

학부때 검색 쉽게 만들어보기

대충 검색 찍먹해보는 이야기

github.com/drakejin

검색: IR(Information Retrieval)

집합적인 정보로부터 원하는 내용과 관련이 있는 부분을 얻어내는 행위
위키백과

검색이란 기능으로 할 수 있는것들이 많다.

단순히 검색을...

“검색어(질의어)로 게시글 찾기”가 아니라

게시글로 사람 찾기 .. 라면?

사람으로 게시글 찾기 라면?

단순히 검색을...

“검색어(질의어)로 게시글 찾기”가 아니라 반대로

게시글로 사람 찾기.. 라면..?

사람으로 게시글 찾기.. 라면?





교과과정 | 컴퓨터공학과 | IT융합대학 X

gachon.ac.kr/major/it/02/curriculum.jsp

검색 0/0

X

가천대학교 Gachon University

IT융합대학 College of IT Convergence

꿈과 희망이 GACH

컴퓨터공학과

학과소개

교과과정

교수진소개

공지사항

우리과 이야기

자료실

학과이력소개

학과 동문회

2019 학과 총보동영상 바로보기

전공기반 경력개발로드맵 바로보기

자랑스런가천동문(GPA) 바로가기

검색

교과과정

✓ 해당 학년의 교과과정만을 참조하시기 바랍니다.
본인의 상세한 교과과정은 학사안내->학사정보->학사안내의 요람에서 확인하십시오.

✓ 교과과정표 상의 학년표기는 1학년은 당해 학년도이고, 2,3,4학년은 각 1년전의 학년도입니다.
(예> 1학년 : 2016학년도, 2학년: 2015학년도, 3학년: 2014학년도, 4학년: 2013학년도)

> 1학기 교과과정

학년	이수구분	교과목명
	전필	컴퓨터공학개론
	교필	College English 1

어려워 마시라~
저는 여러분의 편!
쉽게 구현해보자



“검색은 사용자가 가장 원하는 걸
찾아주는 기능이다”

“학점” 찾아줘

“월급” 찾아줘

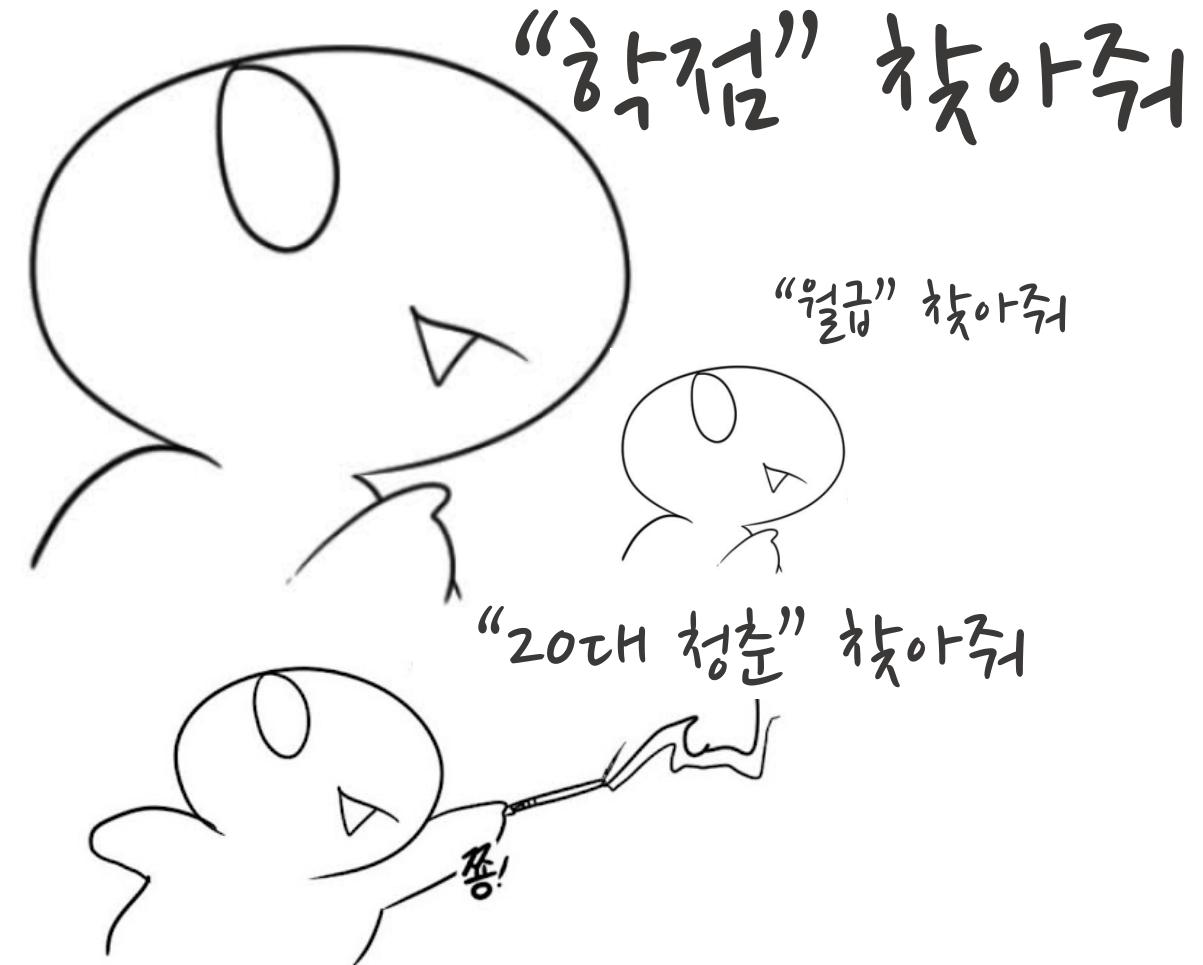
“20대 청춘” 찾아줘

쫑!

검색결과 “하고”



검색결과 “(결과없음)”



“20대 청춘” 찾았습니다

“하고” 찾았습니다

“월급” 찾았습니다

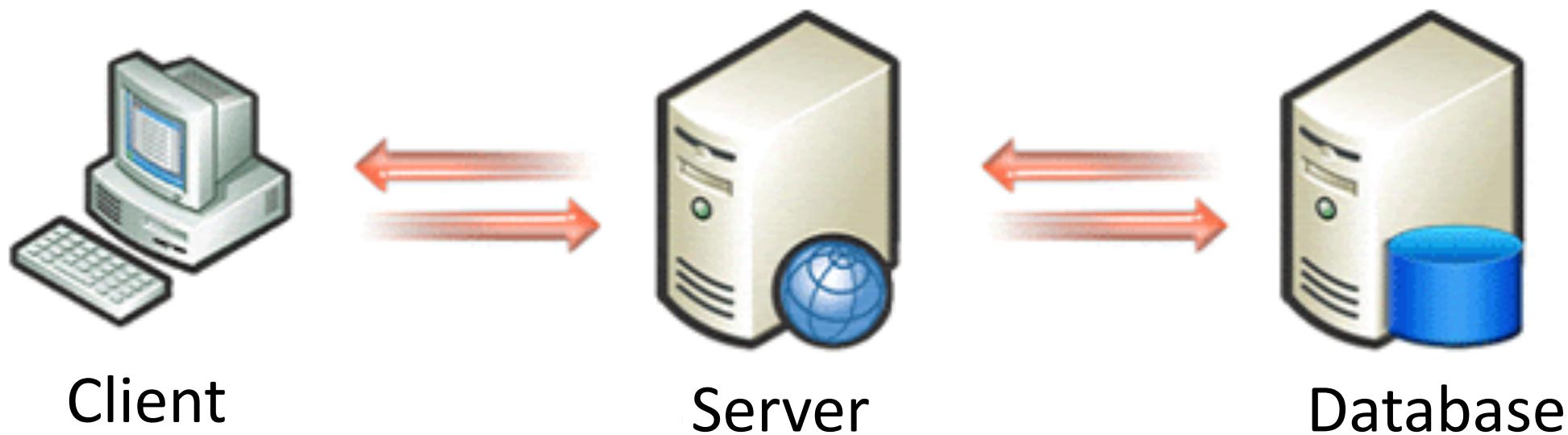
“종!”

검색결과 “흑역사”



키워드를 입력하면
키워드에 해당하는 문서들을 가져와서
어떤 기준에 의해 정렬한 결과를 만든다.

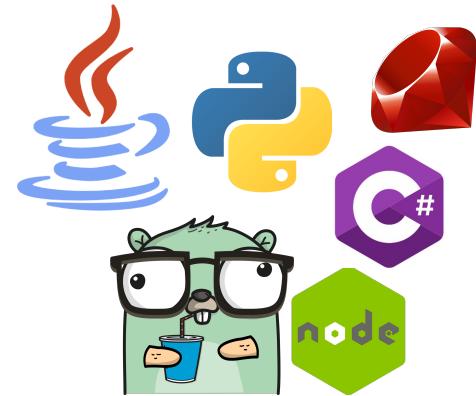
이게 여러분들의 서비스라고 가정해보자.



아마.. 기술은 이렇게 되어있겠지?



Mobile



Server



DATABASE



DATABASE

DBMS: DataBase Management System:

데이터를 관리해주는 시스템



DATABASE

DBMS에는 데이터를 넣고 꺼내는데
엄청나게 많은 유틸기능을 제공한다.

단순하게 엑셀마냥 값을 넣고 빼는것만 있는게 아니다.

우린 이 녀석을 최대한 활용해볼것이다

“검색은 사용자가 가장 원하는걸
찾아주는 기능이다”

```
SELECT * FROM posts P
  WHERE
    (P.title LIKE "%아이폰%") AND
    (P.content LIKE "%아이폰%") AND
    (P.hashtags LIKE "%아이폰%") AND
    (P.recommands LIKE "%아이폰%")
  ORDER BY
    P.created_at DESC
  LIMIT 20
```



인간주장

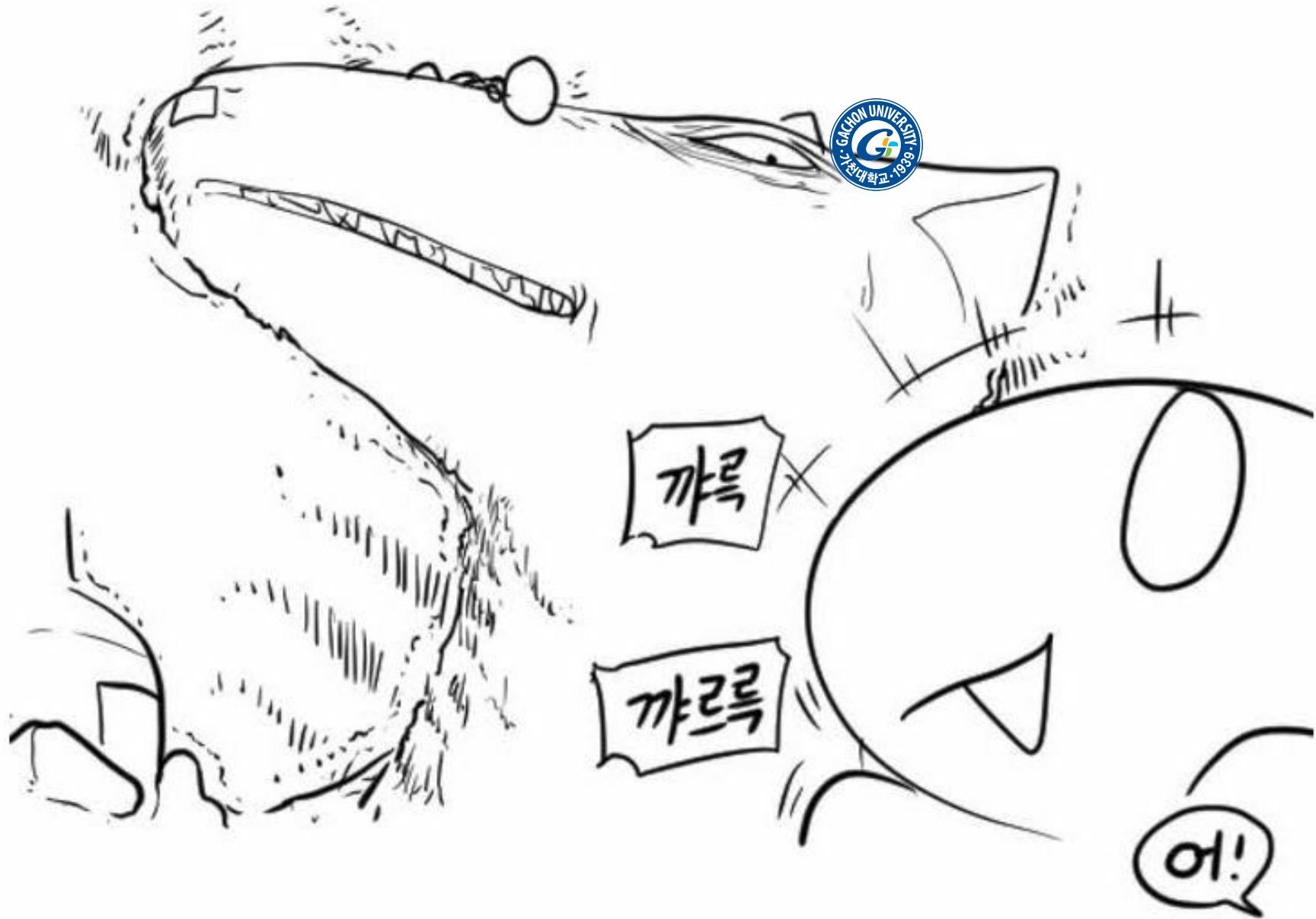
꾸?

그럼 안녕!

Q & A

LIFE
OF
PURPOSE

당근이세영?
영갑니당!



우리 아이 폰 뭐쓰는게좋을까요?

10분 전 2021-09-24 12:32:44

11살 자녀를 둔 아이 아빠입니다. 우리 아이는 모바일 배틀그라운드를 좋아해요.
트위치에서 마인크레프트? 라는 게임방송 보면서 시간을 보내곤 합니다. 괜찮은 핸드폰 추천...

인테리어가 좋은 카페에서 친구들과 아아 한잔

1시간 전 2021-09-24 11:32:44

여기 신논현역 앞에 교보타워빌딩에 있는 스타벅스 인테리어가 정말 좋아요. 여기에서 친구들하고 커피한잔 하려고 들어왔는데, 자리가 딱 있어서 제 아이폰을 올려두고, 커피를 주문하러 갔어요

특가 오지게 한다! 아이폰케이스 가져가라!

2시간 전전 2021-09-24 12:32:44

아이폰케이스 싸게 구매했다. 케이스 가져가라! 야 참 싸다 싸다! 다 사가라 다! 사! 가!

```
SELECT * FROM posts P
WHERE
    (P.title LIKE "%아이폰%") AND
    (P.content LIKE "%아이폰%") AND
    (P.hashtags LIKE "%아이폰%") AND
        (날먹 오지네..)
    (P.recommands LIKE "%아이폰%")

ORDER BY
    P.created_at DESC
LIMIT 20
```

Where의 Like 구문은 0과 1의 느낌으로
일치 하냐 일치 하지 않느냐만 바라본다.

그러므로 연관의 정도를 알 수 없다.

우린 겁나게 많은 기능중에
fulltext search라는 기능을 써야한다

“엥 RDBMS로 like 말고 키워드를 찾는 기능이 있어요?

MySQL DB라면?

- FULLTEXT Index

PostgreSQL DB라면?

- GIN(General Inverted Indexer)
 - tsvector, tsquery

MySQL DB라면? – FULLTEXT Index

이 두개는 무료라서 제가 써봐서 알아요 나머지 DB는 유료인데
미래에 취업하면 사수님에게 물어봐서 알아보도록 합시다 :>

– GIN(General Inverted Indexer)
– tsvector, tsquery

데이터베이스에는 연관도를 연산할 수 있는
FullTextSearch가 있다.

데이터베이스에는 연관도를 연산할 수 있는
FullTextSearch가 있다.

보통 RDBMS로 like연산으로 만들어 쓰다가,
엔진 성능 및 결과 퀄리티가 구리면 사용하는 방식

데이터베이스에는 연관도를 연산할 수 있는
FullTextSearch가 있다.

보통 RDBMS로 like연산으로 만들어 쓰다가,
엔진 성능 및 결과 퀄리티가 구리면 사용하는 방식

RDBMS를 사용하는 서비스라면
아주 쉽게 검색기능을 만들어 쓸 수 있다.

검색 랭킹스코어 =
(연관점수 * 연관가중치) + (작성한날짜 * 날짜가중치)

이러면 연관도도 보고, 최신순도 함께 볼 수 있겠지?

```
SELECT
    *,
    ((MATCH(P.title, P.content, P.hashtags, P.recommands) AGAINST("아이폰") * 2(텍스트가중치))
     +
     (unix_timestamp(P.created_at) * 4(시간가중치))) as ranking_score
FROM
    posts P
ORDER BY
    ranking_score DESC
LIMIT 20
```



인간주장

꾸?

그럼 안녕!

Q & A

LIFE
OF
PURPOSE

당근이세영?
영갑니당!



뭔가 기대하는 표정으로
아무 말 안하고
머뭇거리는데

“음…. 뭔가 iphone검색해도
아이폰 문서가 나오게 하고 싶고,
IPHONE 검색해도 아이폰 문서가 나오게 하고 싶고
핸드폰 해도 아이폰 나오게 하고 싶어요.”



“음…. 뭔가 iphone검색해도
아이폰 문서가 나오게 하고 싶고,
IPHONE 검색해도 아이폰 문서가 나오게 하고 싶고
핸드폰 해도 아이폰 나오게 하고 싶어요.”

“아니 한국에서 하는 서비스인데,
‘나이키 신발’, 이라 검색하나 ‘나이키신발’ 이라 검색하나
‘nike 신발’ 이라 검색하나 문서가 같아야 하는 거 아닌가?”



뭔가 기대하는 표정으로
아무말안하고
머뭇거리는데

“음… 뭔가 iphone검색해도
아이폰 문서가 나오게 하고 싶고,
IPHONE 검색해도 아이폰 문서가 나오게 하고 싶고
핸드폰 해도 아이폰 나오게 하고 싶어요.”

“아니 한국에서 하는 서비스인데,
‘나이키 신발’, 이라 검색하나 ‘나이키신발’ 이라 검색하나
‘nike 신발’ 이라 검색하나 문서가 같아야 하는 거 아닌가?”



“아니 오타 좀 날 수 있지,
태플릿이나 테블릿이나 패라가모나 페라가모나,
핑크솔트나 핑크소금이나….”

“음.... 뭔가 iphone검색해도
아이폰 문서가 나오게 하고 싶고,
IPHONE 검색해도 아이폰 문서가 나오게 하고 싶고
핸드폰 해도 아이폰 나오기

“아니 한국에서 하는
‘나이키 신발’, 이라 검색하나 ‘나이
‘nike 신발’ 이라 검색하나 문서가

“아니 오타 좀 날
태플릿이나 테블릿이나 패라
핑크솔트나 핑크소금



“음.... 뭔가 iphone검색해도
아이폰 문서가 나오게 하고 싶고,
IPHONE 검색해도 아이폰 문서가 나오게 하고 싶고
나오기



하는
나
서가
좀 날
패라
핑크솔트나 핑크소금



“음.... 뭔가 iphone검색해도
아이폰 문서가 나오게 하고 싶고,
IPHONE 검색해도 아이폰 문서가 나오게 하고 싶고



그렇다. 이 수준 이상을 가야된다면...
DB로는 불가능하다.

태블릿이나 테블릿이나 패라
핑크솔트나 핑크소금



뭔가 기대하는
아무말안하
머뭇거리

호에에에에엥에
넘모좌송해요



좀 날

Example: 아버지가방에들어가신다.

Hannanum	Kkma	Komoran	Mecab	Twitter
아버지가방에 들어가 / N	아버지 / NNG	아버지가방에 들어가신다 / NNP	아버지 / NNG	아버지 / Noun
이 / J	가방 / NNG		가 / JKS	가방 / Noun
시느다 / E	예 / JKM		방 / NNG	예 / Josa
	들어가 / VV		에 / JKB	들어가신 / Verb
	시 / EPH		들어가 / VV	다 / Eomi
	느다 / EFN		신다 / EP+EC	

AI/ML 친구들은 kolnpy나 komoran 같은 것들을
많이 봐왔을 건데, 이런 녀석이다.

형태소 분석기가 필요하다.

학부때 검색 열심히 만들어보기

검색 파보는 이야기

elasticsearch를 써보자

github.com/drakejin

미래를 여는 검색 솔루션

Elastic은 강력한 검색 기능을 통해 사람들이 차별화된 방식으로 데이터를 탐색하고 분석할 수 있도록 도와줍니다.

[시작하기](#)

고객 사례

Cisco: IT 투명성을 위해 20억 자산 관리

Elastic Enterprise Search를 통한 종합적인 IT 가시성을 통해 법률, 재무 및 운영 리스크를 줄일 수 있습니다.

유료 광고 포함

새로운 소식

예산이 필요 없는 프리 티어

Workplace Search 7.9: Gmail 통합, 새로운 Kibana 인터페이스, 그리고 새로운 무료 티어를 제공합니다.



이벤트

사전예약하거나 지금 시청하세요.

진행 예정인 이벤트를 검색하거나 온디맨드 콘텐츠에서 지난 이벤트를 찾아보세요. 로드맵 세션, 사용자 이야기, 커뮤니티 멋업, 전문가 심층 분석 등 다양한 콘텐츠가 제공됩니다.



Elastic Stack

기능

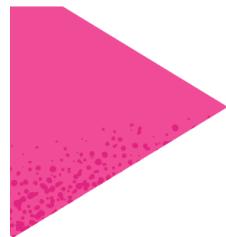
Elasticsearch Service

Elasticsearch

Kibana

Beats

Logstash



The Elastic Stack

핵심 제품을 만나보세요. 모두 무료이며 오픈 소스입니다.

바로 Elasticsearch, Kibana, Beats, Logstash([ELK Stack](#)이라고도 하죠)입니다. 모든 형식의 모든 소스에서 안정적이고 보안이 유지되는 방식으로 데이터를 가져온 다음 실시간으로 검색, 분석, 시각화하세요.

[무료 체험판 시작](#)

신용카드가 필요 없습니다. 간편한 설정.

↳ [또는 다운로드하여 시작하세요](#)



Elasticsearch 엔지니어링 트레이닝을 통해 Elasticsearch 작업의 강력한 기반을 마련해 보세요.

유료 광고 포함
과정 보기 →

Elastic Stack을 사용할 준비가 되셨나요? 가동 중단 시간 없이 업그레이드할 수 있는 모범 사례를 알아보세요.

동영상 보기 →

Elastic Stack을 업그레이드하여 각 릴리즈의 새로운 기능을 활용할 수 있는 전문가의 팁을 확인하세요.

동영상 보기 →



elasticsearch

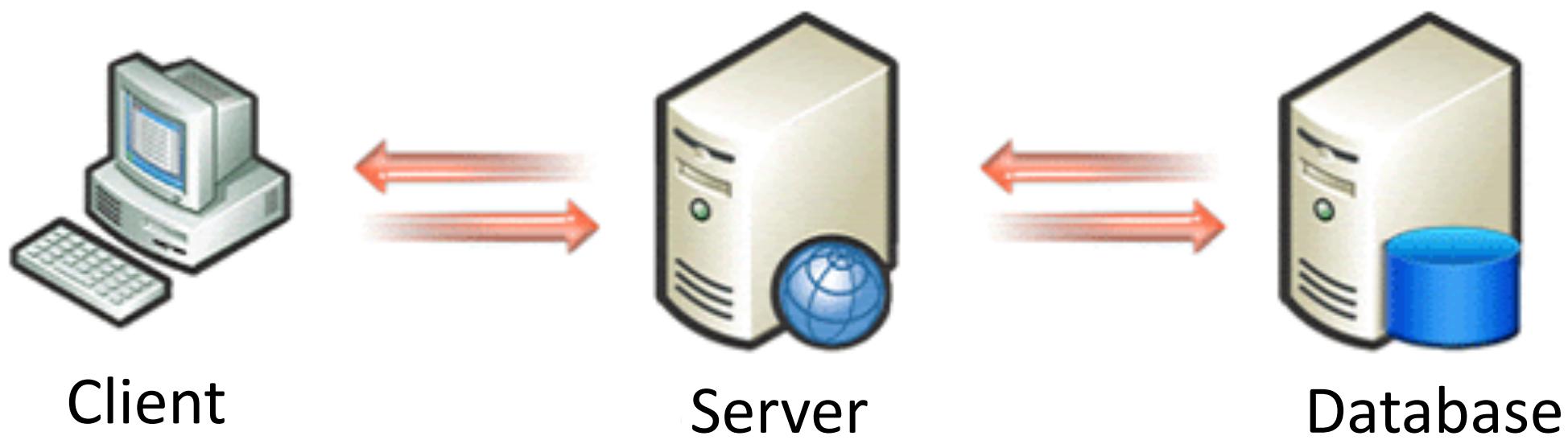
검색서비스를 위한 DB, Elasticsearch
Elasticsearch에는 형태소 분석기가 내장되어 있다.

검색에 관련된 모든 기능이 여기 다 있다.

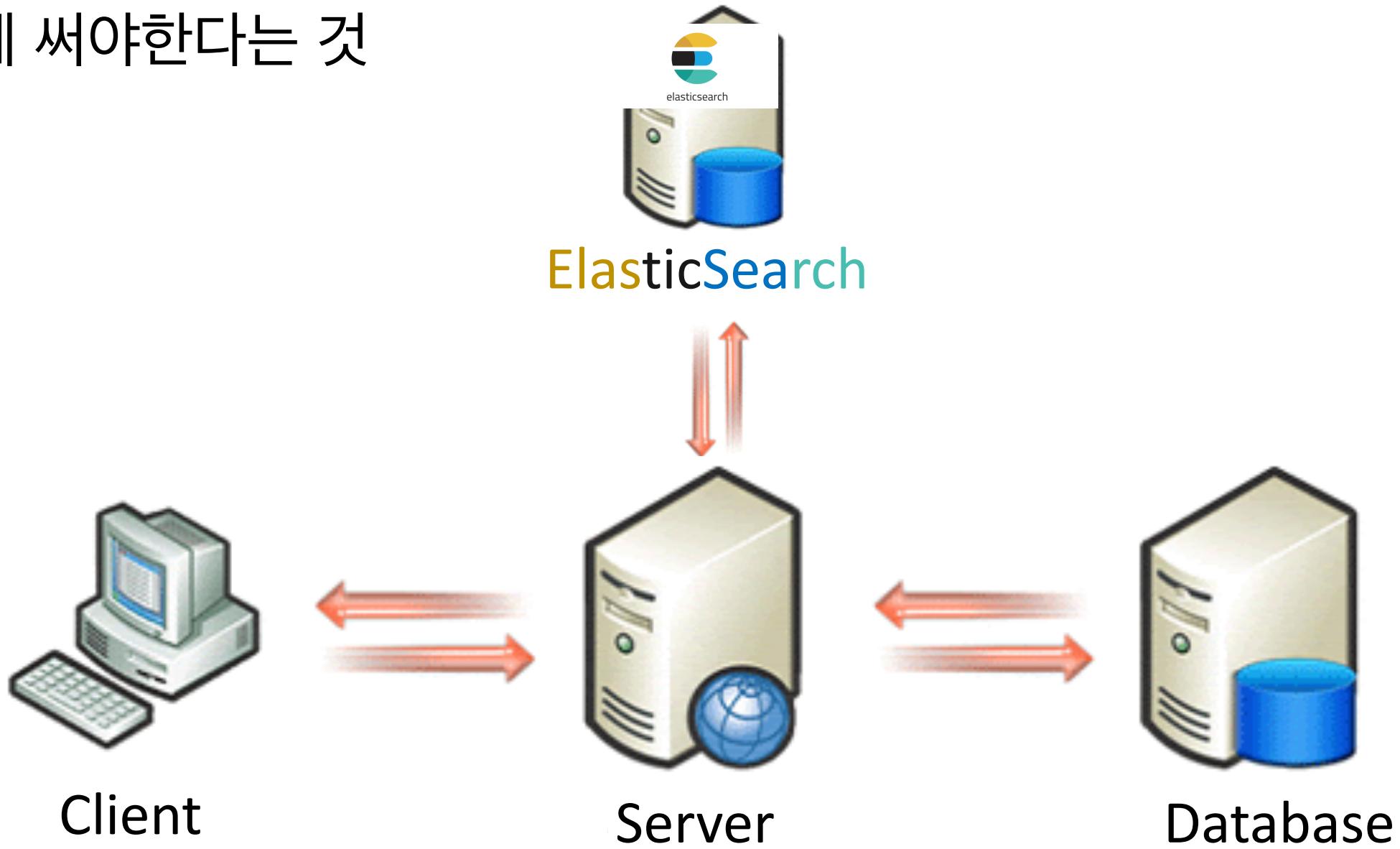
먼저, Elasticsearch를 이용해서 기능을 만들기 앞서
우선 Elasticsearch에 대해 알아보도록 하자

https://docs.google.com/presentation/d/1K5BJEtL0kjA33jaF1nyGhDNgwAeyPz4unjU45bWpdpg/edit#slide=id.g8e6bc3cf2f_5_0

음.. 그러니까.. 이게..



이렇게 써야한다는 것



ElasticSearch에다가 데이터는 어떻게 밀어넣어요?

1. 데이터 추가 및 수정시

DB에다가도 하고 ElasticSearch에다가도 하고

2. 30초~1분 마다 DB를 읽어서 ElasticSearch에
변경사항을 넣는 방식

1번은 코드를 짜서하는 구현이라 다른거 배우지 않고
바로 적용할 수 있다.

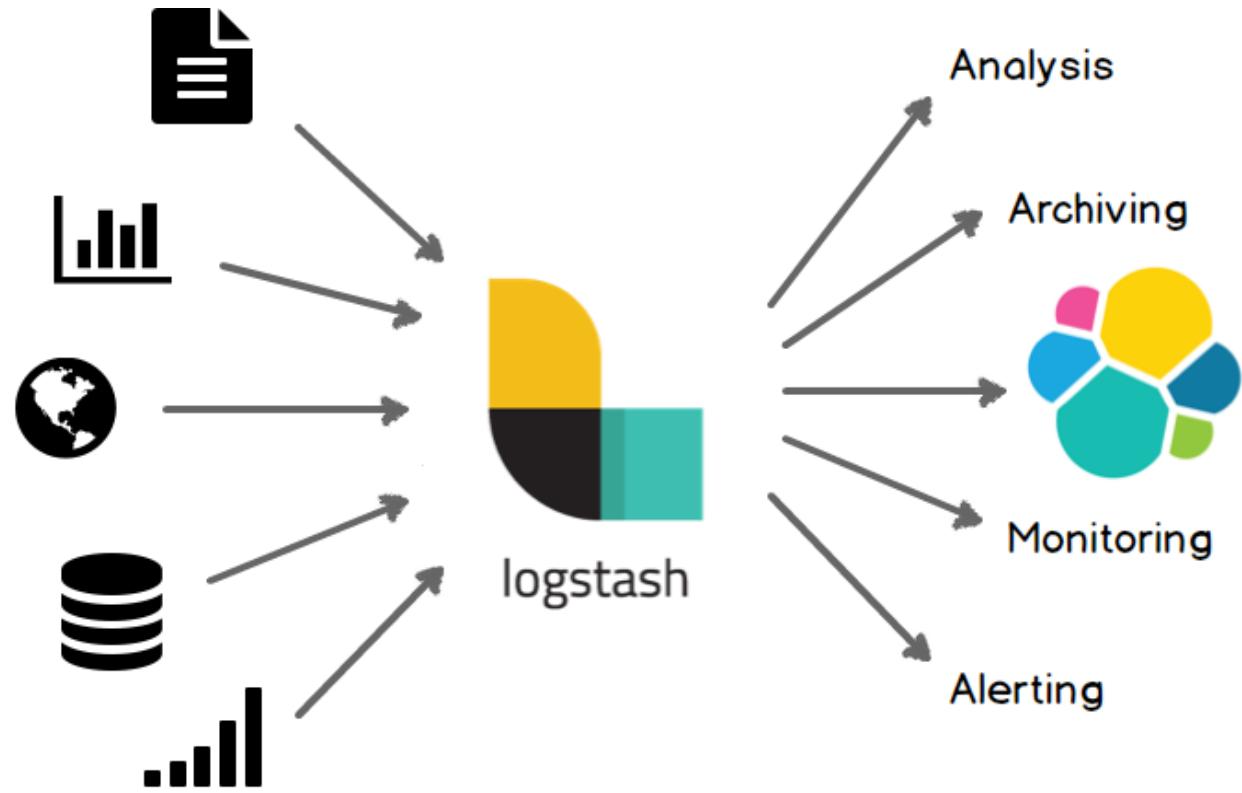
2번은 다른 프로그램을 쓰는 게 일반적이다.
다른 프로그램을 배워야하는게 귀찮지만,
사용설명서가 잘 나와있어 금방 써먹을 수 있다.
나중 실무에서도 잘 써먹을 수 있다.

1. Just Coding...
2. Use LogStash

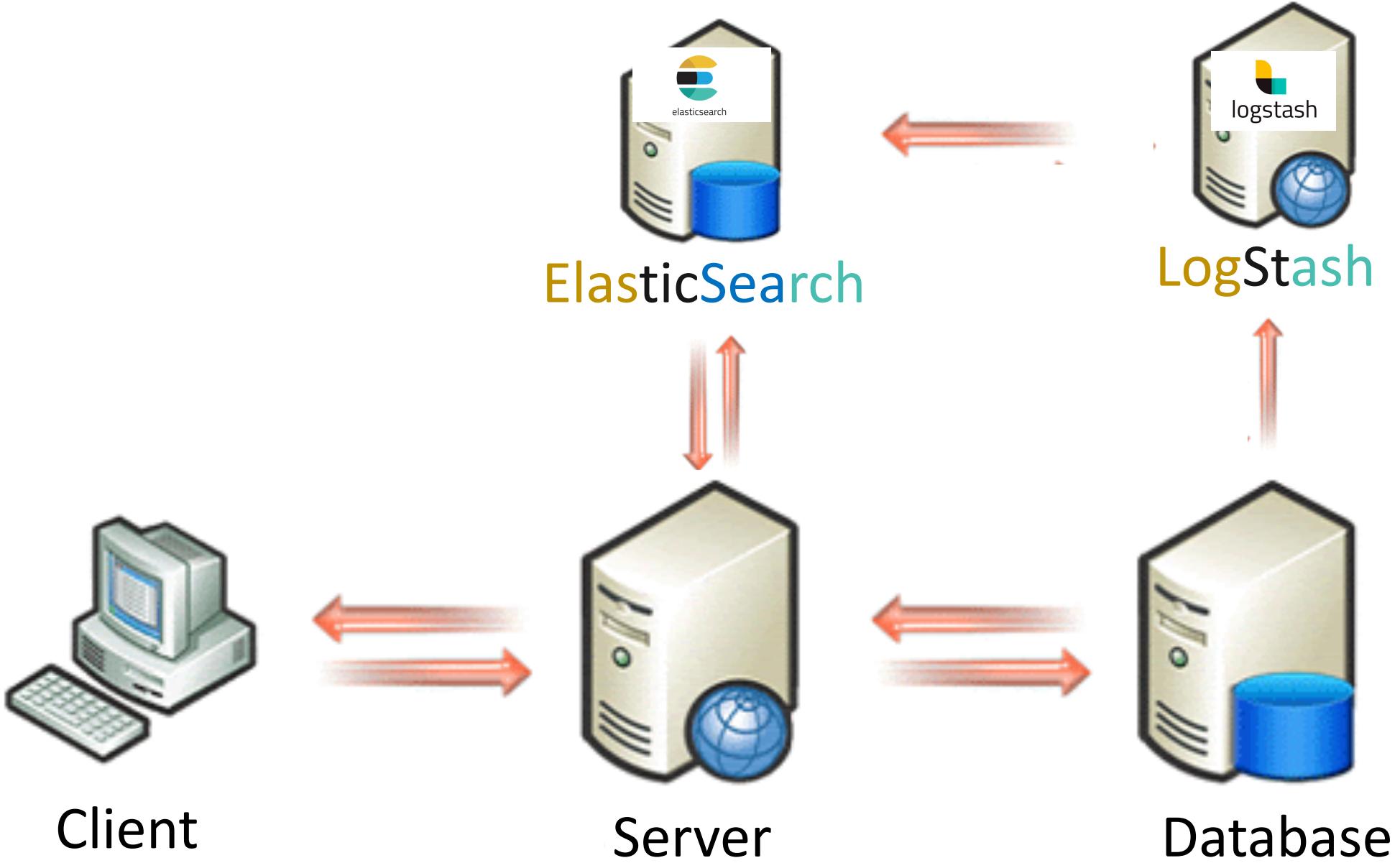


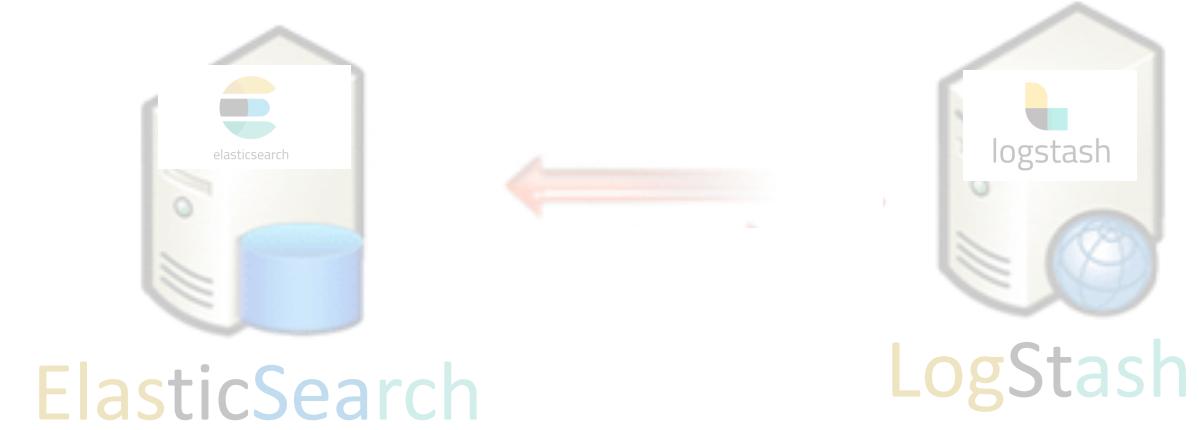
logstash

Elasticsearch에다가 데이터 넣어주는 친구

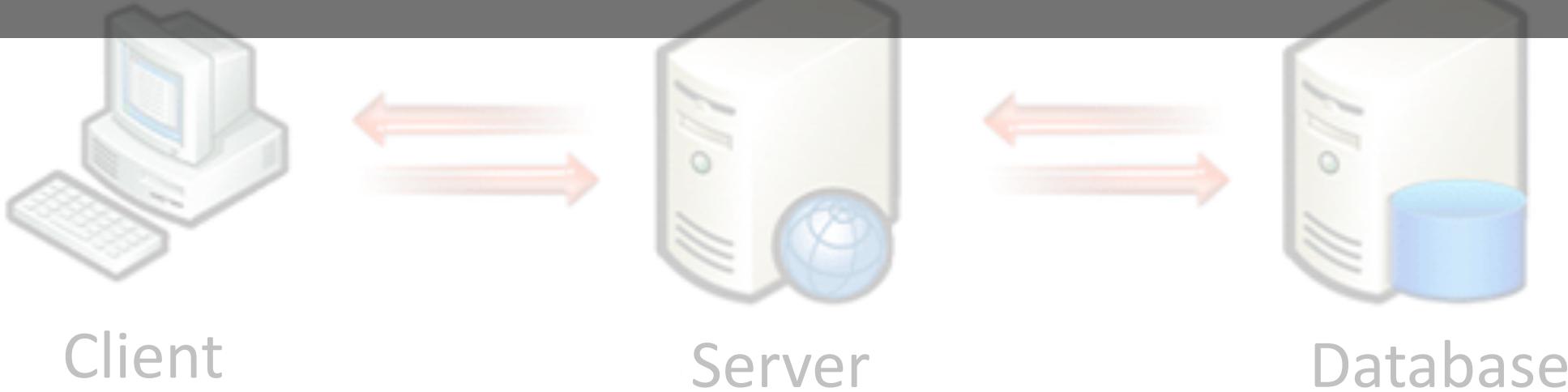


다양한 타입의 데이터를 받고 다양한 타입으로
전송 해주는 유틸기능들이 많다.
데이터파이프라인(여러처리) 구성시에 자주 사용된다





잠깐, 학생이라 이렇게 수 많은 컴퓨터를
호스팅할 컴퓨터가 없다구요?





Client



Server



Database



ElasticSearch



LogStash





ElasticSearch



LogStash

내 컴퓨터에 이렇게 많은 프로그램들 깔기 귀찮으시다구요?

나도 모르는 환경설정 건들까봐 무섭다구옵?



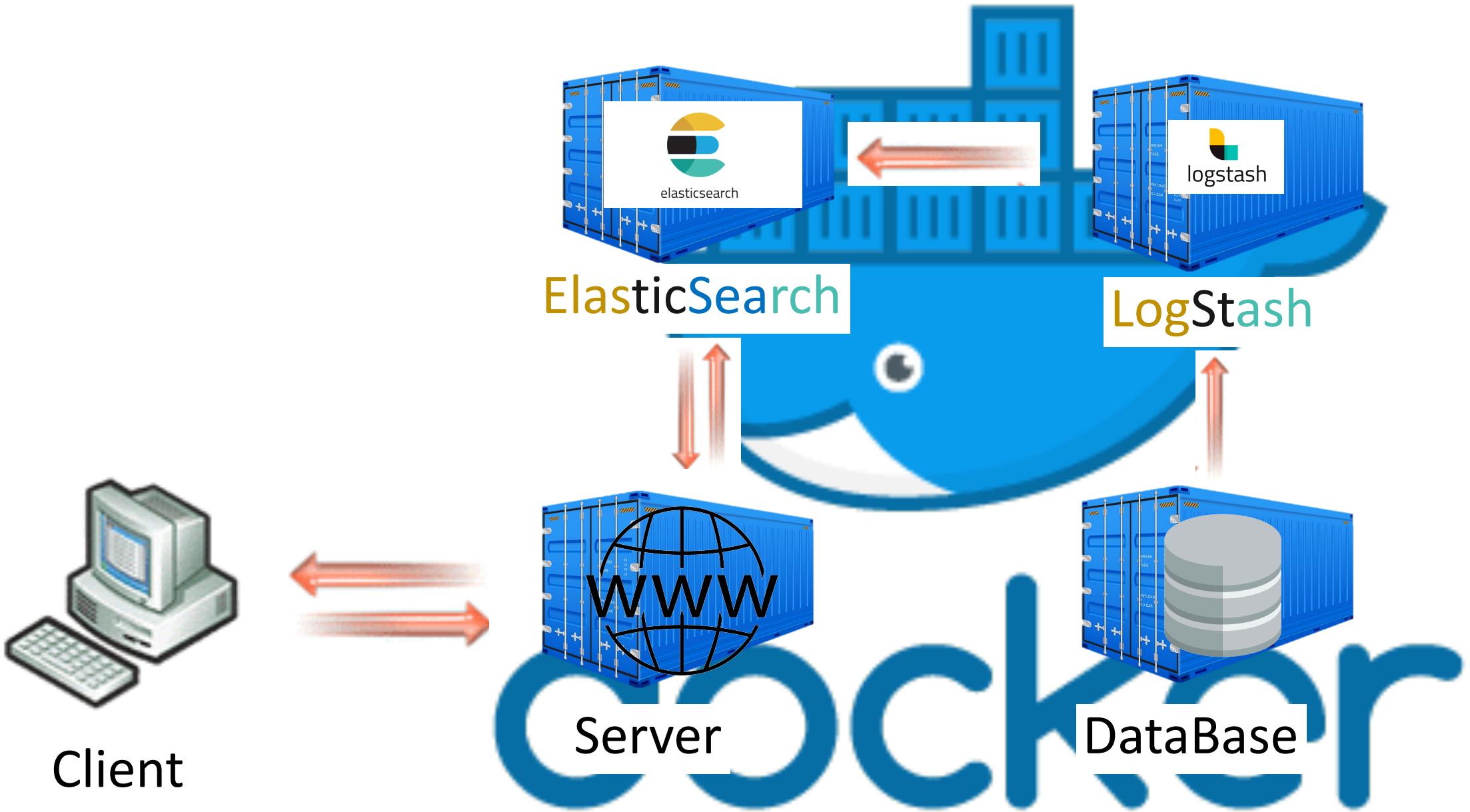
Client



Server



Database





옛날에는 운영체제 통째로 CD로 꾸워서 배포하던 걸
Container라는 개념으로, 쉽게 업로드하고 인스톨할 수
있는 환경을 구성하는 오픈소스 도구.

특, 현업에서 불티나게 쓰인다.

앞으로 보여줄 예제코드들은 전부 docker를 이용해서
실행할 수 있도록 되어있다.

<https://www.44bits.io/ko/post/easy-deploy-with-docker>

<http://pyrasis.com/docker.html>

Docker는 이 발표에서는 다루지 않는다..!

유용한 학습 link를 남기도록 한당

<https://github.com/drakejin/search>

실습해보기 (이거 하면 다들 집 못감)

실습과제: 여기에 있는거 다 따라해보기
그리고 elasticsearch 홈페이지 보면서 나름 자신있는
index schema와 search query 만들어보기

테스트로 넣어놓은 wiki 데이터를 넣어두고

어떻게 형태소분석기와 랭킹을 만들어야
검색경험이 좋아지는지 연구를 시작할때다.

분명 100% 마음에 들진 않겠지만,
그것은 여러분의 랭킹이 구린 탓이 아니다.

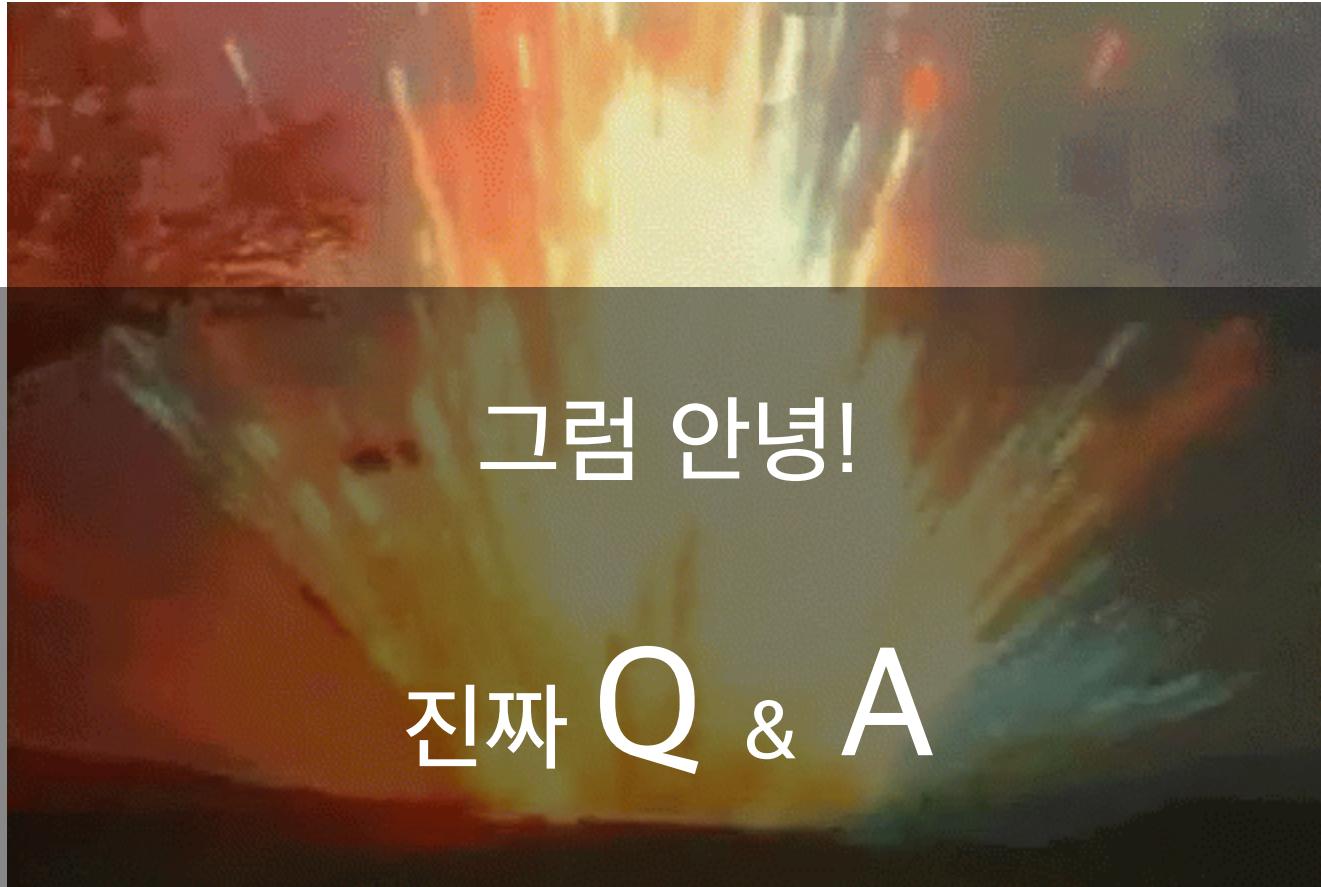
애초에 문서가 구린 탓 이다.



아으 내가 뭔 혀소리들을 한건진 잘 모르겠다.



아.. 후 히읕ㄴ미허ㅓ 둔우ㅓㅓ 쿠쥬덟ㅇㄴ퓨ㅓ버ㅓ스ㄷ
미라ㅓㅓ;ㅁ릉 만



그럼 안녕!

진짜 Q & A

아.. 후 히읗 미허나 둔우나 그 쿠드온 퓨나 버나스
미라나 ; 몬 룽 만