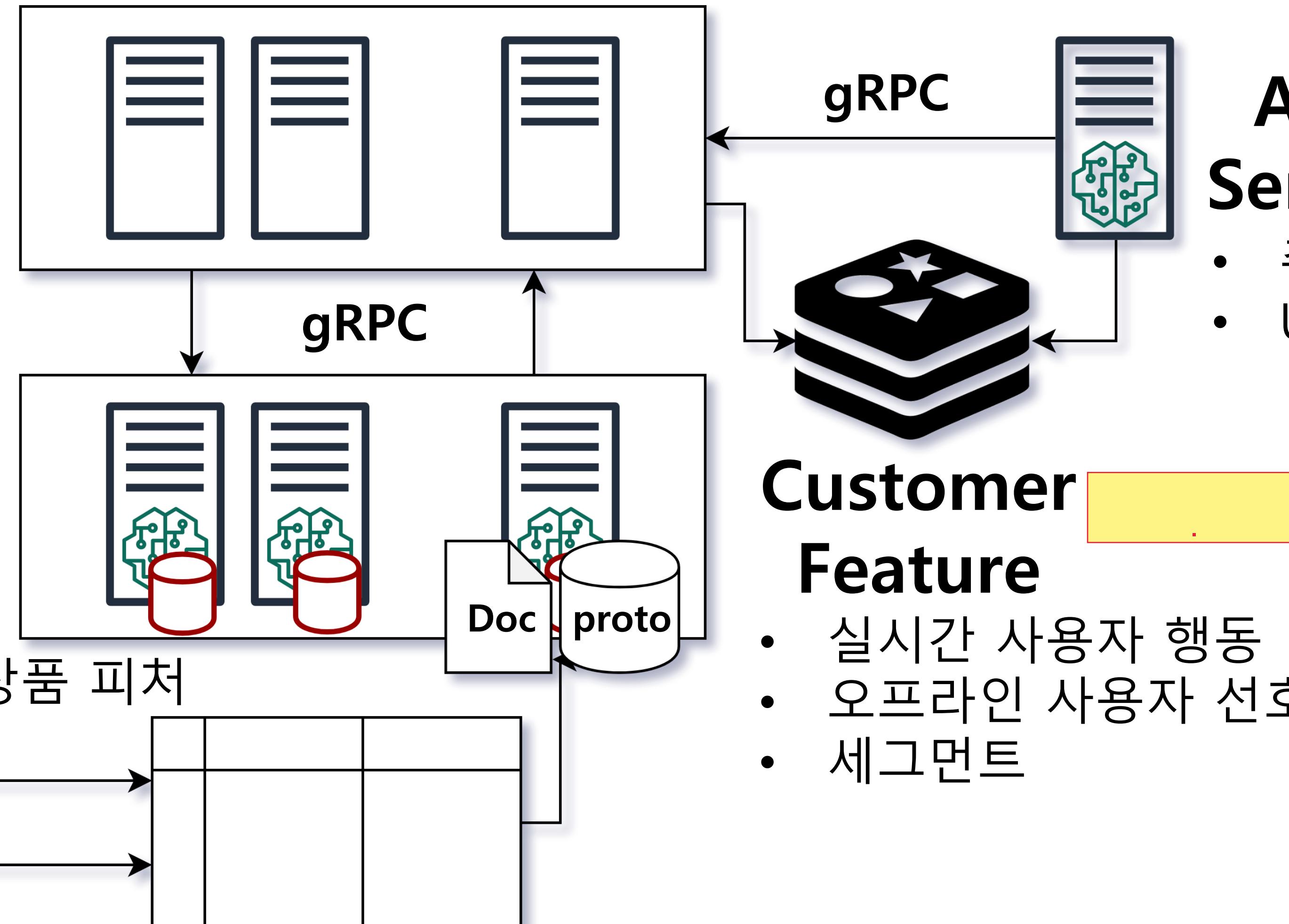


Feature

- Spark, ML, ...

HBase

- 통합 상품 정보



Customer Feature

- 실시간 사용자 행동
- 오프라인 사용자 선호
- 세그먼트

Raw Feature vs Processed Feature

- # review
- rating
- rating score

Raw feature	correlation	.
processed feature	.	.

복잡한 LTR 모델 vs 단순한 LTR 모델

- RNN
- CNN

- Decision tree
- MLP

feature

피처 엔지니어링 vs 파라미터 튜닝

step

가

?

HDFS

(

)

Q & A

Thank You