



MongoDB 세션

초개인화 시대 데이터가 답이다.

Data Enrichment Personalization at MongoDB



김 준 상무, MongoDB
Senior Solutions Architect,
Korea Technical team Lead

나도 모르는 내 취향까지 읽는다... 초개인화, 끝 어딘가



“지금내 감정 원하는 것” 답러닝... 고객성향 실시간 파악해 서비스

입력 2020-09-26 04:04

경영환경

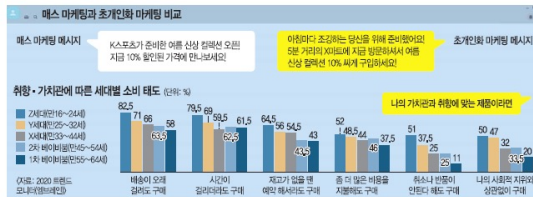
초(超)개인화, 0.1명 단위로 세그멘테이션 하다

2020-02-26 · 권세환 · 조회수 4492



‘취향 저격’ 넷플릭스의 비밀

초개인화가 등장하게 된 배경엔 ‘생산자에서 소비자 중심으로’ 시장의 패러다임 변화가 있다. KB금융경영연구소 권세환 책임연구원은 최근 발표한 보고서에서 “자신의 욕구를 희생했던 예전의 소극적인 소비자들과 달리 자신에게 최적화된 제품과 서비스를 적극 요구하는 소비자들의 등장으로 초개인화 경계가 촉발됐다”고 분석했다. ‘소품종 대량생산’의 획일적 소비 패턴이 저물고 다양한 소비자 기호에 맞춘 ‘다품종 소량 생산’ 시대에 들어섰다는 것이다.



초개인화 서비스는 넷플릭스를 빼놓고 설명하기 힘들다. ‘취향 저격’이라는 수식어를 달고 다니는 넷플릭스는 콘텐츠 장르를 7만6000여 개로 세분화했다. 고객 선호도를 2000개 유형으로 분류해 맞춤형 추천 콘텐츠를 선보이고 있다. 특히 소비자들이 넷플릭스의 영화 추천 알고리즘이 ‘시네매치’가 추천해 준 작품을 선택하는 비율이 전체 콘텐츠의 75~80%에 달한다. 최근에는 사용자의 선호도나 감정까지 분석하는 ‘답러닝’기술까지 도입했다.

이밖에 구글의 AI스피커 ‘구글 홈은 “음악을 들려줘”라는 동일한 명령에도 가족 목소리를 각각 따로 인식해 각자의 취향에 맞는 음악을 틀어준다. 세계 시장 점유율 1위의 무료 음악 스트리밍 서비스 업체인 ‘스포티파이’는 사용자의 음악 선호도를 정확히 분석해 내는데 탁월하다. 기존 사용자의 청취 습관(위치, 시간, 선택 음악, 건너뛴 음악)을 분석하고 비슷한 취향을 가진 타인의 데이터를 혼용한 알고리즘이 ‘1등 서비스’의 비결이다.

1. 초 개인화(hyper-personalization)

필자는 마케팅의 전략도 트렌드를 따라다니기 보다는 소처럼 다소 우직하게 기존의 전략 **이행과 업그레이드**가 더욱 중요하다고 본다. **2021년에**는 기존의 개인화가 더 진화되어 초개인화로 다가가 듯하다. 우리가 이미 알고 있는 개인화는 단순한 구매 고객에게 이름을 부르는 것이 아니라, **소셜** 활동 데이터와 니즈를 파악해 최대한의 친함을 끌어내는 것을 말한다.

그러나 최근에는 ‘개인화(personalization)’를 넘어서 ‘**초개인화(hyper-personalization)**’라는 말이 대두되고 있다. 개인화가 ‘고객 정보 및 온라인 행동 데이터를 기반으로 사용자를 이해하고, 이를 바탕으로 사용자의 경험을 최적화하는 것’이라면, 초개인화는 여기서 한 걸음 더 나아가, ‘**온라인 데이터뿐 아니라 실제 생활 패턴 및 취향 정보를 바탕으로 상품과 타이밍에 적절한 메시지로 사용자의 소비 경험을 가이드 하는 것**’이라고 설명할 수 있다.

초개인화는 소비자의 취향을 분석해 상품, 서비스 등을 예측해서 제공하기도 하고, 소비자의 취향을 반영한 커스터마이징 서비스 등 개인의 개성이 가장 중심된다. 즉 초개인화는 **실시간으로 소비자의 맥락과 현재 처한 상황을 이해해 궁극적으로는 고객이 필요한 부분을 예측해서 개인에게 정확히** 맞춘 서비스와 상품을 제공하는 기술이다.

우리가 그동안 경험했던 기존 고객 마케팅 활동이 시장 세분화(Segmentation)를 통해 타겟팅 그룹의 보편적 선호도 파악을 중시하였다면, 초개인화는 ‘**특정 고객이 현재 원하는 구체적인 혜택이 무엇인가**’를 파악하게 된다. 즉 고객이 원하는 **혜택이 무엇이고, 어느 시점에 무엇을 해야 하는지** 고객 관점에서 세밀하게 분석하여 고객과의 연결 고리를 강화하는 것이 초개인화의 핵심이다.

또한 ‘개인화’가 이름, 나이, 성별 등 인구 통계학적 정보와 온라인 행동 데이터를 바탕으로 개인의 특성에 집중한 서비스라면, ‘초개인화’는 개인의 **특성뿐만 아니라 소비자가 처한 상황과 맥락까지도** 중시한다. 예를 들어 평소 스템으로 취급되던 푸시 알림 메시지가 초개인화 기술을 통해 정교화된 메시지로 구체화하여 최적의 순간에 전달된다면 고객이 수용하는 정보의 가치는 달라질 수 있는 것이다.



개인화 (Personalization)

Analytical Personalization

ML/AI 기반의 기술을 통한 workloads 분석을 통한
개인화 서비스 제공

 고객 행동 기반의 Single View

 개인화된 타겟 광고

 맞춤형 제안

 채팅 봇을 통한
맞춤형 메시지

 제품 추천

Digitized Personalization

디지털 기술의 차별을 통한 개인화 서비스의 혁신을
제공/경험

 얼굴인식

 GPS 추적

 AR/VR

 위치 인식

 생체 인식 센서

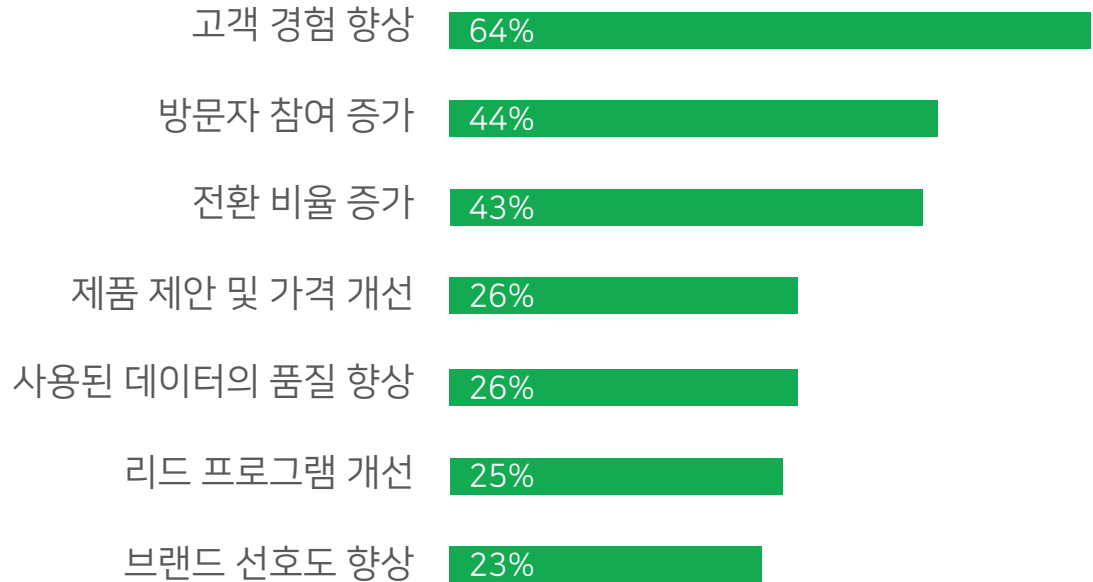


데이터 기반(Data-Driven)의 개인화

달성해야 할 데이터

기반의 개인화

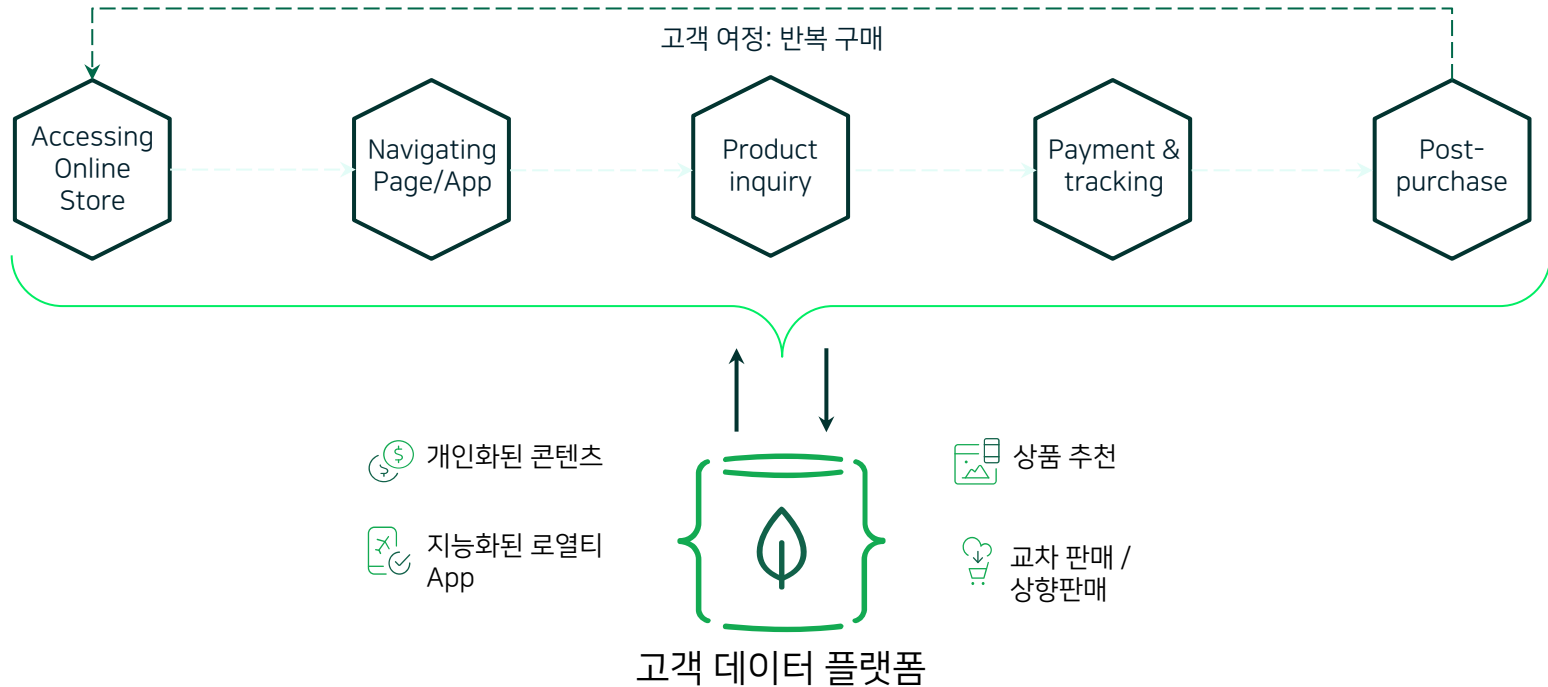
전략의 목표는 ?





개인화 서비스 예

실시간 데이터를 사용하여 고객의 전체 여정을 지원하고 고객 당 평균 수익 증대 모델

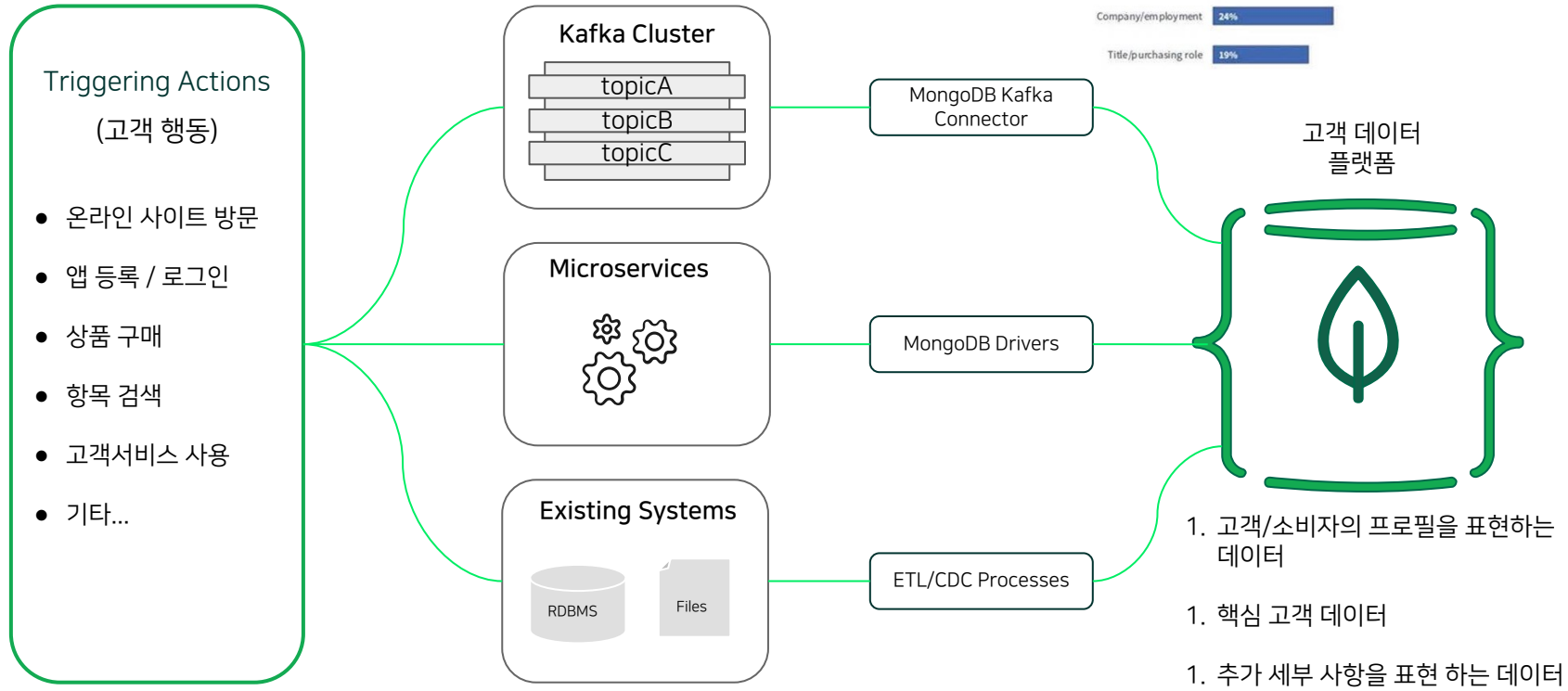


What are the most important types of MARKETING DATA utilized for personalization?



데이터 수집

고객 프로필 생성을 위한 데이터 수집

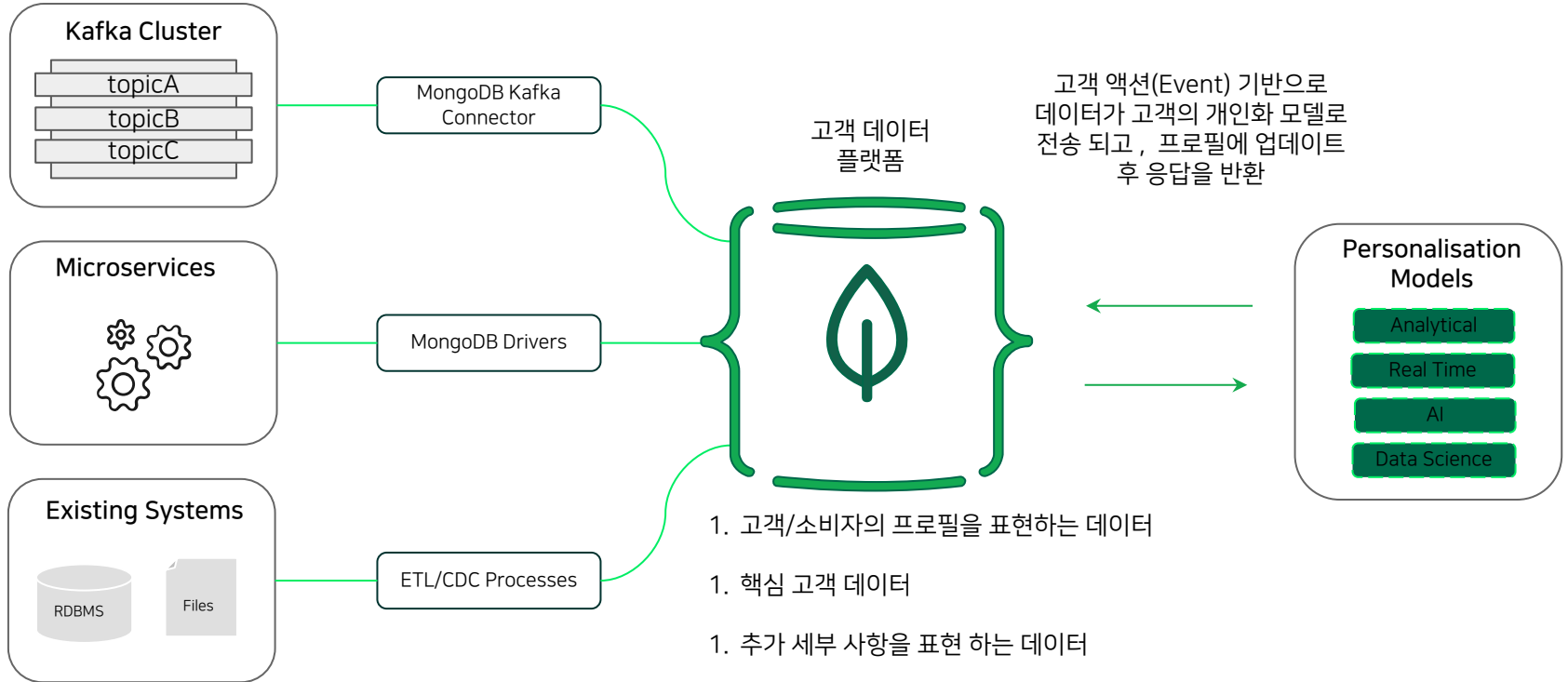


- 1. 고객/소비자의 프로필을 표현하는 데이터
- 1. 핵심 고객 데이터
- 1. 추가 세부 사항을 표현 하는 데이터



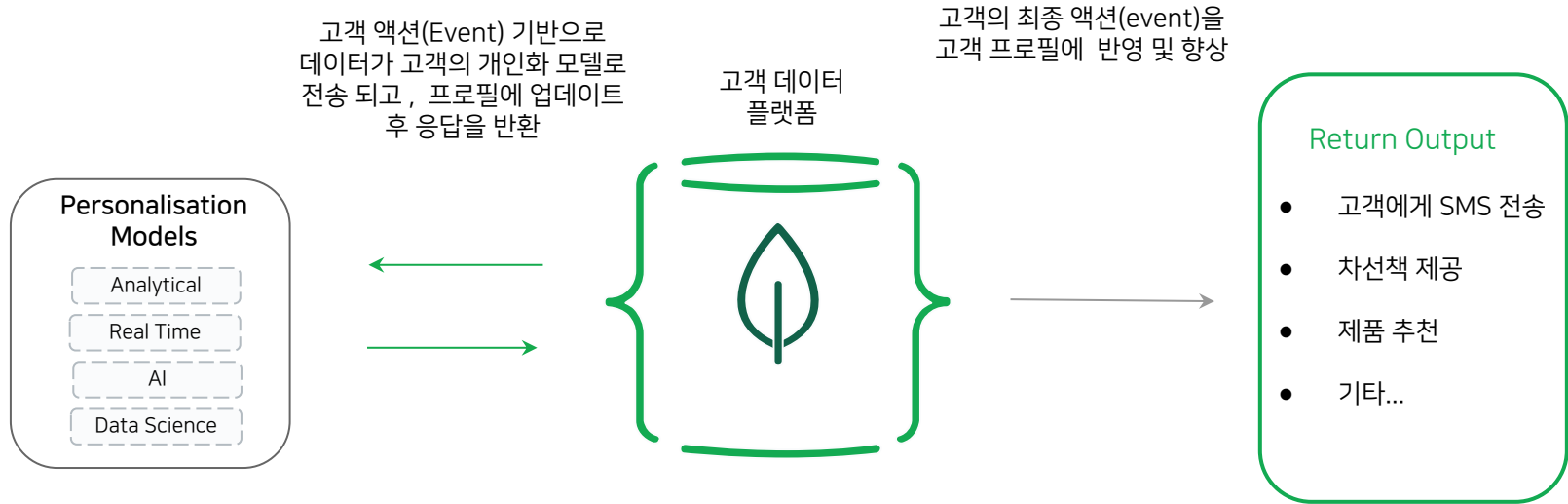
데이터 강화(Enrichment)

고객 프로필 생성을 위한 다양한 Data 소스로 데이터 수집





데이터 결과(Output)



Customer Data gets updated with trigger action.
For example:

"Last web visit" : 33.44.2021

"page" : "https://www.mongodb.com/"

개인화 요구 사항

- 고객 개인정보 보호
- 민첩한 변경 기능
- 실시간 개인화
- Dev & DevOps 친화적 구조





고객 개인정보 보호

개인화 서비스 개발에 고객 데이터를 완전히 보호 하는 것은 필수 아이템이다.

Authentication (인증)



- LDAP 인증 통합
- X.509
- SCRAM

Authorization (권한 관리)



- 룰 베이스 접근 권한 관리
- 필드 레벨 Masking
- Read Only 뷰

Auditing (감사 데이터)



- Fully configurable
- CRUD 감사
- Security
- Cluster

Encryption (암호화)



- TLS (In-flight)
- At Rest (database + volume)
- Field level
- Key management 통합

Certified (인증)



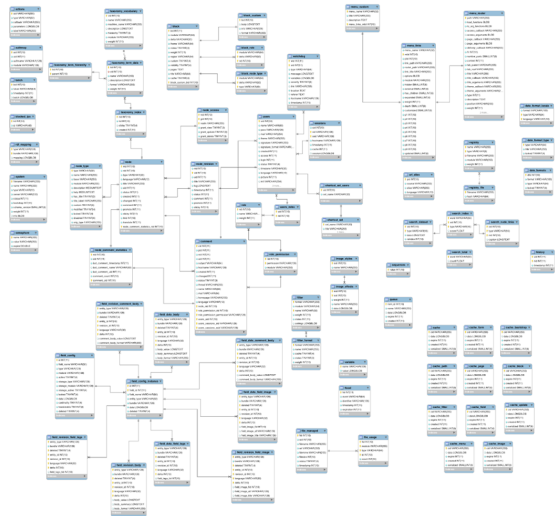
- ISO/IEC 27001
- PCI DSS
- SOC 2
- HIPPA



민첩한 변경 가능

도큐먼트 모델은 RDBMS의 복잡한 스키마 설계의 필요성을 제거함

Extract of a catalog...



The catalog...

```
{
  id: (id of product [text or integer]),
  name: (product name [text]),
  text: (product description [text]),
  locale: (locale (language/country) of product [text]),
  variants: [
    {
      id: (product variant id/SKU [text or integer]),
      name: (product variant name [text]),
      stock: (amount of stock available [integer]),
      size: (size [text]),
      color: (color [text]),
      display-color: (hexadecimal RGB color code, for example: "#000000"),
      display-thumbnail: (thumbnail image, instead of color code [url]),
      price: (price [integer, in cents]),
      currency: (currency [text]),
      pictures: [(picture url [url]), ...]
    },
    ...
  ]
}
```



민첩한 변경 가능

문서 모델은 신속한 적응을 위한 동적 스키마 변경을 지원함

```
{
  "_id" : ObjectId("5ad88534e3632e1a35a58d00"),
  "Player Info" : {
    "Display Name" : "coJack03",
    "PlayerUUID" : "2020-03-10-19-00-00-3467-2345",
    "Balance" : NumberDecimal("1421"),
    "ShopCardID" : "99990000123400009876"
  },
  "salesitem" : [
    { "stockname" : "Magic Hammer",
      "stockID" : "356-56789-12",
      "price" : NumberDecimal("34")},
    "quantity": 1 },
  ],
  + {...}
  ],
}
```

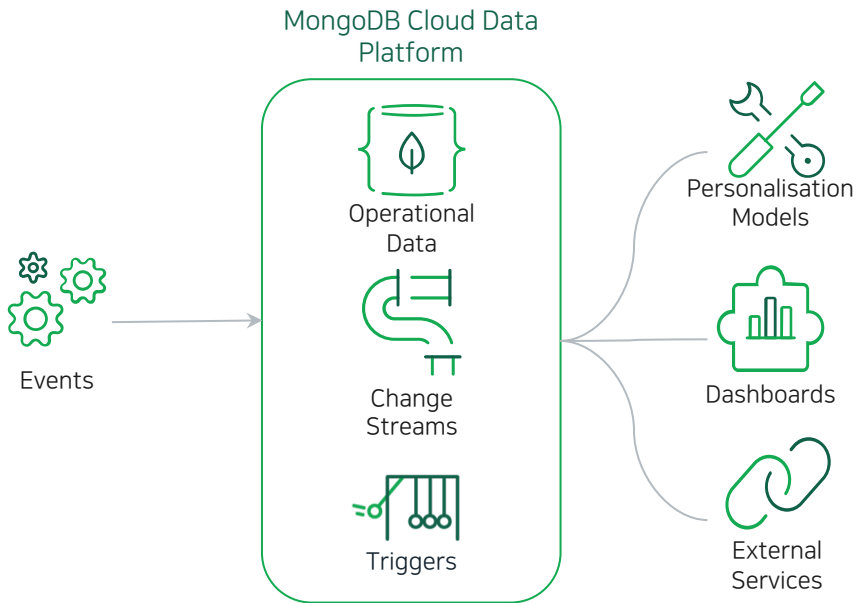
→ Add new fields dynamically at runtime

```
{
  "_id" : ObjectId("5ad88534e3632e1a35a58d00"),
  "Player Info" : {
    "Display Name" : "coJack03",
    "PlayerUUID" : "2020-03-10-19-00-00-3467-2345",
    "Balance" : NumberDecimal("1421"),
    "ShopCardID" : "99990000123400009876"
  },
  ],
  "salesitem" : [
    { "stockname" : "Magic Hammer",
      "stockID" : "356-56789-12",
      "price" : NumberDecimal("34")},
    "quantity": 1 },
  ],
  + {...}
  "geo" : { "type" : "Point", "coord" : [
    51.5065752,-0.109081]}},
  + {...}
  ],
}
```



실시간 개인화

실시간으로 외부 이벤트에 대응 가능



데이터를 소비하는 입장에서는 일괄적이고 정기적인 읽기에 의존하지 않고, 관심있는 데이터 이벤트를 구독하여 사용 할 수 있음

변경 스트림(change stream)을 통해 특정 이벤트에 대한 리스너 등록

데이터베이스 이벤트 (데이터 삭제, 변경, 등록..) 에 대한 응답을 사용자 정의 서버리스 기능을 통해 구성 할 수 있음



실시간 개인화

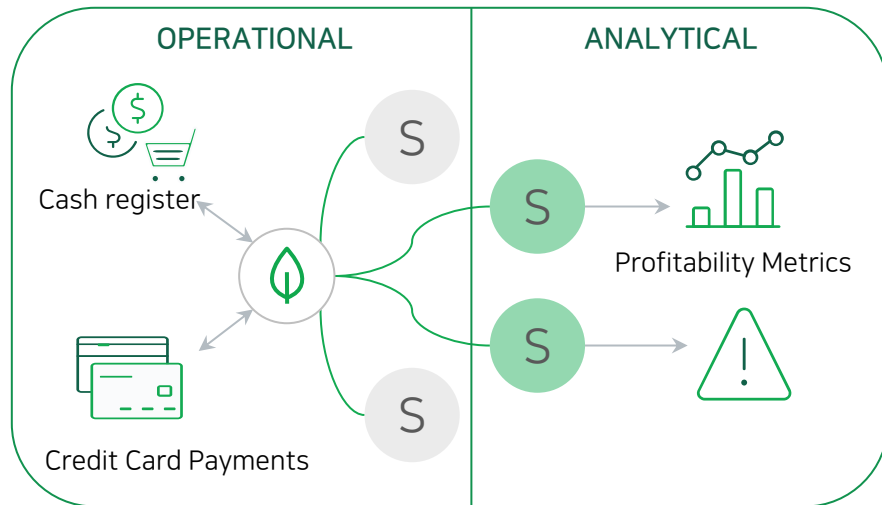
단일 플랫폼 내에서 다양한 워크로드를 활성화 하면 격리 보장

데이터는 기본(Primary)노드 에서
보조(Secondary) 노드로 실시간으로 복제

가용성을 높이는 것 외에도 이런 노드를 통해
특정 워크로드를 분리하여 사용 할 수 있음.

보조(Secondary)노드에 대한 쿼리는 다른
노드의 활동과 완전히 격리.

MongoDB Cluster 외부로 데이터의 이동/복제
가 필요 없음



A single replica set



Dev & DevOps 친화적 구조

고객 / 팀과 함께 혁신

혁신을 위한 인재 pool 관리

개발자가 생각하고 코딩하는 방식에 매핑
되는 문서 모델 사용

테이블/Key-Value/Columnar 모델등
다양한 데이터 모델을 포함하는 SuperSet
모델

24 by 7 , 365 무 중단 서비스 제공

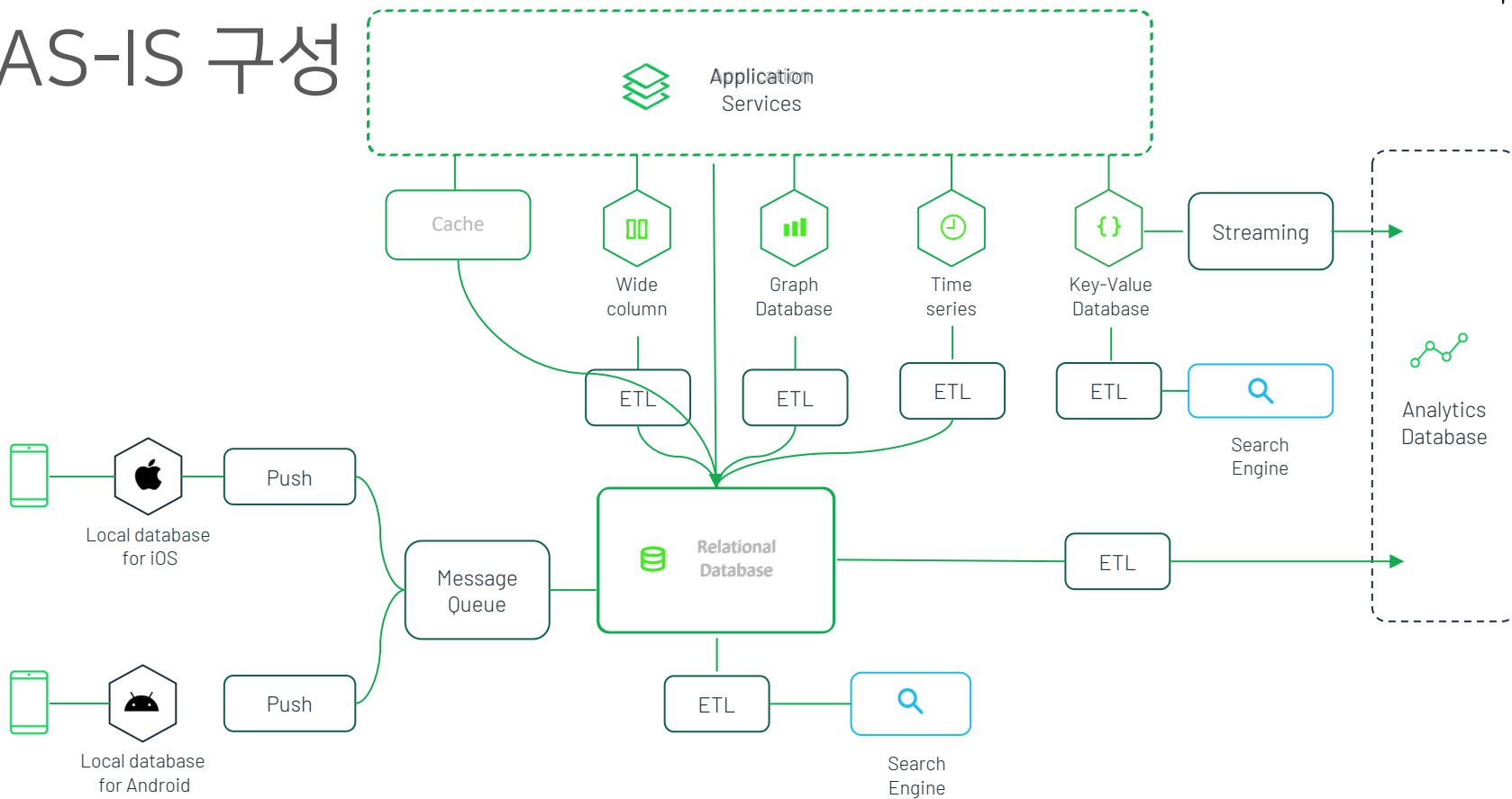


개인화 With MongoDB

- AS-IS 구성
- TO-BE 구성
- Why MongoDB
- Use Cases



AS-IS 구성





데이터 아키텍처의 복잡성으로 인한 혁신에 대한 세금 이 발생 — a **Data & Innovation Recurring Tax** (DIRT).

파편화된 개발자 경험

다중 운영환경하에서
합리적 보안 모델 구축

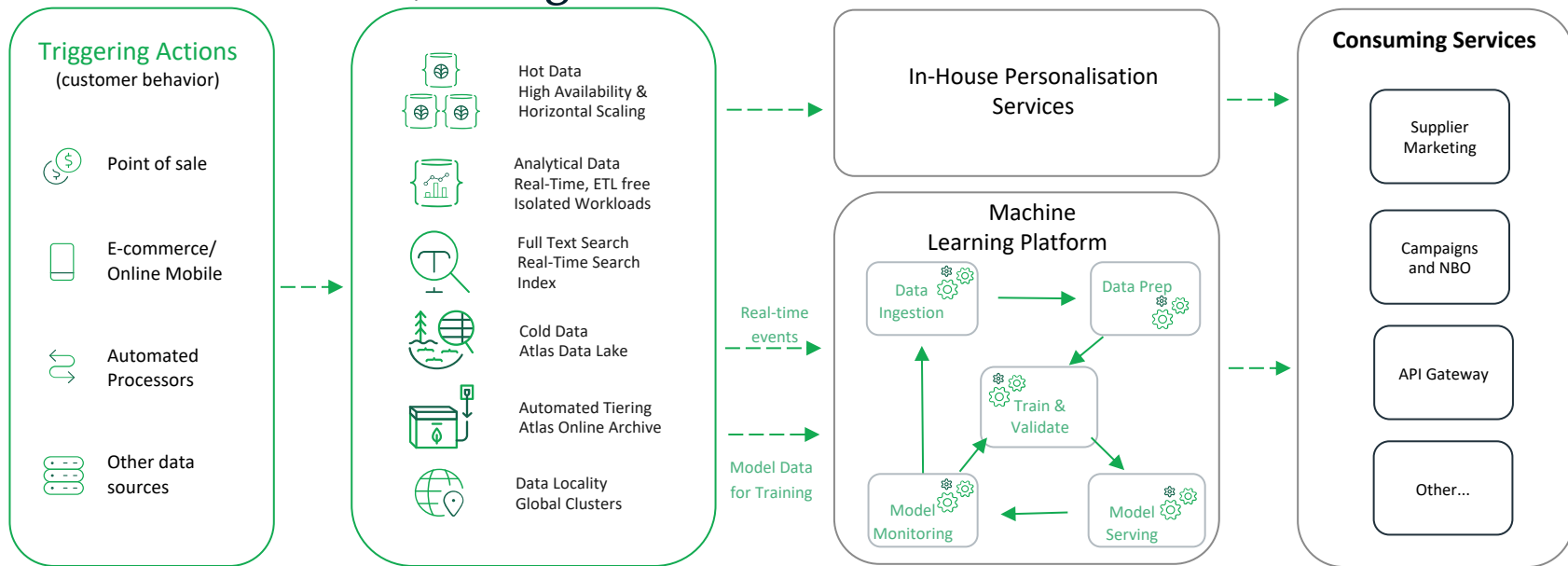
중요 데이터에 대한 통
합 노력 필수

불필요한 데이터 중복



T0-BE : MongoDB와 ML플랫폼 통합

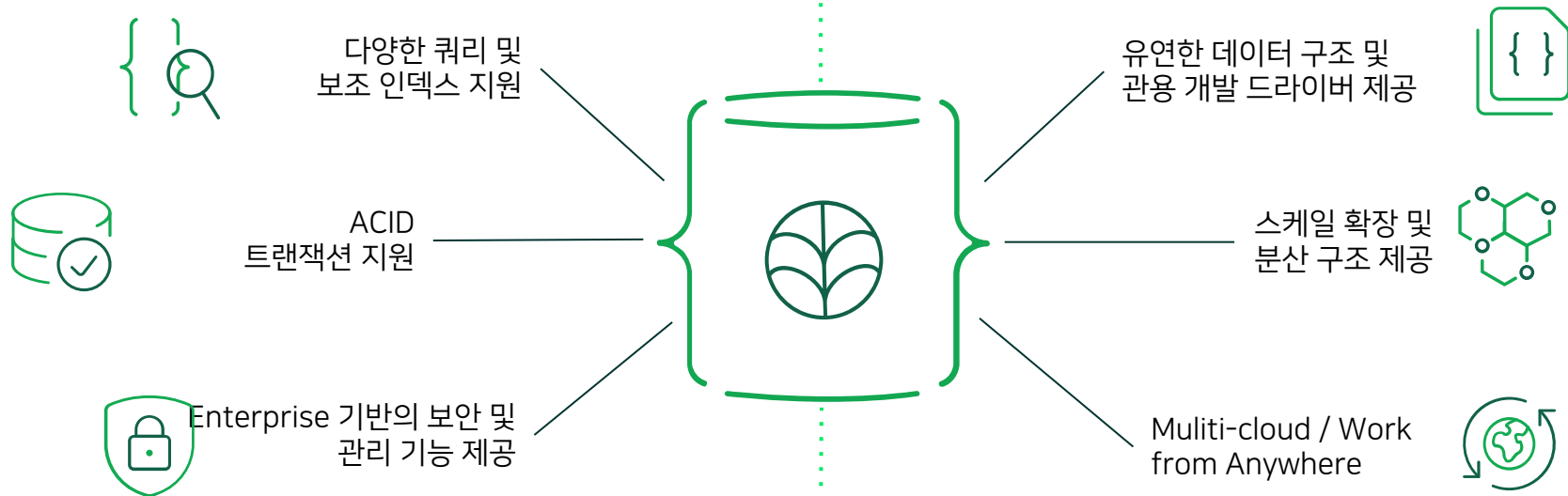
MongoDB 와 내부 ML 플랫폼과의 효과적 통합





Why MongoDB Works?

Mission Critical / 범용 데이터베이스 확장



전통적인 Relational 강점

추가적인 NoSQL & Big Data 강점/혁신 기능 제공



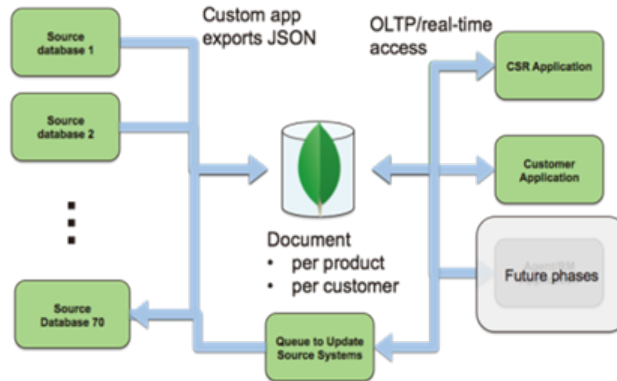
이슈 사항

145년 동안 70개 이상의
보험사를 인수 합병 하면서
동일한 고객의 정보가 서로
다르게 중복 저장 되어 있어
분석/활용에 제한

해결을 위해 2년 동안 2천
500만 달러를 투자하여 DB2로
고객 통합 프로젝트를 실행
하였지만 최종 실패

Solution

MongoDB의 유연한 데이터 모델을 활용하여 90일 만에 고객
DB통합 솔루션 'The Wall'을 구축



도입 효과

145년 이상의 약관, 청구, 고객
DB 정보(24TB)를 통합 관리,
분석하여 콜센터 담당자에게
실시간 정보 제공

고객 경험 향상

고객 이탈 감소

상품 판매 향상

교차 판매 향상





AIA HK - 차세대 빅데이터 분석 플랫폼 구축



이슈 사항

10개 이상의 Legacy 개별 시스템들의 데이터들이 산재

정형,비정형 데이터 통합 저장 불가능

실시간 Data 분석 요건 불 충족

느린 성능의 원인을 찾을 수 없었음

Solution

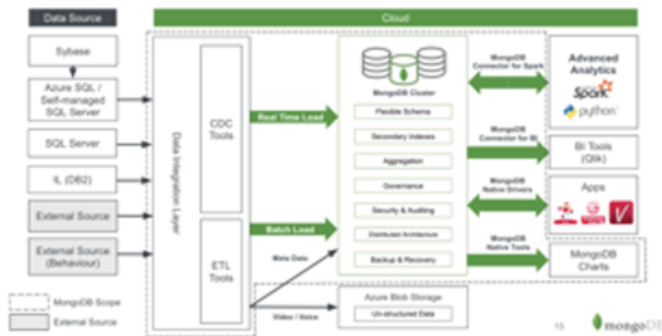
차세대 빅데이터 분석 플랫폼에 MongoDB를 도입
유연한 모델로 통합 저장 구현(정형,비정형,반정형)
실시간 Data 분석 구현

도입 효과

모든 LOB가 시기반 인사이트와 이벤트 트리거를 활용

개인 맞춤형 상품 판매, 서비스 조치, 제안 유도 실현

매출 증대 기여



10+
Legacy 데이터 소스를 통합

2 weeks
Prototype 개발



이슈 사항

110억 달러 이상의 자산관리를 위한 새로운 데이터베이스 필요

체계적인 투자거래 시장에서 보다 민첩하고 경쟁우위를 확보하기 위한 새로운 기술 모색

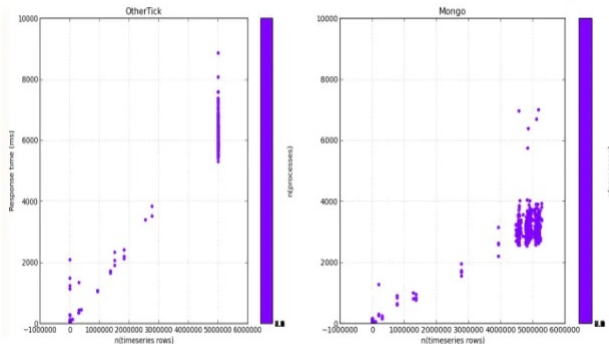
Legacy RDBMS는 고비용 및 경직된 환경 제약 극복

초당 2억 5천만건 이상의 틱 데이터 처리 요건

Solution

MongoDB 도입
유연한 데이터 모델, 확장성을 바탕으로 단일 플랫폼 구축
풍부한 쿼리 언어 표현, 보조 인덱스, 강력한 일관성을 바탕으로 마이그레이션 수행

Results - TickStore - 8 parallel



도입 효과

기존 Legacy 대비 100배 빠른 데이터 검색

초당 2억 5천만 건의 틱 데이터 처리

처리량 25배 증가

스토리지 60% 감소

40% 비용 절감



Thank you.

