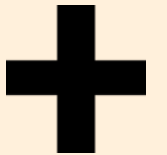




AI 시대, 비즈니스 크리티컬 앱을 위한 개발자 중심 데이터 플랫폼



손광락 | Solutions Engineer





지금까지 Couchbase 는,



Couchone



2011년
합병



Json Database

데이터 모델에
유연한



Memory Database

빠른



2014

어디든 설치되는

Embedded Mobile
Database

2018

하이브리드 검색의



Full Text Search
Database

2021



Couchbase
CAPELLA

Cloud Database

2021



Nasdaq

Listed on Nasdaq Database

2024



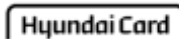
AI-Powered

Vector Database





Couchbase 고객은,



Customers

.....

Application

.....

Performance

Linked in

캐싱 & 싱글뷰를 위한 세션 스토어

2백만+
읽기/초당

1천만
쿼리/초당

TESCO

실시간 가격, 제품, 카탈로그, 재고 관리

1천만+
SKUs

3만5천
요청/초당

amadeus

비행편 가용성, 예약, 가격분석 등

8백만
Ops / 초당

<2.5ms
반응시간

COMCAST

Customer 360 싱글뷰, 'Unified notes' App지원

2.1억개
다큐먼트

10만
사용자

UNITED

리얼타임 승무원 분석, 일정 및 리소스 관리

4.1만
종업원

1.5억+
이용객

Walmart



cars.com



Fanatics



Office DEPOT



STAPLES



TESCO



QVC

BUY

Marriott



amadeus



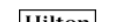
Sabre



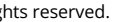
ORBITZ



tripadvisor



Hilton



FICO



Revolut



PayPal



WELLS FARGO



VISA



Pfizer



everyday HEALTH



BD



Cochlear



MONSENSE



COMCAST



KDDI



at&t



Viber



Telefonica



sky



THOMSON REUTERS



FANDANGO



GANNETT



Disney



nielsen



EA



betfair



zynga



BLIZZARD



Playtika



SCN



verizon



SIERRA WIRELESS



CHAMBERLAIN



GE



GE



3 of the Top 10
eCommerce
Companies

3 of the Top 3
GDS
Companies

3 of the Top 3
Credit Reporting
Companies

3 Fortune 500
Healthcare
Companies

6 of the Top 10
Broadcast Companies

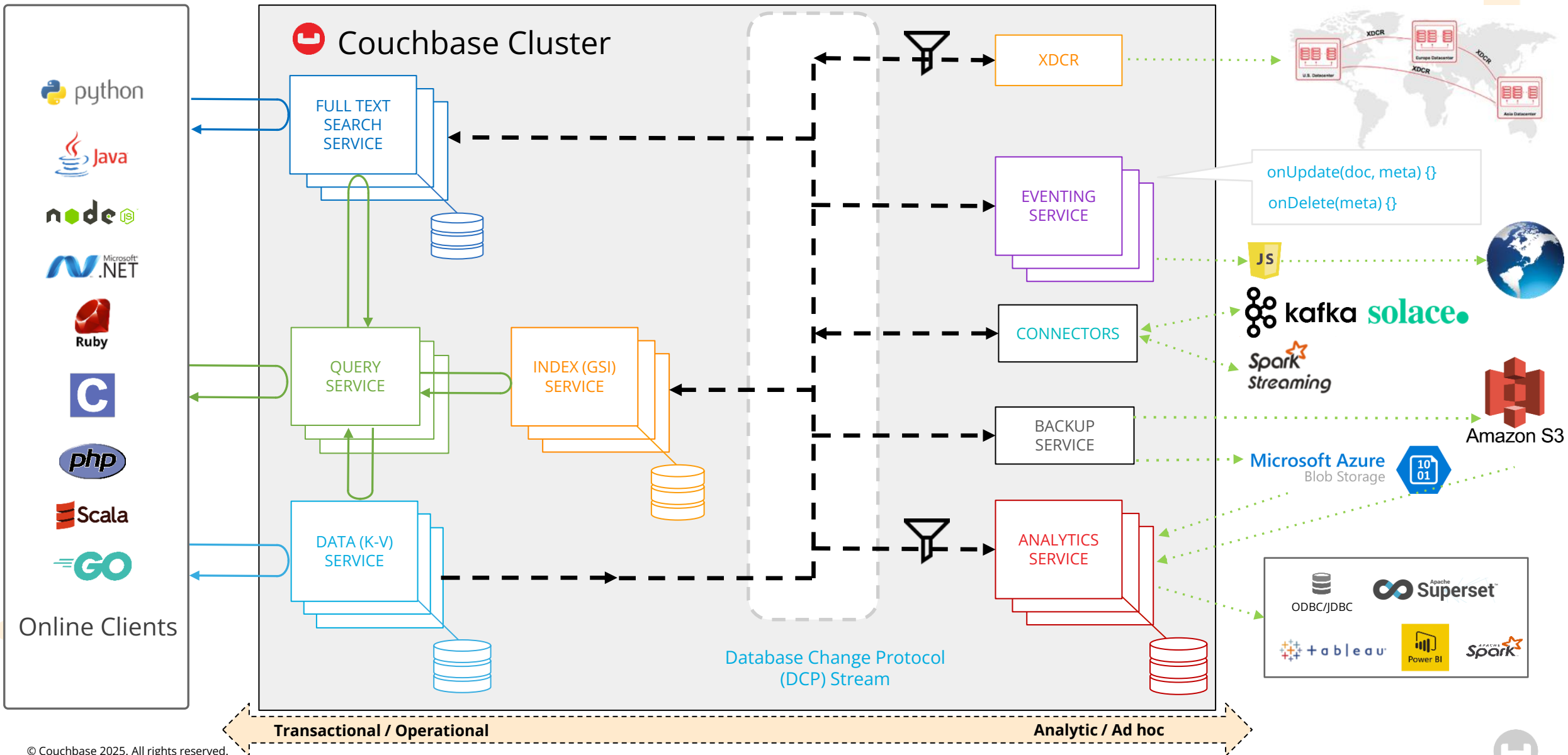
6 of the Top 10
Online Casino
Gaming Companies

2 of the Top 2
IoT
Platforms





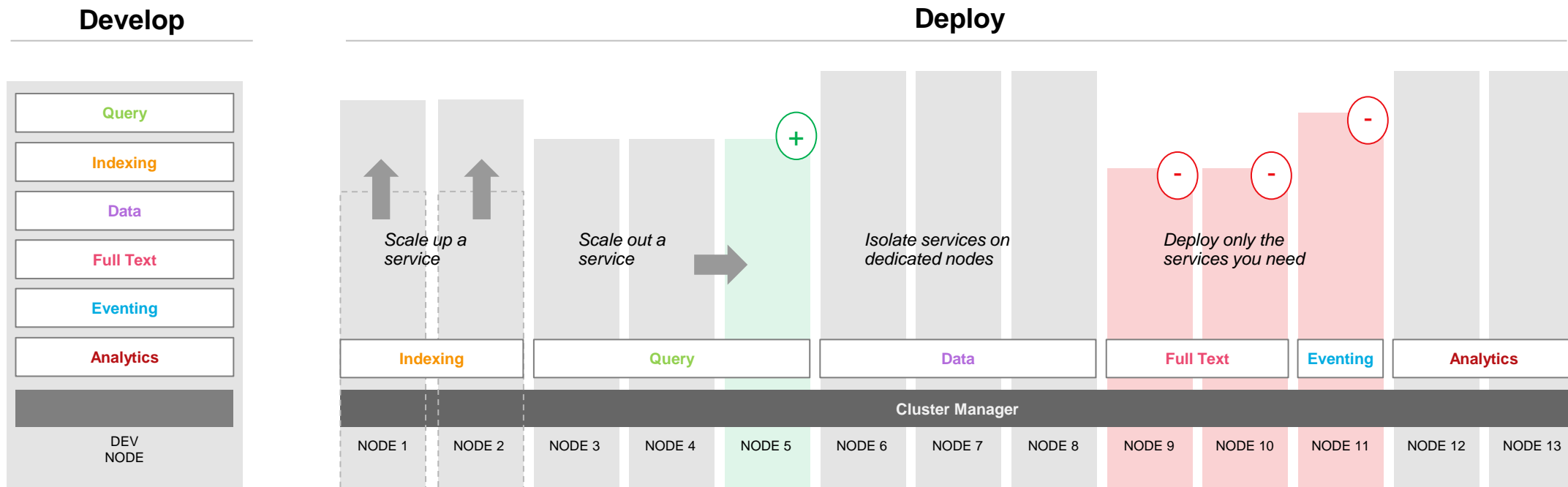
Couchbase { Memory based Micro Service Architecture





Couchbase { 자원 절약형 Multi-Dimensional Scaling Architecture

Couchbase는 Multi-Dimensional Scaling 기능으로 서비스 별 Workload 분산 및 독립성을 보장합니다.



Service 단위 하드웨어 자원 최적화

각 Service 별 시스템 자원을 독립적으로 할당

각 **Service**에서 수행되는 작업이 다른 **Service**에 영향을 최소화, 예를 들어 Analytics Service에서 복잡한 작업을 수행하여 시스템 자원을 많이 사용해도 Data Service 혹은 Query Service에서 수행하는 Operational 작업에는 영향이 없는 구조





Couchbase { NoSQL Database 벤치마크에서 탁월한 성능 우위

<Yahoo! Cloud Serving Benchmark(YCSB-NoSQL Benchmark)>



MongoDB

- Struggles to scale, strongest at 3 nodes
- Price performance makes it worse
- Weakest performer

DynamoDB

- Excels at scans, but throws excessive errors
- Challenged across other workloads

Redis

- Excels at read-only, Capella meets at scale
- Regularly used as cache for Atlas or DynamoDB
- “Fails” Pagination workload

Capella(Couchbase)

- Excels at Read and Write
- Scales effectively for multiple workloads
- Best price-performance

	Couchbase Capella	MongoDB Atlas	Azure CosmosDB (MongoDB API)
	90 /116	84 /116	71 /116
DEPLOYMENT	28 /32	27 /32	21 /32
MANAGEMENT	19 /25	19 /25	19 /25
SUPPORT	11 /16	12 /16	9 /16
PERFORMANCE	20 /26	13 /26	17 /26
PRICING	12 /17	13 /17	5 /17

출처 : <https://www.altoros.com/blog/couchbase-capella-vs-mongodb-atlas-vs-amazon-dynamodb-vs-redis-enterprise-cloud/>

Altoros, April 2023





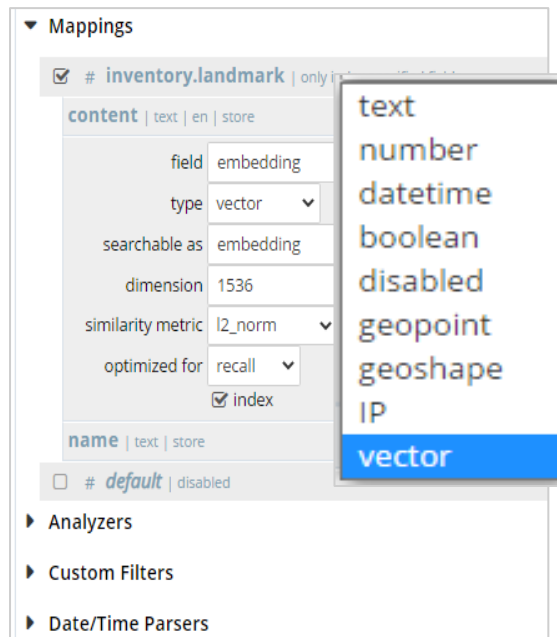
Couchbase { Hybrid Query Support

SQL 쿼리에 Full Text 검색 쿼리와 벡터 검색 쿼리를 하나로 수행

JSON Document

```
{ "type": "shoes",
  "productId": "CP123456",
  "category": "Gym Shoes",
  "name": "Beach Sneakers",
  "brand": "Ultimate Surf",
},
  "description": "The ultimate companion for beach adventurers, designed to seamlessly transition from sandy shores to urban landscapes. This innovative sneaker features a water-resistant, quick-drying mesh upper, allowing your feet to breathe while keeping them dry.",
  "descriptionVector": [0.131, 0.339, -0.611, 0.981,...]
```

Indexes



SQL + 자연어 검색 + AI Vector Search

```
SELECT meta().id, t.name
FROM `product` as t
WHERE
SEARCH(t,
  {"query":{"match_phrase": "sneaker", "field":
"description"}}
)
ORDER BY search_score() DESC
LIMIT 10;
```

```
SELECT *
FROM product
WHERE LOWER(product.type) = 'shoes'
AND product.size = 11
AND product.price between 50 and 80
/* desc SIMILAR TO 'blue running shoes' */
ORDER BY GSI_VECTOR_ORDER(desc_embedding, {
  "knn": [{
    "field": "desc_embedding",
    "vector": [0.1, 0.334, -0.604, 0.985] }]})
LIMIT 4
```

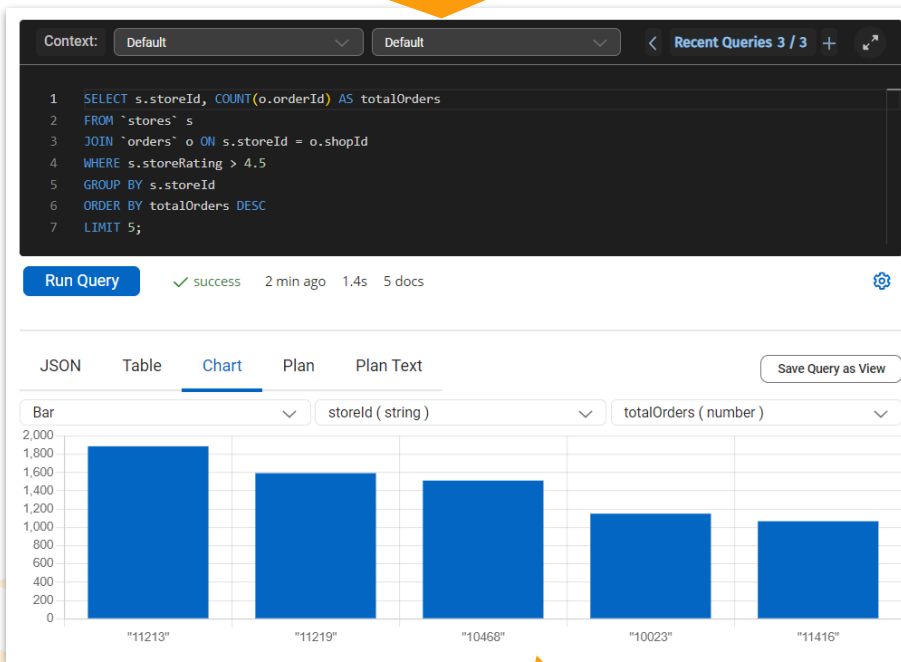




Couchbase { GenAI Based Query/Coding Support

- Generative AI의 LLM을 활용한 Couchbase Capella 전용 Code Assistant
- 자연어로 SQL 및 소스 코딩 지원
- Couchbase 내부 스키마 정보를 활용하여 실제적인 코딩 지원

iQ works across transactional and analytical services



Automatic chart generation

History iQ

Recent Chats (3/3)

Request analysis

Field relationships

Provide any information about the relationships between different fields across collections (print)

Calculate the total orders for stores with a storeRating higher than 4.5, show the top 5

iQ

```

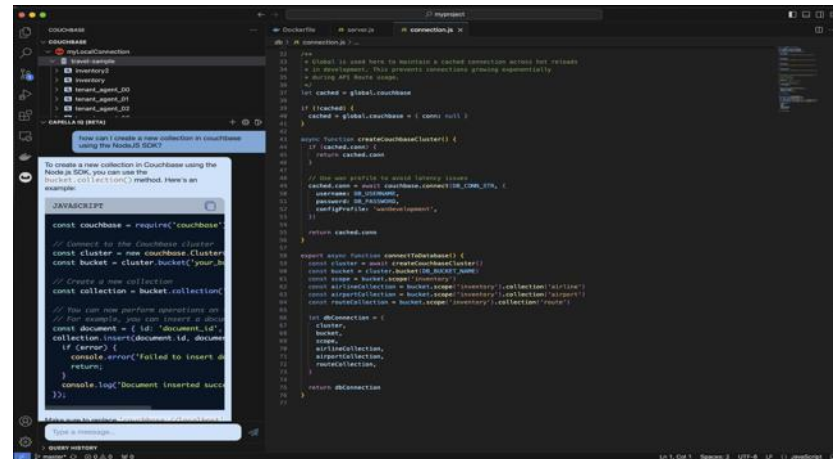
SELECT s.storeId, COUNT(o.orderId) AS
totalOrders
FROM `stores` s
JOIN `orders` o ON s.storeId = o.shopId
WHERE s.storeRating > 4.5
GROUP BY s.storeId
ORDER BY totalOrders DESC
LIMIT 5;

```

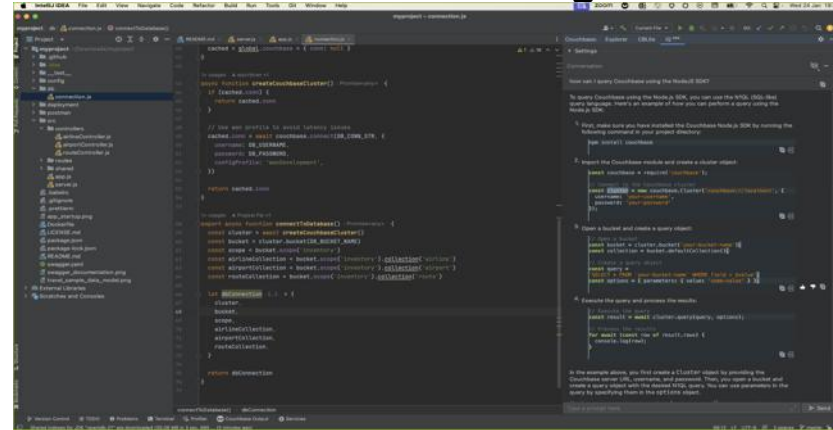
Run Copy Chart Favorite

Iterate on the generated query...

Visual Studio Code Plug-in



IntelliJ IDEA Plug-in



Couchbase { RAG GenAI Support

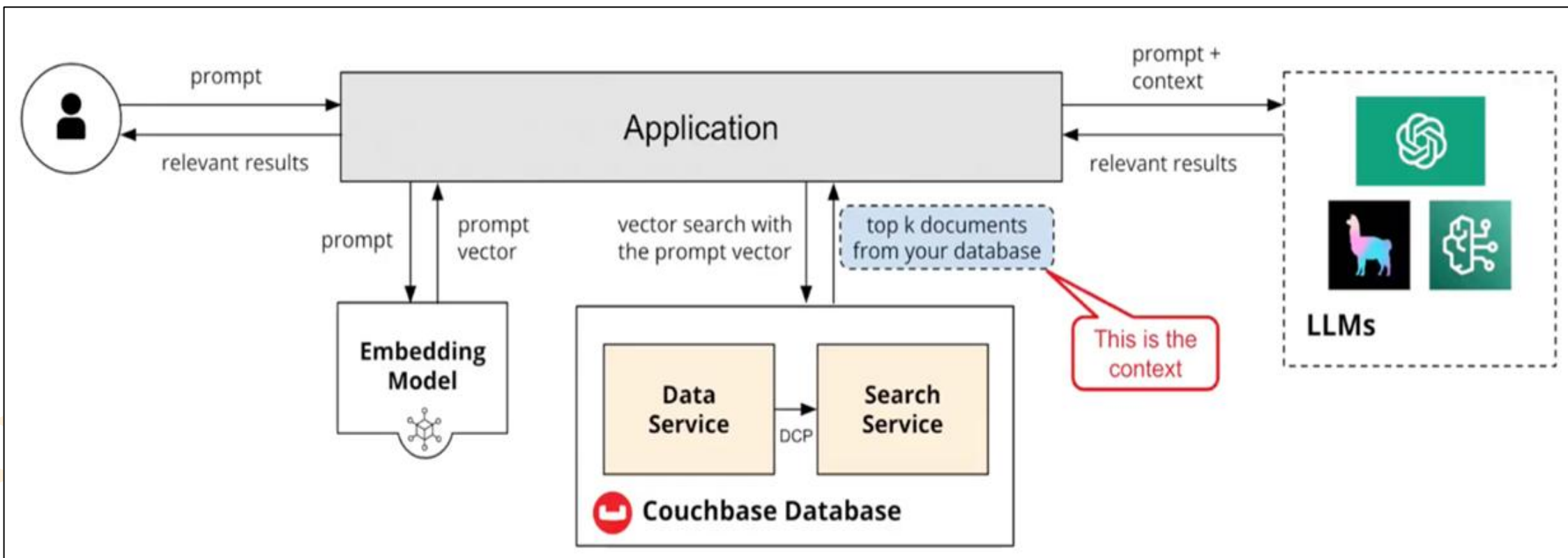
RAG은 내부 지식(문서)를 기반으로 하여 LLM의 환각(Hallucinations)의 위험을 감소시킬 수 있음.

Content Generation

Data Analysis: Classification / Anomalies

Advanced Semantic / Hybrid Search

AI-powered Chatbots and Applications



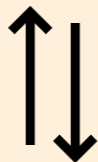


앞으로의 개발 환경은, **New Paradigm**



GenAI 이전 패러다임
GenAI 이전의 패러다임은
간단했습니다

Today



- UI 중심: 사람이 직접 데이터를 입력 및 관리
- 워크플로우: 전환 비용이 높은 사람 중심의 수동 워크플로우

AI가 더해진 새로운 패러다임(AI 에이전트)으로 인해 상호작용이 3배 더 복잡해 집니다!

Tomorrow



[RAG]



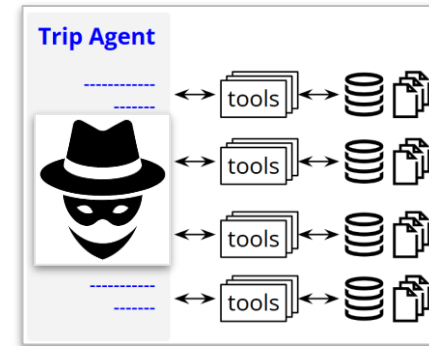
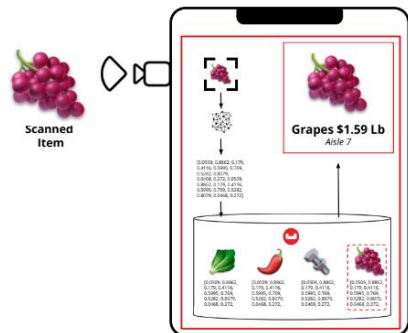
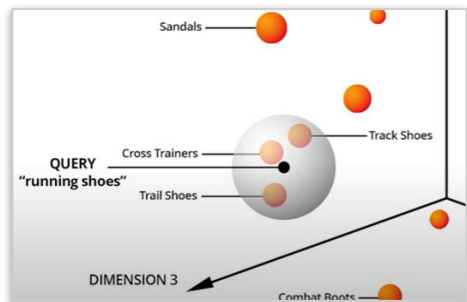
- 자연어 기반 UX: AI 에이전트가 대규모로 실시간으로 비정형 데이터를 처리합니다.
- 워크플로우: AI는 데이터를 캡처할 뿐만 아니라 워크플로우도 생성합니다.

다양한 데이터의 자율적인 수집, 저장, 처리로 초점 이동





Vector Search and Advancing AI Use Cases



Vector Search

Hybrid Search

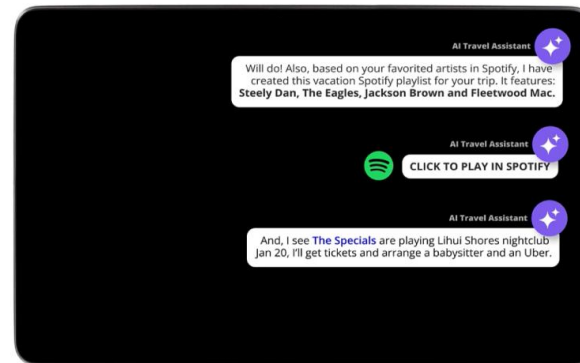
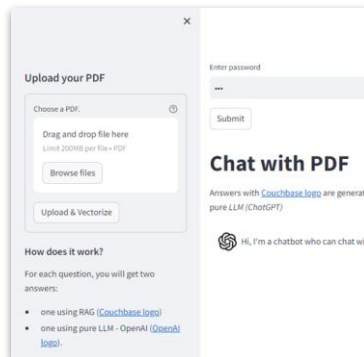
On Device Image Vector Search

RAG with Unstructured Data

Simple RAG Chatbot

Agentic Chatbot

Agentic Applications





에이전트 AI를 통한 전 산업 워크플로우 자동화가 예상



Travel & Hospitality



Retail Banking



Media & Entertainment



Gaming



Teleco



Manufacturing



AGENTIC AI

Healthcare



Government



Retail eCommerce



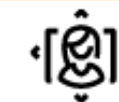
Financial Services



Energy



High Tech





유연한 확장성과 이를 뒷받침 해줄 성능이 중요한 이유는 무엇인가요?

데이터 소스, 형식 및 통합이 크게 증가할 것입니다.



사용자 프로필

(사용자 기본 설정, 과거 여행, 즐겨 찾는 활동)
JSON, XML
NoSQL



기업 데이터

조직을 고유하고 가치 있게 만드는 독점 데이터
JSON, XML



날씨 정보

실시간 및 예보 날씨
JSON, XML
시계열 DB



여행 제한 사항

비자 요건, 코로나19 제한 사항, 안전
JSON, CSV



리뷰 및 추천

상품, 호텔, 여행지에 대한 사용자 리뷰
JSON, 텍스트



현지 언어 및 번역

목적지에 대한 언어 번역 데이터
JSON, XML



상담원 기록 및 아티팩트

버전이 지정된 프롬프트, 응답, 메타데이터, 유효성 검사 기준, 쿼리, 응답, 세션 데이터
텍스트, JSON
에이전트



멀티미디어 데이터

목적지의 이미지, 동영상
JPEG, PNG, MP4
개체 데이터



항공편 정보

실시간 항공편 예약 가능 여부, 가격, 스케줄
JSON, CSV



교통 정보

렌터카, 대중교통 스케줄
JSON, CSV
RDBMS



목적지 정보

관심 장소, 활동, 지역 이벤트
JSON, 텍스트
그래프 DB



환율 정보

실시간 통화 전환율
JSON, CSV
시계열



예약 및 거래

예약 세부 정보, 결제 내역
PDF, JSON, CSV



맞춤형 제안

AI 기반 여행 추천
JSON, CSV



여행 여정

여행 일정, 업데이트
JSON, XML

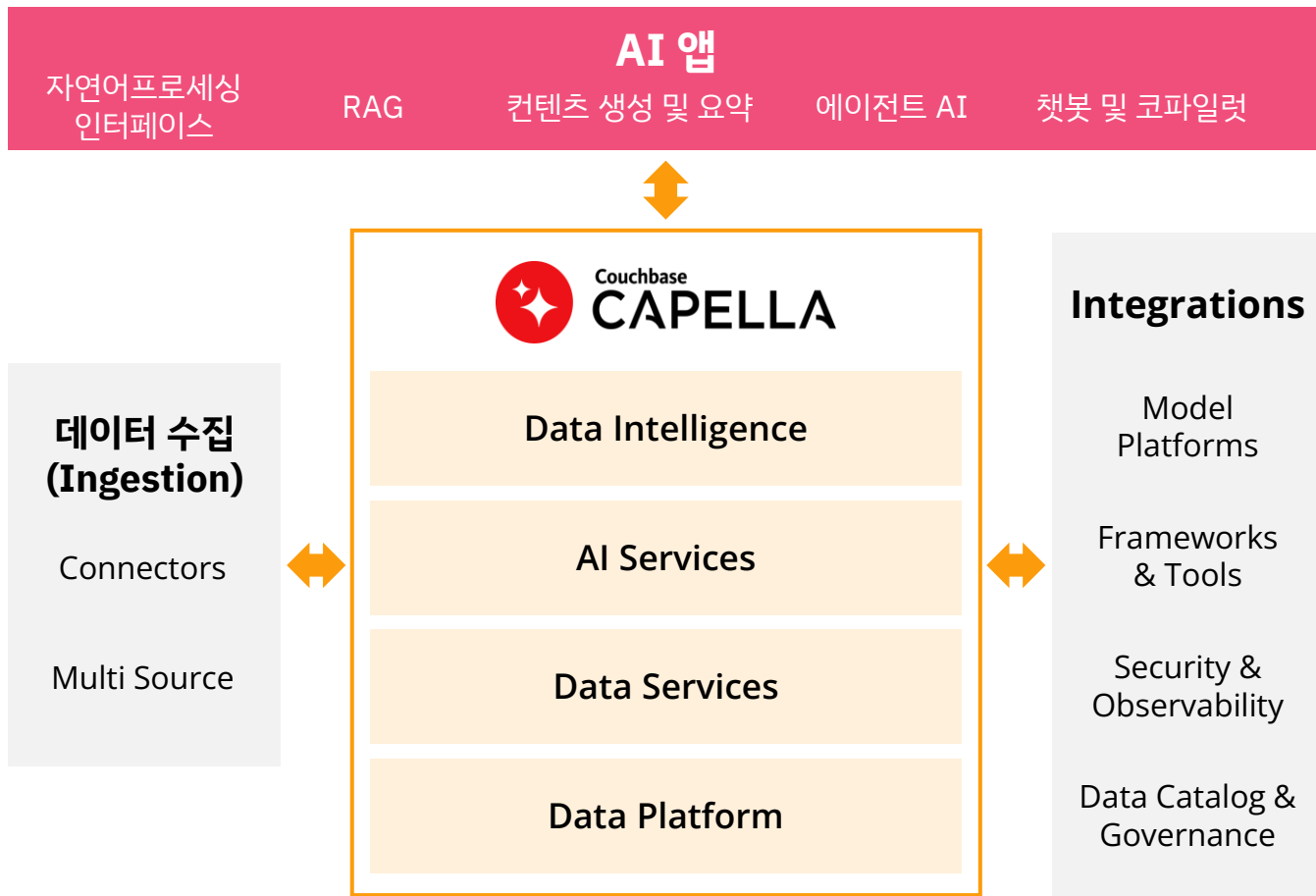




AI 혁신(“AX”)을 위한 데이터 플랫폼의 요구 사항

요구 사항

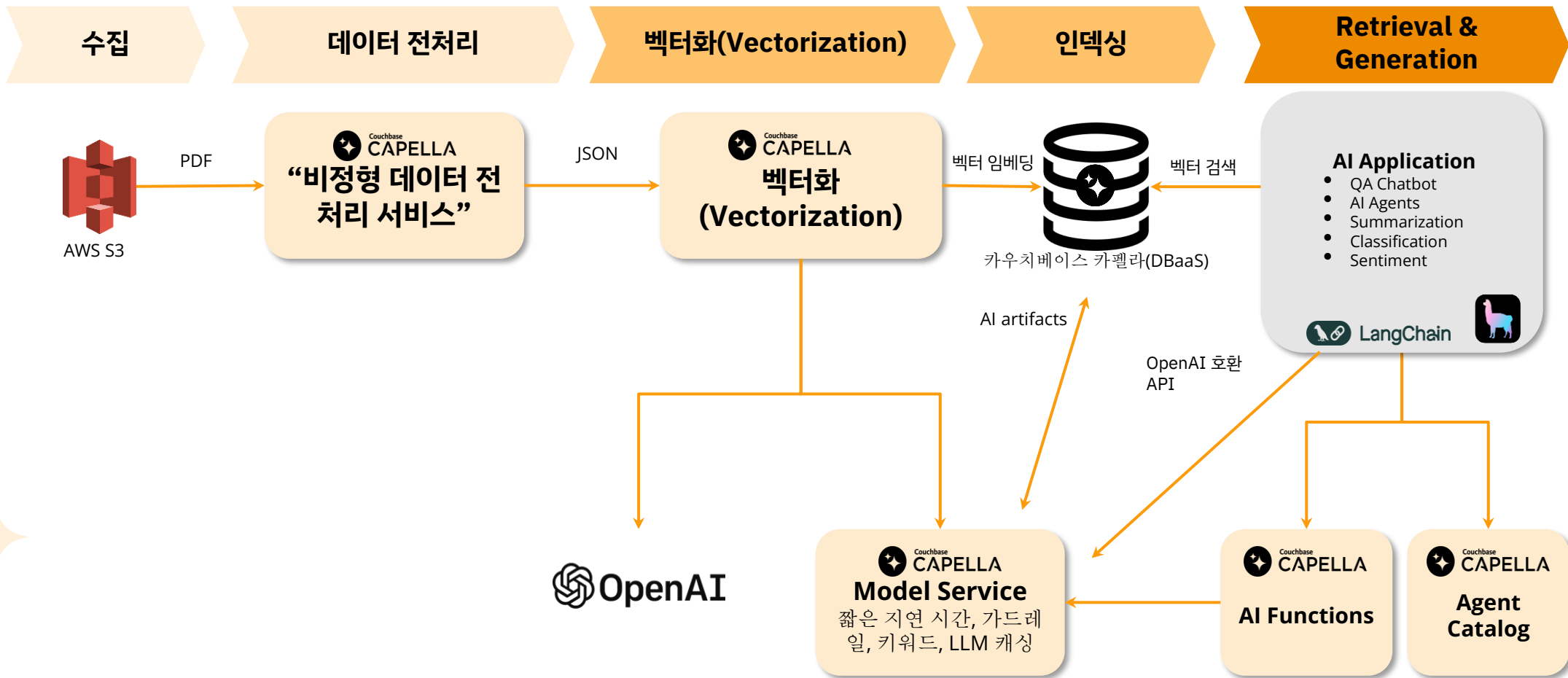
통합 및 단일화된 플랫폼	통합 데이터 관리	✓
	확장성 및 성능	✓
	고급 실시간 분석	✓
	강력한 통합	✓
	온/오프라인 어디든 배포	✓
	성능으로 비용 절감	✓
	안전, 보안, 책임감	✓
	네이티브 AI 기능	✓





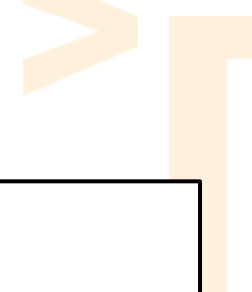
카우치베이스 { AI 서비스로 RAG 앱을 한층 더 강화

개발자가 구조화된 데이터와 구조화되지 않은 데이터를 제어하고, 워크플로를 간소화하고, 단일 보안 플랫폼에서 AI 모델을 사용하여 에이전트 RAG 애플리케이션을 구축할 수 있도록 지원.





데모 동영상 { AI Services Demo Link





Example { Agentic Customer 360

Trusted Data

Trusted AI

GenAI Augmented Cust360

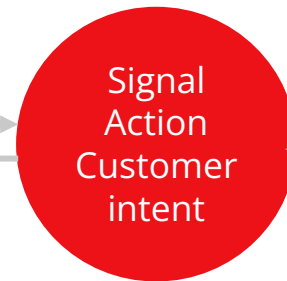
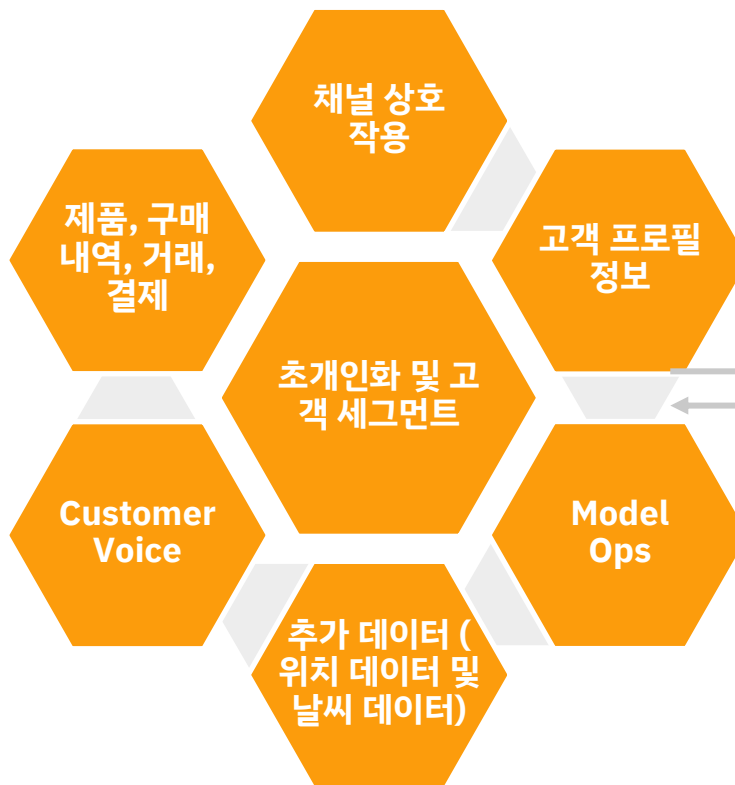
Trusted Signal

Execution
Orchestration

채팅
음성
텍스트
트랜스크립트
문서
이메일
소셜 미디어
프로필

감정 분석
토픽 모델링
분류
클러스터링
요약
음성 분석

Trusted AI: 모델 적용에 앞서
Trusted AI 기반 데이터 합성 선
행 필요





Couchbase { Edge AI [Lite, Sync Gateway, EdgeServer]

Purchase Items



Scan and purchase goods using a mobile app

Price Lookup



In-store kiosks for price and related information for goods

Insurance Claims



Assess images of damage and recommend replacement parts and service

Self Check In



Scan a pass or use facial recognition for instant check-in

Smart Mirror



In-store mirror device shows user in different outfits and accessories

Smart City



Use camera images to assess street conditions and recommend services

Security Surveillance



Visually detect and identify possible security issues, find missing persons, etc.

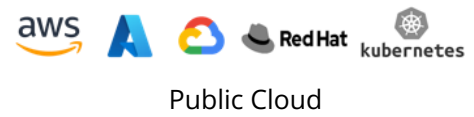
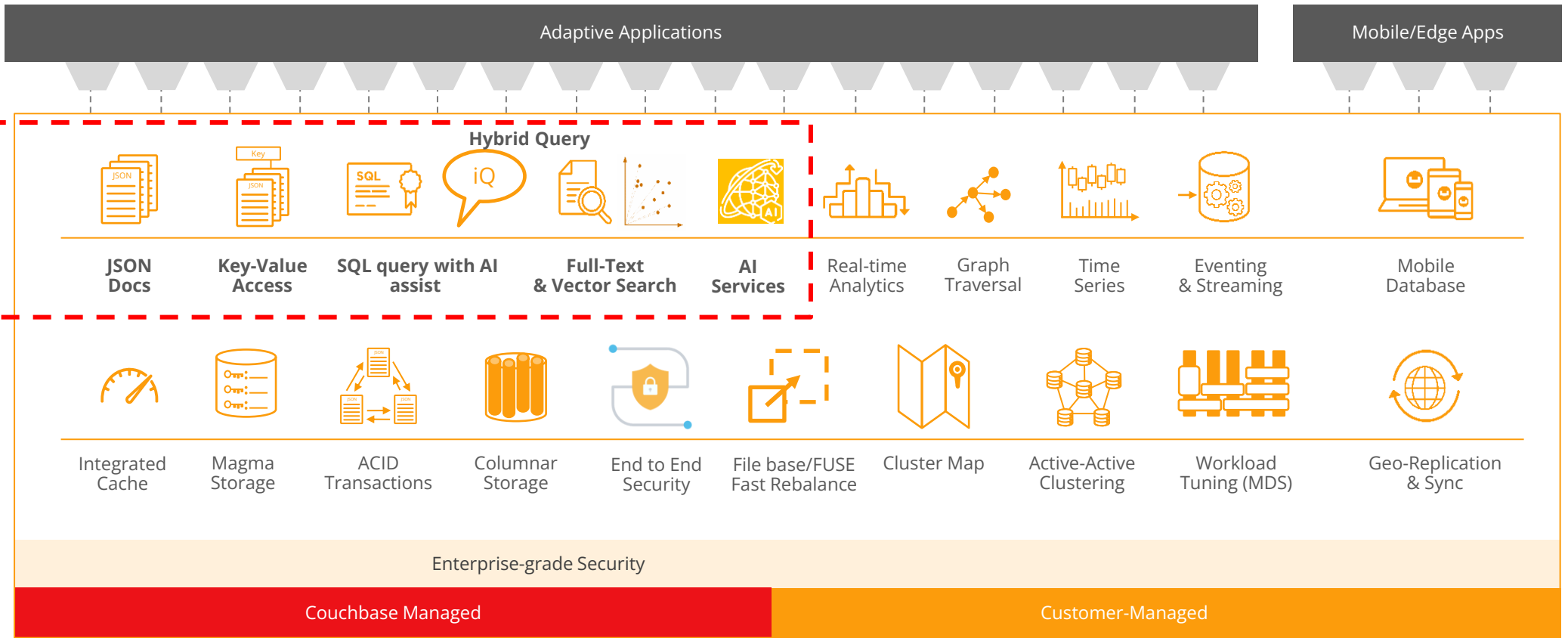
Defense



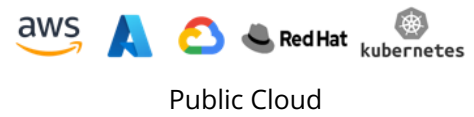
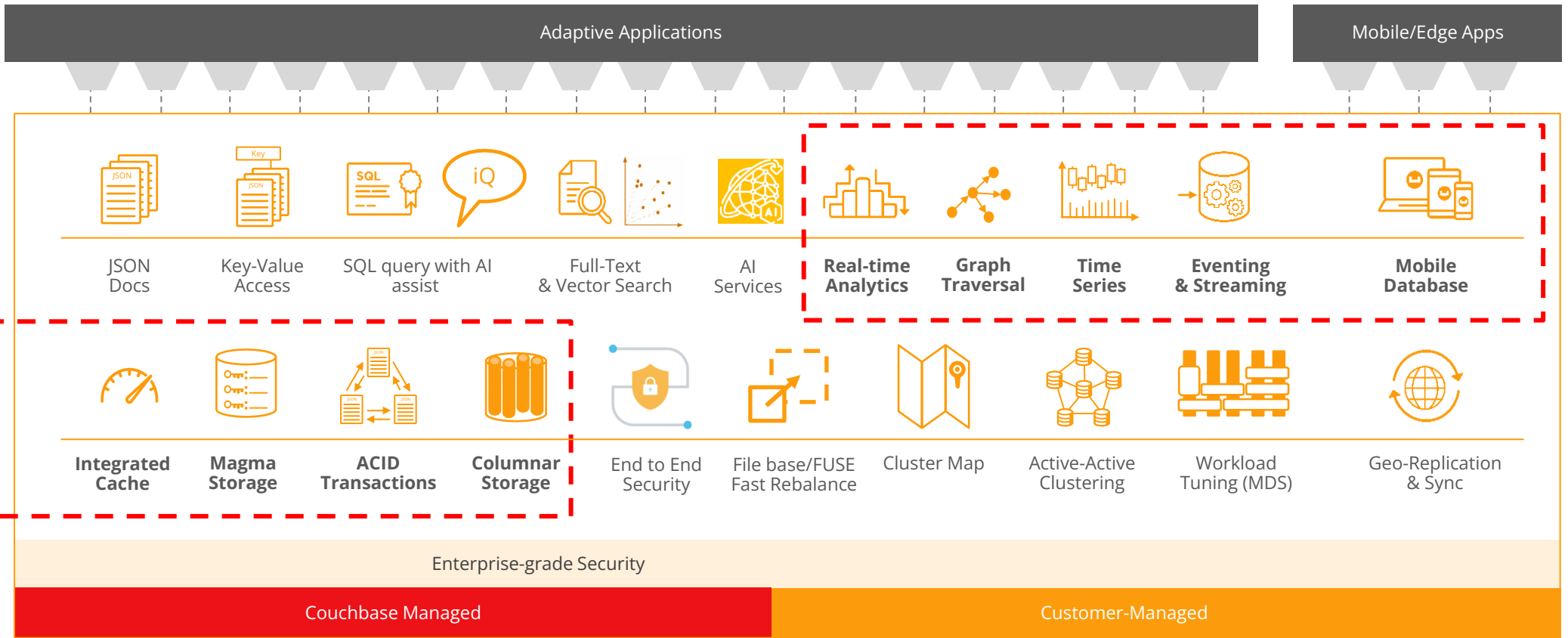
Cameras on drones enable autonomous threat detection



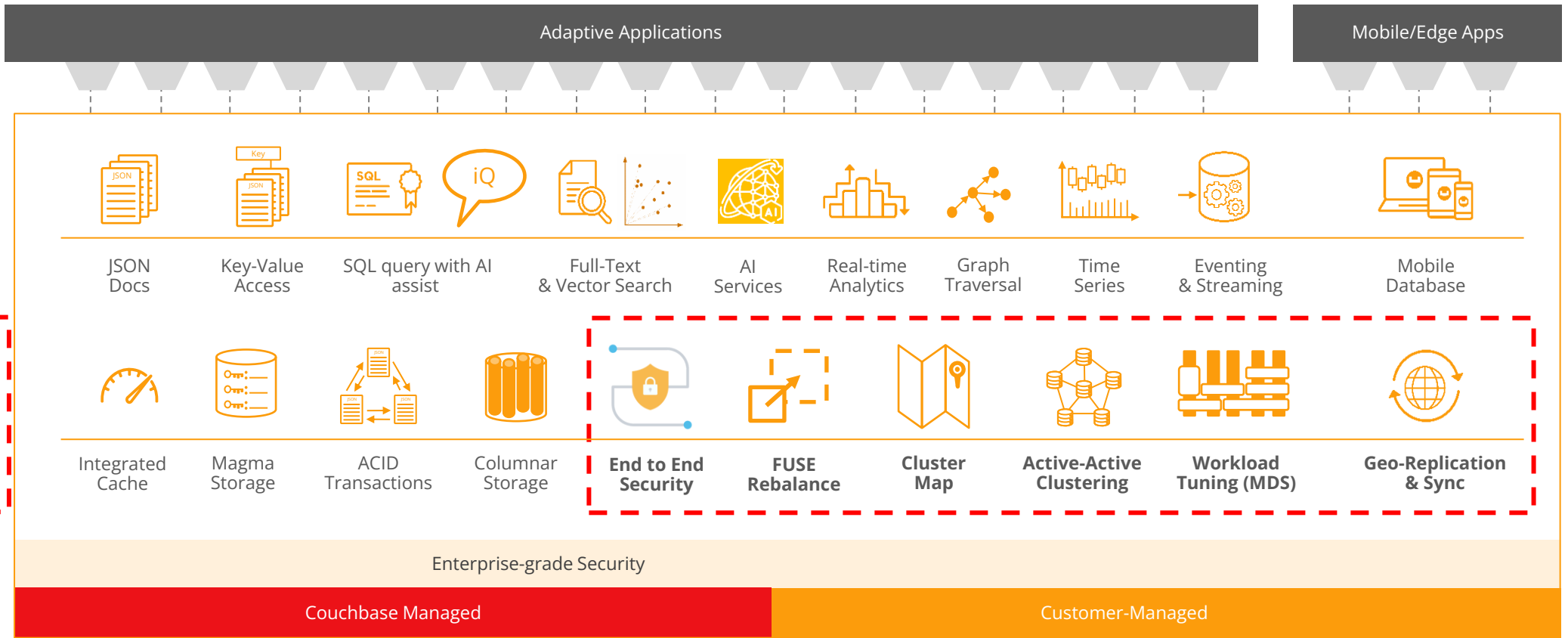
The **Developer** Data Platform for Critical Applications in Our AI World



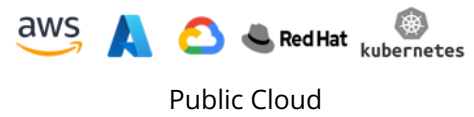
The Developer **Data Platform** for Critical Applications in Our AI World



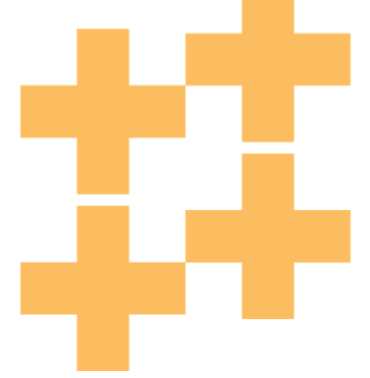
The Developer Data Platform for **Critical Applications** in Our AI World



Enterprise Operations

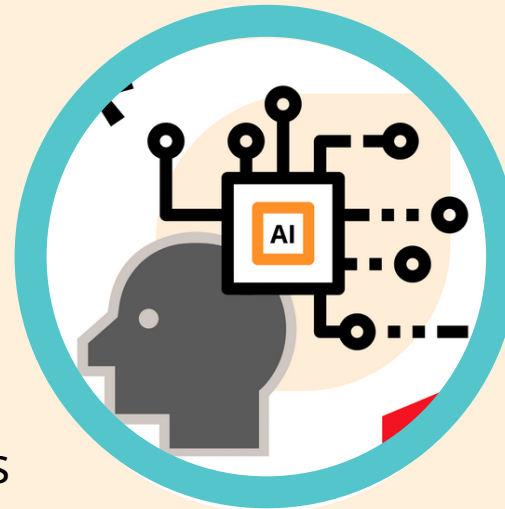


The Developer Data Platform for Critical Applications in Our **AI World**



Part of Everything

- + AI built into and throughout our platform
- + Self-hosted LLM, SLM Agentic applications
- + Attach to emerging AgentOps best practices



FICO[®]

SWARM
ENGINEERING

tando

Seenit



앞으로의 Couchbase는,

- Lightweight Edge Server
- CC Billing
- Enhance Marketplace
- Data API
- Rust SDK
- CB Lite Web Support
- Data Streaming API

Developer Data Platform



Critical Applications



- Fine-Grain RBAC
- Encryption-at-rest
- Query Workload Snapshots
- App Telemetry
- Auto-Scaling
- Continuous Backup and PITR
- XDCR/Mobile Compatibility
- Sync Gateway Optimizations



Artificial Intelligence

- AI Services
- NVIDIA NIM and NeMo
- 1B+ Scale Vector Search
- Columnar Vector Search
- Feature Stores
- UX and Agentic Experiences





Couchbase

The Developer Data Platform
for Critical Applications
in Our AI World



paul.son@couchbase.com
www.couchbase.com
cloud.couchbase.com