

오픈소스 DBMS 활용동향



위데이터랩(주)
김정수 andy@wedatalab.com



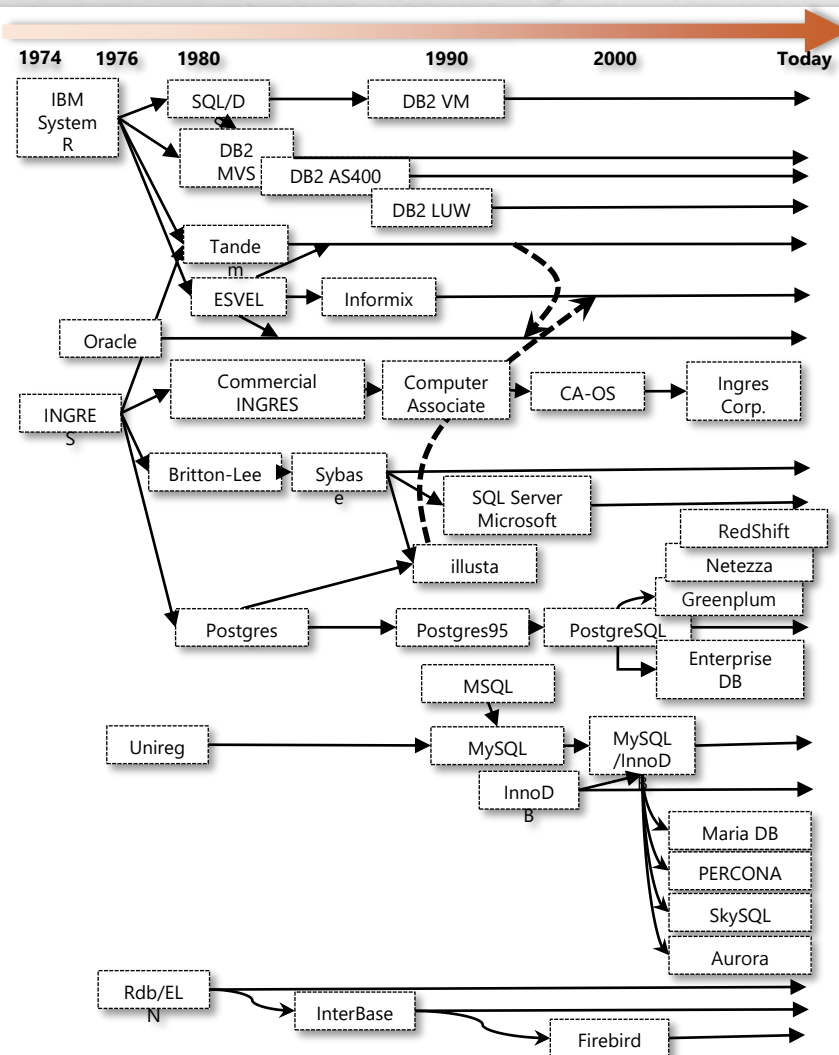
mongoDB



cassandra

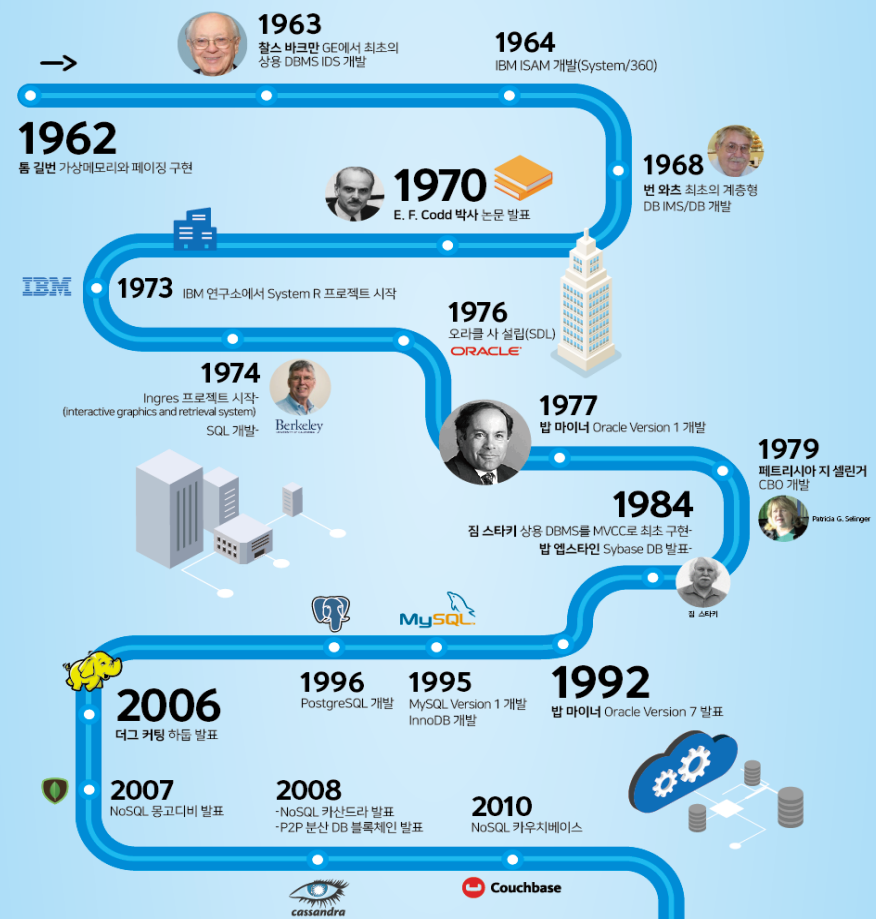


Genealogy of DBMS



source : Andrew Mendelsohn, Rich Niemiec
<https://en.wikipedia.org>

History of Database Systems



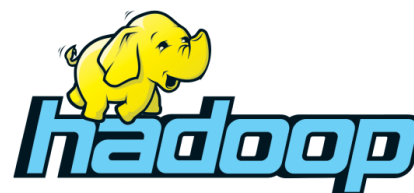
History of DBMS

ORACLE



IDS

IBM DB2



APACHE HBASE

VOLTD B

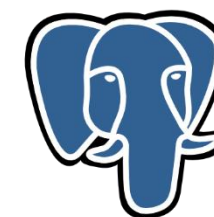
MySQL

ORACLE RAC

BLOCKCHAIN

cassandra

mongoDB



PostgreSQL

RDBMS, NoSQL, VectorDB 현황

RDBMS

- Oracle Database
- MySQL
- Microsoft SQL Server
- PostgreSQL
- SAP HANA DB

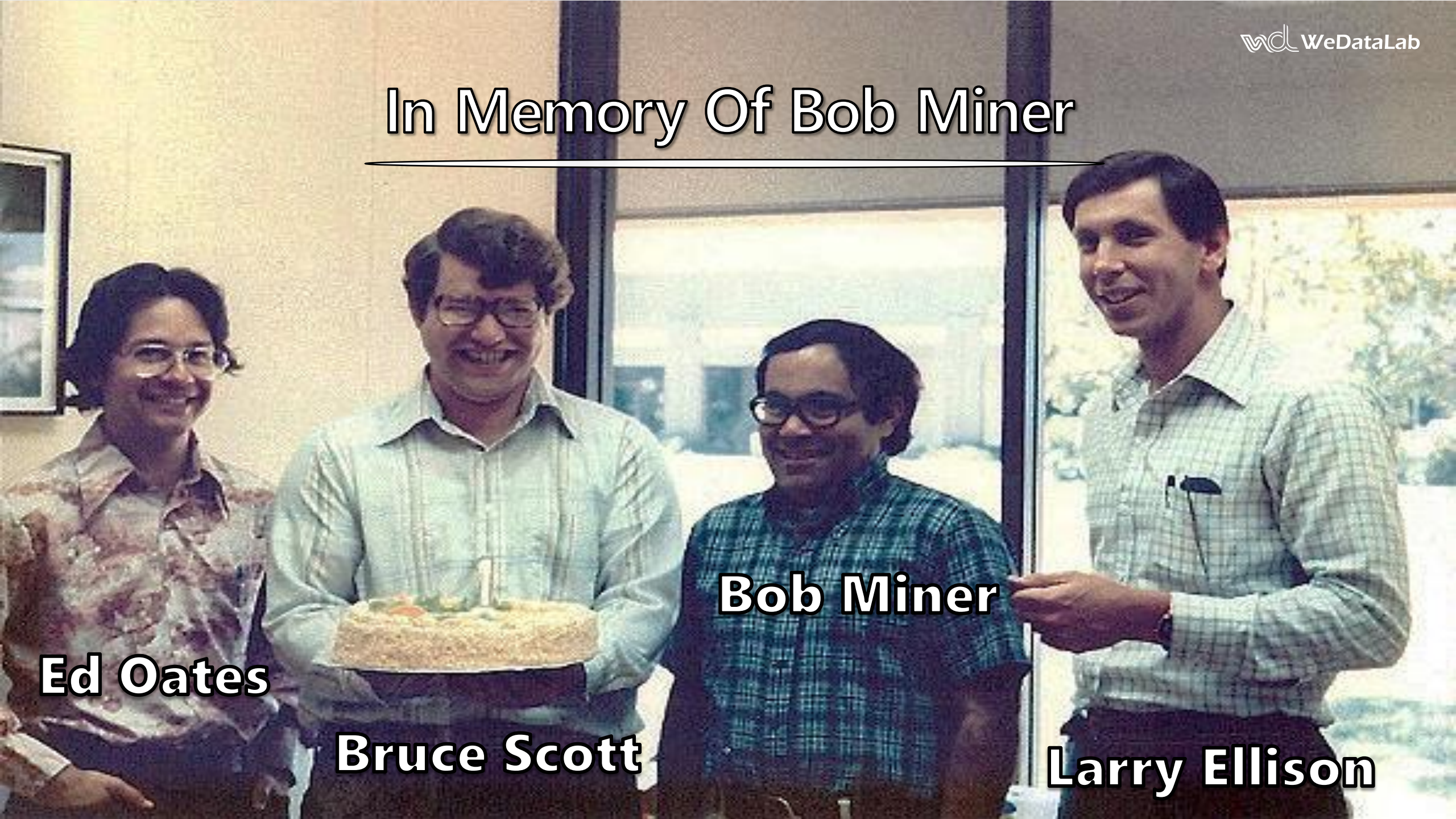
NOSQL

- MongoDB
- Cassandra (Apache)
- DynamoDB (Amazon)
- Couchbase
- Redis

VectorDB

- Milvus (Zilliz)
- Faiss (Facebook AI)
- Pinecone
- Weaviate
- Qdrant

In Memory Of Bob Miner



Ed Oates

Bruce Scott

Bob Miner

Larry Ellison

A man with glasses, wearing a dark suit jacket over a light-colored striped shirt, is standing on a stage. He is holding a small black device in his right hand and gesturing with his left hand. The background is a plain white wall.

Andy Mendelsohn

오픈소스 DBMS

PostgreSQL



Postgres Global Development Core-Team



오픈소스 DBMS



MySQL 명칭

My

Monty

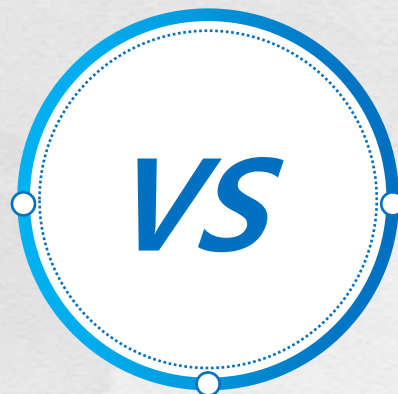
Maria

A photograph of two men sitting on a stage during a panel discussion. The man on the left is wearing a light blue and white striped button-down shirt and glasses. The man on the right is wearing a light blue button-down shirt with a small logo on the chest. They are both looking towards the right side of the frame. The background is a solid red wall.

Heikki Turri

Monty Widenius

오픈소스DBMS 활용동향 , 어떤 DBMS를 선택할 것인가?



오픈소스DBMS 활용동향

공공 클라우드 사업 참여 오픈소스 DBMS 기업 · 제품

선재소프트

골디락스

클라우드 환경 지원 탄력적인 수평 확장(스케일 아웃)

알티베이스

알티베이스

고객 장애 대응 능력, 기술 지원, 비용 경쟁력

인젠트

엑스퍼DB 플랫폼

활용성과 편의성, 서비스 다양성

큐브리드

큐브리드

원천 기술 기반 기술경쟁력, 고객밀착형 기술 지원

티맥스소프트

하이퍼SQL

보안, 기술 지원, 모니터링 및 운영관리 기능 강화

마리아DB

마리아DB

모든 클라우드 지원 능력, 높은 경제성

EDB

EDB 포스트그레

포스트그레SQL 전문기업, 오라클 호환성

오픈소스DBMS 활용동향

위데이터랩

서비스팩

오픈소스 DB 출시하다!

ezisMDB



mariadb + 설치, 패치, 유지보수, 아키텍처링 + 모니터링 솔루션 패키징

ezisPDB



postgresql + 설치, 패치, 유지보수, 아키텍처링 + 모니터링 솔루션 패키징

	basic	pro	enterprize
 ezisMDB	<ul style="list-style-type: none"> ● 설치, 패치 유지보수 	<ul style="list-style-type: none"> ● 설치, 패치, 유지보수, 모니터링솔루션 	<ul style="list-style-type: none"> ● 설치, 패치, 유지보수, 모니터링솔루션, 백업복구, 이중화 컨설팅
 ezisPDB	<ul style="list-style-type: none"> ● 설치, 패치 유지보수 	<ul style="list-style-type: none"> ● 설치, 패치, 유지보수, 모니터링솔루션 	<ul style="list-style-type: none"> ● 설치, 패치, 유지보수, 모니터링솔루션, 백업복구, 이중화 컨설팅

추가 컨설팅

● 성능 튜닝

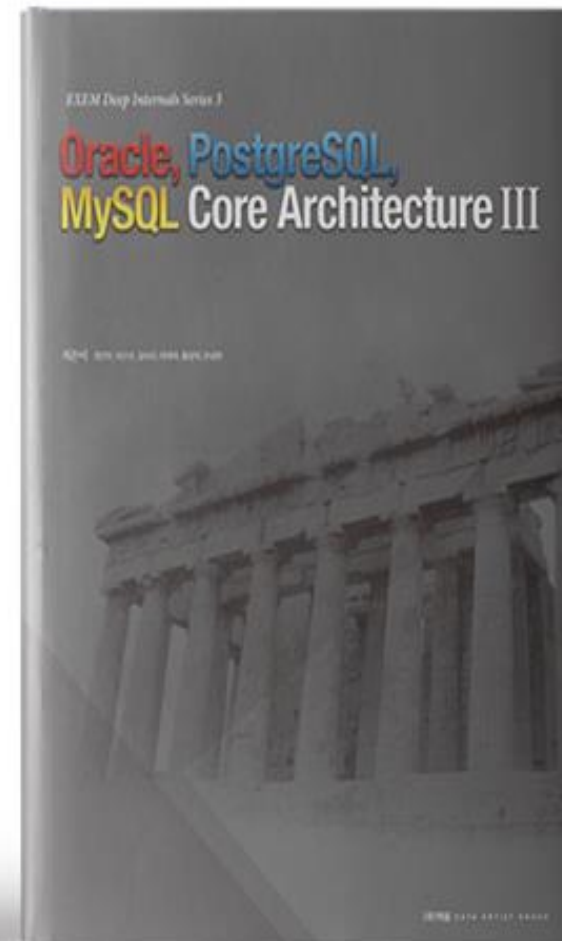
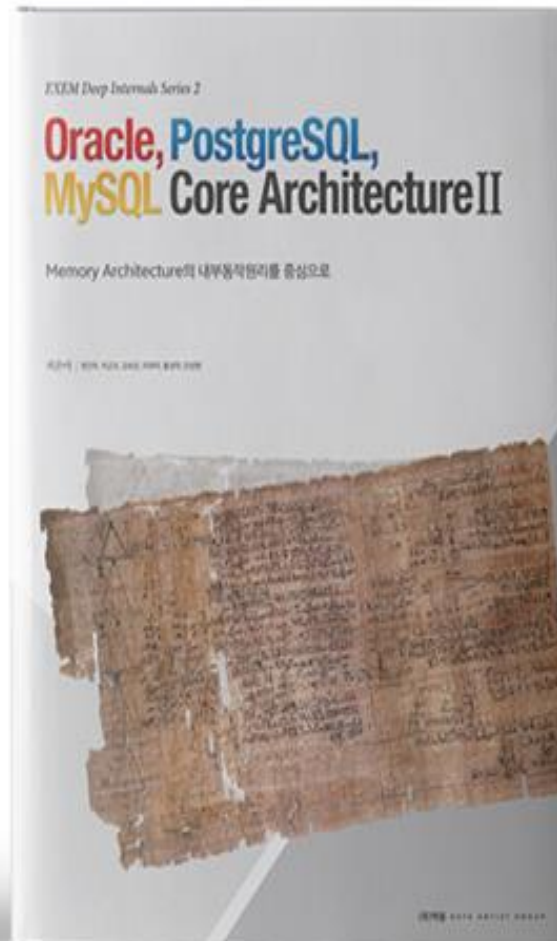
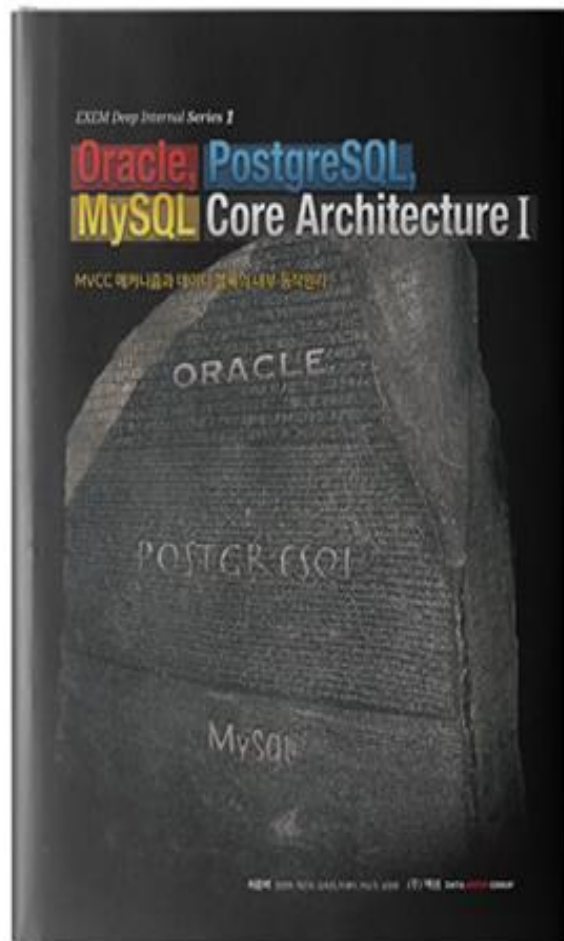
● migration

● 오픈소스 DB교육

● 정보계 Architecture

* 자세한 사항은 권건우 010-6400-9127 로 문의바랍니다

오픈소스DBMS 활용동향



KB국민은행, 'KB 원 클라우드' 통합 운영체제 구축 나서

디지털데일리 | 발행일 2022-04-16 12:39:34

이상일



오픈소스DBMS 활용동향



오픈소스DBMS 활용동향



S-Core

오픈소스DBMS 활용동향



KOREAN AIR



오픈소스DBMS 활용동향



오픈소스DBMS 활용동향



오픈소스DBMS 활용동향

우아한
형제들

오픈소스DBMS 활용동향



서울아산병원
Asan Medical Center



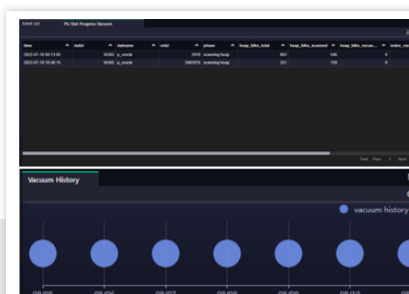
삼성서울병원

오픈소스DBMS 활용동향 , 어떤 DBMS를 선택할 것인가?

Vacuum 작업예측과 처리, 이력정보 관리
실시간 Vacuum 작업수행을 간단히



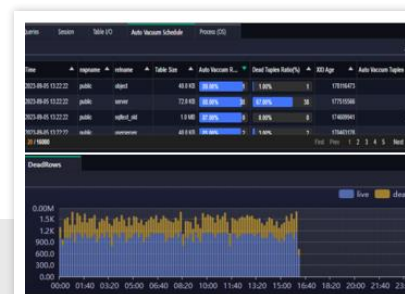
Vacuum 작업예측



Vacuum 처리와 이력정보



Dead Rows 정보



실시간 Vacuum 작업수행

- 예측 가능한 Vacuum 관리로 성능 측면에서 안정적인 운영이 가능하도록 Dead Tuples 비율, Auto Vacuum 예상 비율 그리고 Vacuum 발생 임계값을 제공
- 대량의 트랜잭션 발생 전 간단한 클릭을 통한 수동 Vacuum 작업 수행으로 성능 이슈 방지
- 테이블의 조회성능에 영향을 주는 Dead Rows정보확인

- Auto Vacuum Schedule
- Dead Rows
- Dead Rows List
- Process Vacuum
- Vacuum History

오픈소스DBMS 활용동향 , 어떤 DBMS를 선택할 것인가?

트랜잭션의 Read View 관리 및 Purge발생 정보관리
History List Length 정보를 기반으로 한 Long Query정리



ReadView와 Purge 시각화



분 단위 ReadView와 Purge 수집

• ReadView



• UP limit id

• ACTIVE TX 중의
가장오래된 TRX_ID

• Low limit id

• ReadView 생성
시점의 TRX_ID

- Long Query와 Long Transaction에 의한 아주 많은 양의 데이터를 Purge에 의한 CPU성능과 DISK Write에 영향을 최소화하기 위한 History List Length와 Read View정보 제공

- History List Length
- Read View
- Purge Done
- 관련 통계 정보 수집

오픈소스DBMS 활용동향 , 어떤 DBMS를 선택할 것인가?

wdl WeDataLab

DB internal 전문가그룹

CAIO 강승우 CEO 권건우 CTO 이근오

위데이터랩
오픈소스 DB 출시하다!

ezisMDB

MariaDB mariadb + 설치, 패치, 유지보수, 아키텍처링 + 모니터링 솔루션 패키징

ezisPDB

PostgreSQL postgresql + 설치, 패치, 유지보수, 아키텍처링 + 모니터링 솔루션 패키징

	basic	pro	enterprise
ezisMDB	설치, 패치, 유지보수	설치, 패치, 유지보수, 모니터링솔루션	설치, 패치, 유지보수, 모니터링솔루션, 백업복구, 이중화 컨설팅
ezisPDB	설치, 패치, 유지보수	설치, 패치, 유지보수, 모니터링솔루션	설치, 패치, 유지보수, 모니터링솔루션, 백업복구, 이중화 컨설팅

추가 컨설팅

- 성능 튜닝
- migration
- 오픈소스 DB교육
- 정보계 Architecture

* 자세한 사항은 권건우 010-6400-9127 로 문의바랍니다

wdl WeDataLab

- 위데이터랩 금요 평풍강의 -

"한국최고의 PostgreSQL의 전문가가 설명하는"

PostgreSQL의 내부구조와 튜닝원리

MySQL MariaDB PostgreSQL TmaxTibero SQL Server ORACLE

강사 이근오CTO, 권건우CEO

주최 위데이터랩

후원 서울데이터과학연구원, 한국전자정보통신기술협회

날짜 5월10일 (금) 오후 2시 - 5시

이력 한위데이터랩 CEO, CTO, Oracle, PostgreSQL, MySQL 코아키텍처 1권, 2권 저술, ezis for PostgreSQL 개발, ezis for Oracle 개발, ezis for MongoDB 개발, 삼성카드사내대, ING생명 차세대, 삼성생명차세대, 삼성카드프로젝트

가격 · 무 료

문의 · 010-6400-9127 권건우 대표

내용

1. DBMS 트랜잭션 처리 코어 아키텍처 비교
 - MVCC 아키텍처의 발전
 - MVCC의 두가지 흐름
 - DBMS 별 MVCC 메커니즘의 비교
2. PostgreSQL Deep Internals
 - Memory Architecture
 - Shared Buffers
 - XLog Buffer, XLog file
 - Clog Buffer

참가방법: 신청서 사전 작성후, 강의실 입장
PostgreSQL 신청서: <https://forms.gle/uLZvBYWZpHrPToy26>

오프라인: 서울특별시 종로구 삼일대로 457, 수문회관 1303호 위데이터랩

무료

QR코드로 신청서 작성

wdl WeDataLab

- 위데이터랩 금요 평풍강의 -

"한국최고의 MariaDB의 전문가가 설명하는"

MariaDB의 내부구조와 튜닝원리

MySQL MariaDB PostgreSQL TmaxTibero SQL Server ORACLE

강사 이근오CTO, 권건우CEO

주최 위데이터랩

후원 서울데이터과학연구원, 한국전자정보통신기술협회

날짜 5월24일 (금) 오후 2시 - 5시 (3시간)

이력 한위데이터랩 CEO, CTO, Oracle, PostgreSQL, MySQL 코아키텍처 1권, 2권 저술, ezis for PostgreSQL 개발, ezis for Oracle 개발, ezis for MongoDB 개발, 삼성카드사내대, ING생명 차세대, 삼성생명차세대, 삼성카드프로젝트

가격 · 무 료

문의 · 010-6400-9127 권건우 대표

내용

1. MySQL/InnoDB Undo Architecture
2. MySQL/InnoDB MVCC Architecture
3. Page Internal Architecture
 - 01) Page Layout
 - 02) Directory slot
 - 03) Page 내부 동작원리
 - 04) IBD File의 구조
4. Purge
 - 01) Purge 정의 및 필요성
 - 02) Purge Update Undo
 - 03) Purge Garbage Space

참가방법: 신청서 사전 작성 및 등록
MariaDB 신청서: <https://forms.gle/94nV9c98mtVKAPe58>

세미나 장소: 서울특별시 종로구 삼일대로 457, 수문회관 1303호 위데이터랩

무료

QR코드로 신청서 작성

RAG와 벡터 DB

웹 문서와 LLM 연동 : 지시 + 컨텍스트(Context)

- 웹 문서를 Context로 제공하여 LLM 답변 요청
 - ✓ 입력된 내용과 연관된 문서 검색
(google, 네이버, ...)
 - ✓ 검색된 내용을 가져오기
 - ✓ 해당 문서 첨부하여 LLM 답변 요청

- 검색 정확도에 따라 답변 정확도가 달라짐

지시 : 프롬프트

컨텍스트 : 웹 문서

웹 조회 기반 생성(RAG) = 프롬프트 + 웹 검색



RAG와 벡터 DB

RAG : 문서 검색 기반 생성 (RAG : Retrieval Augmented Generation)

초록마을, AI가 상품 찾아준다

| 마이크로소프트 GPT-4 검색엔진 장착

유통 | 입력 : 2023/08/10 08:38



안희정 기자 | ✉ 기자 페이지 구독 | 📖 기자의 다른기사 보기



[웨비나] ROHM | EEPROM의 기초와 특징, 성능을 최대화시킬 수 있는 테크닉 등을 소개합니다!

D2C 푸드테크 스타트업 정육각이 마이크로소프트의 애저(Azure) 오픈AI GPT-4를 적용한 검색엔진을 자체 개발하고 친환경 유기농 전문 초록마을의 모바일 앱에 전격 도입했다고 10일 밝혔다.

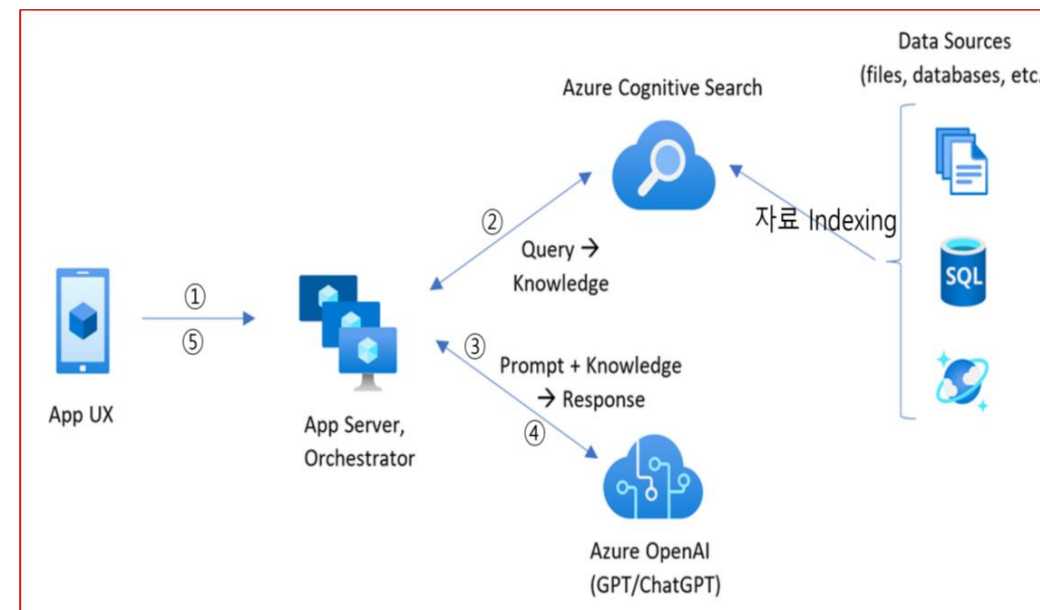
지난달 초 모바일 앱을 네이티브 앱 방식으로 전면 재개발한 데 이어 고객 편의 극대화에 주력한다는 복안이다.

새롭게 적용한 검색엔진은 학습한 검색 패턴을 바탕으로 고객의 의도를 파악



새로운 검색엔진은 정육각 개발팀과 마이크로소프트가 협업한 결과물로 구상부터 적용까지 채 한 달이 걸리지 않았다. 약 2만 개가 넘는 초록마을 상품마스터의 전처리 데이터 생성 및 정리에 GPT-4를 적용하는 것을 구상한 후 마이크로소프트의 전문가 지원 프로그램인 패스트트랙을 통해 한국, 호주의 엔지니어들과 접근 방식을 논의했다.

24년 된 초록마을이 빠르게 검색엔진을 갈아끼울 수 있었던 배경에는 클라우드 기반으로 경영 환경을 전면 전환 하면서 지난 달 초 도입을 완료한 마이크로소프트 애저가 있다. 애저 CosmosDB로 이전한 기존 상품 정보에 GPT-4로 생성한 원천 데이터를 결합하고 애저 Cognitive Search를 이용하는 등 애저 생태계 내에서 매끄럽고 신속하게 새로운 기능을 구현했다.



RAG와 벡터 DB

기업 데이터와 LLM 연동 : 지시 + 컨텍스트(Context)

기업 데이터를 Context로 제공하자.

- ✓ 토큰 제한 : 데이터가 크면?? → 쪼개자 (split)
- ✓ 쪼개지면 필요한 문서를 어떻게 찾을까? → 질문과 유사한 문서를 찾자 (embedding 비교)
 - 질문 : 무엇을 임베딩하면 답변에 필요한 문서를 정확히 찾을까?
- ✓ 문서가 많으면 찾는 속도가 늦지 않을까? → 인덱싱(indexing)을 하자.
 - 벡터 데이터베이스(Vector Database) : 인덱싱된 벡터 관리 시스템

지시 : 프롬프트

컨텍스트 : 벡터 데이터베이스 내의 문서

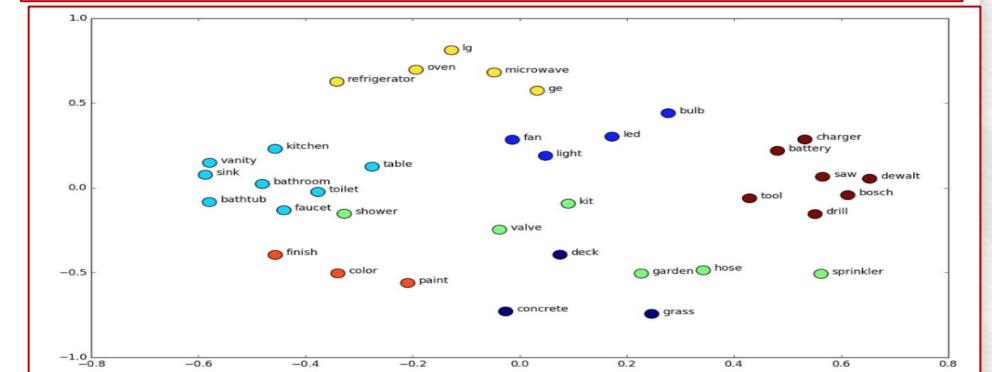
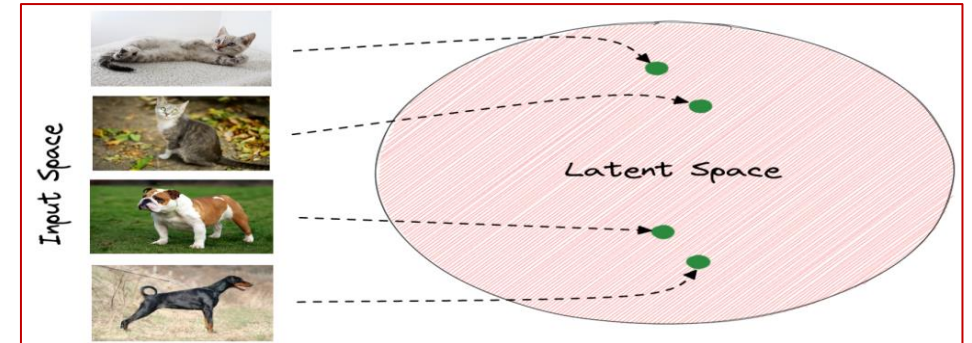
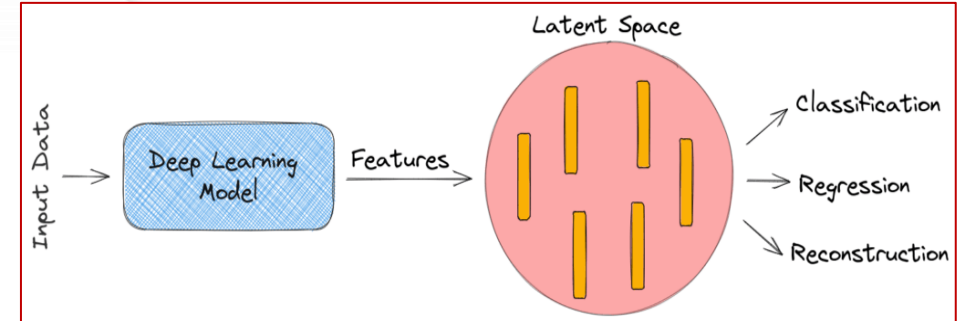
내부 문서 조회 기반 생성(RAG) = 프롬프트 + 벡터데이터베이스



RAG와 벡터 DB

임베딩 → 벡터화

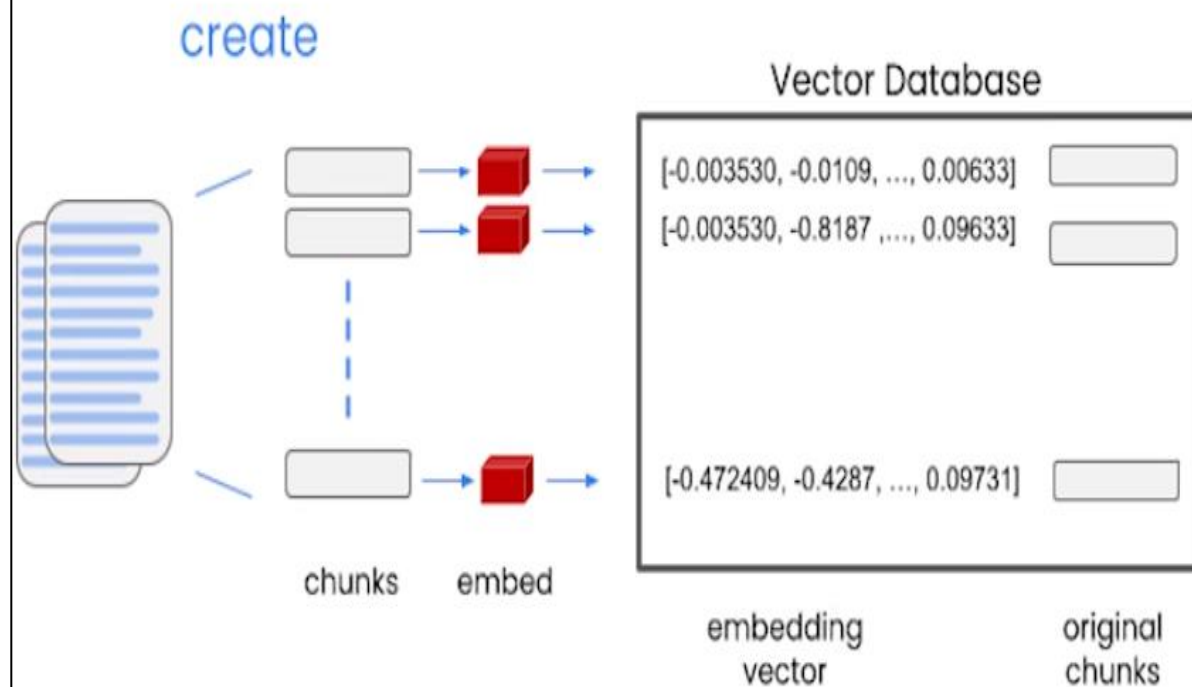
- 비정형 데이터의 임베딩(벡터화) : 잠재 공간 생성
 - ✓ 언어 (문장)
 - 의미적 탐색, 질의 응답, 통역
 - ✓ 이미지(그림)
 - 객체 탐지, 제품 탐지
 - ✓ 소리(음악)
 - 장르 분류, 기계 고장 탐지



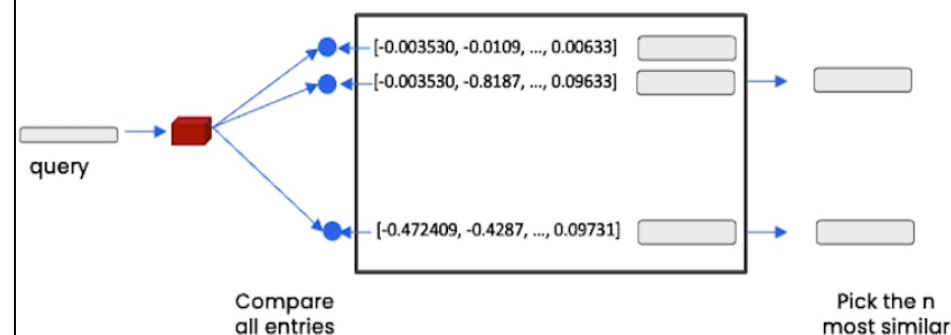
RAG와 벡터 DB

문서 벡터화

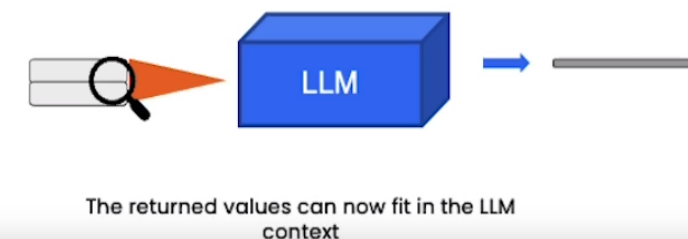
Vector Database



index



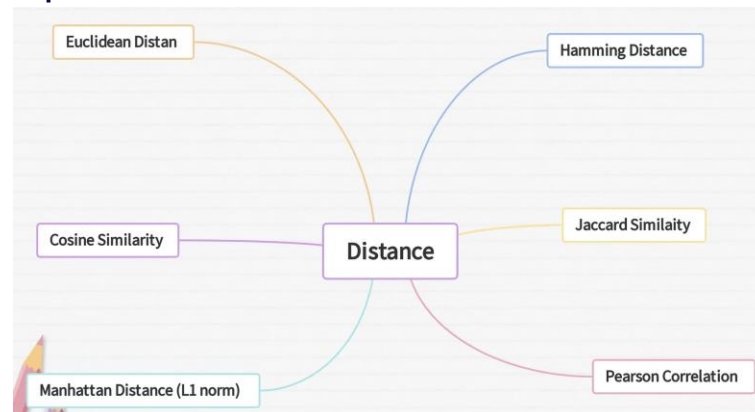
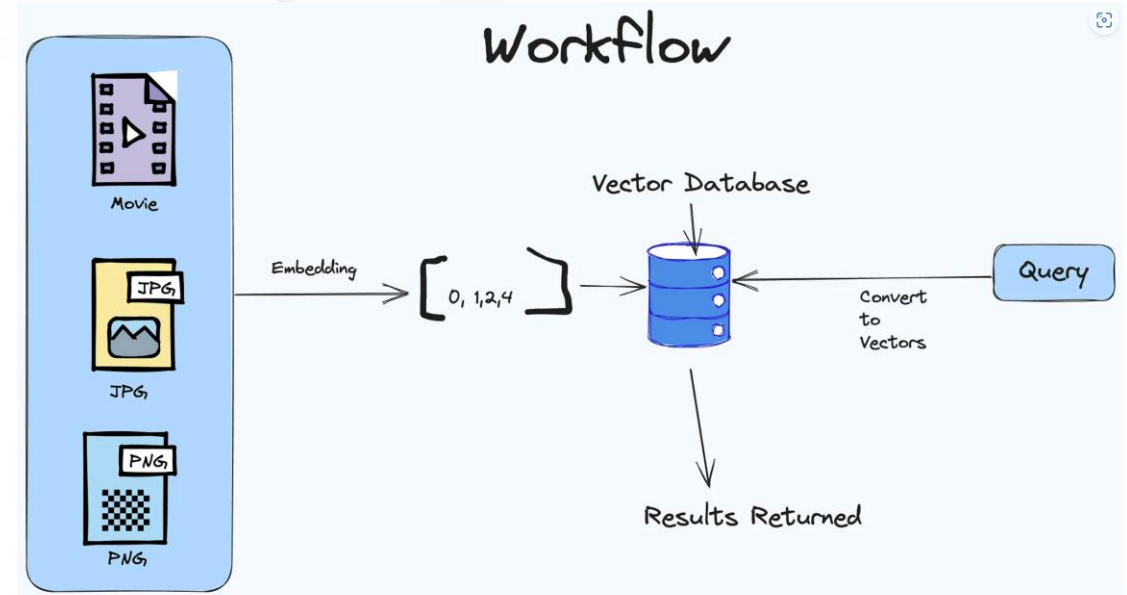
Process with llm



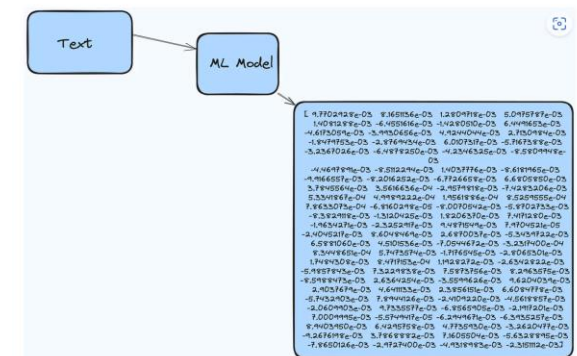
RAG와 벡터 DB

벡터 데이터 베이스

- 벡터를 저장, 탐색 기능 제공
- 탐색 방법 (유클리드 거리, 코사인 거리,...)
- 빠른 탐색을 위한 인덱싱 (Index)
- 메타 데이터에 대한 필터링
- 대용량 벡터에 대한 확장성
- 장애 대응 (Fault Tolerant)
- 벡터 양자화(quantization) 기반 압축
(예 : $\text{round}(3.12410389) = 3$)



Encoding data into vectors:



RAG와 벡터 DB

벡터 데이터 베이스 vs RDBMS

<input type="checkbox"/>	Relational Database (SQL)	Vector Database
1	Database	Collection
2	Table	Vector Space (or sometimes Collection)
3	Row	Vector
4	Column	Dimension
5	Index	Index
6	Select statement	Query/Search
7	Insert statement	Insert
8	Update statement	Not typically used (Embeddings are usually immu
9	Delete statement	Delete
<input type="checkbox"/> ↺↻	Join	Not typically used (Vector operations like nearest
10 records		



Download CSV View larger version

Image by the Author

RAG와 벡터 DB

벡터 탐색 증가세

Papers on Arxiv.org that mention "similarity search" or "semantic search"

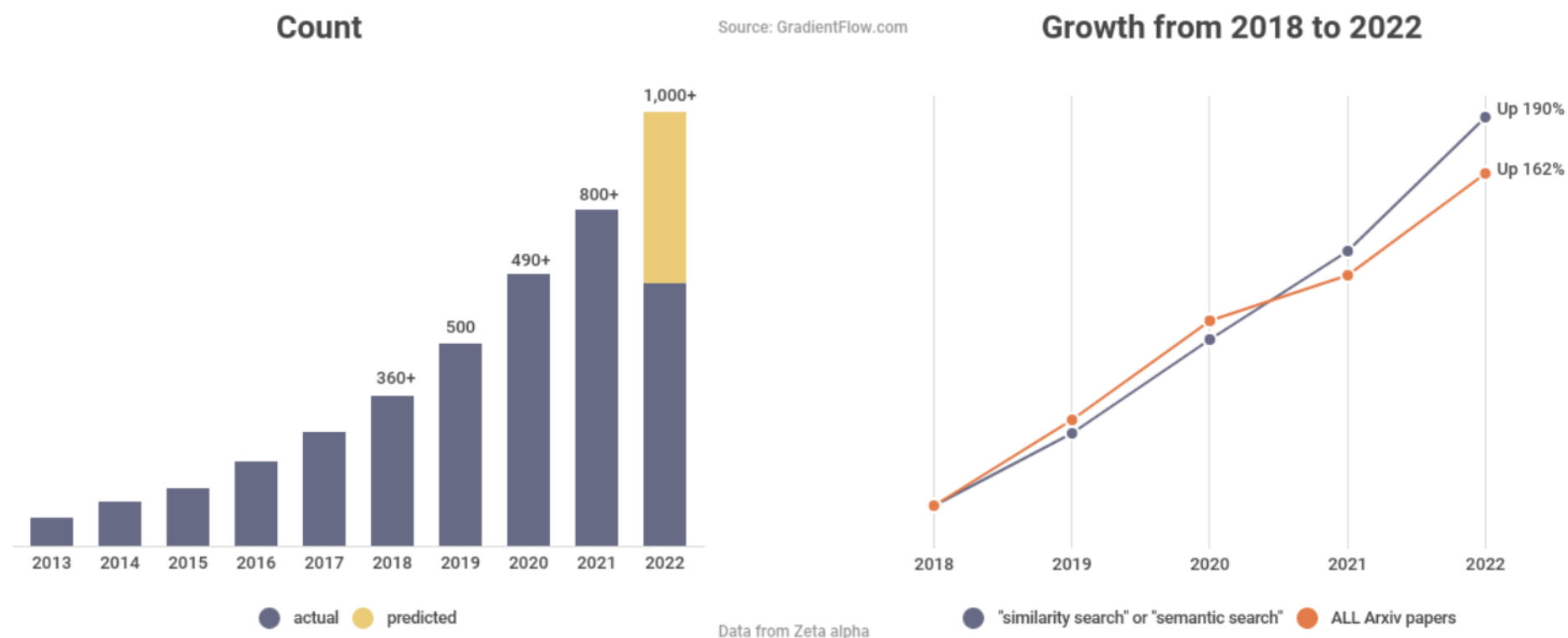
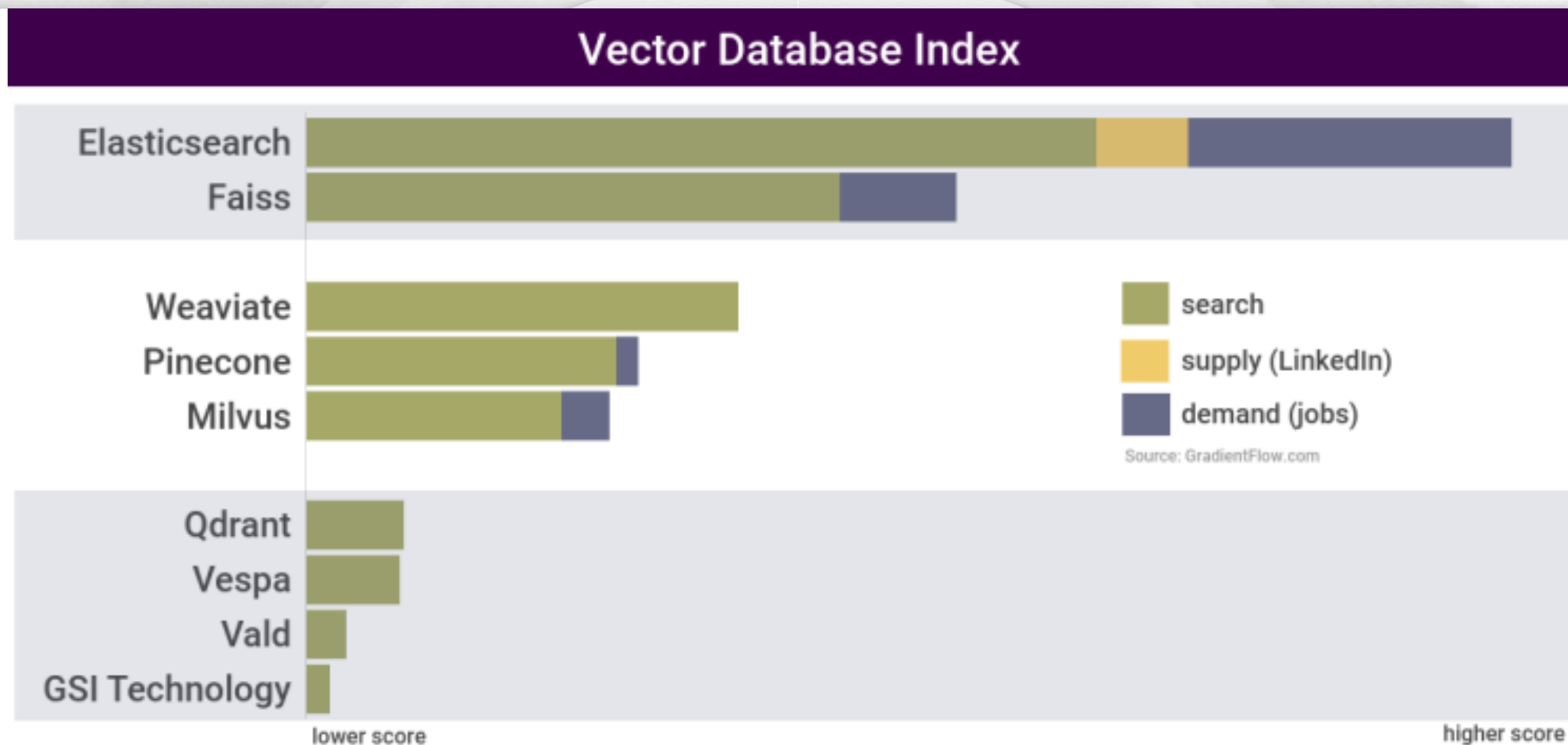


Figure 3: Researchers are publishing more papers on "similarity search" and "semantic search".

RAG와 벡터 DB

벡터 데이터 베이스 사용량, 구직, 구인

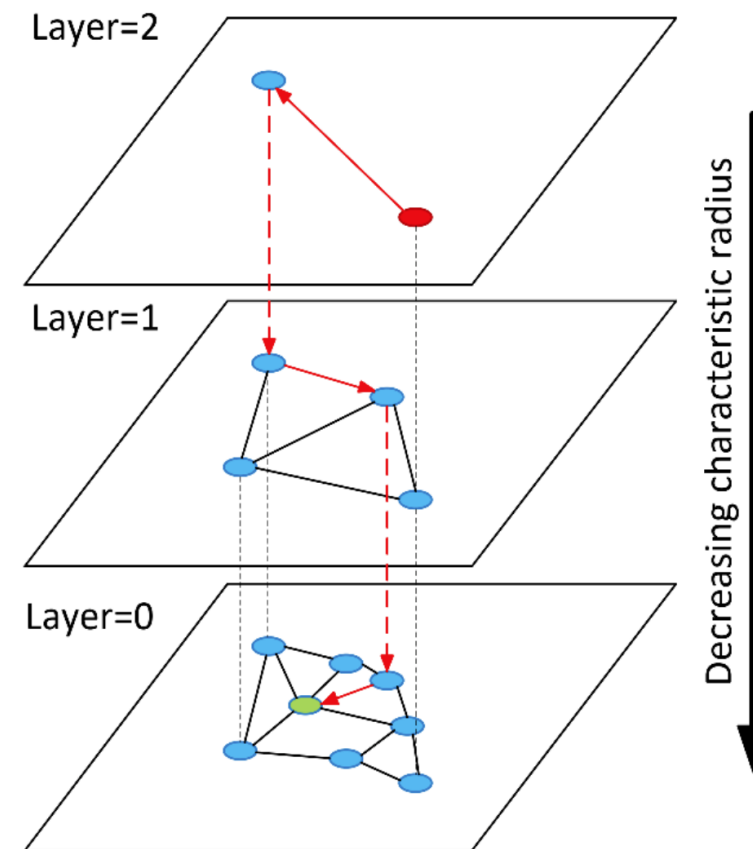
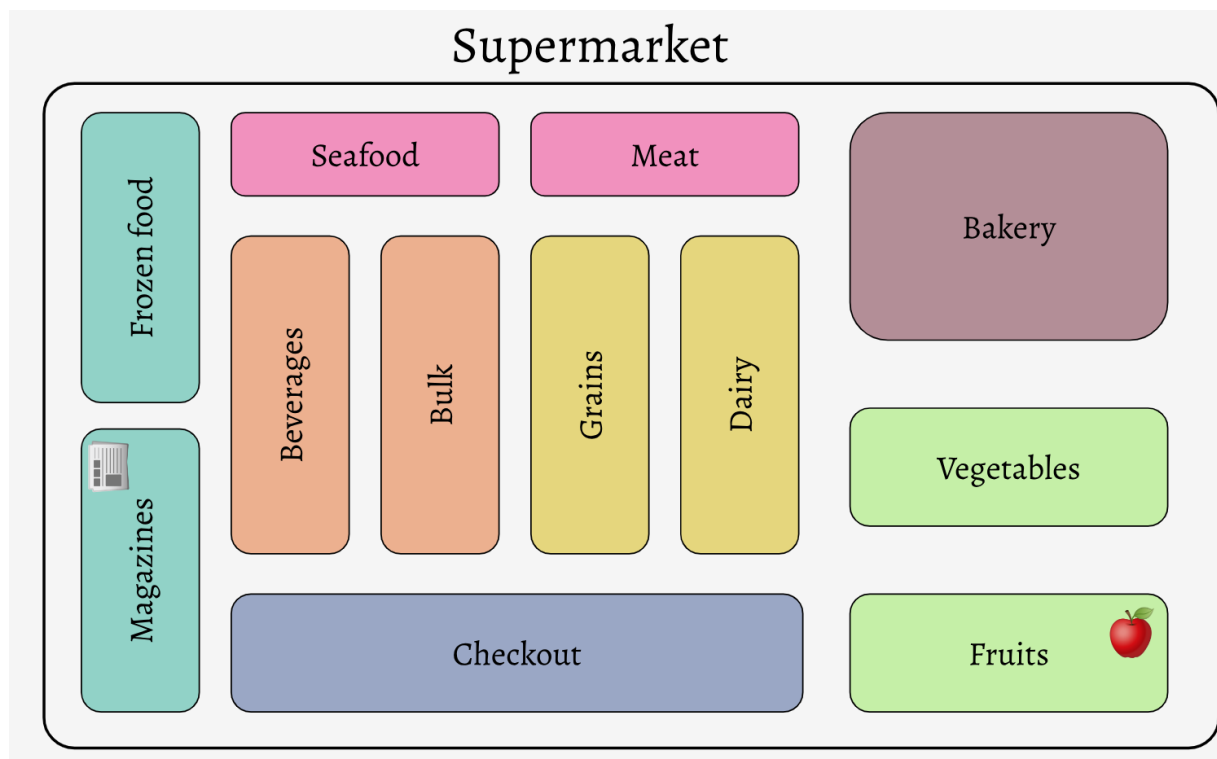


(*) Elasticsearch is a multimodal database; above score is based only on its "vector database" capabilities

RAG와 벡터 DB

벡터 인덱싱 – ANN (Approximate Nearest Neighbors)

- 검색 대상을 근사(?)하게 찾는다
- HNSW (Hierarchical Navigable Small World)
- 다른 방법은?



RAG와 벡터 DB

벡터 인덱싱 성능 튜닝

성능 튜닝 - M

- 한 노드가 가지는 이웃 노드 갯수
- 추천 범위는 5~48
- 인덱스 크기에 직접적인 영향을 준다.
- 빌드 시간 및 재현율에 관여

M	색인 빌드 시간	응답시간	재현율
16	317초	56.6 밀리 초	76.1%
32	688초	97.3 밀리 초	92.4%
48	1166초	98.8 밀리 초	96.9%
60	1585초	119.4 밀리 초	98.6%

표1. M값에 따른 성능 변화 표

256 차원 랜덤 벡터 100만개, ef_construction = 500,

K = 1000, ef = 10000

성능 튜닝 - ef_construction

- 색인 구성 시 이웃 노드를 저장하는 동적 큐의 크기
- 재현율과 빌드 속도에 관여

ef_con	색인 빌드 시간	응답시간	재현율
500	688초	88.7 밀리 초	88.1%
1000	1149초	70.8 밀리 초	89.5%
2000	1942초	66.1 밀리 초	91.0%
3000	2681초	59.6 밀리 초	92.1%

표2. ef_construction 값에 따른 성능 변화 표

256 차원 랜덤 벡터 100만개, M = 32,

K = 1000, ef = 7000

성능 튜닝 - ef

- 검색 시 방문한 이웃 노드를 저장해 놓는 동적 큐
- 해당 값 만큼의 노드를 탐색한다.
- 검색 개수인 K보다 커야 함
- 응답 시간과 재현율에 관여

ef	응답 시간	재현율
1000	14.8 밀리 초	53.3 %
3000	41.0 밀리 초	76.1 %
5000	75.8 밀리 초	85.0 %
7000	78.8 밀리 초	89.5 %
10000	103.9 밀리 초	93.4 %









표3. ef 값에 따른 성능 변화 표


256 차원 랜덤 벡터 100만 개, M = 32, ef_con = 1000

K = 1000

RAG와 벡터 DB

대표적인 벡터 데이터 베이스 비교

Listing	Vector Databases	Type
 Weaviate	Weaviate	Managed / Self-hosted vector database / Open source
 Pinecone	Pinecone	Managed vector database / Close source
 Milvus	Milvus	Self-hosted vector database / Open source
 Chroma	Chroma	Buyer-based open source
 Vespa	Vespa	Managed / Self-hosted vector database
 Vald	Vald	Self-hosted / Open Source
 Qdrant	Qdrant	Open Source
 zilliz	Zilliz Cloud	Open Source

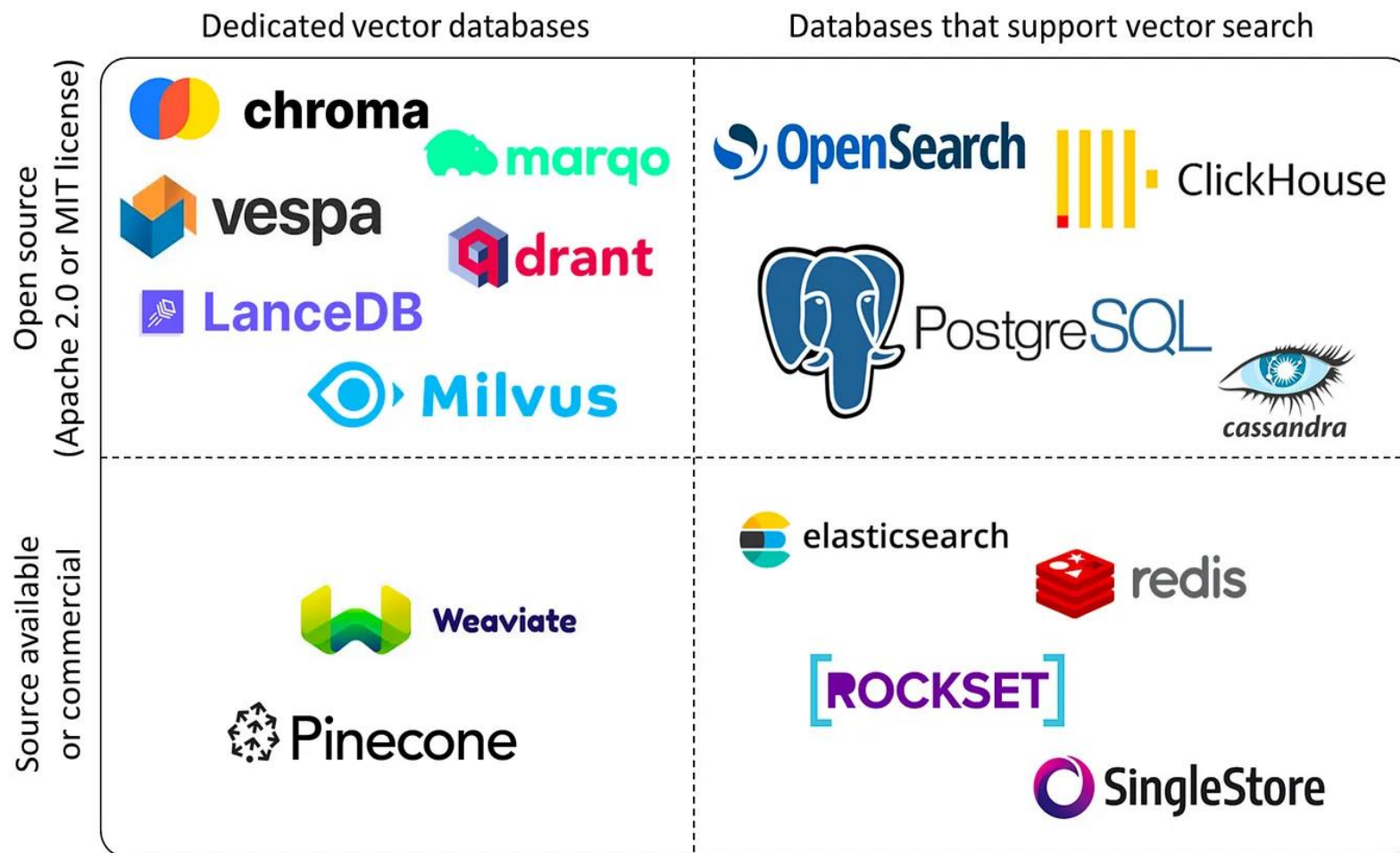
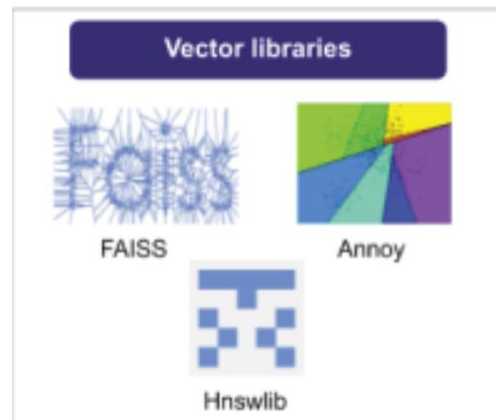
	Released	Billion-scale vector support	Approximate Nearest Neighbor Algorithm	LangChain Integration
Open-Sourced				
Chroma	2022	No	HNSW	
Milvus	2019	Yes	FAISS, ANNOY, HNSW	
Qdrant	2020	No	HNSW	
Redis	2022	No	HNSW	
Weaviate	2016	No	HNSW	
Vespa	2016	Yes	Modified HNSW	
Not Open-Sourced				
Pinecone	2021	Yes	Proprietary	Yes

*Note: the information is collected from public documentation. It is accurate as of May 3, 2023.

<input type="checkbox"/>	Vector Databases	Python API	Java API	Other API's
1	Milvus	Python SDK	Java SDK	Go SDK, N
2	Pinecone	Python SDK	No	No
3	Faiss (Facebook AI Similari...	Python SDK	No	C++
4	Weaviate	Python SDK	Java SDK	Go SDK
5	Qdrant	Python SDK	No	Rust SDK

RAG와 벡터 DB

벡터 데이터 베이스 산업 지형



ELROP (ezis for LLM RAG Orchestration Platform)



| Solution for managing and controlling LLM for optimizing performance

원본 데이터에 대한 chunking, Embedding, vector DB화 부터 LLM 관리까지,
LLM 어플리케이션의 성능을 최적화하기 위한 솔루션입니다. LLM 모델에 대한 효과적인 관리 및 제어가 가능합니다.

주요기능 및 특징



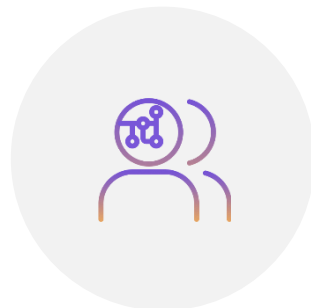
데이터 벡터 DB화

원본데이터에 대한
특징과 출처를 포함하여
벡터DB화 지원



In-Context Learning

환각현상을 최소화해 특정 도메
인에 대한 정교한 답변 제공



LLM 연결

GPT, Gemini와 같은 API부터
Alpaca와 같은 오픈소스 모델까
지 다양한 Foundation Model에
대한 연결 지원



벡터 DB 모니터링

자체 개발 벡터DB 엔진과 자연
어 기반 모니터링 지원

ELROP (ezis for LLM RAG Orchestration Platform)

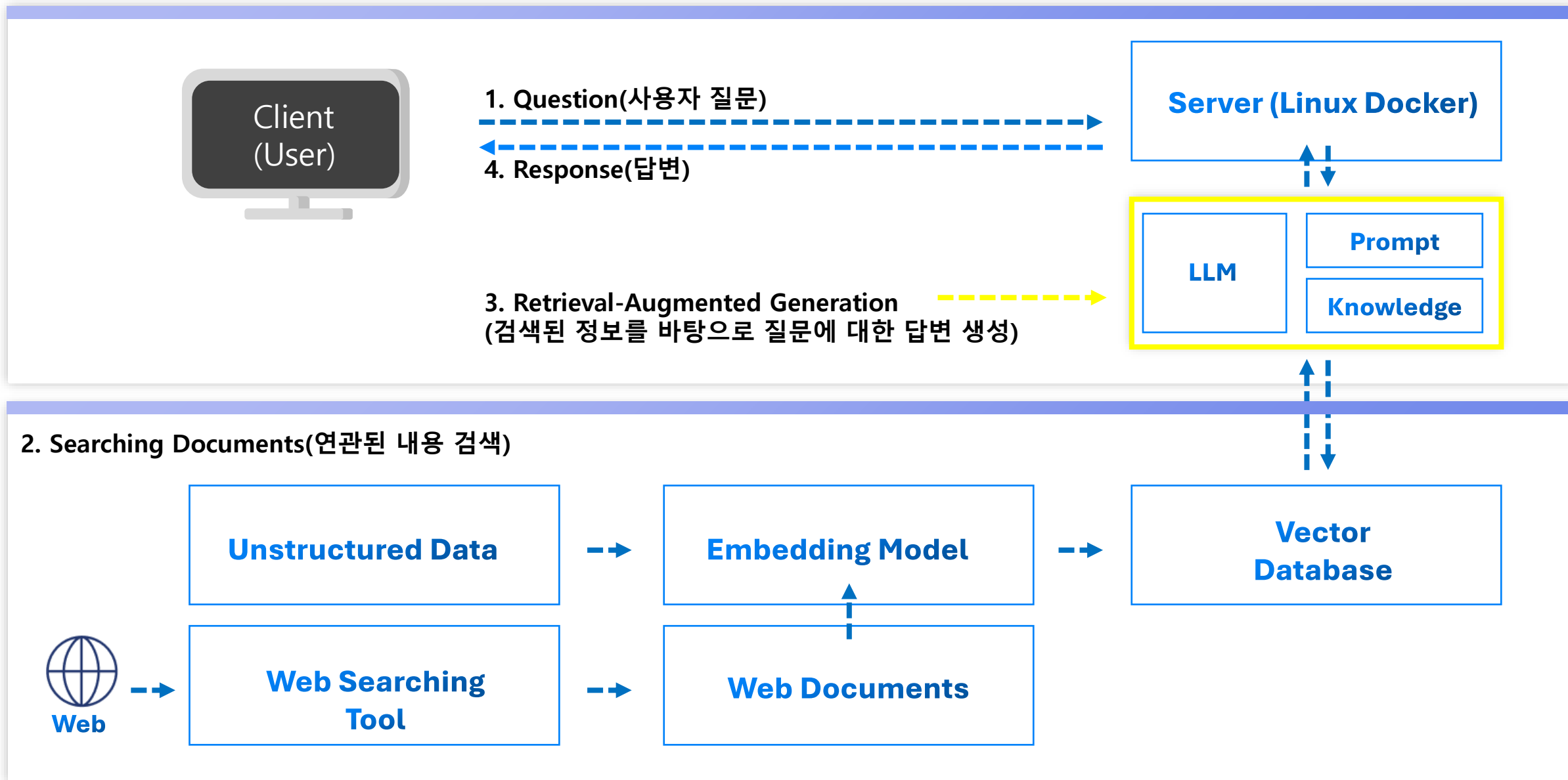


| Solution for managing and controlling LLM for optimizing performance

| 구축효과

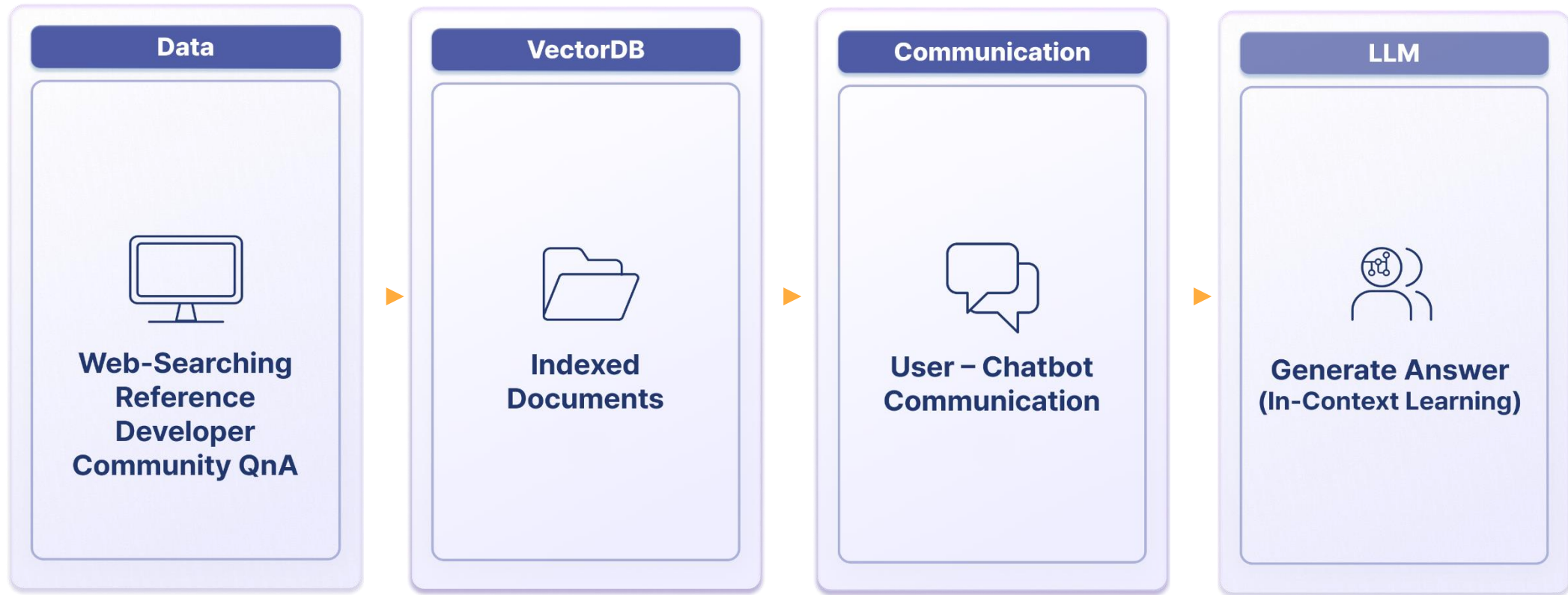
환각현상 최소화	<ul style="list-style-type: none">• RAG(Retrieval Augmentation Generation) 방식을 통한 환각 현상(거짓 정보 생성) 최소화• 답변에 대한 원본 데이터의 출처 표시• 원본 문서의 적절한 chunking, Embedding을 통해 LLM에 정확한 문맥 제공
AI End-to-End 플랫폼	<ul style="list-style-type: none">• 원본 문서 전처리, 벡터DB화, LLM 연결까지 LLM Application의 모든 과정을 한번에 처리할 수 있는 통합 플랫폼• 자체 벡터데이터베이스 엔진 이용을 통한 비용 절감• 사용자 편의성을 위한 자연어 기반 벡터DB 모니터링 기능 지원
고객 맞춤형 레이어 구성	<ul style="list-style-type: none">• 고객의 도메인 맞춤형 생성형 AI Orchestration layer 구성• 오픈형 LLM 구축방식(API 이용)과 폐쇄형(독립 LLM) 구축 방식 등 다양한 구축방식 지원• 답변 요구사항에 맞는 정교한 프롬프트 구성 지원

LLM RAG Orchestration Platform



LLM RAG Architecture

Architecture



Thank you!

